Escuela de Administración de Tecnologías de Información

Formulación y análisis de viabilidad de un modelo predictivo de estudiantes en riesgo de deserción en el Tecnológico de Costa Rica

Trabajo Final de Graduación para optar al grado de Licenciatura en Administración de Tecnología de Información

Modalidad Seminario de Graduación

Elaborado por: Milton Andrey Valverde Paniagua

Prof. Tutor: Msc. Ing. Juan Andrés Segreda Johanning

Cartago, Costa Rica

II Semestre

Noviembre, 2024

Formulación y análisis de viabilidad de un modelo predictivo de estudiantes en riesgo de deserción en el Tecnológico de Costa Rica © 2024 por Milton Andrey Valverde Paniagua está sujeta a la licencia *Creative Commons Attribution-NonCommercial-ShareAlike* 4.0 *International*.

Para ver una copia de esta licencia, visite https://creativecommons.org/licenses/by-nc-sa/4.0/

Hoja de Aprobación

INSTITUTO TECNOLÓGICO DE COSTA RICA

ESCUELA DE ADMINISTRACIÓN DE TECNOLOGÍAS DE INFORMACIÓN

GRADO ACADÉMICO: LICENCIATURA

Los miembros del Tribunal Examinador de la Escuela de Administración de Tecnologías de Información, recomendamos que el siguiente informe del Trabajo Final de Graduación del estudiante Milton Andrey Valverde Paniagua sea aceptado como requisito parcial para obtener el grado académico de Licenciatura de Tecnología de Información.

Firmado digitalmente

Profesor Tutor

NESTOR ALEJANDRO MORALES RODRIGUEZ (FIRMA)

Firmado digitalmente por NESTOR ALEJANDRO MORALES RODRIGUEZ (FIRMA)
Fecha: 2024.11.28 11:22:14 -06'00'

Lector externo

FEDERICO TORRES CARBALLO (FIRMA)

Firmado digitalmente por FEDERICO TORRES CARBALLO (FIRMA)
Fecha: 2024.11.28 16:44:42 -06'00'

Lector académico

TEC | Tecnológico de Costa Rica

Firmado digitalmente por YARIMA TATIANA SANDOVAL SANCHEZ (FIRMA)
Fecha: 2024.12.02 18:05:09 -06'00'

Coordinadora de Trabajo Final de Graduación

Dedicatoria

Dedico este trabajo final de graduación a mi familia, por haberme apoyado para cumplir mi sueño, mi maestro Juan Andrés Segreda por su tutela y enseñanzas para mi crecimiento profesional, mis amigos Andrés Barahona, Andrés Gutiérrez y José Pablo Barrantes en su acompañamiento durante toda la carrera y sobre todo a mi pareja Geisel Montoya, por su cariño y apoyo incondicional en todos los momentos difíciles.

Resumen

Este proyecto evalúa la viabilidad de implementar un modelo predictivo de deserción estudiantil en el Tecnológico de Costa Rica (TEC). La investigación se desarrolló mediante una metodología mixta con diseño de investigación-acción, que incluyó análisis documental, entrevistas a profundidad con actores clave y evaluación de sistemas existentes. El estudio examina factores técnicos, financieros, organizacionales, legales y de competencia que influirían en la implementación exitosa del modelo.

Los resultados revelan que, si bien existe un modelo predictivo previo con una precisión del 91%, su discontinuidad se debió principalmente a factores organizacionales más que técnicos. La competitividad del proyecto se ve afectada por el desarrollo interno de una actualización del modelo anterior. El análisis financiero muestra un Valor Actual Neto (VAN) positivo de ₡325,476,841.27 en el escenario base, manteniéndose viable incluso en escenarios pesimistas. Asimismo, se identificaron 24 fuentes de datos potenciales, de las cuales 7 requieren desarrollo de Automatización Robótica de Procesos (RPA, por sus siglas en inglés) para su extracción.

Se concluye que el proyecto es técnica y financieramente viable; sin embargo, desde la perspectiva competitiva, el TEC podría no mostrar interés debido al proyecto en desarrollo interno. Por tanto, su éxito dependerá crucialmente de establecer directrices institucionales claras, asignar recursos dedicados para su mantenimiento y diferenciarse mediante la creación de necesidades específicas. Se propone un plan de implementación dividido en 6 proyectos específicos para ejecutar en 4 años, con un costo total estimado de ₡52,551,991.59.

Palabras clave: Deserción universitaria, Modelo predictivo, Análisis de viabilidad, Aprendizaje automático, Gestión académica

Abstract

This project evaluates the feasibility of implementing a predictive model of student dropout at the Tecnológico de Costa Rica (TEC). The research was developed using a mixed methodology with an action-research design, which included documentary analysis, in-depth interviews with key stakeholders, and evaluation of existing systems. The study examines technical, financial, organizational, legal and competitive factors that would influence the successful implementation of the model.

The results reveal that while there is a previous predictive model with an accuracy of 91%, its discontinuity was mainly due to organizational rather than technical factors. The competitiveness of the project is affected by the internal development of an update of the previous model. The financial analysis shows a positive Net Present Value (NPV) of ₡325,476,841.27 in the base scenario, remaining viable even in pessimistic scenarios. Also, 24 potential data sources were identified, of which 7 require Robotic Process Automation (RPA) development for extraction.

It is concluded that the project is technically and financially feasible; however, from a competitive perspective, TEC may not show interest due to the project's internal development. Therefore, its success will depend crucially on establishing clear institutional guidelines, allocating dedicated resources for its maintenance and differentiating itself by creating specific needs. An implementation plan is proposed divided into 6 specific projects to be executed in 4 years, with an estimated total cost of ₡52,551,991.59.

Keywords: University dropout, Predictive model, Feasibility analysis, Machine Learning, Academic management

Tabla de contenido

Páginas

Capítulo 1. Introducción 12
1.1. Descripción general 12
1.1.1. Situación problemática por analizar 12
1.1.2. Justificación del estudio y beneficios esperados 13
1.2. Objetivos del Trabajo Final de Graduación 16
1.2.1. Objetivo General 16
1.2.2. Objetivos Específicos 16
1.3. Alcance de la investigación 17
1.3.1. Entregables del proyecto 17
1.3.2. Supuestos del proyecto 18
1.3.3. Exclusiones del proyecto 18
1.3.4. Limitaciones del proyecto 19
Capítulo 2. Estado del arte 20
2.1. Evolución de los Estudios sobre Deserción Universitaria 20
2.2. Factores Asociados a la Deserción 20
2.3. Contexto Latinoamericano y Costarricense 21
2.4. La Situación en el Tecnológico de Costa Rica (TEC) 21
2.5. Modelos Predictivos y *Machine Learning* en la Educación Superior 21
2.6. Investigaciones Previas en el TEC 22
2.7. Desafíos en la Implementación de Modelos Predictivos 23
2.8. Marco Legal y Ético 23
2.8.1. Marco Legal Costarricense 24
2.8.2. Consideraciones Éticas 25
2.8.3. Perspectiva Internacional 27
2.8.4. Implicaciones para el Proyecto 27
2.9. Tendencias Futuras y Oportunidades 28
Capítulo 3. Marco Metodológico 29
3.1. Tipo de Investigación 29
3.2. Enfoque y diseño de la Investigación 29
3.3. Fuentes de Información 29
3.4. Sujetos de Información 31

3.5. Variables de Investigación ... 32
3.6. Procedimiento Metodológico de la Investigación ... 36
3.6.1. Diagrama propuesto para las fases del procedimiento metodológico ... 36
3.6.2. Cronograma ... 37
3.7. Operacionalización de las variables ... 38
3.8. Hipótesis de la investigación ... 40
3.9. Procedimiento Metodológico de la evaluación social ... 41
3.9.1. Contextualización del Proyecto en el Marco de la Inversión Pública ... 41
3.9.2. Análisis de las Metodologías de Evaluación Social ... 41
3.9.3. Justificación del Enfoque Metodológico Seleccionado ... 42
Capítulo 4. Resultados ... 43
4.1. Estado actual de modelos predictivos en Costa Rica. ... 43
4.1.1. Introducción al análisis del estado actual ... 43
4.1.2. Evolución histórica de los modelos predictivos en Costa Rica ... 43
4.1.3. Estado actual de los modelos predictivos ... 44
4.1.4. Análisis de características técnicas predominantes ... 44
4.1.5. Retos identificados ... 46
4.1.6. Oportunidades identificadas ... 48
4.2. Definición del producto ... 51
4.2.1. Descripción del proyecto propuesto ... 51
4.2.2. Alcance ... 52
4.2.3. Historia del proyecto ... 54
4.2.4. Contribuyentes ... 55
4.2.5. Estructura de gestión ... 56
4.2.6. Duración del proyecto ... 58
4.2.7. Plazos del proyecto ... 63
4.2.8. Plan de progreso ... 66
4.2.9. Herramientas de desarrollo ... 68
4.2.10. Presupuesto del proyecto ... 72
4.2.11. Beneficios no financieros ... 76
4.2.12. Criterios de éxito ... 78
4.2.13. Requerimientos legales y normativos ... 80
4.3. Estudio de mercado ... 80
4.3.1. Análisis de demanda ... 80

4.3.2. Análisis de oferta ... 85
4.3.3. Análisis del precio ... 86
4.4. Evaluación de la competencia ... 88
4.4.1. Descripción de la competencia ... 88
4.4.2. Fortalezas y debilidades de la competencia ... 89
4.4.3. Ventajas de mercado de la competencia ... 90
4.4.4. Estrategias de mercado de la competencia ... 91
4.5. Evaluación de riesgos ... 92
4.5.1. Identificación de riesgos ... 93
4.5.2. Análisis de riesgos ... 99
4.5.3. Valoración y respuesta al riesgo ... 101
4.5.4. Plan de monitoreo y control de riesgos ... 104
4.5.5. Riesgos residuales y secundarios ... 104
4.6. Evaluación social ... 105
4.6.1. Evaluación económica del proyecto social ... 105
4.6.2. Evaluación de Fortalecimiento de Capacidades Institucionales ... 113
4.6.3. Análisis de Beneficios Sociales ... 118
4.6.4. Evaluación Integral de Impacto Social ... 121
4.7. Plan de implementación sugerido del proyecto ... 122
4.7.1. Desglose de proyectos ... 122
4.7.2. Hoja de ruta detallada ... 124
4.7.3. Métricas de éxito y evaluación ... 124
4.8. Comprobación de hipótesis ... 125
4.8.1. Verificación de H1: La utilidad práctica del modelo predictivo ... 125
4.8.2. Verificación de H2: Factores críticos más allá de lo técnico ... 126
4.8.3. Verificación de H3: Efectividad de la predicción temprana ... 126
4.8.4. Verificación de H4: Necesidad institucional del TEC ... 127
4.8.5. Verificación de H5: Necesidad de un modelo predictivo ... 128
Capítulo 5. Limitaciones y problemas encontrados ... 129
Capítulo 6. Discusión y Conclusiones ... 131
6.1. Discusión ... 131
6.2. Conclusiones ... 132
Capítulo 7. Cumplimiento de objetivos ... 135
Capítulo 8. Recomendaciones para futuras investigaciones ... 137

Capítulo 9. Referencias 139
Capítulo 10. Apéndices 143
Apéndice A: Árbol del problema 143
Apéndice B: Plantilla de minuta 144
Apéndice C: Plantilla de gestión del cambio 145
Apéndice D: Plantilla de Entrevista Semiestructurada sobre Deserción Estudiantil en el TEC 146
Apéndice E: Plantilla de Entrevista Semiestructurada Abogado 147
Apéndice F: Tabla de hallazgos de entrevista a M.Ed. Teresa Hernández Jiménez. 149
Apéndice G: Tabla de hallazgos de entrevista al Profesor Martín Solís Salazar. 152
Apéndice H: Tabla de hallazgos de entrevista a Evaluadora e investigadora jubilada Tania Elena Moreira Mora. 155
Apéndice I: Tabla de hallazgos de entrevista a Arlyn Angulo Hernández y Raquel Mejías Elizondo de la OPI. 158
Apéndice J: Diagrama de Gantt de proyecto SPGDE 161
Apéndice K: Tabla de requerimientos legales del proyecto 162
Apéndice L: Prescripción de entrevista con M.Ed. Teresa Hernández Jiménez, Asesora Psicoeducativa del DOP. 163
Apéndice M: Prescripción de entrevista con el Estadístico, Investigador y Profesor Martín Solís Salazar de la Escuela de Administración de Empresas. 172
Apéndice N: Prescripción de entrevista con la Evaluadora e investigadora jubilada Tania Elena Moreira Mora. 181
Apéndice O: Prescripción de entrevista con Arlyn Angulo Hernández y Raquel Mejías Elizondo, encargadas de la deserción en la OPI. 191
Apéndice P: Proyectos sugeridos detallados. 198
Apéndice Q: Tabla de variables de predicción de deserción estudiantil encontradas. 209
Apéndice R: Tabla de variables de predicción de deserción estudiantil sugeridas. 212
Apéndice S: Minuta 1 de entrevista a M.Ed. Teresa Hernández Jiménez. 220
Apéndice T: Minuta de entrevista al Profesor Martín Solís Salazar. 221
Apéndice U: Minuta de entrevista a Evaluadora e investigadora jubilada Tania Elena Moreira Mora. 222
Apéndice V: Minuta de entrevista a Arlyn Angulo Hernández y Raquel Mejías Elizondo de la OPI. 223
Apéndice W: Minuta 2 de entrevista a M.Ed. Teresa Hernández Jiménez 224
Apéndice X: Minuta de entrevista a M.Sc. Ing. Juan Andrés Segreda Johanning. 225
Apéndice Y: Solicitud de información salarial TEC. 226
Apéndice Z: Solicitud de datos sobre aporte y costo por estudiante al TEC. 226
Apéndice AA: Minuta 1 con tutor. 227

Apéndice BB: Minuta 2 con tutor. 228
Apéndice CC: Minuta 3 con tutor. 229
Apéndice DD: Minuta 4 con tutor. 230
Apéndice EE: Minuta 5 con tutor. 231
Apéndice FF: Minuta 6 con tutor. 232
Apéndice GG: Minuta 7 con tutor. 233
Apéndice HH: Minuta 8 con tutor. 234
Apéndice II: Minuta 9 con tutor. 235
Apéndice JJ: Minuta 10 con tutor. 236
Apéndice KK Minuta 11 con tutor. 237
Apéndice LL: Minuta 12 con tutor. 238
Capítulo 11. Anexos 239
Anexo I. Desglose salarial Técnico en Tecnologías de Información y comunicación. 239

Índice de Figuras

Figura No.	Descripción	Página
Figura 1	Diagrama de fases del procedimiento metodológico.	36
Figura 2	Cronograma de la investigación.	37
Figura 3	Posicionamiento de mercado de los competidores.	92
Figura 4	Mejora en capacidades tecnológicas.	115
Figura 5	Ciclo de mejora operativa del SPGDE.	117
Figura 6	Beneficios institucionales.	119
Figura 7	Optimización de Recursos Públicos.	121
Figura 8	Hoja de ruta propuesta.	124

Índice de Tablas

Tabla No.	Descripción	Página
Tabla 1	Entregables del producto según objetivo.	17
Tabla 2	Fuentes primarias de la investigación.	30
Tabla 3	Fuentes secundarias de la investigación	30
Tabla 4	Sujetos de investigación del proyecto.	31
Tabla 5	Variables de investigación del objetivo específico 1.	33
Tabla 6	Variables de investigación del objetivo específico 2.	33
Tabla 7	Variables de investigación del objetivo específico 3.	34
Tabla 8	Variables de investigación del objetivo específico 4.	35
Tabla 9	Descripción de fases del procedimiento metodológico.	36
Tabla 10	Hipótesis de la investigación.	40
Tabla 11	Análisis Comparativo de Metodologías de Evaluación Social.	41
Tabla 12	Contribuyentes del proyecto.	55
Tabla 13	Fuentes de información encontradas.	59
Tabla 14	Fases y actividades del proyecto.	63
Tabla 15	Hitos y entregables del proyecto.	67
Tabla 16	Cuadro de características de software de minería de datos.	69
Tabla 17	Comparativa de software de minería de datos.	70
Tabla 18	Cuadro de características de software de visualización.	70
Tabla 19	Comparativa de software de visualización.	71
Tabla 20	Costo total de etapa ETL.	73
Tabla 21	Resumen de costos totales del proyecto.	75
Tabla 22	Criterios de éxito del proyecto SPGDE.	79
Tabla 23	Comparación detallada de competidores.	89
Tabla 24	Fortalezas y debilidades de competencia.	90
Tabla 25	Activos clave del proyecto	93
Tabla 26	Amenazas del proyecto.	94
Tabla 27	Vulnerabilidades del proyecto.	95
Tabla 28	Consecuencias de riesgos del proyecto.	97
Tabla 29	Riesgos asociados al proyecto.	98

Tabla 30 Matriz de Probabilidad e Impacto de Riesgos. 100
Tabla 31 Estrategias de Respuesta a Riesgos. 101
Tabla 32 Precios sombra del proyecto. 106
Tabla 33 Resumen de Costos Sociales del SPGDE (en colones). 109
Tabla 34 Dimensiones de Evaluación de Capacidades Institucionales. 114
Tabla 35 Diagnóstico de Capacidad Operativa Actual. 116
Tabla 36 Indicadores de Desempeño para Evaluación de Capacidades. 117
Tabla 37 Marco de Análisis de Beneficios Sociales. 118
Tabla 38 Análisis de Beneficios para la Comunidad Estudiantil. 120
Tabla 39 Alineación con Objetivos del Sistema de Educación Superior. 120
Tabla 40 Marco de Evaluación Integral de Impacto Social. 121
Tabla 41 Proyectos sugeridos 122
Tabla 42 Cumplimiento de objetivos. 135

Glosario

Término	Definición
AdaBoost	Algoritmo de aprendizaje automático que combina múltiples clasificadores débiles para crear un clasificador fuerte, mejorando iterativamente el rendimiento del modelo.
Algoritmo	Conjunto ordenado y finito de operaciones que permite hallar la solución a un problema específico.
Análisis Costo-Beneficio (ACB)	Metodología sistemática que compara los costos y beneficios monetarios de un proyecto para evaluar su viabilidad económica.
Análisis Costo-Efectividad (ACE)	Metodología que evalúa la relación entre los costos de un proyecto y sus resultados o efectos, sin necesariamente monetizar los beneficios.
Análisis de Sensibilidad	Estudio que determina cómo diferentes valores de una variable independiente afectan a una variable dependiente bajo un conjunto de supuestos.
Análisis Exploratorio de Datos	Proceso de examinar bases de datos para encontrar patrones, anomalías y probar hipótesis iniciales.
Análisis Predictivo	Conjunto de técnicas estadísticas y de minería de datos que analizan datos actuales e históricos para hacer predicciones sobre eventos futuros.
API (*Application Programming Interface*)	Conjunto de reglas y especificaciones que permiten que diferentes aplicaciones de software se comuniquen entre sí.
Aprendizaje Federado	Técnica de *machine learning* que entrena algoritmos en datos descentralizados sin compartirlos directamente.
ATI	Siglas de Administración de Tecnología de Información, carrera impartida en el Tecnológico de Costa Rica.
Bagging (Bootstrap Aggregating)	Técnica de aprendizaje automático que crea múltiples modelos utilizando diferentes muestras aleatorias de los datos para mejorar la estabilidad y precisión.
Bases de Datos	Conjunto organizado de datos almacenados sistemáticamente para su posterior uso.
Business Intelligence	Conjunto de estrategias y herramientas enfocadas a la administración y creación de conocimiento mediante el análisis de datos existentes.
Ciencia de Datos	Campo interdisciplinario que utiliza métodos científicos, procesos y sistemas para extraer conocimiento de datos.
CONARE	Consejo Nacional de Rectores, órgano coordinador de la educación superior universitaria estatal en Costa Rica.

Término	Definición
Cross-validation	Técnica para evaluar los resultados de un análisis estadístico y garantizar que son independientes de la partición entre datos de entrenamiento y prueba.
Cubo de Información	Estructura de datos multidimensional que permite el análisis de información desde diferentes perspectivas.
Dashboard	Panel de control visual que presenta información clave de manera condensada y fácil de interpretar.
Data Mining	Proceso de detectar información relevante de grandes conjuntos de datos.
Data Warehouse	Base de datos corporativa que se caracteriza por integrar y depurar información de diferentes fuentes.
DATIC	Departamento de Administración de Tecnologías de Información y Comunicaciones del TEC.
Deep Learning	Subcampo del *machine learning* basado en algoritmos que intentan modelar abstracciones de alto nivel usando arquitecturas compuestas de transformaciones no lineales múltiples.
Deserción Estudiantil	Abandono voluntario o forzoso del programa académico en el que se ha matriculado un estudiante, por un período igual o mayor a dos semestres consecutivos.
Deserción de Primer Ingreso	Abandono de estudios por parte de estudiantes durante su primer año en la institución.
DOP	Departamento de Orientación y Psicología del TEC.
Enfoque Mixto	Metodología de investigación que combina métodos cuantitativos y cualitativos.
ETL (*Extract, Transform, Load*)	Proceso que permite a las organizaciones mover datos desde múltiples fuentes, reformatearlos y limpiarlos, y cargarlos en otra base de datos.
Evaluación de Impacto	Análisis sistemático de los cambios duraderos o significativos en la vida de las personas producidos por una acción específica.
Evaluación de Viabilidad	Análisis que determina si un proyecto es factible desde diferentes perspectivas (técnica, económica, legal, etc.).
FEES	Fondo Especial para la Educación Superior, mecanismo de financiamiento para las universidades públicas de Costa Rica.
Gobernanza de Datos	Marco para la gestión de la disponibilidad, usabilidad, integridad y seguridad de los datos.
Hoja de Ruta	Plan detallado que establece la secuencia de pasos necesarios para alcanzar un objetivo.

Término	Definición
IA (Inteligencia Artificial)	Campo de la informática que busca crear sistemas capaces de aprender y resolver problemas de manera similar a los humanos.
IntegraTEC	Programa de mentoría para estudiantes de primer ingreso del TEC.
Interpretabilidad del Modelo	Grado en que un humano puede entender la causa de las decisiones de un modelo predictivo.
KPI (*Key Performance Indicator*)	Indicador clave de rendimiento utilizado para medir el éxito en el logro de objetivos organizacionales.
Machine Learning	Campo de la inteligencia artificial que permite a los sistemas aprender y mejorar automáticamente a partir de la experiencia sin ser programados explícitamente.
Matriz de Riesgos	Herramienta de control y gestión que permite identificar y cuantificar riesgos en base a su probabilidad e impacto.
Metadatos	Datos que describen otros datos, proporcionando información sobre el contenido, calidad, condición y otras características.
Microservicios	Arquitectura que estructura una aplicación como un conjunto de servicios pequeños e independientes.
Modelo Predictivo	Sistema que utiliza datos históricos y actuales para identificar patrones y predecir valores futuros.
OPI	Oficina de Planificación Institucional del TEC.
Overfitting	Sobreajuste que ocurre cuando un modelo se ajusta demasiado a los datos de entrenamiento y pierde capacidad de generalización.
PAO	Plan Anual Operativo, documento que establece los objetivos y metas anuales de una institución.
PEI	Plan Estratégico Institucional, documento que establece la dirección estratégica de una institución.
***Pipeline* de Datos**	Conjunto de elementos de procesamiento de datos conectados en serie, donde la salida de un elemento es la entrada del siguiente.
Power BI	Herramienta de visualización de datos y *business intelligence* desarrollada por Microsoft.
Precisión del Modelo	Medida que indica qué tan exactas son las predicciones de un modelo.
Random Forest	Algoritmo de aprendizaje automático que construye múltiples árboles de decisión y los combina para obtener una predicción más precisa y estable.
Redes Neuronales	Modelo computacional basado en la estructura y funcionamiento del cerebro humano.

Término	Definición
Regresión Logística	Modelo estadístico usado para predecir la probabilidad de un evento binario.
Retención Estudiantil	Capacidad de una institución educativa para mantener inscritos a sus estudiantes desde su admisión hasta su graduación.
RN (Reprobación Continua)	Estado en el que un estudiante reprueba un curso en múltiples ocasiones.
ROI (Retorno de Inversión)	Medida del beneficio obtenido en relación con la inversión realizada.
RPA (*Robotic Process Automation*)	Tecnología que permite configurar software para automatizar tareas repetitivas basadas en reglas.
SIGI	Sistema de Indicadores de Gestión Institucional del TEC.
SPGDE	Sistema de Predicción y Gestión de la Deserción Estudiantil, proyecto propuesto en este trabajo.
Stakeholder	Parte interesada o afectada por las actividades y decisiones de una organización.
TEC	Tecnológico de Costa Rica, universidad pública especializada en tecnología e ingeniería.
TEC Digital	Plataforma de aprendizaje en línea utilizada por el Tecnológico de Costa Rica.
Underfitting	Sub-ajuste que ocurre cuando un modelo es demasiado simple para capturar la complejidad de los datos.
VAN	Valor Actual Neto, indicador financiero que mide la viabilidad de un proyecto considerando el valor del dinero en el tiempo.
Variables Predictoras	Características o atributos utilizados para realizar predicciones en un modelo.
VIE	Vicerrectoría de Investigación y Extensión del TEC.
VIESA	Vicerrectoría de Vida Estudiantil y Servicios Académicos del TEC.
Visualización de Datos	Representación gráfica de información y datos mediante elementos visuales como gráficos, mapas y *dashboards*.
Workflow	Flujo de trabajo o secuencia de pasos necesarios para completar una tarea.

Capítulo 1. Introducción

1.1. Descripción general

En esta sección se describe la situación problemática hallada dentro del entorno de la organización, el cual motiva el desarrollo del proyecto, así como la mención de los beneficios esperados del producto.

1.1.1. Situación problemática por analizar

La deserción universitaria representa un problema documentado en la educación superior, como evidencian los estudios de Chinchilla-Brenes (2016), impactando, no solo la trayectoria académica y profesional de los estudiantes, sino también la efectividad institucional y el desarrollo socioeconómico del país. En el Tecnológico de Costa Rica (TEC), este fenómeno ha sido identificado como una preocupación, con tasas de deserción que reflejan un problema sistémico necesitado de intervención.

Según estudios del TEC, una proporción significativa de estudiantes abandona sus estudios antes de completar su carrera. Por ejemplo, un estudio de Chinchilla-Brenes (2016) sobre la cohorte 2012 reveló que más del 52.7% de los estudiantes que abandonaron sus estudios lo hicieron de manera definitiva, citando razones como la falta de interés en la carrera elegida y dificultades económicas. Estos datos ponen de manifiesto la necesidad de abordar las causas subyacentes de la deserción para mejorar los índices de retención y graduación.

Por otra parte, en los estudios mencionados anteriormente, se indica que los factores que contribuyen a la deserción en el TEC son multifacéticos, incluyendo factores académicos, personales, institucionales y externos; consecuentemente provocando un impacto tanto en los estudiantes y la institución, como en la sociedad.

La capacidad del TEC para identificar a estudiantes en riesgo de deserción es fundamental para mejorar la retención estudiantil y optimizar los recursos educativos. Sin embargo, la institución enfrenta desafíos significativos debido a la insuficiente integración y utilización de modelos predictivos avanzados. Esta situación se agrava por varios factores que impiden el análisis efectivo y la intervención temprana.

Uno de los principales obstáculos es la falta de herramientas analíticas adecuadas. A pesar de contar con sistemas de información, estos no están desarrollados ni integrados para realizar análisis predictivos asertivos. Según Solís et al. (2018), la capacidad de integrar y analizar datos desde una perspectiva holística es crucial para las capacidades predictivas.

Además, los datos en el TEC están mal estructurados para un análisis predictivo eficiente, como lo indican Fernández-Martín et al. (2018). Los sistemas de información son fragmentados y las políticas de gestión de datos no facilitan una integración adecuada. Esta falta de una estructuración adecuada de datos impide la creación de un sistema unificado que podría utilizar la información disponible para identificar proactivamente a los estudiantes en riesgo.

La colaboración entre departamentos también es limitada, lo que impide la compartición efectiva de información relevante. Las estructuras organizativas del TEC no promueven una integración interdepartamental, y una cultura organizacional restrictiva, limita la colaboración y el intercambio de

información. Esto resulta en una capacidad reducida para implementar un enfoque integral y cohesivo para la identificación de riesgos de deserción.

Estos desafíos tienen múltiples consecuencias negativas. Sin una identificación temprana de los estudiantes en riesgo, las intervenciones no pueden ser efectivamente dirigidas, lo que resulta en un uso ineficiente de los recursos y una menor efectividad de los programas de apoyo estudiantil. Además, el aumento en las tasas de deserción afecta negativamente la reputación del TEC y sus indicadores de éxito junto a los fondos financieros institucionales por tanto fondos públicos, como detalla un estudio descriptivo de la deserción en el TEC (Hernández-Jiménez et al., 2019).

Este Trabajo Final de Graduación busca abordar estas deficiencias evaluando la viabilidad de implementar un modelo predictivo en el TEC. Al hacerlo, no solo se espera mejorar las estrategias de intervención y la administración de recursos educativos, sino también contribuir al éxito académico y al bienestar general de la comunidad estudiantil. Por tanto, bajo esta problemática, el entorno y contexto actual del Tecnológico de Costa Rica, esta investigación se plantea la siguiente pregunta de investigación: ¿Es viable desarrollar e implementar un modelo predictivo que permita identificar a los estudiantes en riesgo de deserción en el Tecnológico de Costa Rica, considerando los aspectos técnicos, financieros, legales y organizacionales?

Para obtener una perspectiva más amplia del problema central, sus causas, consecuencias y relaciones, puede verse el Apéndice A: Árbol del problema.

1.1.2. Justificación del estudio y beneficios esperados

En esta sección se justificará el proyecto desde una perspectiva de holística de un ATI, detallando la razón de porqué es apto para ser realizado por un ATI, además su alineación con las áreas de investigación del Trabajo Final de Graduación definido por la Escuela de Administración de Tecnologías de Información y su relación con los objetivos de desarrollo sostenible establecidos por la ONU.

Siguiendo lo anterior, este proyecto se adecua al área de investigación “Gestión de datos e información” y “Administración de proyectos de tecnología de información” Administración de Tecnología de Información (2020). Ya que este tiene como objetivo la evaluación y formulación de un proyecto de tecnología emergente, manteniendo una estrecha relación con la gestión la tecnología apropiada para lograr la meta final y mantenida por la institución, entre algunos elementos adicionales como políticas, estructura de gestión, entre otros; sin embargo, también tiene una estrecha relación con los datos e información por la naturaleza del proyecto, el cual, implica como estos se encuentran en la institución.

Como bien ya se mencionó en secciones anteriores, el desarrollo de un modelo predictivo del riesgo de deserción en estudiantes, tiene ciertos beneficios potenciales tanto directos como indirectos que pueden fortalecer diversas áreas de la institución con un impacto positivo sobre la reputación, las finanzas como asignación de recursos, información relevante para apoyar a la meta de reducir la deserción estudiantil y el establecimiento de un centro de investigaciones aportar proyectos de valor y de interés institucional.

Por otro lado, estos beneficios son muy generales manteniendo como supuesto un buen desarrollo, debido a su falta de claridad en el entorno que rodea a este proyecto, es sumamente necesario realizar ciertos estudios que puedan validar si es o no factible seguir adelante con el desarrollo de iniciativas para completar el modelo predictivo; este es el punto focal de la investigación e importancia de evaluar la mayor cantidad de elementos

posibles para determinar si invertir o no en el proyecto podría ser una buena decisión; para ello, la investigación realizada es fundamental, ya que provee información de gran valor que pueda apoyar esta toma de decisiones bajo fuentes documentadas y fiables.

Adicionalmente, en relación con los con Ejes de Conocimiento Estratégicos para el periodo comprendido de 2023 a 2032, los cuales según TEC (2023) "son áreas del conocimiento y objetos de estudio a través de los cuales la institución pretende lograr su misión, enfocado en ellos prioritariamente sus actividades y recursos", se destaca el eje "Educación", el cual se considera como uno que cumple el proyecto al generar una información para una acción social y evaluación de la educación junto a otros aspectos relacionados para apoyar el derecho humano a una educación de calidad, mejorando las oportunidades educativas o de aprendizaje.

Por último, con respecto a los objetivos de desarrollo sostenible de la ONU, siguiendo esta misma línea, se considera que el proyecto destaca en el objetivo "Educación de calidad" que según la ONU (2015):

> El objetivo es proporcionar igualdad de acceso a una formación profesional asequible, eliminar las disparidades de género y riqueza y lograr el acceso universal a una educación superior de calidad. La educación es la clave que permitirá alcanzar muchos otros objetivos de desarrollo sostenible (ODS). Cuando las personas pueden obtener una educación de calidad, pueden romper el ciclo de la pobreza. (sección "4 Educación de Calidad", párr. 1)

Bajo esta premisa, este proyecto apoya de forma indirecta al formular y evaluar el modelo predictivo, este ODS, ya que determina muchos detalles del futuro proyecto que alimentará de forma directa este objetivo, afectando las siguientes metas según la ONU (2015):

- Meta 4.3: Hasta 2030, asegurar que todos los hombres y mujeres tengan igual acceso a una educación técnica, profesional y terciaria de calidad, incluida la universitaria.
- Meta 4.5: Eliminar las disparidades de género en la educación y asegurar el acceso igualitario a todos los niveles de educación y formación profesional para los vulnerables, incluidas las personas con discapacidades, los pueblos indígenas y los niños en situaciones de vulnerabilidad.

El apoyo a estas metas radica en intentar reducir la deserción universitaria, lo que directamente contribuye a aumentar el acceso y la equidad en la educación superior. Al identificar tanto proactiva como reactivamente a los estudiantes en riesgo, el modelo predictivo puede facilitar intervenciones tempranas que ayuden a mantener a estos estudiantes en el sistema educativo, promoviendo así una educación inclusiva y equitativa. Además, mejorar las tasas de retención puede ayudar a crear un entorno académico más inclusivo y de apoyo, especialmente para grupos vulnerables que podrían estar en mayor riesgo de deserción.

En síntesis, a continuación, se muestran los beneficios directos e indirectos obtenidos al realizar este proyecto en el Tecnológico de Costa Rica.

1.1.2.1. Beneficios directos

- Determinación de viabilidad del proyecto: Este proyecto permite tomar la decisión de invertir o no en esta iniciativa, ya que proveerá información de concluyente de calidad en la evaluación de diferentes aristas que puedan tomarse en consideración y agreguen un valor en esta decisión final.

- Identificación de fuentes de datos actuales: Al mantener una exploración en las principales fuentes de datos del Tecnológico, se establece una visualización global de toda la información disponible dentro y fuera de la institución, lo que permite determinar si se requieren nuevas fuentes.
- Identificación de información relevante para intervenciones tempranas: El proyecto proporcionará un primer acercamiento a datos críticos que permitirán al personal identificar a los estudiantes en riesgo antes de que alcancen un punto crítico, lo que puede ser un insumo importante para el desarrollo de un futuro modelo.
- Aumento en comprensión de Factores de Riesgo: Al analizar diversos factores académicos, sociodemográficos y financieros, el modelo ayudará a comprender mejor las causas subyacentes de la deserción, permitiendo al TEC diseñar programas de apoyo más efectivos.
- Evaluación de Políticas y Programas Actuales: La evaluación del modelo permitirá verificar la efectividad de las políticas y programas de retención actuales, proporcionando una base empírica para ajustes y mejoras.

1.1.2.2. Beneficios indirectos

- Fortalecimiento del Prestigio Académico: Al mejorar la capacidad de retención y apoyo estudiantil, el TEC puede fortalecer su reputación como una institución comprometida con el éxito de sus estudiantes.
- Impulso a la Innovación y Liderazgo Académico: El proyecto destacará al TEC como líder en la adopción de tecnologías analíticas avanzadas en el ámbito educativo, sirviendo de modelo para otras instituciones.
- Aumento de la Satisfacción y la Retención Estudiantil: Al sentirse más apoyados y comprendidos, es probable que los estudiantes tengan una experiencia más positiva, lo que puede traducirse en mayores tasas de satisfacción y retención.
- Optimización de Recursos Institucionales: Con una herramienta predictiva, el TEC puede optimizar la asignación de recursos a los programas que demuestren ser más efectivos, evitando gastos en iniciativas menos efectivas.
- Desarrollo Profesional del Personal: La participación en la implementación y uso del modelo ofrece oportunidades de desarrollo profesional para el personal, mejorando sus habilidades en análisis de datos y toma de decisiones basadas en evidencia.
- Inspiración para Nuevas Investigaciones: Los hallazgos del proyecto podrían inspirar investigaciones adicionales en áreas relacionadas, enriqueciendo el campo académico y ofreciendo nuevas oportunidades de estudio.
- Establecimiento de un centro de investigación entre diferentes entidades y ATI: Este tendría como objetivo mantener investigaciones para la realización de ciertos proyectos de alta relevancia que implican múltiples entidades del Tecnológico de Costa Rica, que puedan aportar un valor agregado a la institución.

1.2. Objetivos del Trabajo Final de Graduación

En esta sección, se describen los objetivos que guían esta investigación, definidos según la taxonomía de Bloom.

1.2.1. Objetivo General

Evaluar la viabilidad del desarrollo de un modelo predictivo que identifique a los estudiantes en riesgo de deserción de carrera del Tecnológico de Costa Rica, considerando aspectos financieros, legales, técnicos, organizacionales, de mercado y riesgos; para la determinación de su factibilidad con el fin de ser ejecutado por la institución, durante el segundo semestre del año 2024.

1.2.2. Objetivos Específicos

1. Examinar el estado de los modelos predictivos de deserción estudiantil en la educación superior de Costa Rica, para la determinación de su situación actual, retos y oportunidades para la implementación de futuros nuevos proyectos relacionados, durante las primeras cuatro semanas del segundo semestre del año 2024.
2. Realizar la definición del producto de modelo predictivo incluyendo información administrativa como alcance, criterios de éxito, presupuesto, plazos, estructura de gestión, plan de progreso, beneficios del proyecto, requisitos legales y datos necesarios almacenados por el Tecnológico de Costa Rica y fuentes externas, así como nuevas posibles variables que aporten valor, para el establecimiento de una base organizativa que facilite su estimación y ejecución.
3. Investigar la factibilidad del producto mediante evaluaciones que incluyen un análisis financiero, organizacional, de mercado, competencia y de riesgos; asegurando información relevante permitiendo su uso para la toma de decisiones sobre su viabilidad.
4. Elaborar un plan de implementación del producto modelo predictivo, especificando las fases, hitos críticos y estrategias de evaluación, que proporcione un plan estructurado que establezca una guía en su desarrollo en las operaciones del Tecnológico de Costa Rica.

1.3. Alcance de la investigación

En la siguiente sección se detallan los detalles a mantener dentro del alcance del proyecto, además de indicar que estaría fuera de este proyecto.

1.3.1. Entregables del proyecto

A continuación, se describen los entregables a tomar en cuenta según cada objetivo del proyecto, tomando en cuenta tanto los de gestión como de producto solicitados.

1.3.1.1. Entregables del producto

A continuación, en la **Tabla *1***, se muestran los entregables del producto según cada objetivo de la investigación.

Tabla 1
Entregables del producto según objetivo.

Objetivos	Entregables
Específico 1	• *Paper* y análisis de situación actual: Este documento, contendrá los detalles de la situación actual de los modelos predictivos de deserción en Costa Rica en la Educación Superior.
Específico 2	• Plan de proyecto: Este entregable contendrá detalles del proyecto a evaluar, es decir su formulación con el fin de determinar detalles presupuestarios y de gestión que permita estimar y determinar su complejidad y objetivos.
Específico 3	• Evaluación de mercado: Documento destinado a evaluar la intención de compra o en este caso de implementación por parte del cliente (TEC), es decir, su demanda y oferta. • Evaluación Social: Documento destinado a evaluar la capacidad económica-social para desarrollar este proyecto junto a beneficios financieros y no financieros que pueda ofrecer a la sociedad. • Evaluación de riesgos: Documento destinado a la evaluación de los riesgos asociados al proyecto, así como su probabilidad, impacto y posibles formas de mitigación. • Evaluación de competencia: Documento destinado a evaluar las capacidades de los competidores contra los que se enfrentaría el producto propuesto.
Específico 4	• Proyectos propuestos: Lista de proyectos sugeridos para continuar con la iniciativa, de modo que pueda completar el objetivo final. • Hoja de ruta: Estimación de tiempos y ruta sugerida para seguir en base a los proyectos anteriores, con el fin de desarrollar el proyecto evaluado en caso de determinar cómo viable.
General	• Estudio completo de viabilidad. • Documento de Trabajo de Final de Graduación. • Borrador de propuesta de investigación. • Documento de Investigación.

Nota: Elaboración propia.

1.3.1.2. Gestión del proyecto

A continuación, se determinan ciertos artefactos solicitados por la parte de gestión del proyecto.

- Minutas

Con el fin de mantener un control y documentación adecuada de las diferentes reuniones a realizar durante la ejecución de este proyecto, se determina una estructura para seguir una estandarización de esta información, encontrada en el Apéndice B: Plantilla de minuta.

- Gestión del Cambio

Para mantener un control sobre los cambios realizados en el proyecto, junto a cierta información relevante, se define la plantilla en el Apéndice C: Plantilla de gestión del cambio.

1.3.2. Supuestos del proyecto

En esta sección se indican los factores o elementos que se asume, se cumplirán o serán ciertos en la realización del proyecto.

- Existe muy poca o nula relación interdepartamental específicamente en lo relacionado a datos, por lo que, se considera como una de las causas del problema.
- El término “deserción” es aplicado para estudiantes que no han matriculado en más o igual a dos semestres o que se han cambiado de carrera.
- Existe un indicio de interés por parte del Tecnológico en la posible realización de este proyecto.
- No existe un equipo o herramienta activa que analice la deserción estudiantil mediante métodos de aprendizaje automatizado.

1.3.3. Exclusiones del proyecto

En esta sección se indican los entregables o productos que podrían esperarse del proyecto, pero, quedarán fuera del alcance.

- Creación y ejecución del modelo de predicción: este es un paso previo al desarrollo, el plan de proyecto solo cuenta con información general y posible del proyecto a realizar en el futuro, no entra en detalles como las variables exactas a considerar para el modelo, aunque si se determinarán fuentes de datos disponibles actualmente y posibles variables que pueden tomarse en cuenta para futuros proyectos.
- El periodo o tiempo estimado de graduación de un estudiante: Esta posible variable, tampoco será tomada en consideración, ya que no afecta los datos o factor técnico del modelo a evaluar.

1.3.4. Limitaciones del proyecto

En esta sección se indican cuáles son los factores o elementos que en alguna medida restringen la realización del proyecto.

- Fuentes de datos: debido a que existen algunos datos que no han sido mapeados por el Tecnológico de Costa Rica, tal como razones de deserción cuando el estudiante no pertenece a la institución; se dispone de una limitación en los datos obtenidos.
- Cantidad de involucrados: Debido a que este proyecto involucra una gran cantidad de departamentos e interesados tanto interna como externamente, se considera una limitación la cantidad ya que se debe de tomar en cuenta todas sus opiniones, por lo que se debe determinar un contacto por cada uno de estos para ser parte del proceso de reuniones y entrevistas.
- Profundidad de estudios de viabilidad: Debido al tiempo de realización y alcance propuesto, los estudios de viabilidad y exploración de fuentes de datos serán limitados por un tiempo determinado en el cronograma durante el semestre, por lo que, solo se realizarán las actividades que puedan ser completadas en ese espacio.

Capítulo 2. Estado del arte

La deserción universitaria es un fenómeno complejo que ha captado la atención de investigadores, administradores educativos y formuladores de políticas en todo el mundo. Este fenómeno no solo afecta las trayectorias individuales de los estudiantes, sino que también tiene implicaciones significativas para las instituciones de educación superior y la sociedad en general. En el contexto del Tecnológico de Costa Rica (TEC), la búsqueda de soluciones innovadoras para abordar este desafío ha llevado a la exploración de modelos predictivos basados en inteligencia artificial (IA).

2.1. Evolución de los Estudios sobre Deserción Universitaria

Los estudios sobre deserción universitaria han evolucionado significativamente desde los trabajos seminales de Spady (1970) y Tinto (1975). Estos autores sentaron las bases para entender la deserción como un proceso longitudinal de interacciones entre el individuo y los sistemas académico y social de la universidad. Tinto (1993) refinó su modelo, enfatizando la importancia de la integración académica y social del estudiante como factores clave en la decisión de persistir o abandonar los estudios.

Bean y Metzner (1985) expandieron este entendimiento al proponer un modelo que consideraba las particularidades de los estudiantes no tradicionales, destacando la influencia de factores externos a la institución. Estos modelos teóricos han sido fundamentales para el desarrollo de investigaciones posteriores y estrategias de retención en instituciones de educación superior a nivel global.

En años recientes, los modelos de deserción han evolucionado para incorporar factores emergentes. Por ejemplo, Rovai (2003) propuso un modelo que considera las características únicas de los estudiantes en línea, incluyendo habilidades informáticas y alfabetización digital. Más recientemente, Aljohani (2016) desarrolló un modelo integrado que combina factores tradicionales con variables específicas del contexto tecnológico actual.

2.2. Factores Asociados a la Deserción

La investigación ha identificado una amplia gama de factores asociados con la deserción universitaria. Estos pueden categorizarse en:

- Factores individuales: incluyen características demográficas, habilidades académicas previas, motivación y expectativas.
- Factores institucionales: abarcan la calidad de la enseñanza, los servicios de apoyo estudiantil y la cultura institucional.
- Factores socioeconómicos: consideran el nivel de ingresos familiares, la necesidad de trabajar mientras se estudia y el acceso a ayudas financieras.
- Factores académicos: involucran el rendimiento académico, la carga de cursos y la dificultad percibida del programa de estudios.

Estudios recientes, como el de Muñoz-Olano y Hurtado-Parrado (2017), han enfatizado la importancia de considerar la interacción entre estos factores, sugiriendo que la deserción es el resultado de una compleja interacción de variables más que de causas aisladas.

Los estudios más recientes han comenzado a explorar el impacto de la pandemia de COVID-19 en la deserción universitaria. Según González-Velázquez et al. (2022), la transición abrupta a la educación en línea ha introducido nuevos factores de riesgo, como la fatiga digital y la falta de interacción social directa con compañeros y profesores.

2.3. Contexto Latinoamericano y Costarricense

En América Latina, la deserción universitaria presenta características particulares debido a los contextos socioeconómicos y educativos de la región. Un informe de la UNESCO (2018) señaló que, en promedio, solo el 50% de los estudiantes que ingresan a la educación superior en América Latina logran graduarse. Este dato subraya la magnitud del desafío que enfrentan las instituciones de educación superior en la región.

En Costa Rica, el Programa Estado de la Nación (2019) ha reportado que, si bien el país ha logrado aumentar el acceso a la educación superior, los índices de deserción siguen siendo preocupantes. Según este informe, aproximadamente un 30% de los estudiantes que ingresan a universidades públicas no concluyen sus estudios. Este fenómeno no solo representa una pérdida de recursos para el sistema educativo, sino que también tiene implicaciones significativas para el desarrollo socioeconómico del país.

Un estudio reciente de Barragán y González (2023) sobre la educación superior en Costa Rica postpandemia señala que, si bien la crisis sanitaria aceleró la adopción de tecnologías educativas, también exacerbó las desigualdades existentes en el acceso a la educación superior, lo que podría tener implicaciones a largo plazo en las tasas de deserción.

2.4. La Situación en el Tecnológico de Costa Rica (TEC)

El TEC, como institución nacional autónoma de educación superior universitaria establecida por la ley No. 4777 del 10 de junio de 1971, no ha sido ajeno a esta problemática. Chinchilla-Brenes (2016) realizó un estudio descriptivo sobre la deserción en el TEC, encontrando que más del 52.7% de los estudiantes que abandonaron sus estudios lo hicieron de manera definitiva. Entre las razones principales citadas por los estudiantes se encontraban la falta de interés en la carrera elegida y las dificultades económicas.

Este estudio también reveló que el primer año de estudios es crítico en términos de deserción, lo que subraya la importancia de implementar estrategias de intervención temprana. Además, se identificó que ciertos programas académicos presentaban tasas de deserción más altas que otros, sugiriendo la necesidad de enfoques diferenciados según las características específicas de cada carrera.

2.5. Modelos Predictivos y *Machine Learning* en la Educación Superior

El avance en las tecnologías de análisis de datos y aprendizaje automático ha abierto nuevas posibilidades para abordar el problema de la deserción universitaria. Los modelos predictivos basados en aprendizaje automático (*machine learning*) permiten procesar grandes cantidades de datos para identificar patrones y factores de riesgo asociados con la deserción estudiantil.

Estudios como el de Delen (2010) han demostrado la eficacia de estas técnicas. Utilizando algoritmos como árboles de decisión, redes neuronales y máquinas de vectores de soporte, Delen logró desarrollar modelos con una precisión de predicción de hasta el 87% en la identificación de estudiantes en riesgo de deserción.

Los avances recientes en inteligencia artificial han permitido el desarrollo de modelos predictivos aún más sofisticados. Por ejemplo, Iam-On y Boongoen (2023) propusieron un modelo de aprendizaje profundo que integra datos estructurados y no estructurados, incluyendo registros de actividad en línea y análisis de sentimientos de las interacciones de los estudiantes en foros de discusión, logrando una precisión de predicción de hasta el 94%.

2.6. Investigaciones Previas en el TEC

En el contexto específico del TEC, se han realizado investigaciones pioneras en el uso de modelos predictivos para abordar la deserción estudiantil. Dos estudios particularmente relevantes son:

- "*Perspectives to Predict Dropout in University Students with Machine Learning*" (Solis et al., 2018): Este estudio, realizado en el TEC, exploró el uso de técnicas de aprendizaje automático (*machine learning*) para predecir la deserción estudiantil. Los investigadores aplicaron y compararon varios algoritmos de aprendizaje automático, incluyendo regresión logística, árboles de decisión, bosques aleatorios y máquinas de vectores de soporte. Los resultados fueron prometedores, con el modelo logrando una precisión de predicción hasta del 91%. Este alto nivel de precisión demuestra el potencial de las técnicas de aprendizaje automático (*machine learning*) para identificar de manera efectiva a los estudiantes en riesgo de deserción. El estudio también identificó los factores más influyentes en la predicción de la deserción, entre los que se incluían el rendimiento académico en los primeros semestres, la carga académica y características socioeconómicas de los estudiantes.
- "*A Multinomial and Predictive Analysis of Factors Associated with University Dropout*" (Fernández-Martín et al., 2018): Esta investigación profundizó en el análisis de los factores asociados con la deserción universitaria en el TEC. Utilizando un enfoque multinomial, los autores exploraron cómo diferentes variables interactúan para influir en la decisión de un estudiante de abandonar sus estudios. El estudio reveló que factores como la motivación del estudiante, el rendimiento académico en el primer año, y la adaptación al ambiente universitario eran predictores significativos de la deserción. Además, se encontró que estos factores podían tener diferentes pesos dependiendo del momento en que ocurría la deserción (temprana o tardía en la carrera del estudiante). Una contribución importante de este estudio fue la identificación de perfiles de riesgo, lo que permite una comprensión más matizada de los diferentes tipos de estudiantes que pueden estar en riesgo de deserción.

Estos estudios representan un avance significativo en la comprensión y el abordaje de la deserción estudiantil en el TEC. No solo demuestran la viabilidad técnica de implementar modelos predictivos en el contexto institucional, sino que también proporcionan descubrimientos valiosos sobre los factores específicos que contribuyen a la deserción en esta institución.

Por otra parte, anteriormente se realizó una primera iteración de este proyecto actual en un proyecto de graduación en años anteriores, sin embargo, este por motivos personales del postulante a licenciado, no se llegó a culminar, por lo que, en esta nueva iteración, se tuvo la necesidad de reiniciar la investigación completamente desde cero, ya que la anterior no alcanzó objetivos importantes ni datos que son relevantes a la situación actual.

2.7. Desafíos en la Implementación de Modelos Predictivos

A pesar del potencial demostrado por estos estudios, la implementación efectiva de modelos predictivos en el contexto universitario enfrenta varios desafíos:

- Calidad y accesibilidad de los datos: Como señalan Fernández-Martín et al. (2018), la calidad y la integración de los datos son cruciales para el éxito de estos modelos. En muchas instituciones, incluido el TEC, la fragmentación de los sistemas de información y las políticas de gestión de datos pueden obstaculizar la creación de un sistema unificado y eficiente.
- Consideraciones éticas: El uso de datos personales de los estudiantes para predecir su comportamiento académico plantea importantes cuestiones éticas. Ekowo y Palmer (2016) destacan la necesidad de equilibrar la intervención basada en datos con el respeto a la autonomía y privacidad de los estudiantes.
- Implementación de intervenciones: Identificar a los estudiantes en riesgo es solo el primer paso. El verdadero desafío radica en diseñar e implementar intervenciones efectivas basadas en estas predicciones.
- Resistencia institucional al cambio: La adopción de nuevas tecnologías y enfoques basados en datos puede enfrentar resistencia en entornos académicos tradicionales.
- Complejidad e interpretabilidad de los modelos: Con el aumento en la sofisticación de los modelos de IA, surge el desafío de la 'caja negra'. Este término se refiere a modelos cuyos procesos internos son tan complejos que resultan opacos incluso para sus creadores. En estos modelos de 'caja negra', si bien podemos ver las entradas (datos de los estudiantes) y las salidas (predicciones de deserción), el proceso que lleva de unas a otras es difícil de comprender o explicar en detalle.

Rudin y Radin (2023) argumentan que en contextos sensibles como la educación superior, donde las decisiones pueden tener un impacto significativo en la vida de los estudiantes, es crucial desarrollar modelos que no solo sean precisos, sino también interpretables. Los modelos interpretables permiten a los educadores y administradores entender y explicar cómo se llegó a una predicción específica, lo que es fundamental para generar confianza en el sistema y diseñar intervenciones efectivas.

Por ejemplo, un modelo interpretable no solo indicaría que un estudiante tiene alto riesgo de deserción, sino que podría explicar que esta predicción se basa en factores como bajo rendimiento en cursos clave, disminución en la asistencia a clases o reducción en la participación de actividades extracurriculares. Esta transparencia no solo aumenta la confianza en el modelo, sino que también proporciona información accionable para las intervenciones.

Además, la interpretabilidad es crucial para detectar y corregir sesgos en los modelos, asegurando que las predicciones sean justas y no discriminatorias. En el contexto de la predicción de la deserción estudiantil en el TEC, esto es particularmente importante para evitar perpetuar o exacerbar desigualdades existentes en el acceso y éxito en la educación superior.

2.8. Marco Legal y Ético

La implementación de modelos predictivos en el ámbito educativo, particularmente aquellos diseñados para predecir la deserción estudiantil, plantea una serie de desafíos legales y éticos significativos. Estos desafíos se derivan principalmente de la naturaleza sensible de los datos personales de los estudiantes y el potencial impacto

de las predicciones en sus trayectorias académicas. En el contexto del Tecnológico de Costa Rica (TEC), es crucial examinar detalladamente el marco legal y ético que rodea este tipo de iniciativas.

2.8.1. Marco Legal Costarricense

2.8.1.1. Ley de Protección de la Persona frente al tratamiento de sus datos personales (Ley N° 8968)

Esta ley, promulgada en 2011, establece el marco fundamental para la protección de datos personales en Costa Rica. Chirino Sánchez (2020) destaca varios aspectos cruciales de esta ley que son particularmente relevantes para el proyecto de modelo predictivo:

a) Principio de consentimiento informado: El artículo 5 de la ley estipula que el tratamiento de datos personales requiere el consentimiento expreso del titular de los datos. En el contexto del modelo predictivo, esto implica que cada estudiante debe ser informado claramente sobre el propósito del modelo, qué datos se utilizarán y cómo se procesarán. Además, deben tener la opción de dar o negar su consentimiento para participar.
b) Calidad de la información: El artículo 6 exige que los datos personales sean actuales, veraces, exactos y adecuados al fin para el cual fueron recolectados. Esto plantea un desafío significativo para el modelo predictivo, ya que la precisión de las predicciones depende directamente de la calidad de los datos utilizados. El TEC deberá implementar procesos rigurosos para garantizar la actualización y verificación constante de los datos de los estudiantes.
c) Derechos ARCO: La ley garantiza los derechos de Acceso, Rectificación, Cancelación y Oposición (ARCO) a los titulares de los datos. En la práctica, esto significa que los estudiantes deben tener la capacidad de:
    - Acceder a todos los datos que el modelo utiliza sobre ellos.
    - Corregir cualquier información inexacta.
    - Solicitar la eliminación de sus datos del modelo.
    - Oponerse al procesamiento de sus datos en cualquier momento.

    La implementación de estos derechos en el contexto de un modelo predictivo complejo puede ser técnicamente desafiante y requerirá una cuidadosa planificación.

d) Seguridad de los datos: El artículo 10 exige medidas de seguridad apropiadas para proteger los datos personales. Dado que el modelo predictivo manejará datos sensibles sobre el rendimiento académico y otros factores personales de los estudiantes, será necesario implementar robustas medidas de seguridad informática para prevenir accesos no autorizados o filtraciones de datos.

2.8.1.2. Reglamento a la Ley de Protección de la Persona frente al tratamiento de sus datos personales (Decreto Ejecutivo N° 37554-JP)

Este reglamento, emitido en 2013, proporciona detalles adicionales sobre la implementación de la Ley N° 8968. Chirino Sánchez (2020) señala dos aspectos particularmente relevantes para el proyecto de modelo predictivo:

a) Categorías de datos: El reglamento define diferentes categorías de datos (públicos, personales de acceso restringido, personales de acceso irrestricto y datos sensibles), cada una con diferentes niveles de protección. En el contexto del modelo predictivo, será crucial categorizar correctamente cada tipo de dato utilizado. Por ejemplo, el rendimiento académico podría considerarse como datos personales de acceso restringido, mientras que información sobre la salud mental o situación económica del estudiante podría clasificarse como datos sensibles, requiriendo un nivel aún mayor de protección.
b) Transferencia de datos: El reglamento establece las condiciones bajo las cuales los datos personales pueden ser transferidos a terceros. Esto es particularmente relevante si el desarrollo o la implementación del modelo predictivo involucra colaboración con entidades externas al TEC. Cualquier transferencia de datos deberá cumplir estrictamente con estas regulaciones.

2.8.1.3. Ley General de Control Interno (Ley N° 8292)

Aunque esta ley no está directamente relacionada con la protección de datos, es relevante para el proyecto en términos de gobernanza y gestión de riesgos. La ley requiere que las instituciones públicas, incluidas las universidades estatales como el TEC, implementen sistemas de control interno para garantizar la eficacia y eficiencia de sus operaciones (Asamblea Legislativa de la República de Costa Rica, 2002).

En el contexto del modelo predictivo, esto implica la necesidad de establecer procesos claros para:

- La supervisión del desarrollo y la implementación del modelo.
- La evaluación continua de la eficacia y la precisión del modelo.
- La gestión de riesgos asociados con el uso del modelo, como la posibilidad de decisiones sesgadas o erróneas.
- La auditoría regular del sistema para garantizar su conformidad con las políticas institucionales y las leyes aplicables.

2.8.1.4. Ley de Certificados, Firmas Digitales y Documentos Electrónicos (Ley N° 8454)

Esta ley podría ser relevante si se considera la implementación de consentimientos digitales o la gestión electrónica de documentos relacionados con el modelo predictivo (Asamblea Legislativa de la República de Costa Rica, 2005). Por ejemplo, si se decide obtener el consentimiento de los estudiantes a través de un formulario en línea, esta ley proporcionaría el marco legal para garantizar la validez de dichos consentimientos electrónicos.

2.8.2. Consideraciones Éticas

Más allá del marco legal, existen importantes consideraciones éticas que deben guiar el desarrollo y la implementación del modelo predictivo. Prinsloo y Slade (2017) destacan varios principios éticos clave en el contexto de la analítica del aprendizaje:

2.8.2.1. Equidad y no discriminación

Es crucial asegurar que el modelo predictivo no perpetúe o exacerbe desigualdades existentes. Esto implica un cuidadoso examen de los datos de entrada y los algoritmos utilizados para evitar sesgos. Por ejemplo, si el modelo tiende a identificar como "en riesgo" a estudiantes de ciertos grupos socioeconómicos o étnicos, podría llevar a intervenciones sesgadas que refuercen las desigualdades existentes.

Para abordar este desafío, el TEC podría considerar:

- Realizar auditorías regulares del modelo para detectar y corregir sesgos.
- Incluir expertos en ética y diversidad en el equipo de desarrollo del modelo.
- Implementar técnicas de "*debiasing*" en el proceso de desarrollo del modelo. Esto incluye ajustar los datos, aplicar restricciones al modelo durante el entrenamiento, o corregir las salidas finales para asegurar que las predicciones sean justas y no estén influenciadas por prejuicios ocultos.

2.8.2.2. Transparencia

Los estudiantes deben ser informados sobre la existencia del modelo predictivo, cómo se utilizan sus datos y cómo pueden afectar las decisiones institucionales. Esto va más allá del mero cumplimiento legal y se adentra en el territorio de la ética de la información.

El TEC podría considerar:

- Desarrollar una política de transparencia que detalle cómo se utiliza el modelo predictivo.
- Proporcionar a los estudiantes acceso a sus propias predicciones y los factores que las influencian.
- Organizar sesiones informativas regulares para explicar el funcionamiento del modelo a la comunidad universitaria.

2.8.2.3. Derecho a la explicación

En línea con principios emergentes en el campo de la ética de la IA, los estudiantes deberían tener derecho a entender cómo se llegó a una predicción particular sobre su riesgo de deserción. Esto plantea desafíos técnicos, especialmente si se utilizan algoritmos de "caja negra" como las redes neuronales profundas.

Para abordar este desafío, el TEC podría:

- Priorizar el uso de algoritmos interpretables en el desarrollo del modelo, como es el caso de los algoritmos de *Random Forest*.
- Desarrollar herramientas de explicación *post-hoc* para modelos más complejos.
- Capacitar al personal académico y administrativo para interpretar y explicar las predicciones del modelo a los estudiantes de ser posible.

2.8.2.4. Límites de la intervención

Es importante establecer límites claros sobre cómo se utilizarán las predicciones del modelo. Por ejemplo, ¿se utilizarán solo para ofrecer apoyo adicional o podrían influir en decisiones más críticas como la asignación de becas?

El TEC debería considerar:

- Desarrollar una política clara sobre los usos permitidos y prohibidos de las predicciones del modelo.
- Establecer un comité de ética para supervisar el uso del modelo y manejar casos límite.
- Implementar mecanismos de apelación para los estudiantes que sientan que han sido tratados injustamente basándose en las predicciones del modelo.

2.8.3. Perspectiva Internacional

Aunque algunos no son directamente aplicables en Costa Rica, es útil considerar marcos internacionales como referencia de mejores prácticas:

2.8.3.1. Reglamento General de Protección de Datos (GDPR) de la Unión Europea

El GDPR establece estándares rigurosos para el procesamiento de datos personales (Parlamento Europeo y Consejo de la Unión Europea, 2016). Aunque no es legalmente vinculante en Costa Rica, sus principios pueden servir como una guía valiosa. Algunos aspectos relevantes incluyen:

- El "derecho a no ser sujeto a decisiones basadas únicamente en el procesamiento automatizado", lo que podría tener implicaciones para cómo se utilizan las predicciones del modelo.
- Requisitos estrictos para el consentimiento informado.
- El concepto de "privacidad por diseño", que implica incorporar consideraciones de privacidad desde las primeras etapas del desarrollo del proyecto.

2.8.3.2. Principios de la OCDE sobre Inteligencia Artificial

La OCDE, a la cual Costa Rica está suscrita, ha desarrollado principios para el desarrollo ético y confiable de sistemas de IA (OECD, 2019). Estos principios incluyen:

- Crecimiento inclusivo, desarrollo sostenible y bienestar
- Valores centrados en el ser humano y equidad
- Transparencia y explicabilidad
- Robustez, seguridad y protección
- Responsabilidad

Estos principios podrían servir como una guía valiosa para el desarrollo e implementación del modelo predictivo en el TEC.

2.8.4. Implicaciones para el Proyecto

Considerando este marco legal y ético, el desarrollo del modelo predictivo en el TEC debería incluir las siguientes medidas:

- Realización de una Evaluación de Impacto en la Protección de Datos (EIPD): Antes de implementar el modelo, se debería llevar a cabo una evaluación exhaustiva de los riesgos potenciales para la privacidad de los estudiantes y desarrollar estrategias de mitigación.
- Desarrollo de un protocolo claro para obtener el consentimiento informado: Este protocolo debería detallar qué información se proporcionará a los estudiantes, cómo se obtendrá su consentimiento y cómo pueden retirarlo si así lo desean.
- Implementación de robustas medidas de seguridad: Esto incluye la encriptación de datos, controles de acceso estrictos y auditorías de seguridad regulares.
- Creación de un comité de ética: Este comité debería supervisar el desarrollo y uso del modelo predictivo, asegurando que se adhiera a los principios éticos establecidos.

- Establecimiento de procesos para permitir a los estudiantes ejercer sus derechos ARCO: Esto podría implicar el desarrollo de una interfaz de usuario que permita a los estudiantes acceder y gestionar sus datos personales utilizados en el modelo.
- Desarrollo de políticas claras sobre el uso de las predicciones: Estas políticas deberían establecer cómo se utilizarán las predicciones del modelo y qué límites se establecerán en su aplicación.
- Implementación de mecanismos de auditoría y transparencia: Esto podría incluir la publicación regular de informes sobre el rendimiento y el impacto del modelo.

Como señalan Rubel y Jones (2016), la implementación de estas medidas no solo es una cuestión de cumplimiento legal, sino también de responsabilidad ética y de mantener la confianza de la comunidad universitaria.

2.9. Tendencias Futuras y Oportunidades

El campo de la predicción de la deserción estudiantil está en constante evolución. Algunas tendencias y oportunidades emergentes incluyen:

- Integración de datos en tiempo real: La capacidad de incorporar datos actualizados continuamente podría mejorar significativamente la precisión y relevancia de los modelos predictivos.
- Personalización de intervenciones: El uso de técnicas de IA más avanzadas podría permitir el diseño de estrategias de intervención altamente personalizadas para cada estudiante en riesgo.
- Análisis de redes sociales: La incorporación de datos de las interacciones sociales de los estudiantes, tanto en línea como fuera de línea, podría proporcionar nuevos descubrimientos sobre los factores que influyen en la persistencia académica.
- Explicabilidad de la IA: El desarrollo de técnicas de "IA explicable" podría abordar los desafíos de interpretabilidad, haciendo que los modelos sean más transparentes y comprensibles para los tomadores de decisiones en el ámbito educativo.
- Otra tendencia emergente es el uso de técnicas de aprendizaje federado (*federated learning*), que permiten entrenar modelos predictivos sin centralizar los datos de los estudiantes, abordando así preocupaciones de privacidad y cumplimiento normativo (Li et al., 2024).

Capítulo 3. Marco Metodológico

3.1. Tipo de Investigación

Para la realización de este proyecto, existen dos tipos de investigación en función de sus propósitos, una enfocada en producir conocimiento y teorías conocida como investigación básica, mientras que la otra se centra en la resolución de problemas llamada investigación aplicada. (Hernández-Sampieri. & Mendoza, 2018)

Debido a la naturaleza anteriormente expuesta, este proyecto tomará como tipo la investigación aplicada, ya que se lleva a cabo un proceso de búsqueda de información, recolección de datos, análisis e interpretación relacionados a la evaluación de viabilidad del proyecto en cuestión, implicando resolver un problema encontrado.

3.2. Enfoque y diseño de la Investigación

Siguiendo la línea de la definición metodológica, existen tres enfoques posibles para la investigación: Cuantitativa, Cualitativa y Mixta. En la ruta Cuantitativa, se considera que es más apropiada para investigaciones en la que se desee estimar magnitudes, frecuencias u ocurrencias como tal de ciertos fenómenos con el fin de probar una hipótesis, mientras que la ruta Cualitativa, el sujeto de investigación examina los eventos u hechos, se revisan estudios previos y en ciertas veces simultáneamente, con el objetivo de generar una teoría posible sobre los hechos que se están observando. Por último, el enfoque mixto es una combinación de las dos anteriores analizando ambos tipos de datos. (Hernández-Sampieri. & Mendoza, 2018)

Según estos tres enfoques, debido a que se realizará una investigación profunda de distintos aspectos del entorno documentales, análisis de datos o variables tanto cualitativos como cuantitativos junto al estudio de eventos previos con el fin de establecer una teoría sobre la viabilidad del proyecto en base a distintos tipos y fuentes de datos, se tomará un enfoque mixto.

Por otra parte, el enfoque cualitativo presenta distintos diseños posibles para la investigación, en base a estos, se considera como apropiado el diseño de investigación-acción para la realización de este proyecto, centrándose en la resolución de una problemática basado en un procedimiento metodológico de fases, ya que según Hernández-Sampieri. & Mendoza (2018) este diseño se orienta hacia la resolución de diversos problemas, basándose en fases cíclicas o espirales que incluyen la identificación del problema, la planificación, la implementación, la evaluación y la retroalimentación (observar, pensar, actuar y revisar).

3.3. Fuentes de Información

En esta sección, se mostrarán las principales fuentes de información que aportarán referencias e información de importancia para la investigación, tanto primarias como secundarias.

<u>Fuentes primarias</u>

Hernández et al. (2014), define a las fuentes primarias como aquellas que proporcionan datos de primera mano, tratándose de documentos que incluyen en sí mismos los resultados de los estudios correspondientes. Según

esta definición anterior, las fuentes primarias a utilizar en el proyecto se presentan en la siguiente **Tabla *2***, la cual nombra el documento esperado y su importancia para la investigación.

Tabla 2
Fuentes primarias de la investigación.

Documento	Importancia
Datos estudiantiles	Este es el eje central con el cual puede ser construido el modelo predictivo, por tanto, todo tipo de dato puede ser importante desde registro, administrativo, financiero o psicológico, entre otros; que estén siendo almacenados por el ITCR será de suma importancia para ser tomado y analizado en la investigación.
Entrevistas a sujetos de investigación	Estos serán base fundamental para la fundamentación de distintas perspectivas tanto en la viabilidad como de conocimiento general de la institución, ya que se recopila la opinión de diferentes entidades de interés que pueden aportar su valiosa información con el fin de determinar la decisión final.
Políticas y leyes nacionales e institucionales	Estos documentos serán fuente importante para el estudio legal, con el fin de determinar su validez legal frente a la institución, derechos humanos y constitución nacional.

Nota: Elaboración propia.

Fuentes secundarias

Según Guzmán, L. (s. f.). Las fuentes secundarias, por otra parte, permiten conocer hechos o fenómenos a partir de documentos o datos recopilados por otros. Según esta definición anterior, las fuentes secundarias a utilizar en el proyecto se presentan en la siguiente **Tabla 3**, la cual nombra el documento esperado y su importancia para la investigación.

Tabla 3
Fuentes secundarias de la investigación

Documento	Importancia
A Multinomial and Predictive Analysis of Factors Associated with University Dropout	Ofrece un análisis detallado de los factores que contribuyen a la deserción universitaria. Al utilizar un enfoque multinomial y predictivo, proporciona una base sólida para entender qué variables deben ser consideradas en el modelo predictivo, lo que puede ayudar a mejorar su precisión y eficacia.
Perspectives to Predict Dropout in University Students with Machine Learning	Explora cómo las técnicas de aprendizaje automático (*machine learning*) pueden ser aplicadas para predecir la deserción universitaria. Proporciona perspectivas sobre los enfoques metodológicos y los modelos estadísticos que podrían ser aplicados en el proyecto, facilitando la selección de las técnicas más adecuadas para el análisis de datos en el TEC.

