



Área Académica de Administración de Tecnologías de Información

***Propuesta de solución de minería de datos basada en la metodología CRISP-DM
utilizando los datos almacenados en Oracle Engagement Cloud, para la
recomendación personalizada de productos a clientes, durante el primer
semestre de 2022***

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

Elaborado por: Emanuel Alvarado Pérez

Prof. Tutor: MBA. María José Artavia Jiménez

Cartago, Costa Rica

I Semestre

Junio, 2022



Esta obra está sujeta a la [licencia Reconocimiento-NoComercial-SinObraDerivada 4.0 Internacional de Creative Commons](http://creativecommons.org/licenses/by-nc-nd/4.0). Para ver una copia de esta licencia, visite <http://creativecommons.org/licenses/by-nc-nd/4.0>

ÁREA ACADÉMICA DE ADMINISTRACIÓN DE TECNOLOGÍAS DE INFORMACIÓN

GRADO ACADÉMICO: LICENCIATURA

Los miembros del Tribunal Examinador del Área Académica de Administración de Tecnologías de Información recomendamos que el siguiente Trabajo Final de Graduación del estudiante Emanuel Alvarado Pérez sea aceptado como requisito parcial para optar al grado académico de Licenciatura en Administración de Tecnología de Información.

TEC | Tecnológico
de Costa Rica

Firmado digitalmente por MARIA
JOSE ARTAVIA JIMENEZ (FIRMA)
Fecha: 2022.06.20 09:18:42
-06'00'

MAE. María José Artavia Jiménez

Profesora Tutora

LUIS ALEXANDER
CALVO
VALVERDE
(FIRMA)

Firmado digitalmente por
LUIS ALEXANDER CALVO
VALVERDE (FIRMA)
Fecha: 2022.06.20
10:22:03 -06'00'

Dr. Luis Alexander Calvo Valverde

Lector de Industria

LORENA
ZUÑIGA SEGURA
(FIRMA)

Digitally signed by LORENA
ZUÑIGA SEGURA (FIRMA)
Date: 2022.06.20 17:23:01
-06'00'

MBA. Lorena Zúñiga Segura

Lectora académica

TEC | Tecnológico
de Costa Rica

Firmado digitalmente
por YARIMA TATIANA
SANDOVAL SANCHEZ
(FIRMA)
Fecha: 2022.06.21
11:33:54 -06'00'

Yarima Sandoval Sánchez

Coordinadora Trabajo Final de Graduación

Dedictoria

A mi familia, mi mamá, mi papá, mi hermana, Mari, Jose por todo su apoyo durante estos años para alcanzar mis metas y lograr concluir mis estudios. Quienes siempre confiaron en mis capacidades y me motivaron a nunca rendirme. Gracias por acompañarme durante este largo camino.

Agradecimientos

A todo el equipo de XUM TECH Alex, Zimri, Sally, Su, Erin que me brindaron la oportunidad de desarrollar el proyecto y me apoyaron durante mi proceso de TFG dándome excelentes consejos para generar los mejores resultados,

A la profesora María José Artavia, que me guió para elaborar un proyecto de calidad y siempre tuvo la disposición de ayudarme para salir adelante y sentirme preparado.

A mis amigos María Jesús, Josué, Nathalie, Manuel, quienes fueron una excelente compañía durante el proceso de TFG y estuvieron apoyándome en las buenas y malas. También a todos mis compañeros y profesores con quienes tuve la oportunidad de compartir y crecer a lo largo de la carrera para ser mejor persona y profesional.

Resumen

Alvarado, Emanuel. (2022). Propuesta de solución de minería de datos basada en la metodología CRISP-DM utilizando los datos almacenados en Oracle Engagement Cloud, para la recomendación personalizada de productos a clientes, durante el primer semestre de 2022. (Trabajo Final de Graduación). Área Académica de Administración de Tecnologías de Información. Instituto Tecnológico de Costa Rica.

Esta investigación tiene como propósito proponer una solución de minería de datos con la información almacenada en la herramienta de Oracle Engagement Cloud, para la recomendación personalizada de productos a clientes mediante campañas de mercadeo.

El estudio tiene un enfoque cualitativo. Para recopilar los datos, se realizaron entrevistas y revisiones literarias a fuentes primarias de información, relacionadas a la problemática. Mediante estos instrumentos, se identificaron los datos almacenados en Oracle Engagement Cloud, los requerimientos necesarios para la propuesta, la situación actual y deseada, así como las herramientas utilizadas previamente y las nuevas. Además, se emplearon pruebas y análisis comparativos para determinar el modelo de minería de datos más adecuado, con base en la información disponible y en los requerimientos recolectados.

La investigación concluyó que es necesario crear un reporte de extracción periódica de datos, elaborar un modelo de minería de datos usando el modelo *Collaborative Filtering* en la herramienta Anaconda, donde es posible la limpieza y procesamiento de datos, y crear el archivo de carga para garantizar que sea totalmente compatible con la herramienta. Asimismo, se determinó que dentro de la herramienta de Engagement es necesaria la configuración de elementos como objetos *custom*, pantallas y campos. Se recomienda a XUM TECH utilizar el modelo de recomendación personalizada de productos propuesto, con el objetivo de aumentar el catálogo de servicios ofrecidos a sus clientes.

Palabras clave: sistemas de información, aprendizaje automático, minería de datos, campañas de mercadeo

Abstract

Alvarado, Emanuel. (2022). Proposal for a data mining solution based on the CRISP-DM methodology using the data stored in Oracle Engagement Cloud, for the personalized recommendation of products to clients, during the first semester of 2022. (Final Graduation Project). Academic Area of Information Technology Administration. Technological Institute of Costa Rica.

The purpose of this research is to propose a data mining solution with the information stored in the Oracle Engagement Cloud tool for the personalized recommendation of products to customers through marketing campaigns.

The study has a qualitative approach. To collect qualitative data, interviews and literary reviews were conducted to primary sources of information related to the problem. Through these instruments it was possible to identify the data stored in Oracle Engagement Cloud, the necessary requirements for the proposal, the current and desired situation, the tools previously used and the new ones. Additionally, comparative tests and analyzes were used to determine the most appropriate data mining model, based on the available data and the requirements collected.

The investigation concluded that it is necessary to create a periodic data extraction report, develop a data mining model using the Collaborative Filtering model in the Anaconda tool, where data cleaning and processing is possible, and create the load file to guarantee that is fully compatible with the tool. Likewise, it was determined that within the Engagement tool it is necessary to configure elements such as custom objects, screens, and fields.

XUM TECH is recommended to use the proposed personalized product recommendation model to increase the catalog of services offered to its customers.

Keywords: Information systems, Machine learning, Data mining, Marketing campaigns.

Tabla de contenidos

1. Introducción.....	1
1.1. Descripción general.....	1
1.2. Antecedentes	2
1.2.1. Descripción de la organización.....	2
1.2.1.1. Misión.....	3
1.2.1.2. Visión	3
1.2.1.3. Valores	4
1.2.1.4. Equipo de trabajo	4
1.2.2. Proyectos similares	6
1.3. Planteamiento del problema.....	9
1.3.1. Situación problemática	9
1.3.2. Justificación del proyecto	12
1.3.3. Beneficios esperados del Trabajo Final de Graduación	14
1.4. Objetivos.....	16
1.4.1. Objetivo general	16
1.4.2. Objetivos específicos.....	16
1.5. Alcance del proyecto.....	17
1.6. Supuestos	21
1.7. Entregables.....	22
1.7.1. Entregables del producto	22
1.7.2. Gestión del proyecto.....	23

1.8.	Limitaciones del proyecto.....	27
1.9.	Exclusiones del proyecto	27
2.	Marco conceptual	29
2.1.	Sistema de Información.....	29
2.1.1.	Sistema de apoyo a las decisiones	29
2.1.2.	Customer Relationship Management	30
2.1.3.	Oracle Engagement Cloud	30
2.1.4.	Oracle Eloqua.....	31
2.2.	Comercio electrónico	32
2.2.1.	Tipos de transacciones	33
2.2.2.	Tipos de comercio electrónico	34
2.3.	Analítica empresarial.....	35
2.3.1.	Analítica descriptiva	36
2.3.2.	Analítica predictiva.....	37
2.4.	Modelos de recomendación de productos	38
2.4.1.	Demographic Filtering	38
2.4.2.	Collaborative Filtering	39
2.4.3.	Content Based Filtering	40
2.4.4.	Técnicas de evaluación de modelos de recomendación.....	41
2.5.	Aprendizaje automático.....	42
2.5.1.	K Means Clustering	43
2.5.2.	Naive Bayes	44
2.6.	Técnicas de evaluación de modelos de clasificación	45

2.7.	Técnicas de limpieza de datos	46
2.7.1.	One Hot Encoding.....	46
2.8.	Minería de datos	46
2.8.1.	Metodología CRISP-DM	46
2.8.1.1.	Fase I Entendimiento del negocio	47
2.8.1.2.	Fase II Entendimiento de los datos	48
2.8.1.3.	Fase III Preparación de los datos.....	49
2.8.1.4.	Fase IV Modelado	50
2.8.1.5.	Fase V Evaluación.....	51
2.8.1.6.	Fase VI Despliegue	52
2.9.	Ingeniería de requerimientos	53
2.9.1.	Clasificación de requerimientos.....	53
2.9.2.	Especificación de requerimientos	53
2.10.	Gestión de procesos	54
2.10.1.	Procesos AS IS	54
2.10.2.	Procesos TO BE	54
2.10.3.	Modelado de los procesos de negocio.....	55
2.11.	Cuadrante Mágico de Gartner.....	55
2.11.1.	Magic Quadrant for Cloud Database Management Systems	56
2.11.2.	Magic Quadrant for Data Science and Machine Learning (DSML) platforms...	57
3.	Marco metodológico.....	60
3.1.	Tipo de investigación	60
3.2.	Enfoque de la investigación	61

3.3.	Alcance de la investigación.....	63
3.4.	Diseño de la investigación.....	65
3.5.	Fuentes de investigación	66
3.6.	Sujetos de investigación.....	68
3.7.	Variables de la investigación.....	70
3.8.	Instrumentos de investigación.....	71
3.9.	Matriz de cobertura de variables vs. diseño de los instrumentos.....	77
3.10.	Procedimiento metodológico de la investigación.....	78
3.10.1.	Identificación de las necesidades del negocio	78
3.10.2.	Desarrollo de los modelos predictivos.....	79
3.10.3.	Sugerencia de herramientas para la propuesta.....	81
3.10.4.	Análisis financiero de propuesta.....	81
3.10.5.	Adaptación de la metodología CRISP-DM.....	83
3.11.	Operacionalización de variables.....	85
3.12.	Tabla resumen del procedimiento metodológico de la investigación.....	87
4.	Análisis de resultados	89
4.1.	Fase I. Identificación de las necesidades del negocio	89
4.1.1.	Identificación del proceso actual de recomendación de productos.....	89
4.1.2.	Identificación de requerimientos de la propuesta	93
4.1.3.	Definición de criterios de evaluación para los modelos de minería	97
4.2.	Fase II. Desarrollo de los modelos predictivos	97
4.2.1.	Revisión de datos disponibles en CX.....	97
4.2.2.	Investigación de modelos predictivos para la recomendación personalizada de productos	102

4.2.3.	Evaluación de los modelos de minería.....	104
4.3.	Fase III. Sugerencia de herramientas para la propuesta.....	108
4.3.1.	Revisión de las herramientas utilizadas anteriormente.....	108
4.3.2.	Investigación de nuevas herramientas para la propuesta.....	110
4.3.2.1.	Herramientas de extracción de datos.....	110
4.3.2.2.	Herramientas de almacenamiento de datos.....	113
4.3.2.3.	Herramientas de minería de datos.....	116
4.3.2.4.	Herramientas de mercadeo.....	118
4.4.	Fase IV. Recolección de insumos para el análisis financiero de la propuesta.....	119
4.4.1.	Identificación de costos asociados.....	119
5.	Propuesta de solución.....	121
5.1.	Procesos TO BE.....	121
5.1.1.	Creación de recomendaciones.....	121
5.1.2.	Envío de campañas personalizadas.....	122
5.2.	Módulo de recomendación de productos.....	123
5.2.1.	Extracción de datos de Oracle Engagement Cloud.....	123
5.2.2.	Creación del modelo de recomendación de productos.....	127
5.2.3.	Creación de objetos y campos.....	132
5.2.4.	Configuración de archivos de importación.....	136
5.2.5.	Visualización de recomendaciones.....	138
5.2.6.	Conversión de recomendación a Lead.....	140
5.3.	Creación de campañas con Eloqua.....	143
5.3.1.	Sincronización de los datos.....	143

5.3.2.	Envío de campañas publicitarias para la recomendación de productos	144
5.4.	Análisis financiero de la propuesta	146
5.4.1.	Cotización del proyecto académico	146
5.4.2.	Cotización de la propuesta	147
5.4.3.	Indicadores financieros	149
5.4.4.	Hoja de ruta de la implementación	151
5.4.5.	Beneficios financieros y no financieros	152
6.	Conclusiones.....	154
6.1.	Conclusiones para el objetivo específico uno	154
6.2.	Conclusiones para el objetivo específico dos.....	154
6.3.	Conclusiones para el objetivo específico tres	155
6.4.	Conclusiones para el objetivo específico cuatro	156
7.	Recomendaciones	157
8.	Referencias bibliográficas	160
9.	Apéndices	166
9.1.	Apéndice A. Plantilla de minutas para reunión.....	166
9.2.	Apéndice B. Plantilla para la gestión de cambios	167
9.3.	Apéndice C. Entrevista inicial TFG	168
9.4.	Apéndice D. Resultados de la Entrevista inicial	169
9.5.	Apéndice E. Entrevista #1 Proceso de recomendación/venta de productos en CX	170
9.6.	Apéndice F. Entrevista #2 Conocer los requerimientos/necesidades del negocio	171
9.7.	Apéndice G. Entrevista #3 Verificar lista de requerimientos encontrados	171
9.8.	Apéndice H. Entrevista #4 Entendimiento de los datos, conocer donde se encuentran alojados en Oracle	172

9.9.	Apéndice I. Entrevista #5 Conocer cuáles herramientas han utilizado para este propósito anteriormente.....	173
9.10.	Apéndice J. Entrevista #6 Conocer aspectos a tomar en cuenta durante una cotización de proyectos en XUM TECH.....	174
9.11.	Apéndice K. Plantilla especificación de requerimientos	175
9.12.	Apéndice L. Observación datos disponibles en Oracle Engagement Cloud	175
9.13.	Apéndice M. Revisión documental proyectos de minería de datos para la recomendación de productos	176
9.14.	Apéndice N. Revisión documental herramientas de minería de datos	176
9.15.	Apéndice O. Tabla comparativa de herramientas.....	176
9.16.	Apéndice P. Plantilla de pruebas	176
9.17.	Apéndice Q. Tabla comparativa de modelos.....	177
9.18.	Apéndice R. Resultados de la Entrevista #1	177
9.19.	Apéndice S. Resultados de la Entrevista #3	179
9.20.	Apéndice T. Resultados de la Entrevista #4.....	180
9.21.	Apéndice U. Resultados de la Entrevista #5.....	182
9.22.	Apéndice V. Resultados de la Entrevista #6.....	184
9.23.	Apéndice W. Resultados de la revisión de datos disponibles.....	187
9.24.	Apéndice X. Resultados de la revisión documental de proyectos de minería de datos para la recomendación de productos	188
9.25.	Apéndice Y. Revisión documental de herramientas de almacenamiento AWS.....	192
9.26.	Apéndice Z. Revisión documental de herramientas de almacenamiento Oracle	194
9.27.	Apéndice AA. Revisión documental de herramientas de almacenamiento Microsoft	
	196	

9.28.	Apéndice AB. Revisión documental de herramientas de minería de datos IBM	198
9.29.	Apéndice AC. Revisión documental de herramientas de minería de datos Amazon Sage Maker.....	200
9.30.	Apéndice AD. Revisión documental de herramientas de minería de datos Anaconda	201
9.31.	Apéndice AE. Evaluación del modelo Demographic Filtering	203
9.32.	Apéndice AF. Evaluación del modelo KMeans	204
9.33.	Apéndice AF. Evaluación del modelo Collaborative Filtering	209
9.34.	Apéndice AG. Evaluación del modelo Naive Bayes (Gaussian).....	211
9.35.	Apéndice AH. Evaluación del modelo Naive Bayes (Bernoulli)	215
9.36.	Apéndice AI. Configuración de campo Fecha de creación	219
9.37.	Apéndice AJ. Configuración de campo Cliente asociado	219
9.38.	Apéndice AK. Configuración de campo Correo cliente.....	219
9.39.	Apéndice AL. Configuración de campo Producto.....	220
9.40.	Apéndice AM. Configuración de campo Puntuación.....	220
9.41.	Apéndice AO. Configuración de campo Estado.....	221
9.42.	Apéndice AP. Configuración de campo Encargado	221
9.43.	Apéndice AQ. Configuración de campo Motivo de rechazo.....	222
9.44.	Apéndice AR. Configuración de campo Comentarios adicionales	222
9.45.	Apéndice AS. Configuración de lista de valores Estado	223
9.46.	Apéndice AT. Configuración de lista de valores Motivo de rechazo.....	223
9.47.	Apéndice AU. Configuración de server script Crear Lead.....	224
9.48.	Apéndice AV. Configuración de server script Asignar Correo.....	224
9.49.	Apéndice AW. Configuración de action Crear Lead.....	225

9.50.	Apéndice AX. Transformación #1 Funciones creadas para generar las recomendaciones	225
9.51.	Apéndice AY. Transformación #2 Creación de nuevos campos para almacenar la recomendación	226
9.52.	Apéndice AZ. Transformación #3 Creación de nuevos dataframes	226
9.53.	Apéndice BA. Transformación #4 Cambio de nombre de columnas	226
9.54.	Apéndice BB. Transformación #5 Unión de dataframes de recomendaciones	227
9.55.	Apéndice BC. Transformación #6 Creación de funciones para la limpieza de datos 227	
9.56.	Apéndice BD. Transformación #7 Creación de nuevas columnas en dataframe	228
9.57.	Apéndice BE. Transformación #8 Filtrado del dataframe.....	228
9.58.	Apéndice BF. Transformación #9 Exportación del dataframe a formato CSV	228
9.59.	Apéndice BG. Entrevista #7 Almacenamiento de datos en Engagement Cloud	228
9.60.	Apéndice BH. Resultados de la Entrevista #7	229
9.61.	Apéndice BI. Revisión documental de herramientas de almacenamiento Oracle CX Custom Objects	230
9.62.	Apéndice BJ. Minuta 001. Reunión inicial.....	232
9.63.	Apéndice BK. Minuta 002. Sesión inicial de TFG con la organización	234
9.64.	Apéndice BL. Minuta 003. Sesión semanal de seguimiento	235
9.65.	Apéndice BM. Minuta 004. Sesión acuerdos iniciales con la organización.....	236
9.66.	Apéndice BN. Minuta 005. Sesión acuerdos iniciales con la organización	237
9.67.	Apéndice BO. Minuta 006. Sesión semanal de seguimiento.....	238
9.68.	Apéndice BP. Minuta 007. Sesión semanal de seguimiento	239
9.69.	Apéndice BQ. Minuta 008. Sesión semanal de seguimiento.....	240

9.70.	Apéndice BR. Minuta 009. Sesión semanal de seguimiento	241
9.71.	Apéndice BS. Minuta 010. Sesión semanal de seguimiento	242
9.72.	Apéndice BT. Minuta 011. Segunda sesión de seguimiento de TFG con la organización	243
9.73.	Apéndice BJ. Minuta 012. Sesión semanal de seguimiento	244
9.74.	Apéndice BU. Minuta 013. Sesión semanal de seguimiento.....	245
9.75.	Apéndice BV. Minuta 014. Sesión semanal de seguimiento.....	246
9.76.	Apéndice BW. Minuta 015. Sesión de seguimiento con la coordinación de TFG ...	247
9.77.	Apéndice BX. Minuta 016. Sesión semanal de seguimiento.....	248
9.78.	Apéndice BY. Minuta 017. Sesión semanal de seguimiento.....	249
9.79.	Apéndice BZ. Minuta 018. Presentación de la propuesta a la organización	250
9.80.	Apéndice CA. Minuta 019. Sesión de cierre con la organización	251
9.81.	Apéndice CB. Cálculo de cargas patronales de la investigación.....	252
9.82.	Apéndice CC. Cálculo de cargas patronales de la cotización.....	252
10.	Anexos	253
10.1.	Anexo I. Historias de usuario	253
10.2.	Anexo II. Cotización de database Oracle	256
10.3.	Anexo III. Cotización de almacenamiento en AWS.....	256
10.4.	Anexo IV. Cotización de almacenamiento en Microsoft	257
10.5.	Anexo V. Cotización de herramienta IBM Watson Studio	258
10.6.	Anexo VI. Cotización de herramienta AWS Sage Maker	258
10.7.	Anexo VII. Versiones disponibles de la herramienta Anaconda.....	259
10.8.	Anexo VIII. Documentación interna de Eloqua	260
10.9.	Anexo IX. Notación BPMN	261

10.10.	Anexo X. Evaluaciones de la empresa	262
10.11.	Anexo XI. Reporte de minutas firmado.....	264
10.12.	Anexo XII. Carta de la filóloga	265

Índice de Figuras

Figura 1. Organigrama	6
Figura 2. Árbol del problema.....	11
Figura 3. Subtipos de analítica.....	35
Figura 4. Collaborative Filtering	39
Figura 5. Content Based Filtering.....	40
Figura 6. Similitud de coseno	41
Figura 7. Distancia Euclideana	41
Figura 8. PCC.....	42
Figura 9. Método del codo	44
Figura 10. Fórmula Naive Bayes	44
Figura 11. Ciclo de metodología CRISP DM	47
Figura 12. Fases del procedimiento metodológico	78
Figura 13. Proceso AS IS de ventas con Engagement Cloud	92
Figura 14. Revisión de datos existentes en Oracle Engagement Cloud.....	100
Figura 15. Resultado de la exportación de datos	101
Figura 16. Técnicas de minería de datos más utilizados.....	102
Figura 17. Lenguajes de programación más utilizados.....	103
Figura 18. Gartner Magic Quadrant for Cloud Database Management Systems 2020	113
Figura 19. Magic Quadrant for Data Science and Machine Learning Platforms 2021.....	116
Figura 20. Magic Quadrant for CRM Lead Management 2020.....	118
Figura 21. Proceso TO BE de creación de recomendaciones	122
Figura 22. Proceso TO BE Campañas de mercadeo	123
Figura 23. Reporte de extracción creado	125
Figura 24. Filtros aplicados al reporte de extracción.....	126
Figura 25. CSV de recomendaciones generado	132
Figura 26. Custom object de Recomendación	133
Figura 27. Campos configurados dentro de Engagment	135
Figura 28. Mapeo de importación.....	137

Figura 29. Plantilla de carga de recomendaciones.....	137
Figura 30. Visualización del objeto custom Recomendación.....	138
Figura 31. Lista de recomendaciones cargadas.....	138
Figura 32. Configuración de pantalla de detalles.....	139
Figura 33. Configuración de expresiones para campos de rechazo	139
Figura 34. Detalle de la recomendación	140
Figura 35. Detalles de rechazo de la recomendación.....	140
Figura 36. Botón de creación de lead.....	142
Figura 37. Evidencia del Lead creado.....	143
Figura 38. Cotización del proyecto académico.....	147
Figura 39. Cotización de la implementación	148
Figura 40. Cotización de ganancias de la implementación.....	149
Figura 41. Fórmula de cálculo del ROI	150
Figura 42. Cálculo del ROI.....	150
Figura 43. Propuesta de hoja de ruta.....	151

Índice de Tablas

Tabla 1. Equipo de trabajo	5
Tabla 2. Proyectos externos similares.....	7
Tabla 3. Beneficios esperados.....	14
Tabla 4. Fases asociadas a objetivos.....	19
Tabla 5. Entregables asociados	22
Tabla 6. Cronograma del proyecto.....	23
Tabla 7. Fuentes de información primaria	66
Tabla 8. Fuentes de información secundarias	67
Tabla 9. Sujetos de investigación.....	68
Tabla 10. Variables de la Investigación	70
Tabla 11. Matriz de entrevistas.....	72
Tabla 12. Matriz de revisión documental	74
Tabla 13. Matriz de cobertura de variables.....	77
Tabla 14. Adaptación de la metodología CRISP DM.....	83
Tabla 15. Operacionalización de las variables.....	85
Tabla 16. Tabla resumen del proceso metodológico	88
Tabla 17. Participantes del proceso	89
Tabla 18. Actividades del proceso	90
Tabla 19. Resumen de requerimientos.....	97
Tabla 20. Subject Areas de información en Oracle Engagement Cloud.....	98
Tabla 21. Resultados de la evaluación de modelos de minería de datos	104
Tabla 22. Resultados de la evaluación de modelos de clasificación.....	106
Tabla 23. Cumplimiento de requerimientos por modelo	107
Tabla 24. Herramientas utilizadas previamente.....	108
Tabla 25. Análisis Oracle Report and Analytics.....	110
Tabla 26. Análisis Oracle Export Management.....	111
Tabla 27. Criterios de evaluación para herramientas de extracción de datos	112
Tabla 28. Comparativa de herramientas de extracción de datos.....	112

Tabla 29. Criterios de evaluación de herramientas de almacenamiento.....	114
Tabla 30. Comparativa de herramientas de almacenamiento	115
Tabla 31. Criterios de evaluación de herramientas de minería de datos.....	117
Tabla 32. Resultado de evaluación de herramientas de minería de datos.....	117
Tabla 33. Aspectos importantes en la cotización.....	119
Tabla 34. Lista de atributos extraídos	124
Tabla 35. Bibliotecas necesarias para el modelo	127
Tabla 36. Campos creados dentro del objeto Recomendación	133
Tabla 37. Detalle de listas de valores.....	136
Tabla 38. Sever scripts configurados	141
Tabla 39. Action configurado	141
Tabla 40. Artículos de conocimiento para la configuración de Eloqua.....	144
Tabla 41. Artículos de conocimiento para la creación de campañas	145
Tabla 42. Talento humano sugerido para la implementación	151

Nota Aclaratoria

Género¹:

La actual tendencia al desdoblamiento indiscriminado del sustantivo en su forma masculina y femenina va contra el principio de economía del lenguaje y se funda en razones extralingüísticas. Por tanto, deben evitarse estas repeticiones, que generan dificultades sintácticas y de concordancia, que complican innecesariamente la redacción y lectura de los textos.

Este documento se redacta de acuerdo con las disposiciones actuales de la Real Academia Española con relación al uso del “género inclusivo”. Al mismo tiempo se aclara que estamos a favor de la igualdad de derechos entre los géneros.

¹ Recuperado de: <http://www.rae.es/consultas/los-ciudadanos-y-las-ciudadanas-los-ninos-y-las-ninas>

1. Introducción

En este capítulo, se presenta el contexto organizacional de la empresa XUM TECH, el equipo de trabajo, la descripción detallada de la situación problemática, proyectos similares, los beneficios esperados del presente proyecto, el planteamiento de objetivos, el alcance, la justificación, limitaciones, supuestos, exclusiones y la lista de entregables del proyecto.

1.1. Descripción general

El presente proyecto busca atender una situación problemática identificada en la empresa XUM TECH. Esta compañía es parte del sector de consultoría de tecnologías de información y brinda servicios a nivel regional en Costa Rica, Centroamérica y el Caribe desde el año 2016. La temática central de este TFG es la minería de datos, una técnica de análisis de información altamente utilizada en la actualidad, para identificar patrones ocultos y comportamientos, mediante la utilización de datos históricos. De esta forma, se pueden realizar predicciones y mejorar el proceso de toma de decisiones.

El problema afrontado por XUM TECH y que se pretende resolver con este proyecto consiste en que los servicios de analítica ofrecidos por la empresa no se ajustan a las necesidades actuales de sus clientes. Diversos clientes de XUM TECH han mostrado interés en adquirir una solución de analítica predictiva que les permita recomendar productos de una manera personalizada a sus clientes, mediante el uso de los datos almacenados en el CRM Oracle Engagement Cloud, brindado por XUM TECH. Sin embargo, debido al desconocimiento en temas de *Machine Learning*, metodologías y herramientas para el procesamiento de datos por parte de los colaboradores, así como por la falta de experiencia en proyectos de minería de datos, no se ha logrado satisfacer esta necesidad.

Tomando como base este problema, se planteó como objetivo general “Proponer una solución de minería de datos basada en la metodología CRISP-DM utilizando los datos almacenados en Oracle

Engagement Cloud, para la recomendación personalizada de productos a clientes, durante el primer semestre de 2022”. Esto permitirá a la empresa XUM TECH obtener beneficios como: el desarrollo de una propuesta de recomendación personalizada de productos que responda a las necesidades de los clientes de XUM TECH, el análisis financiero de la viabilidad de la propuesta, y el desarrollo de conocimiento interno sobre modelos de *Machine Learning*, procesos y herramientas adecuadas tanto para la manipulación como para el procesamiento de datos.

Adicionalmente, algunos beneficios indirectos son el aprovechamiento de nuevas oportunidades, la ampliación de los servicios ofrecidos, el mejoramiento de la calidad de las soluciones, el aumento de la cartera de clientes y el incremento en la satisfacción de los clientes.

1.2. Antecedentes

1.2.1. Descripción de la organización

XUM TECH es una empresa dedicada al sector de consultoría de tecnologías de información. Fue fundada en el año 2016 y se encuentra en proceso de expandir su frontera a otros mercados regionales. Actualmente, brinda servicios a países como Costa Rica, El Salvador, Colombia, Guatemala, Honduras, Nicaragua, Panamá, Perú, entre otros países de Latinoamérica.

De acuerdo con XUM TECH (s.f.):

XUM TECH se destaca en el ecosistema de *partners* de Oracle como un *partner* especialista en el diseño de estrategias para impactar la experiencia de los clientes e implementación de la solución Oracle Customer Experience en los distintos procesos de mercadeo, ventas, servicio, lealtad, comercio digital; mediante la metodología de implementación *Cloud Application Services Oracle Unified Method*, la cual asegura no solo una implementación exitosa, si no, también lograr los objetivos de negocio y el impacto en los distintos procesos que buscan nuestros clientes. (párr. 1)

A nivel comercial XUM TECH se especializa en la implementación de sistemas enfocados en la experiencia al cliente, automatización de ventas, procesos de servicio al cliente, mercadeo digital y comercio electrónico. Dentro de los principales productos que comercializa e implementa la organización, se encuentran Oracle Engagement Cloud, Oracle Eloqua, Oracle Policy Automation y Oracle Field Service.

A nivel de organizacional, XUM TECH posee un enfoque en los resultados y la excelencia. Además, se caracteriza por ofrecer una estructura organizacional plana, donde la gestión es descentralizada, esto quiere decir que cada persona administra su tiempo y se empodera de sus tareas con el fin de cumplir los objetivos, por lo tanto, no existen jefes, gerentes ni supervisores.

1.2.1.1. Misión

“Hacer que las tecnologías de información más innovadoras sean accesibles a todas las organizaciones sin importar su tamaño o familiaridad con la tecnología” (XUM TECH, 2020).

1.2.1.2. Visión

De acuerdo con XUM TECH (2021), la visión actual de la organización se encuentra basada en cuatro perspectivas:

- Operación interna y estructura: se busca renovar la imagen digital de XUM TECH y fortalecer su equipo interno mediante el mejoramiento de los procesos internos, con el fin de reforzar la presencia comercial de la empresa y aumentar la calidad de los proyectos.
- Relaciones con clientes: desarrollar instrumentos y destinar recursos con el propósito de fortalecer la relación con los clientes, medir los niveles de satisfacción de los servicios recibidos e identificar oportunidades de mejora.
- Conocimiento y crecimiento profesional: capacitar continuamente al personal de XUM TECH en los diferentes productos Oracle que ofrece la empresa, adicionalmente, crear y dar seguimiento a un plan de carrera para los colaboradores de XUM TECH. Desarrollar alianzas con instituciones educativas con el fin de dar a conocer la organización a estudiantes universitarios.

- Crecimiento financiero: expandir las operaciones de XUM TECH a nivel regional, a fin de favorecer los procesos financieros. Acumular una autonomía financiera de 8.5x y obtener ganancias netas de \$1.200.000 para diciembre de 2022. (XUM TECH, 2021, diapositiva 2)

1.2.1.3. Valores

Los principios culturales o valores que caracterizan y distinguen tanto a la organización como sus colaboradores son los siguientes:

- Empoderamiento: los colaboradores son responsables de dar seguimiento a sus asignaciones y tomar control de las situaciones para alcanzar los resultados deseados.
- Enfoque: el propósito de la organización es generar un impacto positivo para sus clientes.
- Excelencia: entregar productos de alta calidad con un enfoque en el costo-eficiencia.
- Evolución: cada día intentar ser una mejor versión, adaptarse a las nuevas necesidades y buscar nuevas formas de realizar los procesos; la organización y sus colaboradores se encuentran comprometidos con la mejora continua.
- Empatía: comprender y aceptar las diferencias tanto entre colaboradores como clientes, con el fin de brindar las mejores soluciones. (XUM TECH, 2019, diapositiva 3)

1.2.1.4. Equipo de trabajo

En esta sección, se describen de manera detallada los miembros del equipo involucrados en el desarrollo del proyecto y sus respectivos roles. Aunque XUM TECH se destaca por ser una organización con una estructura plana u horizontal, los colaboradores cumplen distintas funciones y aportan conocimiento en diversas áreas. En la **Tabla 1. Equipo de trabajo** se detalla el puesto y el rol desempeñado por los miembros del proyecto.

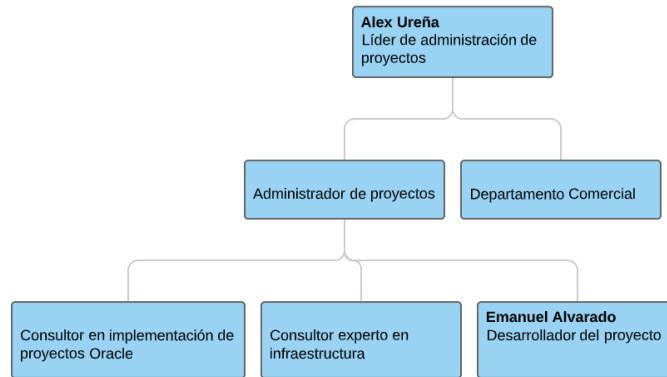
Tabla 1. *Equipo de trabajo*

Puesto en la organización	Rol dentro del proyecto
Líder de Administración de Proyectos (Alex Ureña)	Encargado de verificar y aprobar el proyecto desarrollado. Brinda información acerca de las necesidades de este.
Departamento Comercial	Ofrece información sobre las necesidades expuestas por los clientes de XUM TECH para la implementación de soluciones.
Administrador de Proyectos	Aporta conocimientos sobre el uso de las herramientas y experiencia basada en proyectos antiguos.
Consultor en implementación de proyectos Oracle	Brinda información acerca de las capacidades y el funcionamiento de las herramientas Oracle ofrecidas por XUM TECH.
Consultor experto en infraestructura	Ofrece conocimientos sobre los componentes de infraestructura utilizados en soluciones implementadas anteriormente.
Desarrollador del proyecto (Emanuel Alvarado)	Encargado de desarrollar el proyecto y proponer una solución que se ajuste a las necesidades del negocio.

Nota. Elaboración propia.

En la Figura 1. Organigrama, se muestra cómo se organiza el equipo involucrado en el desarrollo del proyecto. Para cada miembro, se detalla el nombre del puesto y el nivel en la jerarquía de XUM TECH según sus responsabilidades.

Figura 1. Organigrama



Nota. Elaboración propia.

1.2.2. Proyectos similares

En esta sección, se describen los proyectos relacionados con el tema, que fueron desarrollados dentro de la organización y proyectos externos, que fueron realizados por otros estudiantes de la carrera Administración de Tecnología de Información. Estos serán utilizados como insumos para el desarrollo del presente TFG.

1.2.2.1. Proyectos Internos

Para identificar los proyectos internos, se realizó la entrevista detallada en el “Apéndice D. Resultados de la Entrevista inicial” al Líder de Administración de Proyectos, donde se le consultó acerca de proyectos relacionados con la temática central de este TFG. A continuación, se describen dos de los proyectos identificados a nivel interno:

- Módulo para recomendación de ventas: de acuerdo con el Líder de Administración de Proyectos, se han planteado proyectos similares donde los clientes han querido utilizar los datos generados por la herramienta Oracle Engagement Cloud, para crear predicciones; sin embargo, por limitaciones técnicas y económicas del cliente interesado no se concretó este proyecto. Si bien no se desarrolló, existe

documentación como *mockups* y un registro de historias de usuario que pueden servir como insumos para esta investigación.

- Desarrollo de analíticos: XUM TECH ha implementado múltiples proyectos de analítica empresarial utilizando productos Oracle; principalmente, se han desarrollado reportes como *dashboards*, analítica descriptiva e inteligencia de negocios para departamentos específicos de sus clientes.

1.2.2.2. Proyectos Externos

Para identificar los proyectos externos, se realizó una revisión documental de antiguos TFG disponibles en la Comunidad ATI de la plataforma Tec Digital. En la Tabla 2. *Proyectos externos similares*, se describen cuatro proyectos parecidos que se utilizaron como insumo para el desarrollo del presente trabajo:

Tabla 2. *Proyectos externos similares*

Título del proyecto	Autor	Relación - Aporte
Desarrollo de herramienta de inteligencia de negocios para interpretación, análisis y evaluación de datos relacionados al procesamiento de criptomonedas Caso: Progressio Digital	José Andrés Barboza González	Se desarrolla una propuesta de procesos de ETL e inteligencia de negocio. Incluye la propuesta de herramientas para el procesamiento y análisis de datos, adicionalmente detalla buenas prácticas en la gestión de datos. Aporte para el presente proyecto: brinda una guía para la selección de herramientas de análisis y procesamiento de datos.
Solución de inteligencia de negocios para el análisis de datos en el ERP Business Pro	Lenynd Bermúdez Delgado	Desarrollo de procesos ETL e inteligencia de negocio en un sistema de información empresarial.

Título del proyecto	Autor	Relación - Aporte
		Aporte para el presente proyecto: insumo guía para el abordaje de una propuesta de solución de análisis de datos en un SIE.
Propuesta de un modelo de predicción para la infraestructura tecnológica en proyectos de migración. Caso: Mobilize.net	Vinicio Jesús Mora Herrera	Se proponen metodologías de analítica empresarial para la predicción de infraestructura tecnológica y se diseñan de modelos de predicción para resolver un caso real del negocio. Aporte para el presente proyecto: referencia sobre cómo proponer una solución de minería de datos basado en metodología CRISP DM.
Desarrollo de una solución basada en inteligencia de negocios para la adopción de los datos que forman parte de la cadena de suministro de Procter & Gamble	Natalia Araya Chaves	Desarrollo de procesos de ETL, inteligencia de negocios e identificación de herramientas en la nube, útiles para la manipulación de gran cantidad de datos. Aporte para este proyecto: insumo para consultar buenas prácticas, así como herramientas utilizadas para el procesamiento y manipulación de grandes volúmenes de datos.
Desarrollo de prueba de concepto sobre las campañas de mercadeo con el canal de Facebook Messenger para su implementación en el catálogo de servicios de la empresa	Nathalie Jiménez Marín	Se utilizan sistemas empresariales de la suite Oracle <i>Customer Experience</i> , como <i>Engagement Cloud</i> y <i>Eloqua</i> para proponer soluciones orientadas a la creación de campañas de mercadeo.

Nota. Elaboración propia.

1.3. Planteamiento del problema

En esta sección, se describe la problemática identificada dentro del entorno de la organización XUM TECH, la cual motiva el desarrollo de este proyecto. Adicionalmente, se mencionan los beneficios y aportes esperados.

1.3.1. Situación problemática

Desde su fundación en el año 2016, XUM TECH ha brindado servicios de consultoría e implementación de proyectos de TI. Su producto principal ha sido el CRM llamado Oracle Engagement Cloud, este sistema es utilizado por los clientes de XUM TECH para gestionar la operación de sus ventas y servicios durante todo su ciclo de vida.

Adicionalmente, esta herramienta ha servido para ofrecer funcionalidades de analítica descriptiva o inteligencia de negocios a los clientes, al analizar la información de una manera más sencilla por medio de gráficos y visualizaciones. Con esto, facilita la toma de decisiones a los gerentes y líderes de la organización.

De acuerdo con Alex Ureña (2021), en los últimos meses, diversos clientes han mostrado interés en utilizar los datos generados por el sistema Oracle Engagement Cloud para hacer predicciones, con el propósito de agilizar sus procesos internos. Uno de los casos que más ha causado interés en los clientes es utilizar el historial de compras de sus clientes, para recomendar productos de una manera personalizada según sus preferencias; esto les permite tener una mayor oportunidad de generar ventas. Esta necesidad surge debido a que el proceso actual de recomendación de productos es genérico y consume una gran cantidad de recursos de los clientes de XUM TECH.

Como se mencionó, la herramienta de Oracle Engagement Cloud ofrece la opción de crear analíticos descriptivos de los datos almacenados (esto incluye reportes de ventas, solicitudes de servicios, clientes, productos, entre otros). Según indicó Alex Ureña (2021), se ha explorado la herramienta, pero no ha sido posible procesar datos y transformarlos para crear predicciones de manera nativa; adicionalmente, el equipo de XUM TECH tiene escasa experiencia en el uso e

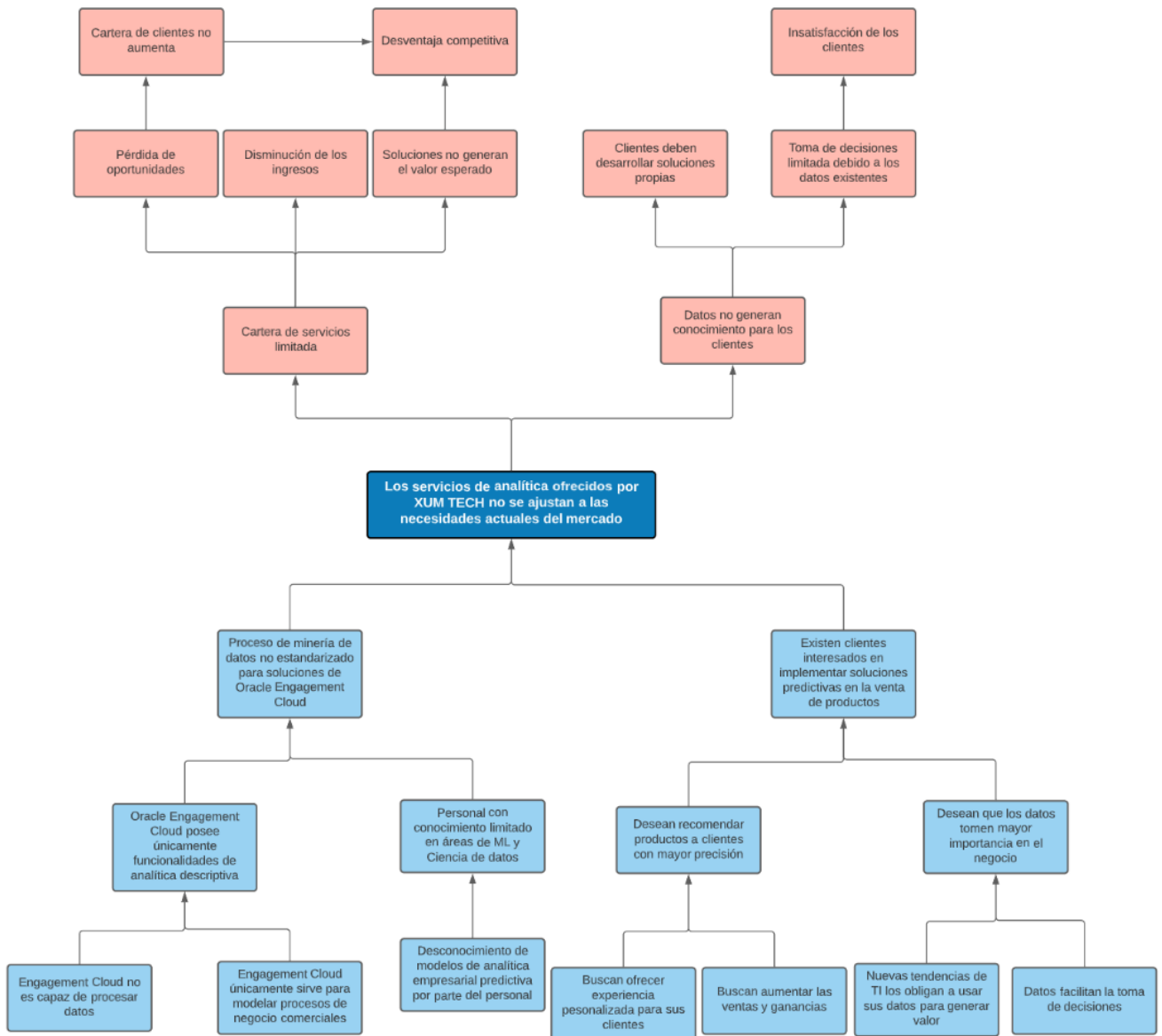
implementación de modelos de *Machine Learning*, ya que desconoce las herramientas, técnicas y metodologías necesarias para diseñar un modelo predictivo.

Este desconocimiento hace que XUM TECH pierda oportunidades de negocio y provoca que sus clientes deban buscar nuevos proveedores de servicios de analítica o, incluso, desarrollar sus propias soluciones, para poder utilizar los datos almacenados en la herramienta Oracle Engagement Cloud, con el propósito de facilitar su proceso de recomendación de productos. Por otra parte, las soluciones ofrecidas por XUM TECH no generan el valor esperado para el cliente, lo cual provoca insatisfacción y una desventaja competitiva.

Luego de exponer las causas y consecuencias de la situación actual, se identifica que el problema de la organización es que los servicios de analítica ofrecidos por XUM TECH no se ajustan a las necesidades actuales de sus clientes. Por este motivo, para la organización es importante diseñar nuevas soluciones dirigidas a sus clientes, con el fin de ofrecer la posibilidad de utilizar los datos almacenados en su CRM para realizar predicciones y, de esta manera, adaptarse a las nuevas necesidades del mercado.

En la Figura 2, se muestra las causas-raíz del problema y sus efectos.

Figura 2. Árbol del problema



Nota. Elaboración propia.

1.3.2. Justificación del proyecto

En esta sección, se presenta la justificación del proyecto. Se describe de manera detallada cómo este proyecto se adapta al perfil de un administrador de tecnología de información (ATI), la importancia que tiene en la organización y las áreas de desarrollo relacionadas con él.

El proyecto presentado se alinea al perfil de un ATI, ya que este se enfoca en la innovación y creación de valor en una empresa de transformación digital como lo es XUM TECH.

Mediante este proyecto se pretende atender una necesidad de XUM TECH. Esto le permitirá a la empresa innovar en los servicios de analítica que ofrece y adaptarse a las necesidades actuales del mercado. Esto debido a que diversos clientes han mostrado interés en soluciones de analítica predictiva con el fin de recomendar productos a sus clientes de una manera personalizada y precisa, para así tener mayores posibilidades de aumentar sus ventas y ganancias, sin embargo, hasta el momento, la empresa no ha logrado satisfacer esa necesidad.

Los beneficios indirectos que se podría obtener XUM TECH al desarrollar este proyecto son la expansión de su cartera de servicios, obtención de mayores ingresos, aumento en la cartera de clientes, mayor valor en los proyectos y mejorar la satisfacción de sus clientes.

En relación con las áreas de desarrollo de la carrera, el proyecto se encuentra asociado específicamente con Sistemas de Información Empresariales y Gestión de Datos e Información, establecidas en el Artículo 14 del Reglamento Específico del Trabajo Final de Graduación.

La asociación con el área Sistemas de Información Empresariales se da por el uso de la herramienta empresarial Oracle Engagement Cloud, la cual es un *Customer Relationship Management* (CRM) que les permite a las empresas gestionar la información de sus clientes, operaciones y todo su ciclo de ventas. En esta herramienta, los clientes de XUM TECH realizan sus procesos de venta y es donde se encuentra almacenada toda la información transaccional que será necesaria para el desarrollo de este proyecto.

Propuesta de solución de minería de datos basada en la metodología CRISP-DM utilizando los datos almacenados en Oracle Engagement Cloud, para la recomendación personalizada de productos a clientes, durante el primer semestre de 2022

Por otro lado, la relación con el área de Gestión de Datos e Información se da porque para el proyecto es necesario investigar sobre herramientas que permitan la extracción, limpieza, manipulación y transformación de los datos, esto incluye el uso de servicios en la nube que garanticen la escalabilidad, integridad y confiabilidad de la solución propuesta. También, se encuentra relacionado con esta área debido al uso de conceptos, modelos y tecnologías de *Machine Learning*, que son necesarios para realizar de manera satisfactoria la predicción de productos recomendados para cada cliente.

Adicionalmente, para el desarrollo de este proyecto, se plantea utilizar el estándar CRISP-DM (*Cross Industry Standard Process for Data Mining*), empleado durante el curso de Analítica Empresarial de la carrera Administración de Tecnología de Información, de acuerdo con IBM (s.f). Este consiste en una metodología altamente probada en la industria, que permite guiar el proceso de minería de datos. La metodología CRISP-DM brinda una descripción detallada de las fases y actividades necesarias para el desarrollo de una solución exitosa, al tiempo que ofrece la flexibilidad precisa para adaptarse a las necesidades de cada organización.

1.3.3. Beneficios esperados del Trabajo Final de Graduación

En la Tabla 3. *Beneficios esperados*, se describen los beneficios directos e indirectos que se esperan al realizar este proyecto.

Tabla 3. *Beneficios esperados*

Perspectiva	Beneficios directos	Beneficios indirectos
Financiera	<ul style="list-style-type: none"> Análisis financiero que permite validar la viabilidad de la solución de minería de datos para la recomendación personalizada de productos. 	<ul style="list-style-type: none"> Ampliación de los servicios ofrecidos por XUM TECH. Mayores ingresos por venta de servicios. Mayores ingresos a los clientes porque saben recomendar productos de manera más precisa.
Clientes	<ul style="list-style-type: none"> Propuesta de solución de recomendación personalizada de productos utilizando los datos almacenados en el CRM Oracle Engagement Cloud. 	<ul style="list-style-type: none"> Mayor satisfacción y retención de clientes. Atraer nuevos clientes interesados en una solución de analítica predictiva. Atención a una necesidad actual de los clientes de XUM TECH. Ofrecer mayor valor agregado a las soluciones mediante la recomendación de productos de manera personalizada.
Capacidades Internas	<ul style="list-style-type: none"> Identificar nuevas formas de utilizar los datos de Oracle Engagement Cloud, para facilitar la toma de decisiones de los clientes (en este caso la recomendación de productos). 	<ul style="list-style-type: none"> Descubrir nuevas herramientas que puedan ser útiles en futuros proyectos que no necesariamente estén relacionados con minería de datos, por ejemplo, herramientas de

Propuesta de solución de minería de datos basada en la metodología CRISP-DM utilizando los datos almacenados en Oracle Engagement Cloud, para la recomendación personalizada de productos a clientes, durante el primer semestre de 2022

Perspectiva	Beneficios directos	Beneficios indirectos
	<ul style="list-style-type: none">Identificar herramientas necesarias a lo largo del análisis de los datos para generar el modelo predictivo.	extracción, limpieza, almacenamiento y análisis de los datos.

Nota. Elaboración propia.

1.4. Objetivos

En esta sección, se indican los objetivos del proyecto, para elaborarlos se empleó la taxonomía de Bloom.

1.4.1. Objetivo general

Proponer una solución de minería de datos basada en la metodología CRISP-DM utilizando los datos almacenados en Oracle Engagement Cloud, para la recomendación personalizada de productos a clientes, durante el primer semestre de 2022.

1.4.2. Objetivos específicos

1. Analizar las necesidades capturadas por el departamento comercial de XUM TECH, para la identificación de los requerimientos necesarios para la propuesta.
2. Evaluar varios modelos predictivos mediante la utilización de los datos obtenidos de Oracle Engagement Cloud, para la selección de un modelo adecuado de minería de datos que permita la recomendación personalizada de productos a clientes.
3. Recomendar las herramientas de extracción, almacenamiento y procesamiento de los datos, para la integración de los datos de Oracle Engagement Cloud y el modelo predictivo.
4. Elaborar un análisis financiero de la solución de minería de datos propuesta, para la verificación de la viabilidad del proyecto y la identificación de beneficios financieros y no financieros.

1.5. Alcance del proyecto

En esta sección, se brinda una descripción detallada de las actividades que se realizarán durante la ejecución del presente proyecto para cumplir con los objetivos propuestos.

En el mercado existen distintas metodologías para desarrollar proyectos de minería de datos. Dentro de las principales, se encuentran KDD (*Knowledge Discovery in Databases*), SEMMA (*Sample, Explore, Modify, Model, Asses*) y CRISP-DM (*Cross Industry Standard Process for Data Mining*). Para el desarrollo de este proyecto, se utilizará la metodología CRISP-DM. Los motivos se exponen a continuación.

De acuerdo con el artículo publicado por Azebedo y Santos (2008), la metodología CRISP-DM posee las siguientes características que la hacen idónea para el desarrollo de proyectos de minería de datos:

- A diferencia del modelo SEMMA y KDD, CRISP-DM contiene una etapa de “Entendimiento del Negocio”. Esto es de suma importancia para el desarrollo del proyecto, ya que permitirá determinar de manera precisa las necesidades que debe satisfacer la solución de minería de datos propuesta, así como los objetivos del cliente.
- El modelo CRISP-DM es sumamente completo, cada una de sus fases se encuentra claramente definida y estructurada, además, detalla los artefactos de documentación y resultados esperados para cada etapa.

Por otro lado, de acuerdo con Data Science Process Alliance (s.f.), se detallan los siguientes puntos a favor de la metodología CRISP-DM:

- Luego de realizar una encuesta en el portal de KDnuggets, sitio encargado de medir el uso de metodologías para el desarrollo de proyectos de minería y ciencia de datos en la industria, la metodología CRISP-DM resultó ganadora durante los años 2002, 2004, 2007 y 2020, por ser la más utilizada en la industria.
- CRISP-DM ofrece beneficios a las organizaciones, como la fácil implementación, por lo tanto, no es necesario realizar grandes cambios en las organizaciones ni capacitaciones.

También, gracias a su primera fase de entendimiento del negocio, ayuda a alinear las necesidades de la empresa con el trabajo técnico y garantiza que el equipo de trabajo entienda correctamente los objetivos del negocio. Además, ofrece flexibilidad para complementarse con otros principios de agilidad o con mejores prácticas utilizadas en la industria.

A continuación, se presentan las principales fases que tendrá el proyecto, basado en sus necesidades y en el artículo publicado por IBM (2014), donde se detalla el proceso de minería de datos y la metodología oficial de CRISP-DM, publicada por Chapman (2000):

- Entendimiento del negocio: consiste en determinar las necesidades expuestas por los clientes de XUM TECH, las cuales deben ser consideradas para el desarrollo del modelo de recomendación de productos. Como resultado de esta etapa, se espera la especificación de requerimientos de la solución, y la identificación de los objetivos de negocio y de la minería de datos.
- Comprensión de los datos: se exploran e identifican los datos existentes en Oracle Engagement Cloud respecto a ventas, historial de compras y características de los clientes que podrían ser útiles para la creación del modelo de recomendación de productos. Dentro de los entregables esperados de esta etapa, se encuentran el reporte de los datos existentes y de la calidad de estos.
- Identificación de herramientas de procesamiento: consiste en investigar y seleccionar las herramientas o servicios de extracción, almacenamiento, procesamiento y análisis de datos que permitan brindar una solución compatible con Oracle Engagement Cloud y crear el modelo predictivo de recomendación de productos. El entregable de esta etapa consistirá en el reporte de costos de cada herramienta seleccionada y el análisis costo-beneficio.
- Preparación de los datos: seleccionar los datos que serán utilizados para la creación del modelo, detectar las inconsistencias y errores en ellos e identificar las transformaciones necesarias para usarlos en el modelo. El resultado de esta etapa corresponde a la lista de actividades de transformación y limpieza de datos.

- Modelado: se prueban diversas técnicas de minería adecuadas para crear el modelo de recomendación de productos y evaluar los resultados. Como producto de esta etapa, se espera la evaluación de los modelos y la selección aquel con el mejor rendimiento obtenido.
- Evaluación: consiste en comparar los resultados obtenidos frente a los objetivos planteados por el negocio, para determinar si cumple con las expectativas. El entregable principal de esta etapa serán las conclusiones obtenidas del modelo de minería de datos propuesto para la recomendación de productos, se toma como base la lista de requerimientos de la etapa de Entendimiento del negocio.
- Análisis financiero: este proyecto incorpora una fase final enfocada en el análisis de costo-beneficio de la propuesta realizada. El objetivo es evaluar la viabilidad del proyecto, así como identificar los beneficios tanto financieros como los que no lo son. El entregable asociado a esta fase será el resultado del análisis financiero.

En la **Tabla 4. Fases asociadas a objetivos**, se detalla la relación existente entre los objetivos del proyecto y cada una de las fases definidas dentro del alcance.

Tabla 4. Fases asociadas a objetivos

Objetivo	Fases asociadas
Analizar las necesidades capturadas por el departamento comercial de XUM TECH, para la identificación de los requerimientos necesarios para la propuesta.	<ul style="list-style-type: none"> • Entendimiento del negocio.
Recomendar las herramientas de extracción, almacenamiento y procesamiento de los datos, para la integración de los datos de Oracle Engagement Cloud y el modelo predictivo.	<ul style="list-style-type: none"> • Identificación de las herramientas de procesamiento
Evaluar varios modelos predictivos mediante la utilización de los datos obtenidos de Oracle	<ul style="list-style-type: none"> • Comprensión de los datos • Preparación de los datos

Propuesta de solución de minería de datos basada en la metodología CRISP-DM utilizando los datos almacenados en Oracle Engagement Cloud, para la recomendación personalizada de productos a clientes, durante el primer semestre de 2022

Objetivo	Fases asociadas
Engagement Cloud, para la selección de un modelo adecuado de minería de datos que permita la recomendación personalizada de productos a clientes.	<ul style="list-style-type: none">• Modelado• Evaluación
Elaborar un análisis financiero de la solución de minería de datos propuesta, para la verificación de la viabilidad del proyecto y la identificación de beneficios financieros y no financieros.	<ul style="list-style-type: none">• Análisis Financiero

Nota. Elaboración propia.