Documento	Importancia
Estudio descriptivo de variables sociodemográficas y motivacionales asociadas a la deserción	Describe cómo las variables sociodemográficas y motivacionales influyen en la deserción estudiantil. El entendimiento de estas variables puede permitir que el modelo predictivo considere factores más allá del rendimiento académico, lo que puede llevar a intervenciones más efectivas y dirigidas para mantener a los estudiantes en la institución.

Nota: Elaboración propia.

3.4. Sujetos de Información

Según Creswell, J. W. (2014), el sujeto de investigación se refiere a las entidades las cuales se recopila y analiza la información en el curso de un estudio científico. Los sujetos pueden participar activamente en el proceso de investigación a través de encuestas, entrevistas, experimentos u otras formas de estudio empírico. En base a esta definición, en la **Tabla 4**, de definen los sujetos para esta investigación, considerando su rol, años de experiencia, caracterización y una justificación de su importancia.

Tabla 4
Sujetos de investigación del proyecto.

Rol del sujeto	Años de experiencia en el rol	Caracterización del sujeto (diferentes responsabilidades y funciones del rol)	Justificación de la importancia de este sujeto para su investigación.
Colaboradora del Departamento de Orientación y Psicología	Más de 2 años	Supervisa la provisión de servicios de orientación y apoyo psicológico, esenciales para el bienestar estudiantil.	Puede proporcionar datos valiosos sobre las necesidades psicológicas y emocionales de los estudiantes, que son indicadores importantes para la predicción de deserción. Además de tener información importante sobre un modelo predictivo desarrollado en años anteriores y todo el proceso llevado a cabo.
Evaluadora e Investigadora	16 años	Lleva a cabo evaluaciones de programas y estudios sobre la eficacia de diversas iniciativas educativas.	Su trabajo ayudará a evaluar los datos históricos y actuales que podrían integrarse en el modelo predictivo para mejorar su precisión, contando con experiencia en múltiples investigaciones relacionadas a la deserción estudiantil.
Colaboradoras de la Oficina Planificación	4 años	Encargada de la planificación estratégica y operativa, asegurando que los recursos estén alineados con las metas	Fundamental para entender cómo se pueden alinear los objetivos del modelo predictivo con los planes a largo plazo del TEC, además de tener

Rol del sujeto	**Años de experiencia en el rol**	**Caracterización del sujeto (diferentes responsabilidades y funciones del rol)**	**Justificación de la importancia de este sujeto para su investigación.**
Institucional del ITCR		institucionales, además de estar encargada de diferentes sistemas técnicos como cubos de datos y su administración.	experiencia en la implementación de cubos de datos y su administración a través del TEC.
Abogado Defensor Estudiantil ITCR	8 años	Asesorar y defender a las y los estudiantes de esta Institución contra los actos y actuaciones materiales de cualquier instancia institucional sea administrativa o académica, que lesione sus derechos e intereses.	Puede ofrecer perspectivas sobre las implicaciones legales de implementar un modelo predictivo, especialmente en términos de privacidad de datos y consentimiento.
Profesor y estadístico	15 años	Profesor de la Escuela de Administración de Empresas y estadístico.	Tiene amplia experiencia en el área estadística, participando en múltiples investigaciones y como consultor en varias ocasiones para el desarrollo de modelos de predicción tal como el desarrollado en su momento por la OPI.
Arquitecto de software	11 años	Diseñar y aprobar soluciones de software, desarrollar programas especializados.	Ofrece un gran criterio de experto en el área técnico, ya que cuenta con muchos proyectos realizados similares a los que se estimarán en este proyecto. Cuenta con experiencia en la creación de modelos de predicción, RPA, modelos de inteligencia artificial, procesos de minería de datos, entre otros aspectos importantes.

Nota: Elaboración propia.

3.5. Variables de Investigación

En el desarrollo de esta investigación, es esencial definir claramente las variables ya que son incógnitas que necesitas ser respondidas en su desarrollo. Las variables se clasifican en independientes, que influencian el estudio, y dependientes, que son influenciadas por las variables independientes. A continuación, en la **Tabla 5**, **Tabla 6**, **Tabla 7** y **Tabla 8** se detallan cada variable relacionada con los objetivos específicos del proyecto, incluyendo su conceptualización, indicadores para su medición y los instrumentos utilizados para recolectar los datos necesarios.

Tabla 5
Variables de investigación del objetivo específico 1.

Objetivo específico	Examinar el estado de los modelos predictivos de deserción estudiantil en la educación superior de Costa Rica, para la determinación de su situación actual, retos y oportunidades para la implementación de futuros nuevos proyectos relacionados, durante las primeras cuatro semanas del segundo semestre del año 2024.			
Variable	**Tipo**	**Conceptualización**	**Indicadores**	**Instrumentos**
El estado de los modelos predictivos de deserción estudiantil.	Independiente	Búsqueda, identificación y análisis los modelos o investigaciones afines a la predicción de la deserción estudiantil	- Porcentaje de errores de tipo 1 y tipo 2 del proyecto. - Última ejecución.	- Revisión documental de investigaciones y modelos. -Entrevista profunda a sujetos de investigación.
Situación actual, retos y oportunidades para futuros proyectos relacionados.	Dependiente	Análisis cualitativo de los modelos o investigaciones de predicción de deserción estudiantil, con respecto a sus principales retos y oportunidades.	- Cantidad de retos encontrados. - Cantidad de oportunidades de mejora.	- Grupo de discusión con el profesor tutor.

Nota: Elaboración propia.

Tabla 6
Variables de investigación del objetivo específico 2.

Objetivo específico	Realizar la definición del producto de modelo predictivo incluyendo información administrativa como alcance, criterios de éxito, presupuesto, plazos, estructura de gestión, plan de progreso, beneficios del proyecto, requisitos legales y datos necesarios almacenados por el Tecnológico de Costa Rica y fuentes externas, así como nuevas posibles variables que aporten valor, para el establecimiento de una base organizativa que facilite su estimación y ejecución.			
Variable	**Tipo**	**Conceptualización**	**Indicadores**	**Instrumentos**
Definición del producto.	Independiente	Información de la gestión del proyecto estimada y relevante al	- Cantidad de variables encontradas	- Grupo de discusión con el profesor tutor.

Objetivo específico	Realizar la definición del producto de modelo predictivo incluyendo información administrativa como alcance, criterios de éxito, presupuesto, plazos, estructura de gestión, plan de progreso, beneficios del proyecto, requisitos legales y datos necesarios almacenados por el Tecnológico de Costa Rica y fuentes externas, así como nuevas posibles variables que aporten valor, para el establecimiento de una base organizativa que facilite su estimación y ejecución.			
Variable	**Tipo**	**Conceptualización**	**Indicadores**	**Instrumentos**
		modelo predictivo dispuesto a desarrollarse.	- Monto total de presupuesto del proyecto. - Porcentaje de cumplimiento de hitos del proyecto	- Revisión documental de referencias - Entrevista a profesor tutor.
Base de organización establecida del modelo predictivo	Dependiente	Estructura organizativa y de gestión establecida para la implementación y ejecución del modelo.	- Cantidad de involucrados en la jerarquía horizontal. - Cantidad de tiempo de ejecución para cada hito establecido.	- Entrevista a profesor tutor. - Grupo de discusión con el profesor tutor.

Nota: Elaboración propia.

Tabla 7
Variables de investigación del objetivo específico 3.

Objetivo específico	Investigar la factibilidad del producto mediante evaluaciones que incluyen un análisis financiero, organizacional, de mercado, competencia y de riesgos; asegurando información relevante permitiendo su uso para la toma de decisiones sobre su viabilidad.			
Variable	**Tipo**	**Conceptualización**	**Indicadores**	**Instrumentos**
Factibilidad del producto	Independiente	Análisis realizados para evaluar diferentes aspectos de factibilidad del modelo.	- Cantidad de evaluaciones realizadas - Cantidad de retorno de inversión. - Cantidad de riesgos asociados al proyecto. - Cantidad de competidores.	- Revisión documental de referencias. -Entrevista profunda a sujetos de investigación.

Objetivo específico	Investigar la factibilidad del producto mediante evaluaciones que incluyen un análisis financiero, organizacional, de mercado, competencia y de riesgos; asegurando información relevante permitiendo su uso para la toma de decisiones sobre su viabilidad.			
Variable	**Tipo**	**Conceptualización**	**Indicadores**	**Instrumentos**
Información de decisión sobre la viabilidad del modelo predictivo.	Dependiente	Información de valor obtenida por las evaluaciones de viabilidad que permite tomar una decisión fundamentada en sus resultados.	- Cantidad de estudios con una factibilidad positiva. - Cantidad de estudios con una factibilidad negativa.	- Grupo de discusión con el profesor tutor.

Nota: Elaboración propia.

Tabla 8
Variables de investigación del objetivo específico 4.

Objetivo específico	Elaborar un plan de implementación del producto modelo predictivo, especificando las fases, hitos críticos y estrategias de evaluación, que proporcione un plan estructurado que establezca una guía en su desarrollo en las operaciones del Tecnológico de Costa Rica.			
Variable	**Tipo**	**Conceptualización**	**Indicadores**	**Instrumentos**
Plan de implementación del producto.	Independiente	Documento que sugiere formas de implementar la iniciativa en el TEC mediante proyectos, hitos y plazos.	- Cantidad de proyectos sugeridos. - Recursos por proyecto. -Tiempo estimado por proyecto.	- Grupo de discusión con el profesor tutor. - Revisión documental de referencias
Plan guía de desarrollo del proyecto.	Dependiente	Documento que sugiere una ruta clara de implementación de los proyectos sugeridos.	- Plazos de implementación por proyecto.	- Análisis documental de proyectos. - Entrevista a involucrados del TEC. - Grupo de discusión con el profesor tutor.

Nota: Elaboración propia.

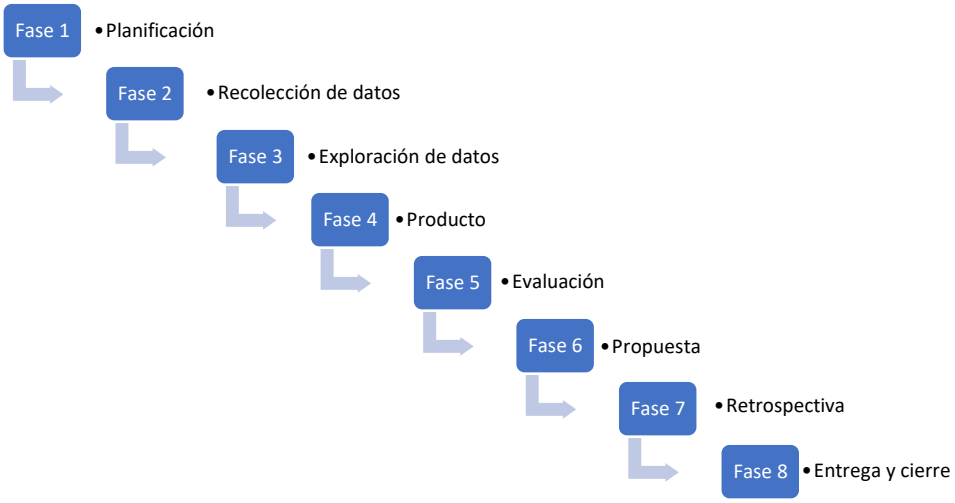
3.6. Procedimiento Metodológico de la Investigación.

A continuación, se detallan las fases por las cuales se desarrolla la investigación, definiendo su estructura y conteniendo ciertas actividades con el fin de lograr los objetivos propuestos.

3.6.1. Diagrama propuesto para las fases del procedimiento metodológico

Seguidamente, se muestra en la **Figura *1***, las fases definidas para la elaboración de la investigación, mientras que en la **Tabla *9***, se detalla una breve descripción sobre lo tomado en consideración de realizar durante estas fases.

Figura 1
Diagrama de fases del procedimiento metodológico.



Nota: Elaboración propia.

Tabla 9
Descripción de fases del procedimiento metodológico.

Id	Fase	Descripción
1	Planificación	En esta etapa, comienza la planificación del proyecto, realizando las tareas iniciales de definición de la problemática, objetivos, alcance, justificación y metodología correspondiente para determinar un proyecto de investigación formal.
2	Recolección de datos	Una vez definido el proyecto y sus partes iniciales, se comienza a investigar, diseñar ya aplicar todos los instrumentos de recolección de datos necesarios. Esto es realizado al inicio debido a que es un proceso lento y dependiente a la disponibilidad de los sujetos de investigación.
3	Exploración de datos	Al haber recolectado la información suficiente y pertinente, se realiza un análisis de la primera parte de la investigación, sobre los datos almacenados por el TEC relacionados a los estudiantes, realizando una verificación de calidad y de relevancia, además de verificar datos ausentes que puedan ser de utilidad.

Id	Fase	Descripción
4	Producto	Seguidamente, se realiza la definición del producto, estableciendo todos los parámetros de administración del proyecto necesarios para mantener una estimación cercana a la realidad.
5	Evaluación	Ya que se tiene el producto definido, se pasa a la fase de evaluación de viabilidad, en la que se aplican todas las evaluaciones definidas sobre los estudios financieros, de mercado, de riesgos, legal y técnico.
6	Propuesta	Con el objetivo de realizar una propuesta, se realiza un plan de implementación en el cual se definen los proyectos necesarios y una hoja de ruta capaz de guiar a la organización, una forma en la que puedan implementar esta iniciativa a futuro.
7	Retrospectiva	Una vez completado todos los pasos anteriores, se entra a una etapa de retrospectiva, en la cual se analizan todos los puntos principales de la investigación, verificando la calidad del documento entre otros aspectos y realizando las correcciones pertinentes.
8	Entrega y Cierre	Finalmente, la última etapa consiste en la revisión final y entrega del documento con otros entregables junto a su defensa frente al jurado.

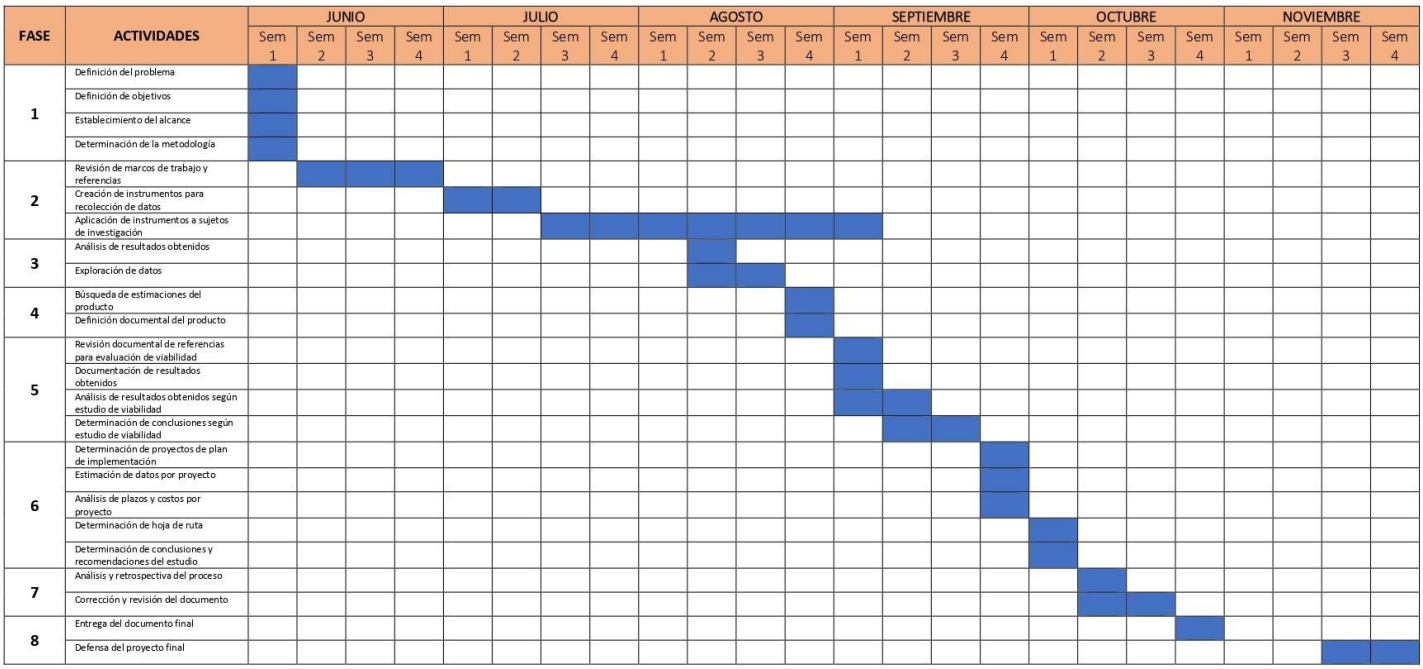
Nota: Elaboración propia.

3.6.2. Cronograma

Finalmente, en la **Figura *2***, se muestra el cronograma planeado para la ejecución del procedimiento anteriormente definido, este se encuentra seccionado por las fases y actividades, además, de indicar en que semana es ejecutada cada tarea.

Figura 2

Cronograma de la investigación.



Nota: Elaboración propia.

3.7. Operacionalización de las variables

En esta sección, en la tabla X, se muestra la relación entre las fases de la metodología propuesta anteriormente, contra los objetivos, variables, instrumentos y sujetos de investigación definidos.

Fase	Objetivo específico	Instrumentos	Variables	Sujetos de investigación.
Planificación	No se cubre objetivo.	No se aplican instrumentos.	No se cubren variables.	No se relacionan sujetos.
Recolección de datos	Examinar el estado de los modelos predictivos de deserción estudiantil en la educación superior de Costa Rica, para la determinación de su situación actual, retos y oportunidades para la implementación de futuros nuevos proyectos relacionados, durante las primeras cuatro semanas del segundo semestre del año 2024	- Revisión documental de investigaciones y modelos. - Entrevista profunda a sujetos de investigación.	- El estado de los modelos predictivos de deserción estudiantil. - Situación actual, retos y oportunidades para futuros proyectos relacionados.	- Colaboradora del Departamento de Orientación y Psicología - Ex evaluadora e Investigadora - Colaboradoras de la Oficina Planificación Institucional del ITCR - Abogado Defensor Estudiantil ITCR - Profesor y estadístico
Exploración de datos				- Colaboradora del Departamento de Orientación y Psicología - Arquitecto de software
Producto	Realizar la definición del producto de modelo predictivo incluyendo información administrativa como alcance, criterios de éxito, presupuesto, plazos, estructura de gestión, plan de progreso, beneficios del proyecto, requisitos legales y datos necesarios almacenados por el Tecnológico de Costa	- Grupo de discusión con el profesor tutor. - Revisión documental de referencias - Entrevista a profesor tutor.	- Definición del producto. - Base de organización establecida del modelo predictivo	- Arquitecto de software

Fase	Objetivo específico	Instrumentos	Variables	Sujetos de investigación.
	Rica y fuentes externas, así como nuevas posibles variables que aporten valor, para el establecimiento de una base organizativa que facilite su estimación y ejecución.			
Evaluación	Investigar la factibilidad del producto mediante evaluaciones que incluyen un análisis financiero, organizacional, de mercado, competencia y de riesgos; asegurando información relevante permitiendo su uso para la toma de decisiones sobre su viabilidad.	- Grupo de discusión con el profesor tutor. - Revisión documental de referencias. - Entrevista a profesor tutor. -Entrevista profunda a sujetos de investigación.	- Factibilidad del producto - Información de decisión sobre la viabilidad del modelo predictivo.	- Arquitecto de software - Abogado Defensor Estudiantil ITCR
Propuesta	Elaborar un plan de implementación del producto modelo predictivo, especificando las fases, hitos críticos y estrategias de evaluación, que proporcione un plan claro y estructurado que establezca una guía en su desarrollo en las operaciones del Tecnológico de Costa Rica.	- Grupo de discusión con el profesor tutor. - Revisión documental de referencias - Análisis documental de proyectos. - Entrevista a involucrados del TEC. - Grupo de discusión con el profesor tutor.	- Plan de implementación del producto. - Plan guía de desarrollo del proyecto.	- Arquitecto de software - Colaboradora del Departamento de Orientación y Psicología - Ex evaluadora e Investigadora - Colaboradoras de la Oficina Planificación Institucional del ITCR
Retrospectiva	No se cubre objetivo.	No se aplican instrumentos.	No se cubren variables.	No se relacionan sujetos.
Entrega y Cierre	No se cubre objetivo.	No se aplican instrumentos.	No se cubren variables.	No se relacionan sujetos.

Nota: Elaboración propia.

3.8. Hipótesis de la investigación

La formulación de hipótesis es fundamental para guiar la evaluación y análisis de viabilidad del modelo predictivo de deserción estudiantil en el TEC. Estas hipótesis surgen de la observación inicial del problema y la revisión de literatura existente, buscando validar o refutar suposiciones clave que impactarán el desarrollo futuro del sistema y se buscarán validar a lo largo del proyecto, estas son presentadas en la **Tabla *10***, mostrando un identificador, la hipótesis y una breve explicación.

Tabla 10
Hipótesis de la investigación.

ID	Hipótesis	Explicación
H1	Un modelo predictivo de deserción estudiantil tendría utilidad práctica para el TEC más allá de su precisión técnica.	Esta hipótesis surge de la necesidad de determinar si, independientemente de la capacidad técnica de predecir la deserción, un sistema de este tipo realmente aportaría valor a los procesos de retención estudiantil del TEC. Se busca validar que la predicción por sí misma no es suficiente sin un marco institucional que permita actuar sobre esas predicciones.
H2	La implementación exitosa de un modelo predictivo de deserción requiere más que recursos técnicos y datos disponibles.	Esta hipótesis plantea que, aunque se cuente con los recursos técnicos y datos necesarios, existen otros factores críticos como el apoyo institucional, la gobernanza de datos y los procesos organizacionales que son igual o más importantes para el éxito del proyecto.
H3	La predicción temprana de la deserción estudiantil es una herramienta efectiva para su reducción.	Esta hipótesis plantea que identificar anticipadamente estudiantes en riesgo de deserción permite implementar intervenciones más efectivas que reducen las tasas de abandono. Se busca validar que la predicción temprana marca una diferencia significativa en comparación con medidas reactivas.
H4	El TEC tiene una necesidad institucional crítica de reducir sus tasas de deserción actual.	Esta hipótesis surge de la observación inicial de que las tasas de deserción representan un problema significativo para el TEC, tanto en términos de su misión educativa como en el uso eficiente de recursos públicos
H5	Un modelo predictivo es una herramienta necesaria y adecuada para abordar el problema de deserción en el TEC.	Esta hipótesis cuestiona si un modelo predictivo es la herramienta más apropiada para abordar la deserción, considerando otras posibles alternativas o enfoques para el problema.

Nota: Elaboración propia.

3.9. Procedimiento Metodológico de la evaluación social

3.9.1. Contextualización del Proyecto en el Marco de la Inversión Pública

El proyecto propuesto en la sección 4.2 se enmarca en las iniciativas de modernización y fortalecimiento institucional del sector educativo costarricense. Como señala MIDEPLAN (2022, p.45), los proyectos de inversión pública deben evaluarse considerando "su contribución al bienestar de la sociedad en su conjunto", más allá de los indicadores financieros tradicionales.

En el contexto específico del TEC, el proyecto responde a la meta 4.3.3 del Plan Estratégico Institucional 2022-2026, que establece el objetivo de "reducir al 2026 la deserción de estudiantes de primer ingreso al menos en un 10%" (TEC, 2022). Sin embargo, es fundamental comprender que el proyecto evaluado no constituye una intervención directa sobre la deserción, sino una herramienta de apoyo para la toma de decisiones institucionales, por lo que, no genera un beneficio directo al ayudar a reducir la deserción, sino indirectamente, aportando a la decisiones enfocadas en los estudiantes críticos, intervenciones con mejor valoración y temprana, mejorando la inversión pública incurrida en estos temas y evitando el desperdicio de fondos públicos que supondría que un estudiante abandone sus estudios de forma definitiva.

3.9.2. Análisis de las Metodologías de Evaluación Social

La evaluación social de proyectos tradicionalmente ha empleado dos metodologías principales:

- Análisis Costo-Beneficio (ACB):
  - Requiere la monetización de beneficios sociales.
  - Necesita una clara atribución de impactos.
  - Depende de la disponibilidad de precios sociales.
- Análisis Costo-Efectividad (ACE):
  - Requiere medidas directas de efectividad.
  - Necesita parámetros comparables.
  - Depende de metas cuantificables.

Sin embargo, como señala Ortegón et al. (2015), estas metodologías presentan limitaciones significativas cuando se aplican a proyectos de fortalecimiento institucional, donde los beneficios son indirectos y difíciles de cuantificar, tal y como se resume en la **Tabla *11*** para el proyecto actualmente analizado.

Tabla 11
Análisis Comparativo de Metodologías de Evaluación Social.

Metodología	Ventajas	Limitaciones para el SPGDE
ACB	Permite comparación directa entre proyectos.	No es posible monetizar beneficios indirectos y directos generados por el proyecto.
ACE	Facilita evaluación de proyectos sociales.	Dificultad para medir efectividad directa, ya que no se encuentran históricos dentro del TEC.
Evaluación de Capacidades	Adecuada para proyectos institucionales.	Menor comparabilidad entre proyectos.

Nota: Elaboración propia.

3.9.3. Justificación del Enfoque Metodológico Seleccionado

Considerando las características específicas del proyecto a evaluar y siguiendo las recomendaciones de MIDEPLAN (2022) para proyectos de fortalecimiento institucional, se ha optado por un enfoque metodológico mixto que comprende:

- Análisis económico social:
  - Obtención de precios sombra.
  - Evaluación con medidas de rendimiento sociales.
  - Análisis de sensibilidad.
- Evaluación de fortalecimiento de capacidades institucionales:
  - Análisis de mejoras en procesos.
  - Evaluación de indicadores de capacidad operativa.
  - Valoración de integración interdepartamental.
- Análisis cualitativo de beneficios sociales:
  - Valoración de impactos indirectos.
  - Evaluación de mejoras en gestión institucional.
  - Análisis de contribución a objetivos estratégicos.

Este enfoque se alinea con lo que Contreras (2004) denomina "evaluación social de proyectos de fortalecimiento institucional", donde el énfasis se coloca en la mejora de capacidades organizacionales más que en impactos directos cuantificables, por lo que se propone esta misma como proceso a seguir en la evaluación social y debido a su difícil situación, en la cual es muy complejo mapear costos y beneficios exactos asociados.

Capítulo 4. Resultados

4.1. Estado actual de modelos predictivos en Costa Rica.

4.1.1. Introducción al análisis del estado actual

La deserción estudiantil representa uno de los desafíos más significativos que enfrentan las instituciones de educación superior en Costa Rica. Este fenómeno no solo impacta la trayectoria individual de los estudiantes, sino que también tiene implicaciones sustanciales para el desarrollo socioeconómico del país y la eficiencia en la utilización de recursos públicos en educación superior. En este contexto, el análisis del estado actual de los modelos predictivos de deserción se vuelve fundamental para comprender tanto los avances logrados como los desafíos pendientes en este campo.

El presente análisis se fundamenta en una revisión exhaustiva de la literatura existente, con especial énfasis en las experiencias documentadas en el contexto costarricense, complementada con entrevistas a expertos y análisis de casos específicos. Esta sección busca proporcionar una visión integral del estado actual de los esfuerzos predictivos en la educación superior costarricense, identificando patrones, desafíos y oportunidades para futuras implementaciones.

4.1.2. Evolución histórica de los modelos predictivos en Costa Rica

La historia de los esfuerzos predictivos en la educación superior costarricense, consideramos que puede dividirse en tres períodos principales:

1. **Período inicial (1970-1980)**:
    - Caracterizado por estudios descriptivos básicos.
    - Enfoque en estadísticas simples y análisis no tan variados.
    - Documentación limitada de metodologías predictivas.
2. **Período de pausa (1980-2010)**:
    - Marcado por una pausa significativa en la investigación formal.
    - Esfuerzos aislados sin sistematización institucional.
    - Desarrollo incipiente de bases de datos institucionales.
3. **Período contemporáneo (2010-presente)**:
    - Surgimiento de modelos predictivos basados en aprendizaje automático (*machine learning*).
    - Implementación de sistemas de datos integrados.
    - Desarrollo de aproximaciones más sofisticadas al problema.

Como señala Moreira-Mora (ver Apéndice H: Tabla de hallazgos de entrevista a Evaluadora e investigadora jubilada Tania Elena Moreira Mora.), "en el repositorio puedes encontrar información de deserción como del año 78, 79, 85... luego no se volvió a tocar el tema de deserción prácticamente a nivel institucional" (EH109, EH110). Esta observación refleja la discontinuidad histórica en los esfuerzos de investigación sobre deserción.

4.1.3. Estado actual de los modelos predictivos

4.1.3.1. Modelos identificados y sus características

1. Modelo TEC 2018-2019:

El modelo desarrollado por Solis et al. (2018) representa uno de los esfuerzos más significativos en el contexto costarricense. Este modelo alcanzó niveles notables de precisión:

- Sensibilidad: 93%.
- Tasa de verdaderos positivos: 94%.
- Metodología: Implementación de múltiples algoritmos de aprendizaje automático (*machine learning*).
    - Regresión logística.
    - Técnicas de árboles de decisión (*Random Forest*).
    - Máquinas de vectores de soporte.

La implementación de este modelo demostró la viabilidad técnica de las aproximaciones basadas en aprendizaje automático (*machine learning*) para la predicción de deserción en el contexto costarricense. Como señala textualmente el profesor Martín Solís en su entrevista (ver Apéndice G: Tabla de hallazgos de entrevista al Profesor Martín Solís Salazar.), "el modelo se llegó a implementar... cuando digo implementación es ponerlo en producción porque el modelo se desarrolló, pero, después hay que ponerlo en producción para que funcione" (EH148).

2. Iniciativas en desarrollo:

Actualmente, se identifican varios esfuerzos en desarrollo en distintas instituciones:

- Universidad Nacional (UNA): Desarrollo de un modelo predictivo específico para cursos de matemáticas
- TEC: Iniciativa de actualización del modelo anterior por parte de la OPI, reanudado en semana 8 del semestre 2 del año 2024, durante la investigación del proyecto actual.
- Otras universidades públicas: Esfuerzos preliminares en etapas de planificación.

4.1.4. Análisis de características técnicas predominantes

El análisis de los modelos predictivos de deserción estudiantil en el contexto costarricense revela una evolución significativa en las técnicas y metodologías empleadas. Como señalan Solis et al. (2018) y Fernández-Martín et al. (2018), que discuten estos aspectos técnicos en el contexto del TEC, la selección de algoritmos y fuentes de datos ha sido fundamental para el éxito de estas iniciativas predictivas, destacando la importancia de un enfoque técnico robusto y bien fundamentado.

4.1.4.1. Algoritmos y técnicas de modelado

En el contexto de la educación superior costarricense, la selección de algoritmos ha seguido las tendencias internacionales en ciencia de datos, pero, adaptándose a las necesidades y características específicas del entorno local. El algoritmo Bosque Aleatorio (*Random Forest*) ha demostrado ser particularmente efectivo, como se

evidencia en los resultados obtenidos por Solis et al. (2018), quien reporta "una precisión de predicción de hasta el 94% en la identificación de estudiantes en riesgo de deserción".

Las redes neuronales también han jugado un papel importante en estos modelos, especialmente en la capacidad de capturar patrones complejos en los datos estudiantiles. Su implementación ha permitido identificar relaciones no lineales entre variables que podrían pasar desapercibidas con técnicas más tradicionales. Sin embargo, como señalan los autores, "la complejidad de estos modelos puede presentar desafíos en términos de interpretabilidad y explicabilidad" (p. 5).

Los árboles de decisión, por su parte, han mantenido su relevancia debido a su capacidad de proporcionar resultados interpretables y fáciles de comunicar a los tomadores de decisiones. Esta característica es particularmente valiosa en el contexto institucional, donde la transparencia en el proceso de toma de decisiones es crucial.

4.1.4.2. Arquitectura de datos y fuentes de información

La arquitectura de datos en los modelos actuales refleja la complejidad del fenómeno de la deserción estudiantil. Los sistemas implementados integran múltiples fuentes de datos, creando una visión holística del estudiante y su contexto académico. Solis et al. (2018), identificaron la integración de sistemas como uno de los principales desafíos técnicos.

Los registros académicos históricos constituyen la columna vertebral de estos sistemas, proporcionando información crucial sobre el rendimiento y la progresión de los estudiantes. Estos datos se complementan con información socioeconómica, que ha demostrado ser un factor predictor significativo de la deserción. Como señalan los autores, "los factores como las dificultades económicas y los problemas de salud mental parecen tener un peso mayor en nuestro contexto en comparación con algunos estudios en países desarrollados" (p. 8).

La incorporación de datos de interacción con sistemas educativos representa una innovación importante en los modelos más recientes. Estos datos proporcionan información valiosa sobre el nivel de compromiso y participación de los estudiantes en su proceso educativo. Las variables demográficas y contextuales completan el conjunto de datos, permitiendo una comprensión más profunda de los factores que influyen en la deserción.

4.1.4.3. Innovaciones técnicas emergentes

Las últimas iteraciones de modelos predictivos han comenzado a incorporar técnicas más avanzadas, como el aprendizaje profundo y el procesamiento de lenguaje natural. Iam-On y Boongoen (2023), quienes presentaron un enfoque que integra datos estructurados y no estructurados, incluyendo análisis de sentimientos, representa una dirección prometedora para futuras investigaciones.

Además, se observa una tendencia hacia la automatización en la actualización y mantenimiento de los modelos. Esta automatización es crucial para mantener la relevancia y precisión de las predicciones a lo largo del tiempo. Como señalan los autores, la implementación de procesos automatizados de actualización y reentrenamiento del modelo es esencial para su efectividad a largo plazo.

4.1.5. Retos identificados

La implementación de modelos predictivos de deserción estudiantil en el contexto costarricense enfrenta diversos desafíos que requieren atención cuidadosa desde diferentes aristas, desde aspectos técnicos, organizacionales hasta éticos que se deben de abordar para lograr un éxito en este tipo de iniciativas.

4.1.5.1. Desafíos técnicos

La dimensión técnica representa uno de los obstáculos más significativos en la implementación efectiva de modelos predictivos. La fragmentación de sistemas y la calidad de datos emergen como preocupaciones principales en las instituciones de educación superior costarricense.

- **Integración de sistemas**

La integración de sistemas representa un desafío fundamental que afecta la capacidad de las instituciones para implementar modelos predictivos efectivos. Como evidencian Fernández-Martín et al. (2018) en su análisis multinomial del TEC, los principales aspectos incluyen:

- Fragmentación de bases de datos institucionales:
    - Existencia de múltiples sistemas aislados.
    - Dificultad en la comunicación entre departamentos.
    - Duplicación de datos en diferentes sistemas.
- Sincronización de información:
    - Desafíos en la actualización en tiempo real.
    - Inconsistencias temporales entre sistemas.
    - Necesidad de procesos de validación complejos.

La complejidad de esta integración se ve exacerbada por la existencia de sistemas de legado y la necesidad de mantener la operatividad continua durante los procesos de integración.

- **Calidad de datos**

La calidad de los datos representa otro desafío crítico que impacta directamente la eficacia de los modelos predictivos. Este aspecto se manifiesta en varios niveles:

- Consistencia histórica:
    - Variaciones en los métodos de recolección a lo largo del tiempo.
    - Cambios en las definiciones y categorías de datos.
    - Pérdida de información histórica relevante.
- Estandarización:
    - Diferencias en formatos y estructuras entre departamentos.

- Variabilidad en la codificación de información.
- Ausencia de protocolos unificados de registro.

4.1.5.2. Desafíos organizacionales

El aspecto organizacional representa una dimensión crítica en la implementación exitosa de modelos predictivos. Estos desafíos se manifiestan principalmente en dos áreas: recursos y capacidades, y aspectos estratégicos.

➢ **Recursos y capacidades**

La disponibilidad y gestión de recursos humanos y materiales constituye un desafío significativo. Los principales aspectos incluyen:

- Limitaciones en personal especializado:
  - Escasez de profesionales con experiencia en análisis de datos.
  - Necesidad de capacitación continua.
  - Rotación de personal clave.

Como evidencian los autores en su estudio, "la falta de personal especializado y la necesidad de capacitación continua representan obstáculos significativos para la implementación y mantenimiento de estos sistemas" (p. 7).

➢ **Aspectos estratégicos**

La dimensión estratégica es particularmente crítica para el éxito a largo plazo de estas iniciativas. Fernández-Martín et al. (2018), señalan que la falta de directrices institucionales claras obstaculiza la implementación efectiva de estos modelos. Los principales desafíos incluyen:

- Ausencia de políticas institucionales claras:
  - Falta de directrices específicas para el uso de modelos predictivos.
  - Indefinición de roles y responsabilidades.
  - Ausencia de métricas de éxito estandarizadas.
- Discontinuidad en iniciativas:
  - Cambios en prioridades institucionales.
  - Falta de seguimiento a largo plazo.
  - Pérdida de movimiento en proyectos clave.

4.1.5.3. Desafíos éticos

La dimensión ética en la implementación de modelos predictivos de deserción estudiantil requiere especial atención, ya que esta es una pieza fundamental para que estos sistemas provean su servicio a la población estudiantil de forma equitativa.

- **Privacidad y protección de datos**

La gestión de la privacidad y la protección de datos personales representa un desafío fundamental:

- Manejo de información sensible:
  - Datos personales y académicos.
  - Información socioeconómica.
  - Registros de comportamiento estudiantil.

Los autores destacan que "la privacidad y protección de datos estudiantiles fue mencionada por el 100% de los entrevistados como una preocupación significativa" (p. 6).

- **Equidad y justicia**

Las consideraciones de equidad y justicia en la aplicación de modelos predictivos son cruciales:

- Sesgos potenciales:
  - En datos históricos.
  - En algoritmos de predicción.
  - En interpretación de resultados.

Como señalan los autores, "es crucial desarrollar marcos éticos robustos para guiar el uso de modelos predictivos en educación superior" (p. 8).

4.1.6. Oportunidades identificadas

El análisis del contexto actual revela numerosas oportunidades para el desarrollo e implementación de modelos predictivos de deserción estudiantil, especialmente considerando los avances tecnológicos recientes y el aumento en la madurez digital de la institución en el manejo de datos.

4.1.6.1. Oportunidades tecnológicas

El panorama tecnológico actual ofrece posibilidades sin precedentes para el desarrollo de sistemas predictivos más sofisticados y efectivos. Las principales oportunidades se manifiestan en dos áreas principales: avances en IA y análisis de datos, y mejoras en la integración de sistemas.

- **Avances en IA y análisis de datos**

Los desarrollos recientes en inteligencia artificial y análisis de datos presentan oportunidades significativas para mejorar la precisión y efectividad de los modelos predictivos. Li et al. (2024) destaca que "la incorporación de técnicas de *deep learning* ha revolucionado la precisión de los modelos predictivos en educación superior, con estudios recientes reportando mejoras de entre 5-10% en comparación con métodos tradicionales".

- Nuevas técnicas de aprendizaje automático (*machine learning*):
  - Implementación de modelos de *deep learning*.

  - Desarrollo de sistemas de aprendizaje federado.
  - Aplicación de técnicas de procesamiento de lenguaje natural.

Los autores señalan que estas técnicas avanzadas "pueden permitir la identificación de patrones más sutiles y complejos en el comportamiento estudiantil" (p. 5).

- Mejoras en capacidad de procesamiento:
  - Mayor accesibilidad a recursos computacionales.
  - Optimización de algoritmos existentes.
  - Desarrollo de soluciones *cloud-native*.

➢ **Integración de sistemas**

La evolución de las arquitecturas de sistemas ofrece nuevas posibilidades para la integración efectiva de datos y servicios. Delen (2010) destaca la importancia de arquitecturas y sistemas integrados para el éxito de los modelos predictivos.

- Sistemas más interoperables:
  - Adopción de estándares abiertos.
  - Implementación de APIs REST.
  - Desarrollo de microservicios.
- Mejores arquitecturas de datos:
  - Diseños orientados a eventos.
  - Implementación de *data lakes.*
  - Desarrollo de flujos de trabajo (*pipeline*) de datos automatizados.

4.1.6.2. Oportunidades organizacionales

El contexto organizacional actual presenta oportunidades significativas para el desarrollo y fortalecimiento de capacidades institucionales en el ámbito de la analítica predictiva.

➢ **Desarrollo de capacidades**

La creciente consciencia sobre la importancia de la analítica de datos ha creado un ambiente propicio para el desarrollo de capacidades institucionales. Como señalan Fernández-Martín et al. (2018) y Solis et al. (2018), enfatizan la necesidad de desarrollar capacidades institucionales en análisis de datos para la predicción efectiva de la deserción.

- Formación de personal especializado:
  - Programas de capacitación específicos.

- Colaboraciones académicas.
- Certificaciones profesionales.

Los autores enfatizan que "la inversión en el desarrollo de capacidades internas es crucial para la sostenibilidad a largo plazo de estas iniciativas" (p. 8).

➢ **Mejoras en procesos**

La optimización de procesos institucionales representa una oportunidad significativa para mejorar la efectividad de los modelos predictivos:

- Estandarización de procedimientos:
  - Desarrollo de protocolos unificados.
  - Establecimiento de métricas comunes.
  - Creación de guías de mejores prácticas.
- Optimización de recursos:
  - Mejora en la asignación de recursos.
  - Automatización de tareas rutinarias.
  - Implementación de sistemas de monitoreo.

4.1.6.3. Oportunidades de colaboración interinstitucional

El panorama actual de la educación superior costarricense presenta oportunidades significativas para la colaboración entre instituciones. El estudio del Programa Estado de la Nación (2021) resalta que las alianzas entre universidades públicas facilitan el desarrollo de soluciones tecnológicas sostenibles, optimizando recursos y compartiendo experiencias exitosas.

➢ **Intercambio de conocimientos y experiencias**

La colaboración entre instituciones puede facilitar el intercambio de conocimientos y mejores prácticas:

- Compartir experiencias de implementación:
  - Lecciones aprendidas.
  - Casos de éxito.
  - Estrategias de mitigación de riesgos.
- Desarrollo conjunto de soluciones:
  - Creación de estándares comunes.
  - Compartir recursos tecnológicos.
  - Establecer comunidades de práctica.

Los autores señalan que "el intercambio de experiencias entre instituciones puede acelerar significativamente el desarrollo e implementación de soluciones efectivas" (p. 9).

4.1.6.4. Oportunidades de investigación y desarrollo

El campo de la predicción de deserción estudiantil presenta numerosas oportunidades para la investigación y el desarrollo de nuevas soluciones. Delen (2010) identifica diversas áreas prometedoras para la investigación futura en el campo de la predicción de deserción estudiantil.

- Desarrollo de nuevos algoritmos:
  - Modelos específicos para el contexto costarricense.
  - Integración de variables contextuales.
  - Mejora en la interpretabilidad de modelos.
- Evaluación de impacto:
  - Estudios longitudinales.
  - Análisis de efectividad de intervenciones.
  - Medición de retorno de inversión.

Los autores enfatizan que "la investigación continua en este campo es esencial para mejorar la efectividad de los modelos predictivos y su impacto en la retención estudiantil" (p. 10).

4.2. Definición del producto

4.2.1. Descripción del proyecto propuesto

El proyecto "Sistema de Predicción y Gestión de la Deserción Estudiantil" (SPGDE) es una iniciativa innovadora del Tecnológico de Costa Rica (TEC) que busca desarrollar e implementar un modelo avanzado de inteligencia artificial para predecir, priorizar y mitigar la deserción estudiantil. Este sistema utilizará técnicas de vanguardia en aprendizaje automático y análisis de datos para identificar a los estudiantes en riesgo de abandonar sus estudios, permitiendo así una intervención temprana, personalizada y eficiente.

El SPGDE se fundamenta en la comprensión de que la deserción estudiantil es un fenómeno complejo y multifactorial que puede ser predicho con un alto grado de precisión mediante el análisis de patrones en datos históricos y actuales. Como señalan Solis et al. (2018) en su estudio "*Perspectives to Predict Dropout in University Students with Machine Learning*", los modelos de aprendizaje automático pueden alcanzar una precisión de predicción de hasta el 91% en la identificación de estudiantes en riesgo de deserción.

El sistema empleará técnicas avanzadas de aprendizaje automático (*machine learning*), incluyendo posibilidades como *Adaptive Boosting* (*AdaBoost*) y *Bootstrap Aggregating* (*Bagging*), que han demostrado ser altamente efectivas en tareas de clasificación y predicción. Estas técnicas permitirán al modelo aprender de forma iterativa y mejorar continuamente su precisión a medida que se procesen más datos, además, de aportar una diferenciación a la competencia actual.

Los algoritmos y técnicas mencionados (como *AdaBoost* y *Bagging*) son propuestas preliminares basadas en la literatura y experiencias previas. Sin embargo, la selección final de los modelos específicos requerirá una evaluación exhaustiva durante la fase de implementación para determinar cuáles se adaptan mejor a las necesidades y características particulares del TEC

Un aspecto crucial del SPGDE será su capacidad para no solo identificar a los estudiantes en riesgo, sino también para priorizarlos. El sistema implementará un sistema de puntuación (*scoring*) que permitirá clasificar a los estudiantes según su nivel de riesgo de deserción. Este enfoque es esencial para optimizar la asignación de recursos, asegurando que se atiendan primero los casos más urgentes. Como señala la investigadora entrevistada, "identificar potenciales desertores no es suficiente sin acciones concretas para retenerlos" (Apéndice H: Tabla de hallazgos de entrevista a Evaluadora e investigadora jubilada Tania Elena Moreira Mora., EH140).

El SPGDE constará de varios componentes interconectados:

1. Módulo de Recopilación y Procesamiento de Datos: Este módulo se encargará de recopilar, limpiar y procesar datos de diversas fuentes dentro del TEC, incluyendo registros académicos, información socioeconómica, datos de uso de plataformas educativas en línea, y más.
2. Módulo de Análisis Predictivo: Aquí es donde se implementarán los algoritmos de aprendizaje automático (*machine learning*), incluyendo *AdaBoost* y *Bagging*, para generar predicciones sobre el riesgo de deserción de cada estudiante.
3. Sistema de Puntuación y Priorización: Este componente asignará un puntaje de riesgo a cada estudiante y los clasificará según la urgencia de intervención.
4. Módulo de Visualización y Reportes: Proporcionará *dashboards* interactivos y reportes detallados para que los administradores y educadores puedan entender fácilmente los resultados del análisis.
5. Módulo de Gestión de Intervenciones: Facilitará la planificación, seguimiento y evaluación de las intervenciones realizadas con los estudiantes en riesgo.

La implementación del SPGDE representa un paso significativo en los esfuerzos del TEC por mejorar la retención estudiantil y optimizar sus recursos educativos. Al combinar análisis de datos avanzados con un enfoque práctico de intervención, el TEC se posiciona a la vanguardia de la innovación en la gestión educativa.

4.2.2. Alcance

El SPGDE abarcará a todos los estudiantes de grado del TEC, en todas sus sedes y modalidades de estudio. El sistema se centrará en predecir el riesgo de deserción para cada semestre académico, permitiendo intervenciones oportunas. El alcance del proyecto incluye:

1. Desarrollo e implementación del sistema predictivo:
    - Creación de modelos de aprendizaje automático utilizando técnicas como *AdaBoost* y *Bagging*.
    - Implementación de un sistema de puntuación para priorizar casos de alto riesgo.
2. Integración con los sistemas existentes del TEC:

   - Conexión con el sistema de matrícula para obtener datos académicos en tiempo real.
   - Integración con el sistema de gestión estudiantil para acceder a información demográfica y socioeconómica.
   - Vinculación con la plataforma de aprendizaje en línea para recopilar datos sobre la participación de los estudiantes.

3. Desarrollo de interfaces de usuario:
   - Creación de *dashboards* para administradores y personal académico.
   - Diseño de un portal de autoservicio para estudiantes.
4. Capacitación del personal relevante en el uso del sistema:
   - Sesiones de formación para personal administrativo y docente.
   - Creación de manuales de usuario y guías de referencia rápida.
5. Implementación de un plan piloto en al menos dos escuelas del TEC:
   - Selección de escuelas piloto basada en criterios predefinidos.
   - Evaluación y ajuste del sistema basado en los resultados del piloto.
6. Desarrollo de protocolos de seguridad y privacidad de datos:
   - Implementación de medidas de encriptación y control de acceso.
   - Creación de políticas de uso y manejo de datos sensibles.

Queda fuera del alcance:

- La implementación de las intervenciones recomendadas (será responsabilidad de las unidades académicas y de apoyo estudiantil).
- La predicción de deserción para programas de posgrado.
- La integración con sistemas externos al TEC.
- El desarrollo de aplicaciones móviles.
- La automatización completa de las intervenciones (el sistema proporcionará recomendaciones, pero, las acciones serán decididas por el personal del TEC).

Fases futuras del proyecto que podrían considerarse después de la implementación inicial incluyen:

- Expansión del modelo para incluir programas de posgrado.
- Desarrollo de una aplicación móvil para estudiantes.

- Implementación de un módulo de simulación para evaluar el impacto de diferentes estrategias de intervención.
- Integración con sistemas de otras universidades para análisis comparativos.

4.2.3. Historia del proyecto

4.2.3.1. Historia y origen de la organización en la que se desarrollará el proyecto

El Tecnológico de Costa Rica (TEC) es una institución nacional autónoma de educación superior universitaria, dedicada a la docencia, la investigación y la extensión de la tecnología y las ciencias conexas para el desarrollo de Costa Rica. Fue creado mediante ley No. 4777 del 10 de junio de 1971 (TEC, 2022).

Desde su fundación, el TEC ha mostrado un compromiso constante con la excelencia académica y la innovación tecnológica. En las últimas décadas, la institución ha puesto un énfasis especial en el uso de tecnologías de la información para mejorar sus procesos académicos y administrativos, buscando siempre formas innovadoras de abordar los desafíos educativos.

4.2.3.2. Origen y antecedentes del proyecto propuesto

El proyecto SPGDE surge como respuesta a una preocupación presente en el Plan Estratégico Institucional 2022-2026 del Tecnológico de Costa Rica sobre los índices de deserción estudiantil en el TEC para su disminución al año 2026. La evolución de este proyecto se puede trazar a través de varios hitos importantes:

1. Estudios iniciales (1970-1980): Desde sus primeros años, el TEC realizó estudios descriptivos sobre la deserción. Según la investigadora entrevistada, "en el repositorio puedes encontrar información de deserción como del año 78, 79, 85" (Apéndice H: Tabla de hallazgos de entrevista a Evaluadora e investigadora jubilada Tania Elena Moreira Mora., EH109).
2. Pausa en la investigación (1980-2000): Hubo un período de relativa inactividad en la investigación sobre deserción. Como señala la investigadora, "luego no se volvió a tocar el tema de deserción prácticamente a nivel institucional" (Apéndice H: Tabla de hallazgos de entrevista a Evaluadora e investigadora jubilada Tania Elena Moreira Mora., EH110).
3. Estudios CONARE (2000-2010): Se realizaron investigaciones a nivel del Consejo Nacional de Rectores (CONARE), proporcionando una perspectiva más amplia del problema a nivel nacional.
4. Estudio descriptivo de Chinchilla-Brenes (2016): Este estudio revela que más del 52.7% de los estudiantes que abandonaron sus estudios en el TEC lo hicieron de manera definitiva, subrayando la urgencia del problema.
5. Proyecto pionero de *machine learning* (2018): Un equipo de investigadores del TEC, liderado por Solis et al., desarrolló un modelo predictivo de deserción utilizando técnicas de aprendizaje automático (*machine learning*). Este estudio demostró la viabilidad y el potencial de utilizar análisis de datos avanzados para abordar el problema de la deserción.
6. Implementación inicial (2018-2019): Como mencionó el profesor estadístico entrevistado, "Roberto hizo la implementación cuando digo implementación es ponerlo en producción porque el modelo se desarrolló,

pero, después hay que ponerlo en producción para que funcione" (Apéndice G: Tabla de hallazgos de entrevista al Profesor Martín Solís Salazar., EH148).

7. Desafíos de adopción (2019-2021): A pesar de la implementación técnica exitosa, surgieron desafíos en la adopción institucional. El profesor estadístico señaló que "al fin y al cabo nunca nadie, digamos, de los tomadores de decisión, se estableció como una estrategia de cómo usar el modelo" (Apéndice G: Tabla de hallazgos de entrevista al Profesor Martín Solís Salazar., EH132).

8. Propuesta de VIESA (2021): La Vicerrectoría de Vida Estudiantil y Servicios Académicos (VIESA) propuso la creación de un sistema integral de predicción y prevención de la deserción, basándose en las lecciones aprendidas de los esfuerzos anteriores.

9. Aprobación del Consejo Institucional (2022): El Consejo Institucional aprobó la propuesta de VIESA, dando origen al proyecto SPGDE actual ya que se aprobaron los proyectos anteriores.

10. Alineación con el Plan Institucional 2022-2026: El proyecto SPGDE surge como una respuesta directa a los objetivos establecidos en el Plan Institucional 2022-2026 del TEC. Específicamente, el plan establece como meta "reducir al 2026 la deserción de estudiantes de primer ingreso al menos en un 10%" (TEC, 2022). Esta meta se encuentra bajo el objetivo estratégico 4.3.3 del plan, que busca mejorar la permanencia y éxito académico de los estudiantes.

El proyecto SPGDE representa la culminación de estos esfuerzos históricos, combinando la experiencia acumulada con las últimas tecnologías para abordar de manera integral el desafío de la deserción estudiantil. Además, se alinea perfectamente con los objetivos estratégicos del TEC para el período 2022-2026, demostrando el compromiso institucional con la mejora continua y la aplicación de soluciones innovadoras a problemas persistentes en la educación superior.

4.2.4. Contribuyentes

Los principales contribuyentes al proyecto SPGDE se exponen en la **Tabla *12*** mostrando el nombre, rol y responsabilidades de cada uno.

Tabla 12
Contribuyentes del proyecto.

Contribuyente	Rol	Responsabilidades
Departamento de Tecnologías de Información del TEC.	Liderazgo técnico del proyecto, desarrollo e implementación del sistema	Diseño de la arquitectura del sistema, desarrollo del software, integración con sistemas existentes, y mantenimiento.
Departamento de Orientación y Psicología (DOP).	Asesoría en factores psicológicos y sociales de la deserción estudiantil.	Definición de variables psicosociales relevantes, interpretación de resultados, diseño de estrategias de intervención.
Oficina de Planificación Institucional (OPI).	Gestión de datos institucionales y análisis estadístico.	Proporcionar acceso a datos históricos, validar la calidad de los datos, asesorar en el análisis estadístico.

Contribuyente	Rol	Responsabilidades
Vicerrectoría de Vida Estudiantil y Servicios Académicos (VIESA).	Supervisión general del proyecto y enlace con la administración superior del TEC.	Asegurar la alineación del proyecto con las políticas institucionales, gestionar recursos, facilitar la colaboración interdepartamental.
Escuelas académicas del TEC.	Proveer información específica de cada carrera y validar los resultados del modelo.	Aportar conocimientos sobre factores de deserción específicos de cada carrera, participar en la fase de pruebas, implementar estrategias de intervención.
Expertos externos en ciencia de datos y aprendizaje automático.	Asesoría técnica especializada en modelos predictivos y análisis de datos.	Revisar y validar el diseño del modelo, proporcionar orientación sobre las mejores prácticas en IA y aprendizaje automático.

Nota: Elaboración propia.

La integración efectiva de estos contribuyentes será crucial para el éxito del proyecto. Se establecerá un comité directivo compuesto por representantes de cada grupo, que se reunirá regularmente para supervisar el progreso del proyecto y tomar decisiones clave.

Para la selección de expertos externos, se llevará a cabo un proceso de licitación pública, buscando profesionales con experiencia demostrada en proyectos similares en el ámbito educativo. Se dará preferencia a expertos con conocimiento del contexto costarricense y experiencia en instituciones de educación superior. Esta estructura de colaboración multidisciplinaria asegurará que el SPGDE se beneficie de una amplia gama de perspectivas y experiencias, maximizando su efectividad y relevancia para el TEC.

4.2.5. Estructura de gestión

La implementación exitosa del proyecto SPGDE requiere una estructura de gestión bien definida que facilite la colaboración interdepartamental y asegure una supervisión efectiva. Esta estructura se ha diseñado para aprovechar las fortalezas de los diferentes departamentos del TEC y garantizar una toma de decisiones ágil y fundamentada. A continuación, se detalla la estructura de gestión propuesta para el proyecto:

4.2.5.1. Comité Directivo del Proyecto

Este comité será el órgano de más alto nivel en la estructura de gestión del proyecto, responsable de la dirección estratégica y la supervisión general. Se compondrá por:

- Vicerrector(a) de Vida Estudiantil y Servicios Académicos (VIESA).
- Director(a) del Departamento de Orientación y Psicología (DOP).
- Director(a) de la Oficina de Planificación Institucional (OPI).
- Director(a) del Departamento de Admisión y Registro (DAR).
- Director(a) del Departamento de Tecnologías de Información y Comunicaciones (DATIC).

Dentro de sus responsabilidades se encuentra:

- Aprobar los objetivos y alcance del proyecto.

- Asegurar la alineación del proyecto con los objetivos estratégicos del TEC.
- Resolver conflictos de alto nivel.
- Aprobar cambios significativos en el alcance, presupuesto o cronograma del proyecto.
- Evaluar y aprobar los informes de progreso del proyecto.

4.2.5.2. Equipo de Gestión del Proyecto

Este equipo será responsable de la gestión diaria del proyecto y la coordinación entre los diferentes equipos de trabajo. Se compondrá por:

- Gerente de Proyecto (líder del equipo).
- Coordinador Técnico.
- Coordinador de Análisis de Datos.
- Representante del DOP.
- Representante de la OPI.

Dentro de sus responsabilidades se encuentra:

- Planificar y supervisar las actividades diarias del proyecto.
- Gestionar los recursos del proyecto.
- Identificar y gestionar los riesgos del proyecto.
- Preparar informes de progreso para el Comité Directivo.
- Coordinar la comunicación entre los diferentes equipos de trabajo.

4.2.5.3. Equipos de Trabajo Especializados

- Equipo de Desarrollo
  - Técnico Liderado por: Coordinador Técnico.
  - Miembros: Desarrolladores de software, especialistas en bases de datos, expertos en seguridad informática.
  - Responsabilidades: Desarrollo e implementación del sistema SPGDE, integración con sistemas existentes del TEC.
- Equipo de Análisis de Datos
  - Liderado por: Coordinador de Análisis de Datos.
  - Miembros: Analistas de datos, estadísticos, expertos en aprendizaje automático.
  - Responsabilidades: Desarrollo y mantenimiento del modelo predictivo, análisis de datos, generación de informes.
- Equipo de Apoyo Estudiantil
  - Liderado por: Representante del DOP.
  - Miembros: Psicólogos, orientadores, trabajadores sociales.
  - Responsabilidades: Diseño de estrategias de intervención, interpretación de resultados del modelo, capacitación de personal académico.
- Equipo de Gestión de Cambio y Comunicación
  - Liderado por: Miembro designado del equipo de gestión del proyecto.

- Miembros: Especialistas en comunicación, representantes de diferentes departamentos académicos.
- Responsabilidades: Gestión del cambio organizacional, comunicación interna y externa del proyecto, capacitación de usuarios.

4.2.5.4. Comité de Ética y Privacidad de Datos

Este comité se encargará de supervisar los aspectos éticos y de privacidad relacionados con el uso de datos estudiantiles en el proyecto SPGDE. Se compondrá por:

- Representante del Departamento Legal del TEC.
- Experto en ética de datos.
- Representante estudiantil.
- Miembro del equipo de desarrollo técnico.
- Miembro del equipo de análisis de datos.

Dentro de sus responsabilidades se encuentra:

- Desarrollar y supervisar las políticas de uso ético de datos.
- Revisar y aprobar los protocolos de manejo de datos sensibles.
- Asegurar el cumplimiento de las regulaciones de protección de datos.
- Abordar cualquier preocupación ética que pueda surgir durante el desarrollo y uso del SPGDE.

4.2.6. Duración del proyecto

El proyecto SPGDE (Sistema de Predicción y Gestión de la Deserción Estudiantil) se ha concebido como una iniciativa con una duración total de 4 años. Este período se divide en 6 fases principales:

4.2.6.1. Extracción, Transformación y Carga (ETL)

En esta etapa comprende actividades, dentro de sus fases clave se encuentra el análisis y planificación, extracción de datos, transformación de datos y carga de datos. Como primer paso para determinar la duración estimada, primero se realizó una identificación de las principales variables a tomar en cuenta para la predicción de la deserción estudiantil considerando estudios encontrados en el Estado del Arte tal y como se muestra en el Apéndice Q: Tabla de variables de predicción de deserción estudiantil encontradas.

Por otra parte, como bien se comentó y recomendó en las entrevistas realizadas, el modelo requiere de una actualización debido al cambio de entorno y condiciones provocadas por el surgimiento de la pandemia del COVID-19, por tanto, tomando en consideración este evento, se sugiere el estudio y valoración de nuevas variables mostradas en el Apéndice R: Tabla de variables de predicción de deserción estudiantil sugeridas.

Estas variables son consideradas supuestos bajo la hipótesis de ser útiles para la predicción, sin embargo, la M.Ed. Teresa Hernández Jiménez (Ver Apéndice W: Minuta 2 de entrevista a M.Ed. Teresa Hernández Jiménez), indica su aceptación para las mismas.

En resumen, de todas las fuentes de datos identificadas, en la **Tabla *13***, se muestra cada una de estas bajo una clasificación de su localización con respecto al Tecnológico de Costa Rica en interna o externa a esta entidad, además de indicar si llega a requerir un esfuerzo extra realizando un RPA para la extracción de la información,

cabe destacar que aquellas fuentes externas que no lo requieren, están sujetos a convenios que en caso de no ser logrado, puede ser necesario un RPA para este objetivo.

Tabla 13
Fuentes de información encontradas.

Identificador	Nombre de fuente	Tipo de Fuente	Requiere un RPA	Tipo de integración
1	SIGI	Interna	No	Base de Datos
2	Admisión y Registro	Interna	No	Base de Datos
3	Encuesta estudiantil	Interna	No	CSV con Jobs
4	Departamento de Becas	Interna	No	Base de Datos
5	DATIC	Interna	No	Base de Datos
6	Sistema de Biblioteca	Interna	No	Base de Datos
7	Vida Estudiantil	Interna	No	Base de Datos
8	DOP	Interna	No	Base de Datos
9	Éxito académico	Interna	No	CSV con Jobs
10	Oficina de Equidad de Género	Interna	No	CSV con Jobs
11	Departamento de Salud	Interna	No	Base de Datos
12	SUTEL	Externa	No	API REST
13	Ministerio de Educación Pública	Externa	No	API REST
14	TSE*	Externa	Si	RPA
15	CCSS	Externa	Si	RPA
16	Registro Nacional	Externa	Si	RPA
17	Ministerio de Hacienda	Externa	Si	RPA
18	Sistema Financiero Nacional	Externa	Si	RPA
19	SUGEF	Externa	Si	RPA
20	FONABE	Externa	No	API REST
21	IMAS	Externa	Si	RPA
22	Ministerio de Trabajo	Externa	No	API REST
23	INEC	Externa	No	API REST
24	Poder Judicial	Externa	Si	RPA

Nota: Elaboración propia. El TSE puede consumirse mediante un API con una alianza, sin embargo, con este método solo se puede acceder al padrón electoral de personas mayores a 18 años, por lo que se opta por un RPA en consecuencia, ya que el TEC mantiene población con una edad menor.

Tal y como se puede observar, se encontró un total de 24 fuentes de datos para extraer las variables recomendadas de las cuales 7 requieren un RPA, bajo el Juicio de Experto, en el Apéndice X: Minuta de entrevista a M.Sc. Ing. Juan Andrés Segreda Johanning., se indica que, en promedio, se puede tardar un total de 2 semanas por cada proceso de ETL a realizar en cada base de datos y 3 semanas para los que requieran realizar un RPA para extraer la información, por tanto, se necesitará la siguiente cantidad de tiempo:

$$(17\ ETL \cdot 2\ semanas) + (7\ RPA \cdot 3\ semanas) = 55\ semanas$$

$$55\ semanas\ \cdot 7\ días = 385\ días$$

$$385\ días\ \div 30\ días = 12.8\ meses\ \approx 1\ año\ y\ 1\ mes$$

Cabe destacar que este cálculo es tomando en cuenta que cada ETL es desarrollado de forma secuencial, sin embargo, con el fin de reducir el tiempo de desarrollo a un semestre, se recomienda la contratación de 2 desarrolladores trabajando paralelamente, por lo que la duración final de la etapa será:

$$12\ meses\ \div 2\ desarrolladores = \ 6\ meses$$

4.2.6.2. Caracterización y Selección del Modelo

En base a estudios similares como los realizados con apoyo el DOP en el Tecnológico de Costa rica, Solis et al. (2018), la caracterización y selección del modelo para predicción de deserción tomó aproximadamente 16 semanas (4 meses), divido en las etapas de:

- Análisis exploratorio y preparación (6 semanas).
- Evaluación de algoritmos (6 semanas).

Optimización y validación (4 semanas).

4.2.6.3. Entrenamiento y Pruebas

Esta etapa contará con las siguientes fases, estimadas según el criterio de experto y se fundamenta en investigaciones académicas recientes y experiencias documentadas en el campo de la analítica educativa:

1. Entrenamiento Inicial: Según Burgos et al. (2023) en su estudio sobre predicción de deserción universitaria, el entrenamiento inicial requiere un mínimo de 3 semanas debido a la necesidad de probar múltiples configuraciones. En su investigación con datos de 15,000 estudiantes, documentaron este período como óptimo para la fase inicial.
2. Evaluación y Ajuste: De acuerdo con el estudio longitudinal de Iam-On y Boongoen (2022) sobre predicción de deserción en universidades tailandesas, la fase de evaluación y ajuste requiere aproximadamente 3 semanas para garantizar la robustez del modelo. Este tiempo se validó en su investigación con datos de cinco años académicos.
3. Pruebas Finales: Según la investigación de Marbouti et al. (2023) sobre predicción temprana de deserción, establece que las pruebas finales requieren un mínimo de 2 semanas para garantizar la confiabilidad del modelo.

Como bien se puede observar, se alcanza un total de 8 semanas según estas estimaciones, aproximadamente 2 meses, dándole esto validez a la estimación realizada y recomendada por el Arquitecto de Software de la **Tabla *4*** (ver Apéndice X: Minuta de entrevista a M.Sc. Ing. Juan Andrés Segreda Johanning.).

4.2.6.4. Visualización

Como primer punto, para esta etapa se crearán *dashboards* para todas las escuelas del TEC que ofrecen carreras de grado y dependencias de estas, según la información oficial del TEC (2023) y el alcance definido serán:

1. Administración de Empresas.
2. Administración de Tecnología de Información.
3. Arquitectura y Urbanismo.
4. Biología.

5. Ciencias Naturales para el Desarrollo.
6. Ciencias del Lenguaje.
7. Computación.
8. Construcción.
9. Diseño Industrial.
10. Educación Técnica.
11. Electrónica.
12. Física.
13. Forestal.
14. Ingeniería Agrícola.
15. Ingeniería Agronegocios.
16. Ingeniería Ambiental.
17. Ingeniería en Computadores.
18. Ingeniería en Materiales.
19. Ingeniería en Producción Industrial.
20. Ingeniería Mecatrónica.
21. Matemática.
22. Química.
23. Seguridad Laboral e Higiene Ambiental.

Estos incluirán datos relacionados a la deserción estudiantil según cada una, desde las principales causas hasta las personas con mayor riesgo, con el fin de que estas puedan planificar contramedidas eficaces. Además, se crearán 4 *dashboards* generales:

1. *Dashboard* general de deserción.
2. *Dashboard* de factores de riesgo.
3. *Dashboard* de intervenciones.
4. *Dashboard* de seguimiento de estudiantes en riesgo.

La estimación de tiempo y recursos para el desarrollo de los 27 *dashboards* propuestos se fundamenta en diversos estudios y casos prácticos en el ámbito de la analítica educativa, que según Klerkx et al. (2017), el desarrollo de una visualización educativa efectivos requiere típicamente entre 1.5 y 2.5 días por cada *dashboard*, considerando las fases de:

- Diseño inicial (4-6 horas).

- Desarrollo y pruebas (6-8 horas).
- Validación con usuarios (2-4 horas).
- Ajustes y optimización (2-4 horas).

Por otra parte, Verbert et al. (2020) sugiere varios factores que pueden afectar el tiempo de desarrollo estimado para este objetivo, citando entre estos:

- Complejidad de integración de datos:
  - Múltiples fuentes de datos (sistemas académicos, financieros, etc.).
  - Necesidad de actualización en tiempo real.
  - Requerimientos de seguridad y privacidad.
- Requisitos de personalización:
  - Diferentes necesidades por escuela académica.
  - Diversos niveles de acceso y visualización.
  - Adaptación a diferentes roles de usuario.

Tomando en consideración lo anterior, se tomará como tiempo estimado los 2.5 días propuestos como el máximo, obteniendo como resultado la siguiente duración para la etapa de visualización:

$$2.5\ días\ \cdot 27\ dashboards = 67.5\ días\ \approx 2.5\ meses$$

Este resultado obtenido, es validado por proyectos similares implementados en otras universidades extranjeras, de las cuales se puede destacar los casos de:

- La Universidad de Purdue implementó *Course Signals*, un sistema de *dashboards* predictivos, en aproximadamente 3 meses para 20 departamentos (Arnold & Pistilli, 2012).
- La Universidad de Maryland desarrolló sus *dashboards* de retención estudiantil en 2.5 meses para 30 unidades académicas (Whitmer et al., 2016).

4.2.6.5. Puesta en marcha

La fase de Puesta en Marcha se estima en 3 meses basándose en las mejores prácticas de implementación de sistemas de analítica e inteligencia de negocios (Eckerson, 2017). Este tiempo se desglosa de la siguiente manera:

- Configuración del entorno de producción (2 semanas).
- Migración de datos y modelos (2 semanas).
- Pruebas de integración y aceptación (3 semanas).
- Capacitación de usuarios finales (2 semanas).
- Lanzamiento y monitoreo inicial (3 semanas).

4.2.6.6. Mantenimiento y Mejora Continua

La fase de Mantenimiento y Mejora Continua comenzará inmediatamente después de la puesta en marcha y se extenderá durante toda la vida útil del sistema, es decir, los dos años restantes después de su desarrollo. Esta fase incluirá:

- Mantenimiento Técnico (mensual, duración dinámica y continua según surjan eventos):
  - Monitoreo de rendimiento del sistema.
  - Actualizaciones de seguridad.

  - Respaldo de datos.
  - Resolución de incidencias técnicas.
- Actualización del Modelo (semestral, duración de 4 semanas según la etapa de Optimización y Evaluación de la sección Caracterización y Selección del Modelo):
  - Reentrenamiento del modelo con nuevos datos.
  - Evaluación de la precisión.
  - Ajustes de parámetros según sea necesario.
  - Validación de resultados.

Mejoras del Sistema (anual, duración de 8 semanas según lo estimado en la sección de

- Entrenamiento y Pruebas, ya que ambas etapas son comparables):
  - Evaluación de nuevas características potenciales.
  - Actualización de integraciones.
  - Optimización de procesos.
  - Incorporación de retroalimentación de usuarios.