1.6. Supuestos

En esta sección, se detallan los factores que fueron necesarios para la ejecución satisfactoria de este proyecto:

- Apoyo por parte de la organización para acceder a los ambientes de Oracle Engagement Cloud, explorar y consultar la información necesaria para el desarrollo del proyecto.
- La modalidad de las reuniones utilizada fue 100% remota durante el desarrollo del proyecto.
- Se brindó acceso a la documentación oficial de Oracle y bases de conocimiento internas de XUM TECH, para conocer las capacidades de Oracle Engagement Cloud y recolectar información de proyectos anteriores.
- Las entrevistas fueron aplicadas únicamente a miembros del equipo de XUM TECH, con el objetivo obtener la información necesaria y aclarar duda sobre aspectos técnicos y administrativos.
- La información brindada por parte de XUM TECH sobre aspectos de costos operativos y salarios está sujeta a políticas de confidencialidad, por esto se definieron montos estimados para el análisis financiero de la propuesta.

1.7. Entregables

En esta sección, se detallan los entregables a realizar durante el proyecto. Se toman en cuenta los entregables del producto y los relacionados con la gestión del proyecto.

1.7.1. Entregables del producto

En este apartado, en la **Tabla 5. Entregables asociados**, se describen los entregables asociados a cada objetivo del proyecto.

Tabla 5. Entregables asociados

Objetivo específico	Entregables
Analizar las necesidades capturadas por el departamento comercial de XUM TECH, para la identificación de los requerimientos necesarios para la propuesta.	1. Especificación de requerimientos.
Evaluar varios modelos predictivos mediante la utilización de los datos obtenidos de Oracle Engagement Cloud, para la selección de un modelo adecuado de minería de datos que permita la recomendación personalizada de productos a clientes.	2. Reporte de datos existentes (origen y descripción de estos). 3. Reporte de calidad de datos. 4. Lista de datos necesarios para desarrollar el modelo. 5. Actividades de limpieza y transformación de datos. 6. Modelos de predicción propuestos. 7. Evaluación de los modelos.

Objetivo específico	Entregables
	8. Modelo de predicción seleccionado.
Recomendar las herramientas de extracción, almacenamiento y procesamiento de los datos, para la integración de los datos de Oracle Engagement Cloud y el modelo predictivo.	11. Lista de herramientas propuestas. 12. Proveedores seleccionados. 13. Costo aproximado de licencias y servicios.
Elaborar un análisis financiero de la solución de minería de datos propuesta, para la verificación de la viabilidad del proyecto y la identificación de beneficios financieros y no financieros.	15. Propuesta de hoja de ruta para la implementación. 16. Análisis financiero. 17. Beneficios financieros. 18. Beneficios no financieros.

Nota. Elaboración propia.

1.7.2. Gestión del proyecto

En esta sección, se describen los artefactos que serán utilizados para la gestión del proyecto.

1.7.2.1. Cronograma

A continuación, en la **Tabla 6. Cronograma del proyecto**, se detalla el cronograma con la distribución de tareas para el desarrollo de este proyecto.

Tabla 6. Cronograma del proyecto

Actividad	Semanas de cronograma TFG																	
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Búsqueda de insumos iniciales																		

Propuesta de solución de minería de datos basada en la metodología CRISP-DM utilizando los datos almacenados en Oracle Engagement Cloud, para la recomendación personalizada de productos a clientes, durante el primer semestre de 2022

Actividad	Semanas de cronograma TFG																	
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Asignación de tutora																		
Primera reunión con tutora asignada																		
Primera reunión de TFG																		
Definición del cronograma del proyecto																		
Reunión inicial con los involucrados del proyecto (empresa, estudiante, tutora)																		
Ajustes y correcciones del anteproyecto																		
Marco conceptual																		
Objetivo específico #1																		
Avance I. Capítulo #1 EMP-PR																		
Correcciones del avance I																		
Definición de la metodología del proyecto																		
Avance II. Capítulo #2 MC																		
Evaluación de la empresa #1																		
Evaluación de tutora #1																		
Segunda reunión de TFG																		
Objetivo específico #2																		
Correcciones del avance II																		
Avance III. Capítulo #3 MM																		
Reunión #2 con los involucrados: empresa, estudiante y tutora																		
Correcciones del avance III																		
Tercera reunión de TFG																		
Avance IV. Capítulo #4 R-PS																		
Evaluación de tutora #2																		
Evaluación de la empresa #2																		
Correcciones del avance IV																		
Objetivo específico #3																		
Avance V. Capítulo #4 R-PS																		
Evaluación de la viabilidad del proyecto																		
Correcciones del avance V																		

Actividad	Semanas de cronograma TFG																	
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Objetivo específico #4																		
Cuarta reunión de TFG																		
Reunión #3 de los involucrados: empresa, estudiante y tutora																		
Evaluación de tutora #3																		
Evaluación de la empresa #3																		
Avance VI. Informe final(correcciones)																		
Finalización del TFG en la organización																		
Entrega del informe a la coordinación																		
Lectura de TFG																		
Defensa Oral de TFG																		

Nota. Elaboración propia.

1.7.2.2. Minutas

Para la gestión del proyecto, se utilizan minutas para el seguimiento de las reuniones donde participe al menos un involucrado de proyecto. La plantilla se encuentra en el “Apéndice A. Plantilla de minutas para reunión” y cuenta con las siguientes secciones:

- Número de reunión
- Lugar
- Objetivo de la reunión
- Participantes
- Hora de inicio y hora de finalización
- Temas tratados (asunto, comentarios y acuerdos)
- Fecha, tema y convocados de la próxima reunión (cuando aplique)

1.7.2.3. Gestión del cambio

Para la gestión del proyecto, se utilizan hojas de control de cambios solicitados, los cuales deben solicitarse con la plantilla que se encuentra descrita en el “Apéndice B. Plantilla para la gestión de cambios”, donde detalla la siguiente información:

- Identificador del cambio
- Solicitante del cambio
- Fecha de solicitud del cambio
- Responsable de la implementación
- Fecha de realización del cambio
- Estado del cambio
- Categoría del cambio
- Descripción del cambio
- Justificación
- Implicaciones de realizar el cambio
- Comentarios
- Encargado de revisar y aprobar el cambio

1.8. Limitaciones del proyecto

Durante el proyecto, se deben considerar factores restringieron su ejecución. A continuación, se detallan estos:

- Únicamente se consultó información con los colaboradores de XUM TECH, los cuales conocen las necesidades expuestas por sus clientes y serán la base de información para este proyecto.
- Para la cotización de las herramientas analizadas se utilizó la información disponible solamente en sitios oficiales del proveedor.
- La información utilizada para la elaboración del proyecto fue extraída de un ambiente de pruebas brindado por XUM TECH.
- Dentro de los ambientes brindados por XUM TECH, no fue posible crear versiones publicables de la aplicación ni modificar los registros e información existente, esto para evitar alteraciones en el ambiente del cliente y pérdidas de información.
- Dentro del proyecto no se incluye cualquier información que sea considerada sensible del registro de clientes (por ejemplo: nombre, correo electrónico, número de teléfono, entre otros), ni de la lista de productos analizados por motivos de confidencialidad.

1.9. Exclusiones del proyecto

A continuación, se enlistan aquellas actividades, detalles y aspectos en general que fueron excluidos del alcance de este proyecto:

- Implementación de la solución en un ambiente de producción o desarrollo: el proyecto consiste en una propuesta, por este motivo, la integración de la solución a las plataformas existentes le corresponderá a la empresa. La carga de datos se realizó en el ambiente de pruebas autorizado únicamente para hacer una prueba de concepto.

- Capacitaciones sobre la propuesta: no se brindaron capacitaciones al personal de XUM TECH sobre la solución propuesta debido a las limitaciones de tiempo y alcance del proyecto.
- No hubo contacto directo entre el desarrollador de este proyecto y los clientes de XUM TECH que han mostrado interés en este tipo de soluciones. La información relacionada con los requerimientos fue obtenida de documentación interna, propuestas realizadas a clientes, proyectos antiguos, y por medio de reuniones con los colaboradores de XUM TECH del área de Ventas y Consultoría.
- Para el desarrollo de los modelos de minería de datos, se utilizó la versión gratuita de la herramienta sugerida, con el objetivo de hacer una prueba de concepto y evaluar los modelos de minería de datos debido a limitaciones tecnológicas y económicas de XUM TECH.

En resumen, a lo largo de este capítulo, se describió el contexto de la organización XUM TECH, el problema existente, los objetivos, el alcance, los supuestos, las limitaciones y exclusiones del presente proyecto. En el siguiente capítulo, se describirán los conceptos necesarios tanto teóricos como prácticos, para abordar el proyecto y así proponer una solución que se ajuste a las necesidades de la organización.

2. Marco conceptual

En este capítulo se listan y detallan los conceptos teóricos y prácticos, así como las definiciones necesarias para abordar de manera integral el problema identificado en la empresa XUM TECH y sustentar las propuestas de solución planteadas para cada uno de los objetivos del proyecto. Los conceptos fueron recopilados por medio de publicaciones, artículos, proyectos similares, libros, sitios web, documentos técnicos, entre otros.

2.1. Sistema de Información

De acuerdo con Stair y Reynolds (2017), un sistema de información se define como un conjunto de componentes interrelacionados que reúnen, procesan, almacenan y distribuyen datos e información y proporcionan un mecanismo de retroalimentación para cumplir un objetivo. Los autores indican que este mecanismo “ayuda a las organizaciones a lograr sus objetivos, como incrementar sus ganancias o mejorar su servicio al cliente”. (p.44)

2.1.1. Sistema de apoyo a las decisiones

Un sistema de apoyo a las decisiones, también conocido como *Decision Support System* (DSS), según Stair y Reynolds (2017), hace referencia a un conjunto organizado de personas, procedimientos, software, bases de datos y dispositivos. Su objetivo principal es favorecer la toma de decisiones para resolver problemas específicos de las organizaciones y ofrecer mayor eficiencia en las decisiones.

De acuerdo con Star y Reynolds (2017) existen algunos conceptos importantes asociados a las etapas de la información dentro de la organización:

- Dato: Consiste en un hecho aislado que representa elementos del mundo real, por ejemplo, un número de empleado, la cantidad de horas trabajadas en un período específico o la cantidad de unidades fabricadas de un producto específico.
- Información: Consiste en un conjunto de hechos o datos organizados de forma que generan valor adicional para la organización, al ser analizados de manera conjunta. También, se puede definir como un conjunto de datos relacionados.

- **Conocimiento:** Consiste en la comprensión de un conjunto de información y de las formas en que esta puede convertirse en algo útil para realizar una tarea específica o tomar una decisión. También, tener conocimiento implica comprender las distintas relaciones entre la información disponible.
- **Bases de datos:** Es un conjunto organizado de hechos e información que, por lo general, consta de dos o más archivos de datos relacionados, por lo tanto, es un componente esencial para el funcionamiento de los sistemas de información. La base de datos de una organización puede contener hechos e información acerca de sus clientes, empleados, inventario, ventas, compras en línea, entre otros.

2.1.2. Customer Relationship Management

Los sistemas de gestión de la relación con los clientes, también conocidos como *Customer Relationship Management* (CRM), de acuerdo con Laudon y Laudon (2022), ayudan a las organizaciones a manejar las relaciones con sus clientes. Además, brindan la información necesaria para coordinar todos los procesos que involucran al cliente, tales como ventas, mercadeo y servicios, con el objetivo de optimizar las ganancias, mejorar la satisfacción de los clientes, ofrecer mejores servicios y aumentar la retención de los clientes.

Adicionalmente, de acuerdo con Romero y Vernadat (2016), los CRM se han convertido en sistemas de información empresarial que respaldan las actividades de ventas, permiten automatizar campañas de marketing, y brindan la posibilidad de emitir y administrar casos o boletos.

2.1.3. Oracle Engagement Cloud

La herramienta Oracle Engagement Cloud es la más utilizada por los clientes de XUM TECH esta plataforma es responsable de permitir la gestión de clientes ya que funciona como un CRM donde se registran solicitudes de servicio, ventas y mercadeo con el objetivo de brindar servicios de calidad a sus clientes. Forma parte de la *suite* de soluciones de Oracle conocida como *Customer Experience* (CX) y permite la integración con otras herramientas como Oracle Eloqua.

De acuerdo con Oracle (2018) Engagement Cloud es una herramienta que combina las capacidades de servicio y ventas en una misma solución. Con lo cual, brinda la posibilidad de automatizar las ventas, interactuar por múltiples canales con los clientes, gestionar las solicitudes de servicio y el conocimiento, y atender clientes de manera digital. La solución también posee la capacidad de tener una vista de 360° del cliente, para así facilitar la colaboración entre profesionales de ventas y servicios. Algunas características que ofrece son las siguientes:

- Completa: ofrece una solución integral de ventas enfocada en mejorar el rendimiento y conectar los procesos comerciales de la organización.
- Innovadora: brinda la posibilidad de adaptarse a las necesidades del cliente, lo cual permite impulsar la productividad.
- Probada: solución altamente reconocida en la industria, garantiza operaciones globales seguras y escalables.

2.1.4. Oracle Eloqua

Eloqua es un sistema empresarial que automatiza campañas de mercadeo, su enfoque se da en soluciones para B2B, pero es mayormente utilizado en B2C. Es una solución tecnológica que permite la configuración de distintos canales y la nutrición de información en cada etapa del proceso de compra. Contiene datos filmográficos en tiempo real, da la posibilidad de crear actividades para las ventas por medio de ciclos cerrados. Asimismo, simplifica la administración de las campañas, realiza la segmentación de los mercados por medio de parámetros asignados y modela los perfiles para obtener oportunidades de ventas más precisas (Oracle, 2021).

Esta herramienta forma parte de la suite de Oracle *Customer Experience (CX)* y es la más utilizada por los clientes de XUM TECH para la creación de campañas de mercadeo y el envío de publicidad por medio de canales electrónicos como correo electrónico y redes sociales.

2.2. Comercio electrónico

De acuerdo con Laudon y Guercio (2020), el e-commerce o comercio electrónico implica el uso de Internet, la *World Wide Web* (Web), aplicaciones móviles y navegadores que se ejecutan en dispositivos móviles para realizar transacciones comerciales. Más formalmente, se puede definir como transacciones comerciales habilitadas de forma digital entre organizaciones e individuos.

De acuerdo con Laudon y Guercio (2020), existen ocho características del *e-commerce* que han impactado la conducta de los negocios y hacen única esta tecnología:

- **Universalidad:** las transacciones comerciales están disponibles en casi todas partes, en todo momento. A diferencia del comercio tradicional, se liberan las restricciones de comercios ubicados en un espacio físico determinado y hace posible comprar artículos desde la propia casa, en el trabajo o incluso desde su automóvil, mediante el comercio electrónico móvil.
- **Alcance global:** la tecnología de comercio electrónico permite que usuarios de todo el mundo realicen compras en sitios ubicados en diferentes zonas geográficas, de manera más conveniente y rentable que el comercio tradicional. Esta característica provoca que para los comerciantes sea más sencillo obtener clientes, así, el número total de usuarios que podría llegar a tener una empresa estará limitado a medida de su alcance.
- **Estándares universales:** corresponde a la existencia de estándares técnicos universales que reducen considerablemente los costos de entrada al mercado, para los comerciantes. Por otro lado, los estándares universales permiten reducir los costos de búsqueda y el esfuerzo requerido por parte de los consumidores al buscar los productos requeridos; esto provoca que el descubrimiento y la comparativa de precios se vuelva más simple, rápida y precisa.
- **Tecnología social:** posibilidad de crear nuevas redes sociales y fortalecer las existentes. El comercio electrónico permite a los usuarios de la web crear y compartir contenido a una comunidad mundial.

- Personalización: ofrecer un mensaje de mercadeo a un segmento específico de clientes, basado en sus características, intereses, comportamiento y compras anteriores.
- Adaptación: cambiar el producto o servicio entregado con base en las preferencias del cliente o en su historial de comportamiento.
- Densidad de información: el comercio electrónico ha disminuido los costos de recolección, almacenamiento y procesamiento de datos. Esto ha implicado un aumento considerable en la densidad de la información, es decir, la cantidad total de datos de calidad disponible para todos los participantes del mercado.
- Interactividad: la tecnología de comercio electrónico permite la comunicación en dos direcciones entre el comerciante y los consumidores; esto se puede lograr mediante el uso de foros, comunidades, comentarios, redes sociales, botones de “Me gusta” y “Compartir”, páginas web y otros espacios donde interactúan los diferentes consumidores con el comerciante.
- Riqueza: permiten ofrecer información de manera interactiva y ajustada a los usuarios de manera individual. Las tecnologías de comercio electrónico habilitan a los comerciantes minoristas y de servicios tanto para comercializar como para vender bienes y servicios "complejos", los cuales antes requerían de una interacción cara a cara por parte de una fuerza de ventas a un público.

2.2.1. Tipos de transacciones

Laudon y Guercio (2020) definen dos tipos de transacciones relevantes dentro del comercio electrónico:

- Transacciones digitales: Las transacciones habilitadas digitalmente incluyen todas las transacciones mediadas por la tecnología digital. En su mayoría, esto significan transacciones que ocurren a través de Internet, la web y/o mediante dispositivos móviles (Laudon y Guercio, 2020).
- Transacciones comerciales: Las transacciones comerciales implican el intercambio de valor (por ejemplo, dinero) a través de fronteras organizacionales o individuales, a

cambio de productos y servicios. El intercambio de valor es importante para comprender los límites del comercio electrónico; sin este, no hay comercio (Laudon y Guercio, 2020).

2.2.2. Tipos de comercio electrónico

A continuación, se describen de acuerdo con lo expuesto por Chaffey y Ellis-Chadwick (2014), en su libro *Marketing Digital: estrategia, implementación y práctica*, los cuatro tipos principales de comercio electrónico:

- *Business to Business (B2B)*: es un tipo de comercio electrónico, donde un negocio se enfoca en venderle productos o servicios a otro negocio.
- *Business to Consumer (B2C)*: sucede cuando las transacciones comerciales se dan entre una organización y una persona quien adquiere los servicios o productos.
- *Consumer to Consumer (C2C)*: ocurre cuando las transacciones comerciales se realizan entre los consumidores, normalmente por sitios de negocios.
- *Consumer to Business (C2B)*: se da cuando los consumidores venden productos o servicios directamente a las organizaciones.

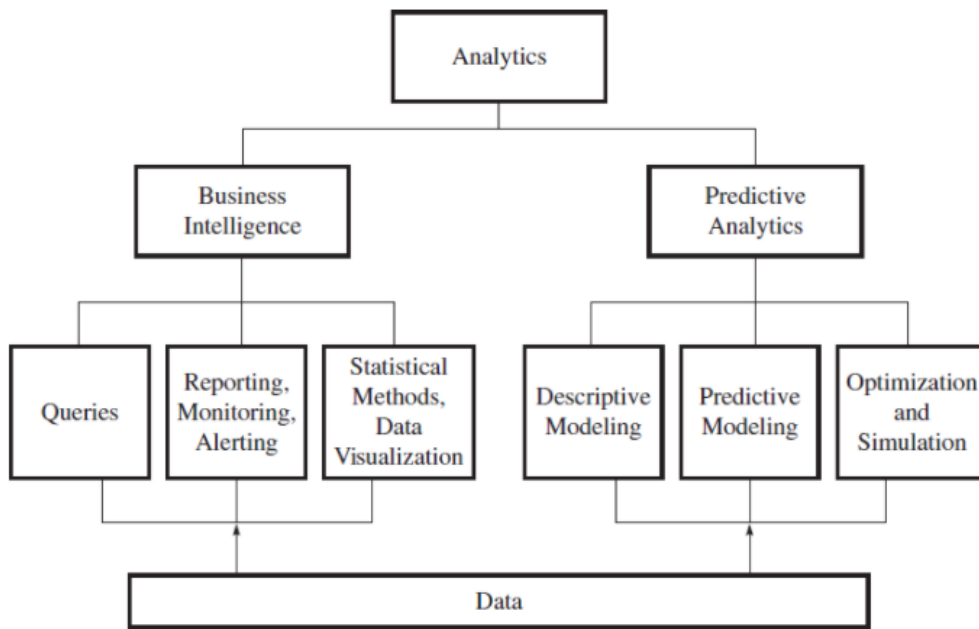
Los conceptos definidos dentro de esta sección (**Error! Reference source not found.**) son importantes dentro proyecto para entender el funcionamiento del negocio y de sus clientes. En el caso de XUM TECH se ubica dentro de la categoría de *Business to Business (B2B)* ya que ofrece servicios principalmente de consultoría, sin embargo, los clientes de XUM TECH en la mayoría de los casos se ubican dentro de la categoría de B2B o B2C ya que comercializan productos y servicios a personas tanto físicas como jurídicas.

Por otro lado, de acuerdo con la problemática identificado el concepto de *E-Commerce* y sus características se debe tomar en cuenta dentro del proyecto, ya que el principal objetivo es brindar una propuesta de solución que permita analizar las transacciones digitales de los clientes de XUM TECH para generar recomendaciones de productos, de esta manera se cumple con sus características de personalización y la adaptación.

2.3. Analítica empresarial

Gopal (2019) describe el término *Analytics* como la aplicación de diferentes métodos de análisis de datos para resolver los problemas de la organización. También, indica que es un conjunto de procesos y técnicas de ciencia de datos aplicados de manera conjunta para obtener información, analizarla, realizar pronósticos para descubrir una posible solución y cuantificar el impacto de estas posibles decisiones, para así optimizar el negocio. Los subtipos de analíticas se muestran en la Figura 3. *Subtipos de analítica*.

Figura 3. *Subtipos de analítica*



Nota. La imagen muestra los principales subtipos de analítica en la actualidad. Tomado de *An Introduction to Analytics*, M. Gopal, 2019, McGraw-Hill Education.

2.3.1. Analítica descriptiva

Esta consiste en un grupo de arquitecturas y tecnologías capaces de transformar datos sin procesar en información que se puede utilizar con fines comerciales, de tal forma que los hace significativos.

Algunas características de la analítica descriptiva son:

- Permite interpretar fácilmente grandes volúmenes de datos y medir el rendimiento histórico. Ayudan a los gerentes comerciales a comprender todo lo que sucede dentro de la organización y con el negocio.
- Incluye el cálculo de estadísticas resumidas para el análisis de datos y la resolución del problema de negocio.
- Los informes revelan información histórica relacionada con la producción, el estado financiero, las operaciones, las ventas, el inventario y los clientes de la organización.
- El término "inteligencia empresarial" suele incluir la creación y el mantenimiento de almacenes de datos, esto consiste en recopilar e integrar la información almacenada en diversas fuentes primarias y secundarias.
- Requiere que quienes toman las decisiones creen hipótesis previas y luego empleen las herramientas de análisis para llegar a las respuestas y confirmar lo percibido originalmente.
- Los entornos de inteligencia de negocios (BI) tradicionales permiten implementar alertas basadas en reglas para informar a quienes toman las decisiones sobre eventos o cambios importantes. Sin embargo, las acciones en BI siempre se definen a través de la interacción humana y son realizadas por humanos (Gopal, 2019).

2.3.2. Analítica predictiva

La analítica predictiva, según Gopal (2019), consiste en el descubrimiento automático de nuevos patrones significativos para el negocio a partir de datos estructurados y no estructurados. Al contrario de la inteligencia empresarial, este tipo de análisis se enfoca en predecir lo que puede suceder, pues busca comprender el futuro y ofrecer una respuesta basada en los datos históricos de la empresa.

Esta analítica, utiliza técnicas cuantitativas sofisticadas, como el modelado predictivo, modelado descriptivo, simulación y optimización para generar nuevo conocimiento, el cual sería imposible de descubrir con enfoques tradicionales como BI. Mediante el modelado predictivo, las empresas son capaces de analizar grandes volúmenes de datos para descubrir patrones ocultos y difíciles de observar por los humanos; de esta manera, pueden identificar riesgos y oportunidades desconocidas.

Algunos ejemplos de donde se puede aplicar este tipo de análisis son la predicción del comportamiento de los clientes, la identificación de patrones de compra, la identificación de tendencias en las actividades relacionadas con ventas, la predicción de demanda en procesos de cadena de suministro, el cálculo de puntaje financiero para la aprobación de créditos bancarios, entre otros.

Han et al. (2012) menciona que existen dos tipos de análisis predictivos principales, los cuales son:

- **Clasificación:** permite encontrar un modelo que describa y distinga entre clases de datos o conceptos. El modelo se obtiene a partir del análisis de un conjunto de datos de entrenamiento, donde se conocen previamente las etiquetas de la clase. Después, el modelo creado se utiliza para predecir la etiqueta de clase de los nuevos registros, donde se desconoce la clase.
- **Regresión:** es empleado con gran frecuencia en la industria para realizar predicciones numéricas sobre los datos, principalmente hace uso de técnicas estadísticas. Además, el

análisis de regresión es útil para identificar tendencias de distribución basadas en los datos existentes.

2.4. Modelos de recomendación de productos

Estos modelos utilizan una serie de cálculos y algoritmos inteligentes que aprenden gradualmente el comportamiento de un cliente y predicen cuáles son sus preferencias. Para realizar la recomendación, la inteligencia artificial toma en cuenta gran cantidad de datos como las consultas de búsqueda, historial de visualización, compras pasadas, artículos colocados en el carrito de compras, ubicación, segmentos de clientes y personas compradoras (Brita Inteligencia Artificial, 2020)

Los principales beneficios obtenidos a partir de las recomendaciones de productos de comercio electrónico basadas en inteligencia artificial son los siguientes:

- Los comercios pueden obtener mayores ganancias y lealtad por parte de sus clientes, gracias a la experiencia personalizada.
- Mayores tasas de conversión y valor medio del pedido.
- Personalización de los canales de comercio electrónico: los sistemas recomendadores permiten mostrar los productos adecuados a los clientes en sitios web y boletines informativos tomando en cuenta sus intereses y preferencias.
- Las empresas pueden promocionar toda su gama de productos y registrar los datos del comportamiento del usuario a medida que se den.
- Estas soluciones también permiten a los profesionales de *marketing* priorizar los gastos de acuerdo con la manera en la que actuarán los consumidores.

2.4.1. Demographic Filtering

Este modelo proporciona la misma recomendación a toda la audiencia según la popularidad, la calificación o el género del artículo. Este tipo de recomendaciones es común en secciones como *Trending Now* o *Popular*. Esta es una de las formas de filtrado más sencillas: supone que los

artículos que son más populares o aclamados por la crítica tendrán mayores posibilidades de ser del agrado de la audiencia en general (Singla, 2020).

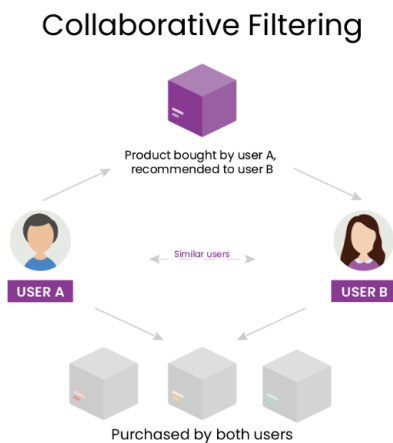
2.4.2. Collaborative Filtering

El más popular de todos los métodos de recomendación de productos. La técnica de filtrado colaborativo se basa únicamente en cómo otros clientes y usuarios han calificado de ante mano un producto que compraron.

El método surge de la idea de que las personas con compras y gustos anteriores similares probablemente querrán usar las mismas cosas en el futuro. Además de esto, confía en las elecciones reales que la gente ha hecho, en lugar de en las calificaciones simples que pueden ser solo una estimación.

Este sistema empareja a personas con intereses similares y proporciona recomendaciones basadas en esta coincidencia. Los filtros colaborativos no requieren metadatos de elementos como en el filtrado basado en contenido. Este tipo de implementación de recomendaciones se puede ver en Amazon o Flipkart en la sección “Comprados juntos con frecuencia” (Brita Inteligencia Artificial, 2020). En la Figura 4, se muestra un ejemplo del funcionamiento del filtrado colaborativo.

Figura 4. Collaborative Filtering

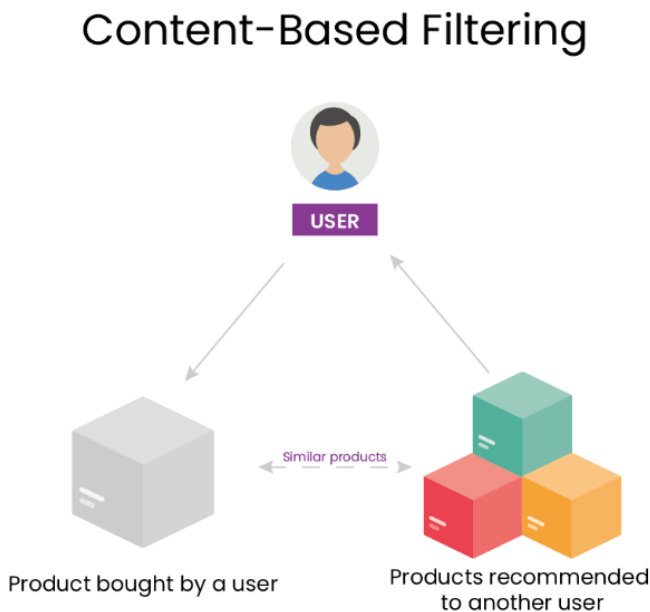


Nota. Tomado de Brita Inteligencia Artificial, 2020.

2.4.3. Content Based Filtering

Esta herramienta sugiere artículos similares, con base en un artículo en particular. Emplea la similitud del elemento con otros guardados previamente, o que le gustaron al usuario en función de los metadatos del elemento o de las palabras clave de búsqueda actuales utilizadas por el usuario para encontrar algún elemento. Esto supone que un usuario tiene mayores posibilidades de que le gusten elementos similares en función de su historial anterior. Esto ayuda a brindarle recomendaciones relevantes en función de las palabras clave buscadas. En la Figura 5 se muestra un ejemplo del funcionamiento del filtrado basado en contenido.

Figura 5. Content Based Filtering



Nota. Tomado de Brita Inteligencia Artificial, 2020.

2.4.4. Técnicas de evaluación de modelos de recomendación

De acuerdo con Deutschman (2022) para evaluar un sistema de recomendación es posible utilizar distintas métricas, dentro de las cuales se encuentran:

- Similitud de coseno: Para calcular la similitud entre un artículo comprado y el artículo nuevo para un sistema centrado en artículos, se calcula el coseno entre 2 vectores que representan esos artículos. La similitud del coseno es la mejor combinación si hay muchas características de alta dimensión, especialmente en la minería de texto.

Figura 6. Similitud de coseno

$$\text{cosine}(x, y) = \frac{\sum_{i=1}^n x_i y_i}{\sqrt{\sum_{i=1}^n x_i^2} \sqrt{\sum_{i=1}^n y_i^2}}$$

Nota. Tomado de Deutschman, 2022

- Distancia Euclideana: La distancia entre dos usuarios en un sistema centrado en el usuario es la longitud de los segmentos de línea que los conectan. El espacio de preferencia son los elementos disponibles y los ejes son los elementos calificados por el usuario. En función de las calificaciones de los usuarios, se buscan elementos que le gusten a usuarios con gustos similares. Cuanto menor sea la distancia entre dos personas, mayor será la probabilidad de que les gusten artículos similares.

Figura 7. Distancia Euclideana

$$d(\text{Person}_i, \text{Person}_j) = \sqrt{(X_1 - X_2)^2 + (Y_1 - Y_2)^2}$$

Nota. Tomado de Deutschman, 2022

- Coeficiente de correlación de Pearson (PCC): Es una medida de la pendiente de la línea que representa la relación entre dos vectores de calificaciones de los usuarios. Puede variar de -1 a 1, 0 significa que no hay correlación lineal. Al usar este enfoque, es posible predecir cómo la persona A calificaría un producto que aún no ha sido calificado. Para hacer eso, simplemente se toma el promedio ponderado de las calificaciones de otros usuarios (incluido el usuario B), donde los pesos se calculan utilizando similitudes de PCC.

Figura 8. PCC

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \times \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}$$

Nota. Tomado de Deutschman, 2022.

2.5. Aprendizaje automático

Esta es una forma de Inteligencia Artificial que permite a los sistemas aprender a través de los datos recolectados. Hace uso de una variedad de algoritmos para mejorar, describir los datos y predecir resultados. Conforme incorporan una mayor cantidad de datos de entrenamiento, es posible producir modelos con resultados más precisos (Hurwitz y Kirsch, 2018).

Han et al. (2012) indican que existen cuatro tipos de aprendizaje automático, los cuales son:

- Supervisado: se cuenta con un conjunto de datos previamente etiquetados, el cual es utilizado posteriormente para enseñar al modelo a realizar las predicciones. A este tipo de aprendizaje también se le conoce como clasificación.
- No supervisado: es conocido como agrupación; a diferencia del aprendizaje supervisado, en este caso no se cuenta con datos previamente etiquetados con las clases

correspondientes. El uso más común en este tipo de aprendizaje es la creación de clústeres para descubrir clases dentro de los datos.

- **Semisupervisado:** utiliza datos etiquetados y no etiquetados para enseñar al modelo. De manera precisa, el conjunto de datos etiquetados se emplea para que el modelo aprenda a clasificar entre las clases y, posteriormente, con ayuda de los datos no etiquetados se refinan los límites entre las clases.
- **Aprendizaje activo:** los usuarios expertos de dominio desempeñan un rol esencial en este tipo de aprendizaje, ya que tienen la posibilidad de etiquetar subconjuntos de datos brindados por el programa de aprendizaje a una clase específica. El fin es optimizar la calidad del modelo, gracias a la adquisición activa de conocimiento de usuarios.

2.5.1. K Means Clustering

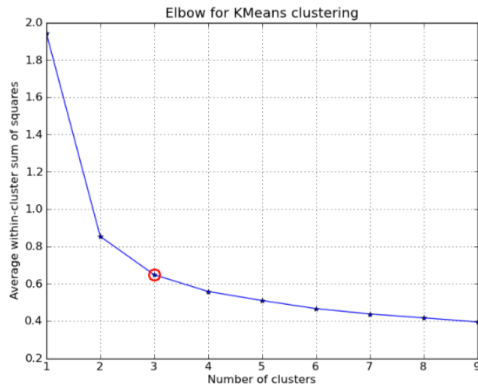
El agrupamiento de K-Means es un algoritmo de aprendizaje no supervisado. En este, no hay datos etiquetados. K-Means realiza la división de objetos en grupos que comparten similitudes y son diferentes a los objetos que pertenecen a otro grupo.

El término “K” es un número. Debe decirle al sistema cuántos clústeres necesita crear. Por ejemplo, $K = 2$ se refiere a dos grupos. Hay una manera de encontrar el mejor u óptimo valor de K para un dato que ya se dio (Kumar, 2021)

Método del codo

El método del codo implica encontrar una métrica para evaluar qué tan bueno es un resultado de agrupación para varios valores de K y localizar el punto del codo. La calidad del agrupamiento mejora de forma rápida al cambiar el valor de K, pero eventualmente se estabiliza. El punto del codo es donde la mejora relativa ya no es muy alta, esto se muestra en la Figura 9.

Figura 9. Método del codo



Nota. Tomado de V. Kumar, 2021.

La técnica de *clustering* K Means, fue una de las técnicas evaluadas durante el proyecto, en la sección 4.2.3 para validar su funcionamiento y verificar su adaptación a las necesidades recolectadas dentro de la sección 4.1.2.

2.5.2. Naive Bayes

Según Kharwal (2020), los clasificadores Naive Bayes se basan en métodos de clasificación bayesianos. Estos se emplean como referencia el teorema de Bayes, que es una ecuación que describe la relación de probabilidades condicionales de cantidades estadísticas. En esta clasificación, se busca encontrar la probabilidad de una etiqueta según algunas características observadas, que se pueden escribir como $P(L | \text{características})$. El teorema de Bayes indica cómo expresar esto en términos de cantidades, las cuales pueden calcularse directamente con la fórmula de la Figura 10.

Figura 10. Fórmula Naive Bayes

$$P(L | \text{features}) = \frac{P(\text{features} | L)P(L)}{P(\text{features})}$$

Nota. Tomado de A. Kharwal, 2020.

Luego de detallar los conceptos necesarios para fundamentar el presente proyecto, en el próximo capítulo se describe el marco metodológico junto con las fuentes y sujetos de información, procedimiento metodológico, variables de investigación y la operacionalización de variables. Esto tiene como fin construir una guía para alcanzar los objetivos planteados en el proyecto.

La técnica de clasificación Naive Bayes fue una de las técnicas evaluadas durante el proyecto, en la sección 4.2.3 para validar su funcionamiento y verificar su adaptación a las necesidades recolectadas dentro de la sección 4.1.2.

2.6. Técnicas de evaluación de modelos de clasificación

De acuerdo con Lahon(2020), algunas métricas de evaluación para modelos de clasificación son:

- *Accuracy*: Define qué tan preciso es el modelo. Por ejemplo. Si predice correctamente 80 muestras correctas de 100 y 20 incorrectamente, entonces el *accuracy* del 80%.
- *Precision*: Cuando se tiene un conjunto de datos que está sesgado, es decir, la cantidad de muestras en una clase supera la cantidad de muestras en otra clase por un margen enorme. No es bueno usar la precisión ya que no es un buen representante de los datos. Es posible que se obtenga una alta precisión en los datos de entrenamiento, pero es posible que el modelo no funcione bien con los datos nuevos. Así que nos fijamos en el *Precision*.
- *F1 Score*: Es una combinación de *Precision* y *recall*. Esta puntuación tiene en cuenta tanto los falsos positivos como los falsos negativos, a diferencia de la precisión y la recuperación, donde solo tomaron uno. La puntuación F1 suele ser más útil que la precisión, especialmente cuando existe una distribución de clases desigual.
- *Recall*: Es un indicador que permite determinar cuantos verdaderos positivos fueron encontrados en una muestra de datos de manera correcta.

2.7. Técnicas de limpieza de datos

Dentro de esta sección se definen las principales técnicas de limpieza de datos implementadas durante la etapa de evaluación de los modelos de aprendizaje automático, para permitir la recomendación personalizada de productos.

2.7.1. One Hot Encoding

De acuerdo con Fawcett (2021), One Hot Encoding es un método que permite transformar los datos categóricos en enteros, para así prepararlos para un algoritmo de *machine learning* y obtener una mejor predicción. Con este método, se convierte cada valor categórico en una nueva columna categórica y a cada una de estas, se le asigna un valor binario de 1 o 0.

2.8. Minería de datos

La minería de datos es un proceso que tiene como objetivo predecir resultados mediante el análisis, la búsqueda de patrones, anomalías y correlaciones sobre grandes conjuntos de datos. En la industria, las empresas comúnmente hacen uso de este tipo de procesos con el objetivo de incrementar sus ingresos, reducir costos y mejorar las relaciones con clientes. Con este proceso, se busca extraer conocimiento útil para las empresas mediante técnicas estadísticas y matemáticas, junto con técnicas de aprendizaje automático y sistemas de bases de datos (Pozas, 2021).

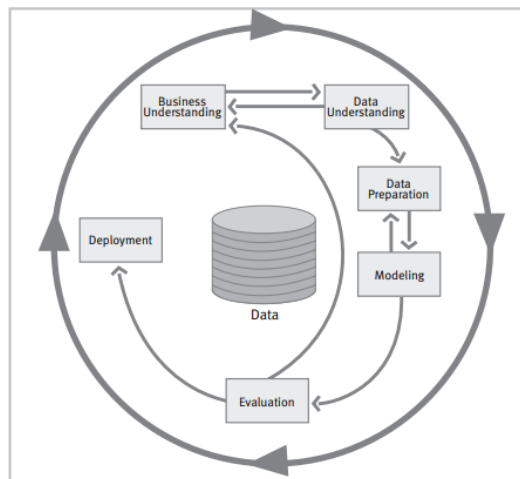
2.8.1. Metodología CRISP-DM

Dentro del proyecto se utilizó como guía una metodología altamente implementada en la industria, de esta manera se pretende hacer uso de las buenas prácticas y artefactos definidos. Lo anterior permite adaptar la solución propuesta a las necesidades reales del negocio y cumplir con los objetivos establecidos dentro de la sección 1.4.

La metodología establecida por Chapman (2000) describe un modelo de proceso estandarizado para proyectos de minería de datos (CRISP-DM). Explica el ciclo de vida de proyectos de este tipo: contiene seis fases principales; para cada una de estas, se describen las tareas, relaciones, salidas y actividades necesarias.

En la Figura 11. Ciclo de metodología CRISP DM, se muestran las fases del proyecto, las cuales se encuentran claramente definidas; sin embargo, la secuencia de estas no es rígida, ya que de ser necesario es posible avanzar y retroceder entre las fases. Adicionalmente, en esa misma figura, se muestra la naturaleza cíclica de la minería de datos: esta no termina al implementar la solución, sino que constantemente evoluciona a causa de la recolección de lecciones aprendidas y el surgimiento de nuevas preguntas comerciales.

Figura 11. *Ciclo de metodología CRISP DM*



Nota. La imagen muestra las principales fases de la metodología CRISP DM. Tomado de *CRISP-DM 1.0*, por P. Chapman 2000.

2.8.1.1. Fase I Entendimiento del negocio

La primera fase del proceso busca comprender los objetivos del proyecto desde una perspectiva empresarial para la definición del problema de minería de datos, así como diseñar un plan preliminar que permita alcanzar los objetivos. Dentro de las principales tareas de esta fase, se encuentran:

- Determinar los objetivos de negocio: comprender desde una perspectiva comercial qué desea lograr el cliente a partir de la minería de datos; se identifican las preguntas que se deben contestar al negocio. Las salidas de esta tarea incluyen:

- Contexto de la organización.
- Objetivos del negocio y lista de preguntas que desea responder.
- Criterios de éxito para el negocio.
- Evaluar la situación: se recolecta información sobre los recursos, restricciones, suposiciones y otros factores que deben considerarse para la ejecución del proyecto. Entre las salidas esperadas de esta tarea se encuentran:
 - Inventario de recursos.
 - Requerimientos, suposiciones y limitaciones.
 - Glosario de terminología relevante para el proyecto.
- Determinar los objetivos de la minería de datos: a partir de los objetivos comerciales, se establecen los objetivos del proyecto en términos técnicos. Las salidas esperadas de esta tarea son:
 - Objetivos de la minería de datos.
 - Criterios de éxito de la minería de datos.
- Producir un plan de proyecto: dentro del plan, se detallan los pasos a seguir durante el proyecto para cumplir los objetivos comerciales y de la minería de datos. Además, se seleccionan preliminarmente las técnicas y herramientas a utilizar durante el proyecto. Algunas salidas de esta tarea son:
 - Plan inicial del proyecto.
 - Evaluación inicial de las herramientas y técnicas a utilizar.

2.8.1.2. Fase II Entendimiento de los datos

Esta fase se centra en la recolección inicial de los datos que serán utilizados para la creación del modelo; posteriormente, se analizan con el fin de conocer los datos, identificar problemas en la calidad de estos, detectar subconjuntos interesantes y formular hipótesis sobre los datos. A continuación, se describen las principales tareas asociadas a esta fase:

- Recolectar los datos iniciales: se recolectan de las distintas fuentes identificadas previamente y se integran en alguna herramienta, con el propósito de comprender y preparar los datos adquiridos. La salida esperada es:
 - Reporte inicial de recolección de datos
- Describir los datos: se examinan las propiedades de los datos adquiridos y se informan sobre los resultados. La salida asociada con esta tarea es:
 - Informe de descripción de los datos existentes junto con sus características.
- Explorar los datos: mediante el uso de herramientas y técnicas de visualización e informes, se examinan los atributos claves y las relaciones entre estos. Adicionalmente, los datos pueden ser sometidos a análisis estadísticos simples con el fin de contribuir a la preparación de los datos. La salida de esta tarea incluye:
 - Informe de exploración de datos con los principales hallazgos
- Verificar la calidad de los datos: se determina si los datos cumplen con los requisitos necesarios para el proyecto y se obtiene respuesta a: ¿los datos están completos?, ¿los datos están correctos o contienen errores?, ¿hay valores faltantes?, entre otras. Dentro de la salida esperada en esta tarea, se encuentra:
 - Reporte de calidad de datos con los problemas encontrados y la creación de una lista de posibles soluciones.

2.8.1.3. Fase III Preparación de los datos

En esta fase, se encuentran todas las actividades de limpieza y transformación necesarias para construir el conjunto de datos final a partir del conjunto de datos original, que posteriormente será utilizado para la creación del modelo de minería de datos definitivo. Algunas tareas incluidas en esta fase son:

- Seleccionar los datos: consiste en decidir cuáles datos serán utilizados para el análisis y modelado. En esta tarea, se descartan tanto registros como atributos, con base en las prioridades del negocio, los objetivos de la minería, la calidad de los datos y las limitaciones técnicas. La salida esperada de esta fase es:

- Lista de datos incluidos y excluidos.
- Limpiar los datos: mejorar la calidad de los datos mediante el uso de técnicas de limpieza, inserción de un conjunto de datos limpios y sustitución aquellos faltantes. La salida esperada de esta fase es:
 - Reporte de limpieza de datos.
- Construir los datos: se crean nuevos atributos, derivados del conjunto de datos original o del conjunto de datos transformado con el propósito de enriquecer el análisis. Como producto de esta tarea, se pueden generar las siguientes salidas:
 - Atributos derivados.
 - Registros generados.
- Integrar datos: combina los datos provenientes de múltiples tablas o registros para así obtener nuevos valores útiles para el modelado. Esta fase tiene como salida:
 - Información combinada proveniente de distintas tablas y registros.
- Formatear datos: consiste en realizar aquellas transformaciones que no cambian el significado de los datos, pero que son necesarias para el modelado y procesamiento de la información. La salida esperada de esta tarea es:
 - Información reformateada.

2.8.1.4. Fase IV Modelado

Durante esta fase, se seleccionan y aplican distintas técnicas de modelado, de tal forma que se definen los parámetros óptimos para obtener los mejores resultados. En esta etapa, podría ser necesario volver a la fase de preparación de datos, ya que en algunos casos las técnicas seleccionadas tienen requisitos específicos sobre la forma de los datos.

- Seleccionar la técnica de modelado: según el problema que se debe resolver, se selecciona la técnica real de modelado en la herramienta elegida previamente. En caso de ser necesario, se seleccionan múltiples técnicas para evaluar distintos escenarios. Las salidas de esta tarea son:

- Técnica de modelado seleccionada.
- Suposiciones de modelado.
- Generar diseño de pruebas: se definen mecanismos que permitan probar la calidad y validez del modelo; por ejemplo, en tareas de clasificación, es necesario definir indicadores para medir la calidad del modelo, y separar los datos en subconjuntos de entrenamiento y pruebas. La salida de esta tarea es:
 - Diseño de pruebas.
- Construir el modelo: se aplica el modelo dentro de la herramienta y se hace uso del conjunto de datos transformado. Dentro de las salidas de esta tarea, se encuentran:
 - Configuración de parámetros.
 - Modelos de minería generados.
 - Interpretación de los resultados del modelo.
- Evaluar el modelo: el responsable de la minería de datos interpreta los resultados obtenidos y los compara tanto con los criterios de éxitos de la minería de datos como con los objetivos del negocio. En esta tarea, también se comparan los resultados obtenidos en todos los modelos desarrollados, para identificar cuál obtuvo mejores resultados según los criterios de evaluación. Las salidas asociadas a esta tarea son:
 - Evaluación de los resultados obtenidos en el modelo.
 - Revisión y ajuste de los parámetros del modelo.

2.8.1.5. Fase V Evaluación

La etapa previa a la implementación final del modelo creado consiste en la revisión a fondo de este, con el propósito de conocer los pasos ejecutados para construirlo y asegurarse que cumpla con los objetivos establecidos por el negocio. En esta etapa, se determina si existe algún aspecto que no se haya considerado y se discuten los pasos futuros en la implementación del modelo. Seguidamente se detallan las tareas principales de esta etapa:

- Evaluación de los resultados: se valora en qué medida el modelo seleccionado cumple o no con los objetivos comerciales y se determina si existe alguna razón comercial para

clasificarlo modelo como deficiente. Además, se incluyen aspectos que no necesariamente están relacionados con el área comercial, los cuales podrían significar un desafío, hallazgo o sugerencia a futuro. Las salidas esperadas como parte de esta tarea son:

- Evaluación de los resultados respecto a los criterios del negocio.
- Modelos aprobados.
- Revisión del proceso: luego de validar los modelos resultantes que cumplen con los objetivos comerciales, es recomendable revisar de manera exhaustiva el proceso de minería de datos. Esto con el fin de determinar si existen factores que no se hayan tomado en cuenta y si se presentan problemas en la calidad del modelo generado. La salida de esta tarea es:
 - Lista de actividades faltantes y pendientes.
- Determinar los siguientes pasos: el equipo de trabajo decide si finaliza el proyecto y comienza la implementación o, por el contrario, si son necesarias más iteraciones para desarrollar el proyecto. Como salida de esta tarea se encuentran:
 - Lista de acciones futuras.
 - Decisiones tomadas.

2.8.1.6. Fase VI Despliegue

La última etapa consiste en presentar el modelo de manera que el cliente pueda utilizarlo dentro de sus procesos operativos, para así satisfacer la necesidad inicial y facilitar la toma de decisiones. Contempla la implementación del modelo en el ambiente productivo del cliente. Este último debe comprender cuáles acciones son requeridas para hacer uso de los modelos creados, sin embargo, el desarrollo de esta integración puede ser realizado por el mismo cliente o por un analista según las necesidades del caso. Las principales tareas de esta etapa se listan a continuación:

- Desarrollar plan de implementación: se determina una estrategia de implementación basada en los resultados obtenidos en la fase de evaluación.
- Monitoreo y mantenimiento del plan: se define una estrategia de preparación para prevenir y evitar períodos largos donde se haga uso incorrecto de los resultados de la minería de datos, principalmente en negocios que los utilizan en su operación diaria.

- Producir el reporte final: se escribe un reporte con las lecciones aprendidas y se presentan los resultados finales al cliente.

2.9. Ingeniería de requerimientos

De acuerdo con Sommerville (2007), los requerimientos de un sistema son “las descripciones de los servicios proporcionados por el sistema y sus restricciones operativas” (p.118). Estos requisitos reflejan las necesidades de los clientes respecto a un sistema que ayude a resolver algún problema empresarial. Al proceso de descubrimiento, análisis, documentación y verificación de estos requisitos se le denomina ingeniería de requerimientos (IR).

Dentro de esta sección del marco conceptual se detalla la clasificación de requerimientos y especificación de los mismo. Esta información será utilizada posteriormente como insumo para la identificación de los requerimientos necesarios para la propuesta.

2.9.1. Clasificación de requerimientos

Sommerville (2007) indica que los requerimientos de un sistema pueden ser clasificados en dos categorías principales:

- Requerimientos funcionales: declaraciones de los servicios que el sistema debe proporcionar, indican cómo debe reaccionar este a entradas y situaciones particulares. En algunas ocasiones, pueden establecer lo que el sistema no debe hacer.
- Requerimientos no funcionales: son restricciones sobre las funciones que ofrece el sistema, por lo general, incluye aspectos de tiempo, proceso de desarrollo, estándares, rendimiento, seguridad, usabilidad, eficiencia y otros.

2.9.2. Especificación de requerimientos

Sommerville (2007) sugiere utilizar el lenguaje natural para transformar las necesidades del cliente en requerimientos del sistema, así pues, la especificación consiste en una forma estandarizada de

escribir los requerimientos del sistema. La principal ventaja de este enfoque es que mantiene la expresividad y comprensibilidad del lenguaje natural, al tiempo que asegura un cierto grado de uniformidad en la especificación. Para esta, se sugiere utilizar una plantilla con la siguiente información:

- Descripción de la función especificada.
- Descripción de las entradas y su origen.
- Descripción de las salidas y su destino.
- Descripción de las entidades utilizadas.
- Descripción de la acción esperada.
- Precondiciones y postcondiciones.
- Descripción de efectos secundarios, si aplican.

2.10. Gestión de procesos

2.10.1. Procesos AS IS

Este consiste en definir la situación actual del proceso o de la compañía. Una buena práctica es solicitar al encargado de ejecutar el proceso que relate cómo lo lleva a cabo o mapear las acciones por medio de un cuestionario para recolectar la información. La identificación del proceso *AS IS* permite analizarlo y compararlo con otros procesos similares o con buenas prácticas, para así mejorarlo o evaluarlo (Madison, 2005).

2.10.2. Procesos TO BE

Este tiene como objetivo mostrar la situación futura del proceso, es decir, el estado ideal al cual desea llegar la organización. Los participantes son personas con alta experiencia en el proceso y se encuentran en condiciones para contribuir con la optimización. Para mapear estos procesos, se utiliza la notación de administración de los procesos de negocios, la cual utiliza notación UML (Madison, 2005).

2.10.3. Modelado de los procesos de negocio

El modelado de los procesos de negocio, comúnmente conocido como BPMN, consiste en utilizar un conjunto de elementos gráficos especializados para describir un proceso y de qué manera es realizado en una organización. El modelador es el encargado de incorporar el nivel de detalle que considere oportuno para que todas las partes involucradas comprendan completamente dicho modelo. (White, 2009).

En el Apéndice 10.9, se detallan los elementos básicos comúnmente utilizados dentro de la notación BPMN, estos elementos son utilizados como insumo para la creación de los diagramas del proceso AS IS y TO BE identificados en este proyecto.

2.11. Cuadrante Mágico de Gartner

De acuerdo con Gartner (s.f.), la metodología de investigación del Cuadrante Mágico proporciona un posicionamiento competitivo gráfico de proveedores de tecnología en mercados de rápido crecimiento. Las capacidades críticas recolectadas por Gartner brindan una visión más profunda de la capacidad y la idoneidad tanto de los productos como de los servicios de TI de los proveedores, en función de casos de uso específicos o personalizados.

El Cuadrante Mágico de Gartner es una herramienta empleada en la industria para comprender inicialmente cuáles proveedores de tecnología podrían ser considerados para una oportunidad de inversión específica. Gartner actualiza sus investigaciones de cada Cuadrante Mágico con cada uno de los distintos mercados a lo largo del año, para lo cual toma en cuenta las últimas novedades de los proveedores de tecnología.

Cada Cuadrante Mágico incorpora un gráfico de dos ejes. El vertical representa el conocimiento de mercado, mientras que el horizontal indica la habilidad de ejecución. Esta herramienta proporciona un posicionamiento competitivo gráfico de cuatro tipos de proveedores de tecnología:

- Líderes: se desempeñan bien de acuerdo con la visión actual del mercado y están posicionados adecuadamente para el futuro.

- Visionarios: entienden hacia dónde se dirige el mercado o tienen una visión para cambiar sus reglas, pero su capacidad de ejecución aún es limitada.
- Jugadores de nicho: se centran con éxito en un segmento pequeño o están dispersos, y no innovan por encima de la media ni superan a los demás.
- Retadores o aspirantes: se desempeñan bien en el momento del análisis e incluso pueden dominar un gran segmento, pero no demuestran una buena comprensión de la dirección del mercado.

El Cuadrante Mágico de Gartner les permite a las organizaciones los siguientes beneficios:

- Obtener rápidamente información sobre los proveedores de tecnología que compiten en el mercado y su capacidad para cumplir con lo que los usuarios finales necesitan en la actualidad y en un futuro.
- Comprender cómo los proveedores de tecnología de un mercado están posicionados de manera competitiva y las estrategias que utilizan para competir por el negocio del usuario final.
- Comparar las fortalezas y desafíos de un proveedor de tecnología con las necesidades específicas de la organización.

El Cuadrante Mágico de Gartner es utilizado a lo largo del proyecto como insumo de información para identificar y sugerir las herramientas de almacenamiento y minería de datos necesarias para la propuesta ya que permite identificar las fortalezas, debilidades de cada herramienta y los líderes del mercado dentro de cada segmento.

2.11.1. Magic Quadrant for Cloud Database Management Systems

De acuerdo con Cook et al. (2021), Gartner define el mercado de *Database Management Systems* (DBMS) en la nube de la siguiente manera: las principales capacidades son el suministro completo por parte de los proveedores de sistemas de software de nube pública o privada, administrados por el proveedor de datos en el almacenamiento en la nube. Los datos se guardan en un nivel de almacenamiento en la nube (como un almacén de objetos en la nube o de datos distribuidos, u otra

infraestructura de almacenamiento en la nube patentada). Opcionalmente, pueden atender múltiples modelos y tipos de datos: relacionales, no relacionales (documento, valor clave, columna ancha, gráfico), geoespaciales, series temporales y otros.

Estos DBMS reflejan estrategias de optimización diseñadas para admitir transacciones y/o procesamiento analítico para uno o más de los siguientes casos de uso:

- Procesamiento de transacciones tradicional y aumentado.
- Almacén de datos tradicional y lógico.
- Exploración de ciencia de datos/aprendizaje profundo.
- Procesamiento de secuencias/eventos.
- Inteligencia operativa.

Este mercado no incluye proveedores que solo brindan DBMS alojados en infraestructura como servicio (IaaS), como en una máquina virtual o contenedor ni administrados por el cliente.

2.11.2. Magic Quadrant for Data Science and Machine Learning (DSML) platforms

De acuerdo con Gartner (2021), este informe evalúa a 20 proveedores de plataformas que pueden ser empleados por los científicos de datos para crear modelos, hacer operativo el aprendizaje automático y obtener datos. Les ayudará a tomar la decisión correcta en un campo abarrotado de plataformas DSML en proceso de maduración y que muestra un rápido desarrollo de productos.

Gartner (2021) define una plataforma de ciencia de datos y aprendizaje automático (DSML) como un producto central y una cartera de soporte de productos, componentes, bibliotecas y marcos integrados coherentemente (incluidos los propietarios, los de socios y los de código abierto). Sus usuarios principales son profesionales de la ciencia de datos, incluidos científicos de datos expertos, científicos de datos ciudadanos, ingenieros de datos, desarrolladores de aplicaciones y especialistas en aprendizaje automático (ML).

Las plataformas DSML cuentan con las siguientes características:

- Ofrecen una combinación de funciones básicas y avanzadas esenciales para crear soluciones DSML (principalmente modelos predictivos y prescriptivos).
- Dan soporte a la incorporación de estas soluciones en los procesos comerciales, la infraestructura circundante, los productos y las aplicaciones.
- Admiten el consumo sostenible de conocimientos derivados de la plataforma y ofrecen funcionalidad para cuantificar y rastrear el valor de los proyectos de ciencia de datos.
- Admiten profesionales de la ciencia de datos con diversas habilidades ("científico de datos" es un título de trabajo y una distinción profesional aplicados de manera inconsistente): la base de usuarios de una plataforma DSML a menudo está compuesta por profesionales con diversos antecedentes técnicos y comerciales.
- Admiten múltiples tareas a lo largo del ciclo de vida de la ciencia de datos, las cuales incluyen:
 - Comprensión del contexto empresarial y del problema.
 - Ingestión de datos.
 - Preparación de datos.
 - Exploración de datos.
 - Ingeniería de características.
 - Creación y formación de modelos.
 - Pruebas de modelos.
 - Despliegue.
 - Seguimiento.
 - Mantenimiento.
 - Gobernanza de datos y modelos.
 - Inteligencia artificial explicable (XAI, siglas en inglés).
 - Seguimiento del valor comercial.
 - Colaboración.

Propuesta de solución de minería de datos basada en la metodología CRISP-DM utilizando los datos almacenados en Oracle Engagement Cloud, para la recomendación personalizada de productos a clientes, durante el primer semestre de 2022

3. Marco metodológico

3.1. Tipo de investigación

De acuerdo con la Pimienta (2012), en la actualidad existen dos tipos de investigación, la básica y la aplicada. A continuación, se describe brevemente cada una y se indica el tipo seleccionado.

La investigación básica tiene como objetivo principal la búsqueda del conocimiento, sin considerar directamente sus posibles aplicaciones prácticas. Pretende tanto ampliar como profundizar los conocimientos acerca de la realidad y se enfoca en la construcción de generalizaciones cada vez mayores (esto incluye hipótesis, leyes, teorías) (Pimienta, 2012, p. 9).

No está interesada por un objetivo crematístico, su motivación es la simple curiosidad, el inmenso gozo de descubrir nuevos conocimientos, se dice que es básica por qué sirve de cimiento a la investigación aplicada o tecnológica; y fundamental porque es esencial para el desarrollo de la ciencia (Ñaupás et al., 2013, p. 93).

La investigación aplicada tiene como objetivo principal la búsqueda y consolidación del saber, así como la aplicación de los conocimientos para acrecentar el conocimiento cultural y científico, y la producción de tecnología al servicio de la sociedad (Pimienta, 2012, p. 9).

Está orientada a resolver objetivamente los problemas de los procesos de producción, distribución, circulación y consumos de bienes y servicios de cualquier actividad humana, principalmente de tipo industrial, infraestructura, comercial, comunicacional, servicios, etc. Se llama aplicada porque con base en la investigación básica, pura o fundamental, en las ciencias fácticas o formales, se formulan problemas e hipótesis de trabajo para resolver los problemas de la vida productiva de la sociedad. Se llama también tecnológica porque su producto no es un conocimiento puro, científico, sino tecnológico (Ñaupás, 2014, p. 93)

El presente proyecto es una investigación aplicada, por las siguientes razones:

- Busca obtener y consolidar nuevos conocimientos con el fin de brindar una solución tecnológica basada en los descubrimientos.

- Pretende resolver una problemática dentro de la organización, relacionada a la mejora del proceso actual de ventas para la recomendación personalizada de productos.
- El producto final de esta investigación no se centra en el conocimiento puro, sino tecnológico.

3.2. Enfoque de la investigación

De acuerdo Hernández et al. (2010), existen dos enfoques de investigación: cuantitativo y cualitativo. En esta sección, se describen las principales características de cada uno, con el propósito de seleccionar el más adecuado para este proyecto:

El enfoque cuantitativo posee las siguientes características:

- Posee un orden riguroso, debido a su naturaleza secuencial, donde cada etapa depende de la anterior y no es posible saltarse pasos.
- Debe ser lo más objetivo posible, esto implica que el investigador no debe permitir que sus temores, creencias o deseos interfieran en los resultados del estudio ni en el proceso utilizado.
- Surge a partir de una idea que se delimita conforme avanza, posteriormente, se definen objetivos y preguntas de investigación. Luego, a partir de estas últimas, se establecen hipótesis y se determinan las variables, más adelante se diseña un plan para probar las variables establecidas y se miden en un determinado contexto.
- En la etapa de recolección de datos, se utilizan procedimientos estandarizados y aceptados por la comunidad científica para garantizar que los resultados obtenidos sean creíbles y cumplan estándares de validez y confiabilidad.
- Para el análisis de las mediciones obtenidas, se emplean métodos estadísticos que permitan extraer las conclusiones de la investigación.
- La finalidad de este enfoque se centra en la formulación y demostración de teorías. Se intentan generalizar los resultados encontrados sobre una muestra y que estos puedan ser replicados.

El enfoque cualitativo posee las siguientes características:

- Es posible desarrollar preguntas de investigación e hipótesis antes, durante o después de la recolección y el análisis de datos.
- El orden en este tipo de investigación es dinámico, esto quiere decir que la secuencia no siempre debe seguir el mismo orden, puede moverse entre los hechos y la interpretación constantemente.
- Ofrece mayor flexibilidad en el planteamiento inicial del problema; en el enfoque cualitativo, no siempre se han definido por completo las variables de investigación.
- En la mayoría de los estudios cualitativos no se prueban hipótesis, sino que se generan durante el proceso y se perfeccionan conforme se recolectan datos, para construir el resultado de la investigación.
- En un enfoque cualitativo el investigador aplica entrevistas a distintos sujetos, posteriormente, analiza los resultados y, a partir de ellos, saca las conclusiones, con el fin de obtener una perspectiva más general.
- Los métodos de recolección de datos no se encuentran completamente estandarizados ni predeterminados, se busca obtener las perspectivas y puntos de vista de los sujetos de investigación. Las preguntas realizadas por el investigador son más abiertas, además, se obtiene datos a través de lenguaje escrito, verbal y visual.
- Dentro de las técnicas utilizadas para la recolección de datos, se encuentran observación no estructurada, entrevistas, revisión documental, discusión en grupo, evaluación de experiencias personales e historias de vida.
- Este tipo de investigación no pretende generalizar de manera probabilística los resultados en poblaciones más amplias ni obtener muestras interpretativas; en la mayoría de los estudios, no busca ser repetibles.

A continuación, se indican las razones por las cuales el presente proyecto es de tipo cualitativo:

- El proceso utilizado no posee un orden riguroso. Se utiliza como guía la metodología CRISP DM, sin embargo, sus fases son flexibles y permiten retroceder a fases anteriores o moverse entre estas según las necesidades del investigador.
- No se plantean hipótesis relacionadas con la problemática de investigación. En su lugar, el problema es analizado durante el desarrollo del proyecto y los resultados se construyen posteriormente.
- Las técnicas de recolección de datos relacionados con la problemática, en su mayoría, son entrevistas, revisión documental y evaluación de experiencias anteriores. Los métodos de recolección de datos buscan ser abiertos, sin hacer uso de métodos estadísticos o estandarizados.
- Los resultados de este proyecto no pretenden ser repetibles ni generalizados de manera probabilística por otros investigadores, debido a las particularidades que presentan las herramientas utilizadas por los clientes de XUM TECH, la información disponible y el contexto empresarial.

3.3. Alcance de la investigación

Según Hernández et al. (2010), existen cuatro tipos de alcance para la investigación cualitativa:

- Exploratorio: se realiza cuando el objetivo es examinar un tema o problema de investigación poco estudiado, del cual se tienen muchas dudas o no se ha abordado antes. Es decir, cuando la revisión de la literatura reveló que tan solo hay guías no investigadas e ideas vagamente relacionadas con el problema de estudio, o bien, si se desea indagar sobre temas y áreas desde nuevas perspectivas.
- Descriptivo: busca especificar las propiedades, las características y los perfiles de personas, grupos, comunidades, procesos, objetos o cualquier otro fenómeno que se someta a un análisis. Es decir, solo pretende medir o recoger información de manera independiente o

conjunta sobre los conceptos o las variables a las que se refieren; por lo tanto, su objetivo no es indicar cómo se relacionan estas.

- **Correlacionales:** asocian variables mediante un patrón predecible para un grupo o población. Su finalidad es conocer la relación o grado de asociación que existe entre dos o más conceptos, categorías o variables en una muestra o contexto en particular. En ocasiones, solo se analiza la relación entre dos variables, pero con frecuencia se reconocen vínculos entre tres, cuatro o más variables.
- **Explicativos:** pretenden establecer las causas de los sucesos o fenómenos que se estudian. Van más allá de la descripción de conceptos o fenómenos, o del establecimiento de relaciones entre conceptos; es decir, están dirigidos a responder por las causas de los eventos y fenómenos físicos o sociales. Su interés se centra en explicar por qué ocurre un fenómeno y en qué condiciones se manifiesta, o por qué dos o más variables se relacionan.

El alcance seleccionado para esta investigación corresponde al exploratorio, por los siguientes motivos:

- El tema central de la propuesta corresponde es la minería de datos, y este tipo de soluciones han sido poco estudiadas dentro de la organización y nunca se han implementado. Por lo tanto, no existe una base de conocimiento previa o algún tipo de guía que oriente al investigador.
- Los entregables generados para cada uno de los objetivos del presente proyecto se basan en la examinación de la problemática, y en la revisión de la documentación disponible en los sitios web oficiales, investigaciones similares y literatura.
- No se identificaron previamente los requerimientos, herramientas, metodologías y datos necesarios para la creación de un modelo de minería de datos orientado a la recomendación de productos.

3.4. Diseño de la investigación

Este término se refiere “al plan o estrategia concebida para obtener la información que se desea con el fin de responder al planteamiento del problema” (Hernández et al., 2014, p.128).

- Teoría fundamentada: las preguntas se centran en los procesos y las relaciones entre los conceptos que conforman el fenómeno de investigación. El producto final de este diseño de investigación consiste en una teoría que explique un fenómeno o responda al planteamiento inicial.
- Etnográfico: las preguntas buscan explorar las características, estructura y funcionamiento de un sistema social (grupo, organización, comunidad, subcultura, cultura), desde una familia, hermandad o hinchada hasta una megaciudad. El producto final más común es la descripción y explicación de un sistema social.
- Narrativa: las preguntas se enfocan en comprender una sucesión de eventos, mediante historias o experiencias de vida siguiendo una secuencia cronológica. Algunos eventos comunes analizados con este tipo de diseño pueden ser catástrofes, elecciones o biografía de individuos. El resultado esperado es una historia secuencia que integra varias narrativas.
- Fenomenológico: las preguntas buscan conocer la esencia de las experiencias: lo que varias personas experimentan en común respecto a un fenómeno o proceso. El producto se centra en la descripción de un fenómeno y la experiencia común de varios participantes con respecto a este.
- Investigación-acción: las preguntas tratan sobre problemáticas o situaciones de un grupo o comunidad específica. El producto final consiste en un diagnóstico de la problemática y un programa o proyecto para resolverla (soluciones específicas).

El presente proyecto utiliza el diseño de Investigación-acción:

- Se analiza una situación problemática actual de la organización XUM TECH. Para esto, se utilizan instrumentos y técnicas que permiten extraer información a partir de las experiencias y el conocimiento adquirido por los colaboradores de la organización.
- Los entregables asociados al proyecto se centran en dar una solución al problema, mediante las tecnologías de información.

3.5. Fuentes de investigación

Fuentes primarias: contienen información original, es decir, son de primera mano. Corresponden a la consecuencia de ideas, conceptos, teorías y resultados de investigaciones. Poseen información directa, antes de ser interpretada o evaluada por otra persona. Las principales fuentes de información primaria son los libros, monografías, publicaciones periódicas, documentos oficiales o informe técnicos de instituciones públicas o privadas, tesis, trabajos presentados en conferencias o seminarios, testimonios de expertos, artículos periodísticos, videos documentales y foros. En la **Tabla 7**. Fuentes de información primaria, se enlistan las fuentes primarias de información utilizadas en este proyecto y su importancia.

Tabla 7. Fuentes de información primaria

Fuentes de información primaria	
Fuente	Importancia para la investigación
Framework de metodología CRISP DM	Estándar de la industria utilizado como guía para el desarrollo de proyectos de minería de datos. Marco de referencia empleado para identificar las necesidades del negocio, transformar los datos disponibles y desarrollar los modelos de minería.
Centro documentación Oracle	Documentación técnica, guías de usuario y referencias relacionadas con las herramientas del ambiente Oracle, integraciones, capacidades, y compatibilidad de los sistemas.
Centro de documentación de proveedores de servicios	Contiene documentación oficial sobre las características, funcionalidades, rendimiento, y costos de los servicios y herramientas necesarias para el desarrollo de la propuesta de minería de datos.

Fuentes de información primaria	
Fuente	Importancia para la investigación
(análisis, procesamiento y almacenamiento de datos)	Adicionalmente, son útiles como insumo para el desarrollo del análisis financiero de la propuesta.
Literatura	Bases teóricas y fundamento para el desarrollo del proyecto, relacionada con temas de <i>Machine Learning</i> , administración, metodología de la investigación, <i>marketing</i> , entre otros.

Nota. Elaboración propia.

Fuentes secundarias: poseen información procesada, que proviene de una fuente primaria. El procesamiento de esta información se pudo dar por una interpretación, un análisis, así como por la extracción y reorganización de los datos de la fuente primaria. La lista de fuentes secundarias y su importancia para este proyecto se encuentra en la **Tabla 8**.

Tabla 8. *Fuentes de información secundarias*

Fuentes de información secundarias	
Fuente	Importancia para la investigación
Proyectos similares dentro de la organización	Conocer experiencias previas de proyectos similares, para identificar las necesidades expuestas por los clientes y así garantizar que la propuesta logre satisfacerlas.
Proyectos similares externos de TFG	Utilizar la información de las propuestas para conocer distintas formas de resolver un problema similar, revisar la literatura relacionada con la temática de proyecto para usarla como insumo.
Sitios web, blogs	Consultar información de diversos autores sobre las temáticas abordadas en el proyecto.
Artículos científicos y académicos	Verificar las sugerencias de distintos autores y las formas de resolver problemas similares de recomendación de productos a clientes, mediante la utilización de técnicas de minería de datos.

Nota. Elaboración propia.

3.6. Sujetos de investigación

Los sujetos de investigación representan a los interesados o personas que aportan la mayor parte de información para comprender el problema de estudio y plantear la solución de inteligencia de negocios adecuada.

Seguidamente, en la Tabla 9. *Sujetos de* , se enlistan estos, su rol, experiencia y caracterización para el presente proyecto.

Tabla 9. *Sujetos de investigación*

Rol del sujeto	Experiencia en el rol	Caracterización del sujeto	Justificación de la importancia de este sujeto para esta investigación
Presidente de la empresa	5 años	Comunicación y negociación con los clientes para identificar sus necesidades y plantear propuestas basadas en los productos ofrecidos por XUM TECH.	Brindar información sobre las necesidades expuestas por los clientes, para la especificación de los requerimientos de la propuesta.
Consultor experto en área de infraestructura	3 años	Conocimiento de infraestructura para sistemas Oracle y amplia experiencia en implementación de proyectos.	Ofrecer conocimiento y sugerencias sobre los componentes de infraestructura utilizados en soluciones ya implementadas.
Consultor especialista en soluciones de analítica Oracle	3 años	Implementación de soluciones de analítica empresarial, <i>dashboards</i> , reportes en distintas áreas para los clientes de XUM TECH.	Dar información acerca de donde se encuentran alojados los datos necesarios para la propuesta de minería.

Propuesta de solución de minería de datos basada en la metodología CRISP-DM utilizando los datos almacenados en Oracle Engagement Cloud, para la recomendación personalizada de productos a clientes, durante el primer semestre de 2022

Rol del sujeto	Experiencia en el rol	Caracterización del sujeto	Justificación de la importancia de este sujeto para esta investigación
Experto del área comercial	2 años	Conocimiento sobre los montos y aspectos tomados en cuenta para realizar una cotización a clientes de XUM TECH.	Ofrecer información para realizar el análisis financiero de la solución planteada.
Consultora especialista en Oracle Engagement	1 año	Conocimiento en las funcionalidades de ventas, leads, manejo de clientes y productos dentro de la herramienta Oracle Engagement Cloud.	Brindar información acerca de las capacidades y el funcionamiento de las herramientas Oracle ofrecidas por XUM TECH.
Administrador de proyectos	3 años	Experiencia amplia en la gestión de proyectos de consultoría que utiliza productos Oracle.	Aportar conocimientos sobre el uso de las herramientas y experiencia basada en proyectos antiguos.

Nota. Elaboración propia.

3.7. Variables de la investigación

En esta sección, se detallan las variables de investigación asociadas a cada uno de los objetivos específicos definidos en el presente proyecto. En la Tabla 10. *Variables de la Investigación*, se indican las variables identificadas y su importancia.

Tabla 10. *Variables de la Investigación*

Objetivo	Variable	Importancia
1. Analizar las necesidades capturadas por el departamento comercial de XUM TECH, para la identificación de los requerimientos necesarios para la propuesta.	<ul style="list-style-type: none"> Proceso Actual de Recomendación de Productos en CX. 	Determinar el contexto actual de la organización y las capacidades a nivel estándar de la herramienta, para comprender el aporte que podría brindar la solución propuesta en la recomendación personalizada de productos.
	<ul style="list-style-type: none"> Necesidades capturadas por el departamento comercial de XUM TECH. 	Identificar los requisitos que debe cumplir la solución.
2. Evaluar varios modelos predictivos mediante la utilización los datos obtenidos de Oracle Engagement Cloud, para la selección de un modelo adecuado de minería de datos que permita la recomendación personalizada de productos a clientes.	<ul style="list-style-type: none"> Datos disponibles en CX. <ul style="list-style-type: none"> Ubicación. Estructura. 	Ser el insumo principal para la creación del modelo de minería de datos.
	<ul style="list-style-type: none"> Modelos predictivos de recomendación personalizada de productos. 	Identificar modelos de minería de datos capaces de recomendar productos de manera personalizada y seleccionar aquel con mejor rendimiento.

Objetivo	Variable	Importancia
3. Recomendar las herramientas de extracción, almacenamiento y procesamiento de los datos, para la integración de los datos de Oracle Engagement Cloud y el modelo predictivo.	<ul style="list-style-type: none"> Herramientas utilizadas/experiencias. 	Conocer las experiencias anteriores de profesionales de XUM TECH, para identificar herramientas o técnicas útiles para el proyecto.
	<ul style="list-style-type: none"> Herramientas de extracción, almacenamiento y procesamiento de datos para el nuevo proceso de minería de datos. <ul style="list-style-type: none"> Características de la funcionalidad. Costos de la solución. 	Permitir la extracción de los datos de Oracle Engagement Cloud, y garantizar que puedan ser procesados y manipulados para la creación del modelo de minería.
4. Elaborar un análisis financiero de la solución de minería de datos propuesta, para la verificación de la viabilidad del proyecto y la identificación de beneficios financieros y no financieros.	<ul style="list-style-type: none"> Análisis financiero de la propuesta. 	Determinar los beneficios directos e indirectos, así como la estructura de costos para verificar la viabilidad de la solución propuesta.

Nota. Elaboración propia.

3.8. Instrumentos de investigación

En esta sección, se describen los instrumentos de investigación utilizados para recolectar los datos e insumos que permiten responder las variables de investigación y garantizar el logro de los objetivos planteados.

Entrevista: de acuerdo con Hernández et al. (2010), este instrumento se define como una reunión para conversar e intercambiar información entre una persona (el entrevistador) y otra (el entrevistado) u otras (entrevistados). En el último caso, podría ser una pareja o un grupo pequeño

como una familia o un equipo. A través de las preguntas y respuestas, se logra una comunicación y la construcción conjunta de significados respecto a un tema (Janesick, 1998). Existen tres tipos de entrevista (Hernández, 2014, p. 403):

- Estructurada: “es propia de la investigación cuantitativa, llamada también entrevista dirigida, controlada, o guiada; sigue un plan preestablecido y se realiza de acuerdo con una guía o formulario previamente preparado con preguntas que responde a las hipótesis formuladas” (Ñaupas, 2014, p. 219).
- Semiestructuradas: se basa en una guía no tan formal y rígida, pues permite que el entrevistador pueda introducir algunas preguntas para esclarecer vacíos en la información; esto quiere decir que no todas las preguntas están predeterminadas.
- No estructuradas: este tipo de entrevista es abierta o libre, en el sentido de que el entrevistador tiene autonomía para hacer las preguntas, pero siempre basándose en una guía general de contenido, aunque no específica. Se ajusta a las necesidades de la investigación de tipo cualitativa. Su mayor debilidad radica en que son difíciles de procesar, por cuanto las respuestas son diferentes entre sí y generalmente extensas.

Seguidamente, en la **Tabla 11**. Matriz de entrevistas, se indican los tipos de entrevistas empleadas durante el desarrollo del proyecto. Para cada una se detalla el identificador, el sujeto de investigación, el objetivo de la entrevista y la referencia al apéndice donde se puede consultar la plantilla de la información recopilada.

Tabla 11. *Matriz de entrevistas*

ID	Objetivo de la entrevista	Tipo de entrevista	Sujeto de investigación	Plantilla de referencia
1	Conocer el proceso de venta de productos en CX	Semiestructurada	Consultora especialista en Oracle Engagement Cloud	Apéndice E. Entrevista #1 Proceso de recomendación/venta de productos en CX
2	Conocer los requerimientos y	Semiestructurada	Experto del área comercial	Apéndice F. Entrevista #2 Conocer los

ID	Objetivo de la entrevista	Tipo de entrevista	Sujeto de investigación	Plantilla de referencia
	necesidades del negocio			requerimientos/necesidades del negocio
3	Verificar la lista de requerimientos encontrados	Semiestructurada	Experto del área comercial	Apéndice G. Entrevista #3 Verificar lista de requerimientos encontrados
4	Comprender los datos y conocer donde se encuentran alojados en Oracle Engagement Cloud	Semiestructurada	Consultor especialista en soluciones de analítica Oracle	Apéndice H. Entrevista #4 Entendimiento de los datos, conocer donde se encuentran alojados en Oracle
5	Identificar cuáles herramientas han utilizado antes para este propósito	Semiestructurada	Consultor experto en área de infraestructura	Apéndice I. Entrevista #5 Conocer cuáles herramientas han utilizado para este propósito anteriormente
6	Conocer aspectos a tomar en cuenta durante una cotización de proyectos en XUM TECH.	Semiestructurada	Experto del área comercial	Apéndice J. Entrevista #6 Conocer aspectos a tomar en cuenta durante una cotización de proyectos en XUM TECH.
7	Conocer alternativas de almacenamiento de datos en Oracle Engagement Cloud	Semiestructurada	Consultora experta en Oracle Engagement Cloud	Apéndice BG. Entrevista #7 Almacenamiento de datos en Engagement Cloud

Nota. Elaboración propia.

Revisión documental: consiste en la consulta de documentos, materiales y artefactos diversos con el propósito de entender el fenómeno central de estudio. Ayudan al investigador a conocer los antecedentes de un ambiente, así como las vivencias o situaciones que se producen en él, y su funcionamiento cotidiano y anormal. En la recolección de documentos, registros, materiales y artefactos, un punto muy importante es que el investigador debe verificar que el material sea auténtico y que se encuentre en buen estado (Hernández et al., 2010).

En la Tabla 12. *Matriz de revisión documental*, se describe el detalle de las revisiones documentales realizadas durante el desarrollo del presente proyecto:

Tabla 12. *Matriz de revisión documental*

ID	Objetivo de la revisión	Fuente de información	Plantilla de referencia
1	Identificar proyectos similares de minería de datos enfocados en la recomendación personalizada de productos.	Artículos académicos. Artículos científicos. Sitios web.	Apéndice M. Revisión documental proyectos de minería de datos para la recomendación de productos
2	Obtener información acerca de las características y costos de las herramientas útiles para la propuesta de minería de datos.	Sitios web. Artículos académicos y científicos.	Apéndice N. Revisión documental herramientas de minería de datos

Nota. Elaboración propia.

Observación: esta técnica permite al investigador recolectar información. De esta manera, se obtiene mayor conocimiento acerca del fenómeno de estudio mediante el contacto directo entre el sujeto cognoscente y el objeto de estudio. Otra definición, brindada por Ñaupás et al. (2013), indica que la observación consiste en el registro sistemático y válido de datos e información de los hechos observados. Dentro de los tipos de observación se encuentran:

- Heurística: sirve para encontrar o descubrir problemas e hipótesis científicas, así como para luego formular el proyecto de investigación.
- De comprobación de hipótesis: reúne la información de diferentes modalidades por medio de la observación, para comprobar la hipótesis.
- Directa, no participante: el contacto entre el investigador y el problema solo está mediado por instrumentos que permiten la observación.
- Participante: el investigador participa en la modalidad, al mismo tiempo registra los resultados y datos para la hipótesis.

- No participante: el investigador es un espectador, requiere de instrumentos para registrar los datos e informaciones siguiendo un plan.
- De campo: recolección de muestras para un análisis utilizando diferentes instrumentos que permitan verificar la hipótesis.
- Experimental: se utiliza un método experimental que consiste en examinar atentamente el efecto que produce la manipulación de las variables.
- Documental: lectura de documentos que permiten recolectar información y registrar los datos.

En el desarrollo del proyecto, se plantea utilizar una observación de tipo Directa, no participante. El objetivo es examinar los datos disponibles en el ambiente proporcionado por XUM TECH, para la creación de los modelos de minería de datos. Para el registro de los hallazgos de esta observación, se utilizó el “Apéndice L. Observación datos disponibles en Oracle Engagement Cloud”.

Pruebas: consiste en cambiar las variables para verificar la funcionalidad. Estas pruebas se van realizando paso por paso y en cada actividad para determinar cuáles son las variables aceptables para su ejecución óptima.

En el caso de este proyecto, se utilizó esta técnica de investigación con el propósito de experimentar con distintos modelos de minería de datos, para así determinar el que obtenga mejor rendimiento. La plantilla utilizada para registrar las pruebas realizadas se encuentra en el “Apéndice P. Plantilla de pruebas”.

Tabla comparativa: de acuerdo con UTEL (2019), un cuadro comparativo es una representación gráfica que sirve para presentar datos, de tal manera que se puedan identificar las diferencias al compararlos. Para elaborar esta herramienta, se deben seguir los siguientes pasos:

- Identificar el propósito de la actividad de aprendizaje.
- Determinar los temas a comparar.

- Dibujar una tabla que contenga los temas y elementos que serán comparados, para ordenarlos en función del tema a analizar.
- Seleccionar las variables más sobresalientes del tema.

En este proyecto, la tabla comparativa se utilizó para realizar comparar las herramientas relacionadas con el proceso de minería de datos y así sugerir las herramientas más idóneas. Esta tabla se detalla en el “Apéndice O. Tabla comparativa de herramientas”.

Adicionalmente, la tabla comparativa fue utilizada para comparar los distintos modelos de minería de datos creados y determinar cuál tenía el mejor rendimiento. La descripción de este instrumento se encuentra en el “Apéndice Q. Tabla comparativa de modelos”.

Cotización: consiste en realizar la recolección de costos fijos y variables necesarios para encontrar el valor monetario del proyecto. Adicionalmente, dentro de la cotización se incluyen los costos relacionados con las herramientas sugeridas, los tiempos de desarrollo y la cantidad de recursos necesarios para la implementación de la propuesta.

3.9. Matriz de cobertura de variables vs. diseño de los instrumentos

A continuación, se muestra la Tabla 13. *Matriz de cobertura de variables*, donde se asocia cada uno de los instrumentos de recolección de datos con las variables identificadas a partir de los objetivos específicos del proyecto.

Tabla 13. *Matriz de cobertura de variables*

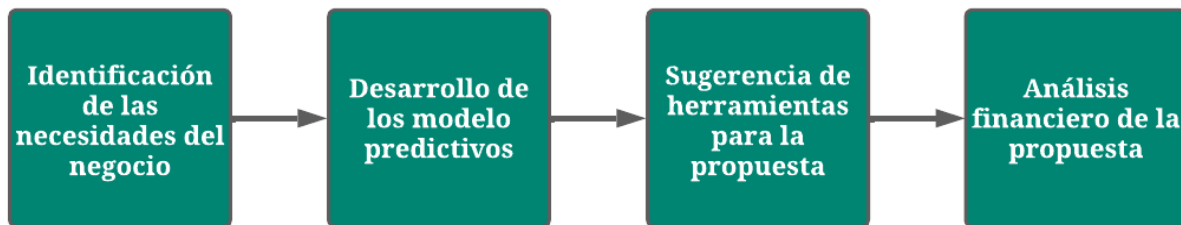
Variable	Entrevistas							Otros instrumentos				
	1	2	3	4	5	6	7	Revisión documental	Pruebas	Cotización	Observación	Tabla comparativa
Proceso actual de recomendación de productos en CX	X											
Necesidades capturadas por Dep. Comercial		X	X					X				
Herramientas utilizadas anteriormente					X		X	X				
Herramientas nuevas para la propuesta							X	X				X
Datos disponibles en CX				X							X	
Tratamiento de datos necesario para modelo								X	X			
Modelos predictivos de recomendación personalizada de productos								X	X			
Análisis financiero de propuesta						X		X		X		

Nota. Elaboración propia.

3.10. Procedimiento metodológico de la investigación

A continuación, la Figura 12. *Fases del procedimiento metodológico* lista las fases metodológicas definidas para este proyecto con base en los objetivos específicos, establecidos en la sección “1.4 Objetivos”. Más adelante, en la sección 3.10.5 se indica la adaptación realizada del procedimiento metodológico con base en la metodología CRISP DM.

Figura 12. *Fases del procedimiento metodológico*



Nota. Elaboración propia.

Seguidamente, se describen las fases del procedimiento metodológico utilizado en este proyecto, junto con el detalle de las actividades necesarias para cumplir cada una de las etapas.

3.10.1. Identificación de las necesidades del negocio

La primera etapa del procedimiento metodológico se enfocó en obtener información acerca de la situación actual y deseada por parte del negocio. Al tratarse de una propuesta de analítica empresarial con los datos almacenados en el sistema Oracle Engagement Cloud, la primera actividad de esta fase consistió en identificar el proceso actual de ventas en el sistema, para así conocer en qué momento se concreta la venta y cuáles datos generados pueden ser útiles para el modelo de minería. La segunda actividad se centró en la recolección de requerimientos necesarios para la propuesta, con el fin de identificar las necesidades y el estado deseado del negocio. A continuación, se describen detalladamente las actividades mencionadas.

Identificación del proceso actual de ventas en CX

Durante esta actividad, se entrevistó a consultores expertos de XUM TECH acerca del proceso de ventas utilizado por los clientes dentro de la herramienta Oracle Engagement Cloud. El propósito fue conocer e identificar el proceso estándar y la forma de gestionar las ventas a través de la herramienta.

Como resultado de esta actividad, se creó un diagrama del proceso *AS IS* de gestión de ventas con la utilización de la herramienta Oracle Engagement Cloud. Para esto, a lo largo del proceso, se identificaron cada una de las actividades y los actores involucrados. El proceso *AS IS* se utilizó como insumo para la creación de la propuesta final del modelo de minería de datos.

Identificación de requerimientos de la propuesta

En esta actividad, se realizó un estudio de la documentación existente acerca de antiguas propuestas de recomendación de productos y se entrevistó a los miembros del área comercial de XUM TECH, con el objetivo de identificar aquellos requerimientos necesarios para la creación de la propuesta de minería de datos. El resultado fue la especificación de requerimientos utilizando el formato establecido en el “Apéndice K. Plantilla especificación de requerimientos”, donde se tomaron en cuenta las necesidades recolectadas por el área comercial y de consultoría de XUM TECH. Adicionalmente, se generó un diagrama del proceso *TO BE*, con el propósito de representar el escenario deseado una vez implementada la propuesta.

3.10.2. Desarrollo de los modelos predictivos

La segunda etapa del procedimiento metodológico se centró en la investigación, creación y aplicación de los modelos de minería de datos seleccionados para evaluar cuál era el más adecuado considerando los datos almacenados en el sistema Oracle Engagement Cloud. En esta etapa, como se mencionó en la sección “1.9. Exclusiones del proyecto”, no se toman en cuenta las herramientas sugeridas por el investigador. A continuación, se detalla la lista de actividades incluidas en esta etapa.

Revisión de datos disponibles en CX

En esta actividad, se consultaron los datos disponibles en el ambiente de Oracle Engagement Cloud brindado por XUM TECH, con el propósito de obtener datos de prueba relacionados con la información de los clientes y las compras realizadas, para evaluar mediante una simulación los modelos de minería de datos propuestos. Los instrumentos utilizados fueron entrevistas a consultores expertos del área de analítica de XUM TECH, enfocadas en ubicar la información disponible en Oracle Engagement Cloud, para extraerla.

Investigación de modelos predictivos de recomendación personalizada de productos

Durante esta actividad, se llevó a cabo una revisión documental de artículos académicos, proyectos similares y literatura para identificar modelos de minería de datos adecuados que permitieran realizar la recomendación personalizada de productos a clientes, con base en la información disponible en Oracle Engagement Cloud. Los modelos de minería encontrados y los resultados se utilizaron como guía para la creación del modelo de minería de datos propuesto.

Sugerencia de tratamiento y transformación de los datos

Los datos recolectados de Oracle Engagement Cloud se examinaron de manera detallada mediante la técnica de observación. El propósito fue identificar las transformaciones necesarias que permitieran aplicar los modelos de minería de datos seleccionados. Asimismo, se seleccionaron y descartaron datos para la creación del modelo. Otros instrumentos utilizados fueron las pruebas y la revisión documental de proyectos similares, así como literatura relacionada con minería de datos.

Creación y prueba de modelos de minería de datos

Seguido de la transformación y preparación de datos, se realizó la creación y evaluación de los modelos de minería de datos seleccionados para verificar el rendimiento y comparar los resultados obtenidos. Posteriormente, se evaluaron los resultados frente a los objetivos planteados. Como resultado, se obtuvo el modelo de minería de datos con el mejor rendimiento.

3.10.3. Sugerencia de herramientas para la propuesta

La tercera etapa del procedimiento metodológico consistió en sugerir herramientas que le permitan a XUM TECH recomendar productos de forma personalizada y ágil, que se integren con los datos almacenados en Oracle Engagement Cloud. A continuación, se detallan las actividades de esta fase.

Revisión de herramientas utilizadas antes

Luego de completar las fases y actividades mencionadas, se revisaron las herramientas empleadas por los consultores de XUM TECH en proyectos anteriores. Esto se hizo mediante entrevistas, para verificar la existencia de herramientas o servicios disponibles en XUM TECH, de forma tal que permitan integrar la solución de minería de datos propuesta, y facilitar las tareas de extracción, transformación, procesamiento y almacenamiento de información.

Investigación de nuevas herramientas para la propuesta

En esta actividad se realizó una revisión documental en la web, literatura y documentos académicos, para sugerir las herramientas necesarias para la implementación de la propuesta de minería de datos. Como resultado, se obtuvo una tabla comparativa entre las herramientas identificadas y la lista definitiva de las herramientas sugeridas, basada en las características, costos y alineamiento con las necesidades del proyecto.

3.10.4. Análisis financiero de propuesta

Finalmente, la etapa de análisis financiero consistió en la evaluación de la factibilidad de la propuesta desde la perspectiva de XUM TECH, esto incluye la cotización del trabajo de investigación realizado durante el TFG y de la implementación de la solución propuesta. Dentro de las actividades realizadas, se encuentran la identificación de costos asociados, y la determinación de beneficios financieros y no financieros. A continuación, se describen detalladamente estas actividades.

Propuesta de solución de minería de datos basada en la metodología CRISP-DM utilizando los datos almacenados en Oracle Engagement Cloud, para la recomendación personalizada de productos a clientes, durante el primer semestre de 2022

Identificación de costos asociados

En esta actividad, se determinaron los costos asociados con la propuesta de solución, se realizó una revisión documental y, adicionalmente, se entrevistó a expertos del área comercial de XUM TECH, para conocer los aspectos que deben ser considerados a la hora de realizar una cotización.

Determinación de beneficios financieros y no financieros

Durante esta actividad, mediante el instrumento de revisión documental, se identificaron los beneficios tanto financieros como no financieros de la propuesta planteada por el investigador. Se emplearon indicadores financieros para determinar la viabilidad de la solución propuesta.

Creación de la hoja de ruta

Adicional al análisis financiero, se realizó una propuesta de hoja de ruta para la implementación, donde se indican los elementos que deben ser configurados en cada ambiente, los recursos necesarios y la cantidad estimada de duración de cada tarea basada en los resultados obtenidos durante la elaboración del proyecto.

3.10.5. Adaptación de la metodología CRISP-DM

Basado en las fases definidas dentro de la metodología del proyecto, en la Tabla 14 se describe detalladamente cuál sección proyecto se encuentra asociada a las fases de la metodología CRISP DM, adicionalmente se hace referencia a los elementos desarrollados como resultado de las actividades.

Tabla 14. Adaptación de la metodología CRISP DM

Fase de CRISP DM	Adaptación dentro del documento		
	Capítulo	Sección	Referencias
Entendimiento de negocio	Análisis de resultados	<ul style="list-style-type: none"> Identificación del proceso actual de ventas en CX (4.1.1) Identificación de requerimientos de la propuesta (4.1.2) Investigación de modelos predictivos para la recomendación personalizada de productos (4.2.2) Definición de criterios de evaluación para los modelos de minería (4.1.3) 	Figura 13
	Propuesta de solución	<ul style="list-style-type: none"> Sugerencia de herramientas para la propuesta. (4.3) 	Tabla 28 Tabla 30 Tabla 32
Entendimiento de los datos	Análisis de resultados	<ul style="list-style-type: none"> Revisión de datos disponibles en CX. (4.2.1) 	Tabla 20
	Propuesta de solución	<ul style="list-style-type: none"> Extracción de datos de Oracle Engagement Cloud (5.2.1) 	Tabla 34
Preparación de los datos	Análisis de resultados	<ul style="list-style-type: none"> Evaluación de los modelos de minería de datos (4.2.3) 	No aplica
	Propuesta de solución	<ul style="list-style-type: none"> Tareas de transformación y limpieza Generación archivo CSV (5.2.2) 	No aplica
Modelado	Análisis de resultados	<ul style="list-style-type: none"> Evaluación de los modelos de minería de datos (4.2.3) <ul style="list-style-type: none"> K Means Naive Bayes Filtrado colaborativo 	Apéndice AH. Evaluación del modelo Naive Bayes (Bernoulli) Apéndice AF. Evaluación del modelo KMeans Apéndice AE. Evaluación del

Fase de CRISP DM	Adaptación dentro del documento		
	Capítulo	Sección	Referencias
			modelo Demographic Filtering Apéndice AF. Evaluación del modelo Collaborative Filtering Apéndice AG. Evaluación del modelo Naive Bayes (Gaussian)
	Propuesta de solución	<ul style="list-style-type: none"> Modelo seleccionado (5.2.2) 	Tabla 21 Tabla 22
Evaluación	Análisis de resultados	<ul style="list-style-type: none"> Evaluación de los modelos de minería de datos (4.2.3) 	Tabla 23
	Propuesta de solución	<ul style="list-style-type: none"> Modelo seleccionado (5.2.2) 	Figura 4
	Conclusiones	<ul style="list-style-type: none"> Conclusiones para el objetivo específico dos (6.2) 	No aplica
	Recomendaciones	<ul style="list-style-type: none"> Recomendaciones (7) 	No aplica
Despliegue	Propuesta de solución	<ul style="list-style-type: none"> Hoja de ruta de la implementación (5.4.4) Creación de objetos y campos (5.2.3) Configuración de archivos de importación (5.2.4) Conversión de recomendación a Lead (5.2.5) 	Figura 21 Figura 22 Figura 25 Figura 26 Figura 27 Figura 29

Nota. Elaboración propia.

3.11. Operacionalización de variables

En esta sección, se detalla la operacionalización de variables en la Tabla 15. *Operacionalización de las variables*, donde se asocian los componentes metodológicos principales del proyecto; esto incluye las fases de la investigación, los objetivos específicos del proyecto, los instrumentos de recolección de datos, las variables y los sujetos de investigación, con el fin de generar una guía que permita alcanzar los objetivos planteados.

Tabla 15. *Operacionalización de las variables*

Fase	Actividad	Objetivo específico	Instrumentos utilizados en la fase	VARIABLES de investigación	Sujetos de investigación
Identificación de las necesidades del negocio.	Identificación del proceso actual de Recomendación de Productos.	1. Analizar las necesidades capturadas por el departamento comercial de XUM TECH, para la identificación de los requerimientos necesarios para la propuesta.	Entrevista #1.	Proceso Actual de Recomendación de Productos en CX.	Consultor especialista en Oracle Engagement.
	Identificación de requerimientos de la propuesta.		Revisión documental. Entrevista #2. Entrevista #3.	Necesidades capturadas por el Departamento Comercial.	Presidente de la empresa. Experto del área comercial.
Desarrollo de los modelos predictivos	Revisión de datos disponibles en CX.	2. Evaluar varios modelos predictivos mediante la utilización de los datos obtenidos de Oracle Engagement Cloud, para la selección de un modelo adecuado de minería de datos que permita la recomendación	Entrevista #4. Observación.	Datos disponibles en CX.	Consultor especialista en Oracle Engagement.
	Investigación de Modelos predictivos de recomendación personalizada de productos.		Revisión Documental #1.	Modelos predictivos de recomendación personalizada de productos.	No aplica.

Fase	Actividad	Objetivo específico	Instrumentos utilizados en la fase	VARIABLES de investigación	Sujetos de investigación
	Sugerencia de tratamiento y transformación de los datos.	personalizada de productos a clientes.	Observación. Pruebas. Revisión. Documental #1.	Tratamiento de datos disponibles.	No aplica.
	Creación y prueba de modelos de minería de datos.		Pruebas. Observación. Tabla comparativa.	Modelos predictivos de recomendación personalizada de productos.	No aplica.
Sugerencia de herramientas para la propuesta.	Revisión de herramientas utilizadas anteriormente.	3. Recomendar las herramientas de extracción, almacenamiento y procesamiento de los datos, para la integración de los datos de Oracle Engagement Cloud y el modelo predictivo.	Entrevista #5. Entrevista #7.	Herramientas utilizadas anteriormente.	Consultor especialista en soluciones de Oracle Engagement. Consultora especialista en Oracle Engagement. Consultor experto en área de infraestructura.
	Investigación de nuevas herramientas para la propuesta.		Revisión documental #2. Tabla Comparativa entre herramientas. Entrevista #7.	Herramientas nuevas para la propuesta.	No aplica.

Fase	Actividad	Objetivo específico	Instrumentos utilizados en la fase	VARIABLES de investigación	Sujetos de investigación
Análisis financiero de la propuesta.	Identificación de costos asociados.	4. Elaborar un análisis financiero de la solución de minería de datos propuesta, para la verificación de la viabilidad del proyecto y la identificación de beneficios financieros y no financieros.	Entrevista #6. Cotización.	Análisis financiero de la propuesta.	Experto del área comercial.
	Determinación de beneficios financieros y no financieros.		Revisión documental.		No aplica.

Nota. Elaboración propia.

3.12. Tabla resumen del procedimiento metodológico de la investigación

Finalmente, en la Tabla 16. Tabla resumen del proceso metodológico, se muestra la relación existente entre los objetivos específicos, los conceptos presentados en el marco conceptual, las fases de la metodología, la propuesta, conclusiones y recomendaciones, así como en los anexos.

Tabla 16. *Tabla resumen del proceso metodológico*

Objetivo específico	Marco conceptual	Marco metodológico	Propuesta	Conclusiones y recomendaciones	Apéndices y anexos
1. Analizar las necesidades capturadas por el departamento comercial de XUM TECH, para la identificación de los requerimientos necesarios para la propuesta.	2.9, 2.10, 2.1, 2.2	3.10.1	5.1	6.1, 7	9.3, 9.4, 9.5, 9.7, 10.1, 9.11, 9.18, 9.19
2. Evaluar varios modelos predictivos mediante la utilización de los datos obtenidos de Oracle Engagement Cloud, para la selección de un modelo adecuado de minería de datos que permita la recomendación personalizada de productos a clientes.	2.8, 2.3	3.10.2	5.2, 5.3	6.2, 7	9.8, 9.12, 9.13, 9.16, 9.20, 9.23, 9.24, 9.31, 9.32, 9.33, 9.34, 9.35, 9.50, 9.51, 9.52, 9.53, 9.54, 9.55, 9.56, 9.57, 9.58.
3. Recomendar las herramientas de extracción, almacenamiento y procesamiento de los datos, para la integración de los datos de Oracle Engagement Cloud y el modelo predictivo.	2.11	3.10.3	5.2	6.3, 7	9.14, 9.15, 9.9, 9.21, 9.25, 9.26, 9.27, 9.28, 9.29, 9.30, 10.2, 10.3, 10.4, 10.5, 10.6, 10.7, 9.61, 9.59, 9.60
4. Elaborar un análisis financiero de la solución de minería de datos propuesta, para la verificación de la viabilidad del proyecto y la identificación de beneficios financieros y no financieros.	2.8.1	3.10.4	5.4	6.4, 7	9.10, 9.22

Nota. Elaboración propia.

4. Análisis de resultados

En este capítulo, se lleva a cabo el análisis de los resultados obtenidos mediante la aplicación de los instrumentos de recolección de información, que se especificaron en la sección de Marco. El análisis de resultados está conformado por cuatro fases, definidas en la sección “**Procedimiento metodológico de la investigación**”. Para obtener la información requerida, se emplearon entrevistas, observaciones y revisiones documentales.

A continuación, se describen los principales hallazgos para cada una de las fases y actividades del procedimiento metodológico definido, luego de aplicar los instrumentos de recolección de datos que sirven como insumo para el diseño de la propuesta de solución.

4.1. Fase I. Identificación de las necesidades del negocio

4.1.1. Identificación del proceso actual de recomendación de productos

Para identificar el proceso actual de ventas y recomendación de productos, se realizó la entrevista “Apéndice R. Resultados de la Entrevista #1” a un consultor experto de Oracle Engagement Cloud. Esta se enfocó en determinar las actividades y a los participantes clave del proceso de venta al emplear la herramienta Oracle Engagement Cloud.

Seguidamente, en la Tabla 17. Participantes del proceso *Participantes del proceso*, se describen a los principales participantes dentro del proceso de ventas actual con Oracle Engagement Cloud.

Tabla 17. *Participantes del proceso*

Participantes	Descripción
Cliente	Muestra interés y adquiere los productos ofrecidos por la organización.
Ejecutivo de ventas	Brinda atención a los clientes de la organización, se encarga de calificar a los clientes potenciales con el fin de vender productos, asegurando que cumplan con los

Participantes	Descripción
	requisitos mínimos establecidos por el negocio.
Oracle Engagement Cloud	Plataforma utilizada para gestionar la relación con los clientes, donde se almacena la información de ellos, los productos adquiridos, las ventas y prospecciones.

Nota. Elaboración propia.

La lista de las principales actividades necesarias para realizar una venta dentro de Oracle Engagement Cloud se detalla en la Tabla 18. *Actividades del proceso.*

Tabla 18. *Actividades del proceso*

Actividad	Participante	Descripción
Mostrar interés en un producto	Cliente	El cliente interactúa con la organización por cualquiera de los canales disponibles (digitales y físicos), para obtener información acerca de un producto específico.
Crear Lead	Ejecutivo de ventas	El ejecutivo de ventas crea una prospección dentro de Oracle Engagement Cloud, donde indica la información del cliente y del producto que desea adquirir.
Completar formularios de precalificación del Lead	Ejecutivo de ventas	El ejecutivo de ventas completa los formularios de precalificación del lead, con

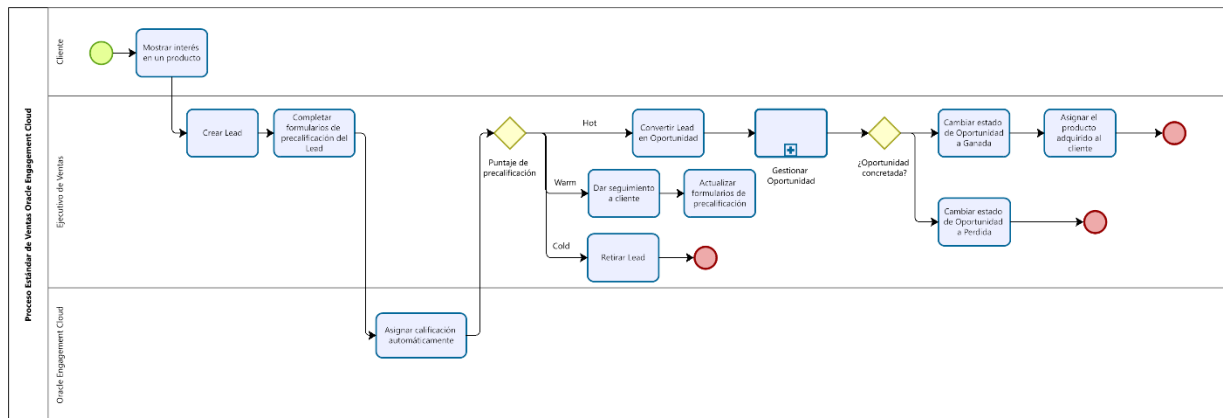
Actividad	Participante	Descripción
		base en la información brindada por el cliente.
Asignar calificación al Lead	Oracle Engagement Cloud	Basado en los resultados del formulario, Oracle Engagement le asigna automáticamente una clasificación (Hot, Warm, Cold).
Convertir Lead en Oportunidad	Ejecutivo de ventas	Cuando el Lead obtiene una buena calificación (el cliente cumple con los requisitos y tuvo un puntaje Warm o Hot) pasa a ser una Oportunidad dentro de Oracle Engagement Cloud, en esta oportunidad se almacena la información proveniente del Lead.
Gestionar oportunidad	Ejecutivo de ventas	Luego de crear la oportunidad, el ejecutivo debe seguir la metodología de ventas definida por el negocio para lograr concretar la venta. Esta metodología puede variar según el tipo de producto.
Cambiar estado de la oportunidad	Ejecutivo de ventas	En caso de concretar la venta, la oportunidad debe pasar a ganada. Cuando un ejecutivo

Actividad	Participante	Descripción
		pierde la oportunidad, pasa a perdida.
Asignar el producto adquirido al cliente	Ejecutivo de ventas	Cuando la oportunidad fue ganada, significa que el cliente adquirió de manera satisfactoria el producto. Por lo tanto, este debe ser asignado al cliente dentro de Oracle Engagement Cloud.

Nota. Elaboración propia.

Finalmente, en la Figura 13. *Proceso AS IS de ventas con Engagement Cloud*, se describe el proceso actual de ventas con la herramienta Oracle Engagement Cloud.

Figura 13. *Proceso AS IS de ventas con Engagement Cloud*



Nota. Elaboración propia.

4.1.2. Identificación de requerimientos de la propuesta

Para completar esta actividad, se realizó una revisión documental de las historias de usuario identificadas previamente por XUM TECH, detalladas en el “Anexo I. Historias de usuario”. En la especificación de los requerimientos, se empleó la plantilla definida en el “Apéndice K. Plantilla especificación de requerimientos”. Para la documentación de estos, se detalla el identificador, el tipo de requerimiento (funcional o no funcional), categoría del requerimiento (modelo de minería, CRM, *Marketing*), historia de usuario asociada y una descripción detallada de la funcionalidad esperada.

Requerimientos del modelo de recomendación

En esta sección, se presenta la especificación de los requerimientos asociados al modelo de recomendación personalizada de productos, estos se utilizarán como base para la construcción de la propuesta de minería de datos.

ID del requerimiento	R001
Tipo de requerimiento	Funcional
Categoría del requerimiento	Requerimiento del modelo de minería
Historia de usuario	HU-01
Descripción de la funcionalidad	El sistema debe recomendar productos considerando el comportamiento de clientes similares, productos adquiridos y relaciones entre las familias de los productos.

ID del requerimiento	R002
Tipo de requerimiento	Funcional
Categoría del requerimiento	Requerimiento del modelo de minería
Historia de usuario	HU-01
Descripción de la funcionalidad	El modelo debe asignar una calificación del 1 al 100, según el nivel de recomendación calculado; donde 1 es el más bajo y 100 el más alto.

ID del requerimiento	R003
Tipo de requerimiento	Funcional
Categoría del requerimiento	Requerimiento del modelo de minería
Historia de usuario	HU-01

Propuesta de solución de minería de datos basada en la metodología CRISP-DM utilizando los datos almacenados en Oracle Engagement Cloud, para la recomendación personalizada de productos a clientes, durante el primer semestre de 2022

Descripción de la funcionalidad	Las recomendaciones generadas por el modelo deben contener el nombre del producto recomendado, la calificación, el monto de venta estimado (si aplica) y la fecha de generación de la recomendación.
---------------------------------	--

ID del requerimiento	R004
Tipo de requerimiento	Funcional
Categoría del requerimiento	Requerimiento del modelo de minería
Historia de usuario	HU-01
Descripción de la funcionalidad	El modelo de recomendación debe permitirle al usuario brindar retroalimentación para su ajuste.

ID del requerimiento	R005
Tipo de requerimiento	Funcional
Categoría del requerimiento	Requerimiento del modelo de minería
Historia de usuario	HU-01
Descripción de la funcionalidad	El sistema debe determinar las tres mejores recomendaciones de productos para cada cliente y excluir los productos rechazados.

Requerimientos de Engagement Cloud

En esta sección, se presenta la especificación de los requerimientos asociados a la herramienta de Oracle Engagement Cloud, el cumplimiento de estos permitirá a los ejecutivos de ventas utilizar las recomendaciones generadas por el modelo dentro de su ambiente productivo.

ID del requerimiento	R006
Tipo de requerimiento	Funcional
Categoría del requerimiento	Requerimiento del CRM
Historia de usuario	HU-01
Descripción de la funcionalidad	El sistema Oracle Engagement Cloud debe almacenar y mostrar las recomendaciones realizadas por el modelo en el módulo de Clientes. Esta información incluye la calificación de la recomendación, el nombre del producto, el monto estimado y la fecha en que se genera la recomendación.

Propuesta de solución de minería de datos basada en la metodología CRISP-DM utilizando los datos almacenados en Oracle Engagement Cloud, para la recomendación personalizada de productos a clientes, durante el primer semestre de 2022

ID del requerimiento	R007
Tipo de requerimiento	Funcional
Categoría del requerimiento	Requerimiento del CRM
Historia de usuario	HU-01
Descripción de la funcionalidad	El sistema Oracle Engagement Cloud debe ofrecer al ejecutivo la opción de aceptar o rechazar la recomendación brindada por el modelo. En caso de aceptarla, se debe convertir en prospectación, pero cuando la rechaza, el ejecutivo debe seleccionar la causa del rechazo.

ID del requerimiento	R008
Tipo de requerimiento	Funcional
Categoría del requerimiento	Requerimiento del CRM
Historia de usuario	HU-01
Descripción de la funcionalidad	El sistema Oracle Engagement Cloud debe ofrecer al ejecutivo la opción de agregar comentarios acerca de la recomendación.

ID del requerimiento	R009
Tipo de requerimiento	Funcional
Categoría del requerimiento	Requerimiento del CRM
Historia de usuario	HU-01
Descripción de la funcionalidad	El sistema Oracle Engagement Cloud debe almacenar la lista de recomendaciones rechazadas junto con el motivo de rechazo, para utilizarlas en futuros análisis y en el modelo.

ID del requerimiento	R010
Tipo de requerimiento	No Funcional
Categoría del requerimiento	Requerimiento del CRM
Historia de usuario	HU-01
Descripción de la funcionalidad	Las recomendaciones de productos generadas deben ser cargadas a Oracle Engagement Cloud cada tres meses (definido por el cliente), por medio de archivos en formato CSV.

ID del requerimiento	R011
Tipo de requerimiento	Funcional

Categoría del requerimiento	Requerimiento del CRM
Historia de usuario	HU-02
Descripción de la funcionalidad	El sistema debe permitir la extracción del registro de recomendaciones rechazadas y aceptadas por los ejecutivos, para asegurar que el modelo pueda mejorar continuamente.

Requerimientos de *Marketing*

En esta sección, se presenta la especificación de los requerimientos asociados a la herramienta de *marketing*, el cumplimiento de estos permitirá enviar información acerca de los productos recomendados a los clientes mediante campañas publicitarias.

ID del requerimiento	R012
Tipo de requerimiento	Funcional
Categoría del requerimiento	Requerimiento de <i>Marketing</i>
Historia de usuario	HU-03
Descripción de la funcionalidad	El sistema debe permitir la creación de campañas de mercadeo, mediante el uso de información generada por el modelo, como los productos recomendados, monto estimado, calificación y estado de la recomendación.

ID del requerimiento	R013
Tipo de requerimiento	No Funcional
Categoría del requerimiento	Requerimiento de <i>Marketing</i>
Historia de usuario	HU-03
Descripción de la funcionalidad	La información utilizada en las campañas de mercado debe estar sincronizada con la información almacenada en Oracle Engagement Cloud.

Posterior a la identificación de los requerimientos, se realizó una entrevista al presidente de la empresa XUM TECH, detallada en el “Apéndice S. Resultados de la Entrevista #3”. Con esta, se validaron todos los requerimientos y se realizaron los ajustes correspondientes. El resultado de esta entrevista se detalla en la Tabla 19. *Resumen de requerimientos*.

Tabla 19. *Resumen de requerimientos*

Cantidad de requerimientos revisados	13
Cantidad de requerimientos aprobados	13
Cantidad de cambios solicitados	1
Cantidad de correcciones realizadas por el investigador	1

Nota. Elaboración propia.

4.1.3. Definición de criterios de evaluación para los modelos de minería

Durante la fase de identificación de las necesidades de los clientes, no se identificaron criterios de evaluación de los modelos, sin embargo, en el caso de los modelos de clasificación, mediante el juicio de experto se establecen los siguientes criterios de evaluación:

- El modelo deberá obtener un mínimo de 80% de precisión.
- El modelo deberá obtener un mínimo de 80% de recall.
- El modelo deberá obtener un mínimo de 75% de F1 score.

4.2. Fase II. Desarrollo de los modelos predictivos

4.2.1. Revisión de datos disponibles en CX

Durante el desarrollo de esta actividad, se realizó una entrevista a un consultor de XUM TECH experto en analítica mediante el sistema Oracle Engagement Cloud. El propósito fue conocer la manera adecuada de consultar y extraer datos del sistema. El detalle se encuentra en el “Apéndice T. Resultados de la Entrevista #4”.