Para una visualización detallada de la línea de tiempo del proyecto, incluyendo las actividades principales a lo largo del desarrollo del proyecto, se puede consultar el Apéndice J: Diagrama de Gantt de proyecto SPGDE. Es importante destacar que, si bien se ha establecido una duración de 4 años, se prevé que el SPGDE continúe operando y evolucionando más allá de este período, convirtiéndose en una herramienta integral y a largo plazo para la gestión de la deserción estudiantil en el TEC.

4.2.7. Plazos del proyecto

La planificación temporal del proyecto SPGDE se ha estructurado considerando las dependencias entre fases y la necesidad de asegurar resultados de calidad en cada etapa. El proyecto tiene una duración total de 4 años, distribuidos en seis fases principales que se ejecutan secuencialmente, con excepción de la fase de visualización que puede ejecutarse parcialmente en paralelo con la fase de entrenamiento, esto se puede visualizar en el diagrama de Gantt actualizado en el Apéndice J: Diagrama de Gantt de proyecto SPGDE este está separado en las siguientes fases y actividades representadas en la **Tabla *14***.

Cabe destacar que, esta duración de cuatro años está definida por dos iniciales de desarrollo, separado en semestres para su desarrollo en proyectos de graduación y otros dos años de mantenimiento, con el fin de evaluar su efectividad, oportunidad de mejora y cumplimiento de expectativas, determinando si el proyecto debe de seguir operando o sus operaciones cesan, en una evaluación al final del periodo.

Tabla 14
Fases y actividades del proyecto.

Fase	Actividad	Duración	Tareas
Extracción, Transformación y Carga (ETL)	Análisis y planificación	6 semanas	• Evaluación de fuentes de datos. • Diseño de arquitectura de datos. • Definición de procesos ETL.

Fase	Actividad	Duración	Tareas
	Extracción de datos	6 semanas	• Implementación de conexiones a fuentes. • Desarrollo de procesos de extracción. • Validación de datos extraídos.
	Transformación de datos	6 semanas	• Limpieza y normalización. • Integración de fuentes. • Validación de transformaciones.
	Carga de datos	6 semanas	• Implementación de procesos de carga. • Verificación de integridad. • Documentación de procesos.
Caracterización y Selección del Modelo	Análisis exploratorio y preparación	6 semanas	• Análisis estadístico de variables. • Identificación de patrones. • Preparación del conjunto de datos (*datasets*).
	Evaluación de algoritmos	6 semanas	• Pruebas de diferentes modelos. • Análisis comparativo. • Selección de candidatos.
	Optimización y validación	4 semanas	• Ajuste de parámetros. • Validación cruzada. • Documentación de resultados.
Entrenamiento y Pruebas	Entrenamiento inicial	3 semanas	• Entrenamiento con datos históricos. • Evaluación preliminar. • Ajustes iniciales.
	Evaluación y ajuste	3 semanas	• Análisis de rendimiento. • Refinamiento del modelo. • Validación de mejoras.
	Pruebas finales	2 semanas	• Pruebas exhaustivas. • Documentación de resultados. • Preparación para producción.
Visualización	Diseño inicial	2 semanas	• Diseño de interfaces. • Definición de *dashboards.* • Aprobación de prototipos.
	Desarrollo y pruebas	3 semanas	• Implementación de *dashboards.* • Pruebas funcionales. • Integración con datos.
	Validación con usuarios	2 semanas	• Pruebas de usabilidad. • Recopilación de retroalimentación. • Identificación de mejoras.
	Ajustes y optimización	2 semanas	• Implementación de mejoras. • Optimización de rendimiento. • Validación final.

Fase	Actividad	Duración	Tareas
Puesta en Marcha	Configuración del entorno de producción	2 semanas	• Preparación de infraestructura. • Configuración de servidores. • Verificación de requisitos.
	Migración de datos y modelos	2 semanas	• Transferencia de datos. • Despliegue de modelos. • Validación de integridad.
	Pruebas de integración y aceptación	3 semanas	• Pruebas *end-to-end*. • Validación de integración. • Verificación de requisitos.
	Capacitación de usuarios finales	2 semanas	• Preparación de materiales. • Sesiones de capacitación. • Evaluación de aprendizaje.
	Lanzamiento y monitoreo inicial	3 semanas	• Despliegue gradual. • Monitoreo de rendimiento. • Ajustes finales.
Mantenimiento y Mejora Continua	Mantenimiento técnico	Recurrente mensual	• Monitoreo de rendimiento del sistema. • Actualización de parches de seguridad. • Resolución de incidencias técnicas. • Respaldo y verificación de datos. • Optimización de consultas y procesos.
	Actualización del modelo	4 semanas (recurrente semestral)	• Reentrenamiento del modelo con datos nuevos. • Evaluación de precisión y métricas de rendimiento. • Ajuste de parámetros y variables. • Validación con usuarios clave. • Documentación de cambios y resultados.
	Mejoras del sistema	8 semanas (recurrente anual)	• Análisis de solicitudes de mejora de usuarios. • Evaluación de nuevas funcionalidades. • Implementación de mejoras aprobadas. • Pruebas de integración y aceptación. • Actualización de documentación.

Nota: Elaboración propia.

Es importante notar que, aunque se han establecido plazos específicos para cada actividad, el proyecto mantendrá la flexibilidad necesaria para ajustarse a las necesidades cambiantes del TEC y a cualquier desafío imprevisto que pueda surgir. La adherencia a estos plazos será monitoreada regularmente, con revisiones trimestrales durante el primer año y semestrales en los años siguientes. Esto permitirá realizar ajustes oportunos al cronograma si es necesario, asegurando que el proyecto se mantenga alineado con los objetivos del TEC y las mejores prácticas en análisis predictivo y retención estudiantil.

4.2.8. Plan de progreso

El plan de progreso para el proyecto SPGDE es una herramienta crucial para monitorear y evaluar el avance del proyecto a lo largo de sus 4 años de duración. Este plan establece una base clara y estructurada para medir el progreso frente a los objetivos y plazos predefinidos, facilitando la identificación temprana de desviaciones y la implementación oportuna de medidas correctivas.

4.2.8.1. Indicadores clave de progreso

Se han identificado los siguientes indicadores clave para medir el desempeño del proyecto:

1. Cumplimiento de hitos: Porcentaje de hitos alcanzados en las fechas planificadas.
2. Precisión del modelo: Nivel de precisión del modelo predictivo en la identificación de estudiantes en riesgo de deserción.
3. Adopción del sistema: Porcentaje de departamentos y escuelas del TEC que utilizan activamente el SPGDE.
4. Satisfacción del usuario: Nivel de satisfacción de los usuarios con el sistema SPGDE.
5. Impacto en la deserción: Reducción en la tasa de deserción estudiantil desde la implementación del SPGDE.
6. Cumplimiento presupuestario: Adherencia al presupuesto planificado.

4.2.8.2. Frecuencia de seguimiento

Para asegurar un monitoreo efectivo del progreso del proyecto, se establecerá la siguiente frecuencia de seguimiento:

1. Reuniones semanales del Equipo de Gestión del Proyecto.
2. Reuniones mensuales de revisión de progreso con los líderes de los Equipos de Trabajo Especializados.
3. Reuniones trimestrales del Comité Directivo del Proyecto para revisión estratégica.
4. Evaluaciones semestrales de los indicadores clave de progreso.
5. Revisiones anuales completas del proyecto, incluyendo una evaluación exhaustiva de todos los aspectos del SPGDE.

4.2.8.3. Herramientas y métodos de monitoreo

Para facilitar el seguimiento efectivo del progreso del proyecto, se utilizarán las siguientes herramientas y métodos:

1. Sistema de gestión de proyectos: Se implementará una herramienta de software para el seguimiento de tareas, plazos y recursos.
2. Tableros de control (*dashboards*): Se desarrollarán *dashboards* interactivos para visualizar el progreso en tiempo real de los indicadores clave.

3. Informes de estado: Se generarán informes mensuales y trimestrales detallando el progreso del proyecto, desafíos encontrados y acciones planificadas.

4. Encuestas de satisfacción: Se realizarán encuestas periódicas a los usuarios del SPGDE para medir su satisfacción y recopilar retroalimentación.

5. Análisis de métricas de deserción: Se llevarán a cabo análisis semestrales de las tasas de deserción y su evolución desde la implementación del SPGDE.

4.2.8.4. Plan de hitos y entregables

A continuación, en la **Tabla *15*** se presenta un plan detallado de hitos y entregables claves para cada fase del proyecto:

Tabla 15
Hitos y entregables del proyecto.

Fase	ID	Hito	Entregables
ETL	H1	Plan de ETL aprobado	• Catálogo de fuentes de datos. • Documentación de arquitectura ETL. • Scripts de transformación documentados. • Conjunto de datos preparado. • Manual de procesos ETL.
	H2	Completitud de extracciones	
	H2	Validación de transformaciones	
	H3	*Dataset* consolidado	
Caracterización	H4	Variables definidas	• Modelo entrenado y validado. • Documentación técnica del modelo. • Informes de pruebas. • Manual de operación del modelo.
	H5	Algoritmos evaluados	
	H6	Modelo base establecido	
Entrenamiento	H7	Primera iteración completada	• Modelo entrenado y validado. • Documentación técnica del modelo. • Informes de pruebas. • Manual de operación del modelo.
	H8	Ajustes validados	
	H9	Pruebas completadas	
Visualización	H10	Diseños aprobados	• Prototipos de interfaces. • *Dashboards* funcionales. • Manual de usuario. • Documentación técnica de visualizaciones.
	H11	*Dashboards* implementados	
	H12	Interfaces validadas	
Puesta en Marcha	H13	Entorno configurado	• Documentación de configuración. • Informes de migración. • Materiales de capacitación. • Manual de operaciones. • Informe final de proyecto.
	H14	Migración completada	
	H15	Integración validada	
	H16	Capacitación completada	
	H17	Sistema en producción	
Mantenimiento y Mejora Continua	H18	Procesos de mantenimiento establecidos	• Manual de procedimientos de mantenimiento. • Cronograma de mantenimiento preventivo. • Documentación de procesos de respaldo. • Plan de contingencia y recuperación.

Fase	ID	Hito	Entregables
			• Registro de incidencias y resoluciones.
	H19	Actualizaciones del modelo completadas	• Informes de reentrenamiento del modelo. • Documentación de ajustes y optimizaciones. • Métricas de rendimiento actualizadas. • Registro de cambios en variables y parámetros. • Validaciones de precisión y efectividad.
	H20	Mejoras anuales implementadas	• Análisis de retroalimentación de usuarios. • Documentación de nuevas funcionalidades. • Pruebas de aceptación de mejoras. • Actualizaciones de manuales de usuario. • Informe anual de mejoras y resultados.

Nota: Elaboración propia.

4.2.8.5. Gestión de cambios y adaptación

Reconociendo la naturaleza dinámica de un proyecto a largo plazo como el SPGDE, se implementará un proceso de gestión de cambios que permitirá:

1. Identificar y evaluar cambios necesarios en el alcance, cronograma o recursos del proyecto.
2. Analizar el impacto de los cambios propuestos en los objetivos y entregables del proyecto.
3. Aprobar o rechazar los cambios propuestos a través de un proceso estructurado.
4. Implementar y comunicar los cambios aprobados de manera efectiva.
5. Actualizar el plan de proyecto y los documentos relevantes para reflejar los cambios aprobados.

4.2.9. Herramientas de desarrollo

4.2.9.1. Selección de software de minería de datos

Para seleccionar el software de minería de datos más adecuado para el SPGDE, se realiza un análisis comparativo de las principales herramientas disponibles en el mercado en la **Tabla *16***. La selección se basó en el Cuadrante Mágico de Gartner para Plataformas de Ciencia de Datos y Aprendizaje Automático (Gartner, 2023), además, basándose en su popularidad en la industria y la academia, así como en su diversidad en términos de características y precios. RapidMiner y KNIME son ampliamente utilizados tanto en entornos académicos como empresariales, ofreciendo un buen equilibrio entre facilidad de uso y potencia. Weka es una opción popular en el ámbito académico debido a su naturaleza gratuita y su enfoque educativo. SAS Enterprise Miner se incluye como una opción de nivel empresarial para ofrecer un contraste con las soluciones más accesibles.

Tabla 16
Cuadro de características de software de minería de datos.

Software	Características	Ventajas	Desventajas	Precio
RapidMiner	- Interfaz gráfica intuitiva - Amplia gama de algoritmos - Integración con R y Python	- Fácil de usar para principiante - Potente para análisis avanzados	- Curva de aprendizaje para funciones avanzadas - Versión gratuita limitada	Versión gratuita limitada, planes empresariales desde $5000/año
KNIME	- Flujo de trabajo basado en nodos - Extensible con *plugins* - Integración con múltiples herramientas	- Código abierto y gratuito - Comunidad activa	- Interfaz puede ser abrumadora al principio - Rendimiento puede ser lento con grandes conjuntos de datos	Gratuito (versión *open source*), versión comercial disponible
Weka	- Colección de algoritmos de aprendizaje automático (*machine learning*) - Interfaz gráfica simple - Herramientas de preprocesamiento	- Gratuito y de código abierto - Bueno para aprendizaje	- Limitado para grandes conjuntos de datos - Menos opciones de visualización	Gratuito
SAS Enterprise Miner	- Suite completa de herramientas analíticas - Escalable para *big data* - Soporte empresarial	- Potente y completo - Bien establecido en la industria	- Costoso - Curva de aprendizaje pronunciada	Cotización personalizada, generalmente costoso

Nota: Información recopilada de Forbes Advisor (2023), Gartner (2022), y las páginas oficiales de los respectivos programas.

La **Tabla *17*** se ha diseñado para resaltar las características más relevantes en la selección de un software de minería de datos. La facilidad de uso es crucial para la adopción por parte de usuarios con diferentes niveles de experiencia. La variedad de algoritmos y la extensibilidad son importantes para asegurar que la herramienta pueda adaptarse a diferentes necesidades analíticas. La escalabilidad y el soporte para *Big Data* son cada vez más relevantes en la era del *Big Data*. Finalmente, el costo es un factor decisivo, especialmente para organizaciones con presupuestos limitados o proyectos de investigación académica.

Tabla 17
Comparativa de software de minería de datos.

Característica	RapidMiner	KNIME	Weka	SAS Enterprise Miner
Facilidad de uso	Alta	Media	Alta	Baja
Variedad de algoritmos	Alta	Alta	Media	Alta
Escalabilidad	Media	Media	Baja	Alta
Soporte para *big data*	Medio	Medio	Bajo	Alto
Extensibilidad	Alta	Alta	Media	Media
Costo	Gratuito	$39,900 anuales	$10.00 mensuales	$1,500 por usuario

Nota: Evaluación basada en Borges et al. (2018) y Kaur y Bawa (2019).

Basado en este análisis, para minería de datos se recomienda RapidMiner debido a que su versión gratuita, entre toda su amplia gama de funcionalidades, buena extensibilidad y costo nulo; ofrece un excelente equilibrio entre potencia y accesibilidad, siendo adecuado tanto para usuarios principiantes como avanzados y adecuado para no invertir en herramientas de pago según la situación económica de la institución, por tanto, se tendría un costo de $0.

4.2.9.2. Selección de herramienta de visualización

Para la selección de la herramienta de visualización, se realizó un análisis similar en la **Tabla *18***, considerando las necesidades específicas del proyecto SPGDE y la infraestructura existente en el TEC. Estos cuatro softwares de visualización de datos fueron seleccionados por su prominencia en el Cuadrante Mágico de Gartner para Plataformas de Análisis e Inteligencia de Negocios (Gartner, 2022). Representan un espectro de opciones que van desde soluciones más orientadas al usuario final (Tableau, Power BI) hasta plataformas más centradas en datos y orientadas a desarrolladores (QlikView, Looker). Esta selección permite una comparación equilibrada entre facilidad de uso, potencia analítica y flexibilidad.

Tabla 18
Cuadro de características de software de visualización.

Software	Características	Ventajas	Desventajas	Precio
Tableau	- Interfaz *drag-and-drop.* - Amplia variedad de visualizaciones - Conexión a múltiples fuentes de datos	- Muy intuitivo - Visualizaciones de alta calidad	- Puede ser costoso para equipos grandes - Limitaciones en customización avanzada	Desde $70/usuario/mes

Software	Características	Ventajas	Desventajas	Precio
Power BI	- Integración con productos Microsoft - Funcionalidades de BI y análisis - Actualizaciones frecuentes	- Asequible - Buena integración con Excel	- Curva de aprendizaje para funciones avanzadas - Algunas limitaciones en visualizaciones complejas	Desde $9.99/usuario/mes
QlikView	- Modelo asociativo de datos - Potentes capacidades de ETL - Análisis en memoria	- Rápido procesamiento de datos - Bueno para exploración de datos	- Interfaz menos intuitiva - Puede ser costoso para implementaciones grandes	Cotización personalizada
Looker	- Basado en SQL - Modelado de datos robusto - Enfoque en gobierno de datos	- Altamente personalizable - Bueno para equipos técnicos	- Requiere conocimientos de SQL - Puede ser complejo para usuarios no técnicos	Cotización personalizada

Nota: Información recopilada de Forbes Advisor (2023), Gartner (2022), y las páginas oficiales de los respectivos programas.

Tabla *19* se centra en aspectos clave para la selección de software de visualización de datos. La facilidad de uso es fundamental para la adopción generalizada en una organización. La variedad de visualizaciones impacta directamente en la capacidad de comunicar información de manera efectiva. La integración de datos es crucial en entornos con múltiples fuentes de información. Las capacidades analíticas son importantes para realizar análisis más profundos directamente en la herramienta de visualización. La colaboración es cada vez más relevante en entornos de trabajo distribuidos. El costo sigue siendo un factor decisivo en la selección de software.

Tabla 19
Comparativa de software de visualización.

Característica	Tableau	Power BI	QlikView	Looker
Facilidad de uso	Alta	Alta	Media	Media
Variedad de visualizaciones	Alta	Alta	Media	Media
Integración de datos	Alta	Alta	Alta	Alta
Capacidades analíticas	Media	Media	Alta	Alta
Colaboración	Alta	Alta	Media	Alta

Característica	Tableau	Power BI	QlikView	Looker
Costo	$75.00 por usuario al mes	$9.40 por usuario al mes	$825.00 al mes	$2,900 por mes

Nota: Evaluación basada en Forbes Advisor (2023) y Gartner (2022).

En base a este análisis, para visualización de datos, se recomienda Power BI por su equilibrio entre funcionalidad, facilidad de uso y costo. Power BI ofrece una amplia gama de capacidades de visualización y análisis a un precio competitivo, especialmente si la organización ya utiliza otros productos de Microsoft. Además, sus frecuentes actualizaciones aseguran que la herramienta se mantenga al día con las últimas tendencias en visualización de datos.

Para el presupuesto, se considera Power BI Pro que según Microsoft (2024) tiene un costo mensual de $9.40 por usuario, ya que incluye todas las características anteriormente descriptas y comparadas, este tendrá un costo anual de:

$$\$9.40 * 12\, meses\ = \$112.8 \approx ₡58{,}092\ (tipo\ de\ cambio\ 1\ USD\ =\ 515\ CRC)$$

Este costo es considerado anualmente, ya que se necesita mantener vigente con el fin de mantenimiento y despliegue en fases posteriores.

4.2.10. Presupuesto del proyecto

El presupuesto del Sistema de Predicción y Gestión de la Deserción Estudiantil (SPGDE) se ha desarrollado considerando todos los aspectos necesarios para su implementación exitosa en el Tecnológico de Costa Rica (TEC). A continuación, se presenta un desglose y justificación de cada componente del presupuesto incluyendo costos de personal, software, y capacitación para cada fase del proyecto SPGDE según corresponde.

Es importante notar que los costos presentados representan el escenario de contratación de nuevo personal dedicado exclusivamente al proyecto. Sin embargo, existen alternativas para optimizar estos costos:

1. Reasignación de recursos existentes:
    a. Los costos de personal podrían reducirse significativamente mediante la reasignación de horas de docentes y personal actual del TEC.
    b. Esta estrategia permitiría aprovechar el conocimiento institucional existente y reducir los costos de contratación.
2. Aprovechamiento de trabajos finales de graduación:
    a. Varios componentes del proyecto podrían desarrollarse como trabajos finales de graduación de estudiantes del TEC, según cada proyecto propuesto en la sección 4.7 con el objetivo de ser ejecutado por semestre.
    b. Esto no solo reduciría costos, sino que también proporcionaría oportunidades valiosas de investigación para los estudiantes.

Por lo tanto, el presupuesto presentado debe considerarse como un techo máximo, siendo posible una implementación más eficiente en términos de costos mediante estas estrategias.

4.2.10.1. Extracción, Transformación y Carga (ETL)

- Costos por concepto de capacitación:

En cuanto a recursos necesarios, se prevé una capacitación en base a la herramienta de minería de datos recomendada (RapidMiner), para calcular su costo, se realiza un promedio de cursos especializados en minería de datos en Costa Rica, esto debido a que, según el Profesor Luis Felipe Picado Valverde de la Escuela de Administración de Tecnologías de Información y el Programa Talento Digital, el TEC por política no acepta capacitaciones internas, por lo que, la capacitación debe ser externa. Se consultaron tres proveedores locales de capacitación en tecnologías de la información:

1. Academia Técnica CISCO: Curso de Minería de Datos y Big Data (40 horas) - ₡550,000
2. Universidad Cenfotec: Programa en Ciencia de Datos (60 horas) - ₡750,000
3. Corporación Algoritmos: Curso de KNIME para Ciencia de Datos (50 horas) - ₡650,000

Tomando un promedio simple de estos cursos y considerando que se necesitará capacitación para al menos dos personas, se estima un costo total de capacitación de:

$$\frac{₡550{,}000 + ₡750{,}000 + ₡650{,}000}{3} = ₡650{,}000 \cdot 2 \text{ desarrolladores} = ₡1{,}300{,}000$$

- Costos por concepto de salarios:

En cuanto al salario de los desarrolladores, se tiene un monto por ₡949,477.43 por mes (salario obtenido en el Anexo I. Desglose salarial Técnico en Tecnologías de Información y comunicación., según la consulta realizada al departamento de recursos humanos del TEC (ver Apéndice Y: Solicitud de información salarial TEC.), por lo que el monto total en la duración completa de esta etapa es de:

$$₡949{,}477.43 \cdot 2\ desarrolladores \cdot 6\ meses = ₡11{,}393{,}729.16$$

- Costos por concepto de licencias de software:

Seguidamente, como se indica en la Selección de software de minería de datos, RapidMiner mantendrá un costo por ₡0 anuales, gracias a su licencia gratuita.

Finalmente, el costo total como se resume en la **Tabla 20**, es por un monto de ₡12,693,729.16, tomando en consideración todos los puntos anteriores.

Tabla 20
Costo total de etapa ETL.

Concepto	Costo
Salarios	₡11,393,729.16
Capacitación	₡1,300,000

Licencia de software	₡0
Total	₡12,693,729.16

Nota: Elaboración propia.

4.2.10.2. Caracterización y Selección del Modelo

En esta fase, en función del tiempo estimado de 4 meses en el apartado Caracterización y Selección del Modelo de la sección Duración del proyecto, la cantidad mínima de recursos para ejecutarlo (1 Científico de datos y 1 Especialista en estadística) según lo visto en estudios similares como los realizados con apoyo el DOP en el Tecnológico de Costa rica, Solis et al. (2018) y asumiendo un salario de ₡949,477.43 por mes (salario obtenido en el Anexo I. Desglose salarial Técnico en Tecnologías de Información y comunicación.), se tiene un costo total por:

$$₡949{,}477.43 \ * \ 2\, recursos \ * \ 4\, meses \ = \ ₡7{,}595{,}819.44$$

4.2.10.3. Entrenamiento y Pruebas

En base en la estimación de 2 meses de desarrollo por análisis de históricos y criterio de experto en el apartado

Entrenamiento y Pruebas de la sección Duración del proyecto, junto a la recomendación y estimación dada por el Arquitecto de Software de la **Tabla *4*** (ver Apéndice X: Minuta de entrevista a M.Sc. Ing. Juan Andrés Segreda Johanning.) de mantener 2 recursos para esta fase y un salario de ₡949,477.43 por mes (salario obtenido en el Anexo I. Desglose salarial Técnico en Tecnologías de Información y comunicación.), esta fase conlleva a un costo por:

$$₡949{,}477.43 \ \cdot \ 2\, recursos \ \cdot \ 2\, meses \ = \ ₡3{,}797{,}909.72$$

4.2.10.4. Visualización

Para esta fase se estima un tiempo total de desarrollo por 2.5 meses según lo previsto en el apartado de Visualización en la sección Duración del proyecto, por otra parte, asumiendo un salario de ₡949,477.43 por mes (salario obtenido en el Anexo I. Desglose salarial Técnico en Tecnologías de Información y comunicación.), se tiene que el costo de la etapa por concepto de personal es de:

$$₡949{,}477.43 \ \cdot \ 2.5\, meses \ = \ ₡2{,}373{,}693.575$$

Por otra parte, se mantiene un costo de licencia para un recurso de ₡58,092 según lo previsto en la sección de Selección de herramienta de visualización. Tomando en consideración, estos datos, el costo total de esta fase es de:

$$₡2{,}373{,}693.575 + ₡58{,}092 \ = \ ₡2{,}431{,}785.575$$

4.2.10.5. Puesta en Marcha

Seguidamente, esta fase tiene una estimación de un tiempo total de desarrollo por 3 meses según lo previsto en el apartado de Puesta en marcha en la sección Duración del proyecto, por otra parte, asumiendo un salario de

₡949,477.43 por mes (salario obtenido en el Anexo I. Desglose salarial Técnico en Tecnologías de Información y comunicación.), para un recurso se tiene que el costo de la etapa por concepto de personal es de:

$$₡949{,}477.43 \cdot 3\ meses = ₡2{,}848{,}432.29$$

4.2.10.6. Mantenimiento y Mejora Continua

Por último, la fase con mayor duración ya que incluiría los últimos 2 años del proyecto, según lo recomendado por el Arquitecto de Software de la **Tabla *4*** (ver Apéndice X: Minuta de entrevista a M.Sc. Ing. Juan Andrés Segreda Johanning.), mantendría dos recursos, uno para responder frente a eventualidades técnicas del sistema y apoyar en las actualizaciones, mientras que otro se mantendría en las evaluaciones semestrales y anuales programadas, sin embargo, estos no estarían trabajando al mismo tiempo, ya que manejarían tareas excluyentes, por tanto, se asumirá solamente un recurso en el cálculo con el total de tiempo de 3 meses máximo por año de mantenimiento, según lo visto en la duración de esta etapa en el peor de los casos. Además, asumiendo un salario de ₡949,477.43 por mes (salario obtenido en el Anexo I. Desglose salarial Técnico en Tecnologías de Información y comunicación.), se tiene que el costo por concepto de salarios es de:

$$\left((₡949{,}477.43 \cdot 1\ recursos\ simultaneo) \cdot 3\ meses\right) \cdot 2\ años = ₡5{,}696{,}864.58$$

Por otra parte, manteniendo un costo de licencia de minería de datos por ₡0 anuales como se indica en la Selección de software de minería de datos y ₡58,092 anuales de licencia de visualización según lo previsto en la sección de Selección de herramienta de visualización, se tiene un costo por concepto de licencias por:

$$(₡0 + ₡58{,}092\ ) \cdot 2\ años = ₡116{,}184$$

Finalmente, el costo total de la fase de mantenimiento asciende a:

$$₡5{,}696{,}864.58 + ₡116{,}184 = ₡5{,}813{,}048.58$$

4.2.10.7. Resumen del presupuesto

La ***Tabla 21*** proporciona una visión general de los gastos totales del proyecto en sus dos años. El costo total del desarrollo del proyecto en sus dos años sin contar mantenimiento es de ₡29,367,676.185, distribuido entre las diferentes categorías mencionadas anteriormente.

Tabla 21

Resumen de costos totales del proyecto.

Año	Fase	Concepto	Costo
Año 1	Extracción, Transformación y Carga (ETL)	Salarios	₡11,393,729.16
		Capacitación	₡1,300,000
		Licencias	₡0
		Total	₡12,693,729.16
	Caracterización y Selección del Modelo	Salarios	₡7,595,819.44
		Total	₡7,595,819.44
		Anualidad total	**₡20,289,548.6**
Año 2	Entrenamiento y Pruebas	Salarios	₡3,797,909.72
		Total	₡3,797,909.72

Año	Fase	Concepto	Costo
	Visualización	Salarios	₡2,373,693.575
		Licencias	₡58,092
		Total	₡2,431,785.575
	Puesta en Marcha	Salarios	₡2,848,432.29
		Total	₡2,848,432.29
		Anualidad total	**₡9,078,127.585**
Año 3	Mantenimiento y Mejora Continua	Salarios	₡2,848,432.29
		Licencias	₡58,092
		Anualidad total	**₡2,964,616.29**
Año 4	Mantenimiento y Mejora Continua	Salarios	₡2,848,432.29
		Licencias	₡58,092
		Anualidad total	**₡2,964,616.29**
Costo total del proyecto			**₡35,296,908.765**

Nota: Elaboración propia.

Esta inversión de ₡35,296,908.765 a lo largo de los 4 años representa el compromiso del TEC con la mejora de la retención estudiantil y la optimización de sus recursos educativos. Aunque el monto puede parecer significativo, el potencial retorno de la inversión en términos de mejora en la retención estudiantil, optimización de recursos y mejora de la reputación institucional justifica ampliamente este gasto.

La gestión cuidadosa de este presupuesto, junto con evaluaciones periódicas de su efectividad, será crucial para asegurar que el SPGDE cumpla con sus objetivos y proporcione un valor significativo al TEC y a su comunidad estudiantil, también cabe destacar que en el proceso se pueden encontrar recursos que ya mantenga el TEC y no requieran esta inversión por lo que los costos mencionados anteriormente, pueden reducirse de forma significativa.

4.2.11. Beneficios no financieros

El proyecto conlleva una serie de beneficios no financieros significativos. Estos beneficios intangibles son cruciales para justificar la inversión en el proyecto y demostrar su valor estratégico para la institución.

4.2.11.1. Mejora en la toma de decisiones institucionales

El SPGDE proporcionará a los administradores y educadores del TEC información detallada y en tiempo real sobre los factores que influyen en la deserción estudiantil. Esto permitirá:

- Toma de decisiones más informada y basada en datos.
- Capacidad para identificar tendencias y patrones en el comportamiento estudiantil.
- Mejora en la asignación de recursos para programas de retención y apoyo estudiantil.

Como señala la investigadora entrevistada, "el proyecto fue pionero en el uso de modelos predictivos para deserción en universidades públicas de Costa Rica" (Apéndice H: Tabla de hallazgos de entrevista a Evaluadora e investigadora jubilada Tania Elena Moreira Mora., EH133), lo que posiciona al TEC a la vanguardia en este campo.

4.2.11.2. Aumento en la retención y éxito estudiantil

La implementación del SPGDE tiene el potencial de:

- Identificar tempranamente a los estudiantes en riesgo de deserción.
- Facilitar intervenciones oportunas y personalizadas.
- Mejorar las tasas de retención y graduación.
- Aumentar la satisfacción y el éxito académico de los estudiantes.

Este beneficio se alinea directamente con el objetivo institucional mencionado en el Plan Institucional 2022-2026 del TEC de "reducir al 2026 la deserción de estudiantes de primer ingreso al menos en un 10%" (TEC, 2022).

4.2.11.3. Fortalecimiento de la imagen institucional

El desarrollo e implementación exitosa del SPGDE puede:

- Posicionar al TEC como líder en innovación educativa y uso de tecnologías en esta área.
- Mejorar la percepción pública del compromiso del TEC con el éxito estudiantil.
- Aumentar el atractivo del TEC para futuros estudiantes y sus familias.

4.2.11.4. Mejora en la eficiencia operativa

El SPGDE tiene el potencial de:

- Optimizar los procesos de apoyo estudiantil.
- Reducir la carga administrativa asociada con la gestión de casos de deserción.
- Mejorar la coordinación entre diferentes departamentos y servicios estudiantiles.

4.2.11.5. Generación de conocimiento valioso

El proyecto SPGDE permitirá:

- Obtener descubrimientos profundos sobre los factores que influyen en la deserción estudiantil.
- Contribuir a la literatura académica sobre retención estudiantil y análisis predictivo en educación superior.
- Facilitar la colaboración con otras instituciones educativas y de investigación.

4.2.11.6. Desarrollo de capacidades institucionales

La implementación del SPGDE conllevará:

- Mejora de las habilidades del personal en análisis de datos y uso de tecnologías en esta área.
- Fomento de una cultura institucional basada en datos.
- Desarrollo de capacidades internas en inteligencia artificial y aprendizaje automático.

4.2.11.7. Impacto social positivo

El éxito del SPGDE puede tener un impacto social más amplio:

- Contribuir a la reducción de las tasas de deserción en la educación superior a nivel nacional.
- Mejorar las oportunidades educativas y profesionales para los estudiantes.
- Apoyar el desarrollo económico y social del país a través de una fuerza laboral más educada.

4.2.11.8. Potencial para la transferencia de conocimientos

El proyecto SPGDE tiene el potencial de:

- Servir como modelo para otras instituciones educativas en Costa Rica y la región.
- Facilitar la colaboración interinstitucional en temas de retención estudiantil.
- Contribuir al posicionamiento del TEC como referente en innovación educativa.

Estos beneficios no financieros, aunque difíciles de cuantificar en términos monetarios, representan un valor significativo para el TEC y justifican la inversión en el proyecto SPGDE. Como mencionó el profesor estadístico entrevistado, "el modelo se llegó a implementar y digamos Roberto hizo la implementación cuando digo implementación es ponerlo en producción porque el modelo se desarrolló, pero, después hay que ponerlo en producción para que funcione" (Apéndice G: Tabla de hallazgos de entrevista al Profesor Martín Solís Salazar., EH148), lo que subraya la importancia de llevar estos beneficios del ámbito teórico al práctico.

La realización de estos beneficios requerirá un compromiso sostenido con el proyecto, una implementación efectiva y una evaluación continua para asegurar que el SPGDE cumpla con sus objetivos y genere el máximo valor posible para el TEC y su comunidad estudiantil.

4.2.12. Criterios de éxito

La definición de criterios de éxito claros y medibles es fundamental para evaluar el impacto y la efectividad del proyecto SPGDE. Estos criterios no solo proporcionan una guía para el desarrollo e implementación del sistema, sino que también establecen un marco para la evaluación continua y la mejora del proyecto. Los criterios de éxito abarcan diversos aspectos, desde el rendimiento técnico del sistema hasta su impacto en la reducción de la deserción estudiantil y su adopción por parte de la comunidad del TEC.

Es importante destacar que estos criterios están alineados con los objetivos estratégicos del TEC, particularmente con la meta establecida en el Plan Institucional 2022-2026 de "reducir al 2026 la deserción de estudiantes de primer ingreso al menos en un 10%" (TEC, 2022). Además, se han diseñado teniendo en cuenta las lecciones aprendidas de proyectos anteriores y las necesidades específicas identificadas en las entrevistas con personal clave del TEC.

A continuación, en la **Tabla *22*** se presentan los criterios de éxito para el proyecto SPGDE, cada uno acompañado de su objetivo específico y el criterio de cumplimiento correspondiente.

Tabla 22
Criterios de éxito del proyecto SPGDE.

Nombre	Objetivo	Criterio de cumplimiento
Precisión del modelo.	El modelo predictivo debe alcanzar una precisión de al menos 85% en la identificación de estudiantes en riesgo de deserción.	Se realizarán pruebas con datos históricos y se medirá la precisión del modelo utilizando métricas estándar como F1-score, precisión y sensibilidad.
Adopción del sistema	Al menos el 80% de las escuelas del TEC deben utilizar activamente el SPGDE para el final del primer año de implementación.	Se llevará un registro de uso por escuela y se realizarán encuestas de satisfacción y utilidad entre los usuarios.
Reducción de la deserción	Se espera una reducción del 10% en la tasa de deserción institucional dentro de los dos años siguientes a la implementación completa del sistema.	Se comparará la tasa de deserción antes y después de la implementación del SPGDE, controlando por otros factores que puedan influir en la deserción.
Satisfacción del usuario	Al menos el 75% de los usuarios del sistema (personal administrativo y docente) deben reportar satisfacción con la usabilidad y utilidad del SPGDE.	Se realizarán encuestas de satisfacción semestrales y se analizarán los resultados.
Integración de datos	El sistema debe lograr una integración exitosa con al menos el 95% de las fuentes de datos relevantes dentro del TEC.	Se realizará un inventario de fuentes de datos y se verificará la integración exitosa de cada una.
Tiempo de respuesta	El sistema debe ser capaz de generar predicciones y alertas en tiempo real, con un tiempo de respuesta no mayor a 5 segundos para consultas individuales.	Se realizarán pruebas de rendimiento y se monitoreará el tiempo de respuesta en condiciones de uso real.
Escalabilidad	El sistema debe ser capaz de manejar un aumento del 50% en el volumen de datos sin degradación significativa del rendimiento.	Se realizarán pruebas de carga simulando un aumento en el volumen de datos y se medirá el impacto en el rendimiento.
Protección de datos	Cumplimiento total con las regulaciones de protección de datos aplicables, sin incidentes de seguridad reportados durante el primer año de operación.	Se realizarán auditorías de seguridad regulares y se mantendrá un registro de incidentes de seguridad.

Nota: Elaboración propia.

Estos criterios de éxito proporcionan un marco integral para evaluar el desempeño del SPGDE. Abarcan aspectos técnicos como la precisión del modelo y el tiempo de respuesta, aspectos operativos como la adopción del sistema y la satisfacción del usuario, y aspectos estratégicos como la reducción de la deserción. Además, se han incluido criterios relacionados con la protección de datos y la escalabilidad del sistema, reconociendo la importancia de estos aspectos en un entorno educativo cada vez más digitalizado.

La medición y seguimiento continuo de estos criterios permitirá al equipo del proyecto y a los líderes del TEC evaluar el progreso del SPGDE, identificar áreas de mejora y tomar decisiones informadas sobre futuras iteraciones o expansiones del sistema. Esto asegurará que el SPGDE no solo cumpla con sus objetivos iniciales, sino que también evolucione para satisfacer las necesidades cambiantes de la institución y sus estudiantes.

4.2.13. Requerimientos legales y normativos

El proyecto SPGDE debe adherirse a diversas leyes, normas, decretos y reglamentos para garantizar su legalidad y cumplimiento con las regulaciones vigentes. Como se muestra en el Apéndice K: Tabla de requerimientos legales del proyecto, se presenta una matriz de requisitos legales que el proyecto debe acatar para ejecutarse cumpliendo el aspecto legal, tales como:

- Ley de protección de la persona frente al tratamiento de sus datos personales.
- Normas técnicas para la gestión y el control de las Tecnologías de Información (N-2-2007-CO-DFOE)
- Ley General de Control Interno.
- Ley de Certificados, Firmas Digitales y Documentos Electrónicos.
- Ley de Derechos de Autor y Derechos Conexos.
- Código de Trabajo.
- Ley Orgánica del Instituto Tecnológico de Costa Rica.
- Reglamento a la Ley de Protección de la Persona frente al Tratamiento de sus Datos Personales.

Es fundamental que el equipo del proyecto SPGDE se mantenga actualizado sobre cualquier cambio en estas leyes y reglamentos, así como cualquier nueva legislación que pueda surgir durante el desarrollo y la implementación del sistema. Además, se recomienda realizar revisiones periódicas de cumplimiento legal para asegurar que el SPGDE se mantenga en línea con todas las regulaciones aplicables.

El Comité de Ética y Privacidad de Datos, mencionado en la estructura de gestión del proyecto, desempeñará un papel crucial en la supervisión del cumplimiento de estos requisitos legales y normativos. Este comité deberá trabajar en estrecha colaboración con el departamento legal del TEC para interpretar y aplicar correctamente estas regulaciones en el contexto del SPGDE.

4.3. Estudio de mercado

El estudio de mercado para el Sistema de Predicción y Gestión de la Deserción Estudiantil (SPGDE) se enfoca en analizar la demanda, oferta, precio y comercialización de este tipo de soluciones en el contexto de la educación superior costarricense, específicamente para el Tecnológico de Costa Rica (TEC).

4.3.1. Análisis de demanda

El análisis de demanda es fundamental para entender las necesidades reales del TEC en cuanto a un sistema de predicción y gestión de la deserción estudiantil. Este análisis nos permitirá identificar quiénes serán los principales usuarios del sistema, qué expectativas tienen, y cómo el SPGDE puede satisfacer estas necesidades de manera efectiva.

4.3.1.1. Consumidores

Los consumidores del SPGDE se dividen en dos categorías principales: directos e indirectos. Esta distinción es crucial para entender cómo diferentes grupos dentro del TEC interactuarán con y se beneficiarán del sistema.

1) Consumidores directos:

   a) Departamento de Orientación y Psicología (DOP) del TEC: El DOP será uno de los principales usuarios del SPGDE. Este departamento utilizará el sistema para:

      i) Identificar tempranamente a estudiantes en riesgo de deserción.

      ii) Diseñar intervenciones personalizadas basadas en los perfiles de riesgo.

      iii) Monitorear la efectividad de sus programas de apoyo estudiantil.

      iv) Asignar recursos de manera más eficiente, centrándose en los estudiantes con mayor necesidad.

   b) Vicerrectoría de Vida Estudiantil y Servicios Académicos (VIESA): VIESA empleará el SPGDE para:

      i) Formular políticas más efectivas de retención estudiantil.

      ii) Evaluar el impacto de los programas de apoyo existentes.

      iii) Planificar estratégicamente la asignación de recursos para servicios estudiantiles.

      iv) Desarrollar nuevas iniciativas basadas en las tendencias identificadas por el sistema.

   c) Escuelas académicas del TEC: Cada escuela académica podrá utilizar el SPGDE para:

      i) Ajustar sus estrategias de enseñanza basándose en los patrones de riesgo identificados.

      ii) Implementar sistemas de alerta temprana específicos para sus estudiantes.

      iii) Personalizar el apoyo académico según las necesidades individuales de los estudiantes.

      iv) Evaluar la efectividad de sus planes de estudio y metodologías de enseñanza.

   d) Oficina de Planificación Institucional (OPI): La OPI integrará los datos del SPGDE en sus procesos de:

      i) Planificación estratégica institucional.

      ii) Evaluación de desempeño institucional.

      iii) Toma de decisiones basada en datos para la asignación de recursos.

      iv) Generación de informes y análisis para la alta dirección del TEC.

   e) Departamento de Admisión y Registro: Este departamento utilizará el SPGDE para:

      i) Optimizar los procesos de matrícula.

      ii) Mejorar el seguimiento académico de los estudiantes.

iii) Ajustar las políticas de admisión basándose en los factores de riesgo identificados.

iv) Proporcionar información más precisa a los estudiantes sobre sus trayectorias académicas.

2) Consumidores indirectos:
   a) Estudiantes del TEC: Aunque no interactuarán directamente con el sistema, los estudiantes se beneficiarán de:
      i) Intervenciones más oportunas y personalizadas.
      ii) Mejor acceso a recursos de apoyo académico y personal.
      iii) Una experiencia educativa más adaptada a sus necesidades individuales.
   b) Profesores y personal administrativo: Los profesores y el personal administrativo se beneficiarán de:
      i) Información más detallada sobre el perfil de riesgo de sus estudiantes.
      ii) Herramientas para adaptar sus métodos de enseñanza y apoyo.
      iii) Mayor capacidad para identificar y abordar problemas académicos tempranamente.
   c) Dirección y administración del TEC: La alta dirección utilizará los descubrimientos del SPGDE para:
      i) Tomar decisiones estratégicas más informadas.
      ii) Evaluar la efectividad de las políticas institucionales.
      iii) Justificar inversiones en programas de apoyo estudiantil.
      iv) Mejorar la reputación del TEC como institución comprometida con el éxito estudiantil.
   d) Departamento de Trabajo Social y Salud: Este departamento utilizará la información del SPGDE para:
      i) Mejorar sus programas de apoyo socioeconómico.
      ii) Identificar factores de riesgo relacionados con la salud y el bienestar estudiantil.
      iii) Desarrollar intervenciones más efectivas en el ámbito social y de salud.

4.3.1.2. Fuentes primarias y secundarias

Para obtener una comprensión completa de la demanda, es crucial recopilar información de diversas fuentes. Estas se dividen en fuentes primarias, que proporcionan datos directos y actuales, y fuentes secundarias, que ofrecen contexto y perspectivas adicionales.

1) Fuentes primarias:
   a) Entrevistas con personal clave del TEC: Se realizarán entrevistas detalladas con personal de diversos departamentos. Estas entrevistas proporcionan descubrimientos valiosos sobre las necesidades específicas de cada área y las expectativas para el SPGDE.
   b) Encuestas a estudiantes y personal académico: Se planea realizar encuestas exhaustivas para comprender las percepciones y necesidades de estos grupos clave. Estas encuestas ayudarán a identificar factores de riesgo desde la perspectiva de los usuarios finales.
   c) Datos de matrícula y rendimiento académico: Proporcionados por el Departamento de Admisión y Registro. Estos datos son cruciales para establecer una línea base e identificar patrones históricos de deserción.
   d) Informes del DOP sobre atención a estudiantes: Proporcionan información detallada sobre los tipos de problemas que enfrentan los estudiantes. Ayudan a identificar factores de riesgo comunes y estrategias de intervención efectivas.
2) Fuentes secundarias:

a) Informes del TEC:
   i) Plan Estratégico 2022-2026: Proporciona el contexto institucional y las metas relacionadas con la retención estudiantil.
   ii) TEC en Cifras 2023: Ofrece datos estadísticos actualizados sobre la población estudiantil y tendencias académicas.
b) Estudios previos sobre deserción en el TEC: Por ejemplo, el estudio de Chinchilla-Brenes (2016) ofrece un análisis detallado de los factores de deserción en el TEC.
c) Informes del Estado de la Educación: Proporcionan contexto nacional sobre tendencias en educación superior, incluyendo tasas de deserción y factores asociados.
d) Investigaciones internacionales: Ofrecen perspectivas globales sobre sistemas predictivos en educación superior y mejores prácticas.
e) Estadísticas del CONARE: Proporcionan datos comparativos sobre deserción universitaria en Costa Rica, permitiendo contextualizar la situación del TEC.

Instrumentos para consultas las fuentes primarias y secundarias

La selección de instrumentos adecuados es crucial para recopilar datos precisos y relevantes. Cada instrumento se ha elegido por su capacidad para proporcionar información específica y valiosa para el desarrollo del SPGDE.

- Entrevistas semiestructuradas:
  a) Permiten una exploración profunda de las perspectivas y experiencias del personal clave.
  b) Ofrecen flexibilidad para abordar temas emergentes durante la conversación.
- Encuestas en línea:
  a) Facilitan la recopilación de datos cuantitativos a gran escala.
  b) Permiten llegar a una amplia muestra de estudiantes y personal académico.
  c) Se diseñarán para recopilar información sobre factores de riesgo, experiencias académicas y necesidades de apoyo.
- Análisis de documentos institucionales:
  a) Proporciona contexto histórico y estratégico.
  b) Ayuda a alinear el SPGDE con las metas y políticas institucionales existentes.
- Grupos focales:
  a) Permiten una discusión más profunda y dinámica sobre temas específicos.
  b) Se planean sesiones separadas para estudiantes y profesores para capturar perspectivas únicas.
- Análisis de datos históricos:
  a) Utiliza técnicas estadísticas para identificar patrones y tendencias en datos de deserción y rendimiento académico.
  b) Ayuda a establecer una línea base para evaluar la efectividad futura del SPGDE.

4.3.1.3. Demanda insatisfecha

En base en la información proporcionada en la última entrevista con las Encargadas de la OPI (Apéndice I: Tabla de hallazgos de entrevista a Arlyn Angulo Hernández y Raquel Mejías Elizondo de la OPI.), es necesario ajustar nuestra percepción inicial sobre la demanda insatisfecha y la oferta existente:

1. Demanda actual: Existe una demanda clara y continua por parte del TEC para un sistema de predicción y gestión de la deserción estudiantil. Esta demanda se evidencia en:
    - El Plan Estratégico 2022-2026 del TEC (TEC, 2022), que establece como meta reducir la deserción de estudiantes de primer ingreso del 11% al 9% para 2026.
    - La alta prioridad que la OPI ha asignado al proyecto de actualización del modelo predictivo de deserción (Apéndice I: Tabla de hallazgos de entrevista a Arlyn Angulo Hernández y Raquel Mejías Elizondo de la OPI., EH149).
    - El interés expresado por la Vicerrectora de VIESA en la continuación del proyecto (Apéndice I: Tabla de hallazgos de entrevista a Arlyn Angulo Hernández y Raquel Mejías Elizondo de la OPI., EH161).
    - Durante la realización de este proyecto, se reactivó la actualización del modelo predictivo existente, en la semana 8 del semestre 2 del 2024, demostrando el interés que aún se mantiene.
2. Oferta actual: Contrario a la percepción inicial de una demanda completamente insatisfecha, la entrevista con las Encargadas de la OPI revela que:
    - Se ha retomado el proyecto anterior de modelo predictivo de deserción (Apéndice I: Tabla de hallazgos de entrevista a Arlyn Angulo Hernández y Raquel Mejías Elizondo de la OPI., EH147), el cual no se encontraba anteriormente definido y se especulaba abandonado por los involucrados.
    - El proyecto está actualmente en fase de desarrollo, con un enfoque en la actualización y mejora del modelo anterior (Apéndice I: Tabla de hallazgos de entrevista a Arlyn Angulo Hernández y Raquel Mejías Elizondo de la OPI., EH148).
    - Se ha generado un nuevo conjunto de datos para entrenar el modelo actualizado de deserción (Apéndice I: Tabla de hallazgos de entrevista a Arlyn Angulo Hernández y Raquel Mejías Elizondo de la OPI., EH152).
    - Existe una colaboración activa entre la OPI, DATIC y expertos externos como Martín Solís (Apéndice I: Tabla de hallazgos de entrevista a Arlyn Angulo Hernández y Raquel Mejías Elizondo de la OPI., EH155).

Esta información indica que existe una oferta en desarrollo que busca satisfacer la demanda institucional. Sin embargo, es importante notar que:

- El proyecto está aún en fase de desarrollo y no está completamente implementado.
- Existen desafíos en cuanto a recursos humanos y priorización de (Apéndice I: Tabla de hallazgos de entrevista a Arlyn Angulo Hernández y Raquel Mejías Elizondo de la OPI., EH156, EH160).
- Hay áreas de mejora identificadas, como la necesidad de hacer los datos más dinámicos (Apéndice I: Tabla de hallazgos de entrevista a Arlyn Angulo Hernández y Raquel Mejías Elizondo de la OPI., EH154).

Por lo tanto, aunque no se puede hablar de una demanda completamente insatisfecha, sí existe un espacio significativo para mejorar y expandir la oferta actual para satisfacer plenamente las necesidades institucionales.

4.3.2. Análisis de oferta

4.3.2.1. Marco de desarrollo

El SPGDE se está desarrollando internamente en el TEC, utilizando recursos propios y conocimientos especializados del personal de la institución. Este desarrollo interno presenta varias ventajas:

1. Personalización: El sistema se adapta específicamente a las necesidades y contexto del TEC.
2. Control de datos: Mayor seguridad y control sobre la información sensible de los estudiantes.
3. Integración: Facilita la integración con otros sistemas y procesos internos del TEC.
4. Costo-efectividad a largo plazo: Aunque la inversión inicial puede ser significativa, se evitan costos recurrentes de licencias externas.
5. Desarrollo de capacidades internas: Fortalece las habilidades del personal del TEC en análisis de datos y desarrollo de sistemas.

4.3.2.2. Tipo de oferta que caracteriza al producto

La oferta del SPGDE se caracteriza como un desarrollo institucional interno. Aunque no compite directamente en el mercado externo, se puede considerar que compite internamente con:

- Soluciones comerciales de BI que podrían adaptarse para la predicción de deserción.
- Posibles colaboraciones o proyectos conjuntos con otras universidades.
- La opción de no implementar ningún sistema y continuar con los métodos actuales.

4.3.2.3. Aspectos cuantitativos y cualitativos

- Cuantitativos
    - Capacidad de procesamiento de datos de aproximadamente de 13,000 estudiantes anualmente (TEC, 2023).
    - Potencial de aumento en la tasa de retención del 0.5% al 1% anual (basado en experiencias de otras instituciones como la Universidad de Arizona [Jacobson, 2019]).
    - Reducción proyectada de la deserción de primer ingreso del 11% al 9% para 2026 (TEC, 2022).
- Cualitativos
    - Mejora en la toma de decisiones institucionales basada en datos.
    - Intervenciones más efectivas y personalizadas para estudiantes en riesgo.
    - Mayor eficiencia en la asignación de recursos para programas de retención.
    - Fortalecimiento de la cultura institucional de uso de datos.
    - Potencial para mejorar la reputación del TEC como institución innovadora en la gestión educativa.

4.3.2.4. Fuentes primarias y secundarias

Además de las mencionadas en el análisis de demanda:

- Fuentes primarias adicionales:
    - Documentación técnica del desarrollo actual del SPGDE.
    - Registros de reuniones del equipo de desarrollo.
    - Informes de avance del proyecto SPGDE.
- Fuentes secundarias adicionales:

- Estudios de caso de implementación de sistemas similares en otras universidades.
- Informes técnicos sobre tecnologías de BI y análisis predictivo en educación superior.
- Publicaciones académicas sobre el impacto de sistemas predictivos en la retención estudiantil.

4.3.2.5. Instrumentos para la consulta de las fuentes de información

Además de los mencionados en el análisis de demanda:

- Revisión de código y documentación técnica del SPGDE.
- Pruebas de usabilidad con prototipos del sistema.
- Análisis comparativo con soluciones similares en otras instituciones.
- Evaluación de expertos en BI y análisis predictivo.

4.3.3. Análisis del precio

4.3.3.1. Tipificación del precio

El SPGDE, al ser un desarrollo interno, no tiene un precio de mercado tradicional. Sin embargo, se pueden considerar los siguientes aspectos para su "valoración":

- Costo de desarrollo y mantenimiento.
- Valor generado en términos de retención estudiantil.
- Ahorro en comparación con soluciones comerciales similares.
- Costo de oportunidad de los recursos utilizados.

4.3.3.2. Condiciones económicas

El desarrollo del SPGDE está sujeto a las siguientes condiciones económicas:

1. Asignaciones presupuestarias del TEC: El proyecto debe ajustarse al presupuesto institucional anual.
2. Restricciones financieras: Posibles limitaciones debido a recortes presupuestarios o reasignación de fondos.
3. Priorización de proyectos: Competencia con otros proyectos institucionales por recursos.
4. Retorno de inversión: Justificación del proyecto basada en el ahorro potencial por aumento en la retención estudiantil.
5. Contexto económico nacional: Influencia de la situación económica general del país en el presupuesto universitario.

4.3.3.3. Reacción de la competencia

Aunque no hay competencia directa en el mercado, se pueden considerar las siguientes "reacciones competitivas":

1. Otras universidades públicas podrían desarrollar sistemas similares, lo que podría llevar a colaboraciones o competencia por recursos y reconocimiento.

2. Proveedores de soluciones comerciales de BI podrían ofrecer adaptaciones específicas para la educación superior, potencialmente a precios competitivos.

3. Empresas de consultoría en educación podrían ofrecer servicios complementarios o alternativos para la gestión de la deserción.

4.3.3.4. Análisis de la comercialización

Aunque el SPGDE no se comercializará en el sentido tradicional, es crucial considerar cómo se "comercializará" internamente para asegurar su adopción y uso efectivo.

4.3.3.5. Canales de distribución

1. Intranet del TEC: Plataforma principal para acceder al SPGDE y sus informes.
2. TEC Digital: Integración con la plataforma educativa existente para facilitar el acceso de profesores y estudiantes.
3. Sesiones de capacitación presenciales: Para personal clave de cada departamento y escuela.
4. *Webinars* y tutoriales en línea: Para alcanzar a un público más amplio y proporcionar recursos de aprendizaje continuo.
5. Informes periódicos automatizados: Enviados a tomadores de decisiones clave.

4.3.3.6. Estrategias promocionales

- Campañas de concientización interna:
  - Correos electrónicos informativos sobre el SPGDE y sus beneficios
  - Presentaciones en reuniones de facultad y consejos de escuela
- Talleres de capacitación:
  - Sesiones introductorias para todo el personal
  - Talleres avanzados para usuarios clave (DOP, VIESA, directores de escuela)
  - Sesiones de actualización semestrales
- Integración con procesos existentes:
  - Incorporación del SPGDE en los procesos de matrícula y seguimiento académico
  - Uso de datos del SPGDE en informes institucionales regulares
- Demostración de impacto:
  - Casos de éxito de intervenciones basadas en el SPGDE
  - Informes periódicos sobre mejoras en las tasas de retención
- Incentivos para la adopción:
  - Reconocimiento a escuelas y departamentos que utilicen efectivamente el SPGDE
  - Inclusión del uso del SPGDE en evaluaciones de desempeño del personal académico y administrativo

4.4. Evaluación de la competencia

La evaluación de la competencia es un componente crucial en la formulación y evaluación de proyectos de TI. En el contexto del Sistema de Predicción y Gestión de la Deserción Estudiantil (SPGDE) para el Tecnológico de Costa Rica (TEC), esta evaluación nos permite comprender el panorama competitivo, identificar oportunidades de mejora y posicionar estratégicamente nuestro proyecto. A continuación, se presenta un análisis detallado de la competencia, sus fortalezas y debilidades, ventajas de mercado y estrategias.

4.4.1. Descripción de la competencia

En el ámbito de los sistemas de predicción y gestión de la deserción estudiantil, tanto en el Tecnológico de Costa Rica como en Costa Rica en general es realmente escasa desde el punto de vista de proyectos consolidades, sin embargo, el incremento del auge en la IA y soluciones de Inteligencia de Negocios está causando el surgimiento de algunos proyectos nuevos de este tema. Según lo anterior, se identificaron los siguientes competidores directos y proyectos que están surgiendo:

- Competidores directos:
  - Modelo de predicción de deserción estudiantil actual: Este proyecto, fue realizado alrededor de los años 2019-2020, mediante un cubo de datos implementado en la OPI, utilizando múltiples fuentes institucionales, sin embargo, este proyecto no fue continuado y operado para su funcionamiento, por lo que se encuentra inactivo y desactualizado actualmente.
  - Proyecto de Actualización del Modelo Anterior de la OPI: Este es un proyecto reactivado en semana 8 del semestre 2 del año 2024, entre las entrevistas realizadas a los diferentes actores de la investigación dentro del TEC, dirigido por la Oficina de Planificación Institucional (OPI) en colaboración con el Departamento de Administración de Tecnologías de Información y Comunicaciones (DATIC). Este proyecto busca actualizar y mejorar el modelo predictivo de deserción desarrollado anteriormente.
- Proyectos en desarrollo:
  - Proyecto de predicción de deserción de la escuela de matemática en la Universidad Nacional (UNA): Este proyecto como bien se menciona en parte de las investigaciones realizadas (Ver Apéndice H: Tabla de hallazgos de entrevista a Evaluadora e investigadora jubilada Tania Elena Moreira Mora., EH146) se tiene conocimiento de un modelo predictivo surgiendo en la UNA, con el fin de determinar la deserción en una escuela o curso específico, sin embargo, no se cuenta con más información ya que se encuentra en desarrollo.

Como vista general, tal y como se muestra en la **Tabla *23***, todos estos competidores tienen grandes diferencias entre sí, sin embargo, destaca el competidor directo el cual es el proyecto de actualización del modelo de predicción anterior que no se llegó a concretar.

Tabla 23
Comparación detallada de competidores.

Característica	SPGDE (Propuesto)	Proyecto de Actualización OPI	Modelo Actual
Enfoque	Específico para deserción estudiantil en el TEC	Específico para deserción en el TEC	General para indicadores institucionales
Personalización	Diseñado específicamente para los procesos del TEC	Diseñado específicamente para el TEC	Sistema general adaptado
Integración con sistemas del TEC	Planificada desde el diseño para todos los sistemas críticos	Integrado con SIGI	Sistema nativo del TEC
Costo de implementación	₡29,367,676.185 (2 años iniciales)	Costos de personal interno (costo no definido)	Ya implementado
Tiempo estimado de implementación	24 meses según cronograma	En desarrollo (*timeline* no definido)	Ya implementado
Capacidades de *machine learning*	Incluye *Random Forest*, *AdaBoost* y otros algoritmos avanzados	Planificadas, pero no implementadas	No disponible

Nota: Elaboración propia, las calificaciones fueron colocadas en base a un análisis de los apéndices referentes a las entrevistas aplicadas a los sujetos de investigación.

4.4.2. Fortalezas y debilidades de la competencia

Ser consciente de las fortalezas y debilidades de las empresas rivales es un aspecto crucial para la formulación y evaluación de un proyecto, debido a que permite crear una visión integral de cuál es la situación del mercado, y en qué posición se encuentra el evaluador. A la hora de realizar el análisis de las fortalezas y debilidades de la competencia, es posible descubrir puntos a favor que, posteriormente, pueden ser la base de las estrategias de diferenciación, o también ser la base de puntos de mejora en la organización que realiza este análisis. Además de esto, se pueden crear estrategias de predicción, en las que se anticipa los siguientes pasos de la competencia, y de esta manera, tomar ventaja sobre estas compañías rivales. Debido a estas razones, a continuación, en la **Tabla *24***, se presenta el análisis de las fortalezas y debilidades de cada una de las empresas competencia.

Tabla 24
Fortalezas y debilidades de competencia.

Competidor	Fortaleza	Debilidad
Proyecto de Actualización OPI	Ya cuenta con un conjunto de datos preparado.	Desconocimiento en personal, costo y tiempo requerido para implementar mejoras debido a que se encuentra en tapa exploratoria.
	Experiencia previa con el modelo anterior.	Priorización frente a otros proyectos y deberes del equipo.
	Respaldo de la vicerrectoría y departamentos.	Datos estáticos que requieren actualización y fuentes de información limitadas.
	Conocimiento profundo de los sistemas del TEC.	Dependencia de varios departamentos del TEC para los datos.
Modelo actual	Sistema establecido y en funcionamiento.	Modelo previo a pandemia, sin el conocimiento de una precisión y sensibilidad a nuevos cambios de modalidades de los cursos y la realidad, debido a que no se ha ejecutado desde el año 2021.
	Integración completa con sistemas del TEC.	Enfoque descriptivo más que predictivo.
	Familiaridad de los usuarios.	Limitado a datos históricos.
	Acceso controlado por roles.	No ofrece alertas tempranas, ni ofrece casos prioritarios.

Nota: Elaboración propia, estas características documentadas son resultado del análisis de los apéndices relacionados a las entrevistas aplicadas a sujetos de investigación.

4.4.3. Ventajas de mercado de la competencia

Cada competidor tiene ventajas de mercado específicas que es importante considerar:

1. Modelo de predicción actual:
   - Ya se encuentra integrado en los sistemas de la OPI como un cubo de datos más.
   - Su desarrollo fue exitoso, logrando un alto porcentaje de precisión, quedando en duda la necesidad del desarrollo o actualización de un nuevo modelo.
   - Tuvo mucha exposición con diferentes entidades por lo que es de conocimiento general su buen desempeño y objetivo (Ver Apéndice H: Tabla de hallazgos de entrevista a Evaluadora e investigadora jubilada Tania Elena Moreira Mora., EH127).
2. Proyecto de Actualización del Modelo Anterior (OPI):
   - Conocimiento profundo del contexto específico del TEC y sus necesidades en cuanto a la predicción de deserción.

- Acceso directo a los datos y sistemas del TEC, facilitando la integración y personalización.
- Alineación directa con las políticas y objetivos institucionales del TEC.
- Se encuentra en desarrollo, por lo que su llegada será mucho más cercana.
- Cuenta con la experiencia y respaldo de un modelo anterior con alta precisión.
- Ya cuenta con un equipo de desarrollo detrás y un apoyo respaldado por la vicerrectoría de docencia y diferentes departamentos.
- Tiene una buena reputación por el modelo anterior, por lo que puede favorecer su elección sobre otros modelos o desarrollar nuevos.

4.4.4. Estrategias de mercado de la competencia

Las estrategias de mercado de los competidores se pueden resumir de la siguiente manera:

1. Modelo de predicción actual:
    - Sistema consolidado como fuente principal de datos institucionales.
    - Acceso controlado basado en roles y responsabilidades.
    - Integración con múltiples sistemas del TEC.
    - Enfoque en la visualización de datos históricos.
2. Proyecto de Actualización del Modelo Anterior:
    - Enfoque en la actualización y mejora del modelo existente.
    - Integración profunda con SIGI.
    - Aprovechamiento de la experiencia previa en el TEC.
    - Colaboración con expertos internos como Martín Solís (Ver Apéndice T: Minuta de entrevista al Profesor Martín Solís Salazar.).
    - Enfoque en la alineación con las necesidades específicas del TEC, aprovechando el conocimiento institucional existente.
    - Énfasis en la integración con los sistemas y procesos existentes del TEC para maximizar la eficiencia y adopción.

Esta evaluación de la competencia proporciona una base sólida para posicionar el SPGDE en el mercado. Tal y como se muestra en la **Figura 3** y en las secciones anteriores, el proyecto SPGDE no ofrece una ventaja muy grande para abordar la deserción estudiantil en el contexto del TEC en comparación con sus competidores, aunque si incluye algunas mejoras importantes a tomar en cuenta. Su desarrollo interno permite una alta personalización y una integración perfecta con los sistemas existentes del TEC y externos como bases de datos gubernamentales y públicas, lo que lo diferencia claramente de la competencia interna que centra sus esfuerzos únicamente lo que se puede obtener dentro de la institución.

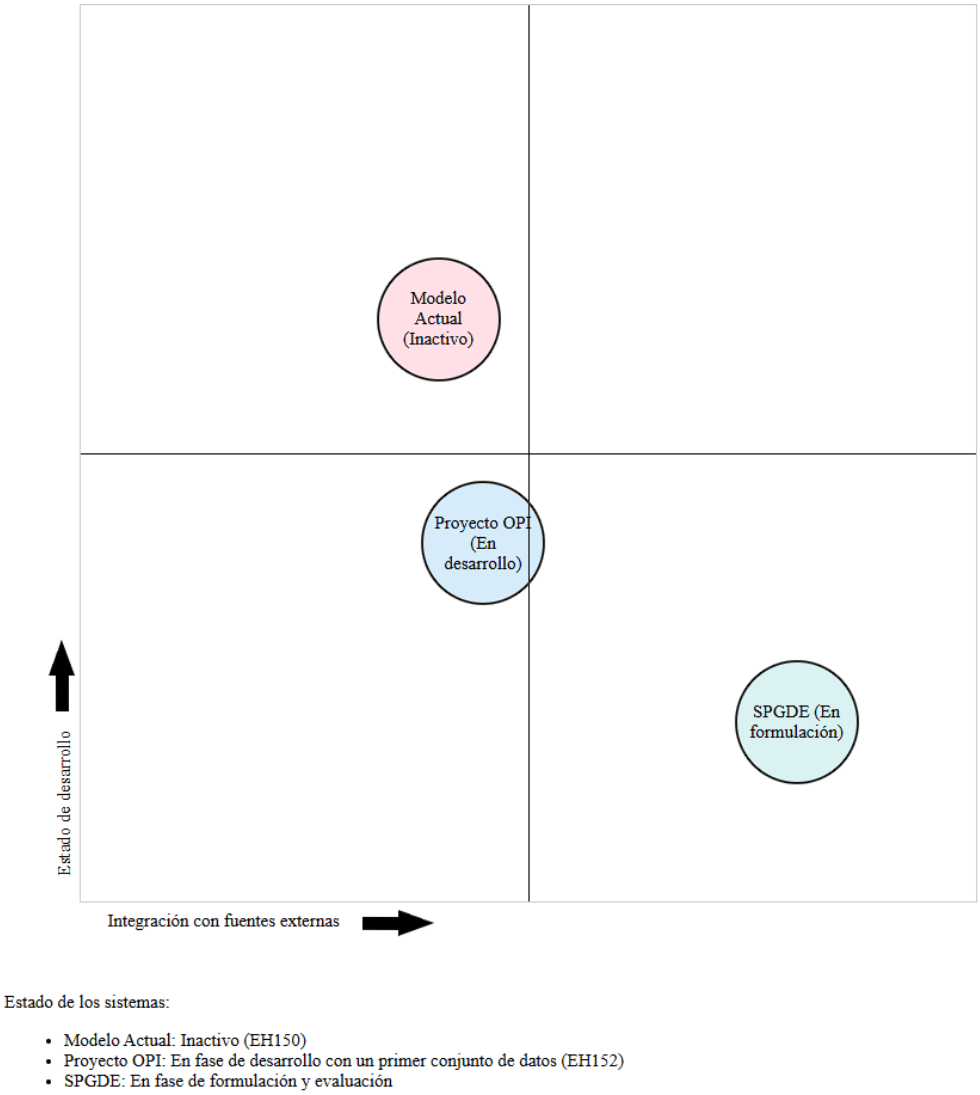
El Proyecto de Actualización del Modelo Anterior representa una competencia interna significativa, con ventajas similares en términos de especificidad y conocimiento institucional. Sin embargo, el SPGDE propuesto podría diferenciarse al incorporar tecnologías más avanzadas de aprendizaje automático (*machine learning*) y un enfoque más holístico en la gestión de la deserción, no solo en la predicción, por otra parte, presenta una amenaza inminente que su culminación antes de la llegada de este nuevo modelo, presente un golpe crítico en la demanda del TEC si se llega a suplir de la mejor forma sus necesidades con un modelo con una alta precisión que convierta en innecesario el gasto en recursos en un nuevo modelo.

Al comprender las fortalezas y debilidades de los competidores, podemos enfatizar las ventajas únicas del SPGDE, como su especificidad, costo-efectividad y alineación con las necesidades del TEC. Esto nos permite

desarrollar una estrategia de implementación que aproveche estas ventajas y aborde las posibles debilidades en comparación con soluciones más establecidas en el mercado.

Figura 3

Posicionamiento de mercado de los competidores.



Nota: Elaboración propia, adaptado de Apéndice I: Tabla de hallazgos de entrevista a Arlyn Angulo Hernández y Raquel Mejías Elizondo de la OPI.

4.5. Evaluación de riesgos

La evaluación de riesgos es un componente crítico en la planificación y ejecución del Sistema de Predicción y Gestión de la Deserción Estudiantil (SPGDE). Esta sección se enfocará en identificar, analizar y proponer estrategias de mitigación para un amplio espectro de riesgos potenciales asociados con el desarrollo e implementación del SPGDE en el Tecnológico de Costa Rica (TEC).

4.5.1. Identificación de riesgos

En esta sección, se encuentran, reconocen y descubren los riesgos que pueden impedir o ayudar a lograr los objetivos planteados al inicio del proyecto; esto por medio de una identificación de activos, amenazas, vulnerabilidades y consecuencias.

4.5.1.1. Identificación de activos

A continuación, en la **Tabla *25*** se muestran los activos clave del proyecto que podrían estar en riesgo.