Como resultado, se identificaron dos mecanismos para consultar los datos almacenados en Oracle Engagement Cloud, los cuales se detallan a continuación:

- *Reports & Analytics*: consiste en un módulo dentro de la herramienta de Oracle Engagement Cloud, donde es posible crear análisis, reportes y *dashboards* utilizando múltiples *Subject Areas* (también llamados objetos) al mismo tiempo. También, en caso necesario, es posible filtrar y exportar la información.
- *Export Management*: este módulo de Oracle Engagement Cloud permite a los usuarios y administradores exportar objetos específicos en formato CSV, únicamente es posible exportar un objeto a la vez.

Además, se le consultó al experto acerca de las *Subject Areas* donde actualmente se almacena la información relacionada a las ventas y los clientes registrados en Oracle Engagement Cloud, el detalle de estas se muestra en la *Tabla 20. Subject Areas de información en Oracle Engagement Cloud*.

Tabla 20. *Subject Areas de información en Oracle Engagement Cloud*

<i>Subject Area</i>	Descripción
<i>Leads</i>	Almacena la información de las prospecciones de venta, la cual incluye detalles del cliente y producto. El <i>Lead</i> representa una etapa previa a la ejecución de la venta de productos.
<i>Assets</i>	Almacena la información de los productos asignados a cada cliente dentro de la herramienta.
<i>Opportunities</i>	Almacena la información de las oportunidades (ganadas, perdidas y en proceso), e incluye información del cliente y los productos asociados a cada oportunidad. Las oportunidades ganadas corresponden al registro de las ventas reales concretadas por la organización.

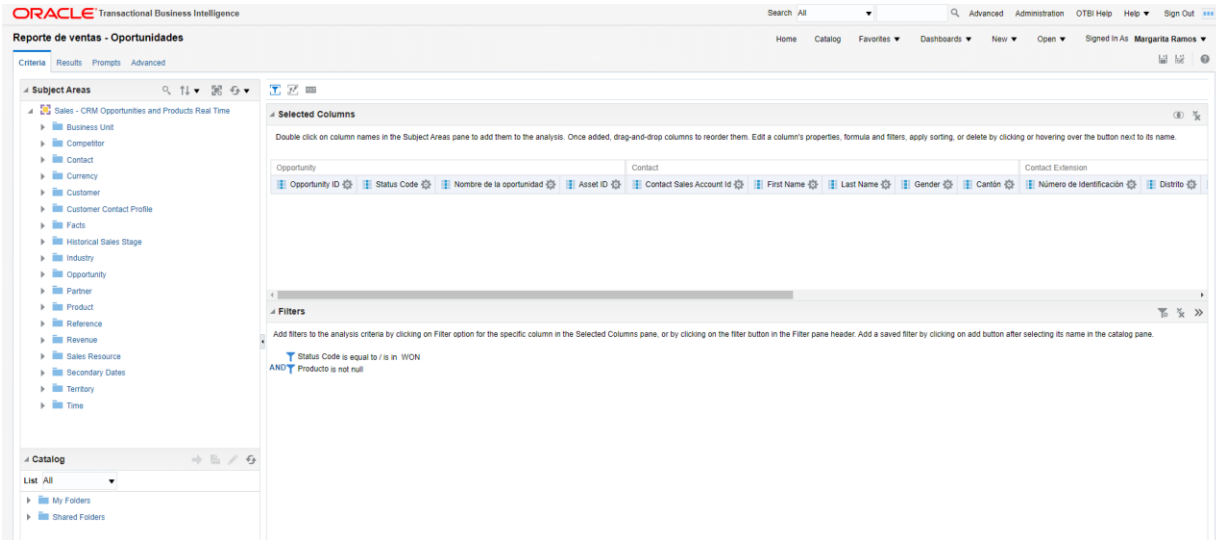
<i>Subject Area</i>	Descripción
<i>Contacts/Accounts</i>	Almacena como datos personales la información relacionada con los clientes. En algunos casos, se almacenan los clientes físicos en el objeto <i>Contact</i> y los jurídicos en el objeto <i>Account</i> .

Nota. Elaboración propia.

Luego de identificar las *Subject Areas* principales, donde se almacena información de las ventas, se determinó que la *Subject Area* a evaluar para la creación del modelo de minería de datos sería *Opportunities*. Esto en razón de que dentro de ella se puede encontrar el registro de las ventas efectuadas, junto con datos del producto y el cliente, los cuales son necesarios para la creación del modelo.

Posteriormente, se aplicó el instrumento de observación detallado en el “Apéndice W. Resultados de la revisión de datos disponibles”, para listar los campos de información disponibles dentro de Oracle Engagement Cloud que se consideran útiles para crear el modelo predictivo, utilizando el módulo de *Reports and Analytics* como se muestra en la Figura 14. *Revisión de datos existentes en Oracle Engagement Cloud.*

Figura 14. Revisión de datos existentes en Oracle Engagement Cloud



Nota. Elaboración propia.

Después de la observación, se creó un reporte personalizado, llamado “Reporte de Ventas-Oportunidades”; los campos se detallan en el “Apéndice W. Resultados de la revisión de datos disponibles”. Finalmente, el reporte fue exportado en formato CSV, como se muestra en la Figura 15. *Resultado de la exportación de datos*, para posteriormente evaluar los distintos modelos predictivos.

Figura 15. Resultado de la exportación de datos

Opportun	Status	Coc	Nombre d	Contact S	Gender	Cantón	Número d	Distrito	Fecha de	Tipo Clie	Estado C	C	Categoría	Tipo de as	Salario Ne	Producto		
3E+14	WON	Tarjeta de	1E+14	Femeninc	SAN CARL	2.05E+08			00:00.0	5.	CONVEI	U.	Union 1.1.	NORMA5.	CUENTA NATURAL	Tarjeta de credito-Black		
3E+14	WON	CDP Desm	3E+14	Masculinc	SAN CARL	8.01E+08			00:00.0	5.	CONVEID.	Divorci	1.	NORMA5.	CUENTA NATURAL	CDP Desmaterializado COL		
3E+14	WON	CDP Desm	3E+14	Masculinc	SAN CARL	8.01E+08			00:00.0	5.	CONVEID.	Divorci	1.	NORMA5.	CUENTA NATURAL	CDP Desmaterializado COL		
3E+14	WON	Credito sc	1E+14	Masculinc	TILARAN	5.01E+08			00:00.0	5.	CONVEI	Casado (a)	1.	NORMA5.	CUENTA NATURAL	Credito sobre CDP COL		
3E+14	WON	Sin Fiador	1E+14	Masculinc	CENTRAL	1.05E+08	San José		00:00.0	1.	REGULA	Casado	1.	NORMA1.	CUENTA REGULAR	Sin Fiador		
3E+14	WON	Tarjeta de	3E+14	Femeninc	POCOCI	7.02E+08			00:00.0	1.	REGULA	U.	Union 1.1.	NORMA1.	CUENTA REGULAR	Tarjeta de credito-Gold		
3E+14	WON	Sin Fiador	3E+14	CENTRAL		1.13E+08			#####		C.	Casado	1.	NORMA1.	REGULAR ACTIVO	:Sin Fiador Especial		
3E+14	WON	Sin Fiador	3E+14	BELEN		2.06E+08			14/11/1982		C.	Casado	1.	NORMA1.	REGULAR ACTIVO	:Sin Fiador Especial		
3E+14	WON	Tarjeta de	3E+14	Femeninc	SAN RAM	1.1E+08			22/07/1971		REGULA	D.	Divorci	1.	NORMA1.	REGULAR ACTIVO	:Tarjeta de credito-Gold	
3E+14	WON	Sin Fiador	3E+14	CENTRAL		3.05E+08			14/10/1991		C.	Casado	1.	NORMA1.	REGULAR ACTIVO	:Sin Fiador		
3E+14	WON	Sin Fiador	1E+14	CENTRAL		2.06E+08			00:00.0		C.	Casado	1.	NORMA1.	CUENTA REGULAR	Sin Fiador Preferencial Ventanilla		
3E+14	WON	Sin Fiador	1E+14	Masculinc	GOICOEC	3.14E+08			00:00.0	1.	REGULA	S.	Soltero	1.	NORMA1.	CUENTA REGULAR	Sin Fiador (Ventanilla)	
3E+14	WON	Sin Fiador	1E+14	CENTRAL		9.01E+08			00:00.0		C.	Casado	1.	NORMA1.	CUENTA REGULAR	Sin Fiador (Ventanilla)		
3E+14	WON	Sobre aho	1E+14	Masculinc	SAN CARL	6.02E+08	Alajuela		00:00.0	1.	REGULA	D.	Divorci	1.	NORMA1.	CUENTA REGULAR	Sobre ahorro corriente	
3E+14	WON	Sin Fiador	1E+14	Femeninc	DESAMPA	1.16E+08	San José		00:00.0		Casado	1.	NORMA1.	CUENTA REGULAR	Sin Fiador			
3E+14	WON	SIN FIADC	1E+14	Femeninc	POCOCI	7.02E+08			00:00.0	1.	REGULA	S.	Soltero	1.	NORMA1.	CUENTA REGULAR	SIN FIADOR VIP	
3E+14	WON	CDP DESM	1E+14	Femeninc	SAN CARL	2.03E+08			00:00.0	5.	CONVEID.	Divorci	1.	NORMA5.	CUENTA NATURAL	CDP DESMATERIALIZADO COL		
3E+14	WON	SOBRE AH	3E+14	Masculinc	CENTRAL	3.03E+08			00:00.0	1.	REGULA	Divorci	1.	NORMA1.	CUENTA REGULAR	SOBRE AHORRO CORRIENTE (VENTANILLA)		
3E+14	WON	CS APOYO	3E+14	Masculinc	CENTRAL	3.03E+08			00:00.0	1.	REGULA	Divorci	1.	NORMA1.	CUENTA REGULAR	CS APOYO FUNERARIO		
3E+14	WON	HIPOOT. PR	3E+14	Masculinc	SAN RAM	2.05E+08			00:00.0	1.	REGULA	C.	Casado	1.	NORMA1.	CUENTA REGULAR	HIPOOT. PROC. CONST. VENTANILLA	
3E+14	WON	AHORRO	1E+14	Femeninc	PEREZ ZEL	1.05E+08			14/07/1951		REGULA	C.	Casado	1.	NORMA1.	REGULAR ACTIVO	:AHORRO MARCHAMO	
3E+14	WON	SOBRE AH	1E+14	Masculinc	GUACIMO	7.02E+08			00:00.0	1.	REGULA	U.	Union 1.1.	NORMA1.	CUENTA REGULAR	SOBRE AHORRO CORRIENTE (VENTANILLA)		
3E+14	WON	AHORRO	1E+14	Masculinc	PALMARE	2.04E+08			00:00.0	5.	CONVEI	C.	Casado	1.	NORMA1.	CUENTA REGULAR	AHORRO PLUS	
3E+14	WON	SIN FIADC	3E+14	Femeninc	DESAMPA	6.04E+08	San José		00:00.0		Divorci	1.	NORMA1.	CUENTA REGULAR	SIN FIADOR VIP	VENTANILLA		
3E+14	WON	SIN FIADC	3E+14	Masculinc	SAN RAM	5.04E+08			00:00.0	1.	REGULA	C.	Casado	1.	NORMA1.	CUENTA REGULAR	SIN FIADOR VIP	
3E+14	WON	SIN FIADC	3E+14	Masculinc	CENTRAL	2.07E+08			00:00.0	1.	REGULA	S.	Soltero	1.	NORMA1.	CUENTA REGULAR	SIN FIADOR VIP	VENTANILLA
3E+14	WON	CDP COLO	1E+14	Masculinc	SAN CARL	2.06E+08			00:00.0	5.	CONVEID.	Divorci	1.	NORMA5.	CUENTA NATURAL	CDP COLONES		
3E+14	WON	SIN FIADC	3E+14	PEREZ ZEL		1.14E+08			00:00.0		C.	Casado	1.	NORMA1.	CUENTA REGULAR	SIN FIADOR (VENTANILLA)		
3E+14	WON	SIN FIADC	1E+14	SAN CARL		2.08E+08			00:00.0		S.	Soltero	1.	NORMA1.	CUENTA REGULAR	SIN FIADOR		
3E+14	WON	AHORRO	1E+14	NICOYA		5.04E+08			00:00.0		S.	Soltero	13.	MENO 23.	CUENTA NATURA	AHORRO MARCHAMO		
3E+14	WON	AHORRO	1E+14	NICOYA		5.04E+08			00:00.0		S.	Soltero	13.	MENO 23.	CUENTA NATURA	AHORRO ESTUDIANTIL		
3E+14	WON	AHORRO	1E+14	NICOYA		5.04E+08			00:00.0		S.	Soltero	13.	MENO 23.	CUENTA NATURA	AHORRO NAVIDENO		
3E+14	WON	AFLIACIO	1E+14	PALMARE		2.06E+08			27/05/1982		C.	Casado	1.	NORMA1.	REGULAR ACTIVO	:AFLIACION REGULAR 2%		
3E+14	WON	SIN FIADC	1E+14	PALMARE		2.06E+08			27/05/1982		C.	Casado	1.	NORMA1.	REGULAR ACTIVO	:SIN FIADOR VIP		
3E+14	WON	CS APOYO	1E+14	PALMARE		2.06E+08			27/05/1982		C.	Casado	1.	NORMA1.	REGULAR ACTIVO	:CS APOYO FUNERARIO		
3E+14	WON	CS BIENES	1E+14	Masculinc	SANTA CR	5.01E+08			00:00.0	4.	R.	SIN F	C.	Casado	1.	NORMA4.	CUENTA ANTES	CS CS BIENESTAR

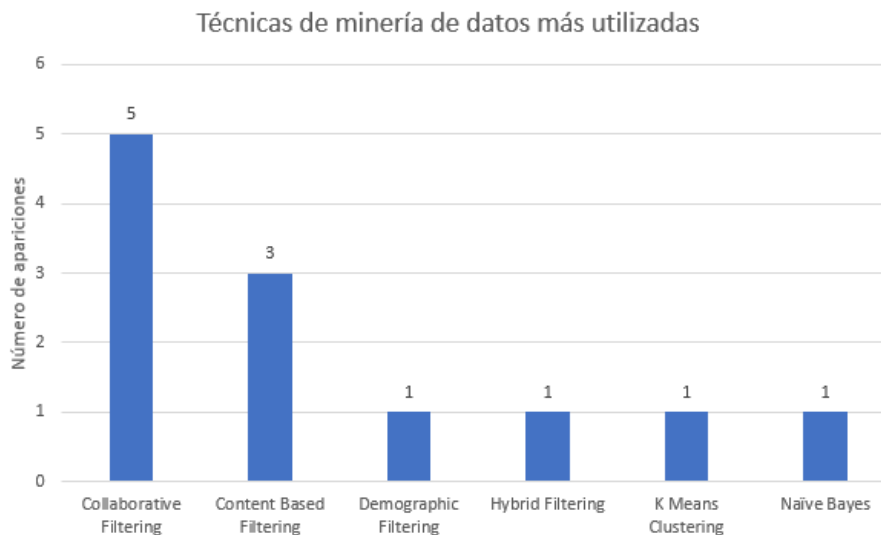
Nota. Elaboración propia.

4.2.2. Investigación de modelos predictivos para la recomendación personalizada de productos

En el desarrollo de esta actividad, se realizó una revisión documental de proyectos de minería de datos, detallada en el “Apéndice X. Resultados de la revisión documental de proyectos de minería de datos para la recomendación de productos”. Aquí se describen los proyectos encontrados en fuentes de información como sitios web de enseñanza en temas de ciencia de datos, artículos académicos y repositorios de proyectos.

Luego de realizar la revisión documental, se determinó que la técnica de minería más empleada en los proyectos consultados de recomendación personalizada de productos fue el *Collaborative Filtering*, seguido del *Content Based Filtering*, como se muestra en la Figura 16. *Técnicas de minería de datos más utilizados.*

Figura 16. *Técnicas de minería de datos más utilizadas*



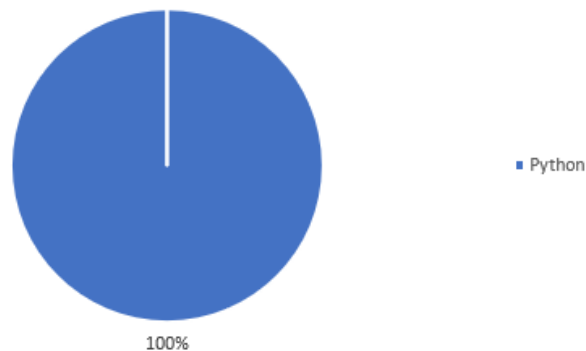
Nota. Elaboración propia.

Asimismo, como se muestra en la Figura 17. *Lenguajes de programación más utilizados*, se determinó que la herramienta más utilizada para el desarrollo de estos modelos de minería de datos

corresponde al lenguaje de programación Python, junto con librerías para el procesamiento de datos como Pandas, SciPy, NumPy, Matplotlib, GraphLab y Sklearn.

Figura 17. *Lenguajes de programación más utilizados*

Lenguajes de programación más utilizados



Nota. Elaboración propia.

Por otro lado, la información necesaria para generar el modelo de recomendación personalizada de productos de acuerdo con la información detallada en el “Apéndice X. Resultados de la revisión documental de proyectos de minería de datos para la recomendación de productos” incluye datos como:

- Información personal del cliente o usuario: detalle de las características que identifican al individuo, como el nombre, género, edad, ubicación, entre otros.
- Información del producto: información relacionada con las características propias del producto, como nombres, categorías, clasificaciones, entre otros.
- Calificaciones brindadas por el usuario: se almacena un registro con el identificador del usuario y del producto, una etiqueta de tiempo, y el puntaje brindado; en este último punto el usuario indica si le gustó o no un producto determinado, se utiliza una escala de Linkert con posibles valores entre el uno y cinco.

4.2.3. Evaluación de los modelos de minería

Tomando como referencia los resultados obtenidos durante la actividad 4.2.2, se realizaron pruebas de algunos de los modelos de minería de datos más utilizados para la recomendación personalizada de productos, dentro de los cuales se encuentran *Demographic Filtering*, *KMeans*, *Collaborative Filtering*. La creación de las pruebas fue con la herramienta Anaconda, junto con el lenguaje de programación Python; cada uno de los modelos fue evaluado de manera independiente haciendo uso de los *Notebooks* brindados por Anaconda, donde el desarrollador puede crear *scripts*, procesar datos y hacer uso de diferentes librerías. Los resultados obtenidos a partir de las pruebas se encuentran detallados en la Tabla 21.

Tabla 21. Resultados de la evaluación de modelos de minería de datos

Modelo	Resultado obtenido	Hallazgos positivos	Hallazgos negativos	Apéndice de referencia
Demographic Filtering (Popularidad)	Recomendación de productos se basa en la popularidad, es decir, los más vendidos. No genera ninguna métrica de evaluación.	<ul style="list-style-type: none"> Tiempo de implementación es corto. No requiere gran cantidad de atributos del cliente o el producto. Poco esfuerzo en tareas de limpieza y transformación de datos. Permite obtener una lista ordenada de los productos más vendidos. 	<ul style="list-style-type: none"> Da una recomendación genérica para todos los tipos de usuarios. No existen métricas de evaluación para este modelo. No toma en cuenta las características del cliente ni del producto para hacer la recomendación, únicamente, la cantidad de veces que se ha vendido un producto. 	Apéndice AE. Evaluación del modelo Demographic Filtering.

Modelo	Resultado obtenido	Hallazgos positivos	Hallazgos negativos	Apéndice de referencia
KMeans (segmentación)	<p>Según la clase determinada de cliente, recomienda los productos más vendidos.</p> <p>Según el método del codo, el número óptimo de clases identificadas dentro del conjunto de datos es de tres.</p>	<ul style="list-style-type: none"> • Toma en cuenta la información del cliente a la hora de realizar la segmentación. • Permite descubrir nuevos grupos de clientes, previamente desconocidos. • Es posible obtener una lista con diferentes productos para recomendar. • Se puede utilizar con nuevos clientes que aún no han realizado compras dentro de la empresa. 	<ul style="list-style-type: none"> • No toma en cuenta las características de los productos. • No existen métricas de evaluación para este modelo. • Requiere gran cantidad de transformaciones principalmente para eliminar valores sucios y transformar datos categóricos a numéricos. 	Apéndice AF. Evaluación del modelo KMeans.
Collaborative Filtering (Implícito)	Permite la recomendación de productos mediante productos similares, basado en las compras realizadas por los clientes de Oracle Engagement Cloud.	<ul style="list-style-type: none"> • Es un algoritmo enfocado en la recomendación de productos. • Busca patrones entre los productos comprados por clientes, de esta manera, genera recomendaciones basadas en los productos en común entre los clientes. 	<ul style="list-style-type: none"> • En casos explícitos, requiere un registro de las calificaciones que el cliente ha brindado a cada producto. 	Apéndice AF. Evaluación del modelo Collaborative Filtering.

Nota. Elaboración propia.

Dentro de la investigación de modelos de recomendación, también se identificó el de clasificación Naive Bayes. Por este motivo, se realizaron pruebas para simular el modelo y obtener métricas de evaluación como *Accuracy*, *F1 Score*, *Recall* y *Precision*; en la Tabla 22, se muestra el resumen de los datos obtenidos para cada evaluación.

Tabla 22. *Resultados de la evaluación de modelos de clasificación*

Enfoques basados en modelos						
Modelo	Resultado obtenido	Métricas de evaluación				Apéndice de referencia
		<i>Accuracy</i>	<i>F1 Score</i>	<i>Recall</i>	<i>Precision</i>	
Naive Bayes (Gaussian)	Se recomienda un solo producto al cliente, basado en sus características.	0.1548	0.07633	0.15118	0.0535	Apéndice AG. Evaluación del modelo Naive Bayes (Gaussian)
Naive Bayes (Bernoulli)		0.1511	0.07633	0.151183	0.053511	Apéndice AH. Evaluación del modelo Naive Bayes (Bernoulli)

Nota. Elaboración propia.

Como se pudo observar en la Tabla 22 ambos modelos incumplen con los criterios de evaluación definidos dentro de la sección 4.1.3. Luego de realizar las pruebas correspondientes con cada uno de los modelos descritos en la Tabla 21. *Resultados de la evaluación de modelos de minería de datos* y en la Tabla 22. *Resultados de la evaluación de modelos de modelos de clasificación*, los resultados obtenidos fueron analizados dentro de la Tabla 23. El objetivo fue determinar si el modelo cumplía o no con los requerimientos, definidos dentro de la sección “Identificación de requerimientos de la propuesta”.

Tabla 23. *Cumplimiento de requerimientos por modelo*

ID del requerimiento	Cumplimiento del requerimiento				
	Demographic Filtering	KMeans	Collaborative Filtering	Naive Bayes (Gaussian)	Naive Bayes (Bernoulli)
R001			✓	✓	✓
R002			✓		
R003	✓	✓	✓	✓	✓
R004			✓		
R005	✓	✓	✓	✓	✓

Nota. Elaboración propia.

Como se observa en la Tabla 23, el modelo que mejor se ajusta a las necesidades actuales es el *Collaborative Filtering*, ya que cumple con todos los requerimientos establecidos para el modelo de minería de datos. Los resultados obtenidos durante esta actividad serán utilizados como insumo dentro de la sección 5.2.2 , donde se describe de manera detallada el modelo de minería de datos propuesto, junto con las tareas de limpieza y transformación necesarias para su desarrollo.

4.3. Fase III. Sugerencia de herramientas para la propuesta

En esta sección, se detallan los principales hallazgos de la revisión de herramientas para la extracción, almacenamiento y procesamiento de datos. Durante esta fase del proyecto de graduación, se presentaron dos actividades principales: la revisión de las herramientas utilizadas previamente y la investigación de nuevas herramientas para la propuesta. Para realizar la selección de las herramientas, estos hallazgos se emplearon como insumos y se establecieron los criterios de evaluación.

4.3.1. Revisión de las herramientas utilizadas anteriormente

Durante esta actividad, se realizó una entrevista al consultor experto del área de infraestructura de TI de XUM TECH, detallada en el “Apéndice U. Resultados de la Entrevista #5”, con el fin de conocer las herramientas utilizadas previamente para almacenar información, procesarla, extraerla y enviarla a los sistemas de Oracle Engagement Cloud.

Como resultado, se identificaron las herramientas que se especifican en la Tabla 24. *Herramientas utilizadas previamente*, para cada una de estas se describen los principales casos de uso o propósitos para los cuales se han empleado.

Tabla 24. *Herramientas utilizadas previamente*

Herramienta	Casos de uso
Oracle Reports and Analytics	<ul style="list-style-type: none">• Reportería específica (BI).• Extracción periódica de datos del ambiente Oracle Engagement Cloud.
Cloud Flare	<ul style="list-style-type: none">• Creación de <i>dashboards</i> y reportería personalizada para monitorear los servicios web ofrecidos por XUM TECH a sus clientes.

Propuesta de solución de minería de datos basada en la metodología CRISP-DM utilizando los datos almacenados en Oracle Engagement Cloud, para la recomendación personalizada de productos a clientes, durante el primer semestre de 2022

Herramienta	Casos de uso
AWS Dynamo DB	<ul style="list-style-type: none"> Almacenamiento de información relacionada con los servicios brindados por XUM TECH, en bases de datos en la nube.
Oracle Integration Cloud (OIC)	<ul style="list-style-type: none"> Integración entre sistemas empresariales, bases de datos, servicios web y la plataforma de Oracle Engagement Cloud.
Oracle Data Integrator (ODI)	<ul style="list-style-type: none"> Creación de procesos de Extracción Transformación y Carga (ETL) con datos almacenados en soluciones de Oracle.

Nota. Elaboración propia.

4.3.2. Investigación de nuevas herramientas para la propuesta

En esta actividad, se utilizaron instrumentos de recolección de datos como entrevistas y revisiones documentales, con el propósito de conocer acerca de las funcionalidades ofrecidas por las herramientas disponibles en el mercado.

4.3.2.1. Herramientas de extracción de datos

El proceso para identificar las herramientas de extracción de datos consistió en entrevistar a consultores expertos en analítica y el sistema Oracle Engagement Cloud. Dentro de las herramientas identificadas, se encontraron *Oracle Reports and Analytics* y *Oracle Export Management*. En la Tabla 25. *Análisis Oracle Report and Analytics* y en la Tabla 26. *Análisis Oracle Export Management*, se detallan las características principales de cada una de estas herramientas, se empleó la guía definida en el “Apéndice N. Revisión documental herramientas de minería de datos”.

Tabla 25. *Análisis Oracle Report and Analytics*

Nombre	Oracle Reports and Analytics
Función principal	Extracción de datos
Fuente de información	<ul style="list-style-type: none"> • Apéndice R. Resultados de la Entrevista #1. • Apéndice T. Resultados de la Entrevista #4. Apéndice R. Resultados de la Entrevista #1
Posición en Cuadrante de Gartner	No aplica
Principales características	<ul style="list-style-type: none"> • Permite la creación de reportes y analíticos personalizados dentro de la herramienta Oracle Engagement Cloud. • Ofrece la posibilidad de almacenar y reutilizar los reportes con la información actualizada en tiempo real. • Brinda la capacidad de exportar reportes en distintos formatos, como CSV, Excel, JSON, entre otros.
Costo aproximado	Sin costo adicional, es una funcionalidad estándar de Oracle Engagement Cloud.

Ventajas	<ul style="list-style-type: none"> • Integración directa con Oracle Engagement Cloud. • Acceso a los datos en tiempo real. • No tiene costo adicional, al estar incluido dentro de la herramienta. • Los reportes son flexibles, es posible agregar campos estándares y personalizados según las necesidades. También, se pueden aplicar filtros sobre los datos.
Desventajas	<ul style="list-style-type: none"> • No permite la transformación y limpieza de los datos dentro del módulo. • Requiere conocimiento técnico en desarrollo de analíticos de Inteligencia Empresarial.

Nota. Elaboración propia.

Tabla 26. *Análisis Oracle Export Management*

Nombre	Oracle Export Management
Función principal	Extracción de datos
Fuente de información	Apéndice T. Resultados de la Entrevista #4
Posición en Cuadrante de Gartner	No aplica
Principales características	<ul style="list-style-type: none"> • Permite extraer los datos almacenados en Oracle Engagement Cloud. • Es posible seleccionar un objeto existente, así como los campos personalizados y estándares de cada uno. • El archivo generado por el proceso de exportación únicamente se encuentra disponible en formato CSV.
Costo aproximado	Sin costo adicional, es una funcionalidad estándar de Oracle Engagement Cloud.
Ventajas	<ul style="list-style-type: none"> • Puede ser utilizado por cualquier usuario de negocio con conocimiento básico en Inteligencia Empresarial. • Es programable, es decir, se puede seleccionar una hora y fecha específica de la exportación.
Desventajas	<ul style="list-style-type: none"> • Se debe configurar cada vez que desean hacer el reporte.

	<ul style="list-style-type: none"> Involucra trabajo manual al momento de exportar los datos, esto podría incluir errores y retrabajos.
--	--

Nota. Elaboración propia.

Para seleccionar el mecanismo de extracción de datos, se tomaron en cuenta los aspectos definidos en la Tabla 27. *Criterios de evaluación para herramientas de extracción*, donde se indica para cada rubro la descripción y su peso dentro de la evaluación de herramientas.

Tabla 27. *Criterios de evaluación para herramientas de extracción de datos*

Rubro	Descripción	Peso
Reutilización	Capacidad de reutilización de los reportes.	20%
Experiencia de usuario	Facilidad para la creación de reportes.	20%
Integración	Capacidad de integrar distintos objetos en un mismo reporte.	20%
Flexibilidad	Capacidad de modificar reportes según las necesidades del usuario.	20%
Costo	Costo aproximado de la implementación.	20%

Nota. Elaboración propia.

Los resultados de la evaluación de herramientas se muestran en la Tabla 28. *Comparativa de herramientas de extracción de datos*.

Tabla 28. *Comparativa de herramientas de extracción de datos*

Rubro	Oracle Reports and Analytics	Oracle Export Management
Reutilización (20 %)	20	0
Experiencia de usuario (20%)	0	20
Integración (20%)	20	0
Flexibilidad (20%)	20	0
Costo (20%)	20	20
Total	80	40

Nota. Elaboración propia.

Como se puede ver en la Tabla 28. *Comparativa de herramientas de extracción de datos*, el mayor puntaje lo obtuvieron *Oracle Reports & Analytics*, debido a su ventaja en los aspectos de reutilización, integración y flexibilidad.

4.3.2.2. Herramientas de almacenamiento de datos

En el caso de las herramientas de almacenamiento de datos, se llevó a cabo una revisión documental en distintas fuentes de información en la web. Asimismo, se entrevistó a una consultora experta en Oracle Engagement, para analizar alternativas de almacenamiento dentro de la herramienta; esta entrevista se encuentra documentada en el “Apéndice BH. Resultados de la Entrevista #7”. La herramienta sugerida fueron los *Custom Object* de Oracle.

La fuente de información utilizada para obtener la lista de posibles herramientas de almacenamiento de datos mediante la revisión documental corresponde a *Gartner Magic Quadrant for Cloud Database Management Systems 2020*, donde se presentan las principales herramientas de almacenamiento de información en el mercado. De acuerdo con la Figura 18. *Gartner Magic Quadrant for Cloud Database Management Systems 2020*, en el cuadrante de líderes se encuentran *Amazon Web Services*, *Oracle Cloud Infrastructure* y *Microsoft*.

Figura 18. *Gartner Magic Quadrant for Cloud Database Management Systems 2020*



Nota. Tomado de Gartner, 2021.

Para evaluar las herramientas identificadas, se emplearon cuatro criterios, detallados en la Tabla 29. Criterios de evaluación de herramientas de almacenamiento junto con el peso correspondiente. El puntaje de cada criterio será asignado solamente a una de las herramientas, la cual será aquella que se ajuste de mejor manera el criterio definido.

Tabla 29. *Criterios de evaluación de herramientas de almacenamiento*

Criterio	Descripción	Peso
Costo	El costo es un factor importante para el negocio, por este motivo, se brindará el mayor puntaje a la herramienta de menor costo.	40%
Experiencia previa	La herramienta evaluada ha sido utilizada por los colaboradores de XUM TECH, de esta manera, se garantiza su correcto funcionamiento.	20%
Capacidad de integrarse con otras soluciones	Se evalúa la capacidad de la herramienta para integrarse con otras soluciones y servicios que faciliten la gestión de campañas de mercadeo.	20%
Calificación brindada en <i>Peer Insights</i>	Considera la calificación general brindada a la herramienta por usuarios de la industria, dentro del portal de Gartner <i>Peer Insight</i> .	10%

Nota. Elaboración propia.

La información considerada para realizar la evaluación se encuentra detallada en el “Apéndice Y. Revisión documental de herramientas de almacenamiento AWS”, “Apéndice Z. Revisión documental de herramientas de almacenamiento Oracle”, “Apéndice AA. Revisión documental de herramientas de almacenamiento Microsoft” y “Apéndice BI. Revisión documental de herramientas de almacenamiento Oracle CX *Custom Objects*”.

Los resultados de la evaluación de las herramientas se muestran en la Tabla 30. *Comparativa de herramientas de almacenamiento*, donde se indica el ganador de cada rubro junto con el puntaje obtenido.

Tabla 30. *Comparativa de herramientas de almacenamiento*

Rubro	AWS Redshift	Oracle <i>Autonomous database</i>	Microsoft <i>Synapse Analytics</i>	Oracle CX – <i>Custom Objects</i>
Costo				40
Experiencia previa				20
Capacidad de integrarse con otras soluciones				20
Calificación brindada en <i>Peer Insights</i>		20		
Total		20		80

Nota. Elaboración propia.

Como se puede ver en la **Tabla 30**, la herramienta con mayor puntaje corresponde a *Oracle CX-Custom Objects*, debido a su ventaja en los aspectos de costo, experiencia previa y capacidad de integración con otras soluciones.

4.3.2.3. Herramientas de minería de datos

En el caso de las herramientas de minería de datos, se hizo una revisión documental en distintas fuentes de información, principalmente en la web. Al igual que las herramientas de almacenamiento de datos, la principal fuente de información fue *Gartner Magic Quadrant for Data Science and Machine Learning Platforms 2021*. En la Figura 19. Magic Quadrant for Data Science and Machine Learning Platforms 2021 se muestran las principales herramientas de la industria.

Figura 19. *Magic Quadrant for Data Science and Machine Learning Platforms 2021*



Nota. Tomado de Gartner (2021).

Para evaluar las herramientas investigadas, se definieron cuatro criterios, detallados en la Tabla 31. El puntaje únicamente será asignado a la herramienta que cumpla de mejor manera el criterio definido.

Tabla 31. *Criterios de evaluación de herramientas de minería de datos*

Criterio	Descripción	Peso
Desarrollos personalizados	Integrar código mediante lenguajes como Python o R, para transformar los datos de acuerdo con el modelo de <i>Collaborative Filtering</i> .	50%
Costo	El factor del costo es un elemento importante para XUM TECH, se le brindará el puntaje a la herramienta más barata.	50%

Nota. Elaboración propia.

La información considerada para realizar la evaluación se encuentra detallada en el “Apéndice AB. Revisión documental de herramientas de minería de datos IBM”, “Apéndice AC. Revisión documental de herramientas de minería de datos Amazon Sage Maker” y “Apéndice AD. Revisión documental de herramientas de minería de datos Anaconda”,

Los resultados de la evaluación de las herramientas se muestran en la Tabla 32, donde se indica el ganador de cada rubro junto con el puntaje obtenido.

Tabla 32. *Resultado de evaluación de herramientas de minería de datos*

Rubro	Amazon Sage Maker	IBM Watson Studio	Anaconda
Desarrollo personalizado	50		50
Costo			50
Total	50		100

Nota. Elaboración propia.

Como se puede ver en la **Tabla 32**, la herramienta con mayor puntaje corresponde a Anaconda, debido a su ventaja en los aspectos de costo y personalización.

Propuesta de solución de minería de datos basada en la metodología CRISP-DM utilizando los datos almacenados en Oracle Engagement Cloud, para la recomendación personalizada de productos a clientes, durante el primer semestre de 2022

4.3.2.4. Herramientas de mercadeo

En el caso de la herramienta de mercadeo, dentro de la sesión de validación de requerimientos detallada en el “Apéndice S. Resultados de la Entrevista #3”, se estableció que se debe emplear la herramienta llamada Oracle Eloqua. Esta es altamente utilizada por los clientes de XUM TECH y permite la creación de campañas de mercadeo, envío de correos electrónicos de manera masiva y la sincronización con la información almacenada dentro de Oracle Engagement Cloud. Adicionalmente, se encuentra listada en la categoría de líder según el *Magic Quadrant for CRM Lead Management 2020*, como se muestra en la Figura 20. Magic Quadrant for CRM Lead Management 2020

Figura 20. Magic Quadrant for CRM Lead Management 2020



Nota. Tomado de Gartner, 2020.

4.4. Fase IV. Recolección de insumos para el análisis financiero de la propuesta

En esta sección, se detallan los principales hallazgos de las actividades relacionadas con el análisis financiero de la propuesta, donde se llevaron a cabo entrevistas para recolectar información importante a la hora de cotizar el proyecto, como los costos de salarios, impuestos y márgenes de ganancia. Posteriormente, en el capítulo 5.4, estos datos son utilizados para la cotización de la implementación.

4.4.1. Identificación de costos asociados

Como parte de esta actividad, se entrevistó a una consultora experta del área comercial de XUM TECH. El propósito fue determinar los aspectos que usualmente se deben tomar en cuenta al cotizar un proyecto de consultoría e implementación de soluciones Oracle. La entrevista completa se encuentra disponible en el “Apéndice V. Resultados de la Entrevista #6”.

La experta del área comercial indicó que en XUM TECH se utilizan plantillas de cotización estandarizadas, con las cuales es posible obtener información acerca de los aspectos financieros tomados en cuenta en una cotización y los costos asociados. A continuación, en la Tabla 33. *Aspectos importantes en la cotización*, se detallan los principales hallazgos de la entrevista.

Tabla 33. *Aspectos importantes en la cotización*

Aspectos de la cotización	Detalle
Costo aproximado del talento humano	Los costos del talento humano de XUM TECH, por motivos de confidencialidad, no serán detallados en este proyecto, sin embargo, se tomará como referencia un costo de \$60 por hora.
Aspectos para tomar en cuenta en la cotización	En el caso de XUM TECH, las cotizaciones toman en cuenta aspectos como: <ul style="list-style-type: none"> • Duración del proyecto.

Aspectos de la cotización	Detalle
	<ul style="list-style-type: none"> • Complejidad del proyecto. • Cantidad de ítems configurados.
Margen de ganancia aproximado	En los proyectos, XUM TECH define un margen de ganancia para asegurar que estos sean rentables; sin embargo, por motivos de confidencialidad el dato no se especifica dentro del proyecto. Para el proyecto, se tomará como referencia un margen del 30%.
Impuestos en la cotización	Todos los proyectos de XUM TECH deben incluir el cálculo de impuestos correspondiente. Debido a que la empresa opera en distintos países, el impuesto se encuentra en un rango de entre el 13% y 29%.

Nota. Elaboración propia.

Durante esta fase, se identificaron los principales aspectos a tomar en cuenta en la cotización e implementación de proyectos dentro de XUM TECH. Posteriormente en la sección 5.4, se utilizarán los datos obtenidos durante esta fase como insumo para realizar el análisis financiero de la propuesta y así determinar los beneficios que podría generar a la organización.

5. Propuesta de solución

En este apartado, se detalla la propuesta de solución planteada por el investigador, basada en los hallazgos del capítulo 4. A lo largo del capítulo, se describe la situación deseada por parte de XUM TECH y la propuesta de solución planteada para cumplir con esta. Adicionalmente, se detallan las decisiones de diseño tomadas junto con sus evidencias, por último, se realiza el análisis financiero de la propuesta y su respectivo plan de implementación para validar la viabilidad del proyecto.

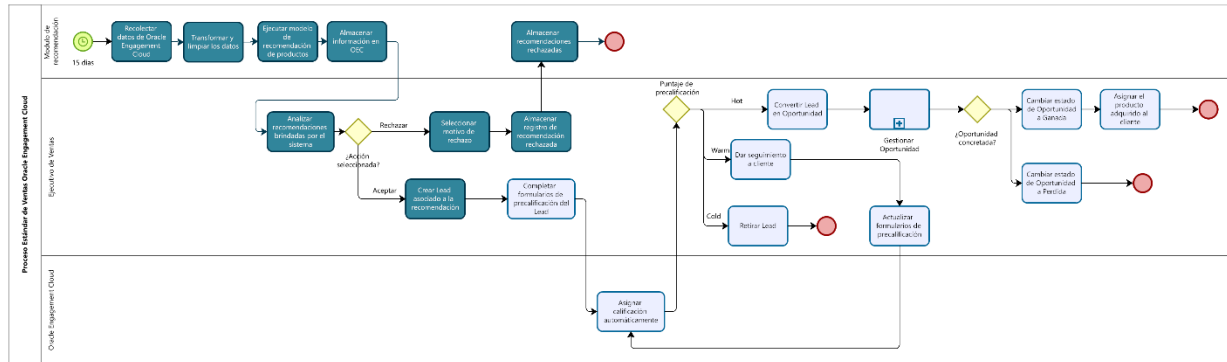
5.1. Procesos TO BE

5.1.1. Creación de recomendaciones

En la **Figura 21**, se muestra el proceso TO BE propuesto, el cual fue diseñado para atender las necesidades planteadas por el departamento comercial de XUM TECH. Respecto al proceso AS IS, se identifican las siguientes mejoras:

- **Módulo de recomendación:** se plantea extraer periódicamente la información almacenada en Oracle Engagement Cloud de manera que el modelo predictivo pueda aprender continuamente de los datos transaccionales generados por la operación de la empresa. Esta información extraída será sometida a tareas de limpieza y transformación para posteriormente utilizarla en la creación de recomendaciones de productos.
- **Análisis de recomendaciones:** las recomendaciones de productos generadas por el nuevo módulo, serán almacenadas dentro de Oracle Engagement Cloud para que puedan ser consultadas por los ejecutivos de ventas, así tendrán la posibilidad de analizar la recomendación generada por el sistema y decidir si les es útil o no. En caso de serlo, servirá como insumo para la creación de una nueva prospectación, donde el ejecutivo contactará al cliente sin necesidad de que este consulte por un determinado producto, como ocurre en el proceso actual.

Figura 21. Proceso TO BE de creación de recomendaciones



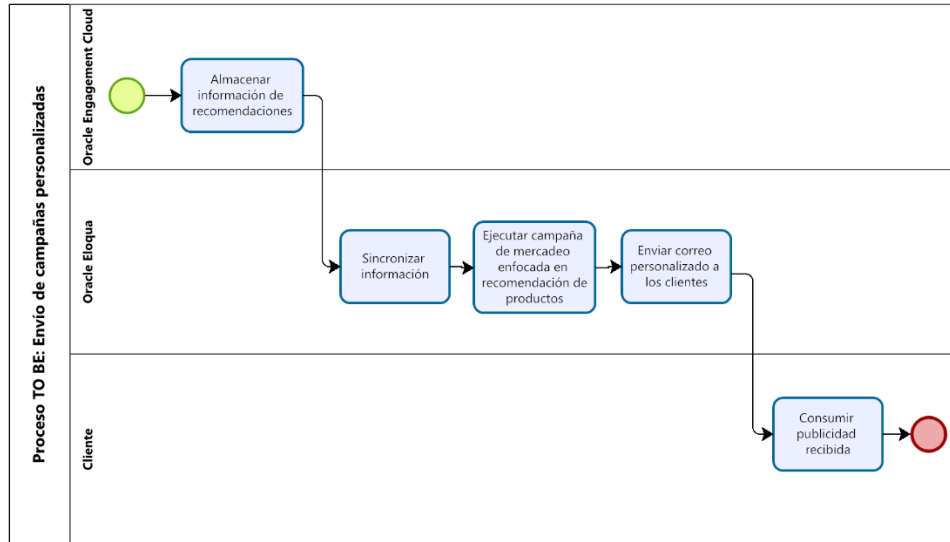
Nota. Elaboración propia.

5.1.2. Envío de campañas personalizadas

Como parte de la propuesta planteada y con base en los requerimientos indicados por el Departamento Comercial de XUM TECH, se planteó el proceso TO BE para el envío de campañas representado en la **Figura 22**. Las sugerencias dadas por el módulo de recomendación podrán emplearse para la creación de campañas de mercadeo con la herramienta de Oracle Eloqua. A continuación, se describen las principales actividades para realizar este proceso:

- La información de las recomendaciones debe almacenarse dentro de Oracle Engagement Cloud, de acuerdo con los requerimientos identificados para la propuesta en la sección 4.1.2 ; la carga de datos debe ser realizada mediante archivos CSV.
- Posteriormente, por medio de un proceso de sincronización, configurado dentro de Eloqua, se extraen los datos de los clientes y las recomendaciones para crear objetos dentro de la herramienta; estos pueden utilizarse en la ejecución de campañas de mercadeo.
- Una vez que los datos de los clientes han sido cargados en Eloqua, se puede enviar campañas de mercadeo automáticas a cada cliente, a través de correo electrónico. En estas, se brindaría información específica acerca de los productos recomendados y se invitaría al cliente a adquirirlos.

Figura 22. Proceso TO BE Campañas de mercadeo



Nota. Elaboración propia.

5.2. Módulo de recomendación de productos

5.2.1. Extracción de datos de Oracle Engagement Cloud

Basado en la comparativa realizada dentro de la sección 4.3.2, la herramienta con mejor calificación, por ende, seleccionada para realizar las tareas de extracción de datos fue *Oracle Reports and Analytics*. Esta herramienta ofrece las siguientes ventajas:

- Permite acceder a los datos en tiempo real, esto quiere decir que los datos mostrados en el reporte representan la actualidad de la empresa.
- Brinda flexibilidad al crear reportes y analíticos, esto permite a los usuarios encargados de la extracción de datos seleccionar nuevos campos o filtrar los datos de una manera ágil, sin generar reportes desde cero, ya que son totalmente reutilizables.

- Ofrece la posibilidad de integrarse directamente con Oracle Engagement Cloud, solo se debe contar con un usuario que tenga asignados los roles para acceder al módulo de *Reports and Analytics*.

Como parte de la propuesta, se creó el analítico “Reporte de ventas-Oportunidades” dentro del ambiente de Oracle Engagement Cloud brindado por XUM TECH, donde se utilizó el *Subject Area Sales-CRM Opportunities and Products* para extraer los atributos descritos en la **Tabla 34** y así crear el modelo de minería de datos.

Tabla 34. *Lista de atributos extraídos*

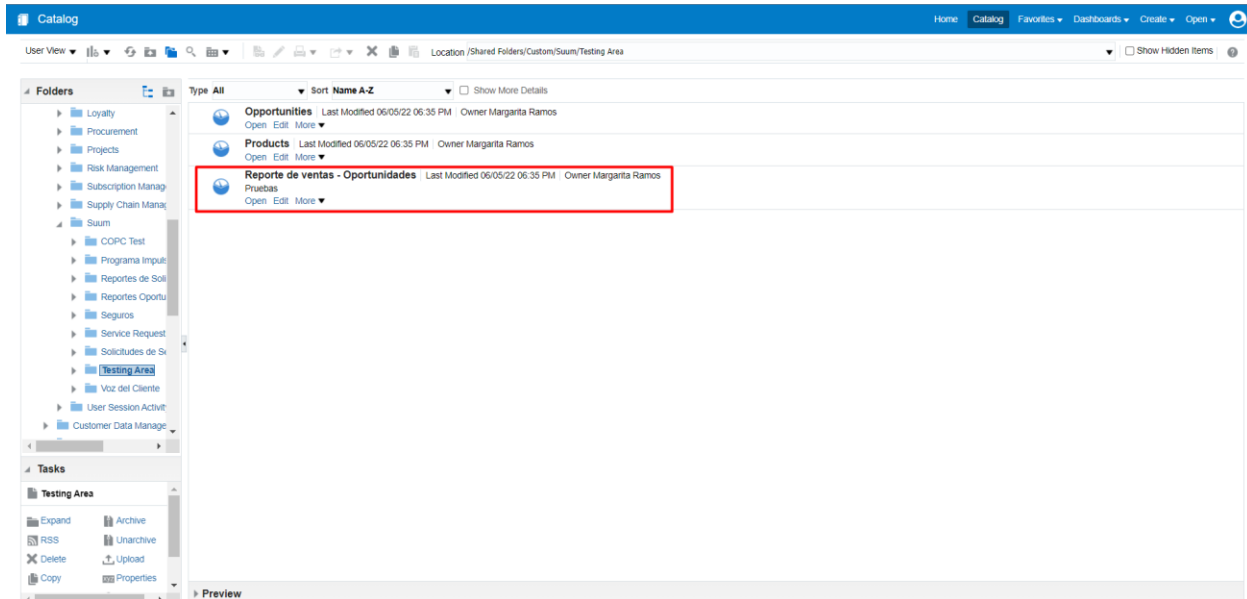
Objeto	Atributo
Contact	<ul style="list-style-type: none"> • First Name • Last Name • Gender • Cantón • Tipo de identificación • Número de identificación • Distrito • Fecha de nacimiento • Tipo cliente • Estado civil • Categoría de asociado • Tipo de asociado • Perfil social • Perfil ambiental • Country • Postal Code
Opportunity	<ul style="list-style-type: none"> • Opportunity ID • Status • Code • Nombre de la oportunidad • Reason Won or Lost • Producto • Contact Sales Account Id, • Primary Contact Party ID • Sales Account ID

Nota. Elaboración propia.

Propuesta de solución de minería de datos basada en la metodología CRISP-DM utilizando los datos almacenados en Oracle Engagement Cloud, para la recomendación personalizada de productos a clientes, durante el primer semestre de 2022

El reporte generado en *Oracle Reports and Analytics* se muestra en la **Figura 23**.

Figura 23. Reporte de extracción creado



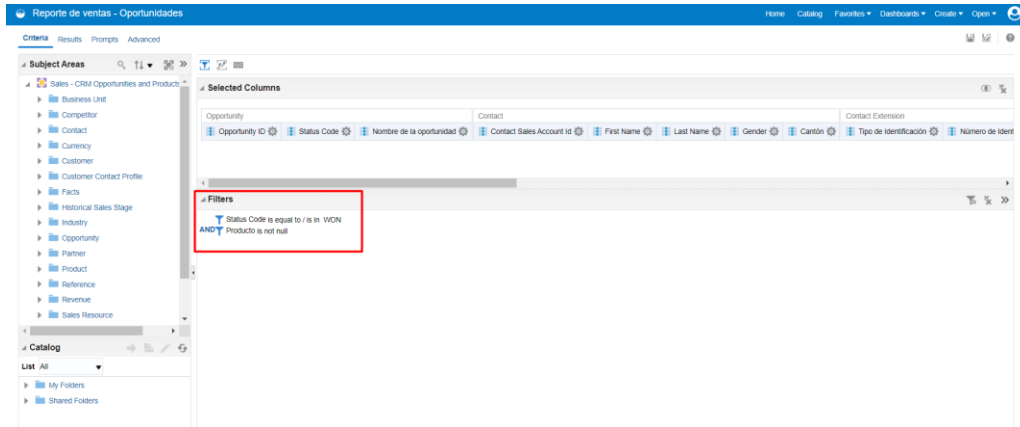
Nota. Elaboración propia.

En la **Figura 24**, se observan las columnas seleccionadas para la creación del analítico, así como los filtros configurados dentro del reporte, los cuales son:

- Filtro #1: filtrar las oportunidades de acuerdo con el *Status Code*; de esta manera, dentro del analítico solo se incluyen oportunidades ganadas (estado *Won*), es decir, las ventas que realmente se concretaron.
- Filtro #2: filtrar los valores nulos dentro de la columna Producto, para evitar aquellos donde no se registró ningún producto como parte de la oportunidad.

Propuesta de solución de minería de datos basada en la metodología CRISP-DM utilizando los datos almacenados en Oracle Engagement Cloud, para la recomendación personalizada de productos a clientes, durante el primer semestre de 2022

Figura 24. *Filtros aplicados al reporte de extracción*



Nota. Elaboración propia.

5.2.2. Creación del modelo de recomendación de productos

Bibliotecas necesarias

La creación del modelo de minería de datos requiere el uso de bibliotecas para el manejo de datos que brindan un conjunto de funciones, que le permiten al desarrollador ahorrar tiempo y esfuerzo. Actualmente, existe una numerosa cantidad de opciones disponibles en el mercado, sin embargo, las librerías utilizadas para la propuesta en lenguaje de programación Python se detallan en la **Tabla 35**.

Tabla 35. *Bibliotecas necesarias para el modelo*

Biblioteca	Función principal	Documentación de referencia
Pandas	Proporciona estructuras de datos conocidas como <i>dataframes</i> , que permiten manipular los datos de manera sencilla y eficiente, además, permite realizar tareas de transformación y limpieza de datos.	pandas - Python Data Analysis Library (pydata.org)
Numpy	Permite elaborar vectores y matrices multidimensionales útiles para la creación, entrenamiento y prueba de los modelos de minería de datos.	NumPy
Matplotlib	Permite crear gráficos para la exploración de la distribución de los datos almacenados en arreglos, <i>dataframes</i> y vectores.	Matplotlib — Visualization with Python
Sklearn	Una de las librerías más utilizadas de la industria para la creación de modelos de <i>Machine Learning</i> , adicionalmente, ayuda a realizar tareas de manipulación y preprocesamiento de los datos.	scikit-learn: machine learning in Python — scikit-learn 1.0.2 documentation

Nota. Elaboración propia.

Modelo seleccionado

De acuerdo con la información obtenida en el Análisis de Resultados, el modelo que mejor se ajusta a las necesidades definidas para este proyecto es el *Collaborative Filtering*, por las siguientes razones:

- Toma en cuenta el comportamiento de clientes similares, el historial de los productos adquiridos por sus clientes y las relaciones entre los diferentes productos.
- Las recomendaciones cuentan con un puntaje del 1 al 100, según la calificación obtenida, con base en la similitud del producto recomendado con el que actualmente tiene el cliente.
- Al ser un *Collaborative Filtering*, permite ser ajustado posteriormente por medio de la recolección de opiniones de los usuarios. Es decir, los clientes podrían evaluar los productos que poseen y esas calificaciones podrían ser utilizadas por el modelo para recomendar productos de manera aún más personalizada, tomando en cuenta sus gustos.
- Ofrece ventajas como facilidad de implementación, requiere menor esfuerzo en la limpieza y transformación de los datos comparado con los otros modelos evaluados en la sección 4.2.3

Tareas de transformación y limpieza

Basado en la evaluación documentada en el “Apéndice AF. **Evaluación del modelo Collaborative Filtering**”, la lista de transformaciones necesarias para la creación de recomendaciones utilizando el modelo *Collaborative Filtering* es la siguiente:

- Eliminación de valores nulos en columnas de *Primary Contact ID* y *Contact ID*.
- Creación de nuevo campo “*Rating*”.
- Filtración de columnas, para que únicamente queden *Contact ID*, Producto y *Rating*.
- Creación de tabla *pivot*, donde los productos pasan a ser las columnas de la tabla y los *Contact ID* pasan a ser las filas.
- Factorización de vectores utilizando *Singular Value Decomposition*.
- Creación de la matriz de correlación entre los distintos productos.

Generar archivo CSV de carga.

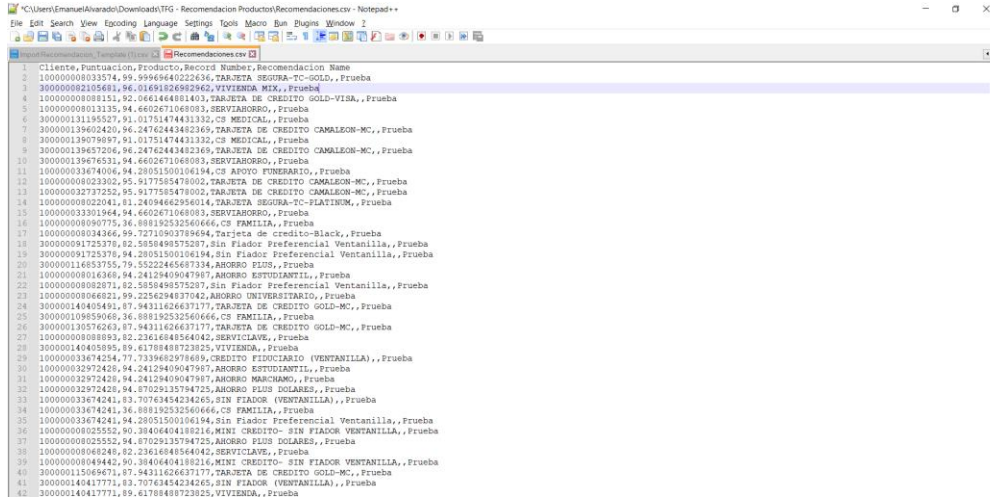
Para generar el archivo CSV, se realizó un trabajo de procesamiento de datos dentro del *notebook*, que permite transformarlos de manera que las columnas resultantes sean el ID del cliente dentro de Oracle Engagement Cloud, el producto recomendado y el puntaje obtenido. A continuación, se listan las tareas de transformación necesarias para crear el archivo CSV de carga:

- Creación de tres funciones que permitieron las tres mejores recomendaciones. Las funciones reciben como parámetro el nombre del producto y retornan el producto recomendado junto con la puntuación. El detalle de estas funciones se encuentra en el “Apéndice AX. Transformación #1 Funciones creadas para generar las recomendaciones”. Seguidamente se indica el nombre de la función y el resultado generado:
 - Obtener_recomendaciones_01: retorna una lista que contiene el nombre del producto y la puntuación calculada de la mejor recomendación.
 - Obtener_recomendaciones_02: retorna una lista que contiene el nombre del producto y la puntuación calculada de la segunda mejor recomendación.
 - Obtener_recomendaciones_03: retorna una lista que contiene el nombre del producto y la puntuación calculada de la tercera mejor recomendación.
- Creación de nuevos campos en el *dataframe*, con el propósito de asignar la lista de productos recomendados utilizando las funciones detalladas en el punto anterior. El código utilizado se encuentra en el “Apéndice AY. Transformación #2 Creación de nuevos campos para almacenar la recomendación”. A continuación, se detalla el nombre de los nuevos campos creados y la información almacenada en cada uno de ellos:
 - Recomendaciones_01: se asigna el resultado de la función Obtener_recomendaciones_01, es decir, el producto y puntuación de la mejor recomendación.