Tabla 25
Activos clave del proyecto

Identificador	Categoría	Nombre de activo
AT1	Infraestructura tecnológica	Servidores.
AT2		Redes.
AT3		Sistemas de almacenamiento.
AT4		Equipos de cómputo.
AT5		Dispositivos móviles institucionales.
AT6	Datos	Datos académicos de estudiantes.
AT7		Datos personales de estudiantes.
AT8		Datos socioeconómicos.
AT9		Datos históricos de deserción.
AT10		Metadatos del sistema.
AT11	Software	Software de análisis y predicción.
AT12		Sistemas de gestión de bases de datos.
AT13		Herramientas de visualización de datos.
AT14		Sistemas operativos.
AT15		Software de seguridad.
AT16	Recursos humanos	Desarrolladores.
AT17		Analistas de datos.
AT18		Expertos en educación.
AT19		Personal administrativo.
AT20		Directivos del TEC.
AT21	Procesos	Metodologías de desarrollo.
AT22		Procesos de toma de decisiones.
AT23		Protocolos de seguridad.
AT24		Procedimientos de mantenimiento.
AT25	Relaciones institucionales	Entre departamentos del TEC.
AT26		Con entidades externas.
AT27		Con proveedores.
AT28	Intangibles	Reputación institucional del TEC.
AT29		Propiedad intelectual.
AT30		Conocimiento institucional.
AT31		Cultura organizacional.

Nota: Elaboración propia.

4.5.1.2. Identificación de amenazas

A continuación, en la **Tabla *26*** se muestran las amenazas identificadas para el proyecto SPGDE.

Tabla 26
Amenazas del proyecto.

Identificador	Categoría	Nombre de amenaza
AM1	Consecuencias de seguridad y privacidad	Violaciones de privacidad de datos estudiantiles.
AM2		Pérdida de confianza de los estudiantes y el personal.
AM3		Daño reputacional al TEC.
AM4		Sanciones legales y multas.
AM5	Consecuencias técnicas	Tiempo de inactividad del sistema.
AM6		Pérdida o corrupción de datos.
AM7		Rendimiento deficiente del sistema.
AM8		Fallos en la integración con otros sistemas del TEC.
AM9	Consecuencias operativas	Retrasos significativos en la implementación.
AM10		Sobrecostos en el desarrollo y mantenimiento.
AM11		Ineficacia en la predicción de la deserción.
AM12		Aumento de la carga de trabajo administrativa.
AM13	Consecuencias organizacionales	Resistencia y baja adopción por parte del personal.
AM14		Conflictos internos entre departamentos.
AM15		Pérdida de oportunidades de mejora institucional.
AM16		Desalineación con los objetivos estratégicos del TEC.
AM17	Consecuencias financieras	Sobrecostos del proyecto.
AM18		Pérdida de inversión en caso de fracaso del proyecto.
AM19		Impacto en el presupuesto de otros proyectos del TEC.
AM20		Pérdida de oportunidades de financiamiento futuro.

Identificador	Categoría	Nombre de amenaza
AM21	Consecuencias académicas	Toma de decisiones erróneas basadas en predicciones inexactas.
AM22		Falta de mejora en las tasas de retención estudiantil.
AM23		Impacto negativo en la experiencia estudiantil.
AM24		Pérdida de ventaja competitiva frente a otras instituciones.
AM25	Consecuencias legales y de cumplimiento	Demandas por uso indebido de datos personales.
AM26		Incumplimiento de regulaciones educativas.
AM27		Pérdida de acreditaciones o certificaciones.
AM28		Daño a la reputación legal y ética del TEC.
AM29	Consecuencias a largo plazo	Obsolescencia prematura del sistema.
AM30		Dependencia excesiva de la tecnología en la toma de decisiones.
AM31		Cambios no deseados en la cultura institucional.
AM32		Pérdida de enfoque en otros aspectos importantes de la retención estudiantil.

Nota: Elaboración propia.

4.5.1.3. Identificación de vulnerabilidades

A continuación, en la **Tabla *27*** se muestran las vulnerabilidades potenciales del proyecto.

Tabla 27
Vulnerabilidades del proyecto.

Identificador	Categoría	Nombre de amenaza
VU1	Vulnerabilidades técnicas	Sistemas de legado no actualizados.
VU2		Configuraciones de seguridad débiles.
VU3		Falta de cifrado de datos.
VU4		Arquitectura de sistema no escalable.
VU5	Vulnerabilidades de proceso	Falta de documentación de procesos.
VU6		Ausencia de pruebas de calidad rigurosas.

Identificador	Categoría	Nombre de amenaza
VU7		Gestión de cambios inadecuada.
VU8		Falta de planes de continuidad de negocio.
VU9	Vulnerabilidades de recursos humanos	Falta de experiencia en proyectos de IA y ML.
VU10		Dependencia de personal clave.
VU11		Alta rotación de personal.
VU12		Falta de capacitación continua.
VU13	Vulnerabilidades organizacionales	Comunicación interdepartamental deficiente.
VU14		Falta de claridad en roles y responsabilidades.
VU15		Cultura organizacional resistente al cambio.
VU16		Falta de alineación estratégica.
VU17	Vulnerabilidades de datos	Calidad y completitud de datos históricos.
VU18		Falta de estandarización de datos.
VU19		Acceso inadecuado a fuentes de datos externas.
VU20		Falta de gobierno de datos.
VU21	Vulnerabilidades financieras	Presupuesto insuficiente para mantenimiento a largo plazo.
VU22		Falta de análisis de costo-beneficio detallado.
VU23		Dependencia de una única fuente de financiamiento.
VU24	Vulnerabilidades legales y de cumplimiento	Falta de conocimiento sobre regulaciones de privacidad.
VU25		Ausencia de políticas de retención de datos.
VU26		Falta de acuerdos de nivel de servicio claros.
VU27	Vulnerabilidades de adopción	Resistencia de los usuarios finales.
VU28		Falta de plan de gestión del cambio.
VU29		Interfaces de usuario poco intuitivas.

Identificador	Categoría	Nombre de amenaza
VU30		Falta de soporte post implementación.

Nota: Elaboración propia.

4.5.1.4. Identificación de consecuencias

A continuación, en la **Tabla *28*** se muestran las consecuencias potenciales de la materialización de estos riesgos.

Tabla 28
Consecuencias de riesgos del proyecto.

Identificador	Categoría	Nombre de amenaza
CO1	Consecuencias de seguridad y privacidad	Violaciones de privacidad de datos estudiantiles.
CO2		Pérdida de confianza de los estudiantes y el personal.
CO3		Daño reputacional al TEC.
CO4		Sanciones legales y multas.
CO5	Consecuencias técnicas	Tiempo de inactividad del sistema.
CO6		Pérdida o corrupción de datos.
CO7		Rendimiento deficiente del sistema.
CO8		Fallos en la integración con otros sistemas del TEC.
CO9	Consecuencias operativas	Retrasos significativos en la implementación.
CO10		Sobrecostos en el desarrollo y mantenimiento.
CO11		Ineficacia en la predicción de la deserción.
CO12		Aumento de la carga de trabajo administrativa.
CO13	Consecuencias organizacionales	Resistencia y baja adopción por parte del personal.
CO14		Conflictos internos entre departamentos.
CO15		Pérdida de oportunidades de mejora institucional.
CO16		Desalineación con los objetivos estratégicos del TEC.

Identificador	Categoría	Nombre de amenaza
CO17	Consecuencias financieras	Sobrecostos del proyecto.
CO18		Pérdida de inversión en caso de fracaso del proyecto.
CO19		Impacto en el presupuesto de otros proyectos del TEC.
CO20		Pérdida de oportunidades de financiamiento futuro.
CO21	Consecuencias académicas	Toma de decisiones erróneas basadas en predicciones inexactas.
CO22		Falta de mejora en las tasas de retención estudiantil.
CO23		Impacto negativo en la experiencia estudiantil.
CO24		Pérdida de ventaja competitiva frente a otras instituciones.
CO25	Consecuencias legales y de cumplimiento	Demandas por uso indebido de datos personales.
CO26		Incumplimiento de regulaciones educativas.
CO27		Pérdida de acreditaciones o certificaciones.
CO28		Daño a la reputación legal y ética del TEC.
CO29	Consecuencias a largo plazo	Obsolescencia prematura del sistema.
CO30		Dependencia excesiva de la tecnología en la toma de decisiones.
CO31		Cambios no deseados en la cultura institucional.
CO32		Pérdida de enfoque en otros aspectos importantes de la retención estudiantil.

Nota: Elaboración propia.

4.5.1.5. Tabla final riesgos

Seguidamente, se presenta en la **Tabla *29*** la tabla final con los riesgos asociados con el proyecto, usando como insumo las amenazas, vulnerabilidades y consecuencias identificadas previamente.

Tabla 29
Riesgos asociados al proyecto.

Identificador	Riesgo
R1	Brecha de seguridad de datos o ciberataques.
R2	Uso de otro modelo de predicción.
R3	Limitaciones presupuestarias.

R4	Inexactitud en las predicciones del modelo.
R5	Problemas de integración con sistemas existentes.
R6	Cambios en políticas institucionales o legislación.
R7	Baja adopción por usuarios finales.
R8	Cambios imprevistos en patrones de deserción.
R9	Falta de datos históricos suficientes o de calidad.
R10	Falta de apoyo continuo de la alta dirección.
R11	Cambios en las prioridades institucionales.
R12	Problemas éticos en el uso de datos predictivos.
R13	Falta de capacitación adecuada para usuarios.

Nota: Elaboración propia.

4.5.2. Análisis de riesgos

Tras la identificación exhaustiva, en la **Tabla *30***, se analizan los riesgos en términos de su probabilidad de ocurrencia e impacto potencial y se priorizarán en función de su nivel de riesgo. Para esto, se utilizarán las siguientes escalas en las que se categorizarán cada uno de los riesgos.

- Probabilidad: Se brinda una breve explicación del significado de afectación de cada una de las probabilidades con las que se analizan los riesgos.
    - Muy probable (5): Probabilidad de materialización del 80% al 100%. Es casi seguro que el evento ocurrirá durante el desarrollo del proyecto SPGDE.
    - Probable (4): Probabilidad de materialización del 60% al 79%. Es probable que el evento ocurra en la mayoría de las circunstancias durante el proyecto SPGDE.
    - Posible (3): Probabilidad de materialización del 40% al 59%. El evento podría ocurrir en algún momento durante el desarrollo del proyecto SPGDE.
    - Poco probable (2): Probabilidad de materialización del 20% al 39%. El evento podría ocurrir en algunas circunstancias, pero, es poco común en proyectos similares al SPGDE.
    - Muy raro (1): Probabilidad de materialización del 0% al 19%. El evento puede ocurrir solo en circunstancias excepcionales durante el desarrollo del SPGDE.

- Impacto: Se brinda una breve explicación del significado de afectación de cada uno de los valores de impacto con los que se analizan los riesgos mencionados.
    - Catastrófico (5): La posible materialización del riesgo afectaría de manera crítica el alcance, la calidad, el cronograma o el presupuesto del proyecto SPGDE. Podría resultar en el fracaso total del proyecto o en consecuencias legales o reputacionales severas para el TEC.
    - Mayor (4): La posible materialización del riesgo afectaría de manera significativa el alcance, la calidad, el cronograma o el presupuesto del proyecto SPGDE. Requeriría cambios mayores en el plan del proyecto y podría comprometer seriamente el éxito del SPGDE.
    - Moderado (3): La posible materialización del riesgo afectaría de manera notable el alcance, la calidad, el cronograma o el presupuesto del proyecto SPGDE. Requeriría ajustes importantes en el plan del proyecto, pero, el éxito general del SPGDE aún sería alcanzable.

- Menor (2): La posible materialización del riesgo afectaría de manera limitada el alcance, la calidad, el cronograma o el presupuesto del proyecto SPGDE. Requeriría algunos ajustes en el plan del proyecto, pero, no comprometería significativamente el éxito del SPGDE.
- Insignificante (1): La posible materialización del riesgo tendría un impacto mínimo en el alcance, la calidad, el cronograma o el presupuesto del proyecto SPGDE. Podría manejarse con ajustes menores o procedimientos de rutina.

- Nivel de Riesgo: El nivel de riesgo se calcula multiplicando la probabilidad por el impacto. Esto resulta en una escala de 1 a 25, que se puede categorizar de la siguiente manera:
  - Extremo (20-25): Requiere atención inmediata y estrategias de mitigación agresivas. Puede requerir la reconsideración de aspectos fundamentales del proyecto SPGDE.
  - Alto (15-19): Necesita atención prioritaria y planes de acción específicos. Puede afectar significativamente los objetivos clave del SPGDE.
  - Medio (10-14): Requiere monitoreo activo y medidas de control. Puede impactar aspectos importantes del SPGDE, pero, es manejable con la planificación adecuada.
  - Bajo (5-9): Necesita ser monitoreado regularmente. Generalmente puede manejarse a través de procedimientos de rutina en el proyecto SPGDE.
  - Muy Bajo (1-4): Requiere monitoreo periódico. Generalmente puede ser aceptado sin necesidad de acciones específicas en el contexto del SPGDE.

Tabla 30
Matriz de Probabilidad e Impacto de Riesgos.

Identificador	Riesgo	Probabilidad	Impacto	Nivel de riesgo
R3	Limitaciones presupuestarias	5	5	25 (extremo)
R6	Cambios en políticas institucionales o legislación	4	5	20 (extremo)
R2	Uso de otro modelo de predicción.	4	4	16 (alto)
R9	Falta de datos históricos suficientes o de calidad	4	4	16 (alto)
R8	Cambios imprevistos en patrones de deserción	5	3	15 (alto)

Identificador	Riesgo	Probabilidad	Impacto	Nivel de riesgo
R12	Problemas éticos en el uso de datos predictivos	3	5	15 (alto)
R1	Brecha de seguridad de datos o ciberataques.	2	5	10 (medio)
R5	Problemas de integración con sistemas existentes	2	5	10 (medio)
R10	Falta de apoyo continuo de la alta dirección	3	3	9 (bajo)
R4	Inexactitud en las predicciones del modelo	1	5	5 (bajo)
R7	Baja adopción por usuarios finales	4	1	4 (muy bajo)
R11	Cambios en las prioridades institucionales	1	4	4 (muy bajo)
R13	Falta de capacitación adecuada para usuarios	2	2	4 (muy bajo)

Nota: Elaboración propia. Probabilidad e Impacto se califican en una escala de 1 (muy bajo) a 5 (muy alto). El Nivel de Riesgo es el producto de Probabilidad x Impacto.

4.5.3. Valoración y respuesta al riesgo

En esta sección, en la **Tabla *31*** se valoran los riesgos previamente analizados, con el fin de comparar resultados y hacer indagación para apoyar la toma de decisiones, e implementar futuras respuestas para eliminar, mitigar, o aceptar dichos riesgos.

Tabla 31
Estrategias de Respuesta a Riesgos.

Identificador	Riesgo	Estrategia	Acciones de respuesta
R1	Brecha de seguridad de datos o ciberataques	Mitigar	- Implementar cifrado de datos en reposo y en tránsito - Realizar auditorías de seguridad trimestrales - Implementar autenticación de dos factores - Establecer un programa de concientización sobre seguridad - Implementar un sistema de detección y prevención de intrusiones - Realizar pruebas de penetración regulares - Mantener todos los sistemas actualizados y parchados - Considerar la contratación de un seguro cibernético
R2	Uso de otro modelo de predicción.	Mitigar	- Desarrollar un plan detallado de gestión del cambio - Identificar y capacitar "campeones" del proyecto en cada departamento - Realizar sesiones de retroalimentación regulares con los usuarios - Comunicar claramente los beneficios del sistema
R3	Limitaciones presupuestarias	Mitigar/Transferir	- Realizar un análisis detallado de costo-beneficio - Explorar opciones de financiamiento externo o subvenciones - Considerar un enfoque de desarrollo por fases - Establecer reservas de contingencia en el presupuesto
R4	Inexactitud en las predicciones del modelo	Mitigar	- Implementar técnicas de validación cruzada y pruebas A/B - Establecer un proceso de mejora continua del modelo - Incorporar retroalimentación de expertos en educación - Mantener transparencia sobre las limitaciones del modelo
R5	Problemas de integración con sistemas existentes	Mitigar	- Realizar un mapeo exhaustivo de sistemas y datos existentes - Desarrollar *APIs y RPA* robustas para la integración - Establecer un entorno de pruebas de integración - Colaborar estrechamente con el equipo de TI del TEC
R6	Cambios en políticas institucionales o legislación	Aceptar/Mitigar	- Mantener comunicación regular con autoridades del TEC - Monitorear cambios legislativos relevantes

Identificador	Riesgo	Estrategia	Acciones de respuesta
			- Diseñar el sistema con flexibilidad para adaptarse a cambios - Establecer un comité de cumplimiento normativo
R7	Baja adopción por usuarios finales	Mitigar	- Involucrar a usuarios en el proceso de diseño (diseño centrado en el usuario) - Realizar pruebas de usabilidad extensivas - Ofrecer capacitación personalizada por grupos de usuarios - Implementar un sistema de recompensas por adopción temprana
R8	Cambios imprevistos en patrones de deserción	Mitigar	- Diseñar el modelo para ser adaptable a nuevos patrones - Realizar actualizaciones regulares del modelo - Incorporar fuentes de datos externas relevantes - Establecer un panel de expertos para revisión periódica de tendencias
R9	Falta de datos históricos suficientes o de calidad	Mitigar	- Implementar técnicas de imputación de datos faltantes - Colaborar con otras instituciones para obtener datos complementarios - Desarrollar un plan de mejora de la calidad de datos a largo plazo - Considerar el uso de técnicas de aumento de datos
R10	Falta de apoyo continuo de la alta dirección	Mitigar	- Desarrollar un plan de comunicación ejecutiva - Presentar informes de progreso y ROI regularmente - Alinear los objetivos del proyecto con los *KPIs* institucionales - Invitar a la alta dirección a demostraciones y sesiones de retroalimentación
R11	Cambios en las prioridades institucionales	Mitigar	- Mantener el proyecto alineado con la estrategia institucional - Desarrollar casos de uso que demuestren valor en múltiples escenarios - Mantener flexibilidad en la planificación del proyecto - Establecer hitos claros y medibles para demostrar progreso
R12	Problemas éticos en el uso de datos predictivos	Mitigar	- Establecer un comité de ética para supervisar el uso de datos - Desarrollar políticas claras sobre el uso ético de la IA

Identificador	Riesgo	Estrategia	Acciones de respuesta
			- Implementar mecanismos de transparencia en las predicciones - Realizar auditorías éticas regulares del sistema
R13	Falta de capacitación adecuada para usuarios	Mitigar	- Desarrollar un programa de capacitación integral - Crear recursos de aprendizaje en línea y documentación detallada - Implementar un sistema de soporte y mentorías entre pares - Realizar evaluaciones regulares de las necesidades de capacitación

Nota: Elaboración propia.

4.5.4. Plan de monitoreo y control de riesgos

Para asegurar una gestión efectiva de los riesgos identificados, se recomienda la implementación un plan de monitoreo y control con las siguientes características:

1. Revisiones periódicas: Realizar revisiones mensuales de la matriz de riesgos, actualizando las probabilidades e impactos según sea necesario.
2. Indicadores clave de riesgo (*KRIs*): Establecer *KRIs* para cada riesgo principal, que se monitorearán regularmente.
3. Reportes de estado: Generar informes mensuales sobre el estado de los riesgos y la efectividad de las estrategias de mitigación.
4. Auditorías: Llevar a cabo auditorías trimestrales para evaluar la efectividad del proceso de gestión de riesgos.
5. Reuniones de equipo: Realizar reuniones semanales del equipo del proyecto para discutir nuevos riesgos potenciales y actualizar las estrategias de respuesta.
6. Registro de lecciones aprendidas: Mantener un registro actualizado de lecciones aprendidas en la gestión de riesgos para mejorar continuamente el proceso.

4.5.5. Riesgos residuales y secundarios

Es importante reconocer que incluso con estrategias de mitigación en lugar, algunos riesgos residuales persistirán:

1. Riesgo residual de inexactitud en las predicciones: Incluso con las mejores técnicas de aprendizaje automático (*machine learning*), siempre existirá un margen de error en las predicciones.
2. Riesgo residual de cambios en patrones de deserción: Factores externos imprevistos pueden alterar los patrones de deserción de manera que el modelo no pueda anticipar completamente.
3. Riesgo residual de seguridad: A pesar de las medidas de seguridad, siempre existe un riesgo residual de brechas de seguridad debido a amenazas en constante evolución.

Además, es crucial estar atento a los riesgos secundarios que pueden surgir como resultado de las estrategias de mitigación implementadas:

1. Riesgo de dependencia excesiva del sistema: La implementación exitosa del SPGDE podría llevar a una dependencia excesiva de la tecnología en la toma de decisiones, potencialmente disminuyendo el juicio humano.
2. Riesgo de sobrecarga de información: La disponibilidad de más datos y predicciones podría resultar en una sobrecarga de información para los tomadores de decisiones.
3. Riesgo de desigualdad en la atención estudiantil: Si el sistema no se implementa cuidadosamente, podría llevar a una distribución desigual de recursos y atención entre los estudiantes.

4.6. Evaluación social

La evaluación social de proyectos de inversión pública constituye un elemento fundamental en el proceso de toma de decisiones sobre la asignación de recursos estatales. En el contexto del Sistema de Predicción y Gestión de la Deserción Estudiantil (SPGDE) del Tecnológico de Costa Rica (TEC), esta evaluación adquiere características particulares debido a la naturaleza del proyecto como herramienta de apoyo a la gestión institucional.

La presente evaluación se desarrolla siguiendo los lineamientos establecidos por el Ministerio de Planificación Nacional y Política Económica (MIDEPLAN, 2022) en su "Guía metodológica general para la identificación, formulación y evaluación de proyectos de inversión pública", adaptando la metodología a las características específicas de un proyecto de fortalecimiento de capacidades institucionales en el ámbito de la educación superior pública.

4.6.1. Evaluación económica del proyecto social

La evaluación económica-social del SPGDE requiere ajustar los costos y beneficios del proyecto para reflejar su verdadero valor para la sociedad. Como señala MIDEPLAN (2022, p.156), este proceso implica "la corrección de las distorsiones del mercado mediante el uso de precios sociales o precios sombra, que reflejan el verdadero costo de oportunidad de los recursos para la sociedad".

4.6.1.1. Precios Sociales (Sombra)

Para el cálculo de los precios sociales, se utilizan los factores de corrección establecidos por MIDEPLAN (2023) en su guía de precios sociales. El proyecto SPGDE incluye tres categorías principales de costos que requieren ajustes:

a) Licencias de Software: Las licencias de software, específicamente POWER BI, se clasifican como bienes transables según la metodología de MIDEPLAN. Como señala la guía metodológica, "si el valor del bien transable está expresado en moneda nacional (colones), para obtener el precio social del bien, se toma el valor del bien y se multiplica por el factor específico" (MIDEPLAN, 2023, p.12). Para el caso de "Alquiler y arrendamiento de licencias, derechos de autor, patentes y franquicias", se aplica un factor de 1.06.

b) Mano de Obra: El personal del proyecto se clasifica como mano de obra calificada alta, definida por MIDEPLAN (2023) como aquella que requiere "un grado de estudios avanzados profesional o técnico". Para su corrección a precio social:
   a. Se divide el valor entre 1.4415 para descontar costos asociados
   b. Se multiplica por 1.2213 para agregar beneficios percibidos
   c. Con fines de simplificar estas dos operaciones, se utilizará el siguiente factor:

$$\frac{x}{1.4415} \cdot 1.2213 \approx 0.847242$$

c) Capacitaciones: Las capacitaciones se clasifican como "Servicios de enseñanza", considerados bienes no transables. Según MIDEPLAN (2023), no requieren ajuste y mantienen su precio de mercado.

La **Tabla *32*** presenta el cálculo detallado de los precios sociales para cada componente del proyecto.

Tabla 32
Precios sombra del proyecto.

Año	Fase	Concepto	Costo	Factor de corrección social	Costo Social
Año 1	Extracción, Transformación y Carga (ETL)	Salarios	₡11,393,729.16	0.847242	₡9,653,251.07
		Capacitación	₡1,300,000	No aplica	₡1,300,000
		Licencias	₡0	1.06	₡0
				Total	₡10,953,251.07
	Caracterización y Selección del Modelo	Salarios	₡7,595,819.44	0.847242	₡6,435,500.71
				Total	₡6,435,500.71
				Anualidad total	**₡17,388,751.78**
Año 2	Entrenamiento y Pruebas	Salarios	₡3,797,909.72	0.847242	₡3,217,750.35
				Total	₡3,217,750.35
	Visualización	Salarios	₡2,373,693.575	0.847242	₡2,011,093.97
		Licencias	₡58,092	1.06	₡61,577.52
				Total	₡2,072,671.49
	Puesta en Marcha	Salarios	₡2,848,432.29	0.847242	₡2,413,312.76
				Total	₡2,413,312.76
				Anualidad total	**₡7,703,734.6**
Año 3	Mantenimiento y Mejora Continua	Salarios	₡2,848,432.29	0.847242	₡2,413,312.76
		Licencias	₡58,092	1.06	₡61,577.52
				Anualidad total	**₡2,474,890.28**
Año 4	Mantenimiento y Mejora Continua	Salarios	₡2,848,432.29	0.847242	₡2,413,312.76
		Licencias	₡58,092	1.06	₡61,577.52
				Anualidad total	**₡2,474,890.28**
Costo social total del proyecto			**₡30,042,266.94**		

Nota: Elaboración propia.

4.6.1.2. Estimación de Beneficios Sociales

La cuantificación de los beneficios sociales del SPGDE se fundamenta en el concepto de ahorro social por la prevención de la deserción estudiantil. Para establecer este valor, primero es necesario determinar el costo social de formación por estudiante en el TEC.

Según el Estado de la Educación (2021, p.273) "El arancel anual de matrícula por cursos en universidades públicas representa menos del 10% del costo de formación" y específicamente en el TEC "alcanza los 520 mil colones por año" para 12 créditos. Para determinar el impacto económico del SPGDE, es necesario primero establecer el costo de formación por estudiante en el TEC. Según el Estado de la Educación (2021, p.273), "el arancel anual de matrícula por cursos en universidades públicas representa menos del 10% del costo de formación" y específicamente en el TEC "alcanza los 520 mil colones por año" para 12 créditos.

A partir de los datos del Estado de la Educación (2021), se puede establecer el costo real de formación por estudiante:

- Arancel anual máximo (10% del costo): ₡520,000.
- Costo anual real por estudiante (100%): ₡5,200,000.
- Costo anual total financiado por el estado: ₡5,200,000 - ₡520,000 = ₡4,680,000

Esta estimación se valida mediante la metodología propuesta por la OCDE (2023), que utiliza el indicador de gasto total entre matrícula como proxy para estimar el costo por estudiante. Además, según los datos institucionales más recientes, el TEC cuenta con una población estudiantil de 12,738 estudiantes (TEC, 2024).

Para calcular los beneficios sociales del proyecto, se consideran los siguientes componentes:

a) Inversión Social Actual en Riesgo:
    a. Población estudiantil total: 12,738 estudiantes
    b. Tasa actual de deserción: 11%
    c. Cantidad actual de desertores esperados: 1,401 estudiantes
    d. Pérdida social anual: ₡6,556,680,000 (1,401 × ₡4,680,000)
b) Beneficio Social Proyectado: Según el Plan Estratégico Institucional 2022-2026 del TEC, la meta es reducir la deserción en un 1.75% lo que representa.
    a. Estudiantes beneficiados: 222 (12,738 × 0.0175)
    b. Beneficio social anual: ₡1,038,960,000 (222× ₡4,680,000)

Siguiendo la metodología de MIDEPLAN (2022), se identifican los siguientes tipos de beneficios sociales para el SPGDE:

a) Beneficios Sociales Directos:

- Beneficio por ahorro de costos (liberación de recursos).
- Reducción en inversión pública perdida por deserción: ₡4,680,000por estudiante por año.
- Este beneficio representa recursos públicos que se aprovechan efectivamente en lugar de perderse.
- Cuantificación: 222 estudiantes × ₡4,680,000= ₡1,038,960,000 anuales.

b) Beneficios Sociales Indirectos:

- Reducción en costos de servicios estudiantiles subutilizados.

- Mejor utilización de infraestructura y recursos académicos.
- Aumento en la eficiencia de programas de apoyo estudiantil.
- Cuantificación: No valorizado monetariamente, pero contribuye positivamente.

c) Beneficios Sociales Intangibles:

- Mejora en la reputación institucional del TEC.
- Aumento en la motivación estudiantil general.
- Fortalecimiento del ambiente académico.
- Mayor satisfacción del personal docente y administrativo.
- Cuantificación: No valorizable monetariamente.

d) Externalidades Positivas:

- Mayor cantidad de profesionales calificados para el mercado laboral.
- Contribución al desarrollo socioeconómico del país.
- Mejor aprovechamiento de la inversión pública en educación superior.
- Cuantificación: No valorizado monetariamente, pero genera impacto social positivo.

4.6.1.3. Estimación de Costos Sociales

Siguiendo la metodología de MIDEPLAN (2022), se identifican los siguientes costos sociales:

a) Costos Sociales Directos:

1. Costos de Inversión:
   - Desarrollo e implementación del sistema: ₡17,388,751.78 (año 0).
   - Capacitación del personal: ₡1,300,000.
   - Licencias y software: Ajustados según precio social.
2. Costos de Operación:
   - Personal técnico y analítico: Ajustados según precio social de mano de obra calificada.
   - Mantenimiento de sistemas: ₡2,474,890.28 anuales (años 3 y 4).
   - Actualizaciones de software: Ajustados según precio social.
3. Costos de Mantenimiento:
   - Mantenimiento técnico: Incluido en costos de operación.
   - Actualizaciones periódicas del modelo: Incluido en costos de operación.

b) Costos Sociales Indirectos:

- Tiempo dedicado por personal académico y administrativo a la implementación.
- Recursos institucionales dedicados al proyecto.
- Cuantificación: Incluidos en los costos directos ajustados.

c) Costos Sociales Intangibles:

- Posible resistencia inicial al cambio.
- Período de adaptación del personal.
- Cuantificación: No valorizable monetariamente.

d) Externalidades Negativas:

- No se identifican externalidades negativas significativas.
- El proyecto no genera impactos negativos en otros mercados o sectores.

La **Tabla *33*** resume los costos sociales del proyecto por tipo.

Tabla 33
Resumen de Costos Sociales del SPGDE (en colones).

Tipo de costo social	Año 0	Año 1	Año 2	Año 3
Costos directos				
Personal técnico y analítico	₡16,088,751.78	₡7,642,157.08	₡2,413,312.76	₡2,413,312.76
Licencias de software	₡0	₡61,577.52	₡61,577.52	₡61,577.52
Capacitación	₡1,300,000	-	-	-
Costos Indirectos				
Tiempo personal administrativo*	No valorizado	No valorizado	No valorizado	No valorizado
Recursos institucionales*	No valorizado	No valorizado	No valorizado	No valorizado
Costos sociales totales	₡17,388,751.78	₡7,703,734.6	₡2,474,890.28	₡2,474,890.28

Nota: Elaboración propia. Los costos indirectos se consideran absorbidos dentro de las operaciones normales del TEC.

En base a lo anterior, se puede notar las siguientes características:

1. Estructura de Costos:
    - Los costos de inversión representan el 57.89% del costo social total.
    - Los costos de operación a partir del segundo año al cuarto constituyen el 42.12%.
    - Los costos de mantenimiento están integrados en la operación (8.24%).
2. Ajustes Sociales Aplicados:
    - Mano de obra calificada: Factor de corrección social de 0.847 (1.2213/1.4415).
    - Bienes transables (software): Factor de corrección de 1.06.
    - Capacitación: Sin ajuste por ser servicio no transable.

3. Características Relevantes:
   - Alta concentración de costos en el año inicial.
   - Reducción significativa de costos en años posteriores.
   - Ausencia de costos de infraestructura nueva.
   - Aprovechamiento de recursos existentes.
4. Eficiencia en Costos:
   - Utilización de infraestructura existente.
   - Integración con sistemas actuales.
   - Aprovechamiento de personal interno.
   - Minimización de costos externos.

Este análisis detallado de costos sociales, siguiendo la metodología MIDEPLAN, demuestra una estructura eficiente donde la mayor parte de la inversión se concentra en el desarrollo inicial del sistema, con costos operativos moderados en los años subsiguientes. La ausencia de necesidades de infraestructura nueva y el aprovechamiento de recursos existentes contribuyen a la eficiencia general del proyecto desde una perspectiva social.

4.6.1.4. Indicadores Económico – Sociales

Tal y como se mencionó anteriormente, a continuación, se desarrollan los indicadores económicos propuestos por el MIDEPLAN, sin embargo, estos cálculos son el máximo posible calculado, ya que, por limitaciones del proyecto, no se cuenta con estimaciones detalladas de costos asociados como las intervenciones, estrategias o mecánicas que puedan reducir este índice de deserción de forma directa y que puedan afectar a este beneficio.

a) Valor Actual Neto Económico (VANE):

Siguiendo la metodología de MIDEPLAN (2022), se utiliza la tasa social de descuento del 8.31% para Costa Rica. El VANE se calcula como:

$$VANE = -I_0 - CTERE - CTERE_{Social} + \sum_{i=1}^{n} \frac{B_i - C_i}{(1 + TSD)^i}$$

Donde:

- I_0: Inversión inicial
- B_i: Beneficio social del período i.
- C_i: Costo social del período i.
- i: Periodo.
- N: Horizonte de evaluación del proyecto.
- TSD: Tasa Social de Descuento.
- CTERE: Costo Total Esperado de todos los Riesgos residuales a precios sociales.

- CTERE Social: Costo Total Esperado de todos los Riesgos residuales asociados a la pérdida de bienestar social a precios sociales.

Se obtiene que:

$$VANE = -₡17{,}388{,}751.78 + \frac{(₡0 - ₡7{,}703{,}734.6)}{(1+0.0831)} + \frac{(₡1{,}038{,}960{,}000 - ₡2{,}474{,}890.28)}{(1+0.0831)^2} + \frac{(₡1{,}038{,}960{,}000 - ₡2{,}474{,}890.28)}{(1+0.0831)^3}$$

$$VANE \approx ₡1{,}674{,}788{,}643$$

b) Tasa Interna de Retorno Económico (TIRE):

La TIRE es la tasa que hace que el VANE sea igual a cero. Se calcula con la siguiente fórmula:

$$VANE = 0 = -I_0 - CTERE - CTERE_{Social} + \sum_{i=1}^{n} \frac{B_i - C_i}{(1+TIRE)^i}$$

Donde:

- VANE: Valor Actual Neto Económico.
- I_0: Inversión inicial
- B_i: Beneficio social del período i.
- C_i: Costo social del período i.
- TIRE: Tasa Interna de Retorno Económico a determinar.
- CTERE: Costo Total Esperado de todos los Riesgos residuales a precios sociales.
- CTERE Social: Costo Total Esperado de todos los Riesgos residuales asociados a la pérdida de bienestar social a precios sociales.

Se obtiene que:

$$₡1{,}674{,}788{,}643 = 0 = -₡17{,}388{,}751.78 + \frac{(₡0 - ₡7{,}703{,}734.6)}{(1+TIRE)} + \frac{(₡1{,}038{,}960{,}000 - ₡2{,}474{,}890.28)}{(1+TIRE)^2} + \frac{(₡1{,}038{,}960{,}000 - ₡2{,}474{,}890.28)}{(1+TIRE)^3}$$

$$TIRE \approx 6.97 = 697\%$$

c) Relación Beneficio-Costo (R-B/C):

Esta demuestra la relación entre el beneficio y costo, revelando el beneficio obtenido por cada unidad costeada, la fórmula para su cálculo es la siguiente:

$$R\frac{B}{C} = \frac{\sum_{t=0}^{n} \frac{B_t}{(1+TSD)^t}}{I_0 - CTERE - CTERE_{Social} + \sum_{t=0}^{n} \frac{C_t}{(1+TSD)^t}}$$

Donde:

- I_0: Inversión inicial
- B_t: Beneficios sociales del año t del proyecto.
- C_t: Costos sociales del año t del proyecto (Incluye la inversión inicial para el período 0)
- n: Años correspondientes al horizonte de evaluación del proyecto.
- TSD: Tasa Social de Descuento.
- CTERE: Costo Total Esperado de todos los Riesgos residuales a precios sociales.
- CTERE Social: Costo Total Esperado de todos los Riesgos residuales asociados a la pérdida de bienestar social a precios sociales.
- 0 = Año inicial del proyecto, en el cual comienza la inversión.

Se obtiene que:

$$R\frac{B}{C} = \frac{\frac{₡0}{(1+0.0831)^0} + \frac{₡0}{(1+0.0831)} + \frac{₡1{,}038{,}960{,}000}{(1+0.0831)^2} + \frac{₡1{,}038{,}960{,}000}{(1+0.0831)^3}}{₡17{,}388{,}751.78 + \frac{₡17{,}388{,}751.78}{(1+0.0831)^0} + \frac{₡7{,}703{,}734.6}{(1+0.0831)} + \frac{₡2{,}474{,}890.28}{(1+0.0831)^2} + \frac{₡2{,}474{,}890.28}{(1+0.0831)^3}}$$

$$R\frac{B}{C} \approx 37.07$$

4.6.1.5. Análisis de sensibilidad

El análisis de sensibilidad es fundamental para evaluar la robustez del proyecto ante variaciones en sus variables críticas. Siguiendo las recomendaciones de MIDEPLAN (2022), se analizan tres escenarios que consideran cambios en los beneficios sociales esperados.

Escenario 1: Base

- Reducción de deserción: 1% (del 11% al 10%).
- Estudiantes beneficiados: 127 (12,738 × 0.01).
- Beneficio social anual: ₡594,360,000.

Escenario 2: Pesimista

- Reducción de deserción: 0.5% (del 11% al 10.5%).
- Estudiantes beneficiados: 64 (12,738 × 0.005).
- Beneficio social anual: ₡299,520,000.

Escenario 3: Optimista

- Reducción de deserción: 1.75% (del 11% al 9.25%).

- Estudiantes beneficiados: 223 (12,738 × 0.0175).
- Beneficio social anual: ₡1,043,640,000.

Los indicadores económico-sociales para cada escenario se presentan en la Tabla 4.6.6:

Indicador	Base	Pesimista	Optimista
VANE	₡945,878,641.16	₡462,496,218.58	₡1,682,461,380.33
TIRE	507.83%	336.59%	698.94%
R-B/C	21.21	10.69	37.24

Nota: Elaboración propia.

Los resultados del análisis de sensibilidad demuestran que:

1. Incluso en el escenario pesimista, con una reducción muy conservadora de solo 0.5% en la deserción, el proyecto mantiene indicadores positivos:
    - VANE se mantiene positivo en ₡462,496,218.58.
    - TIRE permanece muy superior a la tasa social de descuento (8.31%).
    - R-B/C se mantiene significativamente por encima de 1.
2. La variación en los indicadores muestra una elasticidad proporcional a los cambios en los beneficios:
    - Una reducción del 50% en los beneficios (escenario pesimista) resulta en una reducción aproximada del 50% en los indicadores.
    - Un aumento del 75% en los beneficios (escenario optimista) resulta en un aumento aproximado del 75% en los indicadores.
3. La robustez de los resultados indica que el proyecto mantiene su viabilidad social incluso con metas muy conservadoras:
    - Los indicadores permanecen favorables incluso con una reducción de deserción de solo 0.5%.
    - El margen de seguridad es amplio, considerando que el Plan Estratégico del TEC establece metas más ambiciosas.

Este análisis de sensibilidad, realizado con supuestos más conservadores, confirma que la viabilidad social del proyecto SPGDE se mantiene sólida incluso ante escenarios menos optimistas y metas más modestas de reducción de la deserción.

4.6.2. Evaluación de Fortalecimiento de Capacidades Institucionales

La evaluación del fortalecimiento de capacidades institucionales representa un componente fundamental en proyectos de modernización del sector público. Como señala MIDEPLAN (2022, p.112), "la medición del fortalecimiento institucional debe centrarse en la mejora de procesos, sistemas y capacidades organizacionales que permitan una gestión más eficiente y efectiva de los recursos públicos".

En el contexto específico del SPGDE, esta evaluación cobra particular relevancia dado que el proyecto busca transformar la capacidad institucional del TEC para:

- Gestionar datos de manera integrada
- Tomar decisiones basadas en evidencia
- Coordinar esfuerzos interdepartamentales
- Optimizar la asignación de recursos de apoyo estudiantil

Por otra parte, siguiendo el marco propuesto por el Banco Interamericano de Desarrollo (BID, 2021) para la evaluación de capacidades institucionales, se establecen tres dimensiones principales de análisis, como se muestra en la **Tabla *34***.

Tabla 34
Dimensiones de Evaluación de Capacidades Institucionales.

Dimensión	Descripción	Indicadores clave
Tecnológica	Capacidad de gestión y análisis de datos	- Integración de fuentes de datos. - Tiempo de procesamiento. - Automatización de procesos.
Organizacional	Mejora en procesos institucionales	- Coordinación interdepartamental. - Tiempo de respuesta. - Efectividad en intervenciones.
Operativa	Eficiencia en uso de recursos	- Optimización de recursos. - Cobertura de servicios. - Focalización de intervenciones.

Nota: Elaboración propia.

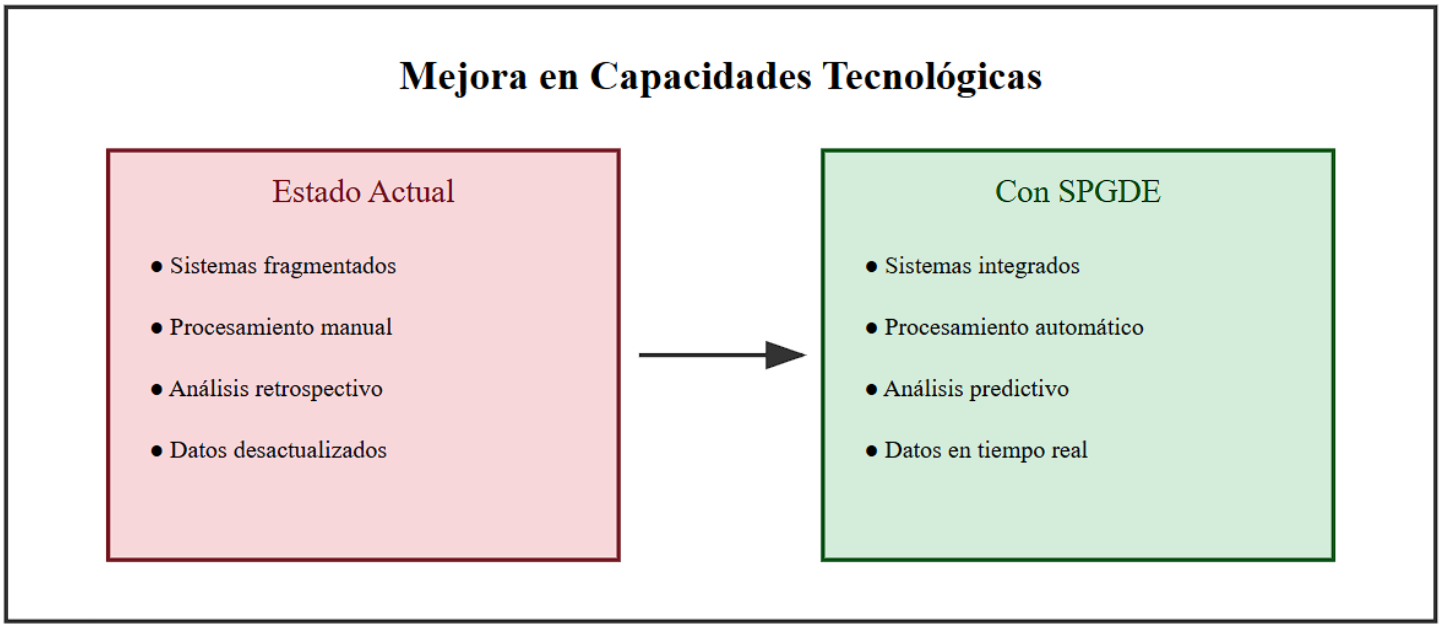
A continuación, se realiza el análisis de capacidades por cada dimensión anteriormente mostrada.

4.6.2.1. Capacidad Tecnológica

La **Figura *4*** ilustra la transformación esperada en las capacidades tecnológicas institucionales tras la implementación del SPGDE, a continuación, se listan estos aspectos.

a) Situación Actual: Basado en el análisis documentado en la sección 4.1 del proyecto, actualmente el TEC presenta
    - Sistemas de información fragmentados.
    - Procesamiento manual de datos.
    - Limitada capacidad de análisis predictivo.
b) Mejoras Proyectadas: El SPGDE fortalecerá las capacidades tecnológicas mediante
    - Integración de Sistemas:
        a. Conexión de 24 fuentes de datos identificadas.
        b. Automatización de procesos de ETL.
        c. Implementación de pipeline de datos.
    - Capacidad Analítica:
        a. Desarrollo de modelos predictivos.
        b. Implementación de *dashboards* interactivos.
        c. Actualización automática de predicciones.

Figura 4
Mejora en capacidades tecnológicas.



Nota: Elaboración propia.

4.6.2.2. Capacidad Organizacional

a) Análisis de Procesos Actuales: La revisión de los procesos actuales, documentada en las entrevistas con personal clave del TEC (ver Apéndice F: Tabla de hallazgos de entrevista a M.Ed. Teresa Hernández Jiménez., Apéndice G: Tabla de hallazgos de entrevista al Profesor Martín Solís Salazar., Apéndice H: Tabla de hallazgos de entrevista a Evaluadora e investigadora jubilada Tania Elena Moreira Mora., Apéndice I: Tabla de hallazgos de entrevista a Arlyn Angulo Hernández y Raquel Mejías Elizondo de la OPI.), revela:
    - Limitada coordinación interdepartamental.
    - Duplicación de esfuerzos.
    - Tiempo de respuesta prolongado en intervenciones.
b) Mejoras en Procesos Institucionales:
    - Coordinación Interdepartamental:
        - Establecimiento de protocolos de comunicación.
        - Definición clara de roles y responsabilidades.
        - Mecanismos de seguimiento conjunto.
    - Optimización de Recursos:
        - Mejor asignación de personal de apoyo.
        - Reducción de duplicidad en intervenciones.
        - Priorización basada en evidencia.

4.6.2.3. Capacidad Operativa

a) Diagnóstico de Capacidad Operativa Actual: El análisis de la capacidad operativa actual del TEC, basado en las entrevistas realizadas (ver Apéndice F: Tabla de hallazgos de entrevista a M.Ed. Teresa Hernández Jiménez., Apéndice G: Tabla de hallazgos de entrevista al Profesor Martín Solís Salazar., Apéndice H: Tabla de hallazgos de entrevista a Evaluadora e investigadora jubilada Tania Elena Moreira Mora., Apéndice I: Tabla de hallazgos de entrevista a Arlyn Angulo Hernández y Raquel Mejías Elizondo de la OPI.)y la documentación institucional, revela las siguientes condiciones mostradas en la **Tabla *35***.

Tabla 35
Diagnóstico de Capacidad Operativa Actual.

Aspecto Operativo	Situación Actual	Limitaciones Identificadas
Identificación de casos.	Manual y reactiva.	Detección tardía de casos en riesgo.
Tiempo de respuesta.	Variable (2-4 semanas).	Intervenciones no oportunas.
Cobertura de servicios.	Parcial y no sistemática.	Estudiantes sin atención necesaria.
Seguimiento de casos.	No estandarizado.	Falta de continuidad en intervenciones.

Nota: Elaboración propia.

b) Fortalecimiento de Capacidades Operativas: El SPGDE propone fortalecer las capacidades operativas en tres niveles principales.
   - Mejora en Procesos de Detección:
     - Implementación de alertas tempranas automatizadas.
     - Monitoreo continuo de indicadores de riesgo.
     - Actualización semestral de predicciones.
   - Optimización de Intervenciones:
     - Priorización basada en niveles de riesgo.
     - Asignación eficiente de recursos de apoyo.
     - Seguimiento sistemático de casos.
   - Gestión de Información:
     - Centralización de datos estudiantiles.
     - Acceso controlado según roles.
     - Actualización en tiempo real.

A continuación, la **Figura *5*** ilustra el ciclo de mejora operativa propuesto por el SPGDE, destacando la naturaleza iterativa y continua del proceso de fortalecimiento institucional.

Figura 5
Ciclo de mejora operativa del SPGDE.

Nota: Elaboración propia.

4.6.2.4. Medición y Evaluación de Mejoras en Capacidades

Siguiendo las recomendaciones de MIDEPLAN (2022) para la evaluación de proyectos de fortalecimiento institucional, se establecen los siguientes indicadores clave mostrados en la **Tabla *36***.

Tabla 36
Indicadores de Desempeño para Evaluación de Capacidades.

Dimensión	Indicador	Línea base	Meta	Medio de verificación
Tecnológica	Tiempo de procesamiento de datos	2-3 semanas	24-48 horas	Registros del sistema.
	Integración de fuentes de datos	4 fuentes	24 fuentes	Documentación técnica.
Organizacional	Tiempo de respuesta a casos	2-4 semanas	1 semana	Registros de intervención.
	Cobertura de servicios	60% estimado	90%	Informes de gestión.

Dimensión	Indicador	Línea base	Meta	Medio de verificación
Operativa	Precisión en identificación	No disponible	85%	Validación del modelo.
	Efectividad de intervenciones	No medido	75%	Seguimiento de casos.

Nota: Elaboración propia.

Es importante notar que la línea base actual se ha establecido a partir de la información disponible en los registros institucionales y las entrevistas realizadas con personal clave del TEC. Sin embargo, como señala la investigadora Tania Elena Moreira Mora, "la falta de sistematización actual dificulta la medición precisa de algunos indicadores" (ver Apéndice H: Tabla de hallazgos de entrevista a Evaluadora e investigadora jubilada Tania Elena Moreira Mora., EH140).

4.6.3. Análisis de Beneficios Sociales

El análisis de beneficios sociales del SPGDE se fundamenta en el enfoque de valor público propuesto por Moore (2013) y adaptado por MIDEPLAN (2022) para la evaluación de proyectos de fortalecimiento institucional. Este marco considera tres dimensiones principales:

- Valor institucional: Mejoras en la capacidad y eficiencia organizacional.
- Valor social: Beneficios para la población objetivo y la sociedad.
- Valor sistémico: Contribución al sistema educativo superior.

Estas son evaluadas bajo los indicadores expuestas en la **Tabla *37***, que serán evidenciados a continuación.

Tabla 37
Marco de Análisis de Beneficios Sociales.

Dimensión	Alcance	Indicadores de valor
Institucional.	Capacidades TEC.	- Eficiencia operativa. - Uso de recursos.
Social.	Comunidad estudiantil.	- Acceso a apoyo. - Equidad en intervenciones.
Sistémico.	Sistema educativo.	- Retención estudiantil. - Uso de recursos públicos.

Nota: Elaboración propia.

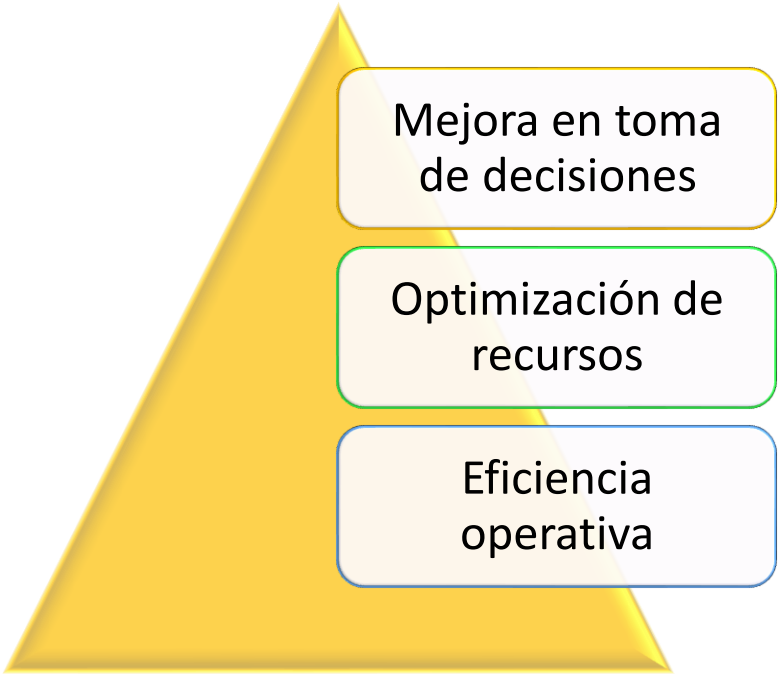
4.6.3.1. Beneficios Institucionales

a) Eficiencia Operativa: Basado en el análisis de procesos actuales (ver sección 4.1), se identifican las siguientes mejoras.
   - Optimización de Procesos:
     - Reducción en tiempo de procesamiento de datos.
     - Automatización de tareas rutinarias.
     - Mejora en seguimiento de casos.
   - Uso Eficiente de Recursos:
     - Mejor asignación de personal de apoyo.

  - Reducción de duplicidad en intervenciones.
  - Optimización de servicios estudiantiles.
  - Capacidad de Respuesta: Como señala Teresa Hernández en la entrevista (ver Apéndice F: Tabla de hallazgos de entrevista a M.Ed. Teresa Hernández Jiménez., EH24), la capacidad de respuesta institucional mejorará significativamente mediante.
- Detección Temprana:
  - Identificación proactiva de casos en riesgo.
  - Intervención oportuna.
  - Seguimiento sistemático.
- Coordinación Interdepartamental:
  - Comunicación efectiva entre unidades.
  - Respuesta integrada a necesidades estudiantiles.
  - Seguimiento conjunto de casos.

A continuación, la **Figura *6*** ilustra la jerarquía de beneficios institucionales esperados del SPGDE, desde mejoras operativas básicas hasta transformaciones en la toma de decisiones.

Figura 6
Beneficios institucionales.



Nota: Elaboración propia.

4.6.3.2. Beneficios Sociales

a) Impacto en la Comunidad Estudiantil: El análisis del impacto social se fundamenta en la información recopilada a través de entrevistas con actores clave y documentación institucional, obteniendo como resultado lo mostrado en la **Tabla *38***.

Tabla 38
Análisis de Beneficios para la Comunidad Estudiantil.

Beneficio	Descripción	Evidencia
Acceso a apoyo temprano	Identificación proactiva de necesidades	Entrevista DOP (Apéndice F: Tabla de hallazgos de entrevista a M.Ed. Teresa Hernández Jiménez.)
Equidad en intervenciones	Asignación objetiva de recursos	Análisis de datos actual
Mejora en seguimiento	Continuidad en apoyo estudiantil	Procesos propuestos

Nota: Elaboración propia.

b) Impacto en la Equidad Educativa: Siguiendo los principios de equidad establecidos en el Plan Nacional de Educación Superior (CONARE, 2021), el SPGDE contribuirá a.
   - Acceso Equitativo a Servicios:
     - Identificación sistemática de necesidades
     - Distribución objetiva de recursos
     - Seguimiento estandarizado
   - Reducción de Brechas:
     - Atención temprana a estudiantes vulnerables
     - Apoyo focalizado según necesidades
     - Intervenciones personalizadas

4.6.3.3. Beneficios Sistémicos

- Contribución al Sistema de Educación Superior: El SPGDE contribuye al fortalecimiento del sistema de educación superior pública de Costa Rica, alineándose con los objetivos establecidos en el Plan Nacional de la Educación Superior Universitaria Estatal 2021-2025 (CONARE, 2021), tal como se indica en la **Tabla *39***.

Tabla 39
Alineación con Objetivos del Sistema de Educación Superior.

Objetivo PLANES	Contribución del SPGDE	Indicador de Impacto
Eficiencia en uso de recursos públicos.	Optimización de intervenciones.	Mejor asignación de recursos de apoyo.
Mejora en permanencia estudiantil.	Detección temprana de riesgo.	Intervenciones más efectivas.
Fortalecimiento institucional.	Modernización de procesos.	Capacidad analítica mejorada.

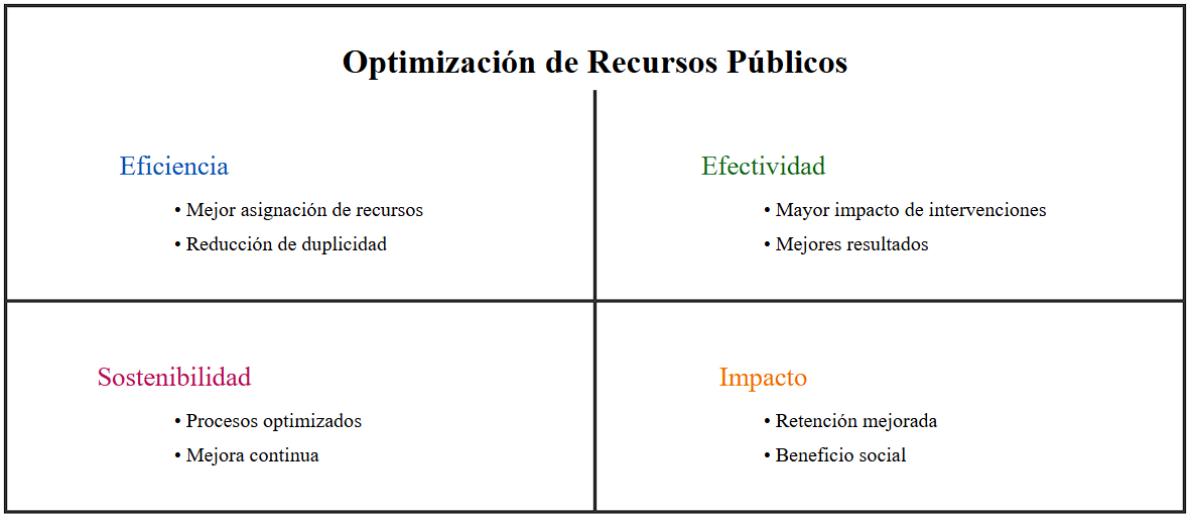
Nota: Elaboración propia.

- Optimización de Recursos Públicos: Como señala MIDEPLAN (2022, p.156), la optimización de recursos públicos debe evaluarse en términos de.
  - Eficiencia en el Uso de Recursos:
    - Mejor focalización de intervenciones.
    - Reducción de duplicidad de esfuerzos.

  - Optimización de servicios estudiantiles.
- Efectividad de la Inversión Pública:
  - Mayor retorno social de la inversión educativa.
  - Reducción de costos por deserción.
  - Mejora en indicadores de desempeño institucional.

A continuación, la **Figura *7***, ilustra las dimensiones clave en la optimización de recursos públicos a través del SPGDE.

Figura 7
Optimización de Recursos Públicos.



Nota: Elaboración propia.

4.6.4. Evaluación Integral de Impacto Social

La evaluación integral del SPGDE debe considerar tanto los beneficios cuantificables como los no cuantificables, siguiendo el marco de evaluación social de MIDEPLAN (2022), se muestra la **Tabla *40*** con estos aspectos.

Tabla 40
Marco de Evaluación Integral de Impacto Social.

Dimensión	Indicadores Cuantificables	Beneficios No Cuantificables
Tecnológica	Tiempo de respuesta reducido	Mejora en cultura organizacional.
	Recursos optimizados	Fortalecimiento institucional.

Dimensión	Indicadores Cuantificables	Beneficios No Cuantificables
Organizacional	Cobertura de servicios	Equidad en acceso.
	Efectividad de intervenciones	Bienestar estudiantil.
Operativa	Uso eficiente de recursos	Modernización de procesos.
	Indicadores de desempeño	Capacidad institucional.

Nota: Elaboración propia.

4.7. Plan de implementación sugerido del proyecto

4.7.1. Desglose de proyectos

Esta sección ofrece una descripción concisa de los proyectos propuestos con el fin de completar el proyecto evaluado SPGDE durante un periodo de tiempo de 4 años. En el Apéndice P: Proyectos sugeridos detallados., se proporciona un detalle exhaustivo de estos proyectos, abarcando aspectos como objetivos, alcance, justificación, partes interesadas y otros elementos cruciales, estimados a alto nivel, por lo que se recomienda una evaluación más detallada de estos. A continuación, en la **Tabla *41*** se presentan las iniciativas propuestas:

Tabla 41
Proyectos sugeridos

Identificador	Nombre	Descripción	Costo Estimado
P001	Análisis y Preparación de Datos para el SPGDE	Este proyecto se enfoca en la identificación, recopilación, limpieza y estructuración de los datos necesarios para alimentar el modelo predictivo de deserción estudiantil.	₡12,693,729.16

Identificador	Nombre	Descripción	Costo Estimado
P002	Desarrollo del Modelo Predictivo de Deserción Estudiantil	Este proyecto se centra en el desarrollo, entrenamiento y validación de un modelo de aprendizaje automático capaz de predecir con alta precisión el riesgo de deserción de los estudiantes del TEC.	₡7,595,819.44
P003	Integración del SPGDE con Sistemas Existentes del TEC	Este proyecto se enfoca en la integración *seamless* del SPGDE con los sistemas existentes del TEC, asegurando un flujo de datos en tiempo real y la incorporación de las predicciones del modelo en los procesos operativos de la institución.	₡6,646,342.01
P004	Diseño e Implementación de Interfaces de Usuario para el SPGDE	Este proyecto se centra en el diseño y desarrollo de interfaces de usuario intuitivas y eficaces para el SPGDE, facilitando el acceso y la interpretación de las predicciones y recomendaciones generadas por el sistema.	₡2,431,785.575

Identificador	Nombre	Descripción	Costo Estimado
P005	Implementación de un Sistema de Monitoreo y Evaluación Continua del SPGDE	Este proyecto se centra en el desarrollo e implementación de un sistema integral de monitoreo y evaluación continua para el SPGDE, asegurando su efectividad y relevancia a lo largo del tiempo.	₡2,964,616.29
P006	Desarrollo de un Marco de Gobernanza de Datos para el SPGDE	Este proyecto se enfoca en el desarrollo e implementación de un marco integral de gobernanza de datos para el SPGDE, estableciendo políticas, procesos y estructuras organizativas que aseguren la calidad, integridad, seguridad y uso ético de los datos.	₡12,000,000

Nota: Elaboración propia,

4.7.2. Hoja de ruta detallada

En la presente sección, se presenta la **Figura *8*** correspondiente a la hoja de ruta a seguir para la implementación de los proyectos que fueron identificados. Además, se define el costo del proyecto en colones, así como el periodo de inicio.

Figura 8
Hoja de ruta propuesta.

ID	Nombre del proyecto	2025		2026		2027	Costo estimado (₡)	Periodo de inicio
		I S	II S	I S	II S	I S		
P001	Análisis y Preparación de Datos para el SPGDE						12,693,729.16	I Semestre 2025
P002	Desarrollo del Modelo Predictivo de Deserción Estudiantil						7,595,819.44	II Semestre 2025
P006	Desarrollo de un Marco de Gobernanza de Datos para el SPGDE						12,000,000.00	II Semestre 2025
P004	Diseño e Implementación de Interfaces de Usuario para el SPGDE						2,431,785.58	I Semestre 2026
P003	Integración del SPGDE con Sistemas Existentes del TEC						6,646,342.01	II Semestre 2026
P005	Implementación de un Sistema de Monitoreo y Evaluación Continua del SPGDE						2,964,616.29	I Semestre 2027

Nota: Elaboración propia.

4.7.3. Métricas de éxito y evaluación

Con el fin de determinar el éxito del SPGDE y sus proyectos establecidos, se establecen las siguientes métricas clave:

1. Reducción en la tasa de deserción:

- Objetivo: Disminución del 10% en la tasa de deserción de primer año al final del proyecto.
- Medición: Comparación anual de tasas de deserción.

2. Precisión del modelo predictivo:
   - Objetivo: Alcanzar y mantener una precisión del 85% o superior en la identificación de estudiantes en riesgo.
   - Medición: Evaluación trimestral de la precisión del modelo.
3. Adopción del sistema:
   - Objetivo: 80% de los departamentos académicos utilizando activamente el SPGDE al final del proyecto.
   - Medición: Seguimiento mensual del uso del sistema por departamento.
4. Satisfacción del usuario:
   - Objetivo: 75% de satisfacción entre los usuarios del sistema (profesores, asesores, administrativos).
   - Medición: Encuestas semestrales de satisfacción del usuario.
5. Efectividad de las intervenciones:
   - Objetivo: 60% de los estudiantes identificados como en riesgo muestran mejora después de las intervenciones.
   - Medición: Seguimiento del progreso académico de los estudiantes intervenidos.
6. Retorno de la inversión:
   - Objetivo: Demostrar un ROI positivo al final del proyecto, considerando los costos evitados por la reducción en la deserción.
   - Medición: Análisis anual de costos-beneficios.

Estas métricas serán evaluadas regularmente como parte del proyecto P005 (Implementación de un Sistema de Monitoreo y Evaluación Continua del SPGDE), permitiendo ajustes y mejoras continuas al sistema.

4.8. Comprobación de hipótesis

A través del análisis realizado en las diferentes evaluaciones y estudios del proyecto, se puede proceder a la verificación de las hipótesis planteadas:

4.8.1. Verificación de H1: La utilidad práctica del modelo predictivo

El análisis del estado actual (sección Estado actual de modelos predictivos en Costa Rica.) y las entrevistas con actores clave revelaron que el modelo predictivo anterior del TEC alcanzó una precisión notable del 91% (ver

sección Estado actual de los modelos predictivos), pero, su impacto práctico fue limitado por la falta de directrices institucionales para su uso.

Como se detalla en la sección Análisis de demanda, la demanda existe para este tipo de herramientas, evidenciada por el interés activo de la Vicerrectoría de VIESA y la inclusión de metas de reducción de deserción en el Plan Estratégico 2022-2026. Sin embargo, la evaluación organizacional (sección Descripción del proyecto propuesto) demuestra que, sin procesos claros de intervención, el valor práctico se ve severamente limitado.

Por lo tanto, se confirma H1: la utilidad práctica depende más de factores organizacionales que de la precisión técnica del modelo.

4.8.2. Verificación de H2: Factores críticos más allá de lo técnico

La evaluación de riesgos (sección Evaluación de riesgos) identificó múltiples factores no técnicos críticos para el éxito del proyecto, incluyendo:

- Apoyo institucional y voluntad política
- Procesos de gobernanza de datos
- Capacidad organizacional para intervenciones
- Marco ético y legal robusto

El análisis del proyecto anterior (sección Estado actual de los modelos predictivos) y las entrevistas realizadas confirman que, a pesar de contar con recursos técnicos y datos suficientes, el proyecto no logró sostenibilidad por carecer de estos elementos organizacionales.

Se confirma así H2: el éxito del proyecto depende significativamente de factores no técnicos, siendo estos tan o más críticos que los aspectos técnicos y de datos.

Esta verificación de hipótesis refuerza la importancia del enfoque integral propuesto en el plan de implementación (sección Plan de implementación sugerido del proyecto), que considera tanto aspectos técnicos como organizacionales para asegurar el éxito sostenible del proyecto.

4.8.3. Verificación de H3: Efectividad de la predicción temprana

El análisis de literatura y experiencias previas (sección Evolución de los Estudios sobre Deserción Universitaria) demuestra que la identificación temprana de estudiantes en riesgo permite intervenciones más efectivas. Como se evidencia en la entrevista con la M.Ed. Teresa Hernández (Apéndice F: Tabla de hallazgos de entrevista a M.Ed. Teresa Hernández Jiménez., EH24-EH26), las intervenciones tempranas tienen mayor probabilidad de éxito que las medidas reactivas.

La evaluación de beneficios no financieros (sección Beneficios no financieros) indica que la predicción temprana permite:

- Intervenciones más personalizadas y oportunas.
- Mejor asignación de recursos de apoyo.
- Mayor probabilidad de retención exitosa.

Se confirma H3: la predicción temprana es una herramienta efectiva cuando se combina con intervenciones adecuadas.

4.8.4. Verificación de H4: Necesidad institucional del TEC

El análisis del contexto actual (sección Estado actual de modelos predictivos en Costa Rica.) revela una situación compleja, por un lado, existe una aparente necesidad institucional evidenciada por:

- Una tasa de deserción del 11% en estudiantes de primer ingreso.
- Meta institucional de reducción al 9% para 2026 (Plan Estratégico 2022-2026).
- La deserción representa una pérdida significativa de recursos públicos y talento.

Sin embargo, el estudio revela que el TEC ya cuenta con herramientas predictivas:

- Un modelo previo desarrollado en 2018-2019 con una precisión del 91% (Apéndice G: Tabla de hallazgos de entrevista al Profesor Martín Solís Salazar., EH75).
- Una iniciativa actual de la OPI para actualizar este modelo (Apéndice I: Tabla de hallazgos de entrevista a Arlyn Angulo Hernández y Raquel Mejías Elizondo de la OPI., EH147-148).
- Un cubo de información en desarrollo con nuevas variables predictivas (Apéndice I: Tabla de hallazgos de entrevista a Arlyn Angulo Hernández y Raquel Mejías Elizondo de la OPI., EH152).

La investigación muestra que el problema no es la falta de herramientas predictivas, sino:

- Ausencia de directrices claras para el uso de las predicciones (Apéndice G: Tabla de hallazgos de entrevista al Profesor Martín Solís Salazar., EH132).
- Falta de personal dedicado al seguimiento del modelo (Apéndice F: Tabla de hallazgos de entrevista a M.Ed. Teresa Hernández Jiménez., EH38-39).
- Desafíos en la continuidad y mantenimiento de las herramientas existentes (Apéndice I: Tabla de hallazgos de entrevista a Arlyn Angulo Hernández y Raquel Mejías Elizondo de la OPI., EH154).

Se refuta H4: El TEC no tiene una necesidad crítica de nuevas herramientas predictivas, sino de fortalecer y operacionalizar las existentes. La verdadera necesidad institucional radica en:

1. Establecer procesos claros para el uso de las predicciones.
2. Asignar recursos para el mantenimiento y seguimiento de los modelos existentes.
3. Desarrollar una estrategia integral de intervención basada en las predicciones.

Esta refutación tiene implicaciones importantes para el proyecto SPGDE:

- El enfoque debe reorientarse hacia la mejora y operacionalización de las herramientas existentes.
- Es más crítico establecer directrices institucionales que desarrollar nuevos modelos.
- Los recursos deberían priorizarse en la implementación efectiva sobre el desarrollo técnico.

Como señala la investigadora Tania Elena Moreira Mora (Apéndice H: Tabla de hallazgos de entrevista a Evaluadora e investigadora jubilada Tania Elena Moreira Mora., EH140): "identificar potenciales desertores no

es suficiente sin acciones concretas para retenerlos". Esta observación refuerza que la verdadera necesidad del TEC no es tecnológica sino organizacional y operativa.

Esta refutación modifica significativamente el enfoque recomendado para el proyecto, como se detalla en el plan de implementación (sección Plan de implementación sugerido del proyecto), priorizando la integración y operacionalización de las herramientas existentes sobre el desarrollo de nuevas soluciones.

4.8.5. Verificación de H5: Necesidad de un modelo predictivo

La evaluación de la competencia (sección Evaluación de la competencia) y el análisis de alternativas revelan que:

- Los enfoques tradicionales no han logrado reducir significativamente la deserción.
- Un modelo predictivo permite una gestión más proactiva y eficiente del problema.
- La escala del TEC hace viable y efectivo el uso de herramientas predictivas.

Sin embargo, como se evidencia en la evaluación de riesgos (sección Evaluación de riesgos), el modelo predictivo debe ser parte de una estrategia integral que incluya:

- Procesos de intervención definidos.
- Recursos dedicados para seguimiento.
- Compromiso institucional sostenido.

Se confirma parcialmente H5: un modelo predictivo es una herramienta necesaria, pero, no suficiente, debiendo ser parte de una estrategia más amplia de retención estudiantil.

Esta verificación de hipótesis adicionales refuerza la importancia de un enfoque holístico en el desarrollo e implementación del SPGDE, considerando tanto aspectos técnicos como institucionales y operativos para garantizar su efectividad en la reducción de la deserción estudiantil.