- Recomendaciones_02: se asigna el resultado de la función `Obtener_recomendaciones_02`, es decir, el producto y puntuación de la segunda mejor recomendación.
- Recomendaciones_03: se asigna el resultado de la función `Obtener_recomendaciones_03`, es decir, el producto y puntuación de la tercera mejor recomendación.
- Creación de tres nuevos *dataframes*, divididos de la siguiente manera, el código utilizado se encuentra en el “Apéndice AZ. Transformación #3 Creación de nuevos *dataframes*”:
 - Recomendaciones_01: contiene el campo *Contact ID* y *Recomendacion01*.
 - Recomendaciones_02: contiene el campo *Contact ID* y *Recomendacion02*.
 - Recomendaciones_03: contiene el campo *Contact ID* y *Recomendacion03*.
- Cambio en el nombre de las columnas en los tres *dataframes*, el código se encuentra en el “Apéndice BA. Transformación #4 Cambio de nombre de columnas”:
 - *Contact ID* mantiene el mismo nombre.
 - Los campos *Recomendación01*, *Recomendación02* y *Recomendación03* pasan a llamarse *Recomendación*, para estandarizar el nombre y posteriormente poder unificar los tres *dataframes* en uno solo.
- Unión de los tres *dataframes* *Recomendaciones_01*, *Recomendaciones_02* y *Recomendaciones_03*, lo cual da como resultado un solo dataframe con los campos *Contact ID* y *Recomendación* llamado “*recomendaciones_def*”. La evidencia de esta transformación se encuentra en el “Apéndice BB. Transformación #5 Unión de *dataframes* de recomendaciones”.
- Creación de funciones para la limpieza de los datos:
 - `Obtener_puntaje`: recibe como parámetro la lista del campo *Recomendación* y retorna únicamente la puntuación.
 - `Obtener_producto`: recibe como parámetro la lista del campo *Recomendación* y retorna únicamente el nombre del producto.

- Creación de nuevas columnas dentro del *dataframe* *recomendaciones_def*, la evidencia de esta transformación se encuentra en el “Apéndice BC. Transformación #6 Creación de funciones para la limpieza de datos”. A continuación, se listan las columnas resultantes luego de aplicar las transformaciones mencionadas:
 - Producto: contiene el nombre del producto recomendado, se asignó utilizando la función de *Obtener_producto*.
 - Puntaje: contiene la puntuación del producto recomendado, se asignó utilizando la función de *Obtener_puntaje*.
 - Recomendación Name: se completó con valores de “Prueba”, ya que será requerido en la creación en Oracle Engagement Cloud.
 - *Record Number*: se completó con valores *string* vacíos, porque no es requerido en la creación en Oracle Engagement Cloud, pero sí debe pasarse como parámetro.
- Los nombres de las columnas se actualizaron, para cumplir con los requerimientos del archivo de importación definido en la sección 5.2.4. La evidencia de esta transformación se encuentra en el “Apéndice BD. Transformación #7 Creación de nuevas columnas en *dataframe*”. Se muestra la actualización:
 - *Contact ID* pasa a llamarse “Cliente”.
 - Puntaje pasa a llamarse “Puntuación”.
- Filtrado del *dataframe*, se creó un nuevo *dataframe* llamado “*recomendaciones_export*”, que es una copia de “*recomendaciones_def*”. En este, se eliminaron las columnas innecesarias para cumplir con los requerimientos del archivo de importación definido en la sección 5.2.4. El *dataframe* final únicamente contiene los campos de Cliente, Puntuación, Producto, *Record Number* y Recomendación Name. La evidencia de esta transformación se encuentra en el “Apéndice BE. Transformación #8 Filtrado del *dataframe*”.
- Finalmente, el *dataframe* llamado *recomendaciones_export* fue exportado a formato CSV. La evidencia de esta transformación se encuentra en el “Apéndice BF. Transformación #9 Exportación del *dataframe* a formato CSV”. El resultado del archivo final se muestra en la **Figura 25**.

Figura 25. CSV de recomendaciones generado



Nota. Elaboración propia.

Posteriormente estos datos pueden ser importados dentro de Oracle Engagement Cloud con el módulo *Import Management*, donde se creó un mapeo para el objeto Recomendación con los atributos como el ID del cliente, producto recomendado, puntaje de la recomendación y nombre de la recomendación, como se muestra más adelante en la sección 5.2.4.

5.2.3. Creación de objetos y campos

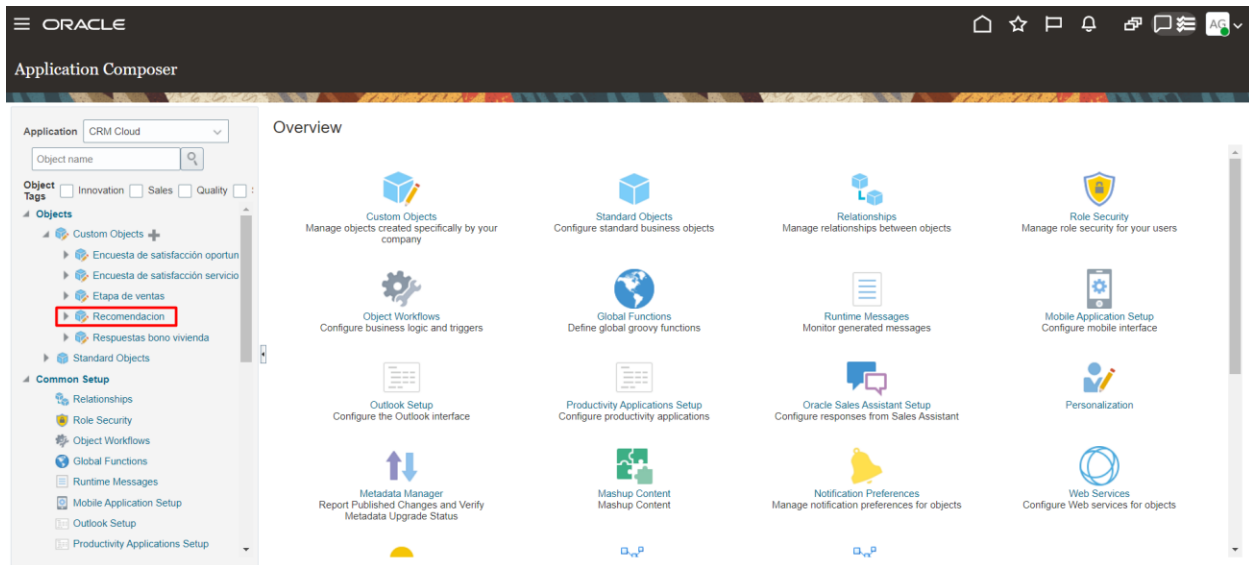
De acuerdo con los resultados obtenidos en la sección 4.3.2, la mejor opción para el almacenamiento de información es crear objetos *custom* dentro de Oracle Engagement Cloud. Las razones son las siguientes:

1. Brinda una opción de almacenamiento a costo nulo, ya que se encuentra integrado dentro de *Oracle Engagement Cloud*.
2. Es posible integrarlo a otros servicios y soluciones para diseñar el sistema de recomendación de productos utilizando sistemas como Eloqua, para crear campañas de mercadeo.

- Es capaz de almacenar la información de las recomendaciones generadas por el modelo, para utilizarla en futuros análisis. También, mediante objetos *custom* es posible realizar la carga de datos con archivos CSV directamente desde el *Import Management*.

Para almacenar y actualizar las recomendaciones de productos generadas por el modelo, se creó un objeto *custom* llamado “Recomendación”, como se puede ver en la **Figura 26**, dentro del ambiente de pruebas brindado por la organización.

Figura 26. Custom object de Recomendación



Nota. Elaboración propia.

Posteriormente, dentro del objeto *custom* “Recomendación”, se crearon los campos descritos en la **Tabla 36**, para almacenar los datos correspondientes a cada recomendación.

Tabla 36. Campos creados dentro del objeto Recomendación

Nombre del campo	Tipo de dato	Descripción	Apéndice de referencia
Fecha de creación	Date	Indica la fecha en que fue generada la recomendación.	Apéndice AI. Configuración de

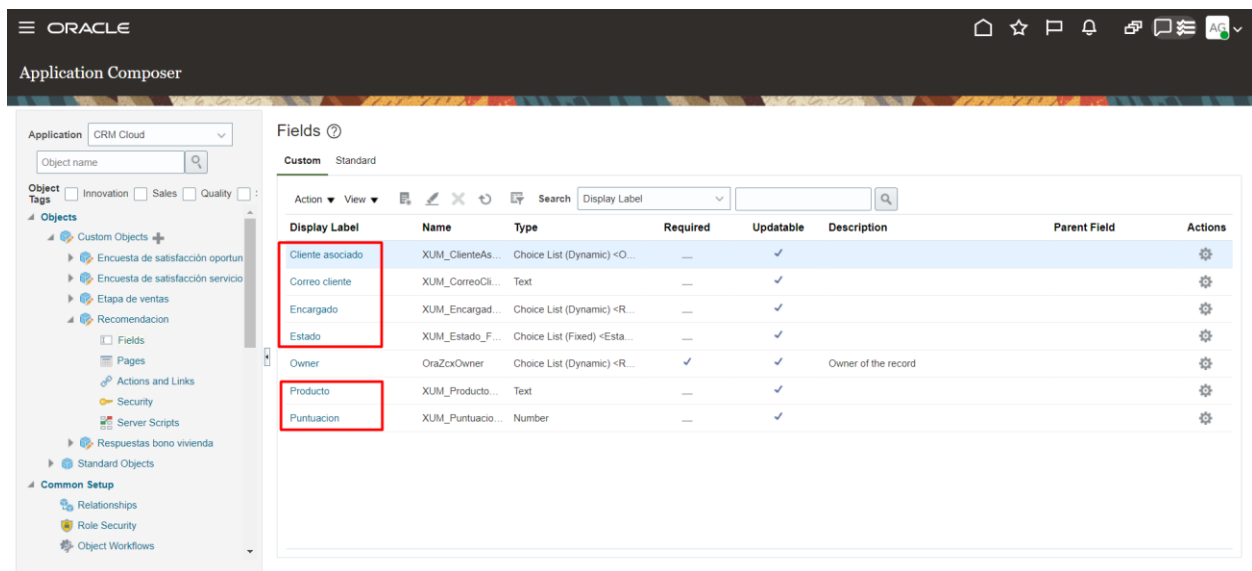
Nombre del campo	Tipo de dato	Descripción	Apéndice de referencia
			campo Fecha de creación.
Cliente asociado	Lista dinámica <Contact>	Almacena el identificador y hace referencia al cliente asociado a la recomendación.	Apéndice AJ. Configuración de campo Cliente asociado.
Correo cliente	Texto	Almacena el correo electrónico del cliente, que será utilizado para enviar información sobre los productos.	Apéndice AK. Configuración de campo Correo cliente.
Producto	Texto	Almacena el nombre del producto asociado a la recomendación.	Apéndice AL. Configuración de campo Producto.
Puntuación	Numérico	Almacena el puntaje de la recomendación, obtenido del modelo.	Apéndice AM. Configuración de campo Puntuación
Estado	Lista fija	Almacena el estado de la recomendación, los posibles valores son “Pendiente”, “Aceptada” y “Rechazada”. Por defecto, todas las recomendaciones nuevas tienen el estado “Pendiente”.	Apéndice AO. Configuración de campo Estado.
Encargado	Lista dinámica <Resource>	Almacena el identificador del ejecutivo a cargo de dar seguimiento a la recomendación.	Apéndice AP. Configuración de campo Encargado.

Nombre del campo	Tipo de dato	Descripción	Apéndice de referencia
Motivo de rechazo	Lista fija	Almacena el motivo de rechazo seleccionado por el ejecutivo de ventas.	Apéndice AQ. Configuración de campo Motivo de rechazo.
Comentarios adicionales	Texto largo	Permite al ejecutivo de ventas añadir cualquier comentario que considere importante para la recomendación.	Apéndice AR. Configuración de campo Comentarios adicionales.

Nota. Elaboración propia.

Durante la creación de los campos *custom*, se utilizó un *sandbox* no publicable dentro del ambiente de Oracle Engagement Cloud brindado por XUM TECH. Estos elementos se pueden ver configurados en la **Figura 27**.

Figura 27. Campos configurados dentro de Engagement



Nota. Elaboración propia.

Además, como parte de la solución fue necesario desarrollar dos listas de valores fijas, las cuales se detallan en la **Tabla 37**.

Tabla 37. *Detalle de listas de valores*

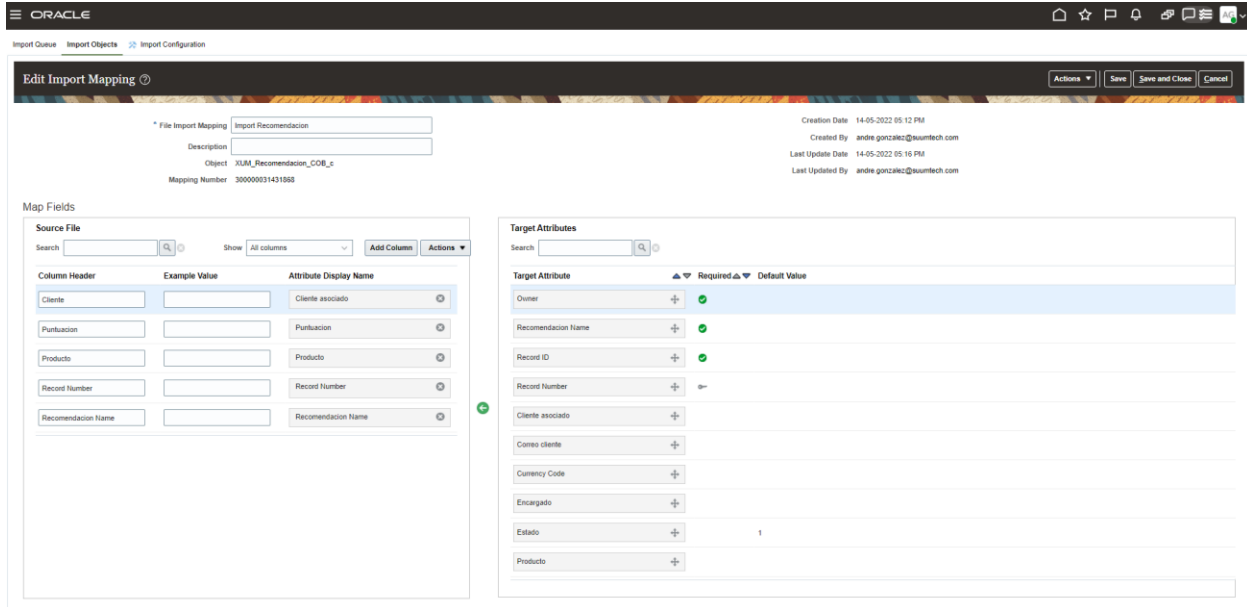
Nombre	Descripción	Valores	Apéndice de referencia
Estado Recomendación	Permite al ejecutivo de ventas cambiar el estado de la recomendación.	1- Pendiente. 2- Aceptada. 3- Rechazada.	Apéndice AS. Configuración de lista de valores Estado.
Motivo Rechazo	Permite al ejecutivo indicar el motivo de rechazo de la recomendación.	1- Cliente no cumple los requisitos mínimos. 2- Riesgo alto. 3- Otro motivo.	Apéndice AT. Configuración de lista de valores Motivo de rechazo.

Nota. Elaboración propia.

5.2.4. Configuración de archivos de importación

Para importar los registros de recomendación a Engagement Cloud, se propone el uso del *Import Management*, este módulo permite la importación de datos a objetos *custom* y estándar por medio de archivos CSV. En el caso de las recomendaciones, se creó un *Import Mapping* que contiene los campos de Cliente, Puntuación, Producto, *Record Number* y *Recomendación Name*, como se muestra en la **Figura 28**.

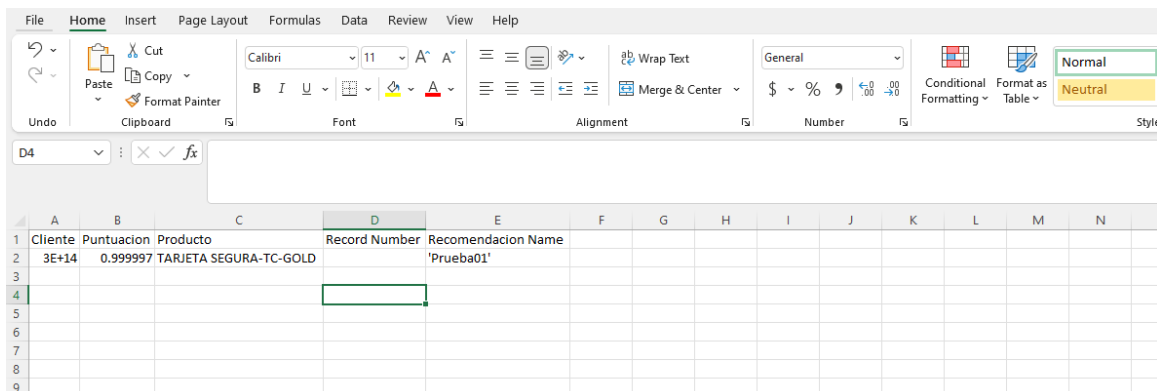
Figura 28. Mapeo de importación



Nota. Elaboración propia.

El *Import Mapping* creado funciona como una plantilla de carga de datos a Oracle Engagement Cloud, la cual puede ser descargada posteriormente para completarla con los datos correspondientes a las recomendaciones generadas, como se muestra en la **Figura 29**.

Figura 29. Plantilla de carga de recomendaciones

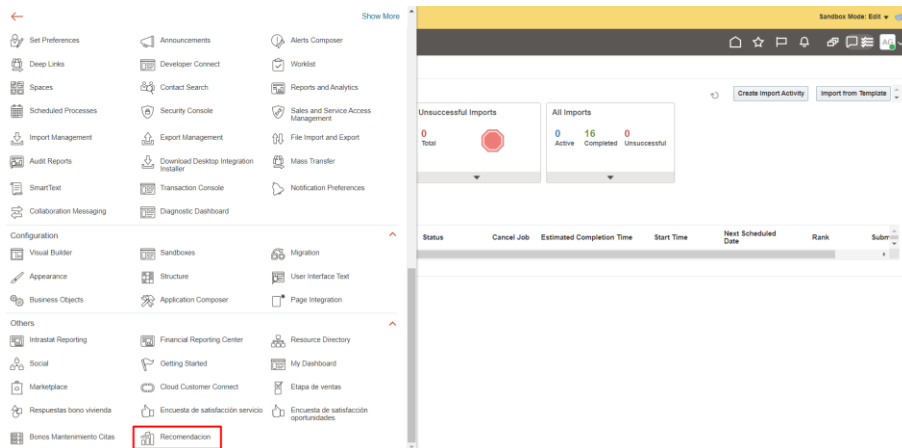


Nota. Elaboración propia.

5.2.5. Visualización de recomendaciones

Dentro de Oracle Engagement Cloud, fue necesario configurar las pantallas del objeto “Recomendación” para añadir los campos creados previamente en pantalla. De esta manera, el ejecutivo de ventas podrá consultar la información en cualquier momento dentro de la herramienta y analizar las recomendaciones generadas. En la **Figura 30**, el objeto de Recomendación creado se muestra en pantalla dentro de Oracle Engagement Cloud.

Figura 30. Visualización del objeto custom Recomendación



Nota. Elaboración propia.

La **Figura 31** muestra la lista de las recomendaciones cargadas en Engagement Cloud, desde la vista del ejecutivo de ventas.

Figura 31. Lista de recomendaciones cargadas

A screenshot of the Oracle Engagement Cloud interface showing a list of recommendations. The table has columns for 'Recomendacion Name', 'Last Update Date', 'Created By', and 'Owner'.

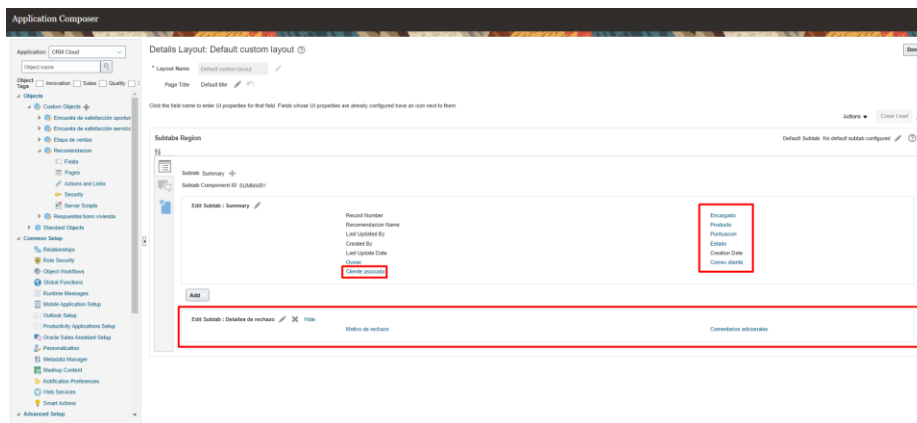
Recomendacion Name	Last Update Date	Created By	Owner
'Prueba01'	14-05-2022	andre.gonzalez@sumtech.com	Andre Go...
REC-00000001	15-05-2022	andre.gonzalez@sumtech.com	Andre Go...

Propuesta de solución de minería de datos basada en la metodología CRISP-DM utilizando los datos almacenados en Oracle Engagement Cloud, para la recomendación personalizada de productos a clientes, durante el primer semestre de 2022

Nota. Elaboración propia.

Como parte de la propuesta, se configuró la pantalla de “Detalles” del objeto *custom* “Recomendaciones” para incluir los campos mencionados. La configuración de esta pantalla se muestra en la **Figura 32**.

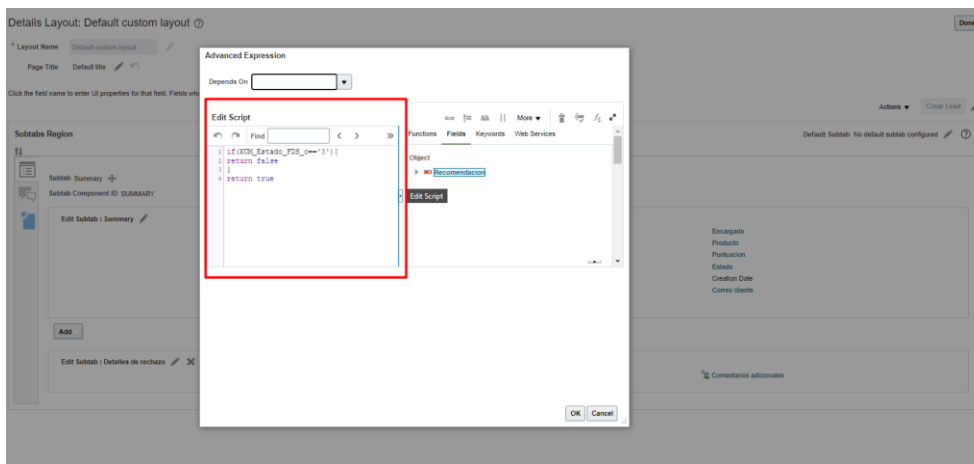
Figura 32. Configuración de pantalla de detalles



Nota. Elaboración propia.

Para las propiedades “Motivo de rechazo” y “Comentarios adicionales”, se configuró una expresión que únicamente muestra estos campos cuando el estado de la recomendación es “Rechazada”. La configuración de esta expresión se encuentra en la **Figura 33**.

Figura 33. Configuración de expresiones para campos de rechazo



Propuesta de solución de minería de datos basada en la metodología CRISP-DM utilizando los datos almacenados en Oracle Engagement Cloud, para la recomendación personalizada de productos a clientes, durante el primer semestre de 2022

Nota. Elaboración propia.

Cuando el ejecutivo de ventas consulta una recomendación, la pantalla se mostrará como en la **Figura 34**; en el caso de las recomendaciones, en estado “Pendiente” o “Aceptada”.

Figura 34. Detalle de la recomendación

The screenshot shows the 'Summary' page for recommendation REC-00000001. The record number is M_570004. The recommendation name is REC-00000001. The last updated by, created by, and last update date are all attributed to andre.gonzalez@suumtech.com on 15-05-2022. The owner is Andre Gonzalez. The associated client is Nathalie Jimenez Marin. The recommendation is assigned to Encargado, with Product 'Prueba' and a score of 89. The status is 'Pendiente' and the creation date is 14-05-2022. The client email is nathalie.jimenez@suumtech.com. A 'Detalles de rechazo' section is visible but empty.

Nota. Elaboración propia.

Para las recomendaciones en estado “Rechazada”, se muestran los campos de “Motivo de rechazo” y “Comentarios adicionales”, como en la **Figura 35**.

Figura 35. Detalles de rechazo de la recomendación

This screenshot is similar to Figure 34 but the status is 'Rechazada'. The 'Detalles de rechazo' section is highlighted with a red box and contains a dropdown menu for 'Motivo de rechazo' with the selected value 'Cliente no cumple con los requisitos del producto' and a text area for 'Comentarios adicionales'.

Nota. Elaboración propia.

5.2.6. Conversión de recomendación a Lead

Con el objetivo de convertir la recomendación a un *Lead* dentro de Oracle Engagement y enviar notificaciones por correo electrónico, se crearon dos *Server Script* en el objeto Recomendación. Estos se encuentran detallados en la **Tabla 38**.

Tabla 38. *Sever scripts configurados*

Nombre	Tipo	Descripción	Apéndice de referencia
Crear_Lead	Object Function	Permite la creación de un Lead a partir de la recomendación, asigna el nombre del producto, el cliente y el dueño del registro.	Apéndice AU. Configuración de <i>server script</i> Crear Lead.
Asignar_Correo	Trigger: BeforeUpdate	Permite asignar el correo electrónico del cliente a la recomendación, para posteriormente utilizarlo dentro de campañas de mercadeo.	Apéndice AV. Configuración de <i>server script</i> Asignar Correo.

Nota. Elaboración propia.

Adicionalmente, para la creación del Lead a partir de la recomendación, fue necesario elaborar un *Action and Link* dentro de *Engagement Cloud*, el cual se detalle en la **Tabla 39**.

Tabla 39. *Action configurado*

Nombre	Descripción	Apéndice de referencia
Crear Lead	Crea un nuevo Lead a partir de la recomendación dentro de	Apéndice AW. Configuración de action Crear Lead.

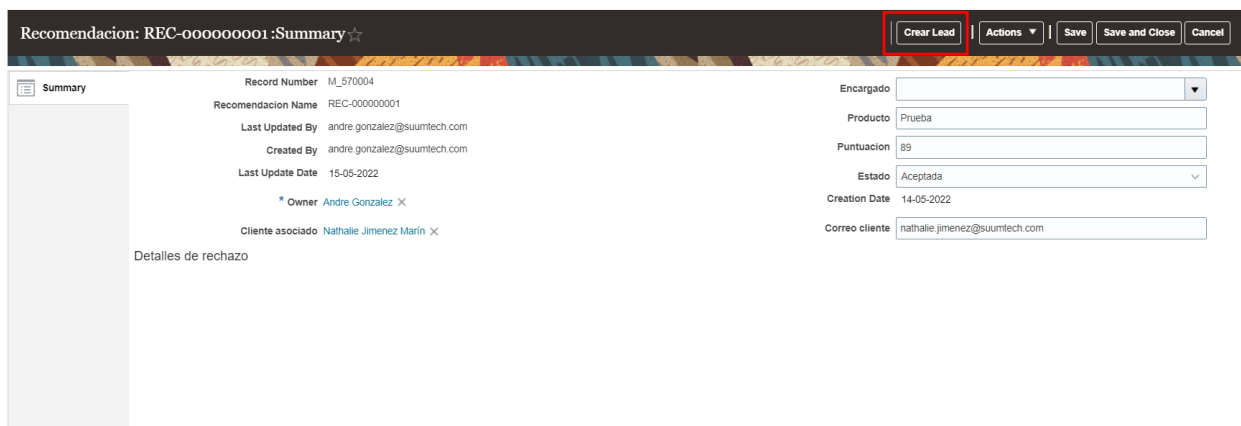
Propuesta de solución de minería de datos basada en la metodología CRISP-DM utilizando los datos almacenados en Oracle Engagement Cloud, para la recomendación personalizada de productos a clientes, durante el primer semestre de 2022

Nombre	Descripción	Apéndice de referencia
	Engagement, con la información del producto, cliente y encargado.	

Nota. Elaboración propia.

Desde la posición del ejecutivo de ventas, al momento de consultar una recomendación se le mostrará una opción llamada “Crear Lead”, como evidencia en la **Figura 36**.

Figura 36. Botón de creación de lead



Nota. Elaboración propia.

Cuando el ejecutivo de ventas selecciona la opción de “Crear Lead”, Engagement Cloud establecerá de manera automática el Lead asociado al cliente, dentro del módulo de prospectaciones. Este podrá ser consultado, tal y como se muestra en la **Figura 37**.

Figura 37. Evidencia del Lead creado

Summary

Lead Number: M_86098

Lead Name: M_86098 - 15/5/2022

Existing Account: Nathalie Jimenez Marín

Tipo de Identificación: CEDULA DE IDENTIDAD

Número de Identificación: [Redacted]

Correo electrónico: nathalie@gmail.com

Número de teléfono: [Redacted]

Attachments: None

Origen: [Dropdown]

Converted Date: dd-mm-yyyy

Expiration Date: dd-mm-yyyy

Tipo de Lead: [Dropdown]

Owner: [Dropdown] [Reject] [Accept]

Description: Prueba

Products

Browse Catalog

Name	Total Amount
No data to display.	

Datos del registro

Creation Date: 15-05-2022

Fecha de última actualización: 15-05-2022

Lead Created By: [Redacted]

Actualizado por: [Redacted]

Nota. Elaboración propia.

5.3. Creación de campañas con Eloqua

Las actividades relacionadas con la creación de campañas de mercadeo se encuentran fuera del alcance de este proyecto, debido a limitaciones técnicas y de tiempo, sin embargo, se proponen los pasos a seguir como parte de la futura implementación de la solución planteada.

5.3.1. Sincronización de los datos

La herramienta de Eloqua ofrece la capacidad de sincronizarse a través de un proceso de importación con Oracle Engagement Cloud configurable. Mediante esta sincronización es posible actualizar los datos en intervalos de tiempo de aproximadamente 15 minutos, de esta manera, se garantiza que la información sea consistente entre ambas herramientas y las campañas de mercadeo cuenten con los datos correctos.

Para el desarrollo de este proceso, se dispone de la documentación detallada en la **Tabla 40**, que puede ser utilizada como guía para realizar el proceso de sincronización de datos entre Engagement

y Eloqua. Esta documentación se encuentra disponible dentro del centro de documentación interna de XUM TECH en Confluence.

Tabla 40. Artículos de conocimiento para la configuración de Eloqua

Nombre del artículo de conocimiento	Descripción	Referencia
Guía Integración Eloqua-Sales para manejo de Leads y Oportunidades	Proporciona una serie de pasos a seguir para crear la sincronización de datos entre Engagament y Eloqua. Resume algunos aspectos clave de la integración para el registro de Leads y/o Oportunidades.	Guía Integración Eloqua-Sales para manejo de Leads y Oportunidades - Oracle - Confluence (atlassian.net)
Guía Custom Objects	Brinda una introducción a los objetos personalizados de Eloqua y guía sobre cómo gestionarlos. En estos objetos, el usuario puede almacenar datos adicionales de manera escalable y vincular esos datos a un contacto o registro de una cuenta.	Guía Custom Objects - Oracle - Confluence (atlassian.net)
Guía Videos de Eloqua	Consiste en una lista de reproducción de capacitaciones internas brindadas por compañeros de XUM TECH sobre el uso correcto de Eloqua para la gestión de correos electrónicos, campañas, segmentos, programas y reportes.	Guía Videos de Eloqua - Oracle - Confluence (atlassian.net)

Nota. Elaboración propia.

5.3.2. Envío de campañas publicitarias para la recomendación de productos

Como parte de la solución propuesta para la campaña de mercadeo de productos recomendados, es necesario crear al menos una plantilla de correo electrónico y una campaña dentro de Eloqua,

que permita informar al cliente los productos recomendados y los invite a adquirirlos por medio de sus canales electrónicos.

Adicionalmente, en la **Tabla 41** se brinda una lista de recursos informativos que pueden ser empleados durante la creación del correo electrónico y el desarrollo de la campaña de mercadeo en Eloqua.

Tabla 41. Artículos de conocimiento para la creación de campañas

Nombre del artículo de conocimiento	Descripción	Referencia
Cómo Crear nuevos campos de contactos	Muestra cómo crear un nuevo campo para un contacto. Estos campos se utilizan para vincular con el contacto información que por defecto Eloqua no posee.	Cómo Crear nuevos campos de contactos - Oracle - Confluence (atlassian.net)
Cómo Subir imágenes	Indica cómo subir imágenes al sitio de Eloqua.	Cómo Subir imágenes - Oracle - Confluence (atlassian.net)
Cómo Crear nuevo grupo de correo electrónico	Señala cómo editar una nueva cabecera para correo electrónico.	Cómo Crear nuevo grupo de correo electrónico - Oracle - Confluence (atlassian.net)
Cómo Editar cabecera correo electrónico	Muestra cómo editar una nueva cabecera para correo electrónico.	Cómo Editar cabecera correo electrónico - Oracle - Confluence (atlassian.net)

Nota. Elaboración propia.

Algunos aspectos que se deben tomar en cuenta durante la configuración de las campañas de mercadeo y correos electrónicos dentro de Eloqua son:

- Analizar en conjunto con el cliente los diseños necesarios para el correo electrónico que será utilizado dentro de las campañas publicitarias. De esta manera, será posible determinar

cuáles elementos de información, imágenes, encabezados, enlaces externos y pie de páginas se deben incluir.

- Examinar junto con el cliente la periodicidad con la que debe enviarse la campaña publicitaria; queda a criterio del cliente según las necesidades del negocio.
- Ajustar los correos electrónicos con el diseño de marca del cliente; como parte de la configuración de la campaña, es importante garantizar que los correos informativos cuenten con los formatos definidos por la empresa, esto incluye aspectos como el tipo, tamaños y colores de letra, así como colores de fondo, entre otros.

5.4. Análisis financiero de la propuesta

5.4.1. Cotización del proyecto académico

Como parte del análisis financiero, se estimó el costo aproximado de realizar la investigación correspondiente a este Trabajo Final de Graduación dentro de XUM TECH, para la creación de la propuesta que atienda la problemática actual.

Para esta estimación, se consideraron las siguientes suposiciones:

- El salario base utilizado corresponde a \$1000 dólares mensuales.
- Los cargos patronales equivalen a un 40.59% del salario mensual, es decir \$405.9, tomando en cuenta aspectos como aguinaldo, vacaciones, CCSS SEM, CCSS IVM, Cuota Patronal Banco Popular, Asignaciones familiares, IMAS, INA, Aporte Patrono Banco Popular, FCL, Fondo de pensiones complementario e INS. En el Apéndice CB. Cálculo de cargas patronales de la investigación se puede consultar el detalle del cálculo de cargos patronales.
- El trabajo de investigación comprende la etapa de anteproyecto y el desarrollo de la propuesta.
- La etapa de anteproyecto tuvo una duración aproximada de 12 semanas.
- La etapa de desarrollo de la propuesta tuvo una duración aproximada de 15 semanas.
- Un mes equivale a cuatro semanas de trabajo.

El resultado del análisis del costo de realizar el proyecto académico, presentado en la **Figura 38**, muestra que su duración aproximada fue de 27 semanas, durante las cuales el costo del salario del investigador correspondería a \$9,489.23.

Figura 38. Cotización del proyecto académico

Suposiciones	
Salario Base (mensual)	\$ 1,000.00
Cargas patronales	\$ 405.90
Etapas	Duración semanas
Anteproyecto	12
Trabajo Final de Graduación	15
Total semanas	27
Cantidad de meses	6.75
Costo total del TFG	\$ 9,489.83

Nota. Elaboración propia.

5.4.2. Cotización de la propuesta

Para esta cotización, se estimaron los costos asociados con el desarrollo e implementación de la solución propuesta en un ambiente productivo de un cliente de XUM TECH. Las suposiciones consideradas fueron las siguientes:

- El costo por hora de consultor corresponde a \$60.
- Los cargos patronales equivalen a un 40.59% del costo por hora del consultor, es decir \$24.35, tomando en cuenta aspectos como aguinaldo, vacaciones, CCSS SEM, CCSS IVM, Cuota Patronal Banco Popular, Asignaciones familiares, IMAS, INA, Aporte Patrono Banco Popular, FCL, Fondo de pensiones complementario e INS. En el Apéndice CC. Cálculo de cargas patronales de la cotización se puede consultar el detalle del cálculo de cargos patronales.
- El día laboral equivale a ocho horas de trabajo.

Propuesta de solución de minería de datos basada en la metodología CRISP-DM utilizando los datos almacenados en Oracle Engagement Cloud, para la recomendación personalizada de productos a clientes, durante el primer semestre de 2022

- El margen de ganancia utilizado es del 30%.
- De acuerdo con la entrevista inicial, documentada en el “Apéndice D. Resultados de la Entrevista inicial” existen al menos cuatro clientes interesados en una solución de recomendación de productos personalizada con los datos de Oracle Engagement Cloud. Estos serán utilizados para el cálculo de los costos e ingresos.

En la **Figura 39** se muestra el análisis financiero de la propuesta. Se realizó la estimación de tiempo en días que podrían tardar las actividades de desarrollo del modelo de minería de datos, configuración de CX, configuración de las campañas de mercadeo, control de calidad y el pase a producción de la funcionalidad para un cliente.

Figura 39. Cotización de la implementación

Modelo de minería de datos	
Entendimiento del negocio	2
Entendimiento de los datos	3
Exploración de datos	4
Limpieza de datos	4
Creación y pruebas del modelo de minería	3
Importación de resultados a CX	1
Configuración de CX	
Creación de analítico de exportación	2
Creación de campos y objetos	2
Configuración de pantallas	2
Creación de archivo de importación(plantilla)	1
Campañas de mercadeo	
Sincronización de datos CX-Eloqua	1
Creación de campaña de recomendación de productos	4
Control de calidad QA	
Pruebas de usuario UAT	4
Pase a producción	
Configuración en CX	2
Duración total (días)	35
Estimación de costos	
Costo de consultores con cargas patronales	\$23,618
Licenciamiento de Anaconda (2 meses)	\$30
Sumatoria	\$ 23,618.00
Impuestos (13%)	\$ 3,070.34
Costo total	\$ 26,688.34
Estimación de ingresos	
Precio de venta	\$ 34,694.84

Nota. Elaboración propia.

El resultado de este análisis indica que el proyecto tendría una duración aproximada de 35 días, con un costo total de \$23,618, tomando en cuenta aspectos como el salario del talento humano requerido, el licenciamiento de herramientas como Anaconda y los impuestos del 13%. El precio de venta se estimó utilizando un margen de ganancia de 30% sobre el costo total, así el resultado sería de \$34,694.84.

En la **Figura 40**, al tomar en cuenta que existen al menos cuatro clientes interesados en la propuesta, se estimaron los costos e ingresos que tendría XUM TECH si realizara la implementación de esta solución con cada uno de ellos.

Figura 40. Cotización de ganancias de la implementación

Lista de Proyectos	Costo	Ingresos	Ganancias
Proyecto I	\$ 26,688.34	\$ 34,694.84	\$ 8,006.50
Proyecto II	\$ 26,688.34	\$ 34,694.84	\$ 8,006.50
Proyecto III	\$ 26,688.34	\$ 34,694.84	\$ 8,006.50
Proyecto IV	\$ 26,688.34	\$ 34,694.84	\$ 8,006.50

Nota. Elaboración propia.

Como resultado del análisis anterior, se obtuvo que realizar la implementación con cuatro clientes implicaría un costo aproximado de \$75.936, generaría ingresos de \$98.716,80 y una ganancia de \$22.780,80.

5.4.3. Indicadores financieros

Dentro del análisis económico, la propuesta se evaluó con la ayuda del indicador financiero retorno sobre la inversión (ROI). Este, de acuerdo con Silva (2020), es una métrica de rendimiento utilizada para evaluar que tan beneficiosa es una inversión en relación con su costo, en otras palabras, el ROI permite a las empresas evaluar la efectividad de una inversión. La fórmula utilizada para el cálculo del retorno sobre la inversión de esta propuesta se encuentra en la **Figura 41**.

Figura 41. *Fórmula de cálculo del ROI*

$$\text{ROI} = (\text{Rendimiento obtenido} - \text{inversión}) \div \text{inversión} \times 100$$

Nota. Tomado de Silva (2020).

Para el cálculo del ROI de este proyecto, se realizaron las siguientes suposiciones:

- El rendimiento corresponde a los ingresos obtenidos luego de implementar la solución en los cuatro clientes mencionados, esto equivale a \$138,779.37.
- La inversión corresponde a la sumatoria de los costos incurridos luego de implementar la solución en los cuatro clientes mencionados y el costo asociado al desarrollo del trabajo final de graduación, esto equivale a \$116,243.19.

El cálculo del ROI se muestra en la **Figura 42**, donde se puede observar que la implementación de esta propuesta generaría un retorno del 19% sobre la inversión. Este resultado es positivo porque indica que la propuesta podría generar más dinero del invertido, es decir, daría ganancias a XUM TECH.

Figura 42. *Cálculo del ROI*

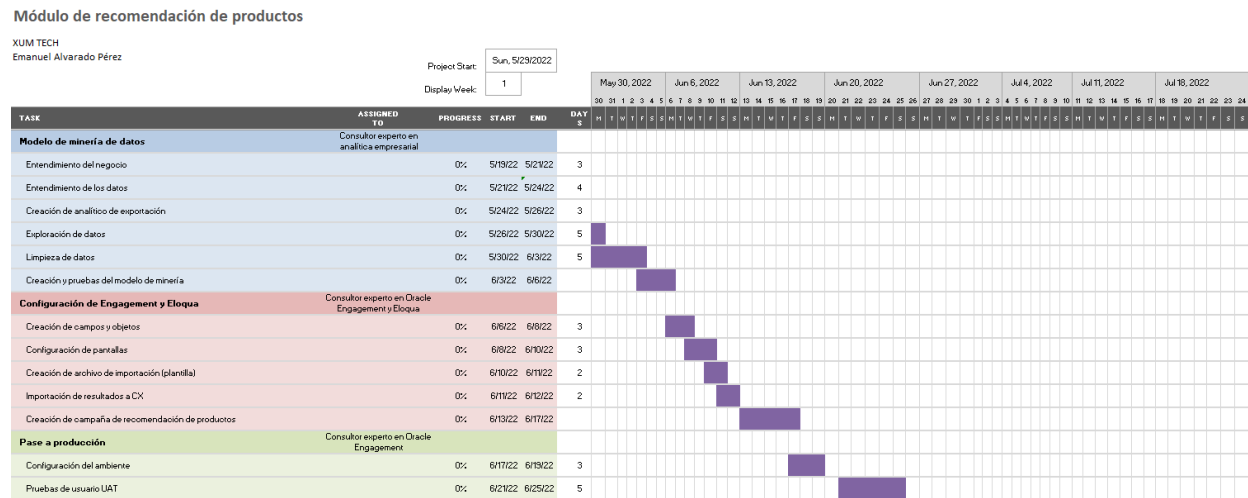
Resumen	
Costo del TFG	\$ 9,489.83
Costo total de la implementación	\$ 106,753.36
Costo de inversión	\$ 116,243.19
Ganancias totales de la implementación	\$ 138,779.37
Rendimiento obtenido	\$ 138,779.37
Cálculo del ROI	
ROI = (Rendimiento obtenido – inversión) ÷ inversión x 100	
19%	

Nota. Elaboración propia.

5.4.4. Hoja de ruta de la implementación

Como parte de la propuesta de solución, en la **Figura 43** se brinda una hoja de ruta a seguir para la implementación real del sistema de recomendación personalizada de productos. Dentro de esta, se listan las actividades necesarias y su duración estimada; las fechas para el inicio del proyecto son ficticias.

Figura 43. Propuesta de hoja de ruta



Nota. Elaboración propia.

Por su parte, en la **Tabla 42** se detallan los roles necesarios para el desarrollo del proyecto.

Tabla 42. Talento humano sugerido para la implementación

Rol	Responsabilidad
Administrador de proyectos	Brinda seguimiento al proyecto con el objetivo de cumplir con los entregables, dentro del tiempo y con la calidad esperada.
Consultor experto en Oracle Engagement Cloud	Debe conocer el módulo de ventas (Lead, Opportunities, Contacts, Accounts) del cliente.

Rol	Responsabilidad
	Es experto en la configuración de campos, pantallas, <i>objetos custom</i> y <i>scripts</i> .
Consultor experto en Eloqua	Configura los procesos de sincronización entre Eloqua y Engagement, desarrolla los elementos necesarios para generar la campaña de mercadeo.
Consultor experto en analítica empresarial	Realiza los procesos de extracción, transformación y carga de datos en el ambiente de Engagement. También, es responsable de realizar las pruebas de los modelos de minería y validar sus resultados.

Nota. Elaboración propia.

5.4.5. Beneficios financieros y no financieros

Basado en los resultados de la presente propuesta y en la bibliografía detallada a lo largo del capítulo 1.3.3 se resumen los principales beneficios financieros y no financieros obtenidos por XUM TECH a partir del desarrollo de este proyecto:

Beneficios financieros

- Se llevó a cabo el análisis financiero de la solución propuesta, como resultado el ROI sería de aproximadamente 19%, como se detalló en la sección 5.4.3.

Beneficios no financieros

- Se verificó la factibilidad de realizar una solución basada en los datos de Oracle Engagement Cloud para la recomendación personalizada de productos.
- Se realizó una propuesta donde se identificaron los elementos más importantes para el desarrollo de un proyecto de minería de datos, enfocado en la recomendación personalizada de productos. Junto con la propuesta se incluyen aspectos como:

- Lista de datos necesarios de Oracle Engagement y configuración de analíticos de extracción.
 - Lista de herramientas sugeridas para las tareas de almacenamiento y procesamiento.
 - Modelo de minería seleccionado.
 - Tareas de limpieza y procesamiento de datos.
 - Documentación de referencia para generar mayor conocimiento interno en temas de minería y ciencia de datos.
- Se diseñó una hoja de ruta de la propuesta, donde se indica de manera explícita la lista de tareas necesarias para el desarrollo del proyecto; adicionalmente, se documentó el talento humano necesario.

Basado en los beneficios directos definidos inicialmente en el capítulo 1.3.3, luego de realizar la propuesta de solución se obtuvieron los siguientes beneficios:

- Se realizó un análisis financiero que permite validar la viabilidad de la solución de minería de datos para la recomendación personalizada de productos.
- Se realizó una propuesta de solución de recomendación personalizada de productos utilizando los datos almacenados en el CRM Oracle Engagement Cloud.
- Se identificaron nuevas formas de utilizar los datos de Oracle Engagement Cloud, para facilitar la toma de decisiones de los clientes (en este caso la recomendación de productos).
- Se identificaron las herramientas necesarias a lo largo del análisis de los datos para generar el modelo predictivo.

Dentro de este capítulo se detalló la solución propuesta para atender la problemática descrita en la sección 1.3 y satisfacer los requerimientos definidos en la sección 4.1.2, también se demostró mediante un análisis financiero la viabilidad del proyecto, adicionalmente se brindó una hoja de ruta le permitirá a XUM TECH la implementación del proyecto. En el próximo capítulo se indican los hallazgos más relevantes asociados a los objetivos planteados dentro de la sección 1.4 obtenidos a partir del desarrollo de esta propuesta.

6. Conclusiones

En este capítulo del documento, se encuentran las conclusiones y principales hallazgos obtenidos del proceso investigativo, los cuales están relacionados a los objetivos específicos.

6.1. Conclusiones para el objetivo específico uno

A continuación, se describen las conclusiones referentes al siguiente objetivo específico:

“Analizar las necesidades capturadas por el departamento comercial de XUM TECH, para la identificación de los requerimientos necesarios para la propuesta”.

1. Tras la recolección de requerimientos, se determinó la necesidad de utilizar la información de Oportunidades, Clientes y Productos para crear un modelo de recomendación personalizada de producto.
2. Mediante la recolección de requerimientos funcionales y no funcionales, se identifica que la propuesta incorpora elementos que deben ser configurados dentro de Oracle Engagement Cloud, Oracle Eloqua y el sistema recomendador de productos.
3. De acuerdo con lo expresado por el departamento comercial de XUM TECH, se determinó la necesidad de adquirir servicios externos a la herramienta Oracle Engagement Cloud, debido a sus limitantes para el procesamiento de información y para realizar predicciones.

6.2. Conclusiones para el objetivo específico dos

A continuación, se describen las conclusiones referentes al siguiente objetivo específico:

“Evaluar varios modelos predictivos mediante la utilización de los datos obtenidos de Oracle Engagement Cloud, para la selección de un modelo adecuado de minería de datos que permita la recomendación personalizada de productos a clientes”.

1. Luego de evaluar distintos modelos predictivos de clasificación, se determinó que el modelo con mejores resultados fue el *Collaborative Filtering*, porque permite cumplir prácticamente todos los requerimientos. Sin embargo, mediante este modelo solo es posible

predecir el producto de interés del cliente y la puntuación, pero no un monto de venta estimado.

2. Las tareas de exploración, limpieza y procesamiento de datos, así como la creación y evaluación de los modelos de recomendación de productos son posibles con el uso de artefactos como *Notebooks*. Estos son brindados por el *software* Anaconda, que permite realizar de manera ágil estas tareas y ofrece una gran cantidad de recursos como bibliotecas, comunidades y documentación útiles para el desarrollo de proyectos.
3. Los modelos de clasificación y de segmentación evaluados como Naive Bayes y KMeans para la recomendación de productos no fueron sugeridos por no cumplir con los requerimientos establecidos por el negocio. Por otro lado, estos modelos representan mayores esfuerzos de transformación de datos como se mencionó en la sección 4.2.3, adicionalmente permiten realizar solamente una recomendación para cada cliente sin asignar un puntaje a diferencia del filtrado colaborativo que crea múltiples recomendaciones y asigna los puntajes correspondientes basado en los patrones encontrados.
4. El uso de marcos de referencia reconocidos mundialmente como CRISP-DM facilita a la organización el desarrollo de modelos de minería de datos y garantiza que estos modelos se encuentren alineados a las necesidades del negocio, gracias a su estructura flexible y enfocada en atender los objetivos empresariales.

6.3. Conclusiones para el objetivo específico tres

A continuación, se describen las conclusiones referentes al siguiente objetivo específico:

“Recomendar las herramientas de extracción, almacenamiento y procesamiento de los datos, para la integración de los datos de Oracle Engagement Cloud y el modelo predictivo”.

1. Según la información recolectada, se determinó que la herramienta ideal para realizar las tareas de extracción de datos es *Oracle Reports and Analytics*. Con la ayuda de esta, es

posible consumir reportes de manera periódica con información en tiempo real, modificarlos y utilizarlos en la creación de modelos de minería de datos.

2. Las herramientas sugeridas como los *Custom Object* de *Oracle Engagement* y *Anaconda* se encargan de realizar las tareas de almacenamiento y procesamiento de datos. Al mismo tiempo, ofrecen la capacidad de integrarse con otras soluciones como *Oracle Eloqua* para adaptarse a las necesidades del cliente y permitir la recomendación personalizada de productos a un bajo costo.

6.4. Conclusiones para el objetivo específico cuatro

A continuación, se describen las conclusiones referentes al siguiente objetivo específico:

“Elaborar un análisis financiero de la solución de minería de datos propuesta, para la verificación de la viabilidad del proyecto y la identificación de beneficios financieros y no financieros.”

1. La cotización de la propuesta brinda una estimación del costo total del proyecto, tomando en cuenta los aspectos fundamentales de XUM TECH, como los salarios, duración de las tareas, las tarifas de licenciamiento e impuestos.
2. El plan de implementación diseñado contiene la hoja de ruta sugerida para el desarrollo de la solución de recomendación de productos en un ambiente productivo e indica los aspectos que deben ser configurados en cada uno de los sistemas.
3. El análisis financiero de la propuesta detallado en la sección 5.4.3 indica que la implementación del proyecto podría generar un retorno sobre la inversión de aproximadamente 19 %.

7. Recomendaciones

En este capítulo, se presentan las recomendaciones generadas a partir del trabajo final de graduación “Propuesta de solución de minería de datos basada en la metodología CRISP-DM utilizando los datos almacenados en *Oracle Engagement Cloud*, para la recomendación personalizada de productos a clientes”. Son las siguientes:

- Utilizar la propuesta planteada en este proyecto como guía para realizar las tareas de exportación, procesamiento y limpieza de datos, y configuración de ambientes de esta manera XUM TECH podrá ofrecer a sus clientes una solución de recomendación personalizada de productos.
- Capacitar al personal actual de XUM TECH en áreas de Ciencia de Datos, con el objetivo de mejorar las habilidades técnicas y poder resolver problemas de mayor complejidad que puedan presentarse con futuros clientes.
- Implementar mecanismos dentro de la herramienta de *Oracle Engagement Cloud*; por ejemplo, validaciones de las entradas de texto en la información del cliente, oportunidad y producto, con el propósito de mejorar la calidad de los datos recolectados y facilitar su análisis.
- Debido a la diferencia entre ambientes de los clientes de XUM TECH y las particularidades de cada uno, indicada durante la etapa de identificación del proceso actual de ventas en *Oracle Engagement Cloud* (sección 4.1.1), para futuros proyectos de minería de datos se recomienda realizar al menos una sesión con el consultor experto en el ambiente. El objetivo sería llevar a cabo las tareas de entendimiento tanto del proceso como de los datos, para garantizar que los datos utilizados en el modelo sean los correctos.
- Durante la etapa de entendimiento de datos, se sugiere contar con la presencia de al menos un miembro del equipo de ventas o consultor experto en el ambiente de *Oracle Engagement Cloud*, para validar el catálogo de productos y evitar registros e información duplicada.

- Cuando las campañas de mercadeo se activen, se recomienda utilizar los mecanismos de monitoreo ofrecidos por la herramienta de mercadeo Oracle Eloqua para dar seguimiento a las campañas de recomendación de productos y validar que estas sean efectivas.
- En el proceso de creación de campañas de mercadeo, se recomienda seguir las buenas prácticas establecidas dentro de XUM TECH, almacenadas en el centro de documentación interna *Confluence* para garantizar que la campaña sea lo más efectiva posible, previendo posibles riesgos como correos no deseados en el buzón del cliente.
- Dentro de la propuesta, no fue posible realizar la creación y configuración de la campaña de mercadeo, sin embargo, se recomienda revisar la sección 5.3, donde se encuentra la hoja de ruta y documentos de referencia útiles para futuras implementaciones.
- En proyectos similares, se sugiere añadir márgenes adicionales de tiempo durante la planificación basado en la complejidad del proyecto, con el objetivo de prevenir imprevistos que puedan presentarse durante la implementación del proyecto.
- Para futuros proyectos donde los clientes estén interesados en implementar una solución de recomendación de productos basado en preferencias, se recomienda desarrollar nuevos mecanismos de recolección de datos, como encuestas, donde los usuarios puedan calificar los productos utilizando escalas numéricas y así sea posible identificar sus gustos reales de una manera más sencilla.
- Incluir mayor cantidad de atributos en los productos almacenados dentro de *Engagement*, para completar con información más detallada acerca de las características del producto, de esta manera XUM TECH podría evaluar nuevos modelos enfocados en la recomendación de productos como por ejemplo el filtrado basado en contenido, que utiliza las características únicas de cada producto y recomienda productos con características similares a los clientes.
- En caso de requerir predicciones adicionales como montos de venta estimados se sugiere explorar otras *Subject Areas* dentro de *Engagement Cloud* para realizar la exportación de los datos históricos correspondientes. De igual manera, es posible utilizar las herramientas

Propuesta de solución de minería de datos basada en la metodología CRISP-DM utilizando los datos almacenados en Oracle Engagement Cloud, para la recomendación personalizada de productos a clientes, durante el primer semestre de 2022

propuestas dentro de este proyecto como *Reports and Analytics* y Anaconda para realizar las predicciones, mediante el uso de modelos de regresión.