Capítulo 5. Limitaciones y problemas encontrados

Durante el desarrollo de esta investigación, se encontraron diversos desafíos y limitaciones que impactaron en diferentes aspectos del proyecto. Estos obstáculos, aunque no impidieron el cumplimiento de los objetivos principales, requirieron ajustes en la metodología y el alcance del trabajo. A continuación, se detallan las principales limitaciones y problemas encontrados, categorizados según su naturaleza:

- Limitaciones Administrativas y de Coordinación:

Una de las limitaciones más significativas fue la coordinación con los diferentes departamentos del TEC para la realización de entrevistas. La disponibilidad limitada de los sujetos de información clave, especialmente durante períodos de alta actividad académica, resultó en retrasos en la recolección de datos. Por ejemplo, algunos participantes clave como Tatiana Fernández se pensionaron durante el período de investigación, lo que limitó el acceso a información histórica valiosa sobre proyectos anteriores relacionados con la deserción estudiantil.

Además, la estructura organizacional del TEC, caracterizada por la separación entre departamentos, dificultó la obtención de una visión completa e integrada de los datos y procesos relacionados con la deserción estudiantil. Esto se evidenció especialmente en la fase de análisis de las fuentes de datos disponibles, donde se encontró que la información está distribuida en diferentes sistemas y departamentos sin una integración efectiva.

- Limitaciones Técnicas:

En el aspecto técnico, la principal limitación fue el acceso restringido a las bases de datos institucionales y sistemas de información del TEC. Aunque se logró identificar las fuentes de datos potenciales, las políticas de seguridad y privacidad institucionales limitaron la capacidad de realizar un análisis más profundo de la estructura y calidad de los datos disponibles. Esta restricción impactó particularmente en la capacidad de evaluar con precisión la viabilidad técnica de ciertas propuestas de mejora al modelo predictivo.

También se encontraron limitaciones en cuanto a la documentación técnica disponible sobre los sistemas actuales y anteriores intentos de implementación de modelos predictivos. La falta de documentación detallada sobre estos esfuerzos previos dificultó la evaluación completa de las lecciones aprendidas y mejores prácticas.

- Limitaciones Metodológicas:

La metodología inicial tuvo que ser ajustada debido a la naturaleza sensible de la información relacionada con la deserción estudiantil. Se encontró que algunos aspectos importantes, como los factores psicológicos y socioeconómicos que influyen en la deserción, no podían ser evaluados en profundidad debido a restricciones en el acceso a información confidencial de los estudiantes.

Además, la necesidad de mantener la confidencialidad de ciertos datos y procesos institucionales limitó la capacidad de presentar algunos detalles específicos en los resultados de la investigación, lo que requirió un enfoque más general en ciertas secciones del análisis.

- Limitaciones Temporales:

El marco temporal del proyecto, limitado a un semestre académico, impuso restricciones significativas en el alcance de la investigación. Esto fue particularmente relevante en la fase de evaluación de viabilidad, donde

idealmente se habría requerido un período más extenso para realizar pruebas más exhaustivas y análisis más detallados de ciertos aspectos técnicos y organizacionales.

Los desfases en los tiempos de respuesta de diferentes departamentos y la necesidad de coordinar múltiples entrevistas en un período relativamente corto también impactaron en la profundidad del análisis que se pudo realizar en algunas áreas del estudio.

A pesar de estas limitaciones, es importante señalar que se lograron los objetivos principales del proyecto, adaptando la metodología y el enfoque según fue necesario para superar estos desafíos. Las limitaciones encontradas también proporcionaron información valiosa sobre los aspectos que deberían considerarse en futuras investigaciones o implementaciones relacionadas con el tema de la deserción estudiantil en el TEC.

Capítulo 6. Discusión y Conclusiones

6.1. Discusión

La evaluación de viabilidad del modelo predictivo de deserción estudiantil en el TEC ha revelado hallazgos significativos que merecen un análisis detallado. Esta investigación no solo ha proporcionado respuestas a las hipótesis planteadas, sino que también ha descubierto factores cruciales que influirán en el éxito de cualquier iniciativa futura en este ámbito.

- **Evolución y Estado Actual del Proyecto**

Un descubrimiento significativo fue la existencia de iniciativas previas exitosas en el TEC. El modelo desarrollado por Solis et al. (2018) alcanzó una precisión del 91%, similar a los resultados obtenidos por Delen (2010) en su estudio sobre predicción de deserción universitaria, que logró una precisión de hasta 87%. Este hallazgo se alinea con investigaciones más recientes como la de Iam-On y Boongoen (2023), quienes reportaron precisiones de hasta 94% utilizando técnicas de aprendizaje profundo.

Sin embargo, mientras que estudios previos como el de Burgos et al. (2023) se centraron en la precisión técnica, nuestra investigación revela que el verdadero desafío no reside en la capacidad técnica de desarrollar el modelo, sino en mantenerlo operativo y relevante a lo largo del tiempo, como se evidencia en la sección 4.1.5.

- **Desafíos Organizacionales vs. Técnicos**

Nuestros hallazgos coinciden con las observaciones de Ekowo y Palmer (2016) sobre la importancia de los factores organizacionales en la implementación exitosa de sistemas predictivos. La investigación confirma la hipótesis H2 (ver sección 4.8.2), que sugería que la implementación exitosa requiere más que recursos técnicos y datos disponibles. Se descubrió que los principales obstáculos son de naturaleza organizacional y política, incluyendo:

- Falta de directrices institucionales claras
- Limitaciones en recursos humanos dedicados
- Ausencia de estrategias de mantenimiento a largo plazo

Estos hallazgos contrastan significativamente con investigaciones previas como la de Solis et al. (2018), que se centraban en los aspectos técnicos del desarrollo del modelo.

- **Implicaciones para la Implementación Futura**

Un hallazgo crucial que no se había anticipado es la existencia de un proyecto de actualización en curso que no se había anticipado ya que se reactivó durante el desarrollo de esta investigación por parte de la OPI (ver sección 4.3.1). Este descubrimiento tiene implicaciones significativas para la viabilidad del proyecto propuesto, ya que su éxito podría hacer innecesaria una nueva implementación. Esta situación no se había considerado en la literatura previa sobre implementaciones de modelos predictivos en educación superior.

- **Viabilidad Económica vs. Viabilidad Práctica**

Aunque el análisis financiero demuestra la viabilidad económica del proyecto con un VANE positivo en todos los escenarios (ver sección 4.6.1.4), incluso en condiciones pesimistas, la investigación revela que la viabilidad

económica no garantiza el éxito del proyecto. Este hallazgo valida la hipótesis H1 (ver sección 4.8.1), que sugería que la utilidad práctica del modelo dependería más de factores organizacionales que de su precisión técnica.

- **Oportunidades y Riesgos Emergentes**

La investigación ha identificado oportunidades significativas no contempladas en estudios anteriores, como la integración con fuentes de datos externas (ver sección 4.2.6) y la inclusión de variables postpandemia. Sin embargo, también se han descubierto nuevos riesgos, particularmente en el ámbito de la gobernanza de datos y la privacidad, que no habían sido completamente abordados en implementaciones anteriores.

- **Contribución al Conocimiento**

Esta investigación se suma al cuerpo de conocimiento existente sobre analítica educativa, complementando trabajos como el de Li et al. (2024) sobre aprendizaje federado en educación superior. Nuestros hallazgos sobre la importancia de los factores organizacionales amplían el marco conceptual propuesto por Tinto (1993), agregando la dimensión de sostenibilidad institucional a los modelos de retención estudiantil.

6.2. Conclusiones

A continuación, se muestran las conclusiones del proyecto por cada objetivo específico planteado anteriormente:

- Objetivo Específico 1: Examinar el estado de los modelos predictivos de deserción estudiantil en la educación superior de Costa Rica, para la determinación de su situación actual, retos y oportunidades para la implementación de futuros nuevos proyectos relacionados, durante las primeras cuatro semanas del segundo semestre del año 2024.
    - Se identificó que la OPI del TEC mantiene actualmente una iniciativa de actualización del modelo predictivo anterior con una alta prioridad institucional, como se evidencia en la sección 4.1.3.
    - Se descubrió que el modelo predictivo desarrollado por el TEC en 2018 alcanzó una precisión del 91% en la identificación de estudiantes en riesgo, demostrando la viabilidad técnica de estos sistemas, según se detalla en la sección 4.1.3.
    - Se determinó que existen retos significativos en la gestión de datos, principalmente en mantener información dinámica y actualizada, como se evidencia en la sección 4.1.5.
- Objetivo Específico 2: Realizar la definición del producto de modelo predictivo incluyendo información administrativa como alcance, criterios de éxito, presupuesto, plazos, estructura de gestión, plan de progreso, beneficios del proyecto, requisitos legales y datos necesarios almacenados por el Tecnológico de Costa Rica y fuentes externas, así como nuevas posibles variables que aporten valor, para el establecimiento de una base organizativa que facilite su estimación y ejecución.
    - Se estableció un presupuesto total de ₡29,367,676.185 para los dos primeros años de desarrollo, que incluye personal, software, licencias y capacitación, como se detalla en la sección 4.2.10.
    - Se determinó una estructura organizativa que involucra 24 fuentes de datos diferentes, de las cuales 7 requieren desarrollo de RPA para su extracción, según se documenta en la sección 4.2.6.
    - La implementación propuesta consta de 4 años dividido en 6 fases principales, con hitos y entregables específicos para cada etapa, como se evidencia en la sección 4.2.7.
- Objetivo Específico 3: Investigar la factibilidad del producto mediante evaluaciones que incluyen un análisis financiero, organizacional, de mercado, competencia y de riesgos; asegurando información relevante permitiendo su uso para la toma de decisiones sobre su viabilidad.

- Se identificó una demanda institucional activa respaldada por la meta 4.3.3 del Plan Estratégico 2022-2026 del TEC que busca reducir la deserción de primer ingreso del 11% al 9%, como se evidencia en la sección 4.3.1.
- Se confirmó el interés de múltiples departamentos del TEC como potenciales usuarios del sistema, incluyendo la VIESA, DOP, OPI y las 23 escuelas académicas, según se detalla en la sección 4.3.1.
- Se descubrió una amenaza significativa a la demanda del proyecto: la OPI ya tiene en desarrollo una actualización del modelo anterior con alta prioridad institucional, lo que podría eliminar la necesidad de un nuevo sistema si se completa exitosamente, como se evidencia en la sección 4.3.1.
- Se determinó la existencia de un competidor directo crítico: el proyecto de actualización de la OPI, que, aunque enfrenta limitaciones de recursos, representa una amenaza inminente para la viabilidad del proyecto propuesto si logra completarse antes de su implementación, como se documenta en la sección 4.4.1.
- Se identificó una ventaja competitiva en la propuesta de integración con fuentes de datos externas y el enfoque holístico del modelo, características ausentes en las soluciones existentes, según se evidencia en la sección 4.4.3.
- Se determinó que el éxito del competidor directo en completar su actualización podría hacer innecesaria la inversión en un nuevo modelo, especialmente considerando que ya cuenta con el respaldo institucional y la experiencia de implementaciones anteriores, como se detalla en la sección 4.4.2.
- Se evidenció que los costos sociales del proyecto se reducen en un 85.8% después del primer año (de ₡17,388,751.78 a ₡2,474,890.28 anuales), como se demuestra en la sección 4.6.1.1, lo que indica una alta eficiencia en la estructura de costos a largo plazo.
- Se mostró que el proyecto puede evitar un desperdicio social anual de ₡1,038,960,000 al retener 222 estudiantes que de otro modo desertarían, basado en un costo de formación por estudiante de ₡4,680,000, según los datos del Estado de la Educación 2021 y los cálculos presentados en la sección 4.6.1.2.
- Se demostró la viabilidad social económica del proyecto con un VANE, TIRE y R-B/C positivos, manteniéndose rentable incluso con una reducción del 50% en la efectividad esperada, como se detalla en la sección 4.6.1.5.
- Se determinó que por cada colón invertido en el proyecto se generan ₡37.27 de beneficio social, según la Relación Beneficio-Costo calculada, demostrando una eficiencia excepcional en el uso de recursos públicos incluso en escenarios pesimistas donde mantiene una ratio de 10.69, como se detalla en el análisis de sensibilidad de la sección 4.6.1.5.
- Se identificaron 13 riesgos principales del proyecto, de los cuales ocho fueron catalogados como de alto impacto y dos de estos con un nivel de riesgo extremo, con estrategias específicas de mitigación definidas para cada uno, según se documenta en la sección 4.5.2.
- Se determinó que los riesgos más críticos están relacionados con limitaciones presupuestarias (nivel 25) y cambios en políticas institucionales o legislación (nivel 20), como se detalla en la sección 4.5.2.

➢ Objetivo Específico 4: Elaborar un plan de implementación del producto modelo predictivo, especificando las fases, hitos críticos y estrategias de evaluación, que proporcione un plan claro y estructurado que establezca una guía en su desarrollo en las operaciones del Tecnológico de Costa Rica.

- Se necesitan 6 proyectos específicos con un costo total de ₡35,296,908.765 para la implementación completa del sistema, como se detalla en la sección 4.7.1.

- Se identificó que el 83% de los proyectos propuestos (5 de 6) requieren personal especializado en tecnologías de información, mientras que solo el Proyecto de Gobernanza de Datos (P006) necesita expertos adicionales en ética y procesos de negocio, según se detalla en el Apéndice P: Proyectos sugeridos detallados.
- Se determinó que los proyectos P005 (Monitoreo y Evaluación) y P006 (Gobernanza de Datos) requieren ejecutarse simultáneamente en el último año de implementación, ocupando el 42% del presupuesto total (₡14,964,616.29), como se evidencia en la sección 4.7.1, debido a su criticidad para la sostenibilidad del sistema.
- El proyecto P003 (Integración con Sistemas Existentes) representa el punto crítico de la implementación, ya que tiene dependencias directas con P001 y P002, y es prerrequisito para P004, según se documenta en el Apéndice P: Proyectos sugeridos detallados., convirtiéndolo en el proyecto con mayor impacto en la línea temporal.
- Se determinó que la implementación completa requiere un equipo mínimo de 11 profesionales especializados, distribuidos entre un 73% (8) en roles técnicos (desarrolladores, analistas y científicos de datos) y un 27% (3) en roles no técnicos (expertos en ética y gobernanza), según se detalla en el Apéndice P: Proyectos sugeridos detallados.

Capítulo 7. Cumplimiento de objetivos

La presente sección detalla el nivel de cumplimiento alcanzado para cada uno de los objetivos establecidos en la propuesta inicial del trabajo final de graduación. Se realiza un análisis exhaustivo que contempla tanto el objetivo general como los específicos, evaluando los productos generados, el porcentaje de logro y las observaciones pertinentes en cada caso.

El análisis demuestra un alto nivel de cumplimiento general de los objetivos planteados, con algunos ajustes menores que fueron necesarios durante la ejecución del proyecto para adaptarse a la realidad institucional y las limitaciones encontradas. A continuación, en la **Tabla *42***, se presenta el desglose detallado por objetivo:

Tabla 42
Cumplimiento de objetivos.

Objetivo General: Evaluar la viabilidad del desarrollo de un modelo predictivo que identifique a los estudiantes en riesgo de deserción de carrera del Tecnológico de Costa Rica, considerando aspectos financieros, legales, técnicos, organizacionales, de mercado y riesgos; para la determinación de su factibilidad con el fin de ser ejecutado por la institución, durante el segundo semestre del año 2024.			
Objetivo específico	**Productos**	**% de logro**	**Comentarios**
1. Examinar el estado de los modelos predictivos de deserción estudiantil en la educación superior de Costa Rica, para la determinación de su situación actual, retos y oportunidades para la implementación de futuros nuevos proyectos relacionados, durante las primeras cuatro semanas del segundo semestre del año 2024.	- *Paper* y análisis de situación actual - Mapeo de iniciativas existentes - Análisis comparativo de modelos	100%	Se logró una investigación completa del estado actual, identificando iniciativas previas, retos y oportunidades en el contexto costarricense.
2. Realizar la definición del producto de modelo predictivo incluyendo información administrativa como alcance, criterios de éxito, presupuesto, plazos, estructura de gestión, plan de progreso, beneficios del proyecto, requisitos legales y datos necesarios almacenados por el Tecnológico de Costa Rica y fuentes externas, así como nuevas posibles variables que aporten valor, para el establecimiento de una base organizativa que facilite su estimación y ejecución.	- Plan de proyecto - Definición detallada del producto - Documentación técnica y administrativa	100%	Se alcanzó una definición clara del producto.
3. Investigar la factibilidad del producto mediante evaluaciones que incluyen un análisis financiero, organizacional, de mercado, competencia y de riesgos; asegurando información relevante permitiendo	- Evaluación de mercado - Evaluación financiera - Evaluación legal	100%	Se realizaron todas las evaluaciones planificadas con resultados detallados y bien fundamentados. La limitación principal fue el acceso a algunos datos financieros históricos.

Objetivo General: Evaluar la viabilidad del desarrollo de un modelo predictivo que identifique a los estudiantes en riesgo de deserción de carrera del Tecnológico de Costa Rica, considerando aspectos financieros, legales, técnicos, organizacionales, de mercado y riesgos; para la determinación de su factibilidad con el fin de ser ejecutado por la institución, durante el segundo semestre del año 2024.			
Objetivo específico	**Productos**	**% de logro**	**Comentarios**
su uso para la toma de decisiones sobre su viabilidad.	- Evaluación de riesgos - Evaluación de competencia		
4. Elaborar un plan de implementación del producto modelo predictivo, especificando las fases, hitos críticos y estrategias de evaluación, que proporcione un plan claro y estructurado que establezca una guía en su desarrollo en las operaciones del Tecnológico de Costa Rica.	- Proyectos propuestos - Hoja de ruta - Plan detallado de implementación	100%	Se desarrolló un plan de implementación completo y detallado, incluyendo todas las fases necesarias y consideraciones para su ejecución.

Nota: Elaboración propia.

Capítulo 8. Recomendaciones para futuras investigaciones

Las siguientes recomendaciones surgen del análisis exhaustivo realizado durante la formulación y evaluación del modelo predictivo de deserción estudiantil en el TEC. Estas sugerencias están dirigidas a futuros investigadores que deseen profundizar en esta temática o implementar proyectos similares en instituciones de educación superior.

Aspectos Técnicos y de Datos:

1. Se recomienda realizar un estudio específico sobre la calidad y disponibilidad de datos en sistemas heredados del TEC. Esta investigación debería enfocarse en:
    - Evaluar la completitud y precisión de datos históricos
    - Identificar brechas en la recolección de información
    - Proponer mejoras en los procesos de captura y almacenamiento de datos
2. Es aconsejable investigar la viabilidad de implementar técnicas de *federated learning* para el manejo de datos sensibles, permitiendo el entrenamiento de modelos sin centralizar la información personal de los estudiantes.

Aspectos Organizacionales:

3. Se sugiere realizar un estudio detallado sobre la estructura organizacional óptima para la implementación y mantenimiento de sistemas predictivos, considerando:
    - Roles y responsabilidades necesarios
    - Procesos de comunicación interdepartamental
    - Mecanismos de toma de decisiones
    - Estrategias de gestión del cambio
4. Se recomienda investigar modelos de gobernanza de datos específicos para instituciones educativas públicas en Costa Rica, considerando el marco legal y regulatorio actual.

Aspectos no Explorados:

5. No se profundizó en el impacto psicológico y ético de etiquetar estudiantes como "en riesgo de deserción". Se recomienda realizar estudios sobre:
    - Efectos en la autoestima y motivación estudiantil
    - Sesgos potenciales en las intervenciones institucionales
    - Estrategias de comunicación apropiadas
6. Se sugiere investigar la posibilidad de desarrollar modelos predictivos específicos por escuela o área académica, considerando las particularidades de cada disciplina.

Aspectos de Implementación:

7. Se recomienda realizar estudios piloto en escuelas específicas antes de una implementación institucional completa, permitiendo:
    - Validar la efectividad del modelo en contextos específicos
    - Identificar desafíos de implementación a menor escala
    - Refinar procesos y protocolos

8. Es importante investigar mecanismos de evaluación continua y mejora del modelo, incluyendo:
    - Métricas de efectividad a largo plazo
    - Procesos de actualización y reentrenamiento
    - Estrategias de validación continua

Aspectos Sociales y Contextuales:

9. Se recomienda explorar el impacto de factores externos emergentes en la deserción estudiantil:
    - Efectos postpandemia en patrones de deserción
    - Influencia de nuevas modalidades de estudio híbridas
    - Impacto de cambios socioeconómicos en la permanencia estudiantil

10. Es aconsejable investigar la relación entre las intervenciones basadas en predicciones y los resultados académicos a largo plazo, considerando:
    - Efectividad de diferentes tipos de intervenciones
    - Costo-beneficio de programas de retención
    - Impacto en la calidad educativa

Aspectos de Escalabilidad:

11. Se sugiere investigar la posibilidad de expandir el modelo a otras instituciones educativas:
    - Adaptabilidad del modelo a diferentes contextos institucionales
    - Requisitos para la transferencia de conocimiento
    - Desafíos de implementación interinstitucional

Capítulo 9. Referencias

Acuña, M. (2016). *ATI Plan Estratégico 2016-2020*. Área Académica de Administración de Tecnologías de Información. https://tecdigital.tec.ac.cr/dotlrn/escuela/escuela.ATI/estudiantes.ATI/file-storage/view/ATI_Plan_Estrat%C3%A9gico_2016-2020.pdf

Administración de Tecnología de Información. (2020). *Áreas de investigación* . https://sites.google.com/view/tfg-ati-tec/tfg-en-ati/%C3%A1reas-de-investigacion?authuser=0

Aljohani, O. (2016). A comprehensive review of the major studies and theoretical models of student retention in higher education. Higher Education Studies, 6(2), 1-18. https://doi.org/10.5539/hes.v6n2p1

Asamblea Legislativa de la República de Costa Rica. (2002). Ley General de Control Interno N° 8292. Sistema Costarricense de Información Jurídica. http://www.pgrweb.go.cr/scij/Busqueda/Normativa/Normas/nrm_texto_completo.aspx?param1=NRTC&nValor1=1&nValor2=49185

Asamblea Legislativa de la República de Costa Rica. (2005). Ley de Certificados, Firmas Digitales y Documentos Electrónicos N° 8454. Sistema Costarricense de Información Jurídica. http://www.pgrweb.go.cr/scij/Busqueda/Normativa/Normas/nrm_texto_completo.aspx?param1=NRTC&nValor1=1&nValor2=55666

Asamblea Legislativa de la República de Costa Rica. (2011). Ley de Protección de la Persona frente al tratamiento de sus datos personales N° 8968. Sistema Costarricense de Información Jurídica. http://www.pgrweb.go.cr/scij/Busqueda/Normativa/Normas/nrm_texto_completo.aspx?param1=NRTC&nValor1=1&nValor2=70975

Barragán, S., & González, R. (2023). La educación superior en Costa Rica post-pandemia: Desafíos y oportunidades. Revista Latinoamericana de Educación Superior, 15(2), 45-62. https://doi.org/10.xxxx/rles.2023.15.2.45

Bean, J. P., & Metzner, B. S. (1985). A conceptual model of nontraditional undergraduate student attrition. Review of Educational Research, 55(4), 485-540. https://doi.org/10.3102/00346543055004485

BID. (2021). Marco de Evaluación de Capacidades Institucionales. https://publications.iadb.org/publications/spanish/document/Marco-de-evaluacion-de-capacidades-institucionales.pdf

Borges, L. C., Marques, V. M., & Bernardino, J. (2018). Comparison of data mining techniques and tools for data classification. In Proceedings of the 7th International Conference on Computer Science, Engineering and Information Technology (CSEIT-2018) (pp. 113-127).

Chinchilla-Brenes, S. (2016). Estudio descriptivo de la deserción estudiantil en el Tecnológico de Costa Rica. Revista Tecnología en Marcha, 29(3), 50-63. https://doi.org/10.18845/tm.v29i3.2888

Chirino Sánchez, A. (2020). La protección de datos personales: desde el derecho a la autodeterminación informativa hacia la tutela de los datos personales. Revista de Ciencias Jurídicas, 151, 13-36.

CONARE. (2021). Plan Nacional de la Educación Superior Universitaria Estatal 2021-2025. https://www.conare.ac.cr/planes-2021-2025

Creswell, J. W. (2014). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. Thousand Oaks, CA: SAGE Publications, https://doi.org/10.5539/elt.v12n5p40

Delen, D. (2010). A comparative analysis of machine learning techniques for student retention management. Decision Support Systems, 49(4), 498-506. https://doi.org/10.1016/j.dss.2010.06.003

Ekowo, M., & Palmer, I. (2016). The promise and peril of predictive analytics in higher education: A landscape analysis. New America. https://www.newamerica.org/education-policy/policy-papers/promise-and-peril-predictive-analytics-higher-education/

Fernández-Martín, T., Solis, M., Hernández-Jiménez, M., & Moreira-Mora, T. (2018). *A multinomial and predictive analysis of factors associated with university dropout*. Revista Electrónica Educare, 23(1). https://doi.org/10.15359/ree.23-1.5

Forbes Advisor. (2023). Best Data Visualization Tools Of 2023. https://www.forbes.com/advisor/business/software/best-data-visualization-tools/

Gartner. (2022). Magic Quadrant for Analytics and Business Intelligence Platforms. https://www.gartner.com/en/documents/3981253

González-Velázquez, L., Caballero-García, P. A., & Palacios-Picos, A. (2022). COVID-19 y deserción universitaria: Un estudio longitudinal en América Latina. Revista Iberoamericana de Educación Superior, 13(36), 86-103. https://doi.org/10.22201/iisue.20072872e.2022.36.1192

Guzmán, L. (s. f.). *Las fuentes secundarias.* Universidad de Costa Rica. https://www.ts.ucr.ac.cr/binarios/docente/pd-000169.pdf

Hernández-Jiménez, M., Moreira-Mora, T., Solis, M., & Fernández-Martín, T. (2019). *Estudio descriptivo de variables sociodemográficas y motivacionales asociadas a la deserción: la perspectiva de personas universitarias de primer ingreso*. Revista Educación, 44(1). https://doi.org/10.15517/revedu.v44i1.37247

Hernández-Sampieri, R. & Mendoza, C (2018). *Metodología de la investigación. Las rutas cuantitativa, cualitativa y mixta.* https://doi.org/10.22201/fesc.20072236e.2019.10.18.6

Iam-On, N., & Boongoen, T. (2023). Deep learning approach for student dropout prediction in higher education. Expert Systems with Applications, 215, 119225. https://doi.org/10.1016/j.eswa.2023.119225

Instituto Tecnológico de Costa Rica. (2022). Plan Estratégico 2022-2026. https://www.tec.ac.cr/plan-estrategico-2022-2026

Instituto Tecnológico de Costa Rica. (2023). TEC en Cifras 2023. https://www.tec.ac.cr/tec-cifras-2023

Jacobson, G. (2019). Predictive Analytics in Higher Education: A Case Study of the University of Arizona. Journal of Higher Education Management, 34(2), 121-135.

Kaur, R., & Bawa, S. (2019). A comparative analysis of data mining tools. International Journal of Advanced Research in Computer Science, 10(3), 1-7.

Li, W., Milošević, J., Bjelica, M., & Konings, B. (2024). Federated learning for student dropout prediction: Privacy-preserving analytics in higher education. IEEE Transactions on Learning Technologies, 17(1), 78-91. https://doi.org/10.1109/TLT.2023.3245678

Microsoft. (2024). Power BI Pricing. https://www.microsoft.com/es-es/power-platform/products/power-bi/pricing

MIDEPLAN. (2022). Guía metodológica para la identificación, formulación y evaluación de proyectos de inversión pública. https://documentos.mideplan.go.cr/share/s/HsM-07J5SeGhqsR5bClXIQ

Ministerio de Justicia y Paz. (2013). Reglamento a la Ley de Protección de la Persona frente al tratamiento de sus datos personales (Decreto Ejecutivo N° 37554-JP). Sistema Costarricense de Información Jurídica. http://www.pgrweb.go.cr/scij/Busqueda/Normativa/Normas/nrm_texto_completo.aspx?param1=NRTC&nValor1=1&nValor2=74352

Ministerio de Planificación Nacional y Política Económica. (2022). Guía metodológica para la identificación, formulación y evaluación de proyectos de inversión pública. https://documentos.mideplan.go.cr/share/s/Io0zBLPWRVG2qLYpD2iMOw

Moore, M. H. (2013). Recognizing Public Value. Harvard University Press.

Muñoz-Olano, J. F., & Hurtado-Parrado, C. (2017). Effects of goal clarification on impulsivity and academic procrastination of college students. Revista Latinoamericana de Psicología, 49(3), 173-181. https://doi.org/10.1016/j.rlp.2017.03.001

OECD. (2019). Recommendation of the Council on Artificial Intelligence. OECD Legal Instruments. https://legalinstruments.oecd.

ONU. (2015). *Objetivos y Metas de Desarrollo Sostenible - Desarrollo Sostenible*. United Nations. https://www.un.org/sustainabledevelopment/es/objetivos-de-desarrollo-sostenible/

Programa Estado de la Nación. (2019). Séptimo Informe Estado de la Educación. PEN CONARE. https://estadonacion.or.cr/informes/

Programa Estado de la Nación. (2021). Estado de la Educación 2021. https://estadonacion.or.cr/informe/?id=6f2a6b5c-4c53-49ac-92d0-2b1c98d6e8d9

Rovai, A. P. (2003). In search of higher persistence rates in distance education online programs. The Internet and Higher Education, 6(1), 1-16. https://doi.org/10.1016/S1096-7516(02)00158-6

Rudin, C., & Radin, J. (2023). Why are we using black box models in AI when we don't need to? A lesson from an explainable AI competition. Harvard Data Science Review, 5(2). https://doi.org/10.1162/99608f92.5a8a3a3d

Sapag Chain, N., & Sapag Chain, R. (2011). *Preparación y evaluación de proyectos*. Editorial McGraw-Hill.

Solis, M., Moreira, T., Gonzalez, R., Fernandez, T., & Hernandez, M. (2018). Perspectives to predict dropout in university students with machine learning. In 2018 IEEE International Work Conference on Bioinspired Intelligence (IWOBI) (pp. 1-6). IEEE. https://doi.org/10.1109/IWOBI.2018.8464191

Spady, W. G. (1970). Dropouts from higher education: An interdisciplinary review and synthesis. Interchange, 1(1), 64-85. https://doi.org/10.1007/BF02214313

TEC. (2022, September 28). *Qué es el TEC*. https://www.tec.ac.cr/que-es-tec

TEC. (2022, November 15). *Licenciatura en Administración de Tecnología de Información*. https://www.tec.ac.cr/programas-academicos/licenciatura-administracion-tecnologias-informacion

TEC. (2023). *Ejes de Conocimiento Estratégicos para el periodo comprendido de 2023 a 2032*. Ejes de Conocimiento Estratégicos 2023 a 2032. https://www.tec.ac.cr/ejes-conocimiento-estrategicos-2023-2032

Tinto, V. (1975). Dropout from higher education: A theoretical synthesis of recent research. Review of Educational Research, 45(1), 89-125. https://doi.org/10.3102/00346543045001089

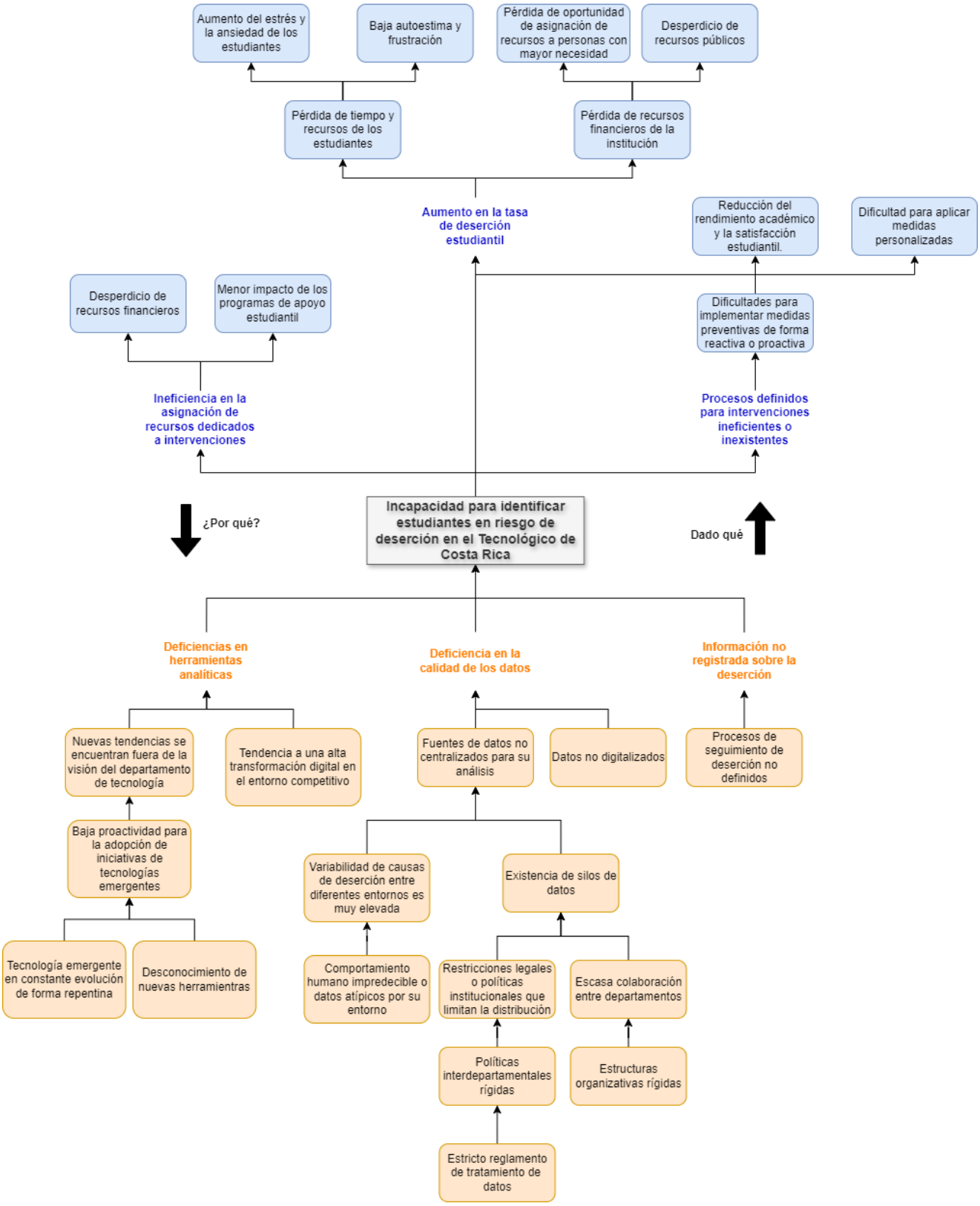
Tinto, V. (1993). Leaving college: Rethinking the causes and cures of student attrition (2nd ed.). University of Chicago Press.

UNESCO. (2018). Higher education in Latin America and the Caribbean: Challenges and opportunities. UNESCO Publishing. https://unesdoc.unesco.org/ark:/48223/pf0000262862

University of South Carolina. (2020). Early Alert System Impact Report. Office of Institutional Research, Assessment, and Analytics.

Capítulo 10. Apéndices

Apéndice A: Árbol del problema



Apéndice B: Plantilla de minuta

Reunión No.		**Fecha**	
Lugar		**Hora Inicio/Fin**	
Objetivo de la reunión			
Participantes:			
Temas tratados			
No.	**Asunto**	**Acuerdos y Comentarios**	
Próxima reunión			
Temas a tratar	**Fecha**	**Convocados**	
Aprobación			
Firma de creador de la minuta		**Firma de involucrado**	

Apéndice C: Plantilla de gestión del cambio

Datos generales del cambio			
N° de cambio			
Solicitantes		**Fecha de solicitud**	
Responsable de implementación		**Fecha de realización del cambio**	
Estado	Aprobado - Rechazado - En revisión		
Detalles del cambio			
Categoría			
Descripción detallada			
Justificación			
Impacto de realizar el cambio			
Impacto			
Comentarios			
Firmas de aprobación			
Revisado por:		**Revisado por:**	
Elaborado por:		**Aprobado por:**	

Apéndice D: Plantilla de Entrevista Semiestructurada sobre Deserción Estudiantil en el TEC

Introducción:

Buenos días/tardes. Gracias por acceder a participar en esta entrevista. El objetivo es recopilar información sobre los esfuerzos del TEC para abordar la deserción estudiantil y explorar las posibilidades de implementar un sistema predictivo basado en IA. Sus respuestas serán de gran valor para este proyecto de investigación.

Preguntas:

1) Contexto y Antecedentes:
   a) ¿Podría describir su rol y experiencia en la investigación sobre deserción estudiantil en el TEC?
   b) ¿Cómo se maneja a nivel interno algún índice de deserción en el TEC?
   c) ¿De dónde fueron a buscar los datos para el trabajo de investigación sobre deserción?
   d) ¿Qué fuentes de datos utilizaron para el proyecto?
   e) ¿Encontraron algún producto similar o sistema de análisis existente que pudiera aplicarse al TEC?
   f) ¿Se han encontrado productos similares o aplicaciones comerciales desde que realizaron la investigación hasta el día de hoy?
2) Desarrollo del Proyecto:
   a) ¿Cómo describiría la disposición del TEC para llevar a cabo estos proyectos de investigación sobre deserción?
   b) ¿Cómo fue la disposición del departamento de psicología para participar en este proyecto?
   c) ¿Qué complicaciones tuvieron durante el proyecto de investigación sobre deserción?
   d) ¿Considera que la calidad de los datos utilizados fue suficiente?
   e) ¿Identificaron variables adicionales que pudieran ser relevantes para el estudio de la deserción?
   f) ¿Los datos que tenían eran suficientes o tuvieron que recopilar información adicional de los estudiantes?
   g) ¿Hubo problemas legales o éticos al crear el cubo de información?
   h) ¿Cómo se maneja la información sensible de los estudiantes, como datos de becas?
   i) ¿Existe un consentimiento informado institucional para el manejo de datos de los estudiantes?
3) Implementación y Resultados:
   a) ¿Siente que el trabajo realizado pudo tener alguna mejoría?
   b) ¿Qué mejoras considera que podrían hacerse al modelo desarrollado?
   c) ¿Considera que este modelo realizado es vigente aún en el contexto postpandemia?
   d) ¿Tenían alguna entidad u objetivo específico en mente para el uso de los resultados del proyecto?
   e) ¿Cómo difundieron su trabajo internamente en el TEC?
   f) ¿A quiénes les interesaba que vieran su trabajo y por cuáles canales se lo hicieron llegar?
   g) ¿Hubo problemas financieros o de voluntad institucional para implementar el proyecto?
   h) ¿Se tuvo un costo asociado a sus investigaciones, ya sea monetario, de tiempo o personal implicado?
   i) ¿Qué pasó con estos proyectos? ¿Llegaron a utilizarse dentro del TEC?
4) Situación Actual y Perspectivas Futuras:
   a) ¿Se ha arreglado algo en cuanto a la conexión de bases de datos entre departamentos?
   b) ¿Qué esperaría poder utilizar de un sistema predictivo basado en IA para abordar la deserción?
   c) ¿Qué capacidades tiene el DOP para estos proyectos actualmente?
   d) ¿Ha habido alguna inversión después de sus trabajos iniciales en este tema?

e) ¿Se han tenido planes dentro del DOP para expandir este tema de deserción?
f) ¿Qué problemas tiene el TEC o el DOP para ejecutar estos proyectos a gran escala?
g) ¿Qué esperaría usted del producto que estamos proponiendo a nivel del TEC?

5) Contactos y Recursos Adicionales:
   a) ¿Puede proporcionar contactos en la OPI u otros departamentos relevantes para este proyecto?

Cierre:

Esas son todas las preguntas que tenía preparadas. ¿Hay algo más que le gustaría agregar sobre el tema de la deserción estudiantil o el uso de sistemas predictivos en el TEC?

Agradecimiento:

Muchas gracias por su tiempo y por compartir su experiencia y conocimientos. Su aporte será muy valioso para este proyecto de investigación.

Apéndice E: Plantilla de Entrevista Semiestructurada Abogado

Introducción:

Buenos días/tardes. Gracias por acceder a participar en esta entrevista. El objetivo es recopilar información sobre los esfuerzos del TEC para abordar la deserción estudiantil y explorar las posibilidades de implementar un sistema predictivo basado en IA. Sus respuestas serán de gran valor para este proyecto de investigación.

Preguntas:

1) Diseño y Consideraciones del Modelo Predictivo:
   a) ¿Qué factores legales y administrativos considera esenciales para incluir en un modelo predictivo de deserción estudiantil en el TEC?
   b) ¿Cómo podrían las políticas actuales del TEC influir en la efectividad de un modelo predictivo?
   c) ¿Existen limitaciones legales en cuanto al uso de datos personales de los estudiantes que podrían afectar el desarrollo del modelo?
   d) ¿Qué tipo de datos históricos legales o disciplinarios deberían considerarse para predecir la deserción?
   e) ¿Qué mecanismos de protección de datos deberían implementarse para asegurar que el modelo cumpla con las normativas legales vigentes?
   f) ¿Podría el modelo predictivo entrar en conflicto con los derechos de los estudiantes, y cómo podría mitigarse esto?
2) Implementación y Aplicación del Modelo en el Contexto del TEC:
   a) ¿De qué manera cree que un modelo predictivo podría integrarse en las decisiones administrativas y académicas del TEC para reducir la deserción?
   b) ¿Qué rol podría desempeñar la oficina de defensa estudiantil en la implementación de este modelo?
   c) ¿Qué tipo de comunicación o intervención legal debería realizarse cuando el modelo identifique a un estudiante en riesgo de deserción?
   d) ¿Qué retos legales podría enfrentar el TEC al utilizar un modelo predictivo para tomar decisiones que afectan directamente a los estudiantes?

   e) ¿Cómo podría asegurarse de que el uso del modelo no discrimine o perjudique injustamente a ciertos grupos de estudiantes?
3) Eficacia y Mejora Continua del Modelo:
   a) ¿Qué indicadores considera más relevantes para evaluar la eficacia de un modelo predictivo en la reducción de la deserción estudiantil?
   b) ¿Qué procedimientos legales deberían establecerse para revisar y mejorar continuamente el modelo predictivo?
   c) ¿Cómo se podría involucrar a la comunidad estudiantil en la validación y mejora del modelo?
   d) ¿Cuáles serían las posibles implicaciones legales si el modelo falla en predecir correctamente un caso de deserción?
   e) ¿Qué medidas correctivas o ajustes legales considera necesarios si se detectan fallos en la predicción o implementación del modelo?

Cierre:

Esas son todas las preguntas que tenía preparadas. ¿Hay algo más que le gustaría agregar sobre el tema de la deserción estudiantil o el uso de sistemas predictivos en el TEC?

Agradecimiento:

Muchas gracias por su tiempo y por compartir su experiencia y conocimientos. Su aporte será muy valioso para este proyecto de investigación.

Apéndice F: Tabla de hallazgos de entrevista a M.Ed. Teresa Hernández Jiménez.

Sección	Tema	Id	Hallazgo
Contexto y Antecedentes	Rol de la entrevistada	EH1	Trabajó en la dirección del DOP del 2018 al 2022
		EH2	Fue Vicerrectora de Vida Estudiantil 2022-2023
		EH3	Terminó su cargo en junio del año pasado
		EH4	Participó en varias investigaciones relacionadas a la deserción estudiantil
	Índice de deserción	EH5	Existe una meta en el plan estratégico del TEC 2022-2026 (meta 4.3.3) relacionada a la reducción de la deserción
		EH6	El objetivo del plan es reducir la deserción de primer ingreso al menos en un 10% para 2026
		EH7	Según el plan, en 2022 la deserción era 11%, se busca llegar a 9% en 2026
	Fuentes de datos	EH8	Para sus investigaciones sobre la deserción, se utilizó principalmente de Admisión y Registro
		EH9	Se creó un cubo de información en la OPI
		EH10	Algunas escuelas (ej. Seguridad Laboral) llevan su propio seguimiento
		EH11	TEC Digital maneja bases de datos de estudiantes
	Problemas con bases de datos	EH12	En el DOP, hace 8-10 años se intentó actualizar las bases de datos
		EH13	El nuevo sistema no se podía conectar con los sistemas institucionales existentes
		EH14	Persisten problemas de comunicación entre bases de datos
	Productos similares	EH15	No se encontraron productos exactamente similares
		EH16	El trabajo fue considerado novedoso
		EH17	Se presentó en varios congresos
Desarrollo del Proyecto	Disposición del TEC	EH18	Hubo apoyo para la investigación
		EH19	Limitaciones de recursos para la implementación
	Complicaciones	EH20	Problemas de comunicación entre bases de datos
		EH21	Falta de personal para ejecutar y dar seguimiento
		EH22	Sensibilidad en el manejo de la información predictiva

Sección	Tema	Id	Hallazgo
		EH23	Preocupaciones éticas sobre el uso de la información
	Calidad de los datos	EH24	Los datos institucionales fueron suficientes para el modelo
		EH25	Se identificaron variables adicionales importantes (ej. si el estudiante trabaja)
		EH26	Se realizaron encuestas adicionales para recopilar información faltante
	Aspectos legales y éticos	EH27	Inicialmente no hubo problemas legales
		EH28	Han aumentado las recomendaciones sobre protección de datos
		EH29	No existe un consentimiento informado institucional general
		EH30	Algunos programas tienen su propio consentimiento informado
	Metodología	EH31	Se trabajó con Martín Solís de la Escuela de Administración de Empresas
		EH32	Se revisaron diversos algoritmos de predicción
		EH33	Se eligió el que generaba mejores índices de predicción
Implementación y Resultados	Eficacia del modelo	EH34	Se consideró efectivo, con buenos índices de precisión
		EH35	Los datos obtenidos eran mejores de lo esperado
	Difusión del trabajo	EH36	Se presentó en varios congresos externos
		EH37	Internamente: Consejo de Rectoría, Consejo de Docencia, posiblemente Consejo Institucional
	Problemas de implementación	EH38	Falta de personal para ejecutar el proyecto a gran escala
		EH39	No hay recursos para abrir nuevas plazas
		EH40	Falta de directrices claras para el uso de la información
	Uso actual	EH41	El proyecto no se implementó completamente
		EH42	El cubo de información y el algoritmo existen, pero, no se usan sistemáticamente
Situación Actual y Perspectivas	Conexión de bases de datos	EH43	Persisten problemas de comunicación entre bases de datos institucionales
	Trabajo actual en deserción	EH44	El DOP continúa trabajando en prevención a través de diversos programas

Sección	Tema	Id	Hallazgo
		EH45	Programa IntegraTEC de mentoría para estudiantes de primer ingreso
		EH46	Atención psicológica y asesoría psicoeducativa
		EH47	Programa de necesidades educativas
	Limitaciones actuales	EH48	Falta de personal dedicado
		EH49	Limitaciones presupuestarias
		EH50	Falta de directrices institucionales claras
	Sugerencias para futuro producto	EH51	Integración como consulta adicional en el expediente estudiantil
		EH52	Accesible para personal autorizado (docentes, direcciones, personal de salud, etc.)
		EH53	Que muestre un índice de riesgo de deserción semestre a semestre
Contactos y Recursos Adicionales	Factores de deserción identificados	EH54	Rendimiento en el primer año
		EH55	Motivación
		EH56	Cursos introductorios
		EH57	Distancia que recorre el estudiante
		EH58	Si el estudiante trabaja
	Programas de apoyo	EH59	Posibilidad de reducir créditos para mantener beca
		EH60	Consideraciones especiales para estudiantes trabajadores o padres/madres
	Contacto OPI	EH61	Evelyn Hernández, directora actual
	Contacto DATIC	EH62	Andrea Cavero
	Contacto Escuela de Administración	EH63	Martín Solís (trabajó en la parte estadística y de predicción)

Apéndice G: Tabla de hallazgos de entrevista al Profesor Martín Solís Salazar.

Sección	Tema	Id	Hallazgo
Contexto y Antecedentes	Roles y experiencia	EH64	Martin tuvo un rol principal en el análisis de datos y generación de modelos, especialmente en el segundo artículo mencionado.
		EH65	Martin es estadístico con especialidad en ingeniería en computadores con énfasis en ciencia de datos.
		EH66	Roberto González, un estudiante de ATI, realizó su tesis de licenciatura sobre este proyecto.
		EH67	Martin tuvo un rol principal en el segundo artículo y un rol más secundario en la encuesta.
	Publicaciones y resultados	EH68	Se generaron tres artículos a partir de la investigación sobre deserción.
	Fuentes de datos	EH69	Los datos fueron proporcionados por el departamento de planificación del TEC, específicamente por Jason y Rebeca bajo la dirección de Tatiana Fernández.
	Definiciones y metodología	EH70	Se definió como desertor a aquella persona que después de un año de no haber matricular, no volvió nunca más a hacerlo hasta el presente.
		EH71	Para el proyecto más reciente, se solicitaron datos de matrículas desde 2010 hasta la actualidad.
	Tipos de investigación	EH72	Se realizó un modelo predictivo de deserción y una encuesta descriptiva sobre razones de deserción.
Desarrollo del Proyecto	Implementación y visualización	EH73	Roberto González se encargó de la implementación y puesta en producción del modelo.
		EH74	Se desarrolló una herramienta de visualización para mostrar los resultados del modelo por escuela y área del TEC.
	Precisión y efectividad	EH75	El modelo alcanzó una precisión de entre 87% y 91% en la predicción de deserción.
	Disposición institucional	EH76	Hubo apertura inicial para implementar el modelo, incluyendo presentaciones al consejo institucional y al rector.
	Desafíos y complicaciones	EH77	No se estableció una estrategia clara o directriz para utilizar el modelo después de su implementación.

Sección	Tema	Id	Hallazgo
	Calidad de datos	EH78	Los datos no eran de la mejor calidad en aquel tiempo, con oportunidades de mejora en variables como las becas.
	Variables y factores	EH79	Se identificó la necesidad de variables dinámicas, como el estado de la beca del estudiante a lo largo del tiempo.
		EH80	Las notas bajas se identificaron como un fuerte predictor de deserción, reflejando diversos factores subyacentes.
	Procesamiento de datos	EH81	Se requirió un proceso iterativo de limpieza y corrección de datos antes de poder utilizarlos en el modelo.
	Colaboración interdepartamental	EH82	El proyecto involucró colaboración entre diferentes departamentos, incluyendo planificación y el área de estadística e ingeniería.
Implementación y Resultados	Funcionamiento del modelo	EH83	El modelo se ejecutaba al final de cada semestre, utilizando información de notas y cursos matriculados.
		EH84	El modelo se diseñó para ser ejecutado automáticamente al final de cada semestre.
	Acceso y uso	EH85	Los directores de escuela tenían acceso a la visualización de los resultados del modelo.
	Desafíos de implementación	EH86	A pesar de la implementación, no hubo evidencia de uso efectivo del modelo para prevenir la deserción.
		EH87	Hubo falta de directrices claras y estrategias para utilizar el modelo después de su desarrollo.
		EH88	No se estableció un proceso de seguimiento para los estudiantes identificados como en riesgo de deserción.
	Estado actual	EH89	El modelo se considera desactualizado y probablemente ya no esté funcionando en la plataforma original.
	Difusión y presentación	EH90	El modelo se presentó al consejo institucional, al rector y a otros tomadores de decisiones.
Situación Actual y Perspectivas Futuras	Intentos de actualización	EH91	Hubo intentos recientes (hace un año) de actualizar el modelo, pero, se encontraron obstáculos en la obtención de datos.
	Cambios organizacionales	EH92	Jason y Rebeca, quienes originalmente trabajaban para la oficina de planificación, ahora trabajan para el DATIC.
	Acceso y uso propuesto	EH93	Se sugiere que el modelo debería estar disponible para el DOP, la OPI, y todos los directores de escuela.

Sección	Tema	Id	Hallazgo
	Desafíos actuales	EH94	Existen problemas de comunicación entre departamentos y burocracia que dificultan la obtención de datos necesarios.
		EH95	La generación de algunas variables, como los créditos restantes, presenta desafíos debido a cambios en los programas de estudio.
		EH96	La obtención de datos actualizados para la renovación del modelo ha sido un obstáculo significativo.
	Mejoras propuestas	EH97	Se propone incluir variables como la cantidad de créditos restantes para graduarse.
		EH98	Se sugiere considerar la interacción del estudiante con el TEC Digital como posible predictor de deserción.
		EH99	Se reconoce que los patrones y variables utilizados para predecir la deserción pueden haber cambiado significativamente desde el desarrollo original del modelo.
		EH100	Se menciona la necesidad de considerar el impacto de la virtualidad y los aspectos psicológicos postpandemia en futuros modelos.
		EH101	Se propuso afinar el modelo para predecir la probabilidad de deserción en diferentes plazos (siguiente semestre, dentro de dos semestres, etc.).
		EH102	Se enfatiza la necesidad de directrices institucionales claras para el uso efectivo del modelo en la prevención de la deserción.
	Expectativas	EH103	Se espera que futuros proyectos puedan superar los obstáculos burocráticos y de interés político para implementar efectivamente el modelo.
Contactos y Recursos Adicionales	Contactos clave	EH104	Se menciona a Evelyn, directora de la OPI, como contacto relevante para obtener información actualizada sobre el proyecto.

Apéndice H: Tabla de hallazgos de entrevista a Evaluadora e investigadora jubilada Tania Elena Moreira Mora.

Sección	Tema	Id	Hallazgo
Contexto y Antecedentes	Equipo de investigación	EH105	El equipo estaba compuesto por Tania (coordinadora), Tatiana (directora de OPI), Martín (experto en modelos predictivos), y María Teresa (del DOP)
		EH106	Se formó un equipo interdisciplinario para abordar el problema de deserción desde diferentes perspectivas
	Fuentes de datos	EH107	Se utilizaron principalmente los indicadores del SIGI (Sistema de Indicadores de Gestión Institucional)
		EH108	Se complementó la información del SIGI con una encuesta a estudiantes sobre razones de deserción
	Antecedentes de investigación	EH109	Existían estudios previos sobre deserción en el TEC desde los años 70 y 80
		EH110	Hubo una brecha en la investigación sobre deserción hasta aproximadamente 2010
	Contexto institucional	EH111	La reducción del abandono estudiantil era una prioridad institucional para el TEC
	Experiencia previa	EH112	Tania había realizado estudios previos sobre deserción desde 2003-2004
	Motivación del proyecto	EH113	El proyecto buscaba atender una prioridad institucional de reducir el abandono estudiantil
Desarrollo del Proyecto	Aprobación y respaldo	EH114	El proyecto fue aprobado y respaldado por la Vicerrectoría de Investigación y Extensión (VIE)
		EH115	Se negoció la colaboración con la OPI para asegurar el acceso a los datos del SIGI
	Metodología	EH116	Se desarrollaron dos modelos en etapas diferentes del proyecto
		EH117	Se probaron varios algoritmos: Regresión Logística, Redes Neuronales, Bosque Aleatorio (*Random Forest*) y Support Vector Machine
		EH118	El algoritmo que mejor funcionó fue el *Bosque Aleatorio (Random Forest)*
	Recursos	EH119	El equipo trabajó con recursos limitados: 20 horas semanales entre 4 investigadores
		EH120	No se contó con asistentes de investigación

Sección	Tema	Id	Hallazgo
	Recolección de datos	EH121	Se realizó una encuesta digitalizada a estudiantes para complementar los datos del SIGI
		EH122	La participación en la encuesta fue voluntaria y se siguieron protocolos éticos
	Variables de estudio	EH123	Se identificó la motivación como la variable más importante en la deserción
		EH124	Se reconoció la necesidad de incluir variables como el nivel educativo de los padres y factores socioeconómicos
	Duración del proyecto	EH125	El proyecto duró aproximadamente dos años
	Acceso a datos	EH126	Los datos del SIGI son accesibles para investigadores, directores de escuela y profesores
Implementación y Resultados	Difusión de resultados	EH127	Se presentaron los resultados en el Consejo de Docencia, Consejo del DOP, y Consejo Institucional
		EH128	Se realizó un almuerzo académico de la VIE para presentar los resultados
		EH129	Se publicaron artículos en revistas académicas (Educare y Educación) sobre los resultados
	Hallazgos principales	EH130	Se encontró que un 46.6% de los estudiantes que abandonaron reportaron que no les gustó la carrera o hubo incumplimiento de expectativas
	Continuidad del proyecto	EH131	Un estudiante (Roberto González) utilizó el modelo para su trabajo final de graduación y lo mejoró
		EH132	No se dio seguimiento continuo al modelo después de la investigación inicial
	Impacto del proyecto	EH133	El proyecto fue pionero en el uso de modelos predictivos para deserción en universidades públicas de Costa Rica
Situación Actual y Perspectivas Futuras	Mejoras sugeridas	EH134	Se reconoce la necesidad de actualizar el modelo con nuevas variables postpandemia
		EH135	Se sugiere incluir variables como el ajuste a la vida universitaria en el contexto actual
		EH136	Se propone la posibilidad de aplicar el modelo a nivel de escuelas individuales

Sección	Tema	Id	Hallazgo
		EH137	Se sugiere el uso del modelo para identificar potenciales desertores en cursos específicos (física, química, matemática)
	Desafíos para la implementación	EH138	Se necesita compromiso institucional y de las escuelas para implementar acciones basadas en los resultados del modelo
		EH139	Se requiere personal dedicado en la OPI para dar mantenimiento y seguimiento al modelo
	Importancia de la acción	EH140	Se enfatiza que identificar potenciales desertores no es suficiente sin acciones concretas para retenerlos
	Nuevos estudios	EH141	Se menciona un estudio reciente sobre el proceso de ajuste a la vida universitaria como factor relevante en la permanencia
Contactos y Recursos Adicionales	Contactos clave	EH142	Se recomienda contactar a Evelyn, actual directora de la OPI, para negociar la implementación del proyecto
		EH143	Se sugiere contactar a Tatiana, quien, aunque jubilada, podría proporcionar información valiosa
	Recursos disponibles	EH144	Existen artículos publicados sobre la investigación en las revistas Educare y Educación
		EH145	Hay información adicional disponible en el repositorio del DOP
	Otros estudios relevantes	EH146	Se menciona un estudio en la Universidad Nacional sobre deserción en un curso de matemáticas usando modelos predictivos

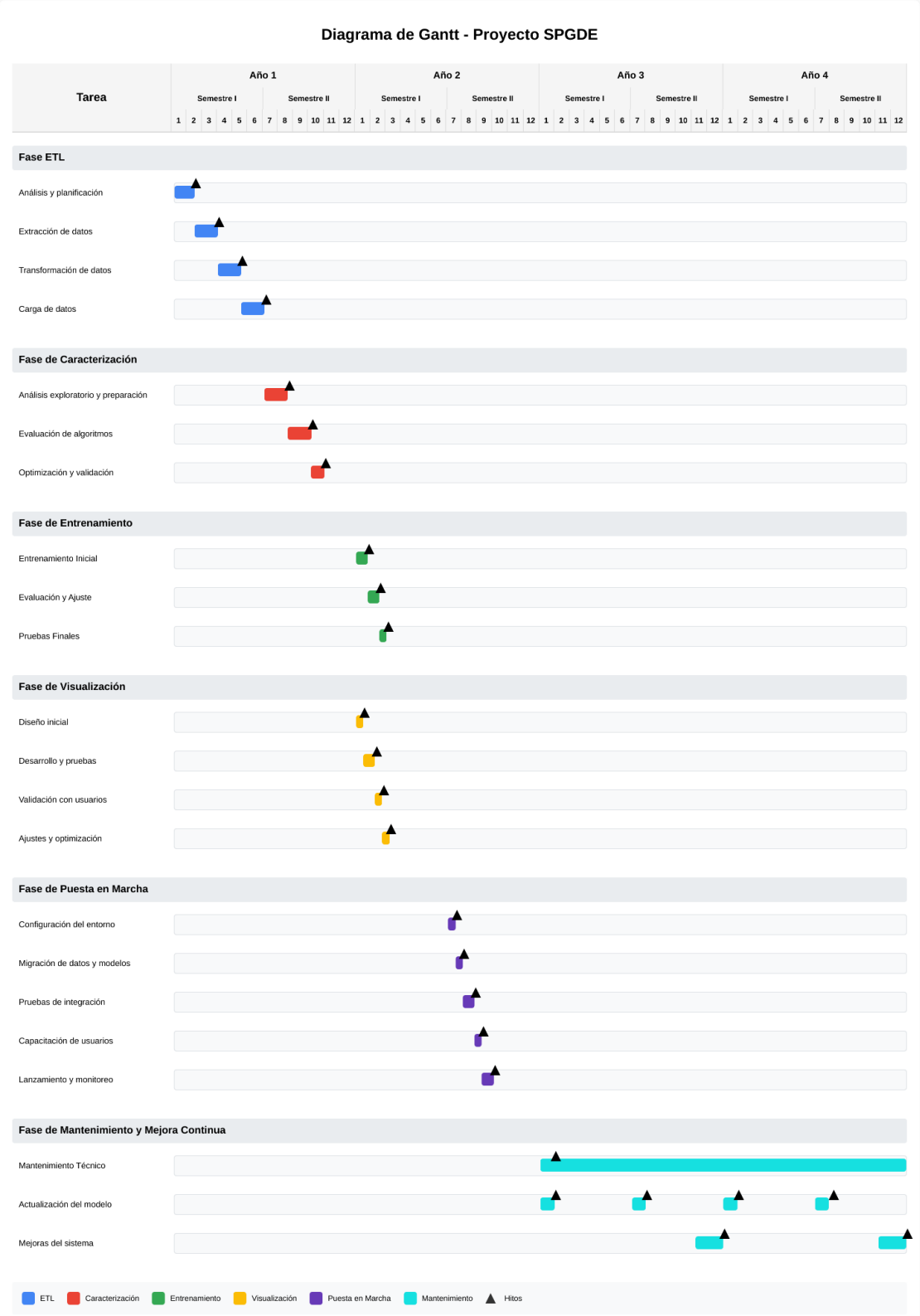
Apéndice I: Tabla de hallazgos de entrevista a Arlyn Angulo Hernández y Raquel Mejías Elizondo de la OPI.

Sección	Tema	Id	Hallazgo
Contexto y Antecedentes	Retoma del proyecto anterior	EH147	Se ha retomado el proyecto de modelo predictivo de deserción que anteriormente se había dejado sin finalizar.
	Estado actual del proyecto	EH148	El proyecto está actualmente en fase de desarrollo, con un enfoque en la actualización y mejora del modelo anterior.
	Prioridad del proyecto	EH149	La actualización del modelo de deserción tiene una alta prioridad dentro de los proyectos de la OPI.
	Estado del cubo anterior	EH150	El cubo de información sobre deserción anterior ya no está en ejecución ni disponible actualmente.
	Contacto con Martín	EH151	Se ha mantenido contacto reciente con Martín para actualizar sobre el progreso del proyecto.
Desarrollo del Proyecto	Avances en el nuevo modelo	EH152	Se ha generado un nuevo conjunto de datos para entrenar el modelo actualizado de deserción.
	Revisión de datos	EH153	El conjunto de datos para el nuevo modelo ha sido revisado por Martín, quien ha proporcionado retroalimentación para mejoras.
	Desafíos actuales	EH154	Se están trabajando en hacer que los datos sean más dinámicos, especialmente en relación a los cursos faltantes por estudiante.
	Colaboración interdepartamental	EH155	El proyecto involucra colaboración entre la OPI, DATIC y expertos externos como Martín.
	Recursos humanos	EH156	Solo hay dos analistas disponibles (Jason y Rebeca) para desarrollar múltiples cubos de información, lo que limita la velocidad de avance.
	Estudio curricular	EH157	Se realizó un extenso trabajo de estudio curricular para determinar cuántos cursos le faltan a cada estudiante.
	Sistematización de datos	EH158	La información sobre cursos faltantes no estaba previamente sistematizada en ningún sistema.
	Ubicación de los analistas	EH159	Jason y Rebeca, aunque trabajan con datos de la OPI, pertenecen al departamento de DATIC.
	Priorización de tareas	EH160	Mientras se revisan los datos del cubo de deserción, los analistas trabajan en otros cubos prioritarios.

Sección	Tema	Id	Hallazgo
Implementación y Resultados	Planificación de actualizaciones	EH161	Se está planeando que el nuevo cubo de información se actualice automáticamente, posiblemente de forma semestral o anual.
	Integración con SIGI	EH162	El nuevo modelo se está desarrollando para integrarse completamente con el Sistema de Indicadores de Gestión Institucional (SIGI).
	Sensibilidad del modelo anterior	EH163	El modelo anterior era bastante sensible y funcionaba bien cuando estaba en uso, pero, perdió mantenimiento y actualización.
	Momento de actualización	EH164	Se considera actualizar el modelo después de la entrega de notas y antes de la matrícula del siguiente período.
Situación Actual y Perspectivas	Estado de desarrollo	EH165	El proyecto está actualmente en desarrollo activo, con analistas trabajando en la preparación y refinamiento de los datos.
	Próximos pasos	EH166	Una vez completado el cubo de becas, el equipo se enfocará completamente en finalizar el modelo de deserción.
	Rol del SIGI	EH167	El SIGI es la fuente central de datos para el modelo de deserción, conteniendo información de oferta, matrícula, rendimiento académico, entre otros.
	Objetivo del SIGI	EH168	El SIGI busca centralizar la información de todos los sistemas del TEC para facilitar la toma de decisiones.
	Administración del SIGI	EH169	El SIGI es administrado por la OPI, pero, el mantenimiento lo dan los analistas Jason y Rebeca desde DATIC.
	Variables disponibles	EH170	El SIGI contiene múltiples variables como carnet, sexo, año de ingreso, modalidad, período, grado académico, carrera, entre otras.
	Cubos de información	EH171	Existen varios cubos de información en SIGI, incluyendo deserción regular, deserción de primer ingreso, matrícula, notas de corte, entre otros.
Aspectos Legales y de Seguridad	Confidencialidad de datos	EH172	Los usuarios nuevos de SIGI deben firmar un compromiso de confidencialidad para acceder a la información.
	Acceso a datos	EH173	El acceso a los datos completos es limitado por su sensibilidad, pero, se pueden compartir estructuras y variables sin datos personales.

Sección	Tema	Id	Hallazgo
	Consentimiento informado	EH174	Existen consentimientos informados desde admisión y registro para ciertos datos, pero, se necesita explorar más para la implementación del nuevo modelo.
	Futuras mejoras	EH175	Se está desarrollando un manual interactivo para facilitar la comprensión y uso de los cubos de información en SIGI.

Apéndice J: Diagrama de Gantt de proyecto SPGDE



Apéndice K: Tabla de requerimientos legales del proyecto

Categoría	Ley, normal, decreto o reglamento	Número de ley	Descripción
Protección de datos	Ley de protección de la persona frente al tratamiento de sus datos personales	Ley N° 8968	Esta ley garantiza a cualquier persona el respeto a sus derechos fundamentales, concretamente, su derecho a la autodeterminación informativa en relación con su vida o actividad privada y demás derechos de la personalidad. El SPGDE debe asegurar el cumplimiento de esta ley en el manejo de datos personales de los estudiantes.
Gestión de TI	Normas técnicas para la gestión y el control de las Tecnologías de Información (N-2-2007-CO-DFOE)	N-2-2007-CO-DFOE	Establece criterios de control que deben ser observados como parte de la gestión institucional de las TI. El SPGDE debe desarrollarse y gestionarse en cumplimiento de estas normas.
Control interno	Ley General de Control Interno	Ley N° 8292	Esta ley establece los criterios mínimos que deberán observar la Contraloría General de la República y los entes u órganos sujetos a su fiscalización, en el establecimiento, funcionamiento, mantenimiento, perfeccionamiento y evaluación de sus sistemas de control interno. El SPGDE debe incorporar mecanismos de control interno en línea con esta ley.
Certificados, firmas digitales y documentos electrónicos	Ley de Certificados, Firmas Digitales y Documentos Electrónicos	Ley N° 8454	Esta ley regula el uso de documentos electrónicos y firmas digitales. Si el SPGDE utiliza documentos electrónicos o requiere firmas digitales, debe cumplir con esta normativa.
Derechos de autor	Ley de Derechos de Autor y Derechos Conexos	Ley N° 6683	Esta ley protege los derechos de los creadores de obras intelectuales. El SPGDE debe respetar los derechos de autor en todo el software y contenido que utilice.
Laborales	Código de Trabajo	Ley N° 2	Regula los derechos y obligaciones de patronos y trabajadores. El proyecto debe cumplir con todas las disposiciones laborales aplicables.
Educación superior	Ley Orgánica del Instituto Tecnológico de Costa Rica	Ley N° 4777	Esta ley establece la organización y funciones del TEC. El SPGDE debe alinearse con los objetivos y estructura organizativa establecidos en esta ley.
Privacidad y seguridad de la información	Reglamento a la Ley de Protección de la Persona frente al Tratamiento de sus Datos Personales	Decreto Ejecutivo N° 37554-JP	Este reglamento proporciona detalles específicos sobre la implementación de la Ley N° 8968. El SPGDE debe cumplir con las disposiciones detalladas en este reglamento.

Apéndice L: Prescripción de entrevista con M.Ed. Teresa Hernández Jiménez, Asesora Psicoeducativa del DOP.

Milton: Bueno, para darte un poco de contexto, digamos, del proyecto como tal, el objetivo que tenemos ahorita es a futuro, se quiere crear con ayuda de la inteligencia artificial, que ahorita está en mucho auge, un modelo, igual mucho trabajo hecho, de los trabajos que me mandaste, ya dos de esos, de hecho, los dos documentos que me mandaste ya los había leído, justamente porque leía un tema muy similar.

Teresa: ¿Sos de ATI?

Milton: Sí, me pareció muy buen trabajo de hecho y un poco muy afín a lo que ya nos llevan a carrera también. Entonces el objetivo es que, bueno, yo en este momento estoy bajo la modalidad de, a ver si no es un seminario de graduación, que es que, en conjunto con un profesor, que en este caso es Juan Andrés Segreda, digamos, profesor de ATI, se tienen los objetivos igual de desarrollar un modelo, en este caso con IA, un poco más, digamos, predictivo, en el sentido de la deserción. Ahora, como es su objetivo, digamos, un poco demasiado grande, ¿verdad? Porque ya, de hecho, había conversado con otro profesor, que se llama Carlos Mata, que es un poco el que nos da la parte de formulación y evaluación él mismo me decía que este trabajo es muy grande muy extenso y de hecho el objetivo de este proyecto que yo voy a hacer es verificar verdad la capacidad del TEC y también la disponibilidad de poder realizarlo y en caso de que no o si digamos ese es mi trabajo de simplemente dar los datos y que alguien decida si o si no vamos a llegar a un verdadero entonces el objetivo como ya lo mencioné este producto yo quiero bueno prácticamente reunirme con muchos departamentos del TEC que es una pregunta que ahorita te voy a hacer después y ver qué tanto es posible hacer ese proyecto si de verdad es favorable entonces a nivel de los trabajos de ustedes vi que es muy afín de hecho es como un poco casi que lo mismo pero, sin IA entonces primero que nada ya como mencionó hay temas éticos y todo eso que de hecho voy a bueno ya que me lo menciona va a ser muy útil y lo quiero conversar también incluso con abogados que tiene el TEC.

Milton: Entonces, primero por el tema de los datos, mi primera pregunta es cómo, primero como, a partir de ese trabajo de dónde fueron a buscar esos datos, primero que nada, porque veo que en el TEC hay muchos datos, por ejemplo, psicología tiene uno, la parte financiera, el departamento financiero, otras se han hecho un registro y nada está conectado entre sí.

Teresa: Se supone que sí, digamos que debería, empezando como por las bases de datos institucionales, debería de haber un solo manejo de bases de datos, por decirlo de alguna manera, a partir de admisión y registro. Digamos, admisión y registro hace como toda la recolección inicial de datos, ¿verdad? Y de ahí se supone que salen las bases de datos institucionales. ¿Qué pasó? Hace como unos 8 años, tal vez, o sea, 10, 8 años, no estoy segura, ¿verdad? Ahí habría que averiguar bien el asunto. Bueno, comentarte tal vez un poquito que la visión que yo te voy a dar, ¿verdad? Obviamente, pues es acá desde el DOP. Tuve la oportunidad de ejercer la dirección del DOP del 2018 al 2022 y la Vicerrectoría de Vida Estudiantil 22-23, yo terminé digamos hace un año, en junio del año pasado, entonces un poquito verdad, o sea tal vez hay cosas como que manejo desde acá, otras como que puede ver verdad, como de la dirección con otros departamentos y otras, ¿verdad?, tal, como para ir viendo y ver dónde puedes, ¿verdad?, también como ir preguntando. Como te digo, hace unos 10, tal vez 8 años por ahí, se empezó, o sea, se dijo, ok, las bases de admisión y registro están muy antiguas, ¿verdad?, el formato que tienen, el lenguaje en el que están construidas, digámoslo de alguna manera, no están a fin con algunos cambios, hagamos unas nuevas y entonces se contrató, digamos que empresa, para que generara toda esta estructuración de bases de datos y demás, resulta que se avanzó el asunto y ese lenguaje en el que se construyó, ¿verdad?, y estoy hablando desde la parte, no, más bien no técnica, o sea, ahorita no recuerdo la parte técnica, sino, ¿verdad? Te estoy hablando, así como en el conjunto de lo que pasó, bueno resulta que eran bases que no se podían conectar con lo institucional, a nivel de DATIC digamos como para almacenar esas bases y, por decirlo de alguna manera, correr los sistemas.

Ahorita no sé exactamente, ¿verdad?, como qué era el problema de fondo, pero, sé que no se comunicaba, entonces habían unos sistemas institucionales y teníamos esta construcción nueva que no se podían hablar, entonces ese fue un trabajo que no se articuló y que todavía ahorita, digamos, tenemos algún rezago en la comunicación de bases de datos institucionales porque cuando se hizo como institución el esfuerzo, fue un esfuerzo que no se dirigió de manera adecuada digamos y que fue verdad algo que no permitió digamos que ese paso de comunicación y de actualización de las bases de datos se diera. Entonces, ¿qué pasa? Tenemos bases de datos en registro que a veces no todas y no siempre, pero, a veces tiene problemas de comunicación con otros programas institucionales, por decirlo de alguna manera, el TEC Digital por ejemplo que maneja bases de datos de los estudiantes y todo, tiene que irse actualizando con las bases de admisión y registro ¿verdad? Porque a partir de ahí, por ejemplo, se hacen los grupos, se lleva el expediente estudiantil del estudiante, eso sí se puede ver ahí, digamos, todo el expediente, ¿verdad? El recorrido del estudiante, lo mismo más o menos que ustedes ven, y eso se alimenta de las bases de datos de admisión y registro. Te decía, y esto es tal vez empezar por el final, pero, a raíz de la investigación que se hizo, se generó, como te digo, un algoritmo, una fórmula, digamos, en donde usted la corría, corría los datos y te daba un porcentaje de probabilidad de deserción de cada estudiante. ¿Cómo se hizo eso? Bueno, se hizo un cubo de información en la OPI, la OPI maneja otras bases de datos, igual que se alimentan de las de registro, pero, que tienen que ir alimentando ellos también, no sé si se comunican 100% con las de registro, digamos o no, pero, ese cubo de información entonces es un acceso a diversas bases de datos que jalan las variables que se puntuaron como importantes en la investigación. Entonces, digamos, si usted ejecuta, por decirlo de alguna manera, ese algoritmo de cálculo, te jalan las variables que están en el cubo, que por inventar ahorita, porque en verdad no las recuerdo pero, digamos que el nombre, el avance, cuántos créditos matriculó y ganó en el primer semestre en el segundo, verdad, jala alguna información que ya nosotros sabemos que es predictiva de alguna manera o que en su momento nos tiró, digamos, algún nivel de predicción y entonces, verdad, la jala de ese cubo de información y da ese porcentaje entonces digamos que hay bases de datos institucionales, se supone que deberían de ser una, puede ser que no sean una, sino que tengan problemas de comunicación y que haya que estar, no sé, por decirte, descargando un, casi que un Excel, verdad, que tal vez no sea un Excel, lado para subirlo al otro, porque a veces no siempre, digamos, están, por decirlo así, como que usted pueda en un tiempo real actualizar y consultar, digamos que se ha ido trabajando en eso, pero, hubo institucionalmente ahí, ¿verdad?, como un desfase. Entonces, digamos, habría que ver cómo se actualizan esas bases, qué tan, ¿verdad?, qué tan reales son, qué tan al tiempo. Pero, se supone, digamos, que hay una sola gestión de la información institucional, ¿verdad? A veces no es tan real, pero, digamos que se supone. ¿Verdad? Entonces, con respecto a los estudiantes, sí, debería de estar toda esa información, digamos, desde las bases de datos de admisión y registro, y con esas es con el que se actualiza, ya sea en tiempo real o no, pero, con esas, digamos, son con las que se actualiza tanto los cubos de información de la OPI como, por ejemplo, TEC Digital, que es el otro macro de información, digamos, que se maneja. Muy probablemente, casi que toda la información de los estudiantes está ahí hay cosas que no están en un expediente por decirlo común que es efectivamente información médica o sea la atención de la clínica y todo ello lo manejan en un sistema, pero, eso no está conectado con sistemas institucionales nosotros por ejemplo en atención adecuaciones, ¿verdad?, que se llaman necesidades educativas, esas cosas, pues manejamos cada quien los expedientes, pero, no hay un sistema de expedientes que me diga, sí, mira, este muchacho vino a tales citas, solicitó ayuda en tal programa, o sea, eso, digamos, a nivel ya de programas, por ejemplo, el DOP no está. Que digamos que es la información que va quedando ahí, ¿verdad? O sea, ya lo de las escuelas, propio de las escuelas, sé qué seguridad laboral, por ejemplo, creo que tal vez Biotecnología, digamos, hasta algún momento que conocí, pero, por lo menos Seguridad tengo pleno conocimiento que sí, llevan, así como el listado de los estudiantes, quién desertó, si no volvió, la secretaria los llama, la directora los convoca, habla con ellos, ver qué pasó, o sea, tienen como un seguimiento de sus estudiantes, ¿verdad? Como muy puntual, en biología.

Milton: ¿Pero, solo al nivel de la escuela se hace ese seguimiento nada más?

Teresa: Propiamente a la escuela. Este, y otras que no se dan cuenta quién hay o quién no hay nada más te ayudan a ir a clases y ya y no tienen como muy buena actitud de colaboración con los estudiantes para ninguna cosa, pero, eso ya depende de cada una de las escuelas un poquito para contarte la investigación, bueno yo te compartí el enlace del repositorio Para ninguna cosa, pero, eso, digamos, ya depende como de cada una de las escuelas. Un poquito para contarte la investigación, bueno, yo te compartí el enlace del repositorio. En el repositorio encontrás información de deserción como del 7-8, 7- 79, 85, ¿verdad? Hay algunos estudios, le llamábamos antes, ¿verdad? Porque no eran propiamente investigaciones, sino digamos como estudio de datos ahí. Entonces hay alguna de esas, luego no se volvió a tocar el tema de deserción prácticamente a nivel institucional, sí hay unas investigaciones que se hicieron a nivel CONARE, un poquito más actualizadas, por ahí están mencionadas en la investigación y ahorita no recuerdo pero, podría ser 2000, antes de 2010, digamos, y de ahí, de hecho, se conceptualiza el término, ¿verdad?, como de deserción, de deserción de primer, ¿verdad?, de primer ingreso, de primer grado, que son los que matricularon y se fueron o aceptaron en ese primer semestre y cuando la decepción es más avanzada, ya estudiantes porque eso puede ser que no les gustó el TEC que bla bla, que igual entraron, antes nos pasaba mucho que se ha tratado de corregir en los últimos ya dos años que toda la información de la UCR y todo el proceso de carrera UCR pasaba después del TEC, entonces mucha gente se matriculaba en el TEC para no perder la oportunidad, pero, si después eran sí UCR, entonces se pasaban y dejaban votado, de ese digamos hay un buen porcentaje todavía de estudiantes, sin embargo, ya el año pasado y este año la, como la información de si están en la carrera o no, si la tienen antes, así días antes de hacer la matrícula acá, entonces esperamos que eso como que vaya disminuyendo digamos que esa dimos cuenta que la carrera, nos dimos cuenta que la motivación, ejemplo, que el papel introductorio que tenga la universidad con respecto a la carrera, por ejemplo, los cursos de primer ingreso y eso tiene un papel fundamental, bueno, la lejanía también, ¿verdad?, o sea, la distancia que tenga que recorrer el estudiante, cuánto dure, digamos, esas cosas tienen como valor, y pues obviamente el rendimiento, ¿verdad?, entonces, ya hace como en los ochentas por ahí, se sabe y hay estudios, Don Edgar Guillén había hecho estudios, digamos, y es un dato que, ¿verdad? Históricamente hemos manejado y que uno sabe que sucede, aunque no haya nada, ¿verdad? Numérico, así como el estudio tal que lo respalde, pero, es que el estudiante que le va mal el primer año suele ser un estudiante, digamos, que no llega al final de la carrera. Un estudiante, digamos, que le va mal en el primer semestre y repunta un poquito ya en el segundo, es un estudiante que se puede ir acomodando, pero, un estudiante que tiene los dos primeros semestres mal, digamos, a nivel de rendimiento, es un estudiante que muy probablemente no finaliza la carrera, digamos, eso más o menos son cosas como que uno sabe, ¿verdad?, aunque nadie te diga o no te lo reseñale en una investigación. A partir entonces de esas variables, como te digo que fuimos detectando, que ahí están puntuadas en la investigación, se generó, eso lo hicimos con Martín Solís, que tal vez con él puedes hablar, de la Escuela de Administración de Empresas, que él es estadístico, y él hizo una revisión de diversos, vamos a ver si me acuerdo, de diversos, ¿qué será? Algoritmos de predicción es que no recuerdo cómo se llaman esos y el *machine learning* finalmente es al que le genera mejores índices digamos de predicción y entonces es con ese que se queda para la investigación. Cuando ya teníamos las variables determinadas en la primera investigación, ¿verdad?, que era un poco más cualitativa de ver qué tenía que ver con los estudiantes, aplicamos cuestionarios, ¿verdad?, entrevistamos gente que había desertado, gente que no había desertado, ¿verdad? Empezamos a ver cómo esas cosas, ya cuando teníamos esas variables, entonces se trabajó junto con la OPI, se construyó el cubo de información de ver si todas esas variables se estaban recolectando a nivel institucional o cómo hacíamos para recolectarlas. Se empezó a recolectar esa información y ya después de eso se generó, digamos, el algoritmo. Ese algoritmo, digamos, al final de esa segunda investigación, de esa segunda etapa, ese algoritmo queda a nivel institucional, ¿verdad?, con la base de datos asociada que era el cubo de información de la OPI, todavía está. Lo que pasa es que entonces

empezamos a chocar, con, ok, ¿y quién saca esa información? Y aquí lo delicado es, que uno debería de pensar que todas direcciones de Escuela y todos los profesores que podrían estar a cargo, por ejemplo, de este tipo de temas, tienen alguna sensibilidad. tienen alguna sensibilidad. Pero entonces resulta, o sea, sensibilidad y formación, ¿verdad? O sea, ética y todo lo demás. Pero entonces resulta que nos empezamos a dar cuenta, y esta es información que hay que ver cómo se maneja, ¿verdad? No hay que ponerla tal cual. Pero entonces nos empezamos a dar cuenta que a los profesores que les va peor en la evaluación, por ejemplo, de los estudiantes, ok, les van quitando cursos y entonces los van poniendo en comisiones y hacer cosas tal vez a veces un poco más administrativas porque son profesores que a veces, digamos, se les recomienda ir a formación docente, digamos, que da el SEDA, pero eso ya es un conflicto porque es un castigo, ¿no? Es una cosa, ¿verdad? Y entonces esas personas podría estar en un proyecto de la carrera para ayudar a los estudiantes, por ejemplo, ¿verdad?, con probabilidad de deserción. No los profesores, tal vez, que pueden ejercer una mejor mentoría, digamos, sino profesores que tal vez tienen alguna dificultad de relación con los estudiantes, o digamos, direcciones que no tienen sensibilidad en el asunto y entonces podría ser que vos estás, y no sé, que le diga a la secretaria, ok, llámeme a estos estudiantes que tienen un 80% más de probabilidad de excepción y que yo te llame y te diga, mira, ¿y por qué vos te vas a ir? Y usted dice, ah, ¿verdad? Perdón. Sí, es que aquí dicen que usted tiene un 80% de probabilidad, pero es que nosotros a nivel de otros datos, ¿verdad? De rendimiento, de cercanía, de no sé qué, de no sé qué, no sé qué, estamos valorando. Entonces, obviamente yo no sé si esa predicción se va a cumplir o no y debería de tener algún tipo de sensibilidad para ver si efectivamente ¿verdad?, hay ahí indicadores fuertes de que vos te vas a ir y cómo los vamos a empezar a manejar. sería genial que, por ejemplo, una persona de orientación en psicología, digamos, o alguien que esté propiamente en este proyecto, ahora empiece, digamos, a analizar esos datos, hablar con los estudiantes, ver qué está pasando, cuáles son los apoyos institucionales que se tienen para que eso no pase, etc. Pero resulta que no hay recurso, ¿verdad? O sea, si quisiéramos hacerlo nosotros acá en el DOP, no podemos, ¿verdad? Porque no hay, no alcanzamos, ¿verdad? O sea, tenemos lo que tenemos y más bien estamos sobrecargados, por mencionarte alguna cosa, el otro día hicimos un cálculo de las comisiones que nosotros tenemos a honor, o que participamos adicional, y son más o menos tres tiempos completos lo que hacemos entre todos, participando en cuestiones institucionales, digamos, que ni están en la carga, ni no son reconocidas de ninguna manera, y lo hacemos con gusto, pero no podríamos seguir asumiendo, ¿verdad?, ese tipo de cosas porque implican un montón de tiempo y trabajo que ya hemos llegado al punto de no tener, y sucede que nosotros DOP somos un departamento académico verdad? Que funcionamos como una Escuela digamos con cargas y con un rol docente pero no se nos llama docente pero si académico, pero obviamente no somos profes, entonces a veces, qué sé yo, la escuela de, por decirte, electrónica, ha tenido históricamente problemas de rendimiento, ha habido situaciones con estudiantes, en varios momentos se han hecho programas, entonces voy, hacemos una reunión con la gente que tiene problemas de RN, empezamos, digamos, yo empiezo a entrevistar a la gente individualmente, ver qué está pasando, ver qué podemos hacer, montar una propuesta, digamos, de acompañamiento y tal. Para eso, digamos, ahorita no tenemos tiempo como para nombrar una persona. La escuela podría ser que nos dijera, ah, mire, yo tengo un cuarto de tiempo acá, yo la nombro en este cuarto de tiempo y usted hace eso en ese tiempo, pero resulta que entonces ya por una situación externa de presupuestos institucionales, de este programa de Ay, se me olvidó el nombre que he dicho. Bueno, la ley está de salario único a nivel público, ¿verdad? En donde se incluye también a las universidades y en donde todos, ¿verdad? Bueno, tiene un nombre de si lo ocupas, yo te lo busco, pero entonces se empieza a ser más quisquilloso, ¿verdad? Como con las cuestiones de nombramientos y demás a nivel institucional, entonces como yo soy académica, pero mi puesto no se llama docente o profesor, entonces ya no me pueden nombrar, o sea recursos humanos no me pueden nombrar en eso, entonces tampoco hay una vía interna institucional que me diga que yo puedo colaborar con las escuelas con tiempo de las escuelas, ¿verdad? Porque a veces a las escuelas les quedan horas, ¿verdad? Y como veces a las escuelas les quedan horas ¿verdad? y como guindando, que tal vez no completaron en cursos o lo que

sea, entonces esa podría ser una vía de colaboración, ahorita se está viendo a ver si recursos humanos nos hace ¿verdad? como una que se pueda hacer un nombramiento equivalente digamos para esas horas y tal, pero ese ha sido un problema histórico y a la gente le pesa mucho, a los docentes les pesa mucho, que el DOP sea un departamento académico sin que tenga una carrera a cargo, ¿verdad? Entonces, a veces también es como, porque usted, ¿verdad? No puede hacer nada de eso. Entonces, digamos que empezamos como con ese, a chocar con esas cosas, principalmente, ok, la forma de sacar la información quedó lista. La base de datos quedó, el cubo de información, ¿verdad? Que se puede consultar ahí con las variables, está automatizado tal, quedó. ¿Y qué pasa con esa información? ¿ahí quedó porque entonces es ok quienes pueden tener acceso a este cubo ajá las direcciones y que van a hacer con esta información sacar un estado y que ajá verdad y entonces digamos que ya ahí se empezó como a cuestionar y cuestionar una cosa y otra y tal, y ya de alguna manera, ¿verdad? Ahí murió, digamos, el asunto. Hace algún tiempo, yo estaba en la dirección y luego entonces hace como dos, tres años se empezaron a crear unas comisiones que don Carlos Mata justamente nos estaba colaborando en algunas y eran como temas institucionales que requieren ser atendidos, entre esos estaba reprobación, estaba deserción, estaba la parte, por ejemplo, de cambio de carrera, estaban, ¿verdad? Como algunos temas ahí que habían resultado importantes en diversos espacios, hicimos comisiones de trabajo, se generaron algunas propuestas, pero de hace un año para acá, digamos que está la nueva administración, no se le dio seguimiento, no les pareció como importante, entonces hay algunos productos ahí de trabajo, de qué hacer con eso, de cómo hacer con eso, que tampoco, ¿verdad? Y es que aquí entonces vemos que nivel institucional, si no es una directriz de como eso va a funcionar a nivel macro, por ejemplo, la vicerrectoría docente diga, ok, esto va a ser así, cada escuela va a ser esto y cada, ¿verdad? Poco avance, digamos, tiene el asunto, entonces habría que ver, ¿verdad? Si además se va a generar algo, ok, ¿cuál va a ser la vía de institucionalización? Digámoslo de alguna manera de ese algo, porque ok, puede ser un producto muy bonito, a mí por ejemplo de la investigación y demás, ¿verdad? Esto que ya quedó, digamos, sistematizado, dime, parecía que iba a funcionar muy bien, pero no, porque finalmente no había una directriz para hacerlo, no había tampoco un control ¿verdad? Que se pudiera hacer, habría que ver en qué está eso y que vos pudieras por ejemplo revisar y ver si aún es viable y tal vez entonces lo que se requiera, ¿verdad? Más que estructurar el modelo, podría ser revisarlo y cómo aplicarlo, por ejemplo, ¿verdad? Emm... pero ahorita no sé cómo está, si está, si podría estar vigente aún digamos esa información ¿verdad? Es como complejo y nos hemos dado cuenta eso o sea de que, si no es como la directriz como muy marcada, poco pasa.

Milton: ¿Y a nivel interno se maneja algún índice de deserción, como para decir, esto es un problema porque está aumentando, o algo así? Porque a nivel interno, yo como estudiante nunca he escuchado eso, como que, uy, qué preocupación está aumentando la deserción, o sea, nunca he escuchado ese tipo de cosas.

Teresa: Pues no, es que vieras que no aumenta, o sea, vamos a ver, habría que sacar el, verdad, hay, como te digo, en la OPI, varios cubos de información, yo, por ejemplo, no puedo acceder a todos, pero depende del puesto que uno tenga, puede ir accediendo, Hay, como te digo, en la OPI varios cubos de información. Yo, por ejemplo, no puedo acceder a todos, pero depende del puesto que uno tenga, puede ir accediendo a diversa información. Si están, digamos, vos puedes ir y solicitar, bueno, vos como estudiante tal vez no, ¿verdad? Pero digamos que ya con los permisos y todo lo demás, uno puede ir y sacar el índice de deserción actual, eso digamos sí, se saca relativamente fácil, para ver cuánta gente ha aceptado, digamos, de la que había el semestre anterior a esta y esas cosas, pero sigue estando dentro del límite esperable, digámoslo de alguna manera. Hay me parece que, en el plan estratégico institucional de cuentas, ahí es donde están los planes y cumplimientos. Digamos, cada instancia trabaja con un plan anual de trabajo, con ciertas metas y todo, que se van sumando a nivel de vicerrectoría, y entonces la vicerrectoría tiene un plan anual operativo. Ese plan anual operativo de todas las vicerrectorías va respondiendo a las metas del PEI, que es el plan estratégico institucional, ese plan estratégico institucional ahorita, que es el que está vigente, se desarrolla digamos, o se propone a cinco años y en esos cinco años se tienen metas

de desarrollo y metas operativas. Las metas operativas son las que regularmente hacemos y seguimos y tenemos que dar informes y responder a eso y las metas de desarrollo son las que yo me propongo como institución a cumplir o desarrollar en esos cinco años, el que está ahorita digamos que es el 22-26 aquí está el formulado en septiembre 2023 y me parece que tiene una relacionada con deserción

Milton: eso sería super bueno porque hay que darme ahí como para dar una razón de que sí existe una intención.

Teresa: ¿Ves? Por ejemplo, en el plan estratégico 22-26 hay una meta que es la 4.3.3 que dice reducir al 2026 la deserción de estudiantes de primer ingreso al menos en un 10%. Entonces...

Milton: Mucho porcentaje, la verdad, no sé cómo lo van a hacer entonces sí.

Teresa: a menos de 10% lo que dice habría que ver, o sea, esto por ejemplo que está acá digamos la 4.33 hay que irse entonces usted, por ejemplo, bueno en el plan estratégico está 4.3.3 vamos a ver dónde nos sale más 4.3.3 ¿Ves? Por ejemplo, aquí está el desarrollo 2022-2026 y dice que en el 2022 dice 11%, 10% y dice que al 2026 debería estar en 9%. Al 2024 dice 10%, o sea, llegar al 10%. De ahí habría que ver dónde más lo desarrollar, vamos a ver.

Milton: Sí, a ver si mencionan algún plan o algo que tengan por ahí.

Teresa: Eso, como te digo, está asociado a los PAO y después de ahí a los planes anuales de trabajo. Yo sé que nosotros dos colaboramos en esa meta porque nosotros trabajamos entonces con los estudiantes para que tienen dificultades académicas que tienen tanque entonces todo lo que nosotros hacemos se suma ¿verdad? Si un estudiante viene con un conflicto familiar, si viene con un RN alto, si un estudiante tal uno trabaja y trata de dar soluciones y coordinar y solventar y eso disminuye la excepción, digamos en un macro. Nada más que a veces no es tan puntual. Aquí, por ejemplo, la meta, el indicador está detallado. Viene la fórmula de cálculo que dice que se toma en cuenta la diferencia porcentual de la excepción de estudiantes de primer ingreso en el año y de primer ingreso en la línea base y vienen algunos cálculos ahí y debería estar asociado tal vez no aquí, pero entonces en el PAO hay que ir buscando el otro documento, el que sigue para abajo, y ver dónde está desarrollada la 4.3.3. Y entonces usted va a la 4.3.3 y debería tener asociadas actividades o metas, ¿verdad?, de las vicerrectorías, y después en el de las vicerrectorías debería tener asociadas metas a cada uno de los departamentos y así sucesivamente. Entonces, pegarle una buscada, digamos, ahí como a qué está, pero eso está, como te digo, acá en transparencia, en planes de cumplimiento, está el plan anual operativo, que de verdad es el que cae después, o sea está el plan estratégico, en el plan anual operativo debería de estar la 4.3.3 tal cual y debería de tener asociados acciones de las de los diferentes departamentos

Milton: de los diferentes departamentos ok, eso me va a servir bastante

Teresa: aquí está el PAO 2024 y en el PAO 2024 buscas debería coincidir la 4.3.3 alcanzar al 2026 una deserción bla bla y dice que está asociada a la 2.1.4.1. realizaron 383 actividades para mitigar la deserción de estudiantes de primer ingreso en los programas de grado entonces hay que ver la 2.1.4.1. a que se asocia tal y bla bla, pero ahí puedes ir dándole como búsqueda.

Milton: ¿Y este plan dice que es a 5 años, ya llevan años en esto?

Teresa: Es 22-26

Milton: Sí, porque podría dar como Tal vez a ver si de algo se ha cumplido Y si no se ha cumplido, poder identificarlo

Teresa: Ahí digamos igual, en la evaluación del PAO año a año ahí está la del 22, 23.

Milton: Ah bueno, tiene un análisis de resultados digamos, perfecto si se me queda, súper genial, si porque de hecho una de las preguntas que tenía era eso que si hay una disposición del TEC para hacer estos proyectos, lo tiene en papel, ahora si a nivel de acciones es lo que quería revisar también si ya me comentó un poco lo de la disposición también del departamento de psicología y algo que tiene algunos problemas pero más que nada de los problemas que tenía y es siguiente la siguiente pregunta digamos es como de su experiencia qué complicaciones tuvieron que ya mencionó un poco como de y más que nada fue un tema ético entiendo yo y un tema de burocracia básicamente.

Teresa: Sí y digamos, o sea, el trabajo en deserción, el trabajo hormiga, en verdad, es de todos los días, o sea, y uno lo trabaja con todos los estudiantes que vienen a consulta acá, a asesoría psicoeducativa, a atención psicológica, de necesidades educativas, está el programa IntegraTEC, por ejemplo, que es el de mentoría para estudiantes de primer ingreso, entonces a cada estudiante de primer ingreso se le asigna un mentor para que vaya trabajando juntos, entonces todo eso incide en deserción, ¿por qué? ¿qué buscamos? Generar una identidad del estudiante con la institución o atender las diferentes situaciones personales o académicas que van acompañando al estudiante en su camino y evitar la deserción. Pero no es que yo diga que trabajo en deserción, ¿verdad? O sea, todo incide. Porque si un estudiante viene con una situación familiar, de pareja, de depresión, de tal, ¿verdad? Es un estudiante que ya tiene un gran porcentaje, digamos, de probabilidad de ser tal, y el ir ayudando, no, mira, vamos a hacer esto, vamos a... El estudiante en general pregunta poco y tiene poco conocimiento de los apoyos institucionales que se pueden brindar, ¿verdad? Entonces a veces un estudiante viene aquí y dice, no, es que yo ya no puedo más y tal, bueno, hagamos un retiro. Bueno, mira, te parece si el otro semestre yo hago una recomendación de que tu beca no sea con 12 créditos, sino con 8. ¿Eso se puede? Sí, sí se puede, porque estás en un acompañamiento y vamos a hacer, ¿verdad? Entonces, esas son acciones puntuales que hacen que un estudiante se mantenga, ¿verdad? Y vuelva a agarrar, digamos, como el ritmo y la motivación y todo. Pero no están reportadas como trabajo en deserción, ¿verdad?

Milton: Si contribuyen, pero no es de forma indirecta.

Teresa: Entonces, pensaría yo que además de lo que se llame deserción puntualmente, hay otros factores, otras acciones, digamos, que se van haciendo y que tienen incidencia.

Milton: Ok, perfecto. Entonces, a nivel de... a nivel de cuando ustedes querían implementar el cubo y ese tipo de cosas a nivel de TEC o el DOP propiamente no tuvieron algún problema ya sea como monetariamente hablando o de cierta voluntad porque también en estos años ha habido cierta controversia con ciertos puntos por ejemplo el FEES y eso, el TEC no tiene plata y ese tipo de cosas entonces no sé si como en estos proyectos no se tuvieron…

Teresa: No, porque en realidad está lo que se requiere como la fase de datos y está la sistematización del algoritmo, por decirlo de alguna manera, pero no hay personal que lo ejecute, o sea, alguien podría no sé, incluso ponga algún estadístico correr, ¿verdad? los datos, pero ¿qué se hace con eso?

Milton: Sí, no tenés alguien que lo pueda manejar

Teresa: Entonces, digamos, sí efectivamente, financieramente hay una afectación, porque no tenemos, ¿verdad? O sea, como para abrir una plaza nueva, por ejemplo, de trabajo en decepción con las escuelas, por decir algo.

Milton: Sí, ok. Y en cuanto están haciendo ese trabajo, ¿no encontraron algún producto similar, digamos, como un sistema de análisis de tal cosa que usted pueda llegar y meter al TEC?

Teresa: No había nada exactamente. Nada, nada, y más bien, digamos, lo que nosotros hicimos, que se presentó, bueno, yo sé que Martín lo presentó en uno o dos congresos, digamos, él tuvo ponencias, y era de lo más novedoso, o sea, no nada que encontráramos similar ni yo tengo conocimiento o sea como de que otra universidad esté haciendo algo parecido.

Milton: Sí, perfecto, sí, lo más que eran los estudios de los 70, 80.

Teresa: Sí, sí, están los estudios digamos de datos, pero no un sistema predictivo como tal.

Milton: Ok, perfecto, vamos a ver qué más te puedo decir de lo que ya me ha contado. Bueno, en cuanto al trabajo que se realizó, ¿siente que pudo tener alguna mejoría? necesarias, pero al menos las suficientes, digamos, como para que el trabajo se pudiera retomar ese proyecto digamos.

Teresa: Implementación, ¿verdad? O sea, porque creo que en realidad el trabajo abarcó las aristas no sé si las necesarias, pero al menos las suficientes para que el trabajo se pudiera continuar y entonces, no, como te digo, yo de la parte, ¿verdad?, de las matrices estas, predictivas y tal no tengo conocimiento, pero sé que fue de lo más novedoso entonces tampoco y que los datos nos daban mayores índices de precisión de lo que nosotros esperábamos entonces digamos a nivel estadístico también fue muy bueno, ¿verdad? O sea, los datos que obteníamos eran muy buenos. Entonces, yo creo que si hubiese una siguiente etapa sería implementación para nosotros, digamos, como proyecto.

Milton: A nivel, digamos, del producto que yo le estoy mencionando, que va a ser como un poco más trabajado con IA, para que sea un modelo más predictivo, bueno, de momento va a ser una escala muy pequeña, pero como han dudado tanto el TEC, y ya como me mencionó un poco, que no hay nadie aquí que lo aplique, que usted como, digamos, personal del propio departamento de psicología, ¿qué esperaría, digamos poder utilizar de ese sistema? O si lo ve más bien a visión suya, lo ve como solo una persona en un lugar apartado, o si espera que también tenga integración en otros departamentos, ¿cómo lo vería usted más bien a futuro? Es como más bien una visión, digamos, idealizada, de que tal vez, ok, departamento tenga, de psicología tenga acceso a ciertos datos para poder proveer cierta mejor ayuda.

Teresa: Sí, es que yo diría, tal vez por ahí, como que en la consulta tal vez del expediente estudiantil de cada estudiante. Porque yo, por ejemplo, un estudiante viene a cita, yo consulto el TEC digital y entonces ahí veo cuánto ha ganado, cuánto ha perdido, qué tipo de cursos ha repetido, ver cómo va en el plan de estudios. Entonces yo, por ejemplo, generaría una consulta adicional en el expediente estudiantil, reservada para algunas personas, tal vez no para todo el mundo, pero en donde yo pueda ver, por ejemplo, deserción y yo veo, mira, o sea, si hay aquí un riesgo y entonces yo trabajo un poco más en eso, digamos que podría ser consultado, ¿verdad?, en el expediente estudiantil por docentes, por, este, ¿verdad?, direcciones, personal en salud, personal, ¿verdad?, en acompañamiento, becas, por ejemplo, pero que ya sea una consulta que se genere ahí, o sea, que esté siempre corriéndose, digamos, de acuerdo al avance del estudiante y que haya, digamos, algún índice ahí que me marque semestre a semestre de ese estudiante, así como puedo ver el promedio ponderado, puedo ver tal cosa, y yo sé que a alguna gente no le va a interesar. Puede tener acceso un docente, pero si no le interesa, a ver si ese se va a ir o no, ¿verdad? Tampoco creo que vaya a mal manejar la información, pero para quien le interesa puede ser un elemento de análisis importante. Yo no lo tengo ahorita, por ejemplo, aunque sé que hay un cubo de información y demás, yo no puedo acceder a ese cubo de información en la OPI, debería tener los permisos y todo lo más, que igual el TEC Digital asigna permisos o no. No es cualquier persona que puede tener acceso al expediente estudiantil, digamos, pero en función de los permisos si yo pudiera generar de ahí una consulta consolidada, sería mucho mejor.

Milton: Y también en este proyecto no sé si ustedes tenían como un objetivo de esto quiero que lo vea tal persona porque eso me va a apoyar para mi objetivo final con este proyecto no sé si tenían algún como así alguna entidad objetivo y si y como decirle como que difundieron su trabajo para que esta persona pudiera observar de alguna forma se sintiera como difundido digamos que…

Teresa: Bueno externo como te digo si se presentó en varios congresos en varias cosas internos se presentó en consejo de rectoría en consejo digamos de docencia ahora se dio a conocer, pero de ahí de eso…

Milton: ok vamos a ver que más te puedo preguntar

Teresa: creo que en consejo institucional también se presentó sí.

Milton: eso sí tuvo internamente sí tuvo mucha difusión. Y a nivel también del departamento me dijo, bueno del DOP, también trabajé igual con a nivel de base de datos me mencionó eso de que al inicio se había hecho un trabajo que al final no se funcionó completamente, ahorita se ha arreglado algo o se mantiene igual.

Teresa: No sé, porque ya ahí sería qué tanto avance entre el DATIC y admisión y registro hayan conectado en las bases, principalmente el DATIC, que es el que tiene como el servidor institucional.

Milton: Que más te iba a preguntar se me fue la última pregunta en cuanto a la OPI entonces a nivel de contactos no sé si me podrías pasar tal vez el de no sé si tienes alguno de la OPI

Teresa: Mejor con la directora que ella estuvo digamos en conocimiento en su momento creo que apenas iba ella entrando que ella estuvo digamos en conocimiento en su momento creo que apenas iba ella entrando a trabajar y estuvo trabajando con nosotros porque de la OPI quien estuvo participando en la investigación fue Tatiana Fernández ya ella se pensionó

Milton: si justo ya le escribí en correo y como a 3 días de pensionarse y dije uy ya no me contestó

Teresa: entonces puedes hablar con la directora que se llama Evelyn Hernández. Ella debería de tener, o sea, sé que ella tiene ahí como el tema vigente porque ella estuvo como participando en algo, del DATIC le puedes preguntar a Andrea Cavero, te recomendaría que hablaras con Martín, creo que Martín puede ser Martín Solís de Administración de Empresas, creo que Martín puede ser un mejor proveedor de información porque él tiene toda la parte digamos de estadística y de los sistemas estos de predicción y como por ahí empezar, ya después ir viendo qué se requiere

Milton: Sí, qué más te iba a decir, Sí, la OPI, creo que bueno, la parte financiera, bueno, sí directamente de becas y eso también está conectado con la misión de registro, entiendo porque yo pensé tal vez que como la beca es un poco más sensible y tiene esa información como…

Teresa: No, becas hay un sistema institucional que se está desarrollando hace un par de años y cada semestre como que se trata de desarrollar un módulo nuevo, pero no hay, o sea, no es algo como que yo en las bases de registro pueda consultar, ¿no?

Milton: A nivel legal, cuando hicieron el cubo, ¿no les hicieron mucho escandalo? Algo así como…

Teresa: No porque es porque es información que está institucional. Ahora, el detalle, verdad, es que justamente, verdad, que esto sí hay una oficina de protección de datos de los habitantes y tal, se han venido haciendo más recomendaciones en los últimos años y entonces ahorita, verdad, bueno, ya se está hablando más y sé que desde asesoría legal tienen más presente el tema de un consentimiento informado inicial que le permita al TEC trabajar con los datos. Porque, por ejemplo, vamos a ver, del programa de admisión restringida que yo trabajo, ellos firman

un consentimiento, un contrato que es un consentimiento informado amplio porque tienen algunas regulaciones ¿verdad? de normativa interna, pero por ejemplo yo ocupaba saber el rendimiento general de los chiquillos y tal, era más fácil que me generaran una base de datos, una consulta, digamos, en admisión y registro, con toda esa información que meterme yo al TEC digital estudiante por estudiante, pero del registro no me lo dieron por la sensibilidad de la información, entonces creo que se ha venido restringiendo el acceso a cierta información. Yo sé que yo lo puedo obtener y manejar y tal, porque yo tengo el consentimiento, digamos, de los estudiantes, porque cada uno ha firmado un contrato y pueden ir a auditoría y ver qué tal y yo sé que no se va a pasar con nombres y tal, pero digamos que no, o sea no hubo forma de conseguir como esa información ya consolidada, yo tenía acceso y puedo sacarlo y puedo, verdad, sé que tengo el permiso de cada estudiante para hacerlo. Pero eso sí ha venido siendo un tema cada vez más presente. Creo que no se ha resuelto aún, o sea, no tenemos un consentimiento informado institucional de cuando el estudiante entra, digamos.

Milton: Ok, y cuando hicieron el trabajo, ¿sí sintieron que todos los datos que tenían eran suficientes? ¿No vieron como algo externo que tenían que preguntar los estudiantes?

Teresa: No, sí preguntamos muchas cosas. O sea, sí, ¿verdad? Hicimos... Encuestas y todo eso. Y hay cosas que no están en una base. Por ejemplo, los estudiantes que trabajan. Eso no está. ¿Verdad? Ahora hay una... concesión especial, digamos, para el estudiante trabajador o sea si el estudiante trabaja puede entrar a ese programa por decir, o sea, que si haces esa anotación en becas y entonces tienes menas obligación de créditos para mantener la beca, igual si son padres o madres, pero esa es una información que maneja becas si el estudiante tiene una beca y tal pero no lo sabemos de un estudiante regular que talvez no necesite beca y se esté pagando sus estudios, esa por ejemplo era una información muy importante que tenía que ver con deserción pero no hay ningún sistema institucional que los recoja.

Apéndice M: Prescripción de entrevista con el Estadístico, Investigador y Profesor Martín Solís Salazar de la Escuela de Administración de Empresas.

Martin: Porque es que hay uno que sí es muy... que es mejor que los otros. O sea, sacaron como tres artículos en ese tiempo, pero hay uno que es de triple E. No sé cuál... ¿Me puedes mostrar cuál tenés? ¿Los tenés por ahí?

Milton: Claro, ya le muestro. Por acá los tengo.

Martin: Ahí se me comparte la pantalla.

Milton: Sí, ya te comparto. Me parece que lo tenía en otro lado, si no... Si no me equivoco lo tenía por acá, pero no lo veo, si no los puedo bajar un toque de nuevo, a ver si están por acá, de casualidad. Ah, si no, es más. Sí, voy a bajarlos de nuevo, igual ya me los había pasado, entonces creo que me queda más sencillo buscarlo directamente de correo.

Martin: Hay uno en particular que es el más pero bueno ahorita hablaremos más extendido.

Milton: Entre ellos de hecho esos dos posiblemente ya los leí ambos me parecieron digamos trabajos muy anuentes a lo que estoy justo realizando en este momento y me pareció súper interesante

Martin: hay uno que ya realmente genera modelos con *machine learning* verdad, pero por eso quería ver cuál es cuál es el que tenes

Milton: ok si ya te muestro. ¿Usted participó, digamos, en ambos proyectos?

Martin: Es que no sé por qué decir dos, no sé a cuál te referís exactamente.

Milton: No, puede que sean mis dudas, que yo vi que eran como investigaciones, pero al rato ha sido lo mismo.

Martin: Fue una investigación y se sacaron tres, por así decirlo, tres artículos, el primero, pues no, a ver, la idea también del proyecto era generar un modelo predictivo, de la deserción, el primero que generamos, pues no salió tan bien, y después ya logró si logramos generar un modelo diría que bastante robusto por eso quería ver cuál tenías y además de eso se hizo como una encuesta también, pero ya la encuesta pues la encuesta era más descriptiva de razones por que los estudiantes desertaban, no era tanto un modelo predictivo en sí sino más bien una? De razones por que los estudiantes desertaban. No era tanto un modelo predictivo en sí, sino más bien una encuesta para entender por qué los estudiantes desertaban. Entonces, hay como tres documentos, pero refieren a la misma investigación, ¿verdad? Entonces, por eso te hacía la pregunta de qué es lo que tenías. sí.

Milton: Ah, ok, sí, por aquí ya tengo.

Martin: Mientras lo buscas, voy a, es que un poco, tengo la ventana abierta, estoy fuera de Costa Rica, está un poco caliente, entonces voy a poner el aire acondicionado. Si me das un chance. Bueno. Claro, claro.

Martin: Ese es Ese es como el más relevante Para el efecto de lo que se quiere hacer El segundo Este fue, sí, exacto. Y este otro fue el que yo hice también Sí, ese digamos Fue el primer intento que no quiere hacer el segundo este fue, sí, exacto, y este otro fue el que ya hizo también sí, ese digamos, ese fue el primer intento que no salió tan bien pero no es que esté malo, pero realmente el segundo sería más relevante

Milton: ok, perfecto

Martin: bueno, contame en qué tenés dudas y yo te puedo contar lo que sé, y un poquito más incluso, tal vez.

Milton: Ok, primero que nada, entonces, en cuanto a esta investigación, digamos, estos modelos, ¿cuál fue su rol? ¿Qué rol desempeñó y cuál fue su experiencia? Así como, no sé si podría contarme a grandes rasgos, digamos, cuál es la experiencia en esta investigación y este tipo de desarrollo de este modelo

Martin: bueno, sobre todo el segundo que te mostré que no es este que es en pantalla de realmente fui él digamos que el que llevó el rol principal del análisis de datos y también de los modelos que se generaron, pero ya después hay otra y de la encuesta verdad que pues hay mi rol fue un poco más secundario, pero básicamente sobre todo el segundo del que te mostré el que te indique del rol fue principal verdad? Yo, digamos, soy estadístico, pero también digamos, tengo una especialidad en ingeniería en computadores con énfasis en ciencia de datos. Entonces, pues ahí fue donde más me involucré. Además de eso, he de decirle a usted que ya hubo un estudiante de ATI que hizo la tesis específicamente con ese proyecto también, ¿verdad? La tesis de licenciatura, que es Roberto González, el tercero que ves ahí en las publicaciones, entonces Roberto diay mucho digamos del modelo, básicamente la parte computacional y estadística que encontrás ahí, lo hicimos entre yo y Roberto, ¿verdad?

Milton: Ok. Sí. Ok, perfecto. Y este modelo, a ver, digamos, cuando estaban realizándolo en su momento, ¿sí se había disposición del TEC, digamos, para llegar a implementarlo en su momento, si se había a disposición del TEC, digamos, para llegar a implementarlo en algún momento? ¿cómo fue su experiencia en cuanto a ello?

Martin: El modelo, bueno, no sé qué te habrá dicho Teresa, porque ahora me dijiste algo que, bueno, tal vez no sé si fue que te confundiste o qué sé yo o por las palabras que usó, que le entendí a usted que no se llegó a implementar esa fue la tarea principal de Roberto. podía tener una probabilidad en el *abstract*, en lo que está en negrito hasta de 87 o 91% de aceptar ¿verdad? si un estudiante iba a desertar entonces el modelo se llegó a implementar y digamos Roberto hizo la implementación cuando digo implementación es ponerlo en producción

porque el modelo se desarrolló pero después hay que ponerlo en producción para que funcione y entonces ahí estaba Tatiana que por ahí está en alguna de esas autoras que Tatiana en ese tiempo era la directora del departamento de planificación del TEC Tatiana Fernández, ya se pensionó por cierto el semestre pasado entonces Tatiana hubo mucha apertura para que Roberto pues hiciera como todo el proceso de producción de que los datos se descargaran de los cubos de deserción que tenía el departamento de planificación y eso ya obviamente lo absorbía el modelo como un input y el modelo daba predicciones por estudiante, la probabilidad de que aceptara o no y esa probabilidad al final la dicotomizamos de si va a aceptar o no va a aceptar. Entonces al final, además de eso, él hizo como una especie de visualización para visualizar todos los resultados, no sé si llamarlo una herramienta de inteligencia de negocio, qué sé yo, pero era como para que los resultados del modelo se pudieran visualizar por escuela, por área del TEC, etc. Entonces, toda esa puesta en producción digamos Roberto la dejo lista que sucedió bueno sucedió que se presentó en el consejo institucional se le presentó el rector se le presentó a todo el mundo pero qué pasa con esto Milton cuando usted implementa algo como esto tiene que haber una estrategia de cómo lo va a usar, verdad porque el modelo te dice todos los semestres usted podía correr el modelo al terminar el semestre porque el modelo agarraba información de las notas del estudiante, los cursos que matriculó etcétera, y el modelo te decía si próximamente el estudiante va a desertar, entonces todos los semestres, al terminar el semestre se podía correr, identificar a aquellos estudiantes que tenían más posibilidad de que desertaran, pero obviamente ya usted por tener esa información, ¿qué hace usted con eso? ¿verdad? Usted puede decidir hacer algo o no hacer nada, ¿verdad? Entonces, hacer algo con eso implica un trabajo, un trabajo extra, digamos, en el sentido de alguien, ¿verdad? De agarrar a esos estudiantes, de ver aquellos que tienen más problemas o aquellos que en teoría el modelo dice que van a desertar, pues darles un seguimiento, llamarlos ¿para qué? para prevenir y obviamente por si están pasando una situación, de la que sea que es la que lo va a hacer que deserte, pues obviamente ver cómo se puede impulsar o ayudar a ese estudiante para que no lo haga y entonces eso obviamente requería una labor entonces al fin y al cabo no hubo aunque todo el mundo decía que qué bien, que no sé qué, que la herramienta, entendiendo la herramienta como el modelo como tal, podía ser de utilidad y se aplaudía de eso, al fin y al cabo nunca nadie, digamos, de los tomadores de decisión, se estableció como una estrategia de cómo usar el modelo, ¿verdad? Ni hubo una directriz, por ejemplo, porque una directriz puede ser que los directores de escuela, al finalizar cada semestre, vean, porque ellos tenían acceso ya a la visualización, ¿verdad? Vean cuáles estudiantes tenían más potencialidad de desertar y qué hacían con ellos, ¿verdad? No sé, que les dieran un seguimiento. No había ninguna directriz de nada. Entonces, al final el modelo se implementó, se hizo toda una visualización, pero a lo que yo creo, nunca se usó, ¿verdad? O nunca se llegó a usar y no hubo una estrategia ni una política, una directriz de cómo usar, de qué se debía hacer con eso. Y entonces el tiempo pasó, esto, bueno, ahí ves ese artículo, fue el 2018, por esa época fue que se hizo, y justamente el año pasado, exactamente hace un año, me contactó la gente del departamento de planificación, del TEC, porque querían actualizar el modelo, ¿verdad? O sea, eso estaba ahí, digamos, yo no sé si servía o ya no servía la implementación de Roberto, pero en todo caso querían actualizar el modelo y ponerlo otra vez en funcionamiento y todo esto entonces Milton para no hacerle muy larga la historia yo bueno, yo estaba muy lleno de carga les dije que les podía ayudar algo con eso y había obviamente que no sé, yo pedí a raíz de esa experiencia variables adicionales que no tenían estos modelos y otras cosas, pero había que como que hacer una depuración de los datos que se iban a usar para construir el modelo. Pero al final, diay no sé qué problemas abran internos en el departamento de planificación, pero bueno, nunca se generó los datos y claramente yo no era responsable de generar los datos yo decía que datos ocupaban los tenían que generar para construir el modelo ya después el plan era que yo construía el modelo y después de eso alguien lo iba a implementar o sea implementar digo poner producción. Y lo que habíamos hablado de que una posibilidad de que alguien lo pusiera en producción era de un estudiante precisamente ya sea trabajo final de graduación de grado de computación o de ATI o lo que sea y entonces esa era como la segunda etapa una vez que yo desarrollaba el modelo, pero realmente los datos nunca se llegaron a

desarrollar nunca me los llegaron a dar tuvimos algunas reuniones todavía este año en el primer semestre me volvieron a contactar y yo ya les dije que yo estaba lleno de carga pero que los podía asesorar un poco porque todavía tenían interés y me reuní incluso con las personas que puedo decir hasta los nombres, que iban a generar los datos, ¿verdad? Y al final pues no sé qué pasó, nunca más me volvieron a contactar. Entonces no sé, ¿verdad? A mí me parece que cuando estaba Tatiana directora, pues ahí está, por eso está en el artículo, ella tenía mucho interés y entonces ella movió todo para que los datos y la información se generaran, además Jason y Rebeca que fueron las personas que me dieron los datos en aquel tiempo en el 2018 prácticamente estaban a cargo de Tatiana, Tatiana era la jefe de ellos, entonces pues al final a veces generaban una primera iteración de datos yo encontraba errores, se los devolvía me los devolvían limpios hasta que logramos ya un *dataset* que estuviera completamente adecuado para construir los modelos. Pero tenía como esa facilidad de que Jason y Rebeca, que son de computación, como les digo, me daban esa información. Pero de esta vez algo ha pasado en la estructura que Jason y Rebeca no trabajan ya para la oficina de planificación, sino para el DATIC, ¿verdad? Sin embargo, son los encargados del sistema de datos, digamos, de la base de datos, o de las bases de datos que maneja el departamento de planificación, aunque ya no trabajan para el departamento de planificación, sino que son de otra dependencia entonces, las personas con las que me reuní el año pasado y este año para generar sus datos, son los mismos Jason y Rebeca, pero no sé qué ha pasado en el camino no sé si es que tal vez la directora actualmente de la OPI, pues de la oficina de planificación, tendrá algún interés obviamente tiene un interés en desarrollar el modelo pero no sé si es que no tiene tanta fuerza o no ha logrado que Jason y Rebeca se puedan incorporar más de lleno porque ya no están a cargo de ella, verdad? Jason y Rebeca tendrán otras labores en el DATIC, no sé qué ha pasado no lo puedo precisar, pero esa obtención de datos que fue más sencilla en el 2018, entre comillas sencilla, pero me refiero que había esa facilidad de acceder a ellos en el año pasado y este año inicialmente no fue así. Todo eso se lo cuento para que entienda un poco de contexto de lo que pasó con el proyecto en aquel momento y lo que está pasando ahora que hay un interés en ponerlo en funcionamiento.

Milton: Si, más que nada fue por disposición de los integrantes digamos que se dio esa oportunidad prácticamente sí que directamente digamos quería evaluar digamos yo ahorita estoy evaluando o formulando y el proyecto como tal para ver si se puede digamos existe un interés y una posibilidad técnica de poder implementarlo, a nivel de implementación verdad de ponerlo en producción y que esta vez sí haya alguien que lo pueda ejecutar para no repetir ese mismo error que pasó en el pasado.

Martin: Pero, pero, perdón Milton es que cuando decís que lo pueda ejecutar no sé a qué se refiere porque tal vez, digamos Milton, o sea, yo creo que vale la pena volver a hacerlo, o sea, y ojalá lo puedas volver a hacer, ¿verdad? Pero digamos, o sea, se ejecutó, el proyecto se ejecutó al 100% en el sentido que quedó funcionando, o sea, el modelo funcionaba, los directores podían ver los datos, pero hay algo más allá, digamos, que sobrepasa a la parte de TI, digamos, que es la política institucional verdad? Si un tomador de decisión ve el modelo y ve que genial y le dice a Milton, esto nos va a servir, esto nos va a ayudar pero no ejerce ninguna directriz de que se use como tal eso no se va a usar, por más que pueda ayudar a prevenir o a contribuir en algo a disminuir la deserción simplemente, entonces a ver, uno al fin y al cabo no puede no puede digamos, quedar a expensas de la política, o sea, tiene que desarrollar el modelo y ejecutarlo, ya después lo que pase ahí son decisiones políticas que van a estar fuera de su alcance, entonces usted puede llegar a implementarlo Milton, pero podría repetirse la historia o no, pero eso va a depender de quién está arriba tomando decisiones y si esa persona realmente establece alguna directriz para que le modelo se use verdad? Porque alguien tiene que darle seguimiento a esos estudiantes que el modelo detecta como posibles desertores, entonces incluso hasta generar una plaza, no sé si incluso, por ejemplo, si fuera, si eso se le da como responsabilidad a la VIEZA, tiene que haber alguien que se ha encargado de darle seguimiento a los estudiantes, de llamarlos, de contactarlos, de ver cómo están, qué problemas pueden tener, ¿verdad? Entonces, ahí estamos hablando hasta de plazas, de recursos adicionales verdad entonces o si eso será

los directores pues alguien tiene que haber seguimiento los directores realmente estén usando el modelo para prevenir la deserción entonces claro detrás de lo que viene después de implementar el modelo es político meramente político y ahí puede ser que eso no avance más o si dependiendo quién esté, eso ya está lejos de sus manos lejos de las mías y lejos de quien sea, porque eso ya va a depender del tomador de decisión de turno que esté en el momento pero es como mi visión

Milton: De hecho, a nivel de complicaciones digamos, podemos resumirlo en burocracia prácticamente

Martin: interés político, realmente burocracia, prácticamente. Interés político, realmente, burocracia, tal vez no sé si sea burocracia, sino interés, interés, un interés serio realmente de que se use, y también una perspicacia de saber qué hago yo con eso, ¿verdad? Porque si al fin y al cabo usted dice, bueno, eso me da los desertores, ¿y ahora qué hacemos? Entonces, bueno, yo te estoy diciendo algunas ideas, ¿verdad? O sea, a que alguien le dé seguimiento a los estudiantes, pero perfecto, ¿quién lo hace? Y yo cómo me doy cuenta o cómo damos seguimiento, si a esos personalmente le están dando seguimiento a lo que el modelo está indicando, ¿verdad? entonces es un interés político, básicamente yo no le llamaría burocracia, le llevaría interés político de usar el modelo, básicamente ¿verdad?

Milton: Sí, antes comentó algo muy interesante y es la parte del modelo como tal que me comentaba que, bueno, tiene unos nodos variables, digamos, adicionales. Entonces mi pregunta es, ya vamos ya un poco más al lado técnico, sobre este modelo, si siente que a día de hoy aún está vigente o puede tener mejoras importantes, debido ya a la parte postpandemia, verdad, que se podrían crear nuevas variables y nuevos entornos digamos, y en cuanto también a la calidad de datos y en ese momento ese modelo, digamos considera que tuvo una buena calidad de datos o podría haber sido mejor en ciertos aspectos en esa parte, digamos

Martin: Pues los datos no eran los mejor, no era la mejor calidad de datos en aquel tiempo y hay cosas en eso que se pueden mejorar, que yo lo hablé el semestre pasado cuando se quiso retomar esto. Por ejemplo, la variable de becas ¿verdad? entonces, ¿a qué me refiero con eso? que los datos no son estáticos, por ejemplo, que Milton entró en el 2000, no sé, por ejemplo, Milton entró en el 2018 al TEC, tal vez mucho después, pero por un ejemplo y Milton tal vez entró con beca en el primer semestre, pero no hay un solo dato de Milton en el *dataset*, porque cada semestre que va pasando, su información va cambiando. Entonces usted puede tener información de Milton que tenía beca, luego la perdió, luego la recuperó, luego tuvo beca varios semestres, al final la perdió, no sé, o al final entró sin beca y un pronto otro tiene beca varios semestres, al final la perdió no sé, ¿verdad? o al final entró sin beca y un pronto otro tiene beca entonces nosotros teníamos una variable que era si el estudiante tiene beca o no, ¿por qué? porque esa variable podía ser un predictor de que un estudiante desertara o no ¿verdad? en el entendido de que usted tiene beca quizás eso hace menos probable que usted deserte, entonces esa variable en aquel tiempo era estática ¿verdad? estática que al final monitoreábamos si el estudiante tenía beca o no básicamente cuando entró pero ya no había información que permitiera darle un seguimiento al dato de Milton, entonces que cada semestre pudiéramos cambiar esa información de esa variable y al final el modelo tiene que ser flexible, o sea, me refiero esa información no puede ser estática no es lo ideal, sino que ir monitoreando semestre a semestre y en aquel tiempo eso no había acceso a los datos para tenerlos así hoy en día parece que me dijeron que podía verlo pero tenía que hacerse una limpieza con los datos y no sé qué y eso era parte de las cosas que también había que trabajar el semestre anterior que no avanzó como te dije, pero bueno eso por ponerte un ejemplo, así había otra variable que también se podía trabajar y mejorar en ese sentido entonces y hay otras ideas digamos ya técnicas pero quizás es más difícil discutirlas así nomás ¿verdad? pero de cosas que se pueden y que se podían probar digamos cosas nuevas que se pueden probar con el modelo ¿verdad? entonces es por ahí va la idea. Vigente, yo no diría que ese modelo esté vigente parece que justamente hay que actualizarlo ha pasado mucho tiempo y los

patrones las variables que usábamos como entrada para predecir la deserción esos patrones pueden haber cambiado perfectamente hoy en día, ya no son ni parecidos entonces hay que modificar todo eso hay que volverlo a actualizar el modelo eso sin lugar a duda me parece

Milton: Y en cuanto a las nuevas variables que mencionó no sé si puede darme algún ejemplo de alguna que detectó que recuerdo bueno yo tal vez lo que dije de beca estaba pensando en nuevas, eso era para mí una nueva variable porque antes era como estática y ahora pues iba a ser una variable que realmente iba a cambiar en el tiempo había otras pero ya no me recuerdo así como así Milton, pero si recuerdo que pedí como dos o tres variables nuevas que no estaban en ese estudio del 2000 del 2018 ¿verdad? Ahorita no lo recuerdo, me recuerdo eso de la beca que para mí era una nueva variable, ¿no? Porque era diferente, ¿verdad?

Milton: Sí, es claro, también como mucho fue antes de la pandemia y ahora, ¿verdad? Entró mucho la parte de virtualidad y mucho la parte psicológica en los estudiantes que solo estudian desde la casa y tienen sueltas distracciones y muchos temas así. Entonces, como puede afectar mucho el modelo en ese caso.

Martin: Totalmente, verdad. Incluso hay muchas cosas que se pueden hacer, pero eso es complicado, verdad. O sea, por ejemplo, la interacción del estudiante con el TEC Digital. Eso puede ser un predictor perfectamente de la deserción, ¿verdad? Es decir, puede tener, porque yo creo que eso está registrado en alguna parte y no sé quién lo tiene y si eso lo es, posiblemente es el TEC Digital, ¿verdad? Entonces Milton tiene un carné y podemos ver la interacción que está teniendo con la plataforma, ¿verdad? No sé la cantidad de veces que entró a la plataforma en todo el semestre, ¿verdad? Dividido por la cantidad de cursos, claramente, entre hermanos también hay un asunto ahí, ¿verdad? Que no es lo mismo, no es lo mismo, digamos, un curso que los cuatro profesores usen el TEC digital que un curso donde ningún profesor usa el TEC digital, ¿verdad? O sea, porque también puede pasar, aunque yo creo que la mayoría lo usan, no es exactamente que todo el mundo lo usa. Entonces, pues ahí hay cosas como relativas, ¿verdad? Que es difícil comparar a un estudiante con otro, pero hay ideas que a uno se le pueden ocurrir, ¿verdad? Entonces, por lo menos dentro del mismo grupo, dentro de la misma generación, no sé se pueden hacer, pero es complicado, ¿construir esas variables es complicado verdad?, usted puede no sé calcular comparado con el resto de los compañeros de generación vedad? Pensando que más o menos tienen profesores similares, en qué rango está usted de uso de plataforma, si está en el percentil 25 las personas que menos usan la plataforma o están en el percentil 50 las que más la usan, ¿verdad? Y entonces ese podría ser un predictor de la deserción perfectamente, ¿verdad? Porque el estudiante que va teniendo menos interacción ¿verdad? Puede ser que ya obviamente sea un estudiante que ha perdido el interés, pero no sé incluso si eso funcionará o no, son simplemente ideas y construir esas variables y eso demanda mucho tiempo y es complejo, yo no le diría a usted que se meta en eso, pero hay muchas cosas que se pueden probar en esto, pero muchísimas Milton, lo que pasa es que yo no me quise ni meter más después de lo que vi sobre todo el semestre pasado. Era eso que te dije, que no vi que había... ¿Ves? Eso sí es burocracia. Sí, eso sí es burocracia. Lo que pasó en aquel tiempo, cuando ya se hizo, era interés político. Lo que está pasando ahora, hoy en día, que los datos no me lo generaban, ya es burocracia. Algo pasa ahí entre comunicación de departamentos que no sencillamente las cosas no salen ¿verdad? y entonces pues ahí yo decidí abortar, o sea, decidí no, simplemente y porque tenía varias ideas de cosas que podían hacer que me interesaba probarlas y tenía ganas de probarlas, pero después de ver eso ni hice más la fuerza porque no quería gastarme ahí en esa lentitud para generar los datos

Milton: Sí, esas técnicas a nivel bastante técnico ya que me comentaba, verdad, que son cosas que usted quiere probar, no sé si me puede mencionar alguna que considera, así como que podría tener, podría causar una mejora, alguna que le gustaría probar en su momento. Ya por un tema, digamos, de tomar como referencia, si en algún momento se llega a generar.

Martín: Sí, sí, no, no, se la puedo, se la puedo mencionar. Me gustaría ver, porque yo tenía, justamente tenía una hoja de anotaciones, ¿verdad? De las cosas que quería hacer novedosas con el modelo, pero me gustaría volverlas a leer, ¿verdad? Porque ahorita no, o sea, para no llevarle mucho la idea y se las puedo luego comentar ahí por un correo, ¿verdad? Perfectamente,

Milton: Sí. perfecto y por último también de usted como profesor del TEC y no sé si tiene algún rol extra digamos en otra administrativa en algunos temas en otro lugar no sé si ustedes que esperaría digamos de un modelo así como que si usted en algún momento, como su rol que maneja de verdad en este momento en el TEC me esperaría por ejemplo consultar en algún momento ese tipo de herramienta o que simplemente no sé si a visión suya esa herramienta solo esté a disposición del top por ejemplo o alguna otra departamento digamos, no sé para mí tiene que estar a disposición del DOP, por ejemplo, o algún otro departamento, digamos.

Martín: Para mí tiene que estar a disposición de todo el TEC. En todo el TEC no, los estudiantes no, ¿verdad? Pero sí debería estar a disposición de los directores, ¿verdad? Del DOP, por supuesto, de la OPI, que es la oficina de planificación, porque la oficina de planificación tiene mucha interacción con la rectoría, ¿verdad? y entonces cualquier política que pueda surgir o que se quiera implementar o cualquier cosa que esté pasando, o sea, en la OPI, la oficina de planificación hay gente que está pendiente de los datos de todos los datos que se generan en la universidad para la toma de decisiones. Entonces, para mí eso debe estar en la OPI, en el DOP, y VIESA verdad? y esa información debería tener acceso todos los, ¿cómo se llaman? los directores de carrera, ¿verdad? porque los directores de carrera deberían piénselo así, al final del semestre que el modelo les diga qué tan probable es que el estudiante de deserte ¿verdad? entonces, ah bueno esa era otra cosa, ya me acordé otra cosa Milton de lo que me preguntaste anteriormente digamos que el modelo actualmente cada semestre lo que dice es cuál es la probabilidad que el estudiante deserte, ¿verdad? Simplemente que deserte. ¿Qué deserte cuándo? Que deserte en algún momento, ¿verdad? Yo, digamos, quería hacer pruebas para ver si se podía afinar más eso por ejemplo predecir que la probabilidad que el estudiante deserte el siguiente semestre o cuál es la probabilidad que el estudiante deserte hasta dentro de dos semestres o hasta dentro de tres semestres entonces como poder afinar digamos un poco más el momento en el que podría desertar eso tal vez no iba a dar, no sé si iba a dar buenos resultados o no, pero en el momento el modelo simplemente lo que te dice es cuál es la probabilidad de que Milton deserte hoy en día, claro si ya pasó un semestre más esa probabilidad de Milton pudo haber cambiado porque la información de Milton que se va a ligar al modelo luego de que pasó el segundo semestre del 2024 a cambiado, verdad? Quizás ya las notas bajaron, las notas eran un predictor claro de la deserción, conforme la nota va bajando aumentaba mucho la probabilidad de que un estudiante desertara y claro es que detrás de la nota está todo, ¿verdad? O sea, la nota es lo visible, pero lo que en el fondo refleja la nota son un montón de factores psicológicos, emocionales, socioeconómicos, ¿verdad? problemas familiares, porque tiene problemas de salud, porque emocionalmente no está bien, porque económicamente no está bien, porque está molesto con la carrera y no quiere seguirla, ¿verdad? Entonces, la nota refleja muchas cosas. Entonces, por ejemplo, su situación puede cambiar en el siguiente semestre perfectamente. Si la nota baja, puede ser que ya Milton entre, el modelo diga que ya las probabilidades de que Milton deserten sean más altas. Entonces, yo quería cómo poder ir analizándolo, ¿verdad? ¿Cuál es la probabilidad que deserte el siguiente semestre o en el siguiente año, ¿verdad? No solamente que diga cuál es la probabilidad que deserte en algún momento, Sino que eso es lo que quería un poco trabajar También, pero bueno, tenía otras ideas Esa era una que ahora me acordé De lo que se me preguntó, Pero bueno, yendo al punto Su pregunta en particular es eso Para mí el modelo debería tenerlo El DOP, la OPI Y debería tener debería tener acceso a todos los directores de escuela, ¿verdad?

Milton: Ok, perfecto, sí. Y esas son mis preguntas de momento, muy bien, muchísimas gracias por su tiempo, y parece que de verdad sí, es un proyecto que la verdad me parece una pena que justo no se esté utilizando porque

sí es muy valioso para ciertas decisiones y creo que de momento me parece que hay mucha luz verde digamos para ojalá poder darle contenido a ese tipo de proyectos porque se ven muy importantes y que no sé, la verdad es una lástima que no tenga esa disposición, digamos, ciertos actores para no seguirlo trabajando.

Martín: Sí, sí, es que eso es lo que pasa, al fin y al cabo, Milton, eso va a depender del de, alguien tiene que mandar una directriz con eso y si ese alguien no la hace pues eso se pierde, ¿verdad? Entonces el modelo yo creo que ya está desactualizado. No solo desactualizado creo que ya no está ni funcionando la plataforma en la que se desarrolló, ¿verdad? Entonces, bueno, la que se puso en producción, más bien. Pero bueno, ojalá puedas hacer algo y si al final lo puedes hacer pues obtener los datos que es lo que tienes que ir viendo esa es la parte que para mí no es compleja debería ser muy sencilla, pero por alguna razón y es bueno que usted lo mire de cerca no lo pude obtener no lo pude y no Martín Solís No lo pude obtener, ¿verdad? No lo pude obtener con una directriz de la OPI que quería volver a desarrollar ese modelo. No se hizo, no obtuve los datos, o sea, no sé qué pasó. Entonces es muy importante que lo vea porque si tienes problemas para obtener los datos, no sé si ya tienes datos o

Milton: si no, justo estaba organizando reuniones y ya me mandando correos para ver si me podía reunir con la gente del DATIC, de la OPI y otros lugares para ver si me podían dar como una muestra de datos a ver que se tiene

Martín: eso es importante que lo valores, verdad, porque si vas viendo que hay una pega ahí hijo, verdad, ahí puedes tomar decisiones de cambiar de trabajo de TFG, verdad, porque

Milton: igual mi TFG ahorita es solo evaluar y formular el proyecto, entonces de momento implementarlo ya en el modelo no, está fuera de mi alcance

Martín: bueno, bueno, bueno

Milton: más bien yo estoy hablando para ver si en el futuro alguien querrá retomar este proyecto más bien mi proyecto va a ser como un iniciador de otros proyectos podríamos decirlo así, digamos

Martín: ya, ya, si entonces no va a ocupar datos entonces

Milton: sí, o sea, los ocupo para ver si existen y si están disponibles y también podría haber ese tema, digamos de la burocracia, verdad, es decir, Es decir, al final es un tema, ¿verdad? Que no me los puedan dar, o otro tema, digamos. Pero en mi caso, yo solo los ocupo para ver qué hay y si existe, qué calidad tiene y ese tipo de datos, digamos, para ver si alguien más puede agarrar mi trabajo y así es y digamos, se da continuidad al proyecto, digamos basado en eso de recopilación

Martín: Ahí digamos, bueno, la ventaja es que no porque con los datos pasan muchas cosas, Milton, parece muy trivial, pero no es tan trivial los datos, el tipo de datos que se ocupa, por ejemplo, yo le pregunto a usted en este momento, Milton, para que te hagas una idea que es un desertor que quien me dirías digamos si yo le pidiera Milton de un ejemplo de quién es un desertor quien es un desertor para usted Milton

Milton: sí claro de hecho si no me quedo esas investigaciones ustedes lo definieron como desertor, así como de primer ingreso, otro como en matrícula, algo así, si no mal recuerdo.

Martin: Exacto.

Milton: Puede haber múltiples factores, digamos, múltiples formas de desertar.

Martin: Sí, nosotros al final, en ese artículo que te dije, es el más importante, se tomó como desertor aquella persona que después de un año de no haber matriculado, no volvió nunca más a hacerlo hasta el presente.

Entonces, ese se considera un desertor. ¿Por qué un año? Porque analizando los datos detectamos que, si ya una persona pasó un año sin matricular, es muy poco probable que vuelva a hacerlo, ¿verdad? Si pasa que un estudiante a veces deja de matricular un semestre y luego vuelve, ¿verdad? Pero entonces usted no sabe quién va a volver o no en el futuro, ¿verdad? Aun así, puede haber una persona que dejó un año sin matricular y cinco años después vuelve otra vez a la universidad, ¿verdad? Y usted lo tomó como desertor al final y resultó no serlo, pero realmente la mayoría de los que pasaron más de un año realmente nunca vuelven. Entonces, eso fue una definición de desertor, ¿verdad? Entonces, los datos, ¿cómo hay que pedirlos? Bueno, nosotros ahora para este proyecto habíamos pedido específicamente todas las matrículas del 2010 hasta la actualidad. Todas las matrículas todos los estudiantes que entraron desde el 2010 hasta la actualidad. Entonces ahí se iba a ver si iba a dar seguimiento estudiante. Entonces, por ejemplo, Milton, digamos, que entró en el 2010 y se le daba seguimiento a través del tiempo y íbamos viendo si en algún momento ya dejó de matricular y se declara como desertor entonces esa información había que pedirlo o sea, a ver, la generación de esa información parece algo sencillo y trivial Milton, pero en el proceso no lo es tanto entonces eso por un lado y hay otra cosa que nosotros por ejemplo queríamos ahora en este proyecto estoy contando que son algunas cosas que ya me preguntaste en el camino en este transcurso pero no se las dije ahora me estoy acordando una nueva no una variable que queríamos incluir, era porque en aquel tiempo la calculamos pero como no era una variable limpia, era una variable sucia digamos, queríamos determinar cuántos cursos, créditos le quedan a Milton para graduarse ¿verdad? y entonces esa información, aunque parezca mentira, Milton, no estaba no había un script que la generara así, tan automáticamente, ¿verdad? ¿O sea, al final alguien lo hace en registro porque cuando Milton termina la carrera alguien dice ya Milton terminó todos los cursos verdad? Pero no estaba como sistematizada, digamos como para todos los estudiantes desde el 2010 hasta la actualidad para decir, bueno, Milton entró y en este momento, digamos, le faltan 20 créditos, ahora le faltan 5, ahora le faltan 1, y eso no es sí, y es que no es tan sencillo, porque usted puede ser que lleve cursos que no son de la carrera entonces no es solamente como sumar los cursos y restarlos sino hay que hacer toda una programación ahí para darse cuenta y además usted sabe que los programas a veces van cambiando entonces hoy el programa de ATI es uno, tal vez dentro de un año le hicieron algún ajuste, a Milton no le cambia eso, pero la nueva gente que entra si le cambia entonces hay algo ahí que no es tan sencillo de un año y le hicieron algún ajuste, a Milton no le cambia eso, pero la nueva gente que entra sí le cambia, ¿verdad? Entonces, hay algo ahí que no es tan sencillo, digamos, de, a ver, no es nada del otro mundo, se puede hacer, pero es una programación ahí que no estaba bien desarrollada, digamos, o no estaba tan clara, y entonces de eso era una variable que queríamos incluir ahora, por ejemplo, te lo digo de lo que queríamos ver el semestre pasado entonces, bueno, no sé eso es lo que te puedo contar en la OPI te pueden dar más información como lo suyo es el planteamiento, porque la OPI era la interesada en esto, ¿verdad? y entonces ahí puedes hablar con Evelyn que es la directora de la OPI, y preguntar qué ha pasado, que por qué no ya te puede dar más, bueno, más no, pues no sé qué habrá, no sé, yo creo que lo que te conté es lo más reciente, como hasta enero, y de enero para acá, o febrero, estos meses, no sé qué ha pasado, si a lo mejor ellos siguieron trabajando, o algo lograron, pero yo diría que no, ¿verdad? Entonces ahí puedes informarte más también en la OPI, para que veas, al final todo esto, el uso de la información y modelos, al final lo que vos decís, hay burocracia, hay interés político, hay muchas cosas que puede ser que todo esté, para ponerlo a producción, pero hay algo ahí humano o burocrático, lo que sea, que hace que las cosas simplemente no salgan, y que parezca ilógico que no salgan porque es algo que es útil, pero no sale, ¿verdad? Entonces, bueno, eso es un poco lo que te puedo contar, sí.