8. Referencias bibliográficas

- Araya, N. (2019). *Desarrollo de una solución basada en inteligencia de negocios para la adopción de los datos que forman parte de la cadena de suministro de Procter & Gamble*. Tec Digital. https://tecdigital.tec.ac.cr/dotlrn/classes/ATI/TI5904/S-1-2021.CA.TI5904.1/file-storage/view/Apuntes%2Fanteproyectos%2Fanteproyecto_Natalia_Araya_Chaves.pdf
- Azevedo, A. y Santos, M. (2008). *KDD, semma and CRISP-DM: A parallel overview*. 182-185. https://www.researchgate.net/publication/220969845_KDD_semma_and_CRISP-DM_A_parallel_overview
- Barboza, J. (2019). *Desarrollo de herramienta de inteligencia de negocios para interpretación, análisis y evaluación de datos relacionados al procesamiento de criptomonedas Caso: Progressio Digital*. Tec Digital. https://tecdigital.tec.ac.cr/dotlrn/classes/ATI/TI5904/S-1-2021.CA.TI5904.1/file-storage/view/Apuntes%2Fanteproyectos%2Fanteproyecto_Jos%C3%A9_Andr%C3%A9s_Barboza.pdf
- Bermúdez, L. (2020). *Solución de inteligencia de negocios para el análisis de datos en el ERP Business Pro*. Tec Digital. https://tecdigital.tec.ac.cr/dotlrn/clubs/Com.ATI/file-storage/view/documentos-trabajo-final-de-graduaci-n%2Fproyectos-finales-de-graduaci-n-p-blicos%2F2020%2FSolucion_Inteligencia_Negocios_para_Analisis_Datos_ERP_Lenynd_Bermudez_I-2020.pdf
- Brita Inteligencia Artificial. (12 de octubre de 2020). *Recomendación de Productos con Inteligencia Artificial*. BRITA México. <https://brita.mx/recomendacion-de-productos-con-inteligencia-artificial-machine-learning>

- Chaffey, D. y Ellis-Chadwick, F. (2014). *Marketing digital: estrategia, implementación y práctica*. Pearson Educación. <http://www.ebooks7-24.com.ezproxy.itcr.ac.cr/?il=3470>
- Chapman, P. (2000). *CRISP-DM 1.0: Step-by-step Data Mining Guide*. SPSS. <https://www.the-modeling-agency.com/crisp-dm.pdf>
- Cook, H., Adrian, M., Greenwald, R., Ronthal, A., y Russom, P. (14 de diciembre de 2021). *Magic Quadrant for Cloud Database Management Systems*. Gartner. <https://www.gartner.com/doc/reprints?id=1-28F8N1L2&ct=211213&st=sb>
- Da Silva, D. (20 de abril de 2020). *¿Cómo calcular el ROI? Fórmula con ejemplos*. Zendesk. <https://www.zendesk.com.mx/blog/roi-como-calcular-sacar/>
- Data Science Process Alliance. (s.f). *What is CRISP DM?*. <https://www.datascience-pm.com/crisp-dm-2/>
- Deutschman, Z. (2022). *Recommender Systems: Machine Learning Metrics and Business Metrics*. Neptune. <https://neptune.ai/blog/recommender-systems-metrics>
- Fawcett, A. (11 de febrero de 2021). *Data Science in 5 Minutes: What is One Hot Encoding?*. Educative. <https://www.educative.io/blog/one-hot-encoding>
- Gartner. (2021). *Market Guide for Multipersona Data Science and Machine Learning Platforms*. Daitaku. <https://content.dataiku.com/market-guides/gartner-multipersona-dsml-market-guide>
- Gartner. (s.f.). *¿Qué es el Magic Quadrant de Gartner?*. Gartner España. <https://www.gartner.es/es/metodologias/magic-quadrant>
- Gopal, M. (2019). "An Introduction to Analytics." Chap. 9.1 in *Applied Machine Learning*. 1st ed. New York: McGraw-Hill Education. <https://www-accessengineeringlibrary-com.ezproxy.itcr.ac.cr/content/book/9781260456844/toc-chapter/chapter9/section/section2>

- Gopal, M. (2019). “The Crisp-Dm (Cross-Industry Standard Process for Data Mining) Model”. Chap. 9.2 in *Applied Machine Learning*. (1ª ed.). New York: McGraw-Hill Education. <https://www-accessengineeringlibrary-com.ezproxy.itcr.ac.cr/content/book/9781260456844/toc-chapter/chapter9/section/section5>
- Han, J., Kamber, M. y Pei, J. (2012). *Data Mining Concepts and Techniques*. <http://myweb.sabanciuniv.edu/rdehkharghani/files/2016/02/The-Morgan-Kaufmann-Series-in-Data-Management-Systems-Jiawei-Han-Micheline-Kamber-Jian-Pei-Data-Mining.-Concepts-and-Techniques-3rd-Edition-Morgan-Kaufmann-2011.pdf>
- Hernández, R., Baptista, P. y Fernández, C. (2014). *Metodología de la investigación*. McGraw-Hill. <http://www.ebooks7-24.com.ezproxy.itcr.ac.cr/?il=285>
- Hurwitz, J. y Kirsch (2018). *Machine Learning for Dummies IBM Limited Edition*. <https://www.ibm.com/downloads/cas/GB8ZMQZ3?msclkid=6369d8b1a8d311ec96dad2318a19297c>
- IBM. (s.f). *CRISP-DM Help Overview*. Documentación de IBM. <https://www.ibm.com/docs/en/spss-modeler/SaaS?topic=dm-crisp-help-overview>
- IBM- (2014). *El proceso de minería de datos*. Documentación de IBM. <https://www.ibm.com/docs/es/db2/11.1?topic=miner-data-mining-process>
- Kharwal, A. (5 de julio de 2020). *Naive Bayes Classification in Machine Learning*. Clever Programmer. <https://thecleverprogrammer.com/2020/07/05/naive-bayes-classification-in-machine-learning/>
- Kumar, V. (23 de julio de 2021). *What is K-Means algorithm and how it works*. Towards Machine Learning. <https://towardsmachinelearning.org/k-means/>
- Lahon, A. (2020). *What is Evaluation metrics and When to use Which metrics?*. <https://medium.com/the-rise-of-unbelievable/what-is-evaluation-metrics-and-when-to-use-which-metrics-23d16446d690>
- Laudon, K. C. y Laudon, J. P. (2022). *Management information systems*. Pearson Educación. <http://www.ebooks7-24.com.ezproxy.itcr.ac.cr/?il=16588>

- Laudon, K. C. y Guercio Traver, C. (2020). *E-commerce*. Pearson Educación. <http://www.ebooks7-24.com.ezproxy.itcr.ac.cr/?il=13877>
- Madison, D. (2005). *Process mapping, process improvement, and process management: a practical guide to enhancing work and information flow*. Paton Press.
- Mora, V. (2018). *Propuesta de un modelo de predicción para la infraestructura tecnológica en proyectos de migración. Caso: Mobilize.net*. Tec Digital. https://tecdigital.tec.ac.cr/dotlrn/clubs/Com.ATI/file-storage/view/documentos-trabajo-final-de-graduaci-n%2Fproyectos-finales-de-graduacion-privados%2F2018%2FMod_predicci_infraest_tecno_proyec_migraci%C3%B3n_Vinicio_Mora-II-2018.pdf
- Jiménez, N. (2021). *Desarrollo de prueba de concepto sobre las campañas de mercadeo con el canal de Facebook Messenger para su implementación en el catálogo de servicios de la empresa*.
- Navarro, I. (21 de agosto de 2020). *Guía / Prevenir que el correo electrónico entre a la bandeja de correo no deseado (SPAM)*. Confluence. <https://xumtech.atlassian.net/wiki/spaces/BCO/pages/502956045/Gu+a+Prevenir+que+el+correo+electr+nico+entre+a+la+bandeja+de+correo+no+deseado+SPAM>
- Ñaupas, H., Mejía, E., Novoa, E., & Villagómez, A. (2013). *Metodología de la investigación Cuantitativa - Cualitativa y Redacción de la Tesis*. Bogotá Colombia: Ediciones de la U.
- Oracle. (2021). *Oracle*. Obtenido de Oracle Eloqua Marketing Automation: https://www.oracle.com/mx/cx/marketing/automation/?source=:ad:pas:go:eng:a_lad:7170000085048216-58700007223387025-p64807255717:RC_WWMK210621P00064:MainAd&SC=:ad:pas:go:eng:a_lad::RC_WWMK210621P00064:MainAd&gclid=CjwKCAjwyvaJBhBpEiwA8d38vHUdFwPmjoVKxK6b4E

- Oracle. (s.f.). *Oracle Engagement Cloud*. <https://www.oracle.com/a/ocom/docs/oracle-engagement-cloud.pdf>
- Oracle. (s.f.). *¿Qué es el aprendizaje autónomo?* <https://www.oracle.com/mx/data-science/machine-learning/what-is-machine-learning/>
- Romero, David y Vernadat, François. (2016). *Enterprise Information Systems State of the Art: Past, Present and Future Trends. Computers in Industry*. 79. 10.1016/j.compind.2016.03.001.
- Pimienta, J. y Orden, A. (2012). *Metodología de la investigación*. Pearson.
- Pozas, J. L. B. (2021, 28 septiembre). *¿Qué es la minería de datos: ¿cuál es su valor comercial, cuánto gana un especialista en data mining?* *CIO MX*. <https://cio.com.mx/que-es-la-mineria-de-datos-cual-es-su-valor-comercial-cuanto-gana-un-especialista-en-data-mining/>
- Stair, R. M. y Reynolds, G. W. (2017). *Principios de sistemas de información*. Cengage Learning. <http://www.ebooks7-24.com.ezproxy.itcr.ac.cr/?il=3979>
- Singla, R. (03 de mayo de 2020). *Recommendation Systems-Demographic Filtering (Part 1)*. Github. https://github.com/rahuls0959/ds-blog/blob/master/_notebooks/2020-05-03_Recommendation%20System_Part%201.ipynb
- Sommerville, I. (2022). *Software Engineering*. Pearson Education. <https://docs.google.com/viewer?a=v&pid=sites&srcid=ZGVmYXVsdGRvbWFpbnxwaXBlnBhc2hhfGd4OjRmODdjNzY3NDU4Njc4ZWE&msclkid=d72642d2a8c711ec883d42c97ec10ec2>
- Universidad Tecnológica Latinoamericana en línea (UTEL). (2019). *Cuadro comparativo*. http://gc.initelabs.com/recursos/files/r162r/w18141w/cuadro_comparativo.pdf
- White, S. A. (2009). *Guía de referencia y modelado BPMN*. Uruguay: Future Strategies.
- XUM TECH. (sf). Inicio [página de LinkedIn]. LinkedIn <https://cr.linkedin.com/company/suumtech>

Propuesta de solución de minería de datos basada en la metodología CRISP-DM utilizando los datos almacenados en Oracle Engagement Cloud, para la recomendación personalizada de productos a clientes, durante el primer semestre de 2022

XUM TECH. (4 de enero de 2020). *Lo que nos hace únicos*. Suumtech.
<https://suumtech.com/sobrenosotros>

XUM TECH. (8 de febrero de 2021). *Reunión mensual febrero 2021*. [Diapositiva de PowerPoint]. SharePoint empresarial.

XUM TECH. (agosto de 2019). *Reunión mensual agosto 2019*. [Diapositiva de PowerPoint]. Sharepoint empresarial.

9. Apéndices

9.1. Apéndice A. Plantilla de minutas para reunión



MINUTA DE REUNIÓN

Proyecto: Propuesta de solución de minería de datos basada en la metodología CRISP-DM utilizando los datos almacenados en Oracle Engagement Cloud, para la recomendación personalizada de productos a clientes, durante el primer semestre de 2022

Reunión No.		Fecha:	
Lugar:		Hora Inicio/Finalización:	
Objetivo de la reunión:			
Participantes:	Presentes:		
	Ausentes:		
Temas Tratados			
No.	Asunto	Comentarios	Acuerdos
Próxima reunión			
Temas a tratar	Fecha	Convocados	

9.2. Apéndice B. Plantilla para la gestión de cambios

Datos Generales del Cambio	
ID	
Solicitante	
Responsable de la implementación	
Fecha de solicitud del cambio	
Fecha de realización del cambio	
Estado de cambio	Aprobado/En Revisión/ Rechazado

Detalles del Cambio	
Categoría	
Descripción detallada	
Justificación	
Implicaciones de realizar el cambio	
Impacto	
Comentarios	

Revisado por:

Nombre tutor

Firma tutor

Aprobado por:

Nombre aprobador

Firma aprobador

Elaborado por:

Nombre del estudiante

Firma estudiante

9.3. Apéndice C. Entrevista inicial TFG

Participantes	
Fecha	
Objetivo	
Preguntas	Respuesta brindada
¿Se ha hecho un proyecto similar a este anteriormente en XUM TECH?	
¿Oracle Engagament tiene alguna limitación para hacer un modelo predictivo de recomendación de productos?	
¿Por qué XUM TECH desea realizar este proyecto?	
¿Algún cliente ha pedido un servicio de este tipo? ¿O ha mostrado interés?	
¿A cuáles miembros de XUM TECH se les podría solicitar información acerca de temas relacionados al proyecto?	
¿Existe algún experto en el área de datos en Oracle Engagaement Cloud u otras herramientas?	
¿Existe documentación relacionada a la necesidad presentada por los clientes?	

9.4. Apéndice D. Resultados de la Entrevista inicial

Participantes	<ul style="list-style-type: none"> Emanuel Alvarado Pérez Alex Ureña Cordero
Fecha	25/10/2021
Objetivo	Conocer el contexto y antecedentes de la problemática actual de la empresa.
Preguntas	Respuesta brindada
¿Se ha hecho un proyecto similar a este anteriormente en XUM TECH?	Se planteó una iniciativa con el Banco Nacional de Costa Rica, sin embargo, no se concretó.
¿Oracle Engagament tiene alguna limitación para hacer un modelo predictivo de recomendación de productos?	No es una herramienta de procesamiento de datos, esto implica que no es capaz de tomar decisiones sobre los datos (estadística, probabilidades) tipo modelos. Es más transaccional y utilizada para modelar procesos comerciales. La reportería disponible es básica, únicamente transaccional y de BI.
¿Por qué XUM TECH desea realizar este proyecto?	<ul style="list-style-type: none"> Ampliar los servicios ofrecidos para los clientes. Obtener mayores ingresos. Mejorar la satisfacción de los clientes.
¿Algún cliente ha pedido un servicio de este tipo? ¿O ha mostrado interés?	Al menos cuatro clientes diferentes han mostrado interés en la implementación de un módulo de recomendación de productos, todos estos clientes pertenecen al sector de banca y finanzas.
¿A cuáles miembros de XUM TECH se les podría solicitar información acerca de temas relacionados al proyecto?	Se puede consultar a miembros con roles como administrador de proyectos, consultor experto de infraestructura, consultores de herramientas Oracle y expertos del área comercial.
¿Existe algún experto en el área de datos en Oracle Engagament Cloud u otras herramientas?	Si existen expertos del área de datos en Oracle, pero únicamente en temas de Inteligencia de Negocios (BI), generación de reportes y <i>dashboards</i> . Nunca se han implementado soluciones de Machine Learning o analítica predictiva.

Propuesta de solución de minería de datos basada en la metodología CRISP-DM utilizando los datos almacenados en Oracle Engagement Cloud, para la recomendación personalizada de productos a clientes, durante el primer semestre de 2022

¿Existe documentación relacionada a la necesidad presentada por los clientes?	Sí, existen documentos donde se especifican las historias de usuario y un mockup del resultado esperado.
---	--

9.5. Apéndice E. Entrevista #1 Proceso de recomendación/venta de productos en CX

Entrevista: Proceso de ventas en CX	
Participantes	
Fecha	
Objetivo	
Preguntas	Respuestas
¿Qué módulos interactúan en la ejecución de ventas?	
¿Cuáles pasos deben realizar los agentes de ventas para realizar una venta mediante la plataforma?	
¿Cómo se concreta una venta en CX?	
¿Existe algún repositorio de documentación donde pueda consultar información relacionada?	
¿Conoce alguna manera de recomendar productos de manera nativa en CX?	
¿Los usuarios de esta plataforma han indicado alguna disconformidad o una oportunidad de mejora en el proceso?	
Otras notas	

9.6. Apéndice F. Entrevista #2 Conocer los requerimientos/necesidades del negocio

Entrevista: Validación de requerimientos expresados por el negocio	
Participantes	
Fecha	
Objetivo	
Preguntas	Respuestas
Objetivo del Negocio	
Requerimientos Funcionales	
Requerimientos No Funcionales	
Datos por tomar en cuenta: <ul style="list-style-type: none"> • Información del cliente • Registro de compras 	
Otras notas	

9.7. Apéndice G. Entrevista #3 Verificar lista de requerimientos encontrados

Entrevista: Verificación de los requerimientos encontrados	
Participantes	
Fecha	
Objetivo	
Preguntas	Respuestas
¿Existe algún requerimiento adicional que no haya sido contemplado durante la especificación de requerimientos?	
¿Considera que los requerimientos abarcan los aspectos fundamentales para la propuesta de solución?	
¿Alguno de estos requerimientos no debería ser considerado para la propuesta?	
Otras notas	

9.8. Apéndice H. Entrevista #4 Entendimiento de los datos, conocer donde se encuentran alojados en Oracle

Entrevista: Entendimiento de los datos disponibles	
Participantes	
Fecha	
Objetivo	
Preguntas	Respuestas
¿Cómo se pueden extraer los datos de clientes y del registro de ventas en CX?	
¿Dónde se encuentran los datos de los clientes y registro de ventas?	
¿Existe algún ID que asocie los clientes con la venta en CX?	
¿Dónde se podría consultar el formato/tipo de dato de los atributos existentes en CX?	
Otras notas	

9.9. Apéndice I. Entrevista #5 Conocer cuáles herramientas han utilizado para este propósito anteriormente

Entrevista: Conocer herramientas utilizadas previamente	
Participantes	
Fecha	
Objetivo	
Preguntas	Respuestas
¿Cuáles mecanismos de extracción de datos han utilizado anteriormente para obtener información de los clientes y el registro de ventas?	
¿Existen formas de hacer extracciones automáticamente, de forma periódica?	
¿XUM TECH cuenta con servicios en la nube o alguna solución similar que permita el alojamiento y procesamiento(transformación) de datos?	
¿Tiene alguna limitante? Espacio disponible, cantidad de solicitudes, compatibilidad.	
¿La solución de Oracle Engagement Cloud tiene alguna restricción para integrarse con otros sistemas?	
¿Existe alguna consideración necesaria para hacer la extracción de datos?	
¿Cuáles mecanismos han utilizado anteriormente para agregar nuevos datos a CX?	
¿Se podría crear nuevos campos custom al objeto Contacto? ¿Sería necesario crear un Objet Custom?	
¿Dónde se podrían almacenar los datos?	

¿Conocen alguna herramienta para hacer esta importación masiva de datos?	
Otras notas	

9.10. Apéndice J. Entrevista #6 Conocer aspectos a tomar en cuenta durante una cotización de proyectos en XUM TECH.

Entrevista: Aspectos a tomar en cuenta en la cotización de proyectos	
Participantes	
Fecha	
Objetivo	
Preguntas	Respuestas
¿Existe alguna cotización previa del proyecto de módulo de recomendación de productos?	
¿Qué aspectos/rubros son tomados en cuenta a la hora de cotizar un proyecto a clientes? <ul style="list-style-type: none"> • Ejemplo: Tiempo, cantidad de recursos, licencias 	
¿Cuál es el costo aproximado de un consultor y administrador de proyectos en XT?	
¿Cuáles son otros costos que deben ser tomados en cuenta?	

¿Se maneja algún tipo de holguras de tiempo para evitar contratiempos en los proyectos? De ser posible brindar un % aproximado.	
¿En casos donde es necesario adquirir un servicio adicional a las soluciones Oracle, ese es adquirido por XUM TECH o por el cliente?	
¿En XUM TECH se maneja algún margen de ganancia para obtener las ganancias?	
Otras notas	

9.11. Apéndice K. Plantilla especificación de requerimientos

ID del requerimiento	
Tipo de requerimiento	
Categoría de requerimiento	
Historia de usuario	
Descripción de la funcionalidad	

9.12. Apéndice L. Observación datos disponibles en Oracle Engagement Cloud

Fecha de la observación		
<i>Subject Area</i>		
Tabla / Objeto	Nombre del atributo	Tipo de dato

9.13. Apéndice M. Revisión documental proyectos de minería de datos para la recomendación de productos

Objetivo					
Fecha de la revisión	Nombre del proyecto consultado	Resumen del proyecto	Herramientas utilizadas	Técnica de recomendación utilizada	Enlace de referencia

9.14. Apéndice N. Revisión documental herramientas de minería de datos

Nombre	
Objetivo principal	
Fuente de información	
Posición en Cuadrante de Gartner	
Principales características	
Costo aproximado	
Ventajas	
Desventajas	

9.15. Apéndice O. Tabla comparativa de herramientas

Criterio	Herramienta 1	Herramienta 2	Herramienta N

9.16. Apéndice P. Plantilla de pruebas

Código	
Modelo utilizado	
Objetivo	
Resultados obtenidos	
Datos de prueba	
Evidencias	

9.17. Apéndice Q. Tabla comparativa de modelos

Modelo/Indicador	Indicador 1	Indicador 2	Indicador N
Modelo 1			
Modelo 2			
Modelo N			

9.18. Apéndice R. Resultados de la Entrevista #1

Entrevista: Proceso de ventas en CX	
Participantes	Carlos Román
Fecha	18/04/2022
Objetivo	Conocer la manera actual de gestionar y concretar ventas dentro de Oracle Engagement Cloud
Preguntas	Respuestas
¿Qué módulos interactúan en la ejecución de ventas?	<ul style="list-style-type: none"> • <i>Contacts/Accounts</i> • <i>Leads</i> • Oportunidades • Metodología de Ventas • <i>Assets</i>
¿Cuáles pasos deben realizar los agentes de ventas para realizar una venta mediante la plataforma?	<ul style="list-style-type: none"> • El cliente llega a sucursal y dice estar interesado en un producto. • El ejecutivo crea el Lead en Oracle Engagement Cloud. Es necesaria la información del cliente para asociarlo al Lead, si no existe se crea primero. Adicionalmente se añaden datos como el canal y monto total. • Se completa la precalificación del Lead basada en el producto. Esta precalificación consiste en un cuestionario que compete el agente para evaluar el potencial de la prospectación. El puntaje obtenido se clasifica en Warm, Cold, Hot. • Basado en el puntaje obtenido en la precalificación el ejecutivo toma la decisión de retirar el Lead (darlo como perdido, en este caso debe agregar una razón) o convertir el Lead en una Oportunidad (se usa la misma información del Lead y se asocia al cliente). • Una vez que se crea la oportunidad, el ejecutivo debe completar la metodología de ventas definida por el negocio. Esta metodología consiste en una serie de etapas y actividades necesarias para concretar la venta. Cada actividad completada representa un aumento en el porcentaje de éxito de la oportunidad, es decir entre más actividades se completen significa que es más probable que se complete satisfactoriamente la venta. • Cuando se completan las actividades de a metodología se cambia el estado de la oportunidad a “Ganado” y la probabilidad de gane a 100%. • Finalmente se asigna el Asset o producto adquirido al cliente.

Entrevista: Proceso de ventas en CX	
¿Existe algún repositorio de documentación donde pueda consultar información relacionada?	<ul style="list-style-type: none"> • Repositorio de documentación interna en <i>Confluence</i>. • Revisar sitio de documentación de solución <i>Edwards</i> (licenciamiento de Oracle para crear recomendaciones en CX).
¿Conoce alguna manera de recomendar productos de manera nativa en CX?	No existe.
¿Los usuarios de esta plataforma han indicado alguna disconformidad o una oportunidad de mejora en el proceso?	En algunos casos los clientes han solicitado y mostrado interés en la recomendación de productos de forma personalizada dentro del sistema.
¿Es posible extraer la información de las ventas realizadas, productos y clientes del sistema?	Si es posible, para esto se puede utilizar el módulo de <i>Business Analytics</i> incluido dentro de <i>Engagement</i> .
¿Existen algunas actividades post venta que los clientes realicen con <i>Engagement</i> ?	<ul style="list-style-type: none"> • Los productos adquiridos se asignan al cliente en <i>Engagement</i>. • Se le ofrecen más productos al cliente. • Se le envían campañas de mercadeo.
Otras notas	
<p>Proceso de Ventas en CX</p> <p>¿Para qué sirve el monto almacenado dentro del Lead?</p> <ul style="list-style-type: none"> • En algunos casos los clientes lo utilizan para hacer cálculos internos dentro del sistema como el PID. Sin embargo, no es un dato requerido. <p>¿Clasificación según puntaje del <i>Lead</i>?</p> <ul style="list-style-type: none"> • <i>Cold</i>: El cliente mostró poco interés y no cumple con los requisitos mínimos para continuar con el proceso de venta. • <i>Warm</i>: El cliente mostró interés, pero tiene algunos requisitos pendientes. En estos casos el ejecutivo debe dar seguimiento durante un período de tiempo definido por el negocio para verificar si termina de completar los requisitos pendientes. • <i>Hot</i>: El cliente mostró interés y cumple con los requisitos, en estos casos el <i>Lead</i> se convierte en una Oportunidad. <p>¿El proceso de ventas es igual para todos los clientes?</p> <ul style="list-style-type: none"> • No siempre, este proceso puede variar según cada cliente. Las actividades de la metodología de ventas cambian completamente, porque es un módulo personalizado para cada cliente. <p>¿En qué momento una oportunidad pasa a ser una venta formal?</p> <ul style="list-style-type: none"> • Cuando la oportunidad cambia al estado “Ganada”. Luego del cambio de estado, la información de la oportunidad no puede ser modificada. 	

Entrevista: Proceso de ventas en CX	
¿La información de clientes siempre se almacena en los <i>Accounts</i> ?	<ul style="list-style-type: none"> Puede depender según el cliente, en algunos casos se decide utilizar <i>Accounts</i> para clientes jurídicos y <i>Contacts</i> para clientes Físicos.
¿Qué información de los <i>Assets</i> es almacenada?	<ul style="list-style-type: none"> Se almacena el grupo, nombre del producto, número y campos personalizados definidos por el cliente.

9.19. Apéndice S. Resultados de la Entrevista #3

Entrevista: Verificación de los requerimientos encontrados	
Participantes	Álex Ureña Cordero
Fecha	21/04/2022
Objetivo	Validar los requerimientos especificados por el investigador, de manera que tomen en cuenta las necesidades de la empresa.
Preguntas	Respuestas
¿Existe algún requerimiento adicional que no haya sido contemplado durante la especificación de requerimientos?	<ul style="list-style-type: none"> Cuando se gana una oportunidad también se debe almacenar esta información para que sea utilizada por el modelo.
¿Considera que los requerimientos abarcan los aspectos fundamentales para la propuesta de solución?	<ul style="list-style-type: none"> Sí, los requerimientos contemplan los aspectos más importantes para la solución.
¿Alguno de estos requerimientos no debería ser considerado para la propuesta?	<ul style="list-style-type: none"> La periodicidad definida en el R010 podría ser flexible para el cliente. No necesariamente tres meses. El requerimiento R004 debe ser revisada a futuro, de acuerdo con Álex este hace referencia a que el modelo pueda ser ajustado y mejorado continuamente. Los requerimientos relacionados a las campañas de mercadeo pueden ser implementados con el uso de la herramienta Eloqua, ya que esta herramienta es la más común entre los clientes de XUM TECH y Oracle por la facilidad de integración con Oracle Engagement Cloud.
Otras notas	
¿Respecto a la versión inicial del proceso TO BE es necesario hacer algún cambio?	

- Cambiar el orden de la creación de campañas de mercadeo basado en la recomendación de productos. La campaña debería lanzarse antes de la creación del Lead, haciendo uso de la información generada por la solución del modelo de recomendación.

9.20. Apéndice T. Resultados de la Entrevista #4

Entrevista: Entendimiento de los datos disponibles	
Participantes	Carlos Román
Fecha	18/04/2022
Objetivo	Identificar el conjunto de datos disponible en Oracle Engagement Cloud y conocer las herramientas y mecanismos de extracción
Preguntas	Respuestas
¿Cómo se pueden extraer los datos de clientes y del registro de ventas en CX?	Es posible extraer estos datos de dos maneras: <ul style="list-style-type: none"> • <i>Reports and Analytics</i> • <i>Export Management</i>
¿Dónde se encuentran los datos de los clientes y registro de ventas?	Subject Areas de Oportunidades y <i>Contacts</i> en modulo de <i>Reports and Analytics</i> .
¿Existe algún ID que asocie los clientes con la venta en CX?	Sería el partyNumber, partyId dentro de la Oportunidad ganada y el <i>Contact/Account</i> .
¿Dónde se podría consultar el formato/tipo de dato de los atributos existentes en CX?	<i>Reports and Analytics</i>
Otras notas	
Ventajas de <i>Reports and Analytics</i>	
<p>¿Cuál es la diferencia de hacerlo por medio del <i>Report and Analytics</i> y el <i>Export Management</i>?</p> <ul style="list-style-type: none"> • Por medio de <i>Reports and Analytics</i> es posible crear un reporte que pueda ser reutilizado posteriormente, también se pueden agregar diferentes <i>Subject Areas</i>(entidades) y sus campos en una misma tabla, así como filtrarlo y convertirlo en CSV o Excel. 	

Entrevista: Entendimiento de los datos disponibles

- *Export Management* únicamente permite exportar una entidad y debe ser configurado cada vez que se ejecute el proceso. En caso de ser necesario combinar o hacer un *join* entre tablas tendría que hacerse en una herramienta externa.

¿Cuáles *Subject Areas* contienen información útil para el modelo?

- *Leads, Assets, Opportunities, Contacts, Accounts.*

¿Cómo se podría crear un analítico en *Reports and Analytics*?

- Entrar en Engagement Cloud.
- Entrar en módulo Reports and Analytics.
- Seleccionar Browse Catalog.
- Seleccionar Create Analysis.
- Seleccionar Subject Area.
- Agregar los campos necesarios en la sección de Criteria.
- Revisar la vista de Results.

9.21. Apéndice U. Resultados de la Entrevista #5

Entrevista: Conocer herramientas utilizadas previamente	
Participantes	Edgar Mendoza
Fecha	20/04/2022
Objetivo	Identificar herramientas utilizadas anteriormente para la extracción de datos y desarrollo de soluciones en la nube.
Preguntas	Respuestas
¿Cuáles mecanismos de extracción de datos han utilizado anteriormente para obtener información de los clientes y el registro de ventas?	<ul style="list-style-type: none"> Usualmente la extracción de datos se hace por medio del módulo <i>Reports and Analytics</i> de <i>Oracle Engagement Cloud</i>.
¿Existen formas de hacer extracciones de datos automáticamente, de forma periódica?	
¿XUM TECH cuenta con servicios en la nube o alguna solución similar que permita el alojamiento y procesamiento(transformación) de datos?	<ul style="list-style-type: none"> Dentro de los servicios en la nube utilizados por XUM TECH se encuentran AWS y <i>Cloud Flare</i>. Principalmente se han utilizado para el desarrollo de servicios y aplicaciones personalizadas para los clientes, así como el monitoreo de estos. Para el almacenamiento de datos únicamente Dynamo DB, sin embargo, se ha utilizado para una pequeña cantidad de datos. No han utilizado soluciones para la transformación y procesamiento masivo de datos, únicamente <i>Oracle Integration Cloud (OIC)</i> para hacer integraciones entre sistemas.
¿Tiene alguna limitante? Espacio disponible, cantidad de solicitudes, compatibilidad.	En los servicios de AWS no existe restricción alguna, la facturación se hace basada en el uso de cada servicio.
¿La solución de Oracle Engagement Cloud tiene alguna restricción para integrarse con otros sistemas?	De momento no se han identificado restricciones para integrarse con otros sistemas, siempre y cuando sea capaz de enviar mensajes a <i>Oracle Engagement Cloud</i> .

Entrevista: Conocer herramientas utilizadas previamente	
¿Existe alguna consideración necesaria para hacer la extracción de datos?	Principalmente se hace por BI
¿Cuáles mecanismos han utilizado anteriormente para agregar nuevos datos a CX?	Algunos clientes han utilizado <i>Oracle Data Integrator (ODI)</i> , para la gestión de datos. Esta solución no es tan común por su alto costo de licenciamiento.
¿Dónde se suelen almacenar los datos en procesos que requieren una solución externa a Oracle Engagement Cloud? ¿Se utiliza una infraestructura de XUM TECH o el cliente?	Depende del cliente, en algunos casos piden el servicio completo por parte de XUM TECH, en estos casos el cliente no es el dueño de la infraestructura, sino que únicamente consume los servicios y XUM TECH administra todo. Por otro lado, algunos clientes prefieren ser dueños de la infraestructura y XUM TECH únicamente realiza el desarrollo necesario.
¿Conocen alguna herramienta para hacer esta importación masiva de datos?	Detalle en sección de “Otras notas”.
Otras notas	
<p>¿Cuáles formas existen de carga masiva de datos conocen para integrar con soluciones de Oracle Engagement Cloud?</p> <ul style="list-style-type: none"> • En los casos donde la carga de datos es muy grande, se utiliza una herramienta de Oracle llamada <i>Oracle Data Integrator (ODI)</i>. Esta herramienta permite hacer procesos de <i>Extract Transform Load</i>. • En casos donde la carga de datos es mediana se puede utilizar <i>Oracle Integration Cloud (OIC)</i> que funciona como orquestador, es decir se encarga de insertar y leer en las Bases de Datos (normalmente son tipo <i>serverless</i>, externas). <ul style="list-style-type: none"> ○ Como consideración, OIC no puede manejar archivos de más de 10MB, es necesario segmentar los datos, normalmente esto se realiza mediante un contador de registros. • Pequeños, cliente provee servicio por API, se extrae del <i>core</i>, se hace la transformación y se mandan. <p>¿Qué tipo de reportería han trabajado con soluciones de Oracle?</p> <ul style="list-style-type: none"> • Reportería personalizada por áreas funcionales en <i>Oracle Engagement Cloud</i>. • Reportería enfocada en la gestión de campañas de mercadeo con la herramienta de Oracle Eloqua. <ul style="list-style-type: none"> ○ Por medio de una integración con Eloqua se permite el envío de mensajes, gracias a la reportería desarrollada es posible conocer cuáles mensajes contestaron respondieron, abrieron o ignoraron. 	

Entrevista: Conocer herramientas utilizadas previamente	
¿Fuera de Oracle han utilizado algún servicio de analítica?	
<ul style="list-style-type: none"> Se han creado <i>dashboards</i> dentro de herramientas como <i>Cloud Fare</i> y AWS para monitorear en tiempo real el estado de los servicios de TI brindados por XUM TECH. 	
¿Existen repositorios donde revisar información relacionada a los servicios en la nube utilizados por XUM TECH?	
<ul style="list-style-type: none"> Documentación oficial de AWS. Documentación de proyectos anteriores disponible en Confluence, contiene información acerca de diseño e implementación de herramientas propias de XUM TECH. (SMS, <i>Communication Broker</i>, <i>Eloqua GreenPlay</i>). Cuentas de AWS: Permite consultar el repositorio de códigos desarrollados. 	

9.22. Apéndice V. Resultados de la Entrevista #6

Entrevista: Aspectos a tomar en cuenta en la cotización de proyectos	
Participantes	Sussana Ramírez
Fecha	19/04/2022
Objetivo	Identificar los principales aspectos tomados en cuenta por el equipo comercial de XUM TECH en la cotización de proyectos.
Preguntas	Respuestas
¿Qué aspectos/rubros son tomados en cuenta a la hora de cotizar un proyecto a clientes? <ul style="list-style-type: none"> Ejemplo: Tiempo, cantidad de recursos, licencias 	<ul style="list-style-type: none"> Cantidad de recursos. Duración (Semanas, Días totales) Cantidad de ítems configurados. Margen de ganancia de XUM TECH. Impuestos de cada país. Tamaño/complejidad del proyecto.
¿Cuál es el costo aproximado de un consultor y administrador de proyectos en XT?	<ul style="list-style-type: none"> Información disponible en el Excel de cotizaciones. (ORACLE CX 2021 V3).
¿Cuáles son otros costos que deben ser tomados en cuenta?	<ul style="list-style-type: none"> Licenciamiento normalmente es gestionado por parte del proveedor.

Entrevista: Aspectos a tomar en cuenta en la cotización de proyectos	
	En este caso XUM TECH no vende licencias de Oracle.
¿Se maneja algún tipo de holguras de tiempo para evitar contratiempos en los proyectos? De ser posible brindar un % aproximado.	<ul style="list-style-type: none"> Únicamente se manejan tiempos predeterminados para cada tarea, según la cantidad de ítems se calcula la duración.
¿En casos donde es necesario adquirir un servicio adicional a las soluciones Oracle, ese es adquirido por XUM TECH o por el cliente?	<ul style="list-style-type: none"> Puede ser que XUM TECH contacte con los <i>partners</i>, sin embargo, el proveedor de la herramienta debe interactuar con el cliente. El cliente lo adquiere, hace el contrato con el proveedor y XUM TECH lo implementa.
¿En XUM TECH se maneja algún margen de ganancia para obtener las ganancias?	Se encuentra definido un margen de ganancia, por temas de confidencialidad no es posible brindarlo. Para efectos del proyecto se puede estimar en un 30%.
Otras notas	
<p>¿Cómo se gestionan las ventas en XUM TECH?</p> <ul style="list-style-type: none"> Se utiliza la herramienta de Oracle Engagement Cloud, donde primero se crean los Lead y posteriormente pasa a ser una Oportunidad. Los encargados de dar seguimiento a estas ventas son los miembros del área comercial. Las ventas pueden darse por demanda interna (XUM TECH contacta y negocia directamente con los clientes) o por demanda externa (un <i>partner</i> contacta a XUM TECH para referir un cliente interesado en implementar sus soluciones). <p>¿Cuál es el proceso de ventas por demanda externa de XUM TECH?</p> <ul style="list-style-type: none"> Un cliente contacta a Oracle o una empresa de consultoría para que lo asesore para resolver una necesidad empresarial. Oracle recomienda una de las herramientas y referencia a XUM como empresa implementadora. XUM TECH realiza una etapa de discovery, donde se conocen las empresas, el ecosistema de Oracle, las funcionalidades de sus sistemas y también se consulta sobre los sistemas actuales del cliente. 	

Entrevista: Aspectos a tomar en cuenta en la cotización de proyectos

- XUM TECH y el cliente realizan un proceso inicial de definición de AS IS y el TO BE.
- Toma y especificación de requerimientos.
- XUM TECH crea una estimación, donde se utiliza la plantilla estándar de Oracle CX.
- Negociación entre ambas partes de la propuesta y estimación brindada por XUM TECH.
- El cliente aprueba la propuesta y se formaliza el acuerdo con la firma del contrato entre ambas partes.
- Una vez firmado el contrato XUM TECH realiza la asignación del equipo responsable de la implementación.

¿Es posible consultar cotizaciones de proyectos pasadas para tomarlo como referencia?

Sí, se pueden consultar cotizaciones anteriores dentro del SharePoint empresarial. Para las estimaciones se utiliza un Excel definido por negocio para determinar los costos.

9.23. Apéndice W. Resultados de la revisión de datos disponibles

Fecha de la observación	26/04/2022	
Subject Area	Sales - CRM Opportunities and Products Real Time	
Tabla/Objeto	Nombre del atributo	Tipo de dato
<i>Opportunity</i>	Channel Type	String - Categórico
	Channel Type Code	Integer - Categórico
	Customer Type	String - Categórico
	Nombre de la oportunidad	String
	Número de la oportunidad	Integer
	Opportunity ID	Integer
	Sales Account ID	Integer
	Sales Channel	String - Categórico
	Status Code	Integer - Categórico
	Status Name	String - Categórico
<i>Opportunity Extension</i>	Nombre de contacto de referencia	String
	Producto	String - Categórico
<i>Contact</i>	Nombre del cliente	String
	Gender	String - Categórico
	First Name	String
	Last Name	String
	Contact Sales Account Id	Integer
	Cantón	String - Categórico
	City	String - Categórico
<i>Contact Extension</i>	Numero de identificacion	Integer
	Estado Civil	String - Categórico
	Nombre del cliente	String
	Tipo Cliente	String - Categórico

9.24. Apéndice X. Resultados de la revisión documental de proyectos de minería de datos para la recomendación de productos

Objetivo de la revisión					
Identificar las herramientas y técnicas de minería de datos más utilizadas en la recomendación de productos personalizada para contar con una base de conocimiento sobre cómo resolver este problema utilizando técnicas de <i>Machine Learning</i> .					
Fecha de la revisión	Nombre del proyecto consultado	Resumen del proyecto	Herramientas utilizadas	Técnica de recomendación utilizada	Enlace de referencia
28/04/2022	<i>Movie Recommendation System</i>	Construcción de un sistema de recomendación de películas de referencia utilizando el conjunto de datos de películas TMDb 5000.	Lenguaje de programación: Python Bibliotecas: <ul style="list-style-type: none"> • NumPy • Matplotlib • Pandas Bases de datos <ul style="list-style-type: none"> • Películas • Créditos 	<ul style="list-style-type: none"> • <i>Demographic Filtering</i> • <i>Content Based Filtering</i> • <i>Collaborative Filtering</i> 	Getting Started with a Movie Recommendation System Kaggle
28/04/2022	<i>Americans preferred lunch</i>	Construcción de un sistema de recomendación de comidas.	Lenguaje de programación: Python Bibliotecas: <ul style="list-style-type: none"> • Pandas • GarphLab, Bases de datos: <ul style="list-style-type: none"> • Usuarios 	<ul style="list-style-type: none"> • <i>Collaborative Filtering</i> • <i>Content Based Filtering</i> 	An Easy Introduction to Machine Learning Recommender Systems - KDnuggets

Propuesta de solución de minería de datos basada en la metodología CRISP-DM utilizando los datos almacenados en Oracle Engagement Cloud, para la recomendación personalizada de productos a clientes, durante el primer semestre de 2022

Objetivo de la revisión					
Identificar las herramientas y técnicas de minería de datos más utilizadas en la recomendación de productos personalizada para contar con una base de conocimiento sobre cómo resolver este problema utilizando técnicas de <i>Machine Learning</i> .					
Fecha de la revisión	Nombre del proyecto consultado	Resumen del proyecto	Herramientas utilizadas	Técnica de recomendación utilizada	Enlace de referencia
			<ul style="list-style-type: none"> Comidas Calificaciones 		
29/04/2022	Sistema de Recomendación de Repositorios Github	Construcción de un motor de recomendación de repositorios Github.	Lenguaje de programación: Python Librerías: <ul style="list-style-type: none"> Pandas NumPy Sklearn Matplotlib Indicadores: <ul style="list-style-type: none"> <i>Sparcity</i> Distancias por coseno <i>Mean Squared Error</i> Bases de Datos <ul style="list-style-type: none"> Usuarios Repositorios Calificaciones 	<ul style="list-style-type: none"> <i>Collaborative Filtering</i> 	Sistemas de Recomendación Aprende Machine Learning Algoritmo k-Nearest Neighbor Aprende Machine Learning

Propuesta de solución de minería de datos basada en la metodología CRISP-DM utilizando los datos almacenados en Oracle Engagement Cloud, para la recomendación personalizada de productos a clientes, durante el primer semestre de 2022

Objetivo de la revisión					
Identificar las herramientas y técnicas de minería de datos más utilizadas en la recomendación de productos personalizada para contar con una base de conocimiento sobre cómo resolver este problema utilizando técnicas de <i>Machine Learning</i> .					
Fecha de la revisión	Nombre del proyecto consultado	Resumen del proyecto	Herramientas utilizadas	Técnica de recomendación utilizada	Enlace de referencia
30/04/2022	<i>Collaborative Filtering for Implicit Feedback Datasets</i>	Construcción de un motor de recomendación de programas de televisión.	<p>Lenguaje de programación: Python</p> <p>Librerías:</p> <ul style="list-style-type: none"> • NumPy • Sklearn • Pandas • Scipy <p>Indicadores de evaluación</p> <ul style="list-style-type: none"> • <i>Recall</i> • <i>Precision</i> • <i>Masked precisión ratio</i> <p>Bases de datos: Interacciones: Contiene el ID usuario, <i>Program ID</i> y <i>Timestamp</i></p>	<ul style="list-style-type: none"> • <i>Collaborative Filtering</i> 	<p>cf.pdf (yifanhu.net) implicit-recommender-system/machine_learning_approach.ipynb at master · billydh/implicit-recommender-system · GitHub</p> <p>implicit-recommender-system/machine_learning_approach.ipynb at master · billydh/implicit-recommender-system · GitHub</p>
30/04/2022	<i>Machine Learning Algorithms for Recommender</i>	Experimento comparativo entre algoritmos de recomendación,	Bases de datos: Usuarios Película Calificaciones	<ul style="list-style-type: none"> • <i>Content Based Filtering</i> 	(PDF) Machine Learning Algorithms for Recommender System - a

Propuesta de solución de minería de datos basada en la metodología CRISP-DM utilizando los datos almacenados en Oracle Engagement Cloud, para la recomendación personalizada de productos a clientes, durante el primer semestre de 2022

Objetivo de la revisión					
Identificar las herramientas y técnicas de minería de datos más utilizadas en la recomendación de productos personalizada para contar con una base de conocimiento sobre cómo resolver este problema utilizando técnicas de <i>Machine Learning</i> .					
Fecha de la revisión	Nombre del proyecto consultado	Resumen del proyecto	Herramientas utilizadas	Técnica de recomendación utilizada	Enlace de referencia
	<i>System - a comparative analysis</i>	utilizando el conjunto de datos de películas de MovieLens.		<ul style="list-style-type: none"> • <i>Collaborative Filtering</i> • <i>Hybrid Filtering</i> • <i>K-Means Clustering</i> • <i>Naive Bayes</i> 	comparative analysis (researchgate.net)

9.25. Apéndice Y. Revisión documental de herramientas de almacenamiento

AWS

Nombre	AWS
Objetivo principal	Almacenar la información generada por la solución propuesta, esta información incluye el registro de datos de entrenamiento para generar el modelo, las recomendaciones generadas y las recomendaciones rechazadas por el ejecutivo de ventas.
Fuente de información	<ul style="list-style-type: none"> • Sitio Oficial de AWS • Reporte de Gartner Magic Quadrant for Cloud Database Management Systems 2020 • Peer Insights <p>Amazon Web Services (AWS) Amazon Redshift Reviews, Ratings, and Features - Gartner 2022</p>
Posición en Cuadrante de Gartner	Líder
Calificación general en Peer Insights	4.4 estrellas
Principales características	<ul style="list-style-type: none"> • Ofrece una gama de servicios de gestión de bases de datos dirigidas a casos de uso operativo, como es el caso de Amazon Relational Database Service (RDS), Amazon Aurora y Amazon DynamoDB. • Ofrece bases de datos dirigidas a casos de uso de análisis, como es el caso de Amazon Redshift, Amazon Athena y Amazon EMR. • AWS puede ofrecer capacidades a nivel de infraestructura para la nube DBMS con, por ejemplo, su sistema AWS Nitro, AQUA (Advanced Query Accelerator) para Amazon Redshift y nodos Redshift RA3 asistidos por hardware. • Sus operaciones están geográficamente diversificadas y sus clientes cubren una amplia gama de industrias y tamaños de implementación.
Costo aproximado	Costo Mensual: 182.80 USD ***Obtenido de <i>Pricing Calculator</i> de AWS.
Ventajas	<ul style="list-style-type: none"> • Presencia líder en el mercado: AWS es el proveedor de servicios de bases de datos en la nube más grande del mundo por ingresos. Ha ofrecido servicios de nube de producción durante más de 10 años, con decenas de miles de clientes, lo que también lo convierte en el proveedor de nube más maduro en términos de ofertas y capacidades. • Amplitud de servicios: AWS ofrece más de una docena de servicios de bases de datos diferentes y una amplia oferta de funcionalidades. Además, AWS tiene una base muy amplia de centros de datos, lo que proporciona una latencia más baja según la proximidad en todo el mundo.

	<ul style="list-style-type: none"> • Rendimiento: El tiempo que los servicios de base de datos de AWS han estado en producción ha llevado al crecimiento de las cargas de trabajo y ha mejorado la capacidad de AWS para manejar estas cargas de trabajo más maduras. Además, el tiempo de actividad y la confiabilidad de AWS son los más confiables de los proveedores evaluados en la investigación de Gartner.
Desventajas	<ul style="list-style-type: none"> • <i>AWS Centric-Focus</i>: AWS es el proveedor de nube más grande. Sin embargo, está detrás de algunos CSP y prácticamente de todos los ISV al ofrecer sus servicios en múltiples plataformas en la nube. Esta centralidad parece estar disminuyendo un poco, dado el soporte de AWS para entornos de nube híbrida, el intercambio de sus datos entre diferentes nubes en ciertos escenarios y una estrategia de datos que reconoce y se está adaptando a la existencia de múltiples plataformas en la nube. • Integración en toda la cartera: AWS utiliza un enfoque de mejor ajuste para los servicios de base de datos, lo que le permite ofrecer soluciones enfocadas para diferentes tipos de casos de uso. Sin embargo, las soluciones de integración de datos, catálogo y metadatos de AWS están surgiendo y actualmente no son tan sólidas como algunos competidores, lo que puede generar más trabajo para los arquitectos. • Costo de administración: AWS ofrece un conjunto profundo y completo de herramientas para que los usuarios puedan ensamblar soluciones óptimas. Este enfoque de ensamblaje requerido a veces puede conducir a que se requieran más gastos generales y de administración.

9.26. Apéndice Z. Revisión documental de herramientas de almacenamiento

Oracle

Nombre	Oracle Autonomous Database
Objetivo principal	Almacenar la información generada por la solución propuesta, esta información incluye el registro de datos de entrenamiento para generar el modelo, las recomendaciones generadas y las recomendaciones rechazadas por el ejecutivo de ventas.
Fuente de información	<ul style="list-style-type: none"> • Cuadrante de Gartner • Sitio oficial Oracle • Peer Insights Autonomous Transaction Processing Oracle Oracle Autonomous Database Reviews, Ratings, and Features - Gartner 2022
Posición en Cuadrante de Gartner	Líder
Calificación general en Peer Insights	4.6 estrellas
Principales características	<ul style="list-style-type: none"> • Oracle Autonomous Database (incluidos los servicios Autonomous Transaction Processing y Autonomous Data Warehouse) están disponibles en Oracle Cloud Infrastructure (OCI) y en la nube privada Oracle Exadata Cloud@Customer (ExaCC). • Oracle también ofrece la base de datos JSON autónoma, el servicio de base de datos MySQL de Oracle, la base de datos NoSQL de Oracle y el servicio Big Data de Oracle. • Sus operaciones están geográficamente diversificadas y sus clientes cubren una amplia gama de industrias y tamaños de implementación.
Costo aproximado	Costo Mensual: 116 USD ***Obtenido de calculadora <i>MyEstimate</i> Oracle
Ventajas	<ul style="list-style-type: none"> • Tecnología DBMS aumentada: la base de datos autónoma de Oracle utiliza ajustes asistidos por IA y ML y aprovecha la infraestructura de la nube de Oracle, incluidos parches y actualizaciones sin tiempo de inactividad. Esta tecnología no solo reduce la necesidad de operación manual, sino que también reduce el costo y aumenta el rendimiento del servicio DBMS. • Nube híbrida: Oracle tiene un entorno híbrido completo para compartir datos y permitir migraciones a la nube. Admite implementaciones locales y ExaCC con recuperación ante desastres en la nube a través de Exadata Cloud Service (ExaCS). ExaCC es compatible con el almacenamiento de datos autónomo y el procesamiento de transacciones autónomo. • Riqueza de la cartera: Oracle Cloud es una buena opción para las empresas que utilizan Oracle como proveedor estándar, donde pueden aplicar las habilidades desarrolladas durante muchos años para usar casos en la nube. Con el nivel gratuito de Oracle Autonomous Database siendo adoptado por clientes nuevos

	<p>en Oracle que valoran un DBMS empresarial autónomo, Oracle está ganando reconocimiento como proveedor de nube empresarial.</p>
Desventajas	<ul style="list-style-type: none"> • Estrategia DBMS centrada en OCI: solo los propios servicios DBMS de Oracle han estado disponibles como servicios administrados en OCI; Oracle ofrecerá un servicio administrado de PostgreSQL en 2022. Si bien Oracle Database generalmente es portátil a otras nubes, los clientes de Oracle Database pagarán el doble de CPU virtuales cuando se ejecuten en otras nubes. Oracle Real Application Clusters (RAC) no es compatible con otras nubes y Oracle Database aún no está certificado para ejecutarse en todos los principales CSP. • Percepción de precio superior: los productos locales de Oracle a menudo se perciben como costosos y difíciles de administrar, y los clientes continúan expresando inquietudes sobre las negociaciones de contratos. Oracle ha progresado en estas áreas a medida que avanza hacia un nuevo modelo de negocios en la nube con precios de pago por uso, la base de datos autónoma, un nivel siempre gratuito, traiga su propia licencia (BYOL) y elasticidad dinámica. Además, los servicios Oracle Cloud Lift gratuitos ayudarán a los clientes a migrar a OCI, y un programa de recompensas de soporte ofrece recompensas por gastar en servicios de OCI. Los clientes deben evaluar cuidadosamente las propuestas para maximizar el valor. • Enfoque de transición a la nube: Oracle tardó mucho en ofrecer un verdadero servicio administrado dbPaaS. Esto ha provocado que las empresas utilicen otras ofertas de dbPaaS de CSP e ISV. Según las cifras de Gartner, la participación de Oracle en el mercado de dbPaaS para 2020 (\$ 875,9 millones de \$ 26,2 mil millones) se mantuvo estable en 3,3%. Los clientes que no son de Oracle deben evaluar cuidadosamente las opciones de migración a la nube a corto plazo.

9.27. Apéndice AA. Revisión documental de herramientas de almacenamiento Microsoft

Nombre	Microsoft Synapse Analytics
Objetivo principal	Almacenar la información generada por la solución propuesta, esta información incluye el registro de datos de entrenamiento para generar el modelo, las recomendaciones generadas y las recomendaciones rechazadas por el ejecutivo de ventas.
Fuente de información	<ul style="list-style-type: none"> • Cuadrante de Gartner • Sitio oficial Microsoft Azure Synapse Analytics Microsoft Azure Pricing Calculator Microsoft Azure
Posición en Cuadrante de Gartner	Líder
Calificación general en Peer Insights	4.4 estrellas
Principales características	<ul style="list-style-type: none"> • Proporciona una amplia gama de ofertas de DBMS en la nube, que incluyen Azure Synapse Analytics, Azure SQL Database, Azure SQL Managed Instance, Azure Cache for Redis, Azure SQL Edge, Azure Cosmos DB, Azure HDInsight y Azure Database para PostgreSQL, MySQL y MariaDB. • Ofrece un ecosistema integral de administración de datos en la nube que abarca todos los casos de uso evaluados para este Cuadrante Mágico. • Sus operaciones están geográficamente diversificadas y sus clientes cubren una amplia gama de industrias y tamaños de implementación.
Costo aproximado	Costo mensual: 1.913,84 USD *** Obtenido de <i>Pricing calculator</i> de Microsoft Azure
Ventajas	<ul style="list-style-type: none"> • Visión integral del ecosistema de datos: Microsoft ha articulado una sólida visión del ecosistema de datos con Azure Synapse Analytics. Esta visión es más refinada y completa que la de la mayoría de sus principales competidores de CSP, e incluye soporte de seguridad y metadatos de extremo a extremo. La fácil integración con otras ofertas de Azure (a través de Synapse Link y Purview) es un punto de venta importante, y el ecosistema también está abierto a ofertas de ISV de terceros. • Ruta de migración clara: casi todas las empresas tienen una relación comercial con Microsoft, a través del uso de sus productos de administración de datos (como Microsoft SQL Server), herramientas de productividad de oficina, software de administración de acceso e identidad y otras ofertas. Azure representa una extensión natural de este ecosistema familiar de Microsoft y una ruta de migración de bajo riesgo a la nube, con un conjunto integral de ofertas en la nube y una sólida comunidad de usuarios con habilidades relevantes.

	<ul style="list-style-type: none"> • Capacidades multimodelo integradas: a diferencia de algunos de sus competidores de CSP, Microsoft ha adoptado una estrategia multimodelo para muchas de sus ofertas de administración de datos, lo que puede simplificar la implementación. Azure Synapse refleja esta estrategia de análisis y Azure Cosmos DB la personifica para DBMS operativos no relacionales. Los dos se pueden usar juntos para transacciones aumentadas de alto rendimiento, con un mínimo movimiento de datos o redundancia.
Desventajas	<ul style="list-style-type: none"> • Madurez del ecosistema inconsistente: algunos componentes clave del ecosistema de la nube de Microsoft aún están en construcción, a saber, Purview, Synapse Link (para Dataverse) y modelos de datos de la industria. Los usuarios potenciales deben verificar cuidadosamente que las capacidades actuales de Microsoft satisfagan sus necesidades y que los plazos para la entrega de la funcionalidad adicional requerida sean claros. • Retos de precios y costos: el modelo general de precios de Microsoft es complejo y desafiante, lo que genera un gran volumen de consultas de los clientes al equipo de administración de proveedores y precios de software de Gartner. • Desafíos de la arquitectura de datos: aunque la cartera de DBMS locales y en la nube de Microsoft es integral, es inconsistente en cuanto a que las opciones de implementación para la nube pública, la nube privada y las instalaciones locales tradicionales varían según el producto. Por el lado positivo, esto permite arquitecturas híbridas para los usuarios que las necesitan. La desventaja es que los usuarios deben decidir cuidadosamente dónde administrarán conjuntos de datos específicos y ejecutarán diversas cargas de trabajo de procesamiento de datos.

9.28. Apéndice AB. Revisión documental de herramientas de minería de datos IBM

Nombre	IBM Watson Studio
Objetivo principal	Brindar la capacidad de crear modelos de minería de datos Data Science and Machine Learning Platform para recomendar productos de manera personalizada.
Fuente de información	<p><i>Gartner Magic Quadrant for Data Science and Machine Learning Platforms 2021</i></p> <ul style="list-style-type: none"> • IBM Watson Studio Reviews, Ratings, and Features - Gartner 2022 • Introduction to IBM Watson Studio - IBM Developer • Adding data from a connection to a project - IBM Cloud Pak for Data as a Service • Adding connections to projects - IBM Cloud Pak for Data as a Service • Connection types - IBM Cloud Pak for Data as a Service • 2021 Gartner MQ DSML (dataiku.com) • Watson Studio - Pricing IBM • Watson Studio - IBM Cloud
Posición en Cuadrante de Gartner	Líder
Principales características	<ul style="list-style-type: none"> • Es una plataforma modular, abierta y extensible para datos e IA que combina un amplio conjunto de capacidades descriptivas, de diagnóstico, predictivas y prescriptivas. • IBM está diversificado geográficamente y su base de clientes abarca muchas industrias y varias funciones comerciales. • La renovación de su oferta ha llevado varios años y la competencia seguirá siendo feroz para IBM. Aún así, IBM ahora ofrece una solución moderna y completa que se basa en sus raíces en SPSS, ILOG CPLEX Optimization Studio y productos anteriores, y que se beneficia de una serie de innovaciones de IBM Research. Estos reflejan una visión completa.
Costo aproximado	Costo mensual: 98 USD. *** Obtenido de <i>IBM Pricing Plans</i>
Ventajas	<ul style="list-style-type: none"> • Soporte multipersona: IBM Watson Studio ofrece una interfaz de flujo de trabajo visual o "lienzo gráfico", así como una selección de cuadernos, lo que permite que los ingenieros de datos, los científicos de datos expertos y los científicos de datos ciudadanos trabajen juntos en el mismo proyecto. Las actividades de canalización de ML, desde la adquisición de datos hasta las operaciones, son compatibles con AutoAI y colaboración, incluido un catálogo para compartir y reutilizar (meta)datos y modelos. • Visión de IA compuesta: la estructura modular de la plataforma IBM Watson Studio contiene, o puede ampliarse con, múltiples componentes para el aumento de decisiones o la automatización. Estos componentes incluyen varios ML y otros marcos de IA, funciones de optimización, análisis espacio-temporal y gráfico, funciones de lenguaje natural y análisis de video/imagen/audío (en modo por lotes o de transmisión). Además, al incluir IBM Decision Optimization, la plataforma

	<p>admite el modelado de decisiones y la gestión de decisiones o el procesamiento de reglas.</p> <ul style="list-style-type: none"> Atención integral a la IA y la gobernanza responsable: IBM ofrece un amplio soporte para la explicación, el sesgo, la equidad, la precisión y el control de la deriva, los datos sintéticos y la privacidad diferencial. Su plataforma también proporciona un sólido soporte de gobernanza (y gestión de riesgos opcional), con linaje, políticas y reglas en su catálogo, así como seguridad frente a adversarios.
Desventajas	<ul style="list-style-type: none"> Alcance de AutoAI: IBM Watson Studio ofrece automatización y aumento de múltiples actividades en la canalización de ML, incluida la selección de datos, la imputación, la visualización, la transformación de características y el modelado. Sin embargo, algunos competidores también aumentan el análisis de series temporales, por ejemplo, mediante el uso de redes neuronales recurrentes y modelos de memoria a corto plazo. Restauración de la marca: con su Watson Studio mejorado, IBM ha alcanzado y, en algunos casos, incluso superado a sus competidores. Sin embargo, los líderes de datos y análisis aún pueden encontrar a sus expertos en ML escépticos sobre la capacidad innovadora de Watson Studio y la capacidad de IBM para mantener el ritmo en un mercado dinámico y competitivo. Claridad en la agrupación de productos: aunque la cohesión del Watson Studio modular en IBM Cloud Pak for Data ha mejorado, sigue existiendo confusión entre los clientes potenciales sobre qué productos y licencias se necesitan para qué configuraciones. Esto aumenta las preocupaciones sobre los costos de licencia.

9.29. Apéndice AC. Revisión documental de herramientas de minería de datos Amazon Sage Maker

Nombre	Amazon Sage Maker
Objetivo principal	Brindar la capacidad de crear modelos de minería de datos para recomendar productos de manera personalizada.
Fuente de información	<ul style="list-style-type: none"> • Cuadrante de Gartner • Sitio oficial de AWS • 2021 Gartner MQ DSML (dataiku.com) • Amazon SageMaker Pricing - Machine Learning - Amazon Web Services • AWS Pricing Calculator
Posición en Cuadrante de Gartner	Visionario
Principales características	<ul style="list-style-type: none"> • Su visión es que los equipos de ciencia de datos utilicen toda la cartera de AWS y la pila de ML, con Amazon SageMaker como núcleo. • Muchos de los componentes y servicios de soporte de AWS se consideraron al evaluar la oferta de AWS. Estos incluyeron SageMaker Studio IDE (que incluye Autopilot, Notebooks, Model Monitor, Experiments y Debugger), Amazon EMR (incluido S3), AWS Glue, Amazon SageMaker Neo, Amazon SageMaker Ground Truth, Amazon SageMaker Clarify, Amazon SageMaker Data Wrangler, Amazon SageMaker Pipelines, AWS CloudWatch, AWS CloudTrail y otros. • Amazon SageMaker continúa demostrando una formidable tracción en el mercado, con un poderoso ecosistema y considerables recursos detrás.
Costo aproximado	Costo mensual estimado: 244.80 USD *** Obtenido de AWS Pricing Calculator
Ventajas	<ul style="list-style-type: none"> • Amplitud y profundidad de la plataforma en la nube: los usuarios pueden aprovechar directamente los servicios de IA preempaquetados de AWS (como Amazon Lex, Polly y Transcribe). SageMaker también se integra de forma nativa con muchas herramientas de análisis y datos en la nube de AWS. Además, SageMaker ofrece una amplia compatibilidad con una amplia gama de marcos y bibliotecas de software de código abierto (OSS) populares y de nicho. • Rendimiento, escalabilidad y granularidad de control: Amazon SageMaker y su cartera de soporte ofrecen el mejor rendimiento y escalabilidad de su clase. La plataforma admite una selección significativa de opciones de hardware optimizadas para varios marcos de ML y aprendizaje profundo, y presenta un modelo de precios de pago por uso sin tarifas mínimas ni compromiso inicial, lo que fomenta la experimentación. • Capacidades de etiquetado de datos y humanos en el ciclo: Amazon SageMaker Ground Truth admite el etiquetado de datos de entrenamiento, y la IA aumentada de Amazon (Amazon A2I) ayuda a crear flujos de trabajo óptimos para la revisión humana de los modelos implementados. AWS conecta a los clientes con proveedores de mercados externos y la fuerza laboral de Amazon Mechanical Turk (MTurk) para el etiquetado humano de datos.

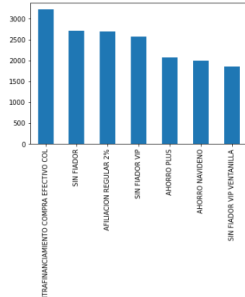
Desventajas	<ul style="list-style-type: none"> • Evolución del atractivo de la ciencia de datos para los ciudadanos: AWS ha hecho que su plataforma sea más accesible, principalmente a través de Autopilot, Data Wrangler, Pipelines y el desarrollo continuo de SageMaker Studio IDE. Aún así, la plataforma es más popular entre los codificadores: no es tan intuitiva para los usuarios no técnicos, en comparación con las herramientas líderes para los científicos de datos ciudadanos. • Se necesita un ritmo rápido de desarrollo para igualar la funcionalidad de los competidores: la ráfaga de nuevos componentes y servicios de AWS está llenando vacíos importantes en su plataforma. Sin embargo, estas nuevas capacidades no están tan probadas ni son tan sólidas como las capacidades de otros proveedores para la preparación de datos, las interfaces de usuario, la colaboración y la coherencia. • Maduración del soporte local, híbrido y multinube: la mayoría de los clientes de Amazon SageMaker operan en entornos puramente en la nube. Algunas capacidades dentro de la cartera de AWS cambian o se vuelven más complicadas en entornos híbridos, multinube o locales. Sin embargo, la compatibilidad con varias nubes está evolucionando y, en la actualidad, la mayoría de los clientes administran datos, modelos y cargas de trabajo de ML dentro de AWS.
-------------	--

9.30. Apéndice AD. Revisión documental de herramientas de minería de datos Anaconda

Nombre	Anaconda
Objetivo principal	Brindar la capacidad de crear modelos de minería de datos para recomendar productos de manera personalizada.
Fuente de información	<ul style="list-style-type: none"> • Cuadrante de Gartner • Sitio oficial de Anaconda 2021 Gartner MQ DSML (dataiku.com) Anaconda Enterprise Review: Pricing, Pros, Cons & Features CompareCamp.com
Posición en Cuadrante de Gartner	Jugador de nicho
Principales características	<ul style="list-style-type: none"> • Permite a los equipos de científicos de datos establecer, supervisar y automatizar canalizaciones y modelos de ciencia de datos basados en inteligencia artificial dentro de clústeres de servidores y redes de producción. • Permite compartir proyectos de ciencia de datos con los miembros del equipo y les permite realizar modificaciones en tiempo real. Los cuadernos en línea, los tableros, las dependencias y los modelos de aprendizaje automático también están disponibles para todos los involucrados en los proyectos del equipo. • Anaconda Enterprise es una herramienta de ciencia de datos escalable, los usuarios pueden proporcionar instantáneamente a sus proyectos y aplicaciones los recursos informáticos necesarios. Las herramientas, paquetes y versiones de modelos utilizados por el equipo se pueden gobernar, administrar y controlar fácilmente.

	<ul style="list-style-type: none"> Los cambios realizados por los miembros del equipo se pueden rastrear en los registros de actividad generados. Sobre todo, Anaconda Enterprise implementa el cifrado SSL y TLS para brindar tranquilidad y seguridad a los miembros de su equipo.
Costo aproximado	Costo de licencia \$15 mensual.
Ventajas	<ul style="list-style-type: none"> Plataforma confiable y flexible: Anaconda ofrece una plataforma popular y confiable dentro de la comunidad de programación, una con opciones tanto para principiantes como para expertos. La GUI es intuitiva, brinda acceso a todas las bibliotecas de R y Python, y ofrece a los usuarios la flexibilidad de trabajar en varios IDE de su elección, incluidos Jupyter y RStudio. Optimización de tecnologías de código abierto: para optimizar las tecnologías de código abierto y admitir la escalabilidad, Anaconda ofrece opciones de mejora mediante GPU, administradas dentro del entorno de Anaconda. Los usuarios también pueden usar clústeres de Apache Hadoop, Apache Hadoop YARN y Kubernetes, en las instalaciones o en la nube. Cultura de colaboración y características complementarias: la comunidad de Anaconda apoya las contribuciones de código abierto de Python, fomentando así una cultura de integridad de código e integración con otros proyectos de ciencia de datos de código abierto. Anaconda Cloud, para científicos de datos, proporciona formas de colaborar, compartir implementaciones e intercambiar bibliotecas de códigos. También permite a los desarrolladores explorar y acelerar el desarrollo y la implementación de modelos.
Desventajas	<ul style="list-style-type: none"> Centrarse en la audiencia técnica: Anaconda se dirige a una audiencia técnica que prefiere codificar en lenguajes R o Python para la ciencia de datos. La plataforma carece de funciones que permitan a los científicos de datos ciudadanos aprovecharla. Falta de algunas capacidades críticas de puesta en funcionamiento del modelo: la plataforma de Anaconda carece de capacidades de gestión de modelos, como gestión de dependencias, explicabilidad y detección de sesgos, así como funciones de inventario de modelos. Anaconda, sin embargo, proporciona algunas características de control y supervisión de modelos, como la programación de implementaciones, información de usuario y consumo de recursos para modelos ML (a través de la interfaz de usuario del programador). Estabilidad: los usuarios de Anaconda destacan los problemas de compatibilidad y tiempo de ejecución con la plataforma. A los usuarios no expertos a menudo les resulta difícil mantener la coherencia de sus proyectos cuando se lanzan nuevas plataformas o actualizaciones de paquetes.

9.31. Apéndice AE. Evaluación del modelo Demographic Filtering

Código	MOD-001
Modelo utilizado	<i>Demographic Filtering</i>
Objetivo	Evaluar el modelo de <i>Demographic Filtering</i> para la recomendación de productos con los datos extraídos de Oracle Engagement Cloud.
Resultados obtenidos	Recomendación basada en la popularidad de los productos, según la cantidad de veces que han sido adquiridos por los clientes. Se genera un conjunto de recomendaciones genéricas para todos los clientes.
Datos de prueba	
Registro de oportunidades ganadas exportados de Oracle Engagement Cloud, detallado en la sección 5.2.1.	
<ul style="list-style-type: none"> Cantidad de filas(original): 38170 Cantidad de columnas(original): 25 	
Evidencias	
<u>Tareas de transformaciones y limpieza</u>	
1. Revisión de registros de venta repetidos	
Resultado: No se encontraron valores repetidos dentro del atributo <i>Opportunity ID</i> .	
<pre>In [24]: df["Opportunity ID"].nunique() Out[24]: 38170</pre>	
2. Revisión de registros sin producto asignado	
Resultado: No se encontraron valores nulos dentro del atributo Producto.	
<pre>In [28]: df['Producto'].isna().sum() Out[28]: 0</pre>	
<u>Recomendación: Productos más adquiridos por los clientes</u>	
Resultado: Se obtienen los productos más vendidos, ordenados de manera descendente.	
<pre>In [34]: df['Producto'].value_counts().head(7).plot(kind='bar') Out[34]: <AxesSubplot></pre>  <pre>In [35]: df['Producto'].value_counts().head(7) Out[35]: INTRAFINANCIAMIENTO COMPRA EFECTIVO COL 3221 SIN FIADOR 2712 AFILIACION REGULAR 2% 2695 SIN FIADOR VIP 2566 AHORRO PLUS 2067 AHORRO INVERSIÓN 1998 SIN FIADOR VIP VENTANILLA 1848 Name: Producto, dtype: int64</pre>	

9.32. Apéndice AF. Evaluación del modelo KMeans

Código	MOD-002
Modelo utilizado	KMeans (Segmentación)
Objetivo	Evaluar el modelo de <i>KMeans</i> para la recomendación de productos.
Resultados obtenidos	Recomendación de productos basada en la similitud entre los clientes pertenecientes a un mismo segmento.

Datos de prueba

Registro de oportunidades ganadas exportados de Oracle Engagement Cloud, detallado en la sección 5.2.1.

- Cantidad de filas(original): 38170
- Cantidad de columnas(original): 25

Evidencias

Transformaciones y limpieza de datos

1. Limpieza de registros sin un cliente asignado, según el campo “Primary Contact Party ID”.

```
df = df.dropna(subset=['Primary Contact Party ID'])
df.head()
```

2. Creación de nuevo campo “Contact ID”, a partir de Primary Contact Party ID.

```
df['Contact ID'] = df['Primary Contact Party ID'].apply(np.int64)
df.head()
```

antón	Tipo de identificación	Número de identificación	Salario Neto	Perfil Social	Perfil Ambiental	Primary Contact Party ID	Sales Account ID	Country	Postal Code	Reason Won or Lost	Producto	Contact ID
SAN RLOS	1 - Persona Física Nacional		NaN	No aplica	No aplica	1.000000e+14	100000008505943	Costa Rica	188-2-10-1	NaN	Tarjeta de credito-Black	100000008033574
SAN RLOS	1 - Persona Física Nacional		NaN	No aplica	No aplica	3.000001e+14	300000082105690	Costa Rica	188-2-10-1	NaN	CDP Desmaterializado COL	300000082105681
SAN RLOS	1 - Persona Física Nacional		NaN	No aplica	No aplica	3.000001e+14	300000082105690	Costa Rica	188-2-10-1	NaN	CDP Desmaterializado COL	300000082105681
SRAN	1 - Persona Física Nacional		NaN	No aplica	No aplica	1.000000e+14	100000008498132	Costa Rica	188-5-8-1	NaN	Credito sobre CDP COL	100000008088151
TRAL	1 - Persona Física Nacional		NaN	No aplica	No aplica	1.000000e+14	100000008557316	Costa Rica	188-1-1-9	Altos tiempos de respuesta	Sin Fidor	100000008013135

3. Eliminar columnas “Contact Sales Account ID”, “Primary Contact Party ID”, “Sales Account ID”, “Status Code”, “Nombre de la oportunidad”.

```
df.drop(['Contact Sales Account Id', 'Primary Contact Party ID', 'Sales Account ID', 'Status Code', 'Nombre de la oportunidad'], axis=1)
df.head()
```

4. Creacion de nuevo campo Edad, a partir de FechaNacimiento.

```
In [346]: df['FechaNacimiento'] = pd.to_datetime(df['Fecha de nacimiento'])

def age(birthdate):
    today = date.today()
    age = today.year - birthdate.year - ((today.month, today.day) < (birthdate.month, birthdate.day))
    return age

df['Edad'] = df['FechaNacimiento'].apply(age)
df.head()
```

Edad	Distrito	Fecha de nacimiento	Tipo Cliente	Salario Neto	Perfil Social	Perfil Ambiental	Country	Postal Code	Reason Won or Lost	Producto	Contact ID	FechaNacimiento	Edad
48.0	NaN	1973-05-14 06:00:00.0	5 CONVENIENCIA (NAT)	NaN	No aplica	No aplica	Costa Rica	188-2-19-1	NaN	Tarjeta de credito-BIACI	100000008033574	1973-05-14 06:00:00	48.0
72.0	NaN	1949-11-08 06:00:00.0	5 CONVENIENCIA (NAT)	NaN	No aplica	No aplica	Costa Rica	188-2-10-1	NaN	CDP Desmaterializado CCL	300000082105681	1949-11-08 06:00:00	72.0
72.0	NaN	1949-11-08 06:00:00.0	5 CONVENIENCIA (NAT)	NaN	No aplica	No aplica	Costa Rica	188-2-10-1	NaN	CDP Desmaterializado CCL	300000082105681	1949-11-08 06:00:00	72.0
70.0	NaN	1951-07-09 05:00:00.0	5 CONVENIENCIA (NAT)	NaN	No aplica	No aplica	Costa Rica	188-5-8-1	NaN	Credito sobre CDP CCL	100000008088151	1951-07-09 05:00:00	70.0
62.0	San José	1959-09-27 08:00:00.000	1.REGULAR ACTIVO 2%	NaN	No aplica	No aplica	Costa Rica	188-1-1-9	Años tiempo de respuesta	Sin Fiaodr	100000008013135	1959-09-27 08:00:00	62.0

5. Tratamiento de valores nulos.

- a. Edad: Sustitución por valor de media.
- b. Cantón: Sustitución por valor de moda.
- c. Estado Civil: Sustitución por valor de moda.
- d. Categoría de asociado: Sustitución por valor de moda.
- e. Gender: Sustitución por valor de moda.
- f. Distrito: Sustitución por valor de moda.

Tratamiento de datos nulos

```
edad_media=df['Edad'].mean()
df['Edad'].fillna(value=edad_media, inplace=True)

canton_moda=df['Cantón'].mode()
df['Cantón'].fillna(value=canton_moda, inplace=True)

estado_moda=df['Estado Civil'].mode()
df['Estado Civil'].fillna(value=estado_moda, inplace=True)

categoria_moda=df['Categoría de asociado'].mode()
df['Categoría de asociado'].fillna(value=categoria_moda, inplace=True)

genero_moda=df['Gender'].mode()
df['Gender'].fillna(value=genero_moda, inplace=True)

distrito_moda=df['Distrito'].mode()
df['Distrito'].fillna(value=distrito_moda, inplace=True)
```

6. Tratamiento de valores duplicados y sucios:

- a. Tipo de Identificación: Sustitución de valores
- b. Estado Civil: Sustitución de valores “Soltero”, “Soltero (a)”, “Casado”, “Casado (a)”, “Divorciado”, “Divorciado”, “Union Libre”.

```
df["Estado Civil"].replace(('Soltero', 'Soltero (a)'), ('S. Soltero (a)', 'S. Soltero (a)'), inplace=True)
df["Estado Civil"].replace(('Casado', 'Casado (a)'), ('C. Casado (a)', 'C. Casado (a)'), inplace=True)
df["Estado Civil"].replace(('Divorciado (a)', 'Divorciado'), ('D. Divorciado (a)', 'D. Divorciado (a)'), inplace=True)
df["Estado Civil"].replace(('Union Libre'), ('U. Union Libre'), inplace=True)
df["Estado Civil"].value_counts()

S. Soltero (a)    953
C. Casado (a)    858
D. Divorciado (a)  184
U. Union Libre   153
V. Viudo         24
Name: Estado Civil, dtype: int64
```

7. Eliminación de columnas: Cantón, Perfil Social, Perfil Ambiental, Postal Code, Contact ID, Producto).

```
df.drop(["Cantón", "Perfil Social", "Perfil Ambiental", "Postal Code", "Contact ID", "Producto"], axis=1, inplace=True)
df.head()
```

8. Creación de columnas *dummy* para los atributos categóricos (*One Hot Encoding*)

- a. Gender
- b. Tipo Cliente
- c. Tipo de Identificación
- d. Estado Civil
- e. Categoría de asociado
- f. Tipo de asociado

Modelado

```
modelo=df
modelo=pd.get_dummies(modelo, prefix='Gender', columns=['Gender'])
modelo=pd.get_dummies(modelo, prefix='Tipo Cliente', columns=['Tipo Cliente'])
modelo=pd.get_dummies(modelo, prefix='Tipo de Identificación', columns=['Tipo de Identificación'])
modelo=pd.get_dummies(modelo, prefix='Estado Civil', columns=['Estado Civil'])
modelo=pd.get_dummies(modelo, prefix='Categoría de asociado', columns=['Categoría de asociado'])
modelo=pd.get_dummies(modelo, prefix='Tipo de asociado', columns=['Tipo de asociado'])
modelo.head()
```

	Gender_Femenino	Gender_Masculino	Tipo Cliente_1 REGULAR ACTIVO 2%	Tipo de asociado_24 CUENTA CS JOVEN	Tipo de asociado_24 REGULAR CS JOVEN	Tipo de asociado_26 CONVENIENCIA CS JOVEN	Tipo de asociado_26 NATURAL CS JOVEN	Tipo de asociado_30 CUENTA PYMES	Tipo de asociado_33 FISICO	Tipo de asociado_4 CUENTA ANTES CS BIENESTAR 2%
1	1	0	0	0	0	0	0	0	0	0
1	0	1	0	0	0	0	0	0	0	0
1	0	1	0	0	0	0	0	0	0	0
1	0	1	0	0	0	0	0	0	0	0
1	0	1	1	0	0	0	0	0	0	0

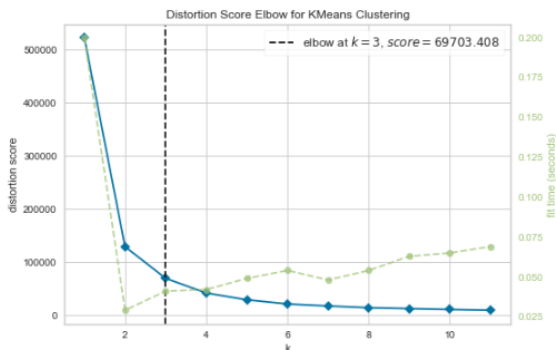
Modelado

Método del Codo

De acuerdo con el análisis del codo, el número óptimo de *clusters* para segmentar los clientes se encuentra en k=3.

```
from sklearn.cluster import KMeans
from sklearn import preprocessing
from yellowbrick.cluster import KElbowVisualizer

model = KMeans()
visualizer = KElbowVisualizer(model, k=(1,12))
visualizer.fit(datos) # Entrenamos con Los datos
visualizer.show()
```



Etiquetar clientes

Luego del análisis del codo, se crea el modelo KMeans, con el k=3 a partir del modelo generado se realizan las predicciones sobre los registros de clientes.

```
modelo.kmeans= KMeans(n_clusters = 3, init = 'k-means++', random_state = 2)
numClust = modelo.kmeans.fit_predict(datos)
modelo["cluster"] = numClust
```

Para cada registro de clientes se le asigno una clase distinta con valores posibles entre 0 y 2.

```
modelo.kmeans= KMeans(n_clusters = 3, init = 'k-means++', random_state = 2)
numClust = modelo.kmeans.fit_predict(datos)
datos["Clase"] = numClust
```

```
datos.head()
```

ipo de lo_24. ULAR OVEN	Tipo de asociado_26. CONVENIENCIA CS JOVEN	Tipo de asociado_26. CUENTA NATURAL CS JOVEN	Tipo de asociado_30. CUENTA PYMES	Tipo de asociado_33. FISICO	Tipo de asociado_4. CUENTA ANTES CS BIENESTAR 2%	Tipo de asociado_4. R. SIN FONDO MUTUAL 2%	Tipo de asociado_5. CONVENIENCIA (NAT)	Tipo de asociado_5. CUENTA NATURAL	Clase
0	0	0	0	0	0	0	0	1	2
0	0	0	0	0	0	0	0	1	0
0	0	0	0	0	0	0	0	1	0
0	0	0	0	0	0	0	0	1	0
0	0	0	0	0	0	0	0	0	0

Recomendación de productos

Identificación de productos más comunes por segmento

```
datos.groupby(['Clase']).Producto.value_counts().groupby(level=0, group_keys=False).head(3)
```

```
Clase  Producto
0      SOBRE AHORRO CORRIENTE    131
      SIN FIADOR VIP              53
      CREDITO SOBRE CDP COL       49
1      AFILIACION REGULAR 2%     126
      SIN FIADOR VIP VENTANILLA  104
      SIN FIADOR VIP              97
2      SOBRE AHORRO CORRIENTE    106
      SIN FIADOR VIP              91
      AFILIACION REGULAR 2%       46
Name: Producto, dtype: int64
```

Creación de recomendaciones para cada segmento de clientes

```

recomendaciones_clase0=filtrado0["Producto"].value_counts().head(3)
recomendaciones_clase1=filtrado1["Producto"].value_counts().head(3)
recomendaciones_clase2=filtrado2["Producto"].value_counts().head(3)

recomendaciones_generadas_clase0=recomendaciones_clase0.index.tolist()
recomendaciones_generadas_clase1=recomendaciones_clase1.index.tolist()
recomendaciones_generadas_clase2=recomendaciones_clase2.index.tolist()
    
```

```

recomendacion.loc[recomendacion["Clase"] == 0, "Recomendacion1"] = recomendaciones_generadas_clase0[0]
recomendacion.loc[recomendacion["Clase"] == 0, "Recomendacion2"] = recomendaciones_generadas_clase0[1]
recomendacion.loc[recomendacion["Clase"] == 0, "Recomendacion3"] = recomendaciones_generadas_clase0[2]

recomendacion.loc[recomendacion["Clase"] == 1, "Recomendacion1"] = recomendaciones_generadas_clase1[0]
recomendacion.loc[recomendacion["Clase"] == 1, "Recomendacion2"] = recomendaciones_generadas_clase1[1]
recomendacion.loc[recomendacion["Clase"] == 1, "Recomendacion3"] = recomendaciones_generadas_clase1[2]

recomendacion.loc[recomendacion["Clase"] == 2, "Recomendacion1"] = recomendaciones_generadas_clase2[0]
recomendacion.loc[recomendacion["Clase"] == 2, "Recomendacion2"] = recomendaciones_generadas_clase2[1]
recomendacion.loc[recomendacion["Clase"] == 2, "Recomendacion3"] = recomendaciones_generadas_clase2[2]
    
```

Creación de dataframe con recomendaciones por cliente

```
recomendacion[recomendacion["Clase"]==0].head()
```

Contact ID	Clase	Recomendacion1	Recomendacion2	Recomendacion3
1	300000002105681	0 SOBRE AHORRO CORRIENTE	SIN FIADOR VIP	CREDITO SOBRE CDP COL
2	300000002105681	0 SOBRE AHORRO CORRIENTE	SIN FIADOR VIP	CREDITO SOBRE CDP COL
3	100000000088151	0 SOBRE AHORRO CORRIENTE	SIN FIADOR VIP	CREDITO SOBRE CDP COL
4	10000000013135	0 SOBRE AHORRO CORRIENTE	SIN FIADOR VIP	CREDITO SOBRE CDP COL
16	100000000034366	0 SOBRE AHORRO CORRIENTE	SIN FIADOR VIP	CREDITO SOBRE CDP COL

```
recomendacion[recomendacion["Clase"]==1].head()
```

Contact ID	Clase	Recomendacion1	Recomendacion2	Recomendacion3
5	300000131195527	1 AFILIACION REGULAR 2%	SIN FIADOR VIP VENTANILLA	SIN FIADOR VIP
6	300000139602420	1 AFILIACION REGULAR 2%	SIN FIADOR VIP VENTANILLA	SIN FIADOR VIP
9	300000139676531	1 AFILIACION REGULAR 2%	SIN FIADOR VIP VENTANILLA	SIN FIADOR VIP
10	100000033674006	1 AFILIACION REGULAR 2%	SIN FIADOR VIP VENTANILLA	SIN FIADOR VIP
11	100000000233302	1 AFILIACION REGULAR 2%	SIN FIADOR VIP VENTANILLA	SIN FIADOR VIP

```
recomendacion[recomendacion["Clase"]==2].head()
```

Contact ID	Clase	Recomendacion1	Recomendacion2	Recomendacion3
0	100000000033574	2 SOBRE AHORRO CORRIENTE	SIN FIADOR VIP	AFILIACION REGULAR 2%
7	300000139657206	2 SOBRE AHORRO CORRIENTE	SIN FIADOR VIP	AFILIACION REGULAR 2%
8	300000139079897	2 SOBRE AHORRO CORRIENTE	SIN FIADOR VIP	AFILIACION REGULAR 2%
12	100000032737252	2 SOBRE AHORRO CORRIENTE	SIN FIADOR VIP	AFILIACION REGULAR 2%
13	100000000022041	2 SOBRE AHORRO CORRIENTE	SIN FIADOR VIP	AFILIACION REGULAR 2%

9.33. Apéndice AF. Evaluación del modelo Collaborative Filtering

Código	MOD-003																								
Modelo utilizado	Collaborative Filtering (Implícito)																								
Objetivo	Evaluar el modelo de <i>Collaborative Filtering</i> para la recomendación de productos con los datos extraídos de Oracle Engagement Cloud.																								
Resultados obtenidos	Recomendación basada en la similitud entre los productos “puntuados” por los clientes.																								
Datos de prueba	<p>Registro de oportunidades ganadas exportados de Oracle Engagement Cloud, detallado en la sección 5.2.1.</p> <ul style="list-style-type: none"> Cantidad de filas(original): 38170 Cantidad de columnas(original): 25 																								
Evidencias	<p><u>Transformación y limpieza de datos</u></p> <p><u>Eliminación de valores nulos en columna Primary Contact ID y creación del campo Contact ID</u></p> <pre>import numpy as np df = df.dropna(subset=['Primary Contact Party ID']) df['Contact ID'] = df['Primary Contact Party ID'].apply(np.int64)</pre> <p><u>Creación de nuevo campo constante Rating</u></p> <pre>df['rating'] = 5</pre> <p><u>Filtrado de columnas</u> Únicamente quedan Contact ID, Producto, rating</p> <pre>test_recomendacion=df[['Contact ID','Producto','rating']] test_recomendacion.head()</pre> <table border="1"> <thead> <tr> <th></th> <th>Contact ID</th> <th>Producto</th> <th>rating</th> </tr> </thead> <tbody> <tr> <td>0</td> <td>100000008033574</td> <td>Tarjeta de credito-Black</td> <td>5</td> </tr> <tr> <td>1</td> <td>300000082105681</td> <td>CDP Desmaterializado COL</td> <td>5</td> </tr> <tr> <td>2</td> <td>300000082105681</td> <td>CDP Desmaterializado COL</td> <td>5</td> </tr> <tr> <td>3</td> <td>100000008088151</td> <td>Credito sobre CDP COL</td> <td>5</td> </tr> <tr> <td>4</td> <td>100000008013135</td> <td>Sin Fiaador</td> <td>5</td> </tr> </tbody> </table> <p><u>Creación de tabla pivot</u> Productos pasan a ser las columnas de la tabla, Contact ID (únicos) pasan a ser las filas. Indica 5 en los productos que el usuario ha adquirido previamente.</p>		Contact ID	Producto	rating	0	100000008033574	Tarjeta de credito-Black	5	1	300000082105681	CDP Desmaterializado COL	5	2	300000082105681	CDP Desmaterializado COL	5	3	100000008088151	Credito sobre CDP COL	5	4	100000008013135	Sin Fiaador	5
	Contact ID	Producto	rating																						
0	100000008033574	Tarjeta de credito-Black	5																						
1	300000082105681	CDP Desmaterializado COL	5																						
2	300000082105681	CDP Desmaterializado COL	5																						
3	100000008088151	Credito sobre CDP COL	5																						
4	100000008013135	Sin Fiaador	5																						


```

collaborative_table = pd.pivot_table(data=test_recomendacion, index='Contact ID', columns='Producto', values='rating', \
aggfunc='sum', fill_value=0)
collaborative_table

```

Producto	AFILIACION CONVENIENCIA	AFILIACION CS JOVEN	AFILIACION CS JOVEN-CONVENIENCIA	AFILIACION PYMES-REGULAR	AFILIACION REGULAR	AHORRO ESTUDIANTL 2%	AHORRO MARCHAMO	AHORRO NAVIDIENO	AHORRO PLUS	AHORRO PLUS DOLARES	...	PL
Contact ID												
10000008001960	0	0	0	0	0	0	0	0	0	0	0	...
10000008002003	0	0	0	0	0	0	0	0	0	0	0	...
10000008011777	0	0	0	0	0	0	0	0	0	0	0	...
10000008011806	0	0	0	0	0	0	0	0	0	0	0	...
10000008011843	0	0	0	0	0	0	0	0	0	0	0	...
...
300000149754496	0	0	0	0	0	0	0	0	0	0	0	...
300000149756517	0	0	0	0	0	0	0	0	0	0	0	...
300000149762899	0	0	0	0	0	5	0	0	0	0	0	...
300000171508976	0	0	0	0	0	0	0	0	0	0	0	...
300000172636529	0	0	0	0	0	0	0	0	0	0	0	...

1582 rows x 65 columns

Singular Value Decomposition

Factorización de los vectores

```

import sklearn
from sklearn.decomposition import TruncatedSVD

X = collaborative_table.T
SVD = TruncatedSVD(n_components=12, random_state=5)
resultant_matrix = SVD.fit_transform(X)
resultant_matrix.shape

```

Creación de matriz de correlación

```

import numpy as np
corr_mat = np.corrcoef(resultant_matrix)
corr_mat.shape

```

(65, 65)

Búsqueda de productos similares

```

col_idx = collaborative_table.columns.get_loc("AFILIACION CONVENIENCIA")
corr_specific = corr_mat[col_idx]
pd.DataFrame({'corr_specific':corr_specific, 'Productos': collaborative_table.columns})\
.sort_values('corr_specific', ascending=False)\
.head(10)

```

	corr_specific	Productos
0	1.000000	AFILIACION CONVENIENCIA
14	0.966029	CDP DESMATERIALIZADO DOL
12	0.744010	CDP COLONES
9	0.722155	AHORRO PLUS DOLARES
2	0.704306	AFILIACION CS JOVEN-CONVENIENCIA
13	0.686339	CDP DESMATERIALIZADO COL
6	0.674319	AHORRO MARCHAMO
11	0.670846	AHORRO VACACIONAL
31	0.667690	SERVICLAVE
60	0.661164	Tarjeta de credito-Black

9.34. Apéndice AG. Evaluación del modelo Naive Bayes (Gaussian)

Código	MOD-003
Modelo utilizado	Naive Bayes (Gaussian)
Objetivo	Evaluar el modelo de <i>Naive Bayes</i> para la recomendación de productos.
Resultados obtenidos	Recomendación de productos basada en los atributos de los clientes con productos similares.

Datos de prueba

Registro de oportunidades ganadas exportados de Oracle Engagement Cloud, detallado en la sección 5.2.1.

- Cantidad de filas(original): 38170
- Cantidad de columnas(original): 25

Evidencias

Transformaciones y limpieza de datos

1. Limpieza de registros sin un cliente asignado, según el campo “Primary Contact Party ID”.

```
df = df.dropna(subset=['Primary Contact Party ID'])
df.head()
```

2. Creación de nuevo campo “Contact ID”, a partir de Primary Contact Party ID.

```
df['Contact ID'] = df['Primary Contact Party ID'].apply(np.int64)
df.head()
```

Intón	Tipo de Identificación	Número de Identificación	Salario Neto	Perfil Social	Perfil Ambiental	Primary Contact Party ID	Sales Account ID	Country	Postal Code	Reason Won or Lost	Producto	Contact ID
SAN RLOS	1 - Persona Física Nacional		NaN	No aplica	No aplica	1.000000e+14	100000008505943	Costa Rica	188-2-10-1	NaN	Tarjeta de credito-Black	100000008033574
SAN RLOS	1 - Persona Física Nacional		NaN	No aplica	No aplica	3.000001e+14	300000082105690	Costa Rica	188-2-10-1	NaN	Desmaterializado COL	300000082105681
SAN RLOS	1 - Persona Física Nacional		NaN	No aplica	No aplica	3.000001e+14	300000082105690	Costa Rica	188-2-10-1	NaN	Desmaterializado COL	300000082105681
IRAN	1 - Persona Física Nacional		NaN	No aplica	No aplica	1.000000e+14	100000008498132	Costa Rica	188-5-8-1	NaN	Credito sobre CDP COL	100000008088151
TRAL	1 - Persona Física Nacional		NaN	No aplica	No aplica	1.000000e+14	100000008557316	Costa Rica	188-1-1-9	Altos tiempos de respuesta	Sin Fidor	100000008013135

3. Eliminar columnas “Contact Sales Account ID”, “Primary Contact Party ID”, “Sales Account ID”, “Status Code”, “Nombre de la oportunidad”.

```
df.drop(['Contact Sales Account Id', 'Primary Contact Party ID', 'Sales Account ID', 'Status Code', 'Nombre de la oportunidad'], axis=1)
df.head()
```

4. Creación de nuevo campo Edad, a partir de FechaNacimiento.

```
In [346]: df['FechaNacimiento'] = pd.to_datetime(df['Fecha de nacimiento'])

def age(birthdate):
    today = date.today()
    age = today.year - birthdate.year - ((today.month, today.day) < (birthdate.month, birthdate.day))
    return age

df['Edad'] = df['FechaNacimiento'].apply(age)
df.head()
```

Índice	Distrito	Fecha de nacimiento	Tipo Cliente	Salario Neto	Perfil Social	Perfil Ambiental	Country	Postal Code	Reason Won or Lost	Producto	Contact ID	FechaNacimiento	Edad
0.0	NaN	1973-05-14 06:00:00.0	5 CONVENIENCIA (NAT)	NaN	No aplica	No aplica	Costa Rica	188-2-19-1	NaN	Tarjeta de credito-BIACI	100000008033574	1973-05-14 06:00:00	48.0
1.0	NaN	1949-11-08 06:00:00.0	5 CONVENIENCIA (NAT)	NaN	No aplica	No aplica	Costa Rica	188-2-10-1	NaN	Desmaterializado CCL	300000082105681	1949-11-08 06:00:00	72.0
2.0	NaN	1949-11-08 06:00:00.0	5 CONVENIENCIA (NAT)	NaN	No aplica	No aplica	Costa Rica	188-2-10-1	NaN	Desmaterializado CCL	300000082105681	1949-11-08 06:00:00	72.0
3.0	NaN	1951-07-09 05:00:00.0	5 CONVENIENCIA (NAT)	NaN	No aplica	No aplica	Costa Rica	188-5-8-1	NaN	Credito sobre CDP-COL	100000008088151	1951-07-09 05:00:00	70.0
4.0	San José	1959-09-27 08:00:00.000	1.REGULAR ACTIVO 2%	NaN	No aplica	No aplica	Costa Rica	188-1-1-9	Atos tiempos de respuesta	Sin Fiaodr	100000008013135	1959-09-27 08:00:00	62.0

5. Tratamiento de valores nulos.

- a. Edad: Sustitución por valor de media.
- b. Cantón: Sustitución por valor de moda.
- c. Estado Civil: Sustitución por valor de moda.
- d. Categoría de asociado: Sustitución por valor de moda.
- e. Gender: Sustitución por valor de moda.
- f. Distrito: Sustitución por valor de moda.

Tratamiento de datos nulos

```
edad_media=df['Edad'].mean()
df['Edad'].fillna(value=edad_media, inplace=True)

canton_moda=df['Cantón'].mode()
df['Cantón'].fillna(value=canton_moda, inplace=True)

estado_moda=df['Estado Civil'].mode()
df['Estado Civil'].fillna(value=estado_moda, inplace=True)

categoria_moda=df['Categoría de asociado'].mode()
df['Categoría de asociado'].fillna(value=categoria_moda, inplace=True)

genero_moda=df['Gender'].mode()
df['Gender'].fillna(value=genero_moda, inplace=True)

distrito_moda=df['Distrito'].mode()
df['Distrito'].fillna(value=distrito_moda, inplace=True)
```

6. Tratamiento de valores duplicados y sucios:

- a. Tipo de Identificación: Sustitución de valores
- b. Estado Civil: Sustitución de valores “Soltero,” Soltero (a)”, “Casado”,” Casado (a)”, “Divorciado”,” Divorciado”, “Union Libre”.

```
df["Estado Civil"].replace(('Soltero', 'Soltero (a)'), ('S. Soltero (a)', 'S. Soltero (a)'), inplace=True)
df["Estado Civil"].replace(('Casado', 'Casado (a)'), ('C. Casado (a)', 'C. Casado (a)'), inplace=True)
df["Estado Civil"].replace(('Divorciado (a)', 'Divorciado'), ('D. Divorciado (a)', 'D. Divorciado (a)'), inplace=True)
df["Estado Civil"].replace(('Union Libre'), ('U. Union Libre'), inplace=True)
df["Estado Civil"].value_counts()
```

S. Soltero (a)	953
C. Casado (a)	858
D. Divorciado (a)	184
U. Union Libre	153
V. Viudo	24

Name: Estado Civil, dtype: int64

7. Eliminación de columnas: Cantón, Perfil Social, Perfil Ambiental, Postal Code, Contact ID, Producto).

```
df.drop(["Cantón", "Perfil Social", "Perfil Ambiental", "Postal Code", "Contact ID", "Producto"], axis=1, inplace=True)
df.head()
```

8. Creación de columnas *dummy* para los atributos categóricos (*One Hot Encoding*)

- a. Gender
- b. Tipo Cliente
- c. Tipo de Identificación
- d. Estado Civil
- e. Categoría de asociado
- f. Tipo de asociado

Modelado

```
modelo=df
modelo=pd.get_dummies(modelo, prefix='Gender', columns=['Gender'])
modelo=pd.get_dummies(modelo, prefix='Tipo Cliente', columns=['Tipo Cliente'])
modelo=pd.get_dummies(modelo, prefix='Tipo de Identificación', columns=['Tipo de Identificación'])
modelo=pd.get_dummies(modelo, prefix='Estado Civil', columns=['Estado Civil'])
modelo=pd.get_dummies(modelo, prefix='Categoría de asociado', columns=['Categoría de asociado'])
modelo=pd.get_dummies(modelo, prefix='Tipo de asociado', columns=['Tipo de asociado'])
modelo.head()
```

	Gender_Femenino	Gender_Masculino	Tipo Cliente_1 REGULAR ACTIVO 2%	Tipo de asociado_24 CUENTAS JOVEN	Tipo de asociado_24 REGULAR CS JOVEN	Tipo de asociado_26 CONVENIENCIA CS JOVEN	Tipo de asociado_26 CUENTA NATURAL CS JOVEN	Tipo de asociado_30 CUENTA PYMES	Tipo de asociado_33 FISICO	Tipo de asociado_4 CUENTA ANTES CS BIENESTAR 2%
1	1	0	0	0	0	0	0	0	0	0
1	0	1	0	0	0	0	0	0	0	0
1	0	1	0	0	0	0	0	0	0	0
1	0	1	0	0	0	0	0	0	0	0
1	0	1	1	0	0	0	0	0	0	0

Modelado

Selección de variable a predecir

La variable por predecir corresponde al nombre del producto vendido a cada cliente, el resto de los campos disponibles corresponde a los atributos, que serán utilizados por el modelo.

```
X = modelo.iloc[:, [0,51]].values
y = modelo.iloc[:, 52].values
```

División de datos en entrenamiento y pruebas

Se realizó una división de 80% para entrenamiento y 20% para pruebas.

```
from sklearn.model_selection import train_test_split

X_train, X_test, y_train, y_test = train_test_split(X,y,test_size= 0.20, random_state=0)
```

Escalamiento de los datos

Los valores numéricos se transforman de manera que todos los datos se encuentran en una escala de cero a uno.

```
# importing standard scaler
from sklearn.preprocessing import StandardScaler

# scaling the input data
sc_X = StandardScaler()
X_train = sc_X.fit_transform(X_train)
X_test = sc_X.fit_transform(X_test)
```

Pruebas del modelo

El modelo creado y entrenado previamente, fue utilizado para realizar predicciones de ejemplo con los valores correspondientes al conjunto de datos de prueba.

```
# import Gaussian Naive Bayes classifier
from sklearn.naive_bayes import GaussianNB

# create a Gaussian Classifier
classifier1 = GaussianNB()

# training the model
classifier1.fit(X_train, y_train)

# testing the model
y_pred1 = classifier1.predict(X_test)
```

Evaluación del modelo

Se obtuvieron cuatro métricas de evaluación del modelo de clasificación Naive Bayes. Las métricas utilizadas corresponden a *accuracy*, *recall*, *f1* y *precision*.

```
# importing accuracy score
from sklearn.metrics import accuracy_score

# printing the accuracy of the model
print(accuracy_score(y_test, y_pred1))
```

0.1480637813211845

```
from sklearn.metrics import f1_score
f1_score(y_test, y_pred, average='weighted')
```

0.07326454132244417

```
from sklearn.metrics import recall_score
recall_score(y_test, y_pred, average='weighted')
```

0.15034168564920272

```
from sklearn.metrics import precision_score
precision_score(y_test, y_pred, average='weighted')
```

C:\ProgramData\Anaconda3\lib\site-packages\sklearn\metrics_classification.py:1318: UndefinedMetricWarning: Precision is ill-defined and being set to 0.0 in labels with no predicted samples. Use `zero_division` parameter to control this behavior.
_warn_prf(average, modifier, msg_start, len(result))

0.05074480147887004

9.35. Apéndice AH. Evaluación del modelo Naive Bayes (Bernoulli)

Código	MOD-004
Modelo utilizado	Naive Bayes (Bernoulli)
Objetivo	Evaluar el modelo de <i>Naive Bayes</i> para la recomendación de productos.
Resultados obtenidos	Recomendación de productos basada en los atributos de los clientes con productos similares.

Datos de prueba

Registro de oportunidades ganadas exportados de Oracle Engagement Cloud, detallado en la sección 5.2.1.

- Cantidad de filas(original): 38170
- Cantidad de columnas(original): 25

Evidencias

Transformaciones y limpieza de datos

9. Limpieza de registros sin un cliente asignado, según el campo “Primary Contact Party ID”.

```
df = df.dropna(subset=['Primary Contact Party ID'])
df.head()
```

10. Creación de nuevo campo “Contact ID”, a partir de Primary Contact Party ID.

```
df['Contact ID'] = df['Primary Contact Party ID'].apply(np.int64)
df.head()
```

Intón	Tipo de Identificación	Número de Identificación	Salario Neto	Perfil Social	Perfil Ambiental	Primary Contact Party ID	Sales Account ID	Country	Postal Code	Reason Won or Lost	Producto	Contact ID
SAN RLOS	1 - Persona Física Nacional		NaN	No aplica	No aplica	1.000000e+14	100000008505943	Costa Rica	188-2-10-1	NaN	Tarjeta de credito-Black	100000008033574
SAN RLOS	1 - Persona Física Nacional		NaN	No aplica	No aplica	3.000001e+14	300000082105690	Costa Rica	188-2-10-1	NaN	Desmaterializado COL	300000082105681
SAN RLOS	1 - Persona Física Nacional		NaN	No aplica	No aplica	3.000001e+14	300000082105690	Costa Rica	188-2-10-1	NaN	Desmaterializado COL	300000082105681
IRAN	1 - Persona Física Nacional		NaN	No aplica	No aplica	1.000000e+14	100000008498132	Costa Rica	188-5-8-1	NaN	Credito sobre CDP COL	100000008088151
TRAL	1 - Persona Física Nacional		NaN	No aplica	No aplica	1.000000e+14	100000008557316	Costa Rica	188-1-1-9	Altos tiempos de respuesta	Sin Fidor	100000008013135

11. Eliminar columnas “Contact Sales Account ID”, “Primary Contact Party ID”, “Sales Account ID”, “Status Code”, “Nombre de la oportunidad”.

```
df.drop(['Contact Sales Account Id', 'Primary Contact Party ID', 'Sales Account ID', 'Status Code', 'Nombre de la oportunidad'], axis=1)
df.head()
```

12. Creacion de nuevo campo Edad, a partir de FechaNacimiento.

```
In [346]: df['FechaNacimiento'] = pd.to_datetime(df['Fecha de nacimiento'])

def age(birthdate):
    today = date.today()
    age = today.year - birthdate.year - ((today.month, today.day) < (birthdate.month, birthdate.day))
    return age

df['Edad'] = df['FechaNacimiento'].apply(age)
df.head()
```

Edad	Distrito	Fecha de nacimiento	Tipo Cliente	Salario Neto	Perfil Social	Perfil Ambiental	Country	Postal Code	Reason Won or Lost	Producto	Contact ID	FechaNacimiento	Edad
48.0	NaN	1973-05-14 06:00:00.0	5 CONVENIENCIA (NAT)	NaN	No aplica	No aplica	Costa Rica	188-2-19-1	NaN	Tarjeta de credito-BIACI	100000008033574	1973-05-14 06:00:00	48.0
72.0	NaN	1949-11-08 06:00:00.0	5 CONVENIENCIA (NAT)	NaN	No aplica	No aplica	Costa Rica	188-2-10-1	NaN	CDP Desmaterializado CCL	300000082105681	1949-11-08 06:00:00	72.0
72.0	NaN	1949-11-08 06:00:00.0	5 CONVENIENCIA (NAT)	NaN	No aplica	No aplica	Costa Rica	188-2-10-1	NaN	CDP Desmaterializado CCL	300000082105681	1949-11-08 06:00:00	72.0
70.0	NaN	1951-07-09 05:00:00.0	5 CONVENIENCIA (NAT)	NaN	No aplica	No aplica	Costa Rica	188-5-8-1	NaN	Credito sobre CDP-CCL	100000008088151	1951-07-09 05:00:00	70.0
62.0	San José	1959-09-27 08:00:00.000	1.REGULAR ACTIVO 2%	NaN	No aplica	No aplica	Costa Rica	188-1-1-9	Años tiempo de respuesta	Sin Fiaador	100000008013135	1959-09-27 08:00:00	62.0

13. Tratamiento de valores nulos.

- Edad: Sustitución por valor de media.
- Cantón: Sustitución por valor de moda.
- Estado Civil: Sustitución por valor de moda.
- Categoría de asociado: Sustitución por valor de moda.
- Gender: Sustitución por valor de moda.
- Distrito: Sustitución por valor de moda.

Tratamiento de datos nulos

```
edad_media=df['Edad'].mean()
df['Edad'].fillna(value=edad_media, inplace=True)

canton_moda=df['Cantón'].mode()
df['Cantón'].fillna(value=canton_moda, inplace=True)

estado_moda=df['Estado Civil'].mode()
df['Estado Civil'].fillna(value=estado_moda, inplace=True)

categoria_moda=df['Categoría de asociado'].mode()
df['Categoría de asociado'].fillna(value=categoria_moda, inplace=True)

genero_moda=df['Gender'].mode()
df['Gender'].fillna(value=genero_moda, inplace=True)

distrito_moda=df['Distrito'].mode()
df['Distrito'].fillna(value=distrito_moda, inplace=True)
```

14. Tratamiento de valores duplicados y sucios:

- Tipo de Identificación: Sustitución de valores
- Estado Civil: Sustitución de valores “Soltero”, “Soltero (a)”, “Casado”, “Casado (a)”, “Divorciado”, “Divorciado”, “Union Libre”.

```
df["Estado Civil"].replace(('Soltero', 'Soltero (a)'), ('S. Soltero (a)', 'S. Soltero (a)'), inplace=True)
df["Estado Civil"].replace(('Casado', 'Casado (a)'), ('C. Casado (a)', 'C. Casado (a)'), inplace=True)
df["Estado Civil"].replace(('Divorciado (a)', 'Divorciado'), ('D. Divorciado (a)', 'D. Divorciado (a)'), inplace=True)
df["Estado Civil"].replace(('Union Libre'), ('U. Union Libre'), inplace=True)
df["Estado Civil"].value_counts()

S. Soltero (a)    953
C. Casado (a)    858
D. Divorciado (a)  184
U. Union Libre   153
V. Viudo         24
Name: Estado Civil, dtype: int64
```

15. Eliminación de columnas: Cantón, Perfil Social, Perfil Ambiental, Postal Code, Contact ID, Producto).

```
df.drop(["Cantón", "Perfil Social", "Perfil Ambiental", "Postal Code", "Contact ID", "Producto"], axis=1, inplace=True)
df.head()
```

16. Creación de columnas *dummy* para los atributos categóricos (*One Hot Encoding*)

- Gender
- Tipo Cliente
- Tipo de Identificación
- Estado Civil
- Categoría de asociado
- Tipo de asociado

Modelado

```
modelo=df
modelo=pd.get_dummies(modelo, prefix='Gender', columns=['Gender'])
modelo=pd.get_dummies(modelo, prefix='Tipo Cliente', columns=['Tipo Cliente'])
modelo=pd.get_dummies(modelo, prefix='Tipo de Identificación', columns=['Tipo de Identificación'])
modelo=pd.get_dummies(modelo, prefix='Estado Civil', columns=['Estado Civil'])
modelo=pd.get_dummies(modelo, prefix='Categoría de asociado', columns=['Categoría de asociado'])
modelo=pd.get_dummies(modelo, prefix='Tipo de asociado', columns=['Tipo de asociado'])
modelo.head()
```

	Gender_Femenino	Gender_Masculino	Tipo Cliente_1 REGULAR ACTIVO 2%	Tipo de asociado_24 CUENTAS JOVEN	Tipo de asociado_24 REGULAR CS JOVEN	Tipo de asociado_26 CONVENIENCIA CS JOVEN	Tipo de asociado_26 CUENTA NATURAL CS JOVEN	Tipo de asociado_30 CUENTA PYMES	Tipo de asociado_33 FISICO	Tipo de asociado_4 CUENTA ANTES CS BIENESTAR 2%
1	1	0	0	0	0	0	0	0	0	0
1	0	1	0	0	0	0	0	0	0	0
1	0	1	0	0	0	0	0	0	0	0
1	0	1	0	0	0	0	0	0	0	0
1	0	1	1	0	0	0	0	0	0	0

Modelado

Selección de variable a predecir

La variable por predecir corresponde al nombre del producto vendido a cada cliente, el resto de los campos disponibles corresponde a los atributos, que serán utilizados por el modelo.

```
X = modelo.iloc[:, [0,51]].values
y = modelo.iloc[:, 52].values
```

División de datos en entrenamiento y pruebas

Se realizó una división de 80% para entrenamiento y 20% para pruebas.

```
from sklearn.model_selection import train_test_split

X_train, X_test, y_train, y_test = train_test_split(X,y,test_size= 0.20, random_state=0)
```

Escalamiento de los datos

Los valores numéricos se transforman de manera que todos los datos se encuentran en una escala de cero a uno.


```
# importing standard scaler
from sklearn.preprocessing import StandardScaler

# scaling the input data
sc_X = StandardScaler()
X_train = sc_X.fit_transform(X_train)
X_test = sc_X.fit_transform(X_test)
```

Pruebas del modelo

El modelo creado y entrenado previamente, fue utilizado para realizar predicciones de ejemplo con los valores correspondientes al conjunto de datos de prueba.

```
# importing classifier
from sklearn.naive_bayes import BernoulliNB

# initializaing the NB
classifier = BernoulliNB()

# training the model
classifier.fit(X_train, y_train)

# testing the model
y_pred = classifier.predict(X_test)
```

Evaluación del modelo

Se obtuvieron cuatro métricas de evaluación del modelo de clasificación Naive Bayes. Las métricas utilizadas corresponden a *accuracy*, *recall*, *f1* y *precision*.

```
# importing accuracy score
from sklearn.metrics import accuracy_score

# printing the accuracy of the model
print(accuracy_score(y_pred, y_test))
```

0.15034168564920272

```
from sklearn.metrics import f1_score
f1_score(y_test, y_pred, average='weighted')
```

0.07326454132244417

```
from sklearn.metrics import recall_score
recall_score(y_test, y_pred, average='weighted')
```

0.15034168564920272

```
from sklearn.metrics import precision_score
precision_score(y_test, y_pred, average='weighted')
```

C:\ProgramData\Anaconda3\lib\site-packages\sklearn\metrics_classification.py:1318: UndefinedMetricWarning: Precision is ill-defined and being set to 0.0 in labels with no predicted samples. Use `zero_division` parameter to control this behavior.
_warn_prf(average, modifier, msg_start, len(result))

0.05074480147887004

9.36. Apéndice AI. Configuración de campo Fecha de creación

Edit Standard Field: Creation Date Save and Close Cancel

Appearance
Configure how this field will appear when displayed to your users.

* Display Label: Creation Date
Help Text: []

Name
Each field requires a unique name. Name and description are for internal use only, and are never displayed to your users.

Name: CreationDate
API Name: CreationDate

Constraints

Required ^{RP}
 Updatable ^{RP}
 Searchable

9.37. Apéndice AJ. Configuración de campo Cliente asociado

Edit Dynamic Choice List : Cliente asociado : Basic Information Back Next Submit Cancel

Appearance
Configure how this field will appear when displayed to your users.

* Display Label: Cliente asociado
Display Width: 40 Characters
Help Text: []

Name
Each field requires a unique name. Name and description are for internal use only, and are never displayed to your users.

Name: XUM_ClienteAsociado_FDS
API Name: XUM_ClienteAsociado_FDS_c
Description: []

Constraints

Required ^{RP}
 Updatable ^{RP}
 Searchable

Depends On: []
 Include in Service Payload

9.38. Apéndice AK. Configuración de campo Correo cliente

Edit Text Field : Correo cliente Save and Close Cancel

Appearance
Configure how this field will appear when displayed to your users.

* Display Label: Correo cliente
Display Width: 30 Characters
Help Text: []
Display Type: Simple Text Box Multiline Text Area

Name
Each field requires a unique name. Name and description are for internal use only, and are never displayed to your users.

Name: XUM_CorreoCliente_FDS
API Name: XUM_CorreoCliente_FDS_c
Description: []

Constraints

Required ^{RP}
 Updatable ^{RP}
 Searchable

Indexed
 Minimum Length: [] Characters
 Maximum Length: 80 Characters

Depends On: []
 Include in Service Payload

Select the fields whose data changes will cause this field's constant expressions to be reevaluated

Default Value
Enter the value you want to set for the field when an object is created. Select Expression if you want to set the default dynamically.

Fixed Value: []
 Expression: []

9.39. Apéndice AL. Configuración de campo Producto

Edit Text Field : Producto Save and Close Cancel

Appearance
Configure how this field will appear when displayed to your users.

* Display Label: Display Width: Characters
Help Text:
Display Type: Simple Text Box Multiline Text Area

Name
Each field requires a unique name. Name and description are for internal use only, and are never displayed to your users.

Name: Description:
API Name:

Constraints

Required Indexed Depends On:
 Updatable Minimum Length: Characters Include in Service Payload
 Searchable Maximum Length: Characters

Default Value
Enter the value you want to set for the field when an object is created. Select Expression if you want to set the default dynamically.

Fixed Value:
 Expression:

9.40. Apéndice AM. Configuración de campo Puntuación

Edit Number Field : Puntuacion Save and Close Cancel

Appearance
Configure how this field will appear when displayed to your users.

* Display Label: Display Width: Characters
Help Text:

Name
Each field requires a unique name. Name and description are for internal use only, and are never displayed to your users.

Name: Description:
API Name:

Constraints

Required Indexed Maximum Value:
 Updatable Minimum Length: Characters Decimal Places: Include in Service Payload
 Searchable Minimum Value: Depends On:

Default Value
Enter the value you want to set for the field when an object is created. Select Expression if you want to set the default dynamically.

Fixed Value:
 Expression:

9.41. Apéndice AO. Configuración de campo Estado

Edit Fixed Choice List : Estado Save and Close Cancel

Appearance
Configure how this field will appear when displayed to your users

* Display Label: Estado Display Type: Single Select Choice List
 Multiple Select Choice List

Help Text:

Name
Each field requires a unique name. Name and description are for internal use only, and are never displayed to your users.

Name: XUM_Estado_FDS Description:
API Name: XUM_Estado_FDS_c

Constraints

Required Updatable Searchable Depends On:
 Include in Service Payload

List of Values
Configure the list of values you want to display in the choice list. Click the search icon to select a predefined lookup type, or create a new one.

* Lookup Type: Estado_recomendacion Constrain list by parent field value selection

Default Value
Enter the value you want to set for the field when an object is created. Select Expression if you want to set the default dynamically.

Fixed Value: Pendiente Expression:

9.42. Apéndice AP. Configuración de campo Encargado

Edit Dynamic Choice List : Encargado : Basic Information Back Next Submit Cancel

Appearance
Configure how this field will appear when displayed to your users

* Display Label: Encargado Display Width: 40 Characters

Help Text:

Name
Each field requires a unique name. Name and description are for internal use only, and are never displayed to your users.

Name: XUM_Encargado_FDS Description:
API Name: XUM_Encargado_FDS_c

Constraints

Required Updatable Searchable Depends On:
 Include in Service Payload

9.43. Apéndice AQ. Configuración de campo Motivo de rechazo

Edit Fixed Choice List : Motivo de rechazo Save and Close Cancel

Appearance
Configure how this field will appear when displayed to your users.

* Display Label: Display Type: Single Select Choice List Multiple Select Choice List

Help Text:

Name
Each field requires a unique name. Name and description are for internal use only, and are never displayed to your users.

Name: Description:

API Name:

Constraints

Required Updatable Searchable Include in Service Payload

Depends On:

List of Values
Configure the list of values you want to display in the choice list. Click the search icon to select a predefined lookup type, or create a new one.

* Lookup Type:

Constrain list by parent field value selection

Default Value
Enter the value you want to set for the field when an object is created. Select Expression if you want to set the default dynamically.

Fixed Value:

Expression:

9.44. Apéndice AR. Configuración de campo Comentarios adicionales

Edit Long Text Field : Comentarios adicionales Save and Close Cancel

Appearance
Configure how this field will appear when displayed to your users.

* Display Label: Display Width: Characters

Help Text:

Name
Each field requires a unique name. Name and description are for internal use only, and are never displayed to your users.

Name: Description:

API Name:

Constraints

Required Updatable Include in Service Payload

Depends On:

Default Value
Enter the value you want to set for the field when an object is created. Select Expression if you want to set the default dynamically.

Fixed Value:

Expression:

9.45. Apéndice AS. Configuración de lista de valores Estado

Edit Lookup Type

* Meaning Description

Lookup Type ESTADO_RECOMENDACION

One or more lookup codes cannot be changed or disabled due to use in a custom choice list.

Lookup Codes

Action View

* Meaning	* Lookup Code	Enabled	Display Sequence	Description
Pendiente	1	<input checked="" type="checkbox"/>	<input type="text"/>	<input type="text"/>
Aceptada	2	<input checked="" type="checkbox"/>	<input type="text"/>	<input type="text"/>
Rechazada	3	<input checked="" type="checkbox"/>	<input type="text"/>	<input type="text"/>

9.46. Apéndice AT. Configuración de lista de valores Motivo de rechazo

Edit Lookup Type

* Meaning Description

Lookup Type MOTIVO_RECHAZO

One or more lookup codes cannot be changed or disabled due to use in a custom choice list.

Lookup Codes

Action View

* Meaning	* Lookup Code	Enabled	Display Sequence	Description
Cliente no cumpi	1	<input checked="" type="checkbox"/>	<input type="text"/>	<input type="text"/>
Riesgo alto	2	<input checked="" type="checkbox"/>	<input type="text"/>	<input type="text"/>
Otro motivo	3	<input checked="" type="checkbox"/>	<input type="text"/>	<input type="text"/>

9.47. Apéndice AU. Configuración de server script Crear Lead

Edit Object Function Validate Save and Close Cancel

Definition

* Function Name:

* Returns:

Description:

Example:

Privileged

Visibility:

Parameters

Function Body

Edit Script

```
1 def leadView = newView('Lead')
2 def nuevoLead = leadView.createRow();
3 nuevoLead.setAttribute('Name', XUM_Producto_FDS_c);
4 nuevoLead.setAttribute('CustomerId', XUM_ClienteAsociado_FDS_Obj_c?.PartyId);
5 nuevoLead.setAttribute('Description', XUM_Producto_FDS_c);
6 nuevoLead.setAttribute('OwnerId', XUM_Encargado_FDS_Obj_c?.PartyId);
7
```

9.48. Apéndice AV. Configuración de server script Asignar Correo

Edit Object Trigger Save and Close Cancel

Name

Trigger: Description:

* Trigger Name:

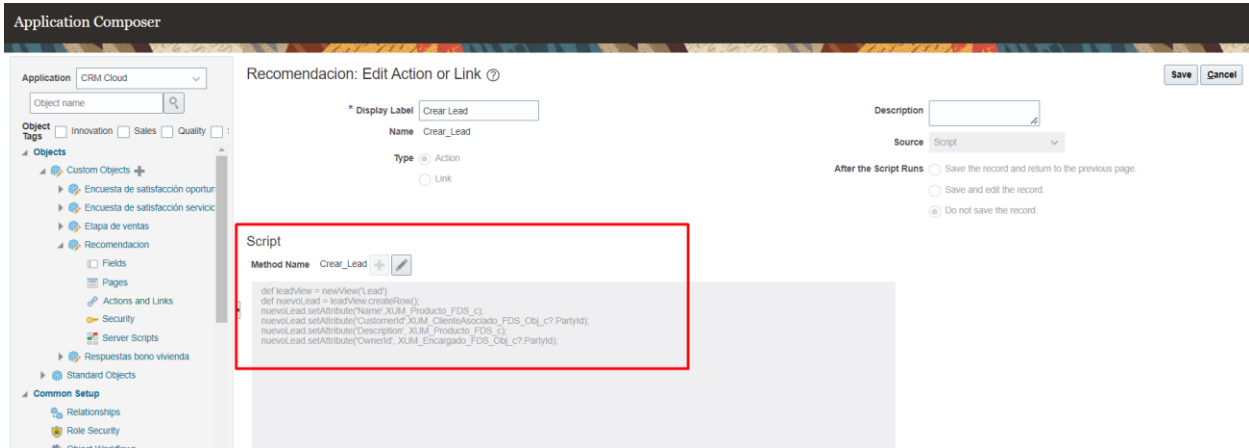
Error Message

* Trigger Definition

Edit Script

```
1 XUM_CorreoCliente_FDS_c=XUM_ClienteAsociado_FDS_Obj_c?.PreferredContactEmailAddress
```

9.49. Apéndice AW. Configuración de *action* Crear Lead



9.50. Apéndice AX. Transformación #1 Funciones creadas para generar las recomendaciones

Obtener mejores 3 recomendaciones

```
def obtener_recomendaciones_01(producto):
    col_idx = collaborative_table.columns.get_loc(producto)
    corr_specific = corr_mat[col_idx]
    recomendaciones=pd.DataFrame({'corr_specific':corr_specific, 'Productos': collaborative_table.columns}).sort_values('corr_spe
    primera=recomendaciones.iloc[1]
    return (primera.tolist())

def obtener_recomendaciones_02(producto):
    col_idx = collaborative_table.columns.get_loc(producto)
    corr_specific = corr_mat[col_idx]
    recomendaciones=pd.DataFrame({'corr_specific':corr_specific, 'Productos': collaborative_table.columns}).sort_values('corr_spe
    segunda=recomendaciones.iloc[2]
    return (segunda.tolist())

def obtener_recomendaciones_03(producto):
    col_idx = collaborative_table.columns.get_loc(producto)
    corr_specific = corr_mat[col_idx]
    recomendaciones=pd.DataFrame({'corr_specific':corr_specific, 'Productos': collaborative_table.columns}).sort_values('corr_spe
    tercera=recomendaciones.iloc[3]
    return (tercera.tolist())
```


9.51. Apéndice AY. Transformación #2 Creación de nuevos campos para almacenar la recomendación

```
copia["Recomendacion01"]=copia.Producto.apply(obtener_recomendaciones_01)
copia["Recomendacion02"]=copia.Producto.apply(obtener_recomendaciones_02)
copia["Recomendacion03"]=copia.Producto.apply(obtener_recomendaciones_03)
copia.head()
```

	Contact ID	Producto	Recomendacion01	Recomendacion02	Recomendacion03
0	100000008033574	Tarjeta de credito-Black	[0.9999969640222637, TARJETA SEGURA-TC-GOLD]	[0.9999947521593926, TARJETA DE CREDITO PLATIN...	[0.9972710903789693, CDP DESMATERIALIZADO COL]
1	300000082105681	CDP Desmaterializado COL	[0.9601691826982962, VIVIENDA MIX]	[0.9456371563062101, AFILIACION PYMES-REGULAR]	[0.889101780461151, TARJETA DE CREDITO CAMALEO...
2	300000082105681	CDP Desmaterializado COL	[0.9601691826982962, VIVIENDA MIX]	[0.9456371563062101, AFILIACION PYMES-REGULAR]	[0.889101780461151, TARJETA DE CREDITO CAMALEO...
3	100000008088151	Credito sobre CDP COL	[0.920661464881403, TARJETA DE CREDITO GOLD-VISA]	[0.8811744422754123, CS MEDICAL]	[0.8799741050734401, TARJETA SEGURA-TC-BLACK]
4	100000008013135	Sin Fiador	[0.946602671068083, SERVAHORRO]	[0.7475236311958727, TARJETA CORPORATIVA GOLD-MC]	[0.7458540481917555, SIN FIADOR PREFERENCIAL P...

9.52. Apéndice AZ. Transformación #3 Creación de nuevos dataframes

Creacion de dataframe con 3 recomendaciones y puntajes

```
recomendaciones_01=copia[["Contact ID","Recomendacion01"]]
recomendaciones_02=copia[["Contact ID","Recomendacion02"]]
recomendaciones_03=copia[["Contact ID","Recomendacion02"]]
```

9.53. Apéndice BA. Transformación #4 Cambio de nombre de columnas

```
recomendaciones_01.columns = ['Contact ID','Recomendacion']
recomendaciones_02.columns = ['Contact ID','Recomendacion']
recomendaciones_03.columns = ['Contact ID','Recomendacion']
```

9.54. Apéndice BB. Transformación #5 Unión de dataframes de recomendaciones

```
recomendaciones_def=recomendaciones_01.append(recomendaciones_02).append(recomendaciones_03)
recomendaciones_def.head()
```

	Contact ID	Recomendacion
0	100000008033574	[0.9999969640222637, TARJETA SEGURA-TC-GOLD]
1	300000082105681	[0.9601691826982962, VIVIENDA MIX]
2	300000082105681	[0.9601691826982962, VIVIENDA MIX]
3	100000008088151	[0.920661464881403, TARJETA DE CREDITO GOLD-VISA]
4	100000008013135	[0.946602671068083, SERVIAHORRO]

9.55. Apéndice BC. Transformación #6 Creación de funciones para la limpieza de datos

```
def obtener_puntaje(lista):
    return lista[0]*100

def obtener_producto(lista):
    return lista[1]

recomendaciones_def["Producto"]=recomendaciones_def.Recomendacion.apply(obtener_producto)
recomendaciones_def["Puntaje"]=recomendaciones_def.Recomendacion.apply(obtener_puntaje)

recomendaciones_def.head()
```

	Contact ID	Recomendacion	Producto	Puntaje
0	100000008033574	[0.9999969640222637, TARJETA SEGURA-TC-GOLD]	TARJETA SEGURA-TC-GOLD	99.999696
1	300000082105681	[0.9601691826982962, VIVIENDA MIX]	VIVIENDA MIX	96.016918
2	300000082105681	[0.9601691826982962, VIVIENDA MIX]	VIVIENDA MIX	96.016918
3	100000008088151	[0.920661464881403, TARJETA DE CREDITO GOLD-VISA]	TARJETA DE CREDITO GOLD-VISA	92.066146
4	100000008013135	[0.946602671068083, SERVIAHORRO]	SERVIAHORRO	94.660267

9.56. Apéndice BD. Transformación #7 Creación de nuevas columnas en dataframe

Creacion de nuevas columnas

```
recomendaciones_def["Recomendacion Name"]='Prueba'
recomendaciones_def["Record Number"]=''
recomendaciones_def
```

9.57. Apéndice BE. Transformación #8 Filtrado del dataframe

Cambio de nombres de columnas

```
recomendaciones_def.rename(columns={'Contact ID':'Cliente', 'Puntaje':'Puntuacion'}, inplace = True)
recomendaciones_def.head()
```

	Cliente	Recomendacion	Producto	Puntuacion	Recomendacion Name	Record Number
0	100000008033574	[0.9999969640222637, TARJETA SEGURA-TC-GOLD]	TARJETA SEGURA-TC-GOLD	99.999696	Prueba	
1	300000082105681	[0.9601691826982962, VIVIENDA MIX]	VIVIENDA MIX	96.016918	Prueba	
2	300000082105681	[0.9601691826982962, VIVIENDA MIX]	VIVIENDA MIX	96.016918	Prueba	
3	100000008088151	[0.920661464881403, TARJETA DE CREDITO GOLD-VISA]	TARJETA DE CREDITO GOLD-VISA	92.066146	Prueba	
4	100000008013135	[0.946602671068083, SERVIAHORRO]	SERVIAHORRO	94.660267	Prueba	

9.58. Apéndice BF. Transformación #9 Exportación del dataframe a formato CSV

Exportación a CSV

```
recomendaciones_export=recomendaciones_def[["Cliente", "Puntuacion", "Producto", "Record Number", "Recomendacion Name"]]
recomendaciones_export.to_csv(r'Recomendaciones.csv', index=False)
```

9.59. Apéndice BG. Entrevista #7 Almacenamiento de datos en Engagement Cloud

Entrevista: Almacenamiento de datos en Engagement Cloud	
Participantes	
Fecha	
Objetivo	Analizar las alternativas de almacenamiento de datos de recomendaciones dentro de Oracle Engagement Cloud

Preguntas	Respuestas
¿Dentro de Engagement Cloud existe alguna manera de almacenar el registro de recomendaciones generadas por el modelo?	
¿Existen formas de importar y exportar los datos, de forma periódica?	
¿Tiene alguna limitante? Espacio disponible, cantidad de solicitudes, compatibilidad.	
¿Los objetos <i>custom</i> tienen algún costo adicional?	
¿La solución de <i>Custom Object</i> tiene alguna restricción para integrarse con otros sistemas?	
¿Han utilizado anteriormente este mecanismo para agregar nuevos datos a CX?	
Otras notas	

9.60. Apéndice BH. Resultados de la Entrevista #7

Entrevista: Conocer herramientas utilizadas previamente	
Participantes	Nathalie Jiménez
Fecha	09/05/2022
Objetivo	Analizar las alternativas de almacenamiento de datos de recomendaciones dentro de Oracle Engagement Cloud
Preguntas	Respuestas
¿Dentro de Engagement Cloud existe alguna manera de almacenar el registro de recomendaciones generadas por el modelo?	Sí, es posible utilizar el <i>Application Composer</i> para la creación de un nuevo objeto <i>custom</i> que almacene todos los campos necesarios para cada recomendación.
¿Existen formas de importar y exportar los datos, de forma periódica?	Sí, es posible importarlos desde el <i>Import Management</i> . Para exportarlos se puede hacer mediante el <i>Export Management</i> .
¿Tiene alguna limitante? Espacio disponible, cantidad de solicitudes, compatibilidad.	Es posible almacenar registros ilimitados dentro de la herramienta.

Entrevista: Conocer herramientas utilizadas previamente	
¿Los objetos <i>custom</i> tienen algún costo adicional?	No implica ningún costo adicional, se encuentra incluido dentro de Engagement Cloud.
¿La solución de <i>Custom Object</i> tiene alguna restricción para integrarse con otros sistemas?	No, es posible integrarlo con los demás sistemas de Oracle como <i>Integration Cloud</i> y <i>Eloqua</i> .
¿Han utilizado anteriormente este mecanismo para agregar nuevos datos a CX?	Sí, este mecanismo ha sido utilizado con distintos propósitos por ejemplo la creación de “Categorías de Servicio”, “Homologación de categorías” y “Aprobaciones” para otros clientes del sector bancario.
Otras notas	
<p>¿Qué tipo de información se puede almacenar dentro de los objetos <i>custom</i>?</p> <ul style="list-style-type: none"> Se pueden almacenar campos de tipo texto, numérico, campos calculados (mediante <i>script</i>), listas de valores y referencias a otros objetos como <i>Contacts</i>, <i>Accounts</i>, <i>Leads</i>, <i>Service Request</i>, entre otros. <p>¿Se pueden asignar acciones a los objetos <i>custom</i>?</p> <ul style="list-style-type: none"> Sí, se pueden crear botones personalizados dentro de Oracle Engagement Cloud utilizando <i>scripts</i> y <i>triggers</i> que permiten utilizar la información del objeto <i>custom</i> para crear un nuevo objeto o ejecutar una acción específica. <p>¿Se pueden modificar las pantallas de un objeto <i>custom</i>?</p> <ul style="list-style-type: none"> Si, se pueden mostrar u ocultar campos, crear validaciones y agregar nuevos campos dentro de las pantallas estándar del objeto dentro de Engagement. 	


9.61. Apéndice BI. Revisión documental de herramientas de almacenamiento Oracle CX Custom Objects

Nombre	Oracle CX – Custom Objects
Objetivo principal	Almacenar la información generada por la solución propuesta, esta información incluye el registro de datos para generar el modelo, las recomendaciones generadas y las recomendaciones rechazadas por el ejecutivo de ventas.
Fuente de información	9.59 Apéndice BG. Entrevista #7 Almacenamiento de datos en Engagement Cloud
Posición en Cuadrante de Gartner	No aplica
Calificación general en Peer Insights	No aplica
Principales características	<ul style="list-style-type: none"> Es posible configurarlo desde Oracle Engagement Cloud, además los objetos pueden relacionarse con otros previamente creados o estándar como <i>Contacts</i> o <i>Accounts</i>.

	<ul style="list-style-type: none"> • Permite el almacenamiento de objetos <i>custom</i>, además es posible configurar los campos y pantallas asociados a este objeto. • De manera nativa es posible integrarlo con herramientas adicionales de mercadeo como Oracle Eloqua. • Ofrece la posibilidad de crear mapeos de importación dentro de Oracle Engagement Cloud para cargar los datos mediante archivos de CSV. • Los datos almacenados pueden ser exportados por medio del Export Management y consultados por sistemas de terceros en cualquier momento.
Costo aproximado	No implica costos adicionales.
Ventajas	<ul style="list-style-type: none"> • Se encuentra totalmente integrado con Oracle Engagament Cloud, los campos <i>custom</i> representa una nueva tabla dentro de la base de datos del CRM. • Pertenece al ecosistema de soluciones Oracle, esto garantiza que se puede usar en conjunto con herramientas como Integration Cloud y Eloqua para la gestión de campañas de mercadeo. • Brinda la posibilidad de configurar pantallas dentro de la herramienta para mostrarle al usuario únicamente el contenido que requiere visualizar.
Desventajas	No se identificaron desventajas.

Propuesta de solución de minería de datos basada en la metodología CRISP-DM utilizando los datos almacenados en Oracle Engagement Cloud, para la recomendación personalizada de productos a clientes, durante el primer semestre de 2022

9.62. Apéndice BJ. Minuta 001. Reunión inicial

 MINUTA DE REUNIÓN			
Proyecto: Propuesta de solución de minería de datos basada en la metodología CRISP-DM utilizando los datos almacenados en Oracle Engagement Cloud, para la recomendación personalizada de productos a clientes, durante el primer semestre de 2022			
Reunión No.	1	Fecha:	14/02/2022
Lugar:	Zoom	Hora Inicio/Finalización:	05:30 pm. / 06:30 pm
Objetivo de la reunión:	Explicación inicial del proceso TFG con la profesora tutora María José.		
Participantes:	Presentes: <ul style="list-style-type: none"> • María José Artavia • Emanuel Alvarado • Josué Solís • Maribel Cordero • Jose Carlos Chaves 		
	Ausentes:		
Temas Tratados			
No.	Asunto	Comentarios	Acuerdos
1	Metodología de trabajo	Elegir entre cascada y enfocada en objetivos.	Definir metodología de trabajo para proyecto de TFG y posible cronograma para 20/02/2022. Responsable: Emanuel Alvarado
2	Marco Conceptual	Comenzar la búsqueda de conceptos relacionados al proyecto.	Montar un borrador con la lista de conceptos que podría incluir para el 20/02/2022. Responsable: Emanuel Alvarado
3	Revisión del Anteproyecto	Retroalimentación del capítulo 1.	Enviar anteproyecto con retroalimentación el 16/02/2022. Responsable: María José Artavia
4	Corregir Anteproyecto	Luego de recibir retroalimentación profesora tutora,	Aplicar ajustes y correcciones en el documento para el 20/02/2022. Responsable: Emanuel Alvarado

Propuesta de solución de minería de datos basada en la metodología CRISP-DM utilizando los datos almacenados en Oracle Engagement Cloud, para la recomendación personalizada de productos a clientes, durante el primer semestre de 2022




MINUTA DE REUNIÓN

Proyecto: Propuesta de solución de minería de datos basada en la metodología CRISP-DM utilizando los datos almacenados en Oracle Engagement Cloud, para la recomendación personalizada de productos a clientes, durante el primer semestre de 2022

5	Agendar Reunión entre partes	Debe estar tutora, contraparte y estudiantes.	Establecer fecha y hora. Responsable: Emanuel Alvarado, Jose Carlos Chaves, Josué Solís.
Próxima reunión			
Temas a tratar		Fecha	Convocados
Reunión entre partes interesadas TFG		21/02/2022	<ul style="list-style-type: none"> Alex Ureña María José Artavia Emanuel Alvarado Josué Solís Jose Carlos Chaves

9.63. Apéndice BK. Minuta 002. Sesión inicial de TFG con la organización




MINUTA DE REUNIÓN

Proyecto: Propuesta de solución de minería de datos basada en la metodología CRISP-DM utilizando los datos almacenados en Oracle Engagement Cloud, para la recomendación personalizada de productos a clientes, durante el primer semestre de 2022


Reunión No.	2	Fecha:	21/02/2022
Lugar:	Teams	Hora Inicio/Finalización:	02:30 pm. / 02:45 pm
Objetivo de la reunión:	Explicar el proceso de TFG a todos los involucrados del proyecto: contraparte de la empresa, tutora y estudiantes.		
Participantes:	Presentes: <ul style="list-style-type: none"> • Alex Ureña Cordero • María José Artavia • Emanuel Alvarado • Josué Solís • Jose Carlos Chaves 		
	Ausentes:		
Temas Tratados			
No.	Asunto	Comentarios	Acuerdos
1	Funcionamiento de TFG	Se explicaron las reglas del proceso y las fechas importantes de evaluación /entrega de informes.	El estudiante es responsable de agendar las próximas reuniones entre los involucrados. Responsable: Emanuel Alvarado
Próxima reunión			
Temas a tratar		Fecha	Convocados
Reunión entre partes interesadas TFG #2		POR DEFINIR	<ul style="list-style-type: none"> • Alex Ureña • María José Artavia • Emanuel Alvarado • Josué Solís • Jose Carlos Chaves

9.64. Apéndice BL. Minuta 003. Sesión semanal de seguimiento


 MINUTA DE REUNIÓN Proyecto: Propuesta de solución de minería de datos basada en la metodología CRISP-DM utilizando los datos almacenados en Oracle Engagement Cloud, para la recomendación personalizada de productos a clientes, durante el primer semestre de 2022			
Reunión No.	3	Fecha:	21/02/2022
Lugar:	Teams	Hora Inicio/Finalización:	05:30 pm. / 06:00 pm
Objetivo de la reunión:	Primera sesión semanal con la tutora de TFG, se revisó la retroalimentación del anteproyecto, el cronograma propuesto y las dudas existentes al momento de la reunión.		
Participantes:	Presentes: <ul style="list-style-type: none"> • María José Artavia • Emanuel Alvarado 		
	Ausentes:		
Temas Tratados			
No.	Asunto	Comentarios	Acuerdos
1	Retroalimentación de anteproyecto	La profesora comentó los principales hallazgos y retroalimentación del anteproyecto presentado.	<ul style="list-style-type: none"> • Cambiar verbo del objetivo específico #2 a uno más similar a Analizar según la taxonomía de Bloom. • Cambiar verbo del objetivo específico #3 a uno similar a Evaluar (que incluya el desarrollo de los modelos). • Modificar alcance, agregar detalle del objetivo específico #4(costo-beneficio). • Hacer tabla Objetivos – Fases del alcance. • Aplicar los demás ajustes sugeridos. Responsable: Emanuel Alvarado
	Objetivo Especifico #3 Desarrollo de Modelos	Revisión del alcance y entregables relacionados al objetivo.	Evaluar la posibilidad de agregar una propuesta de Plan de Implementación. Por el momento queda tentativo de incluir la simulación de los modelos con datos para la recomendación de productos, según avance el proyecto se evaluará si se realiza o no. Responsable: Emanuel Alvarado

Uso Confidencial

Pág. 1/2


 MINUTA DE REUNIÓN Proyecto: Propuesta de solución de minería de datos basada en la metodología CRISP-DM utilizando los datos almacenados en Oracle Engagement Cloud, para la recomendación personalizada de productos a clientes, durante el primer semestre de 2022		
Próxima reunión		
Temas a tratar	Fecha	Convocados
Reunión entre partes interesadas TFG #2	POR DEFINIR	<ul style="list-style-type: none"> • María José Artavia • Emanuel Alvarado

9.65. Apéndice BM. Minuta 004. Sesión acuerdos iniciales con la organización

 MINUTA DE REUNIÓN			
Proyecto: Propuesta de solución de minería de datos basada en la metodología CRISP-DM utilizando los datos almacenados en Oracle Engagement Cloud, para la recomendación personalizada de productos a clientes, durante el primer semestre de 2022			
Reunión No.	4	Fecha:	21/02/2022
Lugar:	Teams	Hora Inicio/Finalización:	05:00 pm. / 05:30 pm
Objetivo de la reunión:	Reunión de acuerdos iniciales entre contraparte y estudiante, para definir temas de sesiones, entrevistas, documentos y ambientes que podrán ser utilizados para el desarrollo del proyecto.		
Participantes:	Presentes: <ul style="list-style-type: none"> Alex Ureña Cordero Emanuel Alvarado 		
	Ausentes:		
Temas Tratados			
No.	Asunto	Comentarios	Acuerdos
1	Selección de Disponible	Se consultó por cuál ambiente podría ser utilizado para el desarrollo del proyecto.	<ul style="list-style-type: none"> Se acordó utilizar un ambiente de pruebas disponible del cliente llamado Coopeservidores. Responsable: Emanuel Alvarado
Próxima reunión			
Temas a tratar	Fecha	Convocados	
POR DEFINIR	POR DEFINIR	POR DEFINIR	

Propuesta de solución de minería de datos basada en la metodología CRISP-DM utilizando los datos almacenados en Oracle Engagement Cloud, para la recomendación personalizada de productos a clientes, durante el primer semestre de 2022

9.66. Apéndice BN. Minuta 005. Sesión acuerdos iniciales con la organización




MINUTA DE REUNIÓN

Proyecto: Propuesta de solución de minería de datos basada en la metodología CRISP-DM utilizando los datos almacenados en Oracle Engagement Cloud, para la recomendación personalizada de productos a clientes, durante el primer semestre de 2022


Reunión No.	5	Fecha:	28/02/2022
Lugar:	Teams	Hora Inicio/Finalización:	05:30 pm. / 06:00 pm
Objetivo de la reunión:	Segunda sesión de seguimiento con entre la profesora tutora y el estudiante, se revisaron consultas del capítulo #1 y el marco conceptual.		
Participantes:	Presentes:		
	<ul style="list-style-type: none"> • María José Artavia Jimenez • Emanuel Alvarado Pérez 		
	Ausentes:		
Temas Tratados			
No.	Asunto	Comentarios	Acuerdos
1	Revisión de Objetivos	Se consultó sobre el cambio en la redacción del objetivo 1 y 2.	Cambiar verbos, en objetivo #1 utilizar "Analizar" y en objetivo #2 utilizar "Recomendar". Responsable: Emanuel Alvarado
2	Consultas de distintos temas	Preguntas sobre fuentes de bibliografía confiable.	Utilizar fuentes de Bases de Datos suscritas del TEC, revistas, web, pero con menos de 10 años de antigüedad. Responsable: Emanuel Alvarado
Próxima reunión			
Temas a tratar	Fecha	Convocados	
Sesión de seguimiento con tutora #3	07/03/2022	<ul style="list-style-type: none"> • María José Artavia Jimenez • Emanuel Alvarado Pérez 	

9.67. Apéndice BO. Minuta 006. Sesión semanal de seguimiento


 MINUTA DE REUNIÓN			
Proyecto: Propuesta de solución de minería de datos basada en la metodología CRISP-DM utilizando los datos almacenados en Oracle Engagement Cloud, para la recomendación personalizada de productos a clientes, durante el primer semestre de 2022			
Reunión No.	6	Fecha:	07/03/2022
Lugar:	Teams	Hora Inicio/Finalización:	05:30 pm. / 06:00 pm
Objetivo de la reunión:	Tercera sesión de seguimiento entre la profesora tutora y el estudiante, se revisaron avances del capítulo #1, pendientes del marco metodológico y otras consultas.		
Participantes:	Presentes: <ul style="list-style-type: none"> • María José Artavia Jimenez • Emanuel Alvarado Pérez 		
	Ausentes:		
Temas Tratados			
No.	Asunto	Comentarios	Acuerdos
1	Consultas de distintos temas	Preguntas sobre avances del TFG.	Completar borrador guía de marco metodológico. Responsable: Emanuel Alvarado
2	Revisión del capítulo #1	La profesora revisará el avance del capítulo #1	Cuando las correcciones se realicen se informará al estudiante. Responsable: María José Artavia.
Próxima reunión			
Temas a tratar	Fecha	Convocados	
Sesión de seguimiento con tutora #4	14/03/2022	<ul style="list-style-type: none"> • María José Artavia Jimenez • Emanuel Alvarado Pérez 	

Propuesta de solución de minería de datos basada en la metodología CRISP-DM utilizando los datos almacenados en Oracle Engagement Cloud, para la recomendación personalizada de productos a clientes, durante el primer semestre de 2022

9.68. Apéndice BP. Minuta 007. Sesión semanal de seguimiento

 MINUTA DE REUNIÓN			
Proyecto: Propuesta de solución de minería de datos basada en la metodología CRISP-DM utilizando los datos almacenados en Oracle Engagement Cloud, para la recomendación personalizada de productos a clientes, durante el primer semestre de 2022			
Reunión No.	7	Fecha:	14/03/2022
Lugar:	Teams	Hora Inicio/Finalización:	05:30 pm. / 06:00 pm
Objetivo de la reunión:	Cuarta sesión de seguimiento entre la profesora tutora y el estudiante, se revisaron avances del trabajo final de graduación, principalmente dudas del capítulo #3 marco metodológico.		
Participantes:	Presentes: <ul style="list-style-type: none"> • Maria José Artavia Jimenez • Emanuel Alvarado Pérez 		
	Ausentes:		
Temas Tratados			
No.	Asunto	Comentarios	Acuerdos
1	Consultas de distintos temas	Preguntas sobre correcciones del capítulo #1.	Entrega de retroalimentación se realizará en el transcurso de la semana. Responsable: Maria José Artavia.
2	Revisión del avances y consultas de Marco Metodológico	La profesora recomienda hacer cambios en metodología seleccionada.	Utilizar una metodología cualitativa en el proyecto, poner en el documento que se utilizará metodología cualitativa con ciertos análisis cuantitativos. Responsable: Maria José Artavia, Emanuel Alvarado.
Próxima reunión			
Temas a tratar		Fecha	Convocados
Sesión de seguimiento con tutora #5		21/03/2022	<ul style="list-style-type: none"> • Maria José Artavia Jimenez • Emanuel Alvarado Pérez

9.69. Apéndice BQ. Minuta 008. Sesión semanal de seguimiento


 MINUTA DE REUNIÓN			
Proyecto: Propuesta de solución de minería de datos basada en la metodología CRISP-DM utilizando los datos almacenados en Oracle Engagement Cloud, para la recomendación personalizada de productos a clientes, durante el primer semestre de 2022			
Reunión No.	8	Fecha:	21/03/2022
Lugar:	Teams	Hora Inicio/Finalización:	05:30 pm. / 06:00 pm
Objetivo de la reunión:	Sesión de seguimiento entre la profesora tutora y el estudiante, se revisaron avances del trabajo final de graduación, principalmente dudas del capítulo #3 marco metodológico.		
Participantes:	Presentes: <ul style="list-style-type: none"> • Maria José Artavia Jimenez • Emanuel Alvarado Pérez 		
	Ausentes:		
Temas Tratados			
No.	Asunto	Comentarios	Acuerdos
1	Consultas de distintos temas	Preguntas sobre marco metodológico.	Responsable: Maria José Artavia, Emanuel Alvarado.
2	Revisión de capítulo #2	Retroalimentación será brindada durante la semana.	Entrega de retroalimentación se realizará en el transcurso de la semana. Responsable: Maria José Artavia.
Próxima reunión			
Temas a tratar	Fecha	Convocados	
Sesión de seguimiento con tutora #6	28/03/2022	<ul style="list-style-type: none"> • Maria José Artavia Jimenez • Emanuel Alvarado Pérez 	

Propuesta de solución de minería de datos basada en la metodología CRISP-DM utilizando los datos almacenados en Oracle Engagement Cloud, para la recomendación personalizada de productos a clientes, durante el primer semestre de 2022


9.70. Apéndice BR. Minuta 009. Sesión semanal de seguimiento

 MINUTA DE REUNIÓN			
Proyecto: Propuesta de solución de minería de datos basada en la metodología CRISP-DM utilizando los datos almacenados en Oracle Engagement Cloud, para la recomendación personalizada de productos a clientes, durante el primer semestre de 2022			
Reunión No.	9	Fecha:	28/03/2022
Lugar:	Teams	Hora Inicio/Finalización:	05:30 pm. / 06:00 pm
Objetivo de la reunión:	Sesión de seguimiento entre la profesora tutora y el estudiante, se revisaron avances del trabajo final de graduación, principalmente dudas del capítulo #3 marco metodológico.		
Participantes:	Presentes: <ul style="list-style-type: none"> • Maria José Artavia Jimenez • Emanuel Alvarado Pérez 		
	Ausentes:		
Temas Tratados			
No.	Asunto	Comentarios	Acuerdos
1	Consultas de distintos temas	Preguntas sobre marco metodológico.	Responsable: Maria José Artavia, Emanuel Alvarado.
2	Reunión de seguimiento entre contrapartes	Se propone hacer la reunión la próxima semana.	Agendar sesión con Álex para revisar avance del proyecto. Responsable: Emanuel Alvarado.
Próxima reunión			
Temas a tratar		Fecha	Convocados
Sesión de seguimiento con tutora		04/03/2022	<ul style="list-style-type: none"> • Maria José Artavia Jimenez • Emanuel Alvarado Pérez

9.71. Apéndice BS. Minuta 010. Sesión semanal de seguimiento


 MINUTA DE REUNIÓN			
Proyecto: Propuesta de solución de minería de datos basada en la metodología CRISP-DM utilizando los datos almacenados en Oracle Engagement Cloud, para la recomendación personalizada de productos a clientes, durante el primer semestre de 2022			
Reunión No.	10	Fecha:	04/04/2022
Lugar:	Teams	Hora Inicio/Finalización:	05:30 pm. / 06:00 pm
Objetivo de la reunión:	Sesión de seguimiento entre la profesora tutora y el estudiante, se revisaron avances del trabajo final de graduación, principalmente dudas del capítulo #3 metodológico.		
Participantes:	Presentes: <ul style="list-style-type: none"> • Maria José Artavia Jimenez • Emanuel Alvarado Pérez 		
	Ausentes:		
Temas Tratados			
No.	Asunto	Comentarios	Acuerdos
1	Consultas sobre entrega del Capítulo #3	Consultas del avance del proyecto y correcciones.	Responsable: Maria José Artavia, Emanuel Alvarado.
2	Cambio en los objetivos del proyecto	Se consulta sobre la posibilidad de cambiar el orden de los objetivos #2 y #3 y añadir una exclusión a proyecto.	Se da autorización para hacer el cambio en el orden de los objetivos. Adicionalmente la profesora tutora brinda la recomendación de añadir una exclusión al proyecto, donde se indique que el modelo de minería de datos se desarrollará en las herramientas disponibles para el investigador, no en las sugeridas. De acuerdo con la profesora, no es necesario hacer un control de cambios. Responsable: Emanuel Alvarado, Maria José Artavia.
3	Consultas sobre el capítulo #4 y capítulo #5	Se consulta sobre los entregables esperados para cada capítulo.	En el capítulo #4 se deben agregar los resultados de aplicar cada uno de los instrumentos definidos en la metodología. En el capítulo #5 se debe agregar la información relacionada directamente con la propuesta. Responsable: Emanuel Alvarado
Próxima reunión			

Propuesta de solución de minería de datos basada en la metodología CRISP-DM utilizando los datos almacenados en Oracle Engagement Cloud, para la recomendación personalizada de productos a clientes, durante el primer semestre de 2022

 **MINUTA DE REUNIÓN**
 Proyecto: Propuesta de solución de minería de datos basada en la metodología CRISP-DM utilizando los datos almacenados en Oracle Engagement Cloud, para la recomendación personalizada de productos a clientes, durante el primer semestre de 2022


Temas a tratar	Fecha	Convocados
Reunión de contrapartes	08/04/2022	<ul style="list-style-type: none"> • María José Artavia Jimenez • Emanuel Alvarado Pérez • Josué Solís Blanco • José Carlos Chaves • Álex Ureña Cordero
Sesión de seguimiento con tutora	18/04/2022	<ul style="list-style-type: none"> • Maria José Artavia Jimenez • Emanuel Alvarado Pérez

9.72. Apéndice BT. Minuta 011. Segunda sesión de seguimiento de TFG con la organización


 **MINUTA DE REUNIÓN**
 Proyecto: Propuesta de solución de minería de datos basada en la metodología CRISP-DM utilizando los datos almacenados en Oracle Engagement Cloud, para la recomendación personalizada de productos a clientes, durante el primer semestre de 2022

Reunión No.	11	Fecha:	08/04/2022
Lugar:	Teams	Hora Inicio/Finalización:	01:00 pm / 01:30 pm
Objetivo de la reunión:	Sesión de seguimiento entre la profesora tutora, contraparte de la empresa y el estudiante. Se consultó a Álex sobre la percepción del avance del proyecto hasta el momento.		
Participantes:	Presentes:		
	<ul style="list-style-type: none"> • María José Artavia Jimenez • Emanuel Alvarado Pérez • Josué Solís Blanco • José Carlos Chaves • Álex Ureña Cordero 		
Ausentes:			
Temas Tratados			
No.	Asunto	Comentarios	Acuerdos
1	Seguimiento del estado actual de TFG	Álex comenta que le gustaría reunirse para ver el estado del proyecto.	Agendar una sesión luego de semana santa para informar sobre el avance del proyecto. Responsable: Emanuel Alvarado, Álex Ureña Cordero.
Próxima reunión			
Temas a tratar		Fecha	Convocados
Reunión de seguimiento con Álex		21/04/2022	<ul style="list-style-type: none"> • Emanuel Alvarado Pérez • Álex Ureña Cordero
Sesión de seguimiento con tutora		18/04/2022	<ul style="list-style-type: none"> • Maria José Artavia Jimenez • Emanuel Alvarado Pérez

9.73. Apéndice BJ. Minuta 012. Sesión semanal de seguimiento

 MINUTA DE REUNIÓN			
Proyecto: Propuesta de solución de minería de datos basada en la metodología CRISP-DM utilizando los datos almacenados en Oracle Engagement Cloud, para la recomendación personalizada de productos a clientes, durante el primer semestre de 2022			
Reunión No.	12	Fecha:	18/04/2022
Lugar:	Teams	Hora Inicio/Finalización:	05:30 pm. / 06:00 pm
Objetivo de la reunión:	Sesión de seguimiento entre la profesora tutora y el estudiante. Se consultó sobre el avance del proyecto, también se revisaron consultas acerca de las correcciones pendientes del capítulo #3.		
Participantes:	Presentes: <ul style="list-style-type: none"> • María José Artavia Jimenez • Emanuel Alvarado Pérez 		
	Ausentes:		
Temas Tratados			
No.	Asunto	Comentarios	Acuerdos
1	Instrumentos de recolección de datos	Se consulta sobre la posibilidad de modificar o quitar instrumentos.	Los instrumentos pueden ser cambiados por el investigador según las necesidades del proyecto. Responsable: Emanuel Alvarado
2	Entrevistas del TFG	Se comenta sobre las entrevistas a realizar	Las entrevistas serán aplicadas en el transcurso de la semana (18/04-22/04). Responsable: Emanuel Alvarado
Próxima reunión			
Temas a tratar	Fecha	Convocados	
Reunión de seguimiento con Álex	21/04/2022	<ul style="list-style-type: none"> • Emanuel Alvarado Pérez • Álex Ureña Cordero 	
Sesión de seguimiento con tutora	25/04/2022	<ul style="list-style-type: none"> • María José Artavia Jimenez • Emanuel Alvarado Pérez 	

9.74. Apéndice BU. Minuta 013. Sesión semanal de seguimiento

 MINUTA DE REUNIÓN			
Proyecto: Propuesta de solución de minería de datos basada en la metodología CRISP-DM utilizando los datos almacenados en Oracle Engagement Cloud, para la recomendación personalizada de productos a clientes, durante el primer semestre de 2022			
Reunión No.	13	Fecha:	21/04/2022
Lugar:	Teams	Hora Inicio/Finalización:	04:00 pm. / 04:30 pm
Objetivo de la reunión:	Sesión de seguimiento entre el estudiante y la contraparte de la empresa.		
Participantes:	Presentes: <ul style="list-style-type: none"> Alex Ureña Cordero Emanuel Alvarado Pérez 		
	Ausentes:		
Temas Tratados			
No.	Asunto	Comentarios	Acuerdos
1	Revisión del proceso AS IS y TO BE	Se validan los procesos de manera conjunta.	Ajustar proceso TO BE. Separar la creación de campañas, y reordenarlas antes de la creación del Lead. Responsable: Emanuel Alvarado
2	Revisión de los requerimientos del proyecto	Se revisan los requerimientos especificados por el investigador.	Se debe añadir un requerimiento adicional sobre a solución. Un dato importante para considerar dentro del modelo es el uso del registro de oportunidades ganadas. Responsable: Emanuel Alvarado
Próxima reunión			
Temas a tratar		Fecha	Convocados
Sesión de seguimiento con tutora		25/04/2022	<ul style="list-style-type: none"> Maria José Artavia Jimenez Emanuel Alvarado Pérez

9.75. Apéndice BV. Minuta 014. Sesión semanal de seguimiento




MINUTA DE REUNIÓN

Proyecto: Propuesta de solución de minería de datos basada en la metodología CRISP-DM utilizando los datos almacenados en Oracle Engagement Cloud, para la recomendación personalizada de productos a clientes, durante el primer semestre de 2022

Reunión No.	14	Fecha:	25/04/2022
Lugar:	Teams	Hora Inicio/Finalización:	04:00 pm. / 04:30 pm
Objetivo de la reunión:	Sesión de seguimiento entre el estudiante y la profesora tutora.		
Participantes:	Presentes: <ul style="list-style-type: none"> María José Artavia Jiménez Emanuel Alvarado Pérez 		
	Ausentes:		
Temas Tratados			
No.	Asunto	Comentarios	Acuerdos
1	Revisión del avance capítulo 4 y 5	Se revisa el estado actual del TFG.	Se brinda una semana adicional respecto al cronograma inicial del proyecto. Responsable: Emanuel Alvarado/Maria Jose Artavia
Próxima reunión			
Temas a tratar		Fecha	Convocados
Sesión de seguimiento con tutora		02/05/2022	<ul style="list-style-type: none"> María José Artavia Jiménez Emanuel Alvarado Pérez

Propuesta de solución de minería de datos basada en la metodología CRISP-DM utilizando los datos almacenados en Oracle Engagement Cloud, para la recomendación personalizada de productos a clientes, durante el primer semestre de 2022

9.76. Apéndice BW. Minuta 015. Sesión de seguimiento con la coordinación de TFG


 MINUTA DE REUNIÓN			
Proyecto: Propuesta de solución de minería de datos basada en la metodología CRISP-DM utilizando los datos almacenados en Oracle Engagement Cloud, para la recomendación personalizada de productos a clientes, durante el primer semestre de 2022			
Reunión No.	15	Fecha:	02/05/2022
Lugar:	Teams	Hora Inicio/Finalización:	04:00 pm. / 04:30 pm
Objetivo de la reunión:	Sesión de seguimiento entre el estudiante, a profesora tutora y a coordinación de TFG.		
Participantes:	Presentes: <ul style="list-style-type: none"> • María José Artavia Jiménez • Josué Solís Blanco • Emanuel Alvarado Pérez • Yarima Sandoval • Jose Carlos Chaves • Maribel Cordero 		
	Ausentes:		
Temas Tratados			
No.	Asunto	Comentarios	Acuerdos
1	Revisión del estado actual de TFG	Se revisa uno a uno los proyectos para analizar la posibilidad de prórroga.	De acuerdo con el avance presentado hasta el momento no es necesario solicitar prórroga. Responsable: Emanuel Alvarado
Próxima reunión			
Temas a tratar		Fecha	Convocados
Sesión de seguimiento con tutora		09/05/2022	<ul style="list-style-type: none"> • María José Artavia Jiménez • Emanuel Alvarado Pérez

Propuesta de solución de minería de datos basada en la metodología CRISP-DM utilizando los datos almacenados en Oracle Engagement Cloud, para la recomendación personalizada de productos a clientes, durante el primer semestre de 2022

9.77. Apéndice BX. Minuta 016. Sesión semanal de seguimiento

 MINUTA DE REUNIÓN			
Proyecto: Propuesta de solución de minería de datos basada en la metodología CRISP-DM utilizando los datos almacenados en Oracle Engagement Cloud, para la recomendación personalizada de productos a clientes, durante el primer semestre de 2022			
Reunión No.	16	Fecha:	09/05/2022
Lugar:	Teams	Hora Inicio/Finalización:	04:00 pm. / 04:30 pm
Objetivo de la reunión:	Sesión de seguimiento entre el estudiante y la profesora tutora.		
Participantes:	Presentes: <ul style="list-style-type: none"> • María José Artavia Jiménez • Emanuel Alvarado Pérez 		
	Ausentes:		
Temas Tratados			
No.	Asunto	Comentarios	Acuerdos
1	Revisión de puntos pendientes capítulo 5	Se informa sobre el estado actual del capítulo	Entregar un avance la próxima semana con los puntos pendientes de capítulo 4, 5, 6 y 7. Responsable: Emanuel Alvarado
Próxima reunión			
Temas a tratar		Fecha	Convocados
Sesión de seguimiento con tutora		16/05/2022	<ul style="list-style-type: none"> • María José Artavia Jiménez • Emanuel Alvarado Pérez

9.78. Apéndice BY. Minuta 017. Sesión semanal de seguimiento


 MINUTA DE REUNIÓN			
Proyecto: Propuesta de solución de minería de datos basada en la metodología CRISP-DM utilizando los datos almacenados en Oracle Engagement Cloud, para la recomendación personalizada de productos a clientes, durante el primer semestre de 2022			
Reunión No.	17	Fecha:	17/05/2022
Lugar:	Teams	Hora Inicio/Finalización:	12:00 pm. / 12:30 pm
Objetivo de la reunión:	Sesión de seguimiento entre el estudiante y la profesora tutora.		
Participantes:	Presentes: <ul style="list-style-type: none"> • María José Artavia Jiménez • Emanuel Alvarado Pérez 		
	Ausentes:		
Temas Tratados			
No.	Asunto	Comentarios	Acuerdos
1	Revisión de dudas sobre el análisis financiero	Se consulta sobre la posibilidad de utilizar suposiciones en el cálculo del ROI.	Se puede utilizar la suposición de que al menos cuatro clientes han mostrado interés en implementar esta solución. Responsable: Emanuel Alvarado
2	Revisión de dudas sobre la propuesta de solución	Se consulta sobre la posibilidad de limitar el alcance de la propuesta de solución	Se descarta la implementación y configuración de la campaña de mercadeo en Eloqua. Responsable: Emanuel Alvarado
Próxima reunión			
Temas a tratar	Fecha	Convocados	
Sesión de seguimiento con tutora	24/05/2022	<ul style="list-style-type: none"> • María José Artavia Jiménez • Emanuel Alvarado Pérez 	
Sesión de seguimiento con contraparte	23/05/2022	<ul style="list-style-type: none"> • Alex Ureña Cordero • Emanuel Alvarado Pérez 	

Propuesta de solución de minería de datos basada en la metodología CRISP-DM utilizando los datos almacenados en Oracle Engagement Cloud, para la recomendación personalizada de productos a clientes, durante el primer semestre de 2022

9.79. Apéndice BZ. Minuta 018. Presentación de la propuesta a la organización

 MINUTA DE REUNIÓN			
Proyecto: Propuesta de solución de minería de datos basada en la metodología CRISP-DM utilizando los datos almacenados en Oracle Engagement Cloud, para la recomendación personalizada de productos a clientes, durante el primer semestre de 2022			
Reunión No.	18	Fecha:	23/05/2022
Lugar:	Teams	Hora Inicio/Finalización:	11:00 am. / 11:30 am
Objetivo de la reunión:	Sesión de seguimiento entre el estudiante y la contraparte de la empresa.		
Participantes:	Presentes: <ul style="list-style-type: none"> Alex Ureña Cordero Emanuel Alvarado Pérez 		
	Ausentes:		
Temas Tratados			
No.	Asunto	Comentarios	Acuerdos
1	Revisión completa de la propuesta de solución	Se valida la propuesta de solución.	Se brinda el visto bueno a la propuesta por parte de la contraparte de la empresa. Responsable: Alex Ureña Cordero.
Próxima reunión			
Temas a tratar		Fecha	Convocados
Sesión de seguimiento con tutora		24/05/2022	<ul style="list-style-type: none"> Maria José Artavia Jiménez Emanuel Alvarado Pérez

9.80. Apéndice CA. Minuta 019. Sesión de cierre con la organización

 MINUTA DE REUNIÓN Proyecto: Propuesta de solución de minería de datos basada en la metodología CRISP-DM utilizando los datos almacenados en Oracle Engagement Cloud, para la recomendación personalizada de productos a clientes, durante el primer semestre de 2022			
Reunión No.	19	Fecha:	23/05/2022
Lugar:	Teams	Hora Inicio/Finalización:	03:30 pm. / 04:00 pm
Objetivo de la reunión:	Sesión de seguimiento entre el estudiante, a profesora tutora y la contraparte de la empresa.		
Participantes:	Presentes: <ul style="list-style-type: none"> • Alex Ureña Cordero • María José Artavia Jiménez • Josué Solís Blanco • Emanuel Alvarado Pérez 		
	Ausentes:		
Temas Tratados			
No.	Asunto	Comentarios	Acuerdos
1	Reunión final de TFG	Se valida el trabajo realizado en la empresa.	Se brinda el visto bueno por parte de la contraparte de la empresa. Completar la valoración final del TFG realizado. Responsable: Alex Ureña Cordero.
Próxima reunión			
Temas a tratar		Fecha	Convocados
Sesión de seguimiento con tutora		24/05/2022	<ul style="list-style-type: none"> • María José Artavia Jiménez • Emanuel Alvarado Pérez

Propuesta de solución de minería de datos basada en la metodología CRISP-DM utilizando los datos almacenados en Oracle Engagement Cloud, para la recomendación personalizada de productos a clientes, durante el primer semestre de 2022

9.81. Apéndice CB. Cálculo de cargas patronales de la investigación

Rubro	Porcentaje (%)	Investigador
Salario		\$1,000
Aguinaldo	8.33	\$83
Vacaciones	5.76	\$58
CCSS SEM	9.25	\$93
CCSS IVM	5.25	\$53
Cuota Patronal Banco Popular	0.25	\$3
Asignaciones Familiares	5	\$50
IMAS	0.5	\$5
INA	1.5	\$15
Aporte Patrono Banco Popular	0.25	\$3
FCL	1.5	\$15
Fondo de pensiones complementario	2	\$20
INS	1	\$10
Total por hora		\$1,406

9.82. Apéndice CC. Cálculo de cargas patronales de la cotización

Rubro	Porcentaje (%)	Rol		
		Project Manager	Consultor de Engagement	Consultor de Analítica Empresarial
Pago por hora		60	60	60
Aguinaldo	8.33	4.998	4.998	4.998
Vacaciones	5.76	3.456	3.456	3.456
CCSS SEM	9.25	5.55	5.55	5.55
CCSS IVM	5.25	3.15	3.15	3.15
Cuota Patronal Banco Popular	0.25	0.15	0.15	0.15
Asignaciones Familiares	5	3	3	3
IMAS	0.5	0.3	0.3	0.3
INA	1.5	0.9	0.9	0.9
Aporte Patrono Banco Popular	0.25	0.15	0.15	0.15
FCL	1.5	0.9	0.9	0.9
Fondo de pensiones complementario	2	1.2	1.2	1.2
INS	1	0.6	0.6	0.6
Total por hora		\$ 84.35	\$ 84.35	\$ 84.35

10. Anexos

10.1. Anexo I. Historias de usuario

Historia de Usuario HU-01-Recomendaciones	
Yo Como:	Ejecutivo de Banca Personas Ejecutivo de Banca Empresas Ejecutivo de Sucursal (Plataforma y Cajas)
Requiero:	<p>Visualizar e interactuar con registros de recomendaciones de productos para clientes, tanto para venta cruzada (colocación de nuevos productos) como para venta aumentada (incremento del valor de un producto existente). Esta información debe facilitarse a nivel de Oracle CX en una nueva pantalla de información.</p> <p>Las recomendaciones de productos deben considerar aspectos como comportamiento de clientes similares, relaciones entre familias de productos, intereses del Banco de promover productos particulares como parte de estrategias comerciales, entre otros. Son el resultado de un proceso que genera las tres mejores recomendaciones para cada uno de los clientes parametrizados en el modelo. Se debe excluir de las recomendaciones productos que han sido previamente rechazados por el cliente.</p> <p>La información que se debe mostrar al usuario a nivel de Oracle CX es:</p> <ul style="list-style-type: none"> • Calificación de la recomendación: En una escala de 1 a 100, indicar la calificación brindada por el modelo de recomendaciones para esa recomendación, siendo 1 el menor nivel de recomendación, y 100 el más alto. • Nombre del producto recomendado, utilizando el nombre entre el catálogo de productos del Banco en CX • Monto de venta estimado para la recomendación, o N/A si no aplica para el tipo de producto • Fecha de generación de la recomendación • Acción por realizar, indicando las posibles opciones que puede elegir el usuario como siguiente paso a la recomendación, la opción "Convertir en prospectación" permite aceptar la recomendación y pasarla al módulo de prospectaciones como una nueva, y la opción "Rechazar recomendación" permite indicar que la recomendación no es viable para este cliente. <ul style="list-style-type: none"> ◦ <i>Si se elige rechazar la recomendación, el sistema debe presentar un combo de motivos con valores como los siguientes: Producto no aplica para el cliente, No aplica por grupo de interés, Producto previamente rechazado.</i> • Comentarios del Ejecutivo: Permite anotar cualquier comentario del ejecutivo antes de aceptar o rechazar la recomendación, en caso de que se deba tomar nota de algún aspecto antes de procesar la recomendación. • Estado de la recomendación, aparece como Pendiente si la recomendación no ha sido procesada por el Ejecutivo, como Rechazada si fue procesada pero descartada por el Ejecutivo, y Aceptada si se convirtió en prospectación. <ul style="list-style-type: none"> ◦ <i>Si la recomendación aparece como Aceptada, la etiqueta corresponde a un vínculo que lleva al usuario a la prospectación correspondiente.</i>
Para:	Conducir acciones de venta a clientes de forma dirigida y efectiva para cada uno de los segmentos de clientes del Banco, con el fin de aumentar la cantidad de productos por cliente y el valor de cada cliente del Banco.

3

Historia de Usuario HU-02-Resultados	
Yo Como:	Ejecutivo de Banca Personas Ejecutivo de Banca Empresas Ejecutivo de Sucursal (Plataforma y Cajas)
Requiero:	Brindar retroalimentación al modelo de recomendaciones de productos. Como usuario del módulo, y considerando que éste debe aprender y mejorar continuamente, debo brindar información relevante que permita determinar si las recomendaciones del módulo son de valor o no. Al rechazar la recomendación, el módulo me solicita indicar la razón por la cual esa recomendación debe ser rechazada, de entre una lista de valores. Esta información debe ser almacenada en el sistema de recomendaciones para su análisis posterior. Al aceptar la recomendación, esta se convierte en una prospectación. Esta debe crearse ligada al cliente en cuestión, su contacto primario, indicar mediante un campo oculto que la prospectación corresponde a una recomendación de productos automatizada, una probabilidad de gane de 5%, una fecha de cierre 2 semanas después de la fecha actual, y la moneda debe ser tomada de la recomendación. La nueva prospectación creada debe ser asignada automáticamente de acuerdo con el modelo previamente definido de acuerdo con el esquema de territorios y segmentación de fuerza de ventas ya implementado.
Para:	Asegurar que el modelo de recomendaciones pueda mejorar continuamente y ser ajustado para aumentar su efectividad.

Criterios de aceptación:

Número	Descripción de la solicitud	Evento	Resultado
1	Captura de retroalimentación sobre efectividad del modelo de recomendaciones	<ul style="list-style-type: none"> - El sistema presenta al usuario el listado de recomendaciones de productos, indicando una lista de opciones que el usuario puede escoger. - El usuario selecciona una de las opciones para cada recomendación, sea rechazarla, convertirla en prospectación, o anotar comentarios y analizar después. - El usuario selecciona la opción de guardar todo. - El sistema almacena en Oracle CX la información indicada por el usuario para su posterior extracción. 	El sistema almacena información relevante para retroalimentar y mejorar el modelo de recomendaciones.
2	Análisis de información brindada por usuarios	<ul style="list-style-type: none"> - Se realiza una extracción de la información de Oracle CX en modalidad de volcado total de datos, indicando aspectos como ID de Cliente, producto recomendado, calificación de la recomendación, acción realizada por el Ejecutivo, comentarios del Ejecutivo, motivo de rechazo, estado de la recomendación. - La información se analiza utilizando herramientas externas a Oracle CX, con lo cual se determina