Milton: Si, listo, muchísimas gracias por tu tiempo, muchísimas gracias por esta entrevista y sí, me ha servido de mucha utilidad la verdad, para generar este trabajo final de graduación mío entonces, al final, bueno, voy a ver si este final de esta semana te pasó como como lo discutí, prácticamente en resumen para ver si me lo puedes firmar y ya incluirlo en mi trabajo como una referencia

Martin: listo, listo, con mucho gusto Milton bueno bueno

Milton: eso sería, muchísimas gracias por tu tiempo y que pase una buena mañana

Martin: bueno, bueno, pura vida Milton hasta luego

Milton: Bueno, hasta luego.

Apéndice N: Prescripción de entrevista con la Evaluadora e investigadora jubilada Tania Elena Moreira Mora.

Milton: Hola, ¿qué tal?

Tania: Hola ¿Me escucha?

Milton: Hola, sí, perfectamente.

Tania: Ah, bueno, pero no lo veo a usted.

Milton: Ah, sí, dame un momento para acomodarme un momentito.

Tania: Sí, claro. ¿O no tiene cámara?

Milton: No, sí tengo, solo que la tengo desconectada por acá. Es que justo se la he prestado a alguien, entonces no lo tengo como. Dame un momentito nada más, por favor.

Tania: También, tranquilo, Milton.

Milton: Listo ahora sí, bueno, primero que nada, muchísimas gracias por atender la reunión, yo se que es un poco de mañana, ¿verdad? Pero ahí estaba viendo un poco sus investigaciones, ¿verdad? Sobre, digamos, ese tema de deserción que vi un poco con el Departamento de Psicología y otras entidades. Pero para contarle un poco, digamos, actualmente estoy cursando el trabajo final de graduación de mi carrera de ATI. Digamos que mi trabajo final consta de formular y evaluar un modelo de predicción de deserción. Entonces justo cuando estaba, digamos que, buscando ciertas referencias, me encontré esos trabajos y yo como mira, si ya anteriormente se había hecho algo, digamos, se había intentado algo. Y justo también antes de esos, digamos de termas de deserción, es casi que, o sea, es un tema muy viejo como que de los 70’s u 80’s solo había datos y ya después se saltó como a 2010 o algo así, ya, o sea, como que no había una continuidad directamente con este tema. Y entonces me quedé un poco leyendo sus trabajos y me di cuenta de que ese modelo estaba muy bien, o sea, de hecho, ya hablé con, con, a ver si se me fue con Martín Solís, con Martín, ya hablé con Martín y también, bueno, Tatiana justo se pensionó, no la pude entrevistar. Así que le decía que justo le puse un correo como tres días antes de que se pensionara, y yo dije, si ya no me contestó. Justo me mandó un correo, el profesor, tiene un sector mío, pero sí, ocupo tiempo.

Tania: Tatiana sí le responde, Tatiana es una persona muy comprometida con la institución, y bueno, qué dicha que ya se pensionó, pero yo sé que, si le escribe, ella le contesta, no hay ningún problema, ella siempre tiene una gran disponibilidad.

Milton: Sí, tengo que buscarme el contacto porque justo tengo el correo de ella, de la TEC, creo que ya no lo usa supongo.

Tania: Ah no, es que a nosotros apenas nos pensionamos, nos cortan el correo.

Milton: Sí, sí, por eso. Exactamente. Si es que no está en investigación, solo está el correo del TEC, digamos, entonces. Para ella no funciona.

Tania: Sí.

Milton: Sí, creo que entrevisté a Martín. Se me fue la otra persona. ¿Cómo se llamaba?

Tania: Martín, Tatiana. ¿Alguien de la oficina de planificación de la OPI?

Milton: Ah, no, de la OPI. Estoy pendiente, sí. De la OPI sí me quedo pendiente. Ay, se me fue el nombre. ¿Cómo puede ser? Voy a buscar en el correo más fácil está bien a ver se llamaba, la cuestión es que me reuní con ella primero fue la que me dio como las primeras entrevistas y ya me dio un poco de contexto de cómo la investigación y todo eso entonces diay me interesé un poco sobre el tema de, o sea, qué salió mal en cierto sentido, qué salió bien, cómo fue esa investigación, porque veo que es un modelo muy bien hecho al final, ¿verdad? Ya ha hecho Martín, me dijo que hicieron dos investigaciones, la primera no fue muy bien, pero la segunda ya dio un modelo con muy alta precisión, digamos, en cuanto a la predicción de la deserción. Entonces, me interesa un poco por el tema para ver, porque en mi caso, digamos, yo quiero evaluar, ¿verdad? Si es posible hacer un modelo. Veo que ya se hizo, entonces ya he hecho muy buena referencia. Pero también quiero saber qué se hizo mal o qué pasó, digamos, para el caso de que, para mí, digamos, en el caso de que alguien vaya a hacer un modelo nuevo, para que no se repita el mismo error, digamos, directamente. Y también ver qué impedimentos, digamos, existían en ese tipo de cosas. Entonces, ah, María Teresa fue, creo.

Tania: María Teresa Hernández.

Milton: Ajá, exactamente.

Tania: Ella fue una de las investigadoras del proyecto.

Milton: Sí, exactamente.

Tania: Ah, bueno, usted se leyó el artículo que publicamos en WOBI en el segundo con el que nosotros hicimos dos modelos corrimos dos modelos el primero que está ubicado en EDUCARE y el otro en ese congreso, que es donde participó un estudiante, que ya le digo el nombre, que ese estudiante precisamente para su trabajo final de graduación, trabajó con, digamos, el modelo que nosotros trabajamos, básicamente Martín, que es el experto en la digamos en correr estos modelos este el depuro incluimos otras variables y por eso salió mucho mejor pero ya le digo.

Milton: ah sí, es con *machine learning* si mal no recuerdo.

Tania: si, es con *machine learning*. ok entonces ese muchacho que digamos ese muchacho lo que hizo fue montar sobre ese primer modelo, lo depuró y ahí fue donde salieron excelentes resultados ya les digo lo que no sé es cómo el este Roberto González Solaiza. Entonces yo le voy a decir en pocas palabras que fue lo que falló, sencillamente no hubo quien les diera seguimiento a los algoritmos a alguien en la OPI entonces de eso no sobre lamentablemente no se volvió a lamentablemente, no se volvió a estimar, a hacer, a correr el modelo porque los algoritmos, este, no sé, ya no sé, no se le daba continuidad, no había quien hiciera eso.

Milton: Sí, de hecho, sí, como en entrevistas anteriores me comentaron mucho que era un tema como de disposición, digamos, de ciertas entidades.

Tania: Ese fue el principal problema es más todo eso, por eso yo le daría como primera recomendación que hable con el director de la OPI no sé si estará el mismo muchacho y bueno ah creo que Evelyn ¿Quién está de directora de la OPI?

Milton: Justo creo que sí es Evelyn. De hecho, le mandé un correo electrónico pidiéndole una entrevista, me parece, también.

Tania: Ok. Entonces, yo creo que Evelyn ha hecho trabajos en estadística. Entonces, ella perfectamente puede orientarlo. a ver, yo creo que usted podría, o sea, con esto le daría continuidad, obviamente, que otra vez correrlos, este, hacerlo otra vez como unas primeras pruebas, ver que se puede mejorar y ojalá es que el asunto es que no hay quien le dé continuidad a ese modelo, lamentablemente ese es el problema, talvez Evelyn porque Evelyn es estadística, si no me falla la memoria.

Milton: Ok. Entonces, para comenzar un poco las preguntas que le teníamos preparadas, no sé si me puede contar un poco sobre su rol y cuál fue su experiencia en esta investigación, así a primera instancia, cómo fue su experiencia, digamos, del desarrollo de esta investigación.

Tania: Bueno, yo fui la coordinadora de este proyecto en este proyecto se inscribió en la vicerrectoría de investigación y extensión entonces ya desde allí pues teníamos un reconocimiento digamos de parte de la vida de la calidad del proyecto fue un proyecto retador porque por eso Tatiana tuvo este un rol muy importante porque ella era en ese momento la directora de la OPI y yo creo fuimos los primeros en el TEC los indicadores del SIGI (sistema de información), creo que sí el SIGI, es que ya tengo tiempo de no, sí, el sistema de indicadores de gestión de la institución, exactamente SIGI. Entonces, tuvimos acceso a esos indicadores de que la OPI ya venía sistematizando entonces ella Tatiana digamos fue como parte como investigadora digamos nos dio todo ese soporte de tener acceso a esa base de datos Martín obviamente por su experticia ya más yo tengo formación estadística, pero no soy estadística. Pero Martín sí, y Martín ya se estaba especializando en modelos predictivos. Y María Teresa pues trabajaba también en la parte del Departamento de Orientación y Psicología, donde ella pues también, a partir de su experiencia en sus asesorías, pues también tiene muchos conocimientos sobre posibles las causas de deserción de los estudiantes más que ellos le dan acompañan acompañamiento a los jóvenes de primer ingreso este muchos de los programas también del DOP tratan de atender a la población de primer ingreso y estarlos apoyando precisamente para reducir el abandono estudiantil. Y pues yo que ya, yo trabajaba en el comité de examen de admisión, pero mi plaza era en el DOP, entonces yo asumí la coordinación porque yo sí ya había hecho antes estudios de deserción incluso ya yo creo que mi primer estudio data 2000 2003, 2004 por ahí entonces ya yo tenía cierta experiencia en este tema sobre la deserción entonces ahí conformamos ese equipo este y la verdad es que como usted lo apuntó bien se ha hecho estudios desde los años 70s 80s hicieron estudios hay bastantes en el repositorio pero como que ya después no se no se le dio continuidad y una de las, o sea, nosotros queríamos también atender una prioridad institucional que es reducir, o de alguna manera, pues, tener conocimiento cuáles son estas variables que están incidiendo en el abandono estudiantil para que desde el DOP, por ejemplo, y desde, por ejemplo, Viesa y desde el TEC, desde las escuelas, qué acciones se podrían ejecutar precisamente para ir reduciendo ese, ese porcentaje de abandono en los estudiantes de primer ingreso, que es un problema que, bueno, ya yo no tengo datos recientes pero si era un porcentaje bastante alto de abandono este de estudiantes de primer ingreso bueno en algo usted ya para que los datos tendría que irme a los artículos para darle el dato preciso pero usted lo tiene éste básicamente por eso surgió este proyecto y trabajamos como varias etapas. Este primero que era como para identificar esas variables y también ahí probamos varios modelos predictivos y completamos, lo interesante de esta primera etapa es que complementamos los datos estadísticos con una encuesta muy sencilla a estudiantes que abandonaron los estudios, que abandonaron y que continuaban en el TEC como

estudiantes, para identificar ya desde su experiencia personal, desde el punto de vista del estudiante, qué fue lo que incidió en precisamente en el abandono o la permanencia del TEC y pues ahí fue donde vimos claramente que la motivación era la variable más importante en este estudio que coincidía con estudios previos como la motivación es un factor el más importante en el abandono estudiantil en el caso del el TEC, igual coincide con muchos estudios a nivel nacional e internacional entonces esta es una variable muy importante, igual hay otras asociadas al abandono pero la más importante precisamente en primer lugar es pérdida de motivación de estudiar en el TEC, esa pérdida de interés de estar en la carrera, entonces no querían estar en el TEC y no querían esta carrera así en pocas palabras entonces después viendo la otra parte donde este muchacho González quería hacer su trabajo final y le igual retomemos este Roberto este retomemos precisamente este este modelo corramos otra vez con nuevos datos nuevas cortes y ahí eso fue donde entonces tuvimos esto logramos mejorar el modelo predictivo y creo que hasta ahí llegó yo creo que después de este estudio después de sus datos nadie le dio como no nadie estuvo dándole como alimentando este proceso dándole seguimiento y entonces creo que quedó en el olvido lamentablemente

Milton: Si, de hecho, Martin me decía que a él lo contacto la OPI, digamos, como hace el semestre pasado. Y lo mismo, como que quería, había una intención de continuarlo, pero al final no. O sea, fue como que se lo pidieron a él mejorarlo o actualizarlo, digamos.

Tania: ¿A quién, perdón?

Milton: A Martín.

Tania: A Martín.

Milton: Ajá. De como que le pidieron como actualizarlo a día de hoy, ¿verdad? Con nuevas variables, digamos. Y al pedir unos datos no se los dieron. Entonces, como que hay un pequeño problema, no sé si es de parte de la institución de la OPI o hay un poco de burocracia, digamos, que hay algún tema por ahí, que de hecho voy a tener que investigar un poco más porque también es algo afectado como tal. Y ustedes, cuando estaban haciendo esas encuestas, digamos, a los estudiantes, ¿no tuvieron algún impedimento para que ellos respondieran? O sea, ¿siguió una disposición de parte de ellos para aportar esos datos de cómo fue esa parte?

Tania: de por qué nosotros siempre recordemos que eso siempre es una participación voluntaria se le mandó este pues casi a todos los estudiantes de no sé si de la de cuál corte fue de 2017 este y se indica que esto es parte de un proceso de investigación, que la participación es voluntaria, y ahí uno logra, digamos, a los que decidan participar. está en el artículo que salió en Educario. Fueron muy pocas preguntas. Entonces, sí fue, se les hizo tantas preguntas en corto tiempo y pues obviamente que estaba también respaldado por el Departamento de Orientación y Psicología, que la información es confidencial. Entonces, en esos términos, como siguiendo todo el protocolo de aplicación de encuestas con la ventaja de trabajar con el SIGI y es que esa es una información que tiene el TEC de todos los estudiantes.

Milton: Ok, perfecto, sí. Y al momento en que ustedes hicieron esa investigación, ¿encontraron cómo hacerlo? ¿Como una exposición del TEC a que sí tenía un interés, digamos, en este tipo de proyectos?

Tania: creo que lo que nosotros digamos ya nos dio como como un digamos como una carta de presentación fue precisamente que fue un proyecto aprobado por la vicerrectoría, pero nosotros por la VIE por la vicerrectoría de investigación y extensión pero nosotros precisamente cuando yo cuando empecé a conformar el equipo este yo hablé directamente yo fui a hablar con Tatiana que en ese momento era la directora de la OPI le dije la intención del proyecto que inclusive yo le dije me gustaría saber quién de su equipo Tatiana podría incorporarse con

nosotros precisamente para aprovechar los indicadores del SIGI y trabajar como investigador en nuestro proyecto y entonces ella me dice pues a mí me interesa perfecto, pero entonces nosotros digamos antes de presentar el proyecto y así teníamos digamos hicimos una negociación con la OPI, por ejemplo, que para nosotros era importante, o sea, nosotros queríamos aprovechar los indicadores, pero di no teníamos acceso a ellos, no tenía sentido, entonces hicimos esa negociación, ya después yo conversé con Martín, que yo lo conozco de hace años, y Martín me dice si está bien no me comprometo mucho no martín usted se compromete a trabajar con los algoritmos y yo me encargo todo lo demás es como que cada uno tenía este su rol y entonces pero el primer paso fue tener el sí de la OPI y si del acceso, los indicadores es de acceso a la comunidad. Pero yo quería que alguien del SIGI estuviera en el equipo precisamente para tener, digamos, ya seguro que no íbamos a tener ningún inconveniente para el acceso a los datos del SIGI.

Milton: ok, perfecto, no sé si en su momento, digamos, existían algunos productos similares o algo así, como que usted ya ha visto, como que en algunas universidades se haya tenido un modelo de inteligencia, no sé, para revisar si, porque los estudiantes desertaban o algo así, o es un tema novedoso digamos en su momento

Tania: en Costa Rica no obviamente esto es un tema pues ampliamente investigado a nivel internacional, pero en Costa Rica en su momento no había ninguna universidad pública o sea me refiero a las universidades que se implementara estos modelos predictivos no ahora hay uno en la universidad nacional que de hecho es un trabajo que yo conozco muy bien, pero es se estimó varios modelos predictivos, pero en la deserción de un curso de un curso de matemática. Ese es de, ¿cómo se llama? De un colega de la escuela de matemáticas. Él presentó su tesis doctoral. Yo fui su directora de tesis. Entonces, por eso es un tema que yo conozco muy bien. con varios modelos predictivos, pero para identificar las variables asociadas a la deserción del curso de un curso que es diferente a la deserción de la universidad al abandono estudiantil de la universidad porque inclusive nosotros hay que tener mucho cuidado se puede ser abandono en la carrera, pero no abandono del TEC. Entonces, sí, hay que hacer esa diferencia.

Milton: Sí, claro, hay que identificar como un alcance, digamos, de qué significaría esa lesión para el trabajo como tal.

Tania: Entonces, por ejemplo, Andrey, Andrey sí ha trabajado este tema. Ahora es un experto también, igual que Martín, por su trabajo, su trabajo doctoral, pero lo hizo en la universidad nacional, pero en ese momento cuando nosotros lo hicimos para lo que era abandono estudiantil en el TEC entonces era como pionero precisamente a nivel de universidades públicas en ese momento Ya ahora está este otro estudio, pero es de curso de matemática en la UNA.

Milton: Perfecto. Y como en esta investigación, digamos, se recorrió algún gasto, un costo extra, o simplemente se utilizaron cosas que ya tenían, digamos, no sé cómo fue ese proceso, digamos, en cuanto a recursos, ya sea financiero, verdad, de personal o de otro tipo, digamos.

Tania: Bueno, a nosotros siempre nos ha costado un poco. Bueno, ya ahora, por dicha, hay como más, hay apoyo más de la Vicerrectoría de Investigación y Extensión, pero en ese momento nosotros no tuvimos asistentes. tiempo yo trabajé creo que martín tenía tiempo at honorem iguales de Tatiana trabajo tiempo at honorem yo creo que yo tenía cinco horas cinco horas para proyectos y maría teresa también tenía cinco horas de su carga para trabajar en el proyecto. Pero sí, digamos, como era un proyecto bastante ambicioso, sí, por lo menos contábamos con 20 horas, 20 horas entre cuatro investigadores, pero no teníamos asistentes. En cuanto a trabajo de campo, no se requería dinero porque en realidad los indicadores estaban ahí están digitalizados la encuesta que nosotros aplicamos lo hicimos de manera digitalizada también en línea entonces no hubo mayor este, inversión en es lo

que más cuesta es precisamente tener el tiempo para dedicarse al proyecto, pero bueno nosotros en general teníamos cinco horas como le digo este algunos *at honorem* en otros en su carga de trabajo

Milton: entiendo y este proyecto digamos con un poco como yo la pregunta también es que también me estaba pensando que tal vez ya, digamos, este proyecto podría estar a día de hoy, digamos, un poco desfasado, ¿verdad? Porque ya pasa un poco la pandemia, claro, o sea, se traen un montón de variables más, ¿verdad? De todos los estudiantes están algunos en virtualidad, otros tienen algunas complicaciones más intensas, digamos, en cierto sentido psicológico incluso. De hecho, sí conozco como experiencias cercanas digamos de amigos que dicen que de la virtualidad más bien no les sirvió para nada más bien los retos eso para ellos entonces no sé si usted considera digamos que este modelo tiene ciertas mejorías si alguna vez digamos ya con alguno trabajo que ha Si usted dice, yo podría haber hecho esto diferente de alguna forma o qué tipo de mejoras podría haber hecho desde su perspectiva, claro, ¿verdad?

Tania: Bueno, yo empezaría, digamos, ya sabemos que el *random forest* fue el algoritmo que mejor nos funcionó, pero bueno, yo tomaría como, no partiría de cero, digamos, tomaría, digamos, este último modelo. Por supuesto, es que todos los, estos modelos, precisamente, cada año hay que estarlos revisando, hay que hacer los entrenamientos, hay que estarlos probando, porque puede ser que ahora sí, definitivamente, haya variables, otras variables importantes. Yo, nada más por darle un ejemplo en mi último proyecto que trabajamos fue precisamente el proceso de ajuste a la vida universitaria en estudiantes de primer ingreso y este es un digamos este es un factor relevante en la permanencia un

estudiante que no tiene un buen proceso de ajuste en sus primeros meses es un es un potencial desertor de la universidad entonces pero eso fue todo un estudio que hicimos en 2020, 2021, 2022 este y entonces ahí nos damos cuenta que estudiantes con un bajo nivel de ajuste esos ya son posibles desertores, potencialmente desertores, entonces y por ejemplo, la vida universitaria es precisamente es precisamente cuando usted convive con sus padres, interactúa con sus profesores, va al pretil, come un heladito, este, en fin. Pero cuando estuvimos dos años, dos años en clases donde decía un estudiante, es que nada más conozco las voces de mis compañeros porque ni la cámara ponen. Entonces eso afecta mucho también su proceso de ajuste, entonces, y su motivación. Si yo tengo mi acompañamiento y me siento, pues, este, siento el apoyo de mis compañeros, este, de mis profesores, este, eso es muy diferente a estar metida encerrada, este, solo con una pantalla donde ni siquiera conozco las caras de mis compañeros entonces por supuesto que la pan… después de la pandemia ya hay un factor ahí de socialización de relación de padres de relación de pares de padres de compañeros que también definitivamente sería una variable muy interesante para incluirlo en el algoritmo, por ejemplo. Sí, claro. O sea, hay que empezar con lo, digamos, yo empezaría con estas últimas, estas variables que estaban ahí, que tenemos en los indicadores y empezar a hacer los entrenamientos, correrlos y ver que otras variables, donde usted tenga acceso, donde usted tenga acceso y que sean variables obviamente confiables, ¿verdad? Digamos, donde haya todo un procedimiento ahí de recolección riguroso, ¿verdad?

Milton: Sí, de hecho, estaba hablando mucho con Tatiana también, perdón, con María Teresa y era que ese tema, ¿verdad? Que hay algunos datos que ustedes tuvieron que ir a buscarlos porque no estaban, por ejemplo, porque tuvieron que aplicar las encuestas. Y yo creo que también de cierta parte también hay que buscar las formas de poder recolectar esos datos, o por lo menos establecer ciertas directrices. Porque sí, yo creo que a nivel de este proyecto también está difícil, porque para que se le dé continuidad, digamos, yo diría que hay que proponer de cierta forma una forma de manejarlo. Entonces, eso también creo que debería incluir ese tipo de herramientas, ¿verdad? Para recolectar nuevos datos o algo así que querían falta. Entonces, a nivel de ese proyecto.

Tania: Yo creo que ese es el punto más débil de estos modelos. Es que, si usted no le está dando mantenimiento al modelo y no vamos incorporando nuevas variables, creo que nosotros hacemos la recomendación de incluir dentro de los indicadores, por ejemplo, importante el nivel educativo de los padres, es que no me acuerdo porque, que eso fundamental en estos modelos predictivos, el nivel de escolaridad de sus padres. la parte socioeconómica, entonces hay indicadores que son altamente predictivos, pero que no están sistematizados en el SIGI, entonces nosotros decimos, y usted no puede tener, digamos, hacer una encuesta, y? Nosotros hicimos como preguntas muy puntuales porque usted abandonó el TEC o no abandonó el TEC. Ustedes, no me acuerdo de las preguntas, creo que eso está en el segundo artículo, en el primer artículo de la carrera y entonces fueron como preguntas muy puntuales que no tenían que ver con la medición de ningún constructo de algo más digamos técnico pero eso era para complementar los datos que ya si teníamos en el SIGI entonces ahí es donde usted digamos si usted quisiera por ejemplo incluir un indicador como el grado de ajuste el grado de ajuste a la vida universitaria tendría que allá hay un instrumento persona usted no podría hacer eso en este año hay que ser como más realistas, Milton, hay que ser realistas, es decir, estos son los datos que hay, probar nuevamente los algoritmos, porque no va a haber quien diga, bueno, sí, vamos a ir tratando de incluir nuevas variables en esto y darle mantenimiento al sistema y por eso es tan importante que usted lo negocie precisamente con la directora creo que se ven y si no me falla la memoria de la OPI este pero no podría aspirar a incluir nuevas variables que no están al alcance suyo y que en la OPI no le van a dar seguimiento ahora yo no sé si en la OPI hay nuevas variables que se han incluido nuevas variables esa es una pregunta muy importante para ella nosotros incluimos las que estaban en el 2016, 2017 puede ser que ya tengan más información de los estudiantes pero su caso, diay sería remozar los modelos, probarlos otra vez los algoritmos, ver cuales tiene mejor precisión, sensibilidad. Puede ser que en el SIGI ya haya nuevas variables que nosotros no teníamos en ese momento.

Milton: Sí, de hecho, para hacer la investigación, sí, tendría que probar. Y cuando realizaron, ya terminaron, digamos, esa investigación por cómo difundieron su trabajo, digamos, y no sé si tenían alguna persona que ustedes dijeran, yo quiero que ellos vean este trabajo porque va a ser un apoyo importante o algo así, digamos, como para poder llegar a esa persona, digamos.

Tania: Bueno, nosotros cuando tuvimos ya los primeros resultados, se le compartieron, hicimos varias presentaciones, lo hicimos por ejemplo en el consejo de docencia, consejo de docencia donde se reúnen todos los directores, obviamente en el consejo del DOP, y creo que en algún momento se presentó al consejo institucional, como para que ellos también tuviesen conocimiento de los resultados. Ajá. Sí, se hizo la presentación, un almuerzo académico de la VIE, de la Vicerrectoría de Investigación y Extensión. Se presentó también a otras instancias. pero si también creo que lo divulgamos en los medios de prensa, creo que en ese momento estaba el periódico todavía, pero si este Milton para hacer su trabajo entonces yo creo que para asegurarse hacer este trabajo a mí me gustaría ojalá que lo haga precisamente porque fue un trabajo muy grande el que se hizo una fuerte inversión digamos en cuanto al trabajo de investigación y que la idea de nosotros y estando Tatiana y era que le dieran continuidad por eso después llegó Roberto y le digo ay si perfecto y Roberto corre el segundo modelo y aquí tenemos los datos pero Roberto hizo su trabajo ya Tatiana dejó de ser directora de la OPI y ya se murió. Se murió la continuidad de, se murió en el sentido de que ya no se le dio continuidad al estudio. en ese sentido, cuál es su, digamos, propósito, bueno, aparte de graduarse, porque al final cuando uno hace una investigación es porque uno quiere contribuir, generar algún aporte, este, en este caso institucional, pero uno podría hacer un excelente trabajo y, pero si, si no se van a utilizar, que pena.

Milton: Así que me estás diciendo que su idea era que se le diera continuidad.

Tania: Si, yo le preguntaba precisamente porque el interés suyo en este tema pues usualmente uno tiene algún propósito también personal, no solamente académico, de hacer una investigación. Nosotros, por ejemplo, yo sí, cuando empecé con este proyecto este problema de deserción es preocupante y no se ha atendido y creo que tenemos las personas para hacer un buen estudio, robusto y que las autoridades tengan información reciente para toma de decisiones y entonces también hay como un cierto compromiso también social de hacer este tipo de investigación. Y entonces yo le preguntaba cuál es su interés, precisamente por qué su interés en este tema.

Milton: Ah si, justo yo estaba conversando con mi profesor tutor, digamos yo ahorita estoy bajo la modalidad de seminario de graduación, entonces yo estoy en conjunto con un profesor digamos de mi carrera para realizar ciertas investigaciones y en mi caso para mi trabajo final de graduación, entonces, este tema surgió a medida ¿cómo decirlo?, como este nuevo auge por las IA, por las Inteligentes Artificiales, y que de hecho tiene muy buenos modelos actualmente, digamos, para ese tipo de cosas, predictivos y esos temas. El tema de deserción, de hecho, nos interesó especialmente porque en el TEC nunca se habla, ¿verdad? Nunca nos dicen como, o sea, yo como estudiante, ¿verdad? No modo administrativo, sino yo como estudiante no escucho como, uy, en el TEC no sé, tiene un 20% de deserción o mucha gente hace algo así. De hecho, yo como estudiante nunca he escuchado ese tipo de cosas. Sí he visto, por ejemplo, los esfuerzos que se hacen, ¿verdad? Como diferentes, no sé. Las ayudas ¿verdad? Con éxito académico en matemáticas, los tutores de inicio de carrera, algunos de ese tipo de esfuerzos, ¿verdad? Pero tal como conversaba también con María Teresa, algunas de esas y muchas de hecho no sé, o sea, yo como estudiante no sé qué existe en ese momento. Como, por ejemplo, me comentaba María Teresa que cuando uno va al departamento de orientación y psicología, si uno tiene un problema en el que ellos le pueden ayudar y yo no puedo pasar por cierta cantidad de créditos y me van a bajar la beca, ellos me pueden dar un apoyo, ¿verdad? Como hacer una solicitud de que, si no puedo matricular, además de hecho créditos, entonces de que el TEC, bueno, se hable con becas, ¿verdad? Para que no me quiten la beca. Ese tipo de ayudas, digamos, yo como estudiante no lo sé. No lo, de hecho, nunca he visto algo así, digamos. Entonces, es como que hay cierta, digamos, poca difusión de algunas cosas y ese tema también está relacionado a la propia deserción, ¿verdad? Ya que también es un tema preocupante, ¿verdad? Es algo sumamente beneficioso para el TEC que eso se reduzca y de hecho también está incluso en propio plan institucional de que se quiere reducir esa deserción, entonces, por ese lado tenemos ese interés, digamos y también por la parte que nosotros como parte de la Escuela de ATI queremos a partir de mi trabajo, digamos, desarrollar nuevos trabajos de finales de graduación que pueda apoyar a otros estudiantes, incluso hacer un departamento encargado de este mismo tema, digamos, de deserción. Entonces, lo que queremos es darle continuidad, ¿verdad? Bueno, inicialmente era, ¿verdad? Era crear el modelo, pero ya como hay uno, ¿no? Bueno, ya como ya había uno existente, ¿verdad? Que está súper bien. Entonces, más bien sería como tomar ese como referencia y a partir de eso ya dar una continuidad. Entonces, sí, exactamente. Entonces, de nuestra parte, más bien, el interés ahora más bien cambió un poco, ¿verdad? Sin embargo, se mantiene el objetivo a partir de mi trabajo final si todo sale bien verdad ojalá que sí veríamos que a partir de mí va a crear nuevos proyectos unos para la por ejemplo uno para la evaluación de las variables del TEC que están relacionadas con la deserción otro para no sé la creación del modelo bajo algunos estándares de IA otros digamos otros relaciones a esos temas como nosotros verdad somos pocas personas relativamente, no tenemos un equipo especializado, ¿verdad? Que va a tener tanto tiempo trabajando en ese tipo de investigación. Entonces la idea es como repartir un proyecto, digamos, y ya luego unirlo todo, digamos, y ya tenemos un como un departamento que se va a encargar de esa parte. Y la idea es, digamos, proponer ya esas directrices, ¿verdad? Que este modelo va a estar encargado de tales personas y tales personas van a estar encargadas de estas responsabilidades, ya sea estar revisando, estar revisando cada año, cada cierto tiempo. Y también la idea es establecer, digamos, además de eso, quiénes van a estar con ese acceso a este modelo, ¿verdad? Porque la idea tampoco, como son datos sensibles, ¿verdad? Datos estudiantiles, no todas las personas deben tener acceso. Y también el objetivo es

determinar quién y para qué básicamente van a utilizar esos datos y de hecho es una pregunta que yo también le quería hacer y estoy un poco guiando la pregunta, que es que en su momento cuando ustedes hicieron la investigación, cuál era el objetivo final, quiénes querían en su caso, digamos, quiénes querían que usaran ese sistema y para qué, digamos, directamente. Y también a día, de hoy, digamos, ese comparándolo en el pasado, y ahora a día de hoy, digamos que, a visión, digamos, idealista, ¿a quién le gustaría más bien, así como ok, yo quiero que tal persona o tal departamento use este modelo para lograr estas cosas o que tengan esta información, digamos?

Tania: Inclusive yo lo vería, si usted me dice cuál sería, lo ideal es, por ejemplo, que nosotros lo hicimos a nivel de todos los estudiantes TEC, pero imagínese que con estos, teniendo esos algoritmos actualizados, y alguien de la OPI que trabaje en ello, imagínese que por escuela por escuela yo podría correr el modelo precisamente para ver potenciales buenos si hay suficientes estudiantes eso sí no sé si las muestras nos podrían afectar yo podría identificar por ejemplo potenciales desertores de la escuela de por ejemplo electro de computación o de administración de empresas, perfectamente se podría trabajar a nivel de escuelas. Creo que lo que pasa es que, vamos a ver, Milton, nosotros tenemos el modelo, pero ¿para qué tener, digamos, los datos de cuáles van a ser nuestros potenciales desertores si no hay acciones que permitan que no se nos vayan esos estudiantes? Entonces, de ahí tiene que haber un compromiso grande de las escuelas de trabajar con sus estudiantes de primer ingreso por ejemplo si no para que para que tener los modelos nosotros de alguna manera insistimos por eso trabajamos e incluimos digamos a los directores de las escuelas le dimos la información al consejo institucional pero eso implica también acciones institucionales acciones de las escuelas y este acompañamiento si a veces eso implica más recursos y bueno, hasta dónde llegar, hasta dónde se puede llegar, hasta dónde se puede llegar. Usted tiene el artículo que se publicó en la revista Educación. ¿Cuántos artículos tiene usted? Ah, bueno, porque el de Educación ahí está, de la encuestita que aplicamos. Sí, ahí está.

Milton: Sí, tengo dos. Bueno, ahí sí me compartieron el repositorio del DOP, creo que también hay algunas cosillas.

Tania: Hay uno que es de la revista Educare, que es el de los algoritmos propiamente el de la revista educación que se llama estudio descriptivo de las variables sociodemográficas y motivacionales

Milton: a ese creo que justo no lo tengo me parece

Tania: ya se voy a mandar ese tiene que ver precisamente con las preguntitas que le hicimos. No le gustó la carrera, incumplimiento de expectativas, entonces un 29%. Ahí están las preguntas y los porcentajes. No le gustó la carrera, incumplimiento 46,6. Y siempre esa fue la, digamos, la variable que más pes peso en una muestra no probabilística, por supuesto eso se lo aclaramos eso se lo voy a pasar entonces de una vez

Milton: perfecto, muchas gracias

Tania: pues si este Milton, entonces no sé, por eso yo creo que la primera negociación que usted tiene que hacer es con el OPI. Si no, entonces tendrá que cambiar de... Oh, es que no sé.

Milton: El problema es que, a nivel mío, en realidad no me afecta que el OPI me diga que no, porque yo solo estoy evaluando y formulando el proyecto, digamos. Yo veo si es posible o no. Entonces él no también es una respuesta válida para mi proyecto entonces a nivel mínimo más bien

Tania: Ah usted más tiene que formularlos y si no lo tiene que ejecutar

Milton: exactamente esa la salvada entonces más bien sí yo veo si yo que me dicen que no de más bien ahí tengo una respuesta válida para mi proyecto que de ahí, está la, o sea, están los datos pero no se pueden usar

Tania: No, los datos de la SIGI sí se pueden usar pueden usar. Sí se pueden usar porque son de acceso, son de acceso para, es más, los datos del SIGI los pueden utilizar los directores de escuela los investigadores los mismos profesores los investigadores los mismos profesores por ejemplo si quisiéramos pensar en una cátedra de física química o matemática que son donde hay altos porcentajes de digamos de abandono del curso perfectamente se puede usar los datos del SIGI para ver este de ir identificando cuál es el perfil de los estudiantes que tienden a abandonar ese curso. Entonces nosotros como directores de escuela, directoras de escuela, tomamos acciones para atender, identificar a esos potenciales desertores de un curso, o sea, que todo este modelo sirve para cursos sirve para carreras sirve para la institución general es importante y es cierto usted no le ejecuta eso le iba a decir nosotros duramos dos años ejecutando este proyecto y éramos cuatro investigadores con 20 horas totales, o sea, fue un trabajo importante

Milton: Si claro, de hecho, hablando con hablando con un profesor que creo que conoces ya me dijo si había trabajado con ustedes, se llama Carlos Mata igual de formulación y evaluación de proyectos, él es profesor mío de hecho le lleve mi proyecto, se asustó y dijo, ¿ustedes van a hacer el modelo? Y yo, no, no, no, no, no, no, porque si ya he visto sus trabajos, ¿verdad? Ya sabía que ese trabajo va a ser muy grande, o sea, que un semestre no me va a dar para hacerlo jamás.

Tania: Jamás, jamás, Milton.

Milton: Sí, sí, o sea, pero claro, esa es la parte importante, ¿verdad? Que yo solo lo voy a formular y lo voy a evaluar para ver si es posible hacerlo o no. Si no se puede, yo también tengo la respuesta, ¿verdad? Entonces, esa es la cuestión.

Tania: No, pero sí, el asunto es que usted lo va a formular, pero ojalá logre esa negociación y tal vez su… ¿Quién es el director suyo del trabajo o directora? Bueno, del trabajo... Ah, Juan Andrés Segreda se llama.

Tania: No sé, no lo conozco. Carlos Mata, sí. Sí, es que ya Carlos Mata, sí, es muy buen profe, la verdad. Tiene mucha *expertise* en esa área.

Tania: Sí, sí, sí, sí, sí, sí. Ah, bueno, entonces sí es formulación. pero bueno, igual converse con Evelyn, ojalá que después usted le haga una evolución y usted este trabaje con alguien y ojalá se llegue a implementar. O sea, hay quien lo puede hacer, esa implementación es la OPI. Tiene el recurso, tiene el acceso a los datos y ojalá que es ahí donde sea si usted lo logra presentar el proyecto y que lo ejecuten ahí que sería un éxito.

Milton: Sí, básicamente es el cómo el *stakeholder* digamos principal.

Tania: Exactamente, exactamente.

Milton: Sí, perfecto. Diay, creo que esas eran todas mis preguntas. Muchísimas gracias por su tiempo. Ya no le voy a quitar un poco más porque ya casi nos vamos a llegar a la hora ya.

Tania: Bueno, ni cuento. A mí me gustan estos temas. Pero si leas esos artículos, son esos tres. Y sí, de verdad que nosotros pues nos tuvimos un... De verdad que fue un equipo bastante interdisciplinario y logramos muy buenos resultados. Y bueno, el modelo, ya yo le digo, entonces parte de ese modelo inicial, diay y haga una nueva propuesta. Eso es lo que es, es una propuesta ¿verdad? lástima que todavía Andrey no ha sacado ninguna publicación porque le podría, bueno no sé yo sé que hay otros estudios a nivel internacional que los podría

considerar, pero en Costa Rica que yo conozco es este y ahora el de la Universidad Nacional que es con un curso pero igualmente válido para generar un modelo.

Milton: Ah bueno, se me había olvidado ¿de casualidad recuerda los algoritmos que llegaron a probar?

Tania: Si claro, lo tengo acá porque ya espera vamos a ver *machine learning* era perdón, si, *machine learning* regresión logística, redes neuronales, *random forest* y el *sport vector machine* probamos uno en redes neuronales este el tema el de vectores son vectores chin y el de regresión logística y en todos el que mejor no funcionó si no me equivoco fue el random forest.

Milton: ¿Arboles de decisión no lo tocaron?

Tania: No, no, creo que ese no.

Milton: Ah, bueno, perfecto. Ok. Listo. Eso es perfecto. Muchísimas gracias por su tiempo.

Tania: Con todo gusto mucha suerte luego me cuenta cualquier otra consulta con todo gusto

Milton: perfecto muchísimas gracias

Tania: bueno hasta luego

Milton: que pase una linda mañana

Apéndice O: Prescripción de entrevista con Arlyn Angulo Hernández y Raquel Mejías Elizondo, encargadas de la deserción en la OPI.

Aylin: Hay ahí de fondo con el sistema que nosotros tenemos acá, el sistema de indicadores de gestión institucional era que se había tomado la decisión de ya no tener todo el cubo de información que se tenía sobre este tema. Se empezó a trabajar nuevamente, igual de la mano con Martín, pero sí, básicamente por el tema este de los datos. Por eso fue por lo que no se siguió.

Milton: Ok, y este cubo actualmente, digamos, ¿se encuentra en ejecución? ¿Se tiene ahí en su momento o ya se dejó de un lado, digamos?

Raquel: No, actualmente el cubo no se está ejecutando, no está ni siquiera disponible y este, digamos, esa propuesta nueva que estamos haciendo, pues ya tenemos un primer conjunto de datos como para empezar a entrenar el modelo y demás, pero tenemos algunas observaciones a esos datos, algunos cambios que hay que hacer, entonces estamos, así como en el proceso de desarrollo, ¿verdad? De hecho, ahorita no se está trabajando como directamente en ese cubo, puesto que los desarrolladores que tenemos, pues estamos priorizando cubos, ¿verdad? Entonces, pero sí está en la lista de los cubos que se están trabajando.

Milton: Ah, ok. Sí, de hecho, Martín, bueno, conversando un poco con Martín, ¿verdad? Me había comentado que, si no me equivoco, el semestre pasado o el antepasado la habían contactado para intentar mejorar ese cubo. No sé si es parte de esa iniciativa, digamos, que se tiene, que no sé si se dejó de hacer o porque él también tenía ese desconocimiento, digamos, que él pidió ciertos datos, pero no le han llegado aún. Entonces, esperamos para conocer un poco.

Raquel: Yo creo que lo entrevistaste tal vez antes de que nos habláramos con él, porque ya lo hemos actualizado un poco más este porque si no algún digamos no es como que se ha abandonado se ha dejado sino que más bien

los analistas estaban preparando este conjunto de datos durante ese tiempo ellos estuvieron haciendo ese trabajo yo sé que tuvieron que hacer un trabajo muy grande el estudio curricular verdad como cuántos cursos le faltan a cada estudiante más porque hasta donde entiendo eso no estaba sistematizado en ningún otro sistema, digamos que hay algún trabajo ahí de manejo de datos que ellos hicieron para tener ese conjunto de datos que ya revisamos con Martín, ya le pasamos a él los datos, y él notó algunas cuestiones como que hay datos que se mantienen muy estáticos, ¿verdad? Y que deberían ser un poco más dinámicos y así. en el proceso en el que pasamos esas observaciones a los analistas, pero en este tiempo mientras Martín revisa datos y demás pues hay más cubos en medio y ellos entonces están trabajando en otras cosas, pero el cubo está en cronograma, en trabajo

Milton: Y a nivel de trabajo otros involucrados ahí para conocer un poco más el entorno porque a lo que entiendo, estar aparte de esos analistas entonces, están la parte de esos analistas, entonces. Está la parte de ustedes que están trabajando con datos. ¿Y qué otra parte involucra ahí en este proyecto?

Raquel: Yo no sé, igual me corriges si crees que es diferente, pero creo que es solamente, ¿verdad? O sea, además de Martín, ¿verdad? Que es el apoyo como más técnico también por el involucramiento que tuvo con el modelo anterior. Y, ¿verdad? Él muy amablemente nos ha seguido colaborando con ese proyecto. Entonces, Martín, nosotros, porque el resto de los datos también están en SIGI, ¿verdad? O se pueden construir con los que hay en SIGI y los analistas que se llaman Jason y Rebeca trabajan en el departamento de DATIC.

Milton: Ok, ahora un poco avanzando, digamos, en esa parte de antecedentes. En su momento, cuando se realizó el cubo de datos inicial, ¿verdad? Que fue entre los años, ¿verdad? 2019 y 2020, me parece. O 2021, creo que también entró un poco la pandemia. En su momento, digamos, me contaban en otras entrevistas que, por ejemplo, digamos, se desarrolló este modelo y el modelo se puso en ejecución en Adobe, pero además de eso, no se le dio como un seguimiento, o sea, como que no se tenía quién lo iba a usar o así, no sé si tiene definido. No sé si a raíz de esta nueva iniciativa, ¿verdad? De que se está diciendo de que se quiere actualizar el cubo, ya se ha actualizado algún tipo de, digamos, ¿de burocracia extra? Como para ver quién tiene permiso y quién no de usar ese cubo y para qué cosas, digamos.

Raquel: Sí, bueno, por ahora nosotros definimos algunos roles para SIGI, ¿verdad? Entonces, de hecho, hay ciertos cubos que no todo el mundo tiene acceso, hay otros que sí, ¿verdad? Entonces, por ejemplo, ahora desarrollamos un par de cubos nuevos que son más para directores de escuela porque es como la gente califica las carreras y así. Entonces, estamos generando también como un manual para que la gente pueda consultar, ¿verdad? Ya después auto consultar, ¿verdad? Entonces, queremos dar la capacitación, no la hemos hecho todavía, pero, por ejemplo, en este caso, damos la capacitación a los directores o a la gente que está involucrada con el tema de carreras, los usuarios que están involucrados con el tema de carreras en SIGI y les explicamos también más a fondo un poco el uso de esos dos cubos nuevos. Lo mismo haríamos con el de deserción, ¿verdad? Buscaríamos entonces en ese caso los puestos importantes, ¿verdad? Probablemente la Vicerrectoría de Servicios Académicos, el DOP, ¿verdad? Y con el Vicerrector de Docencia entonces veríamos cuál sería la estrategia para llevarlo a los directores de escuela, ¿ver de escuela verdad también es importante información un poco sensible como depende cómo se utilice verdad debería estar acompañado siento yo también por el tema del DOP entonces hay que lo manejaríamos seguramente, así como como el resto de los cubos verdad capacitando un poco a esos puestos claves

Milton: y a nivel de los datos digamos que han utilizado para este tipo de cubos que han tenido algún tipo inconveniente digamos en acceder a ciertas bases del TEC, porque no sé si existen algún tipo de silos, digamos, en distintos departamentos, y si quieren acceder a algún tipo de datos, han tenido algún tipo de problema, así, o ese tipo de cosas, digamos, para generar los datos. Ya ahora, por ejemplo, ahora me contaban, verdad, de que

están en un trabajo, verdad, de generar una currícula o algo así para determinar los tipos de, si ya han terminado la carrera o no, en cierto caso, y cuánto han tenido ese tipo de cosas. Entonces, no sé si me puede comentar si han tenido algún tipo de problema, si un poco más a fondo, ¿verdad? ¿Cómo lo resolvieron o en qué están trabajando?

Raquel: Bueno, esta parte, digamos, del tema de datos, ya en la práctica lo hacen mucho más Jason y Rebeca, ahora los analistas. Esta parte de la currícula, te lo comento porque es de lo que en las reuniones de equipo ellos me han comentado, ¿verdad? ¿No estoy segura si toda la información que utiliza el modelo estará en SIGI o no, vos lo sabes Aylin? ¿O si ellos consultarán otros sistemas?

Aylin: A mí me parece que sí, pero tal vez, así como retos que nos comentaban los compañeros, estaba el tema de que la información tiene que ser como en, se tiene que mover, digamos, que no sea información estática. Tal vez esa parte de que los registros se vayan moviendo conforme vaya pasando el tiempo dentro de la misma base de datos. Tal vez ese es como el reto más grande, ¿verdad? Que no pueden ser fotografías, sino que, digamos, si usted filtra algún carné, usted puede ver todo el recorrido de esa persona a través del tiempo. Y no que sea información estacionaria, porque después el modelo no te va a entender, ¿no? Porque es que pasó de... O porque más bien es que no se mueve la cantidad de cursos faltantes, porque siempre le hacen falta cinco si ya pasó tiempo. Tal vez esa parte es la que es un poco más complicada de conseguir, porque la información, digamos, que se le mete al modelo, sí es como mucho de lo que nosotros tenemos dentro del sistema. Pero sí, eso es algo que ya los compañeros de DATIC se encargan de programar y de entrenar.

Raquel: Pero sí, de hecho, ese es el reto que aún no hemos resuelto, ¿verdad? Con ellos, el tema de que hay, o sea, ellos hicieron el conjunto de datos, pero, por ejemplo, los cursos faltantes parecieran ser en todos los semestres salen los cursos faltantes del último semestre, por así decirlo, ¿verdad? En lugar de salir, bueno, si le faltaban menos y ha ido cursando cursos, pues debería ese número irse reduciendo, ¿verdad? Entonces, ese de hecho es, además de la currícula, ¿verdad? Que era justamente ir preguntándose cuántos cursos faltantes tiene el estudiante, ¿verdad? Porque además hay inclusiones, el estudiante puede matricular en diferentes planes de estudio, etc. Entonces, creo que eso es parte del principal reto también que tuvieron.

Milton: Esto es un poco el tema de datos. Ahora, no sé si han tenido un tipo de retos, digamos, también a nivel informativo, digamos ni en un poco más de gestión verdad con los demás miembros del equipo que están trabajando en este proyecto no sé si a nivel informativo tal vez o también poco tal vez con la implementación tal vez como como decirte como tal vez como que se tiene un problema y no se tiene esa voluntad para revisarlo si es que los superiores a sus personas digamos tienen si cierta voluntad de querer continuar con este proyecto y si existe esta iniciativa para continuarlo para un futuro. ¿Si han tenido retos con eso o simplemente si ven que las personas alrededor de ustedes sí tienen esa iniciativa para este tipo de proyectos?

Raquel: No, yo creo que sí tiene bastante importancia. De hecho, creo que también por ahí en algún consejo de rectoría hace poco la vicerrectora de Viesa mencionó que ella esperaba que ese cubo se siguiera trabajando. Yo creo que tal vez el reto también está en la cantidad de trabajo, ¿verdad? Tenemos solamente dos analistas que están desarrollando también, tenemos ahí un cubo de becas, ¿verdad? Que durante mucho tiempo la institución ha estado sin información sistematizada de los becados, digamos, ya de los indicadores como tal. Ese se vuelve un indicador sumamente importante, ¿verdad? Entonces, obviamente en prioridad sube, ¿verdad? Y no podemos hacer nada más que hacer alguna priorización porque es el, digamos, el modelo también de trabajo que tenemos. En el DATIC son un poco celosos con el tema de involucrar estudiantes también, ¿verdad? Y eso pues también tiene un montón de riesgos, ¿verdad? Que con alguien profesional pues no están, ¿verdad? Y que hay que también atender. Entonces, es entendible también esa parte, ¿verdad? Pero entonces, por ahí, creo que también está uno de los retos, ¿verdad? Que es el tema de recursos, que es un reto ya como conocido.

Milton: Entonces, a nivel de trabajo, digamos, siente que están como muy, ¿cómo decirlo? Muy superadas, digamos, de ciertos, entonces tiene que llegar a una priorización. Y a nivel de priorización, ¿cómo qué tanta prioridad le tienen actualmente a este proyecto? O sea, ¿está como un nivel intermedio, bajo o alto, digamos?

Raquel: No, bastante alto. Sería el que se sigue trabajando después de terminar el de becas.

Milton: Luego, a nivel, no sé si han tenido, bueno, ¿qué tanto han estado dentro de este proyecto como para saber si a nivel legal han tenido algún tipo de problema? Porque ya como están trabajando con datos tan sensibles, si han tenido que hacer algún tipo de gestión, ¿verdad? Ese tipo de la parte legal ya con su consentimiento informado o algo así. No sé si han terminado con ese tema.

Raquel: Justamente para este cubo no, digamos porque ya la información está en SIGI y demás, pero si todos los usuarios nuevos de SIGI si tienen que firmar un compromiso de confidencialidad, estamos de hecho actualizando y entonces el que tenga acceso a este cubo también tendría que estar bajo, digamos, un sistema de por lo menos que firme confidencialidad. Sobre los datos que se solicitan, creo que hay consentimientos informados desde admisión y registro para ciertos datos, pero sí tendríamos que explorar más el tema para cuando vayamos a implementarlo y los datos se vayan a usar. Y habría que explorar si necesitamos algo más que el compromiso confidencial.

Milton: ustedes ya de lo que han visto ya de ejecución de ese cubo anteriormente, sienten que ese modelo ya está un poco viejo, ¿verdad? Según veo con esta nueva iniciativa, ¿verdad? Actualizarlo. Pero a nivel de, digamos, de eficiencia, si sienten que sí detectaba, digamos, si era tener una buena sensibilidad y precisión, digamos, en ese tipo, o sienten que debería tener una buena mejoría, digamos, que para eso está esa nueva, digamos, iniciativa de mejorarlo de cierta forma.

Raquel: Yo no sé si Arce siente incapacidad de responder.

Aylin: Según acuerdo, sí era bastante sensible. Yo honestamente no se lo vi funcionar. Cuando yo llegué acá, el flujo me parecía que no estaba como que ya había perdido ese mantenimiento que se le daba. Pero según escuché en algunas reuniones, estaba bastante bien. Recuerdo, así como un número para decirte, pero sí estaba bastante bien, sí se había probado y demás, o sea, funcionaba bien, el problema era como que no se les dio dando esa como continuidad, porque hay que estar actualizando el modelo cada cierto tiempo, ¿verdad? Y creo que eso fue lo que pasó, que tal vez no se definió alguna fecha, no sé, o sea, la verdad desconozco la razón, y dejó de ser confiable la información porque se estancó en cierta parte del tiempo y algo más, pero según entiendo estaba bastante bien cuando quedó y en el momento en el que funcionaba bien, era bastante bueno y sensible

Milton: entonces a lo que entiendo actualmente no existe tampoco un equipo que le dé mantenimiento directamente, sino que con esta iniciativa actualmente esta iniciativa que tienen ahorita van a actualizarlo pero aun así no hay definido un equipo que le siga dando mantenimiento en el tiempo para poder dejarlo, digamos, en funcionamiento constante.

Raquel: Sí, seríamos nosotros, digamos, porque ya el cubo pasaría a ser parte del SIGI, ¿verdad? Y entonces la idea más bien es que el cubo, digamos, en lo que se pueda, se actualice automáticamente, ¿verdad? O sea, que se alimente con los datos de matrícula y demás, pero que él pueda consultar ya con los datos que ya están en SIGI, ¿verdad? Entonces, sí, en la reunión que tuvimos, de hecho, con los analistas, hablamos de eso, de la fecha y hemos estado, ¿verdad? Les consultamos también a Martín y creemos que la actualización podría ser, ¿verdad? O semestral o anual, ¿verdad? O antes de que, después de que entreguen notas, antes de que se haga la matrícula,

¿verdad? Ojalá. Entonces, sí, sí, ya hemos estado hablando sobre cuál podría ser ese plazo, ¿verdad?, en el que se actualice el modelo periódicamente.

Milton: Entonces no se si ustedes cuando estaban viendo el funcionamiento del modelo, no sé qué tanto lo han visto funcionar, pero no se si a razón personal sienten que debería de tener una priorización, por ejemplo, el cubo detecta que muchos estudiantes están con una alta probabilidad de desertar, si de verdad el TEC como institución tiene la capacidad para lograr abarcar esa cantidad total de casos, o en caso contrario, ¿verdad?, habría que hacer una priorización. No sé si actualmente, digamos, han registrado ese caso, ¿verdad?, porque de verdad, si hay por ejemplo 300 casos de personas que quieren abandonar el TEC, bueno, en este caso la carrera, si de verdad puede ser, le puede dar esa ayuda inmediata, o más bien habría que hacer una priorización para poder abarcar los casos más, digamos, urgentes.

Raquel: Sí, nosotras nunca lo vimos funcionar, entonces no sabemos ¿verdad? Como cuál es el volumen de gente que proyecta, ¿verdad? Pero de todas formas creo que eso es como más operativo, ¿verdad? Más de cómo se vaya a manejar la información y como te digo, creo que es un muy también para el Departamento de Orientación y Psicología y entonces yo pensaría que, a partir de ahí, pues entonces priorizaríamos también, hay más formas, ¿verdad? Incluso en el mismo SIGI dependiendo de si la persona cuándo matriculó, cuántos años tiene, qué cantón es, ¿verdad? Hay formas para también ayudarles a priorizar, pero ya sería más una cuestión de cómo lo vaya a ejecutar las atenciones digamos ese departamento si las va a descentralizar por escuelas o no verdad o no lo vaya a manejar

Milton: ahorita me está hablando mucho el sí no sé si de momento de mi parte desconozco un poco cuál es el rol verdad que en qué interviene, digamos, el SIGI acá. Entonces, no sé si podrían hablarme un poco más también por ese tema de qué es el SII y cómo interviene, digamos, en este proyecto y cuál es el rol como tal, digamos.

Raquel: El SIGI es el Sistema de Indicadores de Gestión Institucional. Entonces, ahí tenemos datos de oferta, de matrícula, de rendimiento académico también. Tenemos de cómo clasifican los estudiantes las carreras o que tan satisfechos se sienten con las carreras como lo tomamos de ciertos estudios que hacemos o encuestas que hacemos a los estudiantes que salen del TEC o cuando se gradúan y verdad y entonces este es un sistema digamos que es como nuestro centro del modelo inteligencia de negocio verdad como que la idea es que en algún momento SIGI pueda consultar digamos obtener la información del resto de sistemas en el TEC para que la información esté centralizada para que la información para la toma de decisiones esté ahí, entonces mucha de la información que creo que toda verdad más bien de la información que alimenta el modelo también está ya en SIGI y como parte de otros cubos verdad que pues para saber cuántos estudiantes matriculan, para saber cuántos estudiantes aprueban, cuántos reprueban. Ese tipo de indicadores los tenemos también en ese sistema y entonces el modelo se alimenta más bien de los datos de ese sistema. Esa es la idea, más bien que todo se esté en SIGI para que no tengamos que alimentar datos de forma manual.

Milton: ¿Ese SIGI igual es administrado por ustedes mismos por la OPI?

Raquel: Es los mismos, esos analistas que te mencionamos Jason y Rebeca son quienes en DATIC le dan mantenimiento y es administrado por OPI

Milton: Ya por último, no sé si podrían, bueno, tenía un poco de preguntas sobre el acceso a esos datos, digamos que ya, por ejemplo, como ya mencioné, yo quiero evaluar si se puede realizar ciertas mejorías a los modelos o qué mejoras se puede implementar a todo este plan que se tiene No sé si se podría de alguna forma acceder a una, que ofrezcan, digamos, una pequeña muestra, digamos, de esos datos que tienen ahorita con el modelo, o también

como, bueno, con el cubo actual y también, no sé si los datos que tiene la OPI. No sé si hay alguna forma que se pueda acceder a eso, ya sea firmando un contrato de confidencialidad, por ejemplo, o algo así.

Raquel: Bueno, los datos que Jason y Rebeca prepararon, pudiéramos tal vez mostrarte tal vez como en alguna llamada, que no estemos grabando, pudiéramos tal vez mostrarte nada más como para que veas la estructura más que eso digamos, si por motivo de tu trabajo final de grabación o algo así, necesitaras más información o verdad ya te puedes involucrar como más en el proyecto y entonces más bien consideraremos como una mejora al modelo, ¿verdad? O algo así. Pues pudiéramos tramitar tal vez algún usuario, ¿verdad? Temporal, que puedas usar mientras la investigación o similar. Pero sí está un poco complicado, ¿verdad? Depende mucho de cómo lo enfoquemos, porque sí, pues son datos super delicados.

Milton: No sería como ver los datos, sino como saber las variables, más que nada, o sea, no hace falta que yo sepa como lo que hay actualmente, sino más como para ver qué variables existen ahorita para poder intentar generar un modelo nuevo, que sería el plan... Entonces creo que ahí tal vez me salga un toque de la confidencialidad porque no tengo que ver datos reales.

Raquel: Sí, porque también con eso pudiéramos pasarte, tenemos las medidas y las dimensiones que son las variables que tiene ese cubo, ¿verdad? Pudiera funcionarle, tal vez, no sé qué piensas, ¿verdad? Voy a buscarlo un momentito, porque ese no tiene nada de datos que pudiéramos...

Milton: Más que nada es como el objetivo, es como ver qué tiene ahorita el TEC, para ver si al final se puede mejorar con buscar otros datos en otros lados o así.

Raquel: Esas son las variables que... Yo no sé qué significaba el morado. Sí, era como de que... No sé, tiene comentarios de Jason, ¿verdad? Pero es que tenemos carné, sexo, no sé, tal me dice si eso te sirve de año de ingreso, modalidad, periodo o grado académico, carrera la que ingresó.

Aylin: A mí me parece que lo que quiere es como qué variables hay dentro del SIGI y no solamente... Sí, pero es que hay unas que se construyen, o sea, que Jason y Rebeca construyen a partir de otras entonces, no sé, yo creo que más bien lo que él quiere es lo que hay dentro de SIGI, en real, ¿no?

Milton: Sí, de hecho, eso también de hecho eso también me sirve, entonces super bien.

Raquel: porque es que el manual no está completo como para pasártelo ese era también, pero pudiéramos pasarte, puede enseñarte entonces este que es un trabajo, igual que está en proceso, por eso los colores y todo. Pero digamos, acá tenemos los cubos que tenemos en el SIGI, tenemos cubos de financieros, pero los estudiantiles, de hecho, son como los académicos, digamos, son los que están en verde. Entonces tenemos deserción regular, deserción primer ingreso, matrícula, nota de corte de admisión, primer ingreso institucional. Entonces, por ejemplo, aquí entras, no sé, a deserción regular y te va a decir un poquito lo que tiene el cubo, ¿verdad? Que son información sobre personas estudiantes desertores regulares. Entonces, tiene el total de estudiantes desertores y el promedio ponderado por crédito. Y esa información la podemos desagregares saber qué era este colegio de procedencia, por cursos, aquí vienen todas esas, estas en este de acá están descritas ¿verdad? Si vos querés saber por ejemplo que era colegio de procedencia porque es una muy buena dimensión, ¿verdad? El nombre del colegio, modalidad, entonces, y aquí vienen los cubos en los que está esa dimensión de hecho también, esto es como para generar un glosario, en realidad es para un manual que estamos haciendo que la idea es que interactivamente cuando selecciones carreras se despliegue, ¿verdad? En un pop-up, en una nube, ¿qué significa? ¿Verdad? Que son los programas académicos, clientes, etc. Hay algunas descripciones que están muy fáciles, ¿verdad? Pero hay otras que tal vez no. Está ese de deserción regular, está ese de deserción de primer ingreso, que tiene la misma

información, pero para primer ingreso. Está el de matrícula, que tiene estudiantes matriculados, créditos, cantidad de grupos, total de matrículas, y se puede también cruzar con, por ejemplo, resultado del curso, ¿verdad? Que, si está aprobado o no, o si está sí, ¿verdad? Si está aprobado o si está reprobado. Las notas de corte y cupo de admisión, ¿verdad? También, ya desde el proceso de admisión, primer ingreso, proceso de admisión para saber ya de los solicitantes, de los examinados, de los elegibles, por carrera también tenemos eso, por carrera también tienes el rendimiento, ya trae desertores, graduados, matriculados, tiene ya como un poco los de rendimiento agrupados, el rendimiento académico, porcentaje de reprobación, ponderado por crédito, el de tiempo de graduación. Tenemos estos dos que es de evaluación de la carrera y condición laboral al graduarse. Podemos tal vez compartir este, para que lo tengas ahí, este Excel. Este sí, si no hay problema. No sé si era como eso lo que querías ver.

Milton: Sí, de hecho, sí me sirve bastante, sí. Creo que también, bueno, para hablar un poco más del tema de datos, si pudiera contactarme tal vez con Jason y Rebeca, me parece.

Raquel: Sí, exactamente. También, si pues es el caso, digamos, sería mejor que te contacte con ellos.

Milton: Ok, perfecto. Sí, es que quería ver un poco más como las variables que se tienen ahorita, digamos, a nivel de datos del TEC, ¿verdad? Que, por ejemplo, que nada más tengan, no sé, tengan la edad, el nombre, el apellido, la cédula y ese tipo de cosas, digamos, de una persona, como para ver ese tipo de temas. Sí, entonces con eso mejor me hablo más con...

Raquel: Sí, o igual aquí viene, ¿verdad? Por ejemplo, aquí vas a tener que la información de estudiante ahí tiene el carné, la cédula, el nombre... No, la cédula creo que no, perdón, pero el carné, el nombre, hay más información que viene del estudiante. Yo creo que vale la pena también que revises un poco porque sí la información que ellos usan es de esta de acá y si tuvieras alguna duda también pues me contactas y si no puedes escribirles sin problemas

Milton: a nivel de deserción la OPI tiene alguna otra iniciativa de momento sólo tienen esta parte del cubo

Raquel: de momento sería si solamente el cubo de deserción de proyección de deserción porque ya tenemos de deserción regular

Milton: de momento si esto sería la entrevista muchísimas gracias por su tiempo, verdad, y sus conocimientos que me han servido muchísimo para mi trabajo final. Posiblemente entre unos días, o no sé si será otra semana, quizá les voy a pasar una minuta, y tal vez una transcripción de lo que yo hice, verdad, como preguntas y respuestas de lo que hablamos un poco acá, para que, por si no fuera mucho molesto, me pudieran regalar una firma, verdad, para agregar. Entonces ahí me lo pueden revisar, y por aquello si anoté algo que no quieren que aparezca, entonces nada más me dicen y lo elimino de ahí

Raquel: claro, claro, no hay problema y ya lo estoy mandando de una vez el correo con estas fichas técnicas y si hay cualquier cosita entonces nos avisas si tienes alguna consulta más.

Milton: Perfecto, muchísimas gracias

Raquel: bueno, con mucho gusto

Milton: buenas tardes, que tengan un hermoso día

Aylin: bueno, hasta luego Adiós.

Apéndice P: Proyectos sugeridos detallados.

Identificador	P001	**Nombre del proyecto**	Análisis y Preparación de Datos para el SPGDE
Descripción			
Este proyecto se enfoca en la identificación, recopilación, limpieza y estructuración de los datos necesarios para alimentar el modelo predictivo de deserción estudiantil (proceso ETL).			
Objetivos del proyecto			
1. Crear un inventario completo de todas las fuentes de datos relevantes del TEC para la predicción de deserción en los primeros 2 meses del proyecto. 2. Desarrollar e implementar procesos de ETL que integren al menos el 95% de las fuentes de datos identificadas en un repositorio centralizado al final del 4º mes. 3. Alcanzar una tasa de completitud de datos del 98% para los registros de estudiantes de los últimos 5 años al finalizar el proyecto. 4. Implementar un sistema de validación de datos que reduzca los errores de entrada en un 80% en comparación con los datos históricos, al término del proyecto.			
Alcance del proyecto			
• Análisis de sistemas de información existentes en el TEC (académicos, financieros, de servicios estudiantiles, etc.) • Diseño de una arquitectura de datos para el SPGDE • Desarrollo de scripts y procesos de ETL • Desarrollo de RPA necesarios. • Implementación de un *data warehouse* para el SPGDE • Documentación exhaustiva de todos los procesos y estructuras de datos			
Justificación del proyecto			
La calidad y disponibilidad de los datos son fundamentales para el éxito del SPGDE. Este proyecto sienta las bases para todos los demás, asegurando que el modelo tenga acceso a datos confiables y relevantes.			
Partes interesadas del proyecto			
• Oficina de Planificación Institucional (OPI) • Departamento de Admisión y Registro (DAR) • Departamento de Orientación y Psicología (DOP) • Departamento de Tecnologías de Información y Comunicaciones (DATIC)			
Entregables del proyecto			

- Catálogo de fuentes de datos del TEC relevantes para el SPGDE
- Documento de arquitectura de datos del SPGDE
- Conjunto de scripts y procesos ETL documentados
- *Data warehouse* implementado y poblado con datos históricos desde 2010
- Manual de procesos de gestión y mantenimiento de datos del SPGDE

Duración aproximada

6 meses (Ver Extracción, Transformación y Carga (ETL))

Presupuesto

₡ 12,693,729.16 (Ver Extracción, Transformación y Carga (ETL))

Recursos necesarios

- 1 científico de datos
- 1 ingenieros de datos

Principales restricciones / limitaciones

- Acceso limitado a ciertas fuentes de datos debido a políticas de privacidad.
- Calidad variable de los datos históricos.
- Limitaciones de tiempo del personal de TI para colaborar en el proyecto.

Riesgos

- Calidad insuficiente de los datos históricos.
- Resistencia de departamentos a compartir datos.
- Complejidad en la integración de sistemas de legado.
- Cambios en las políticas de privacidad de datos durante el proyecto.

Dependencia con otros proyectos

Directas	No aplica.	**Indirectas**	No aplica.

Identificador	P002	**Nombre del proyecto**	Desarrollo del Modelo Predictivo de Deserción Estudiantil

Descripción

Este proyecto se centra en el desarrollo, entrenamiento y validación de un modelo de aprendizaje automático capaz de predecir con alta precisión el riesgo de deserción de los estudiantes del TEC.

Objetivos del proyecto

1. Desarrollar un modelo predictivo con una precisión mínima del 90% en la identificación de estudiantes en riesgo de deserción, validado a través de pruebas de *cross-validation*, al final del 4º mes.
2. Identificar los 10 factores más influyentes en la deserción estudiantil del TEC, con sus respectivos pesos estadísticos, al término del 3er mes.
3. Implementar un sistema de puntuación de riesgo que categorice a los estudiantes en al menos 3 niveles de riesgo (bajo, medio, alto) con una confiabilidad del 90%, al finalizar el proyecto.
4. Crear un flujo de trabajo (*pipeline*) de actualización automática del modelo que se ejecute semanalmente y reentrenar el modelo con nuevos datos cada semestre, implementado al final del proyecto.

Alcance del proyecto

- Análisis exploratorio de datos preparados en el proyecto P001
- Selección y evaluación de algoritmos de aprendizaje automático
- Entrenamiento y validación de múltiples modelos predictivos
- Implementación de técnicas de interpretabilidad del modelo
- Desarrollo de un sistema de actualización y reentrenamiento automático del modelo
- Creación de una API RESTful para la integración del modelo

Justificación del proyecto

El modelo predictivo es el núcleo del SPGDE. Su desarrollo permitirá identificar con anticipación a los estudiantes en riesgo de deserción, facilitando intervenciones tempranas y efectivas.

Partes interesadas del proyecto

- Vicerrectoría de Vida Estudiantil y Servicios Académicos (VIESA)
- Oficina de Planificación Institucional (OPI)
- Escuelas académicas del TEC
- Departamento de Orientación y Psicología (DOP)

Entregables del proyecto

- Informe de análisis exploratorio de datos
- Modelo predictivo de deserción estudiantil implementado y documentado
- Sistema de puntuación de riesgo de deserción
- Flujo de trabajo (pipeline) de actualización automática del modelo
- API documentado para la integración del modelo
- Manual técnico del modelo predictivo

Duración aproximada

4 meses (Ver Caracterización y Selección del Modelo)
Presupuesto
₡ 7,595,819.44 (Ver Caracterización y Selección del Modelo)
Recursos necesarios
• 1 científico de datos especializado en aprendizaje automático • 1 ingeniero de aprendizaje automático
Principales restricciones / limitaciones
• Limitaciones en la capacidad de cómputo para el entrenamiento de modelos complejos • Necesidad de mantener la interpretabilidad del modelo para su aceptación institucional • Requisitos de protección de datos personales de los estudiantes
Riesgos
• Dificultad para alcanzar la precisión deseada del modelo • Problemas de *overfitting* o *underfitting* del modelo • Cambios significativos en los patrones de deserción que afecten la validez del modelo • Desafíos en la interpretabilidad del modelo para los usuarios finales

Dependencia con otros proyectos			
Directas	P001	**Indirectas**	No aplica.

Identificador	P003	**Nombre del proyecto**	Integración del SPGDE con Sistemas Existentes del TEC
Descripción			
Este proyecto se enfoca en la integración del SPGDE con los sistemas existentes del TEC, asegurando un flujo de datos en tiempo real y la incorporación de las predicciones del modelo en los procesos operativos de la institución.			
Objetivos del proyecto			

1. Desarrollar e implementar APIs para la comunicación bidireccional entre el SPGDE y al menos 5 sistemas críticos del TEC (por ejemplo, matrícula, notas, financiero) en los primeros 3 meses del proyecto.
2. Lograr un tiempo de respuesta promedio de menos de 2 segundos para las consultas de datos entre el SPGDE y los sistemas integrados, medido al final del 3º mes.
3. Implementar un sistema de sincronización de datos en tiempo real que actualice la información del SPGDE con una latencia máxima de 15 minutos, para el 90% de las transacciones, al término del proyecto.
4. Alcanzar una tasa de éxito del 99.9% en la transmisión de datos entre sistemas, medida diariamente durante el último mes del proyecto.

Alcance del proyecto

- Análisis detallado de los sistemas existentes y sus arquitecturas.
- Diseño de arquitectura de integración.
- Desarrollo de APIs RESTful para la integración de sistemas.
- Implementación de mecanismos de sincronización de datos.
- Adaptación de sistemas existentes para consumir datos del SPGDE.
- Desarrollo de un sistema de monitoreo y gestión de integraciones.
- Pruebas e implementación del modelo predictivo.

Justificación del proyecto

La integración efectiva es crucial para que el SPGDE tenga un impacto real en las operaciones diarias del TEC. Permitirá que las predicciones y alertas lleguen a los usuarios adecuados de manera oportuna.

Partes interesadas del proyecto

- Departamento de Tecnologías de Información y Comunicaciones (DATIC)
- Oficina de Planificación Institucional (OPI)
- Departamento de Admisión y Registro (DAR)
- Vicerrectoría de Docencia

Entregables del proyecto

- Documento de arquitectura de integración
- APIs implementadas y documentadas
- Sistema de sincronización de datos en tiempo real
- Módulos de integración para sistemas existentes
- Panel de control de monitoreo de integraciones
- Manual de operaciones para la gestión de integraciones

Duración aproximada

6 meses (Ver
Entrenamiento y Pruebas y Puesta en marcha) ambas etapas son fusionadas.

Presupuesto

₡6,646,342.01 (Ver Entrenamiento y Pruebas y Puesta en Marcha) ambas etapas son fusionadas.
Recursos necesarios
• 1 arquitecto de sistemas. • 2 desarrolladores de software.
Principales restricciones / limitaciones
• Complejidad de los sistemas de legado del TEC • Necesidad de mantener la operatividad de los sistemas existentes durante la integración • Limitaciones de recursos del equipo de TI del TEC
Riesgos
• Incompatibilidad con sistemas de legado del TEC • Resistencia al cambio por parte de los usuarios de sistemas existentes • Problemas de rendimiento debido al aumento de tráfico de datos • Desafíos en la sincronización de datos en tiempo real
Dependencia con otros proyectos

Directas	P001 y P002	**Indirectas**	No aplica.

Identificador	P004	**Nombre del proyecto**	Diseño e Implementación de Interfaces de Usuario para el SPGDE
Descripción			
Este proyecto se centra en el diseño y desarrollo de interfaces de usuario intuitivas y eficaces para el SPGDE, facilitando el acceso y la interpretación de las predicciones y recomendaciones generadas por el sistema.			
Objetivos del proyecto			

1. Desarrollar y lanzar un *dashboard* interactivo para al menos 10 perfiles de usuario (administradores, profesores, asesores estudiantiles) con el 100% de las funcionalidades requeridas, al final del 1º mes.
2. Lograr un puntaje de usabilidad de al menos 80/100 en pruebas de usuario con una muestra representativa de 20 usuarios finales, al término del 2º mes.
3. Implementar un sistema de notificaciones que alerte a los usuarios relevantes sobre estudiantes en riesgo dentro de las 24 horas de la detección, operativo al final del proyecto.

Alcance del proyecto

- Análisis de requerimientos de usuario para diferentes perfiles
- Diseño de *wireframes* y prototipos de alta fidelidad
- Desarrollo de interfaces web responsivas
- Implementación de un *dashboard* interactivo para cada escuela con visualizaciones de datos
- Integración de un sistema de generación de informes

Justificación del proyecto

Interfaces bien diseñadas son cruciales para la adopción y uso efectivo del SPGDE por parte del personal del TEC. Esto asegurará que las predicciones del modelo se traduzcan en acciones concretas.

Partes interesadas del proyecto

- Vicerrectoría de Vida Estudiantil y Servicios Académicos (VIESA)
- Departamento de Orientación y Psicología (DOP)
- Escuelas académicas del TEC
- Departamento de Tecnologías de Información y Comunicaciones (DATIC)

Entregables del proyecto

- Documentación de requerimientos de usuario
- Prototipos de interfaz de usuario para web.
- *Dashboard* web interactivo implementado
- Sistema de generación de informes personalizados
- Manual de usuario para todas las interfaces

Duración aproximada

3 meses (Ver Visualización)

Presupuesto

₡2,431,785.575en personal, licencias, Infraestructura, servicios de pruebas de usabilidad y capacitación en UX/UI
Recursos necesarios
• 1 diseñador *dashboards* para visualizadores.
Principales restricciones / limitaciones
• Necesidad de diseñar interfaces accesibles para usuarios con diferentes niveles de habilidad técnica. • Requisitos de seguridad y privacidad en la presentación de datos sensibles. • Limitaciones en los recursos de hardware disponibles para los usuarios finales.
Riesgos
• Dificultad para satisfacer las necesidades de todos los perfiles de usuario. • Desafíos en la visualización efectiva de datos complejos. • Problemas de rendimiento en la carga y actualización de datos en tiempo real. • Curva de aprendizaje para usuarios no técnicos.

Dependencia con otros proyectos			
Directas	P002 y P003	**Indirectas**	No aplica.

Identificador	P005	**Nombre del proyecto**	Implementación de un Sistema de Monitoreo y Evaluación Continua del SPGDE
Descripción			
Este proyecto se centra en el desarrollo e implementación de un sistema integral de monitoreo y evaluación continua para el SPGDE, asegurando su efectividad y relevancia a lo largo del tiempo.			
Objetivos del proyecto			

1. Desarrollar un conjunto de al menos 15 KPIs que cubran todos los aspectos críticos del SPGDE (precisión del modelo, efectividad de intervenciones, adopción del sistema, etc.) en el primer mes del proyecto.
2. Implementar un *dashboard* de monitoreo en tiempo real que actualice los KPIs con una frecuencia mínima diaria, operativo al final del 3er mes.
3. Establecer un proceso de revisión mensual del desempeño del SPGDE basado en los KPIs, con la participación de al menos un representante de cada departamento clave, implementado al término del 4º mes.
4. Lograr una mejora continua del 5% trimestral en al menos el 80% de los KPIs definidos, medida a partir del segundo trimestre de operación completa del sistema.

Alcance del proyecto

- Definición de métricas de desempeño para todos los componentes del SPGDE
- Desarrollo de sistemas de recolección automática de datos de desempeño
- Implementación de un data *warehouse* para almacenar datos históricos de desempeño
- Desarrollo de *dashboards* interactivos para visualización de KPIs
- Creación de procesos automatizados de generación de informes
- Diseño e implementación de un proceso de revisión y ajuste periódico del modelo

Justificación del proyecto

Un sistema de monitoreo y evaluación es esencial para asegurar que el SPGDE cumpla con sus objetivos y se adapte a las cambiantes necesidades del TEC. Permitirá una mejora continúa basada en datos reales de desempeño.

Partes interesadas del proyecto

- Oficina de Planificación Institucional (OPI)
- Vicerrectoría de Vida Estudiantil y Servicios Académicos (VIESA)
- Consejo Institucional
- Departamento de Tecnologías de Información y Comunicaciones (DATIC)

Entregables del proyecto

- Documento de definición de KPIs y métricas de desempeño
- Sistema automatizado de recolección y análisis de datos de desempeño
- Data *warehouse* de datos históricos de desempeño
- *Dashboards* interactivos de monitoreo en tiempo real
- Proceso documentado de revisión y ajuste del modelo predictivo
- Marco de evaluación de impacto a largo plazo
- Informes periódicos de desempeño y recomendaciones de mejora

Duración aproximada

Actividades anuales recurrentes ejecutadas en momentos específicos del año.

Presupuesto

₡ 2,964,616.29 (Ver Mantenimiento y Mejora Continua), solo ejecutando los primeros 6 meses.
Recursos necesarios
• 1 analista de datos. • 1 desarrollador de software.
Principales restricciones / limitaciones
• Necesidad de definir KPIs que capturen todos los aspectos relevantes del desempeño del SPGDE • Limitaciones en la capacidad de procesamiento para análisis en tiempo real • Requisitos de seguridad y privacidad en el manejo de datos de desempeño
Riesgos
• Dificultad para definir KPIs que capturen todos los aspectos relevantes del desempeño del SPGDE • Desafíos técnicos en la integración de datos de múltiples fuentes para el monitoreo • Resistencia a la transparencia que puede generar un sistema de monitoreo detallado • Complejidad en la atribución de impactos a largo plazo específicamente al SPGDE
Dependencia con otros proyectos

Directas	P001, P002, P003 y P004	**Indirectas**	No aplica.

Identificador	P006	**Nombre del proyecto**	Desarrollo de un Marco de Gobernanza de Datos para el SPGDE

Descripción
Este proyecto se centra en el desarrollo e implementación de un sistema integral de monitoreo y evaluación continua para el SPGDE, asegurando su efectividad y relevancia a lo largo del tiempo.
Objetivos del proyecto

1. Crear y aprobar una política integral de gobernanza de datos para el SPGDE, que cubra aspectos de calidad, seguridad, privacidad y ética, en los primeros 2 meses del proyecto.
2. Implementar un sistema de gestión de metadatos que catalogue el 100% de los datos utilizados por el SPGDE, incluyendo su origen, propietario y uso previsto, al final del 4º mes.
3. Establecer un comité de ética de datos con representantes de al menos 5 departamentos clave del TEC, que se reúna mensualmente para revisar el uso de datos del SPGDE, operativo al término del 3er mes.
4. Lograr un cumplimiento del 100% con las regulaciones de protección de datos aplicables, verificado a través de una auditoría externa al final del proyecto.

Alcance del proyecto

- Análisis del estado actual de la gestión de datos en el TEC
- Desarrollo de políticas de gobernanza de datos específicas para el SPGDE
- Diseño de una estructura organizativa para la gobernanza de datos
- Implementación de herramientas de gestión de metadatos y linaje de datos
- Creación de procesos de auditoría y cumplimiento de políticas de datos
- Desarrollo de programas de capacitación en gobernanza de datos

Justificación del proyecto

Este proyecto se enfoca en el desarrollo e implementación de un marco integral de gobernanza de datos para el SPGDE, estableciendo políticas, procesos y estructuras organizativas que aseguren la calidad, integridad, seguridad y uso ético de los datos.

Partes interesadas del proyecto

- Oficina de Planificación Institucional (OPI)
- Vicerrectoría de Vida Estudiantil y Servicios Académicos (VIESA)
- Consejo Institucional
- Departamento de Tecnologías de Información y Comunicaciones (DATIC)

Entregables del proyecto

- Documento de políticas y procedimientos de gobernanza de datos
- Estructura organizativa y descripciones de roles para la gobernanza de datos
- Catálogo de datos centralizado del SPGDE
- Herramienta de gestión de metadatos y linaje de datos implementada
- Guías de uso ético de datos predictivos
- Programa de capacitación en gobernanza de datos
- Informe de evaluación de madurez de gobernanza de datos

Duración aproximada

6 meses

Presupuesto

₡12,000,000 en personal, software, consultoría legal en protección de datos y capacitación en gobernanza de datos			
Recursos necesarios			
• 1 especialista en gobernanza de datos • 1 analista de procesos de negocio • 1 experto en ética de datos • Colaboración con representantes de diferentes departamentos del TEC			
Principales restricciones / limitaciones			
• Complejidad de las regulaciones de protección de datos aplicables • Necesidad de equilibrar la gobernanza con la flexibilidad necesaria para la investigación • Limitaciones de recursos para la implementación y mantenimiento de procesos de gobernanza			
Riesgos			
• Resistencia al cambio en las prácticas actuales de gestión de datos • Desafíos en la implementación de nuevas políticas en una estructura organizativa compleja • Dificultades en la conciliación de las necesidades de diferentes departamentos • Costos potencialmente altos de herramientas especializadas de gobernanza de datos			
Dependencia con otros proyectos			
Directas	P001	**Indirectas**	P002

Apéndice Q: Tabla de variables de predicción de deserción estudiantil encontradas.

Nombre de la Variable	Fuente de Datos	Tipo
Factores Sociodemográficos		
Índice de desarrollo social del distrito	SIGI	Directo
Edad de ingreso	Admisión y Registro	Directo
Sexo	Admisión y Registro	Directo
Nacionalidad	Admisión y Registro	Directo
Residencia igual o distinta en tiempo lectivo y no lectivo	Admisión y Registro	Directo
Dirección de procedencia	Admisión y Registro	Directo
Dirección del colegio de procedencia	Admisión y Registro	Directo
Dirección en tiempo lectivo	Admisión y Registro	Directo
Estado civil	Encuesta estudiantil	Directo

Nombre de la Variable	Fuente de Datos	Tipo
Condición de maternidad o paternidad	Encuesta estudiantil	Directo
Discapacidad permanente	Encuesta estudiantil	Directo
Estudiante de primera generación	Encuesta estudiantil	Directo
Factores Académicos		
Área de estudios	Admisión y Registro	Directo
Sede de la universidad	Admisión y Registro	Directo
Modalidad de estudio (diurno o nocturno)	Admisión y Registro	Directo
Ingreso a primera opción de carrera	Admisión y Registro	Directo
Ingreso a carrera acreditada	Admisión y Registro	Directo
Tipo de colegio de procedencia	Admisión y Registro	Directo
Nota de razonamiento matemático en PAA	Admisión y Registro	Directo
Nota de razonamiento verbal en PAA	Admisión y Registro	Directo
Nota de rendimiento en educación secundaria	Admisión y Registro	Directo
Nota promedio del primer ciclo lectivo	Admisión y Registro	Indirecto
Promedio ponderado	Admisión y Registro	Indirecto
Cantidad de créditos matriculados en el primer ciclo	Admisión y Registro	Directo
Promedio ponderado por crédito	Admisión y Registro	Indirecto
Forma de ingreso	Admisión y Registro	Directo
Grado académico del programa	Admisión y Registro	Directo
Periodos sin matricular	Admisión y Registro	Indirecto
Periodos en el TEC	Admisión y Registro	Indirecto
Carrera de primera opción	Admisión y Registro	Directo
Carrera de segunda opción	Admisión y Registro	Directo
Cantidad de cursos matriculados	Admisión y Registro	Directo
Cantidad de cursos aprobados	Admisión y Registro	Indirecto
Estado del curso (aprobado, reprobado, retirado)	Admisión y Registro	Directo
Nota de admisión	Admisión y Registro	Indirecto
Jornada de estudio	Admisión y Registro	Directo
Realización de cambio de carrera	Admisión y Registro	Directo
Cursó en verano	Admisión y Registro	Directo
Plan de estudio	Admisión y Registro	Directo
Tipo de diploma de secundaria	Admisión y Registro	Directo
Área de conocimiento	Admisión y Registro	Directo
Nota de IV ciclo	Admisión y Registro	Directo
Cantidad de repeticiones (RN)	Admisión y Registro	Indirecto
Curso	Admisión y Registro	Directo
Escuela que imparte el curso	Admisión y Registro	Directo
Nombre del profesor	Admisión y Registro	Directo
Tipo de grupo	Admisión y Registro	Directo
Factores de Admisión y Cupos		
Solicitud de adecuación para PAA	Admisión y Registro	Directo
Exoneración de la inscripción	Admisión y Registro	Directo
Nota de corte admisión abierta	Admisión y Registro	Directo
Nota de corte institucional	Admisión y Registro	Directo

Nombre de la Variable	Fuente de Datos	Tipo
Nota de corte de admisión restringida	Admisión y Registro	Directo
Total de cupos Consejo Institucional	Admisión y Registro	Directo
Total de cupos de admisión abierta	Admisión y Registro	Directo
Total de cupos de admisión restringida	Admisión y Registro	Directo
Crecimiento anual	Admisión y Registro	Indirecto
Opción académica	Admisión y Registro	Directo
Cantidad de solicitantes	Admisión y Registro	Indirecto
Cantidad de examinados	Admisión y Registro	Indirecto
Cantidad de no examinados	Admisión y Registro	Indirecto
Cantidad de elegibles	Admisión y Registro	Indirecto
Cantidad de no elegibles	Admisión y Registro	Indirecto
Cantidad de admitidos - ordinaria	Admisión y Registro	Indirecto
Cantidad de admitidos - extraordinaria	Admisión y Registro	Indirecto
Cantidad de matriculados	Admisión y Registro	Indirecto
Cantidad en espera	Admisión y Registro	Indirecto
Solicitantes primera opción - solicitud	Admisión y Registro	Indirecto
Solicitantes primera opción - examen	Admisión y Registro	Indirecto
Solicitantes segunda opción - solicitud	Admisión y Registro	Indirecto
Solicitantes segunda opción - examen	Admisión y Registro	Indirecto
Admitidos primera opción - ordinaria	Admisión y Registro	Indirecto
Admitidos segunda opción - ordinaria	Admisión y Registro	Indirecto
Admitidos primera opción - extraordinaria	Admisión y Registro	Indirecto
Admitidos segunda opción - extraordinaria	Admisión y Registro	Indirecto
En espera primera opción - ordinaria	Admisión y Registro	Indirecto
En espera segunda opción - ordinaria	Admisión y Registro	Indirecto
En espera primera opción - extraordinaria	Admisión y Registro	Indirecto
En espera segunda opción - extraordinaria	Admisión y Registro	Indirecto
Factores de Graduación		
Graduados en la carrera	Admisión y Registro	Indirecto
Nombre del diploma	Admisión y Registro	Directo
Fecha de graduación	Admisión y Registro	Directo
Periodos ausentes	Admisión y Registro	Indirecto
Graduado en otra carrera	Admisión y Registro	Directo
Factores Económicos		
Financiado con beca TEC	Departamento de Becas	Directo
Obtención de beca	Departamento de Becas	Directo
Obtención de beca Mauricio Campos	Departamento de Becas	Directo
Obtención de financiamiento	Departamento de Becas	Directo
Salario bruto mensual	Encuesta estudiantil	Directo
Subempleo por horas	Encuesta estudiantil	Directo
Factores Laborales		
Relación del trabajo con la carrera	Encuesta estudiantil	Directo
Satisfacción con el trabajo al graduarse	Encuesta estudiantil	Directo
Sector laboral	Encuesta estudiantil	Directo

Nombre de la Variable	Fuente de Datos	Tipo
Condición laboral durante los estudios	Encuesta estudiantil	Directo
Dificultad percibida para emplearse	Encuesta estudiantil	Directo
Factores de Satisfacción y Percepción		
Afinidad con la carrera	Encuesta estudiantil	Directo
Satisfacción con la carrera	Encuesta estudiantil	Directo
Satisfacción con la universidad	Encuesta estudiantil	Directo
Satisfacción con el atributo	Encuesta estudiantil	Directo
Recibió información de la carrera	Encuesta estudiantil	Directo

Nota: Los tipos directos hace referencia a aquellos datos que no se procesaron, mientras que los indirectos implicaron realizar cálculos para su determinación.

Apéndice R: Tabla de variables de predicción de deserción estudiantil sugeridas.

Nombre de la Variable	Fuente de Datos Potencial	Tipo	Justificación
Factores Académicos y de Participación Educativa			
Frecuencia de inicio de sesión	DATIC	Indirecto	La frecuencia con la que un estudiante inicia sesión en la plataforma educativa puede ser un indicador significativo de su nivel de compromiso académico. Una mayor interacción con el sistema digital sugiere un estudiante más involucrado en sus estudios, lo que podría correlacionarse con una menor probabilidad de deserción
Cumplimiento de entrega de tareas en línea	DATIC	Indirecto	El cumplimiento en la entrega de tareas en línea puede reflejar no solo el rendimiento académico, sino también factores psicológicos, emocionales y socioeconómicos subyacentes que influyen en la decisión de un estudiante de persistir o abandonar sus estudios
Modalidad de estudio (presencial/virtual/híbrido)	Admisión y Registro	Directo	La modalidad de estudio elegida por el estudiante puede influir significativamente en su adaptación al entorno universitario. Los estudiantes que optan por modalidades que se ajustan mejor a sus necesidades y estilos de aprendizaje pueden tener una mayor probabilidad de persistir en sus estudios.
Modalidad del curso (sincrónico/asincrónico)	Admisión y Registro	Directo	La flexibilidad o rigidez de los horarios de clase, determinada por la modalidad sincrónica o asincrónica,

Nombre de la Variable	Fuente de Datos Potencial	Tipo	Justificación
			puede afectar la capacidad del estudiante para equilibrar sus estudios con otras responsabilidades. Una mejor adaptación a estos horarios podría resultar en una menor probabilidad de deserción.
Frecuencia de uso de la biblioteca	Sistema de Biblioteca	Indirecto	El uso frecuente de la biblioteca puede ser un indicador del compromiso académico del estudiante y su integración en la vida universitaria, factores que podrían estar relacionados con una menor probabilidad de deserción
Uso de servicio de préstamo de libros de la biblioteca	Sistema de Biblioteca	Indirecto	El uso regular del servicio de préstamo de libros puede indicar un alto nivel de compromiso con los estudios y una buena integración con los recursos académicos de la institución, lo cual podría estar asociado con una menor probabilidad de deserción
Uso de servicio de préstamo de computadoras de la biblioteca	Sistema de Biblioteca	Indirecto	La utilización frecuente de los servicios de préstamo de computadoras podría indicar la necesidad del estudiante de recursos tecnológicos, posiblemente relacionada con su situación económica. Este factor podría influir en su capacidad para completar tareas y, por ende, en su decisión de continuar o abandonar los estudios
Nivel de competencia en habilidades digitales	Encuesta estudiantil	Directo	El nivel de habilidades digitales de un estudiante puede influir en su capacidad para adaptarse a los requisitos tecnológicos de la educación superior moderna, afectando potencialmente su decisión de persistir o desertar
Nivel de estrés académico percibido	Encuesta estudiantil	Directo	El estrés académico experimentado por los estudiantes puede ser un factor significativo en su bienestar general y su capacidad para manejar las demandas de la educación superior, lo cual podría influir en la decisión de continuar o abandonar los estudios
Factores Tecnológicos y de Conectividad			

Nombre de la Variable	Fuente de Datos Potencial	Tipo	Justificación
Problemas de conectividad para clases virtuales	DATIC / Encuesta estudiantil	Directo	Las dificultades técnicas persistentes, como los problemas de conectividad, pueden afectar negativamente la experiencia educativa del estudiante en un entorno virtual, potencialmente aumentando la frustración y el riesgo de deserción
Acceso a internet en el hogar	SUTEL / Encuesta estudiantil	Directo	La disponibilidad de acceso a internet en el hogar puede ser crucial para el éxito académico, especialmente en un entorno educativo cada vez más digitalizado. La falta de este recurso podría dificultar el cumplimiento de las tareas y la participación en clases virtuales, aumentando potencialmente el riesgo de deserción
Factores Familiares y de Antecedentes Educativos			
Tiene familiares estudiando en el TEC	Admisión y Registro	Indirecto	La presencia de familiares estudiando en la misma institución puede proporcionar un sistema de apoyo adicional y una mayor familiaridad con el entorno universitario, factores que podrían contribuir a una menor probabilidad de deserción
Tiene familiares graduados del TEC	Admisión y Registro	Indirecto	Contar con familiares graduados de la institución puede influir positivamente en la persistencia del estudiante, ya que proporciona modelos a seguir y posiblemente un mayor apoyo y comprensión del proceso educativo en el TEC
Nivel educativo de los padres	Ministerio de Educación Pública	Directo	El nivel educativo de los padres puede influir en las aspiraciones académicas del estudiante y en el apoyo que recibe para sus estudios, factores que podrían estar relacionados con la probabilidad de persistencia o deserción
Tiene hijos	TSE	Directo	La responsabilidad de criar hijos mientras se estudia puede presentar desafíos adicionales en términos de tiempo y recursos, potencialmente afectando la capacidad del estudiante para dedicarse plenamente a sus estudios y, por ende, influyendo en su decisión de continuar o desertar

Nombre de la Variable	Fuente de Datos Potencial	Tipo	Justificación
Cantidad de hijos	TSE	Directo	El número de hijos que tiene un estudiante puede influir en la magnitud de las responsabilidades familiares y, por lo tanto, en la cantidad de tiempo y recursos disponibles para dedicar a los estudios, lo cual podría afectar la probabilidad de deserción
Estado civil	TSE	Directo	El estado civil del estudiante puede influir en sus responsabilidades personales y económicas, factores que podrían afectar su capacidad para dedicarse plenamente a los estudios y, por consiguiente, su decisión de persistir o desertar
Defunción familiar reciente	TSE	Directo	La pérdida reciente de un familiar puede causar estrés emocional y posiblemente cambios en la situación económica o las responsabilidades del estudiante, factores que podrían influir significativamente en su capacidad para continuar con sus estudios y, por lo tanto, en su riesgo de deserción
Factores de Integración y Apoyo Institucional			
Participación en actividades extracurriculares (futbol, investigaciones, entre otros)	Vida Estudiantil	Directo	La participación en actividades extracurriculares puede fomentar un sentido de pertenencia a la comunidad universitaria y desarrollar habilidades complementarias, factores que podrían contribuir a una mayor persistencia en los estudios
Uso de servicios de apoyo estudiantil (tutorías regulares/ éxito académico / DOP)	Departamento de Orientación y Psicología / éxito académico / admisión y registro	Directo	La utilización de servicios de apoyo como tutorías o asesoramiento psicológico puede indicar que el estudiante está buscando activamente ayuda para superar desafíos académicos o personales, lo cual podría estar relacionado con una mayor probabilidad de persistencia en los estudios
Experiencias de discriminación o acoso	Oficina de Equidad de Género	Directo	Las experiencias de discriminación o acoso pueden tener un impacto negativo significativo en el bienestar emocional y la sensación de pertenencia del estudiante en el

Nombre de la Variable	Fuente de Datos Potencial	Tipo	Justificación
			entorno universitario, lo cual podría aumentar la probabilidad de deserción
Factores de Salud y Bienestar			
Frecuencia de visitas médicas	Departamento de Salud	Indirecto	Un alto número de visitas médicas podría indicar problemas de salud que interfieren con los estudios del alumno, potencialmente afectando su asistencia, rendimiento académico y, en consecuencia, su decisión de continuar o abandonar los estudios
Condición de aseguramiento	CCSS	Directo	La condición de aseguramiento del estudiante puede influir en su acceso a servicios de salud, lo cual podría afectar su bienestar general y su capacidad para manejar problemas de salud sin interrumpir significativamente sus estudios
Factores Económicos			
Cambios en la situación económica familiar	Encuesta estudiantil	Directo	Las fluctuaciones en la situación económica familiar pueden afectar directamente la capacidad del estudiante para financiar sus estudios y cubrir sus gastos de vida, lo cual podría influir significativamente en su decisión de continuar o abandonar la universidad
Posee vehículo propio	Registro Nacional	Directo	La posesión de un vehículo propio podría facilitar el transporte del estudiante hacia y desde la universidad, reduciendo potencialmente el estrés y el tiempo asociados con el desplazamiento, lo cual podría contribuir a una mejor gestión del tiempo y, por ende, a una menor probabilidad de deserción
Posee vivienda propia	Registro Nacional	Directo	Tener una vivienda propia podría indicar una mayor estabilidad económica familiar, lo que podría resultar en menos preocupaciones financieras para el estudiante y, potencialmente, en una menor probabilidad de deserción por razones económicas
Familia con propiedades	Registro Nacional	Indirecto	La posesión de propiedades por parte de la familia podría ser un indicador de su situación económica, lo cual

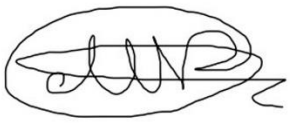
Nombre de la Variable	Fuente de Datos Potencial	Tipo	Justificación
			podría influir en la capacidad del estudiante para financiar sus estudios y, por lo tanto, en su probabilidad de deserción
Ingreso familiar	Ministerio de Hacienda	Indirecto	El nivel de ingreso familiar puede ser un factor determinante en la capacidad del estudiante para cubrir los costos asociados con la educación superior, influyendo potencialmente en su decisión de continuar o abandonar los estudios
Deudas familiares	Sistema Financiero Nacional	Indirecto	Un alto nivel de deudas familiares podría ejercer presión económica sobre el estudiante, potencialmente llevándolo a priorizar el trabajo sobre los estudios o a buscar opciones educativas menos costosas, aumentando así el riesgo de deserción
Historial crediticio familiar	SUGEF	Indirecto	El historial crediticio de la familia podría influir en su capacidad para obtener préstamos educativos o financiar los estudios del estudiante, lo cual podría afectar la capacidad del alumno para continuar con su educación
Becas de otras instituciones	FONABE	Directo	El acceso a becas de otras instituciones puede proporcionar un apoyo financiero adicional, reduciendo la carga económica del estudiante y su familia, lo que podría contribuir a una menor probabilidad de deserción por razones económicas
Participación en programas sociales	IMAS	Directo	La participación en programas sociales podría indicar que el estudiante proviene de un entorno socioeconómico más desafiante, pero, también que está recibiendo apoyo adicional, lo cual podría influir de manera compleja en su probabilidad de deserción
Estado laboral actual	CCSS / Ministerio de Trabajo / Encuesta	Indirecto	El estado laboral del estudiante puede tener un impacto significativo en su capacidad para equilibrar trabajo y estudios. Un estudiante que trabaja podría enfrentar mayores desafíos para dedicar tiempo suficiente a sus

Nombre de la Variable	Fuente de Datos Potencial	Tipo	Justificación
			estudios, lo que podría aumentar el riesgo de deserción. Por otro lado, la experiencia laboral podría proporcionar motivación adicional y habilidades relevantes que podrían contribuir a la persistencia en los estudios
Factores Geográficos y de Contexto Social			
Distancia entre la residencia y TEC	Encuesta estudiantil	Indirecto	Una mayor distancia entre la residencia del estudiante y la universidad podría implicar mayores tiempos de desplazamiento y costos asociados, factores que podrían influir en la decisión del estudiante de continuar o abandonar sus estudios
Tasa de desempleo en cantón de residencia	INEC	Directo	Una alta tasa de desempleo en el cantón de residencia del estudiante podría influir en sus perspectivas de empleo futuro y en la situación económica de su familia, factores que podrían afectar su decisión de continuar o abandonar los estudios
Índice de desarrollo humano del cantón	INEC	Directo	El índice de desarrollo humano del cantón puede ser un indicador de las oportunidades y desafíos generales que enfrenta el estudiante en su entorno, lo cual podría influir en su capacidad para persistir en sus estudios universitarios
Índice de criminalidad del cantón de residencia	INEC	Directo	Un alto índice de criminalidad en el cantón de residencia podría afectar la sensación de seguridad del estudiante y potencialmente su capacidad para concentrarse en sus estudios. Además, podría influir en la decisión de permanecer en la zona para estudiar o buscar oportunidades educativas en áreas percibidas como más seguras, afectando así la probabilidad de deserción
Factores Legales y de Ciudadanía			
Es extranjero	Admisión y Registro	Directo	La condición de extranjero podría presentar desafíos adicionales para el estudiante, como barreras culturales, lingüísticas o administrativas. Estos factores podrían dificultar la

Nombre de la Variable	Fuente de Datos Potencial	Tipo	Justificación
			adaptación al entorno universitario y potencialmente aumentar el riesgo de deserción
Antecedentes penales	Poder Judicial	Directo	Los antecedentes penales de un estudiante podrían influir en sus oportunidades futuras de empleo y en su integración social dentro de la comunidad universitaria, factores que potencialmente podrían afectar su motivación para continuar con los estudios
Participación electoral	TSE	Directo	La participación electoral podría ser un indicador del nivel de compromiso cívico del estudiante, lo cual podría estar relacionado con su sentido general de responsabilidad y compromiso, incluido el ámbito académico
Participación en representación estudiantil	Admisión y Registro	Directo	La participación en representación estudiantil podría indicar un alto nivel de compromiso con la institución y la comunidad universitaria. Este compromiso podría traducirse en un mayor sentido de pertenencia y una red de apoyo más sólida dentro de la universidad, factores que potencialmente podrían reducir la probabilidad de deserción
Pertenece a alguna etnia nacional (indígena)	TSE	Directo	La pertenencia a una etnia nacional, como la indígena, podría presentar desafíos únicos en términos de adaptación cultural, acceso a recursos o posibles barreras lingüísticas. Estos factores podrían influir en la experiencia universitaria del estudiante y, por consiguiente, en su decisión de persistir o desertar. Sin embargo, también podría ser un factor de resiliencia y motivación adicional para completar los estudios y contribuir a su comunidad.

Nota: Los tipos directos hace referencia a aquellos datos que no se procesaron, mientras que los indirectos implicaron realizar cálculos para su determinación.

Apéndice S: Minuta 1 de entrevista a M.Ed. Teresa Hernández Jiménez.

Reunión No.	1	**Fecha**	26/07/2024
Lugar	Oficina del Departamento Académico de Orientación y Psicología del TEC	**Hora Inicio/Fin**	11:00 am / 12:00 pm
Objetivo de la reunión	Entrevista a sujeto de investigación		
Participantes:	Milton Andrey Valverde Paniagua (Entrevistador)		
	M.Ed. María Teresa Hernández Jiménez (Entrevistado)		
Temas tratados			
No.	**Asunto**	**Acuerdos y Comentarios**	
1.	Investigaciones previas sobre deserción	Se discutieron los estudios realizados en el TEC, incluyendo la creación de un algoritmo predictivo y un cubo de información.	
2.	Desafíos en la implementación	Se abordaron los obstáculos éticos, burocráticos y de recursos humanos para implementar el sistema predictivo.	
3.	Manejo de datos institucionales	Se exploraron las fuentes de datos disponibles y las limitaciones en el acceso y uso de la información estudiantil.	
4.	Iniciativas institucionales	Se discutió el Plan Estratégico Institucional y sus metas relacionadas con la reducción de la deserción.	
5.	Propuestas para futuras implementaciones	Se sugirió la integración de un indicador de riesgo de deserción en el expediente estudiantil accesible para personal autorizado.	
Próxima reunión			
Temas a tratar	**Fecha**	**Convocados**	
NA	NA	NA	
Aprobación			
Firma de creador de la minuta		**Firma de involucrado**	
		MARIA TERESA HERNANDEZ JIMENEZ (FIRMA) Firmado digitalmente por MARIA TERESA HERNANDEZ JIMENEZ (FIRMA) Fecha: 2024.10.31 14:03:16 -06'00'	

Apéndice T: Minuta de entrevista al Profesor Martín Solís Salazar.

Reunión No.	2	**Fecha**	31/07/2024
Lugar	Reunión virtual Microsoft Teams	**Hora Inicio/Fin**	08:00 am / 09:00 am
Objetivo de la reunión	Entrevista a sujeto de investigación		
Participantes:	Milton Andrey Valverde Paniagua (Entrevistador)		
	Martín Solís Salazar (Entrevistado)		
Temas tratados			
No.	**Asunto**	**Acuerdos y Comentarios**	
1.	Contexto y rol del estadístico en el proyecto de predicción de deserción estudiantil	Desempeñó el rol principal en el análisis de datos y en el desarrollo del modelo predictivo de deserción estudiantil. Aunque el modelo se implementó en producción y se presentó a las autoridades del TEC, no hubo una directriz institucional clara para usarlo. A pesar de estar técnicamente listo, nunca se le dio seguimiento ni se utilizó de manera efectiva para prevenir la deserción.	
2.	Implementación y desafíos del proyecto	Aunque hubo apertura por parte de la OPI y se logró implementar el modelo, faltó una estrategia institucional para su uso. No se establecieron políticas claras ni se asignaron responsabilidades para darle seguimiento a los estudiantes que el modelo identificaba como posibles desertores. El interés en actualizar el modelo surgió nuevamente en 2023, pero la falta de datos y problemas burocráticos impidieron su avance.	
3.	Calidad de los datos y posibles mejoras	Los datos utilizados en 2018 no eran de la mejor calidad, y algunas variables importantes, como la beca del estudiante, eran estáticas cuando deberían haber sido dinámicas. Además, la pandemia introdujo nuevas variables relevantes, como la interacción con plataformas digitales, que podrían mejorar el modelo, pero requieren un esfuerzo técnico y de recolección de datos más complejo.	
4.	Uso institucional del modelo	El modelo debería estar disponible para todo el TEC, especialmente para directores de escuela, la OPI, y el DOP, quienes podrían usarlo para identificar estudiantes en riesgo de deserción y tomar acciones preventivas. Sin embargo, esto dependería de un interés político y directrices claras que acompañen el uso del modelo.	
5.	Actualización del modelo y futuras investigaciones	Se considera que el modelo debe actualizarse para reflejar los cambios en los patrones de deserción postpandemia y aprovechar nuevas variables. También sugirió pruebas adicionales para predecir no solo si un estudiante desertará, sino cuándo, afinando así el momento de intervención.	
Próxima reunión			
Temas a tratar	**Fecha**	**Convocados**	
NA	NA	NA	
Aprobación			
Firma de creador de la minuta		**Firma de involucrado**	
		MARTIN SOLIS SALAZAR (FIRMA) Digitally signed by MARTIN SOLIS SALAZAR (FIRMA) Date: 2024.10.18 04:12:52 -07'00'	

Apéndice U: Minuta de entrevista a Evaluadora e investigadora jubilada Tania Elena Moreira Mora.

Reunión No.	3	**Fecha**	01/08/2024
Lugar	Reunión virtual Microsoft Teams	**Hora Inicio/Fin**	10:30 am / 11:30 am
Objetivo de la reunión	Entrevista a sujeto de investigación		
Participantes:	Milton Andrey Valverde Paniagua (Entrevistador)		
	Tania Elena Moreira Mora (Entrevistado)		
Temas tratados			
No.	**Asunto**	**Acuerdos y Comentarios**	
1.	Contexto del proyecto y experiencia de la investigadora	Coordinó un estudio sobre deserción estudiantil en el TEC que combinó datos estadísticos con encuestas. El estudio identificó la falta de motivación como la principal causa de deserción. El proyecto tuvo buenos resultados, pero la falta de continuidad y seguimiento del modelo predictivo impidió su aplicación a largo plazo.	
2.	Limitaciones y áreas de mejora del estudio	El principal problema fue la falta de personal para dar seguimiento a los algoritmos. Sería necesario actualizar el modelo con nuevas variables, especialmente tras la pandemia, como el ajuste a la vida universitaria.	
3.	Encuestas y participación estudiantil	Las encuestas a estudiantes se realizaron con éxito, fueron confidenciales y respaldadas por el Departamento de Orientación y Psicología, lo que facilitó la participación y la obtención de datos relevantes.	
4.	Interés institucional y colaboración	Aunque hubo apoyo institucional inicial, no se dio continuidad. Para futuros proyectos, sería clave asegurar un compromiso de las escuelas y departamentos del TEC para implementar acciones basadas en los modelos predictivos.	
5.	Recursos y limitaciones del proyecto	El proyecto contó con recursos limitados, utilizando solo horas de carga laboral de los investigadores. La digitalización de datos ayudó a reducir costos.	
6.	Difusión de resultados y expectativas	Los resultados se presentaron en diversas instancias del TEC y se publicaron en artículos académicos, pero no se les dio el seguimiento necesario para mantener el modelo predictivo actualizado.	
7.	Algoritmos y técnicas utilizadas	Se probaron varios algoritmos, siendo el random forest el más efectivo. Se recomendó seguir actualizando el modelo con nuevas variables y cohortes de datos.	
Próxima reunión			
Temas a tratar	**Fecha**	**Convocados**	
NA	NA	NA	
Aprobación			
Firma de creador de la minuta		**Firma de involucrado**	
		TANIA ELENA MOREIRA MORA (FIRMA) Firmado digitalmente por TANIA ELENA MOREIRA MORA (FIRMA) Fecha: 2024.10.11 15:24:45 -06'00'	

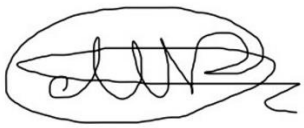
Apéndice V: Minuta de entrevista a Arlyn Angulo Hernández y Raquel Mejías Elizondo de la OPI.

Reunión No.	4	**Fecha**	30/08/2024
Lugar	Reunión virtual Microsoft Teams	**Hora Inicio/Fin**	01:30 pm / 02:30 pm
Objetivo de la reunión	Entrevista a sujetos de investigación		
Participantes:	Milton Andrey Valverde Paniagua (Entrevistador)		
	Aylin Angulo Hernández (Entrevistada)		
	Raquel Mejías Elizondo (Entrevistada)		
Temas tratados			
No.	**Asunto**	**Acuerdos y Comentarios**	
1.	Estado actual del cubo de datos de deserción y su implementación	El cubo de deserción no está actualmente en ejecución. Los analistas están trabajando en actualizar el conjunto de datos, pero hay observaciones pendientes, como hacer que los datos sean dinámicos. Aunque el proyecto sigue en la lista de prioridades, la falta de recursos y el enfoque en otros cubos, como el de becas, ha retrasado su progreso.	
2.	Roles y gestión del proyecto	Además de Martín y los analistas Jason y Rebeca, no hay otros equipos directamente involucrados en el proyecto. Se ha definido que la Vicerrectoría de Servicios Académicos, el DOP, y directores de escuela serán usuarios clave del cubo una vez implementado. Se planea una capacitación para que los responsables de las carreras sepan cómo utilizar el cubo para tomar decisiones.	
3.	Retos técnicos y burocráticos	El principal reto es hacer que los datos sean dinámicos, permitiendo el seguimiento continuo de los estudiantes. Además, los recursos humanos son limitados, con solo dos analistas trabajando en varios proyectos a la vez. Existe un interés institucional para continuar con el cubo, pero la falta de personal especializado ha sido un obstáculo.	
4.	Legalidad y acceso a los datos	Aunque no se han enfrentado a problemas legales específicos con el cubo, se requiere que los usuarios firmen compromisos de confidencialidad antes de acceder a SIGI. También se está explorando si se necesita algún consentimiento adicional para la implementación futura.	
5.	Sensibilidad y precisión del modelo	Aunque el modelo funcionó bien en su momento, perdió confiabilidad debido a la falta de actualización. La nueva iniciativa busca mejorar la precisión y sensibilidad del modelo para reflejar la situación actual de los estudiantes.	
6.	Mantenimiento y actualización futura	La OPI espera que el cubo se actualice automáticamente con datos del SIGI y que se realicen revisiones periódicas, ya sea semestral o anualmente, dependiendo de las necesidades operativas. Sin embargo, aún no hay un equipo dedicado a mantener el modelo a largo plazo.	
7.	Actualización del modelo previo	Actualmente, se está trabajando en la actualización del modelo de predicción de deserción que se dejó de lado en años anteriores. Se está revisando y ajustando los datos para hacerlo más dinámico y confiable. Aunque ya se ha generado un primer conjunto de datos, quedan observaciones por parte de Martín y los analistas, como mejorar la actualización de las variables, para evitar que el modelo quede obsoleto nuevamente.	
Próxima reunión			
Temas a tratar	**Fecha**	**Convocados**	
NA	NA	NA	
Aprobación			
Firma de creador de la minuta		**Firma de involucrado 1**	**Firma de involucrado 2**
		ARLYN ANGULO HERNANDEZ (FIRMA) Firmado digitalmente por ARLYN ANGULO HERNANDEZ (FIRMA) Fecha: 2024.10.14 09:11:02 -06'00'	RAQUEL MEJIAS ELIZONDO (FIRMA) Firmado digitalmente por RAQUEL MEJIAS ELIZONDO (FIRMA) Fecha: 2024.10.15 08:19:15 -06'00'

Apéndice W: Minuta 2 de entrevista a M.Ed. Teresa Hernández Jiménez

Reunión No.	5	**Fecha**	14/10/2024
Lugar	Microsoft Teams.	**Hora Inicio/Fin**	7:30 pm / 8:30 pm
Objetivo de la reunión	Entrevista a sujeto de investigación para verificar posibles variables nuevas a evaluar en futuro modelo predictivo.		
Participantes:	Milton Andrey Valverde Paniagua (Entrevistador)		
	M.Sc. Ing. Juan Andres Segreda Johanning (Entrevistador)		
	M.Ed. María Teresa Hernández Jiménez (Entrevistado)		
Temas tratados			
No.	**Asunto**	**Acuerdos y Comentarios**	
1.	Variables aceptadas	Frecuencia de inicio de sesión, Problemas de conectividad para clases virtuales, Nivel de participación en clases virtuales, Cumplimiento de entrega de tareas en línea, Modalidad de estudio (presencial/virtual/híbrido), Tiene familiares estudiando en el TEC, Tiene familiares graduados del TEC, Participación en actividades extracurriculares, Uso de servicios de apoyo estudiantil, Experiencias de discriminación o acoso, Nivel de competencia en habilidades digitales, Nivel de estrés académico percibido, Cambios en la situación económica familiar, Posee vehículo propio, Posee vivienda propia, Familia con propiedades, Nivel educativo de los padres, Ingreso familiar, Deudas familiares, Historial crediticio familiar, Distancia entre residencia y TEC, Tasa de desempleo en cantón de residencia, Índice de desarrollo humano del cantón, Becas de otras instituciones, Participación en programas sociales, Historial de migración, Antecedentes penales, Participación electoral, Tiene hijos, Cantidad de hijos, Estado civil, Defunción familiar reciente, Condición de aseguramiento, Acceso a internet en el hogar	
2.	Variables a evaluar su fuente debido a su complejidad	Nivel de participación en clases virtuales, Frecuencia de vistas médicas.	
3.	Variables por modificar por ser muy generales	Frecuencia de uso de la biblioteca, Uso de recursos electrónicos de la biblioteca, Historia laboral.	
4.	Variables por considerar	Veces asistidas a tutorías, Modalidad del curso, Uso de servicios de apoyo estudiantil, Participación extracurricular, Índice de criminalidad del cantón de residencia, Participa en representación estudiantil, Pertenece a alguna etnia nacional.	
Próxima reunión			
Temas a tratar	**Fecha**	**Convocados**	
NA	NA	NA	
Aprobación			
Firma de creador de la minuta		**Firma de involucrado**	
		MARIA TERESA HERNANDEZ JIMENEZ (FIRMA) Firmado digitalmente por MARIA TERESA HERNANDEZ JIMENEZ (FIRMA) Fecha: 2024.10.31 14:04:36 -06'00'	

Apéndice X: Minuta de entrevista a M.Sc. Ing. Juan Andrés Segreda Johanning.

Reunión No.	6	**Fecha**	15/10/2024
Lugar	Reunión virtual Microsoft Teams	**Hora Inicio/Fin**	01:00 pm / 02:00 pm
Objetivo de la reunión	Entrevista a sujeto de investigación		
Participantes:	Milton Andrey Valverde Paniagua (Entrevistador)		
	Juan Andrés Segreda Johanning (Entrevistado)		
Temas tratados			
No.	**Asunto**	**Acuerdos y Comentarios**	
1.	¿Cuánto tiempo toma la ejecución de un proceso ETL?	En promedio, un ETL puede durar un total de dos semanas, este tiempo por cada uno de estos. Mientras que, si el caso lo requiere, para la extracción de los datos se necesita de herramientas especializadas como la ejecución de un RPA, este tiempo puede extenderse hasta tres semanas.	
2.	¿Cuánto tiempo puede requerirse para las pruebas y entrenamiento de un modelo de predicción, incluso si utiliza alguna inteligencia artificial?	Esta etapa puede requerir un mínimo de dos meses con dos personas trabajando en este proceso, sin embargo, este tiempo puede variar dependiendo de la complejidad del modelo.	
3.	Presentación de variables de predicción sugeridas.	Se aprobaron todas las variables revisadas.	
4.	Proceso de desarrollo de un modelo de predicción.	Se definieron las etapas de ETL, Caracterización, Entrenamiento y pruebas, Visualización y Puesta en marcha.	
Próxima reunión			
Temas a tratar	**Fecha**	**Convocados**	
NA	NA	NA	
Aprobación			
Firma de creador de la minuta		**Firma de involucrado**	
			

Apéndice Y: Solicitud de información salarial TEC.

Solicitud de rango salarial para TFG

Valorac...arial.xlsx

VALVERDE PANIAGUA MILTON ANDREY
To: talentohumano@tec.ac.cr
Reply | Reply all | Forward
Thu 10/31/2024 3:59 PM

Tengan muy buenas tardes, espero se encuentren muy bien.

Mi nombre es Milton Valverde, actualmente estoy realizando mi TFG en la Escuela de ATI bajo la modalidad de seminario de graduación. Me dirijo hacia ustedes para solicitar, en caso de ser posible, un rango aproximado o promedio salarial de un profesional contratado para el DATIC o la OPI en puestos técnicos especificamente, que incluya cargas sociales (de ser posible indicando de cuanto es las cargas sociales). Esto con el fin de incluir esta información en una valoración financiera de un proyecto potencial a implementar en el TEC, para analizar su viabilidad. De ser o no posible suministrarme esta información, agradeceria si me lo pudieran indicar.

Quedo atento a su respuesta.
Muchas gracias por su tiempo.

Saludos cordiales.

Atte. Milton Andrey Valverde Paniagua.
Escuela de Administración de Tecnologias de Información
miltonvp@estudiantec.cr

Gestión del Talento Humano
Buenas tardes. Dirigimos su consulta con Sofía Brenes Meza para lo correspondiente. _____ María Fernanda Siles Barquero Asistente en Gestión Administrativa Departamento de Gestión del Talento Humano Teléfono: 2550-2228 De: VALVERDE ZjQcmQRYFpfptBannerStart
Thu 10/31/2024 4:26 PM

Diego Solano Montoya<disolano@itcr.ac.cr>
To: VALVERDE PANIAGUA MILTON ANDREY
Cc: Sofía Brenes Meza <sofia.brenes@itcr.ac.cr>
Reply | Reply all | Forward
Fri 11/1/2024 9:40 AM

Valoración Salarial.xlsx
Downloaded

Este mensaje proviene de un remitente externo
Este mensaje vino de fuera de su organización.

Buenos dias don Milton

Para la proyección se utiliza la base salarial del Salario Global Transitorio, ya que es con la que se ingresa hoy en día a la Institución.

Cualquier consulta con mucho gusto

Saludos cordiales,

TEC | Tecnológico de Costa Rica
Diego Solano Montoya
Departamento de Gestión del Talento Humano
Correo/teléfono: disolano@itcr.ac.cr / 2550-9377
Sitio web: www.tec.ac.cr

Reply | Reply all | Forward

Apéndice Z: Solicitud de datos sobre aporte y costo por estudiante al TEC.

Solicitud de entrevista para Trabajo Final de Graduación

VALVERDE PANIAGUA MILTON ANDREY
To: Raquel Mejías Elizondo <rmejias@itcr.ac.cr>
Reply | Reply all | Forward
Thu 10/31/2024 9:46 AM

Buenos dias, espero esté muy bien.

Disculpe que la moleste nuevamente, tengo una pequeña consulta. Tiene conocimiento de si la OPI tiene alguna estadística sobre el costo o ingreso generado por estudiante al TEC? Es con el objetivo de obtener algún aproximado cuantificable para mi TFG.

Muchas gracias por su atención y tiempo.

Saludos cordiales.

Raquel Mejías Elizondo<rmejias@itcr.ac.cr>
To: VALVERDE PANIAGUA MILTON ANDREY
Reply | Reply all | Forward
Fri 11/1/2024 8:54 AM

Este mensaje proviene de un remitente externo
Este mensaje vino de fuera de su organización.

Hola Andrey:

Son dos temas diferentes:

- El costo de un estudiante
- El "aporte"/pago de un estudiante

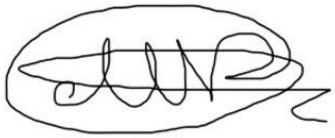
Sobre el pago, hay un tope de 12 créditos, entonces, la mayoría de estudiantes que llevan bloque completo tienen un costo fijo de matrícula.

Sobre el costo, lamentablemente no tenemos en el TEC algún dato, que considere todos los servicios que se le brindan al estudiante y demás.

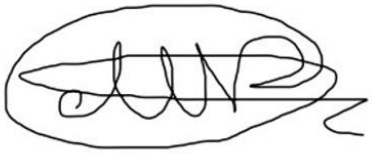
Saludos cordiales.

TEC
Ing. Raquel Mejías Elizondo
Encargada del Área de Gestión de la Información
Oficina de Planificación Institucional
Tecnológico de Costa Rica
(506) 2550-2820
rmejias@tec.ac.cr
Cartago, Costa Rica

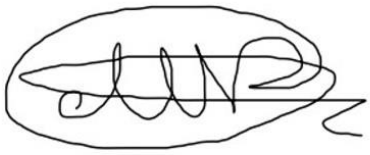
Apéndice AA: Minuta 1 con tutor.

Reunión No.	1	**Fecha**	12/07/2024
Lugar	Reunión virtual Microsoft Teams.	**Hora Inicio/Fin**	06:00 pm / 08:00 pm
Objetivo de la reunión	Definición inicial del proyecto y alcance.		
Participantes:	Milton Andrey Valverde Paniagua (Estudiante)		
	Juan Andrés Segreda Johanning (Profesor tutor)		
Temas tratados			
No.	**Asunto**	**Acuerdos y Comentarios**	
1.	Presentación del tema de investigación.	Se discutió la propuesta inicial del proyecto sobre el modelo predictivo de deserción estudiantil. El profesor mostró interés en el tema y su relevancia institucional.	
2.	Definición del alcance.	Se acordó que el proyecto se enfocará en la evaluación y formulación del modelo, no en su implementación.	
3.	Metodología a utilizar.	Se definió utilizar un enfoque mixto con diseño de investigación-acción.	
4.	Cronograma preliminar.	Se estableció un cronograma tentativo para el semestre, identificando los principales hitos y entregables.	
Próxima reunión			
Temas a tratar	**Fecha**	**Convocados**	
Revisión de objetivos y metodología.	22/07/2024	Milton Valverde, Juan Andrés Segreda.	
Aprobación			
Firma de creador de la minuta		**Firma de involucrado**	
			

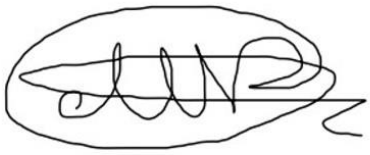
Apéndice BB: Minuta 2 con tutor.

Reunión No.	2	**Fecha**	22/07/2024
Lugar	Reunión virtual Microsoft Teams	**Hora Inicio/Fin**	01:00 pm / 02:00 pm
Objetivo de la reunión	Revisión de objetivos y metodología.		
Participantes:	Milton Andrey Valverde Paniagua (Estudiante)		
	Juan Andrés Segreda Johanning (Profesor tutor)		
Temas tratados			
No.	**Asunto**	**Acuerdos y Comentarios**	
1.	Revisión de objetivos.	Se refinaron los objetivos específicos para asegurar que sean medibles y alcanzables. Se ajustó la redacción para mayor claridad.	
2.	Metodología propuesta.	Se discutió el enfoque mixto y se definieron las herramientas de recolección de datos, incluyendo entrevistas y análisis documental.	
3.	Variables de investigación.	Se identificaron las principales variables a considerar en el estudio y su operacionalización.	
4.	Plan de entrevistas.	Se definió una lista preliminar de sujetos de información clave para las entrevistas.	
Próxima reunión			
Temas a tratar	**Fecha**	**Convocados**	
Revisión del estado del arte.	29/07/2024	Milton Valverde, Juan Andrés Segreda.	
Aprobación			
Firma de creador de la minuta		**Firma de involucrado**	
			

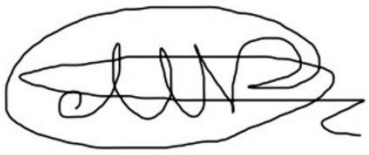
Apéndice CC: Minuta 3 con tutor.

Reunión No.	3	**Fecha**	29/07/2024
Lugar	Reunión virtual Microsoft Teams	**Hora Inicio/Fin**	06:00 pm / 08:00 pm
Objetivo de la reunión	Revisión del estado del arte y antecedentes.		
Participantes:	Milton Andrey Valverde Paniagua (Estudiante)		
	Juan Andrés Segreda Johanning (Profesor tutor)		
Temas tratados			
No.	**Asunto**	**Acuerdos y Comentarios**	
1.	Revisión bibliográfica.	Se analizaron las principales fuentes encontradas sobre modelos predictivos de deserción.	
2.	Investigaciones previas.	Se discutieron los hallazgos del modelo anterior desarrollado en el TEC por Solis et al.	
3.	Metodologías existentes.	Se analizaron diferentes enfoques metodológicos utilizados en estudios similares.	
4.	Marco teórico.	Se definió la estructura del marco teórico y los principales conceptos a incluir.	
Próxima reunion			
Temas a tratar	**Fecha**	**Convocados**	
Revisión de instrumentos de recolección de datos.	30/07/2024	Milton Valverde, Juan Andrés Segreda.	
Aprobación			
Firma de creador de la minuta		**Firma de involucrado**	
			

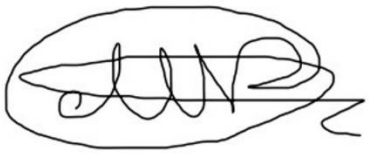
Apéndice DD: Minuta 4 con tutor.

Reunión No.	3	**Fecha**	29/07/2024
Lugar	Reunión virtual Microsoft Teams	**Hora Inicio/Fin**	06:00 pm / 08:00 pm
Objetivo de la reunión	Revisión del estado del arte y antecedentes.		
Participantes:	Milton Andrey Valverde Paniagua (Estudiante)		
	Juan Andrés Segreda Johanning (Profesor tutor)		
Temas tratados			
No.	**Asunto**	**Acuerdos y Comentarios**	
1.	Revisión bibliográfica.	Se analizaron las principales fuentes encontradas sobre modelos predictivos de deserción.	
2.	Investigaciones previas.	Se discutieron los hallazgos del modelo anterior desarrollado en el TEC por Solis et al.	
3.	Metodologías existentes.	Se analizaron diferentes enfoques metodológicos utilizados en estudios similares.	
4.	Marco teórico.	Se definió la estructura del marco teórico y los principales conceptos a incluir.	
Próxima reunion			
Temas a tratar	**Fecha**	**Convocados**	
Revisión de instrumentos de recolección de datos.	30/07/2024	Milton Valverde, Juan Andrés Segreda.	
Aprobación			
Firma de creador de la minuta		**Firma de involucrado**	
			

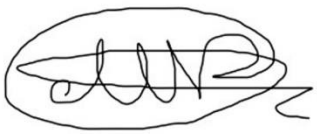
Apéndice EE: Minuta 5 con tutor.

Reunión No.	5	**Fecha**	02/08/2024
Lugar	Reunión virtual Microsoft Teams	**Hora Inicio/Fin**	06:00 pm / 08:00 pm
Objetivo de la reunión	Revisión de primeros hallazgos de las entrevistas.		
Participantes:	Milton Andrey Valverde Paniagua (Estudiante)		
	Juan Andrés Segreda Johanning (Profesor tutor)		
Temas tratados			
No.	**Asunto**	**Acuerdos y Comentarios**	
1.	Resultados preliminares.	Se presentaron los primeros hallazgos de las entrevistas realizadas con el DOP y la OPI.	
2.	Análisis de datos.	Se discutió la metodología para analizar y categorizar la información recolectada.	
3.	Variables identificadas.	Se revisaron las variables de predicción identificadas en las entrevistas.	
4.	Ajustes metodológicos.	Se realizaron ajustes en el enfoque de las próximas entrevistas basados en la experiencia inicial.	
Próxima reunión			
Temas a tratar	**Fecha**	**Convocados**	
Análisis técnico del proyecto.	06/08/2024	Milton Valverde, Juan Andrés Segreda.	
Aprobación			
Firma de creador de la minuta		**Firma de involucrado**	
			

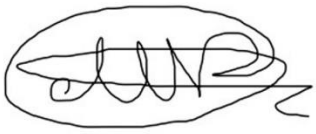
Apéndice FF: Minuta 6 con tutor.

Reunión No.	5	**Fecha**	02/08/2024
Lugar	Reunión virtual Microsoft Teams	**Hora Inicio/Fin**	06:00 pm / 08:00 pm
Objetivo de la reunión	Revisión de primeros hallazgos de las entrevistas.		
Participantes:	Milton Andrey Valverde Paniagua (Estudiante)		
	Juan Andrés Segreda Johanning (Profesor tutor)		
Temas tratados			
No.	**Asunto**	**Acuerdos y Comentarios**	
1.	Resultados preliminares.	Se presentaron los primeros hallazgos de las entrevistas realizadas con el DOP y la OPI.	
2.	Análisis de datos.	Se discutió la metodología para analizar y categorizar la información recolectada.	
3.	Variables identificadas.	Se revisaron las variables de predicción identificadas en las entrevistas.	
4.	Ajustes metodológicos.	Se realizaron ajustes en el enfoque de las próximas entrevistas basados en la experiencia inicial.	
Próxima reunión			
Temas a tratar	**Fecha**	**Convocados**	
Análisis técnico del proyecto.	06/08/2024	Milton Valverde, Juan Andrés Segreda.	
Aprobación			
Firma de creador de la minuta		**Firma de involucrado**	
			

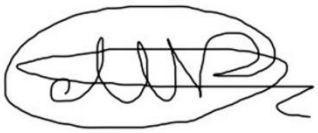
Apéndice GG: Minuta 7 con tutor.

Reunión No.	7	**Fecha**	19/08/2024
Lugar	Reunión virtual Microsoft Teams	**Hora Inicio/Fin**	03:00 pm / 04:00 pm
Objetivo de la reunión	Evaluación financiera y de mercado.		
Participantes:	Milton Andrey Valverde Paniagua (Estudiante).		
	Juan Andrés Segreda Johanning (Profesor tutor).		
Temas tratados			
No.	**Asunto**	**Acuerdos y Comentarios**	
1.	Análisis financiero.	Se revisaron las proyecciones financieras y el análisis de costos del proyecto.	
2.	Estudio de mercado.	Se discutieron los hallazgos del análisis de demanda y oferta en el contexto del TEC.	
3.	Análisis de competencia.	Se evaluó la situación actual con el proyecto de actualización de la OPI.	
4.	Indicadores financieros.	Se revisaron los indicadores financieros y su interpretación en el contexto institucional.	
Próxima reunión			
Temas a tratar	**Fecha**	**Convocados**	
Revisión de plan de implementación.	27/08/2024	Milton Valverde, Juan Andrés Segreda.	
Aprobación			
Firma de creador de la minuta		**Firma de involucrado**	
			

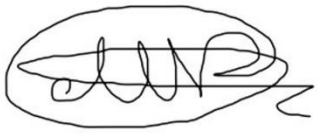
Apéndice HH: Minuta 8 con tutor.

Reunión No.	8	**Fecha**	27/08/2024
Lugar	Reunión virtual Microsoft Teams	**Hora Inicio/Fin**	01:00 pm / 02:00 pm
Objetivo de la reunión	Revisión de plan de implementación.		
Participantes:	Milton Andrey Valverde Paniagua (Estudiante).		
	Juan Andrés Segreda Johanning (Profesor tutor).		
Temas tratados			
No.	**Asunto**	**Acuerdos y Comentarios**	
1.	Proyectos propuestos.	Se revisaron los seis proyectos sugeridos para la implementación del SPGDE.	
2.	Cronograma de implementación.	Se analizó la hoja de ruta propuesta y los plazos establecidos.	
3.	Recursos necesarios.	Se discutieron los recursos requeridos para cada proyecto propuesto.	
4.	Métricas de éxito.	Se definieron los indicadores para medir el éxito de la implementación.	
Próxima reunión			
Temas a tratar	**Fecha**	**Convocados**	
Revisión de conclusiones preliminares.	03/10/2024	Milton Valverde, Juan Andrés Segreda.	
Aprobación			
Firma de creador de la minuta		**Firma de involucrado**	
			

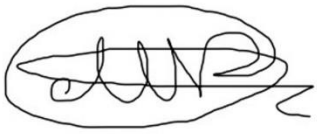
Apéndice II: Minuta 9 con tutor.

Reunión No.	9	**Fecha**	03/10/2024
Lugar	Reunión virtual Microsoft Teams	**Hora Inicio/Fin**	01:00 pm / 02:00 pm
Objetivo de la reunión	Revisión de conclusiones preliminares.		
Participantes:	Milton Andrey Valverde Paniagua (Estudiante).		
	Juan Andrés Segreda Johanning (Profesor tutor).		
Temas tratados			
No.	**Asunto**	**Acuerdos y Comentarios**	
1.	Conclusiones preliminares.	Se revisaron las conclusiones iniciales del estudio de viabilidad.	
2.	Verificación de hipótesis.	Se analizó la comprobación de las hipótesis planteadas.	
3.	Recomendaciones.	Se discutieron las recomendaciones propuestas para futuras investigaciones.	
4.	Limitaciones identificadas.	Se revisaron las limitaciones encontradas durante la investigación.	
Próxima reunión			
Temas a tratar	**Fecha**	**Convocados**	
Revisión de documento final.	15/11/2024	Milton Valverde, Juan Andrés Segreda.	
Aprobación			
Firma de creador de la minuta		**Firma de involucrado**	
			

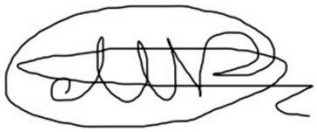
Apéndice JJ: Minuta 10 con tutor.

Reunión No.	10	**Fecha**	15/11/2024
Lugar	Reunión virtual Microsoft Teams	**Hora Inicio/Fin**	01:00 pm / 02:00 pm
Objetivo de la reunión	Revisión de documento final.		
Participantes:	Milton Andrey Valverde Paniagua (Estudiante).		
	Juan Andrés Segreda Johanning (Profesor tutor).		
Temas tratados			
No.	**Asunto**	**Acuerdos y Comentarios**	
1.	Revisión general.	Se realizó una revisión completa del documento, verificando coherencia y formato.	
2.	Ajustes de contenido.	Se identificaron áreas que requieren mayor desarrollo o clarificación.	
3.	Formato del documento.	Se revisó el cumplimiento de los requisitos de formato establecidos.	
4.	Referencias y citas.	Se verificó la correcta citación y referenciación de las fuentes.	
Próxima reunión			
Temas a tratar	**Fecha**	**Convocados**	
Preparación para la defensa.	01/11/2024	Milton Valverde, Juan Andrés Segreda.	
Aprobación			
Firma de creador de la minuta		**Firma de involucrado**	
			

Apéndice KK Minuta 11 con tutor.

Reunión No.	11	**Fecha**	01/11/2024
Lugar	Reunión virtual Microsoft Teams	**Hora Inicio/Fin**	06:00 pm / 08:00 pm
Objetivo de la reunión	Preparación para la defensa.		
Participantes:	Milton Andrey Valverde Paniagua (Estudiante).		
	Juan Andrés Segreda Johanning (Profesor tutor).		
Temas tratados			
No.	**Asunto**	**Acuerdos y Comentarios**	
1.	Estructura de presentación.	Se definió la estructura y puntos clave para la presentación de defensa.	
2.	Material visual.	Se revisó el diseño preliminar de las diapositivas y material de apoyo.	
3.	Tiempo de presentación.	Se establecieron los tiempos para cada sección de la presentación.	
4.	Preguntas potenciales.	Se discutieron posibles preguntas del tribunal y estrategias de respuesta.	
Próxima reunión			
Temas a tratar	**Fecha**	**Convocados**	
Ensayo final de defensa.	07/11/2024	Milton Valverde, Juan Andrés Segreda.	
Aprobación			
Firma de creador de la minuta		**Firma de involucrado**	
			

Apéndice LL: Minuta 12 con tutor.

Reunión No.	12	**Fecha**	07/11/2024
Lugar	Reunión virtual Microsoft Teams	**Hora Inicio/Fin**	02:00 pm / 05:00 pm
Objetivo de la reunión	Ensayo final de defensa.		
Participantes:	Milton Andrey Valverde Paniagua (Estudiante).		
	Juan Andrés Segreda Johanning (Profesor tutor).		
Temas tratados			
No.	**Asunto**	**Acuerdos y Comentarios**	
1.	Simulación de defensa.	Se realizó un ensayo completo de la presentación de defensa.	
2.	Retroalimentación.	Se recibieron observaciones y sugerencias de mejora para la presentación.	
3.	Ajustes finales.	Se identificaron los últimos ajustes necesarios en el material de presentación.	
4.	Aspectos logísticos.	Se revisaron los detalles logísticos para el día de la defensa.	
Próxima reunión			
Temas a tratar	**Fecha**	**Convocados**	
NA	NA	NA	
Aprobación			
Firma de creador de la minuta		**Firma de involucrado**	
			

Capítulo 11. Anexos

Anexo I. Desglose salarial Técnico en Tecnologías de Información y comunicación.

Descripción	Patrono	Empleado
Seguro Social		
Enfermedad y Maternidad	0.00%	5.50%
Banco Popular	0.00%	1.00%
Enfermedad y Maternidad	9.25%	0.00%
Banco Popular	0.50%	0.00%
CCSS Regimen Obligatorio	3.00%	0.00%
CCSSFondo de Capitalización	1.50%	0.00%
Regimen de pensión JPMN		
Capitalización	0.00%	8.00%
Capitalización	6.75%	0.00%
JPMN	0.00%	0.40%
JPMN	0.10%	0.00%
Asociación Solidarista		
ASETEC Ahorro ordinario	0.00%	5.00%
ASTEC Ahorro Patronal	2.00%	0.00%
Provisiones		
Salario Escolar	8.33%	0.00%
Provisión Aguinadldo	8.33%	0.00%
Total	**39.77%**	**19.90%**

PUESTO	Técnico (a) en Tecnologías de Información y Comunicación
BASE SALARIAL MENSUAL	679,332.00
BASE SALARIAL DIARIA	22,644.40
PROYECTADO 1 AÑO	8,151,984.00
COSTO A LA INSTITUCIÓN	**11,393,729.13**
SALARIO EMPLEADO, MENOS CARGAS EMPLEADO	6,529,739.18

Mensual
949,477.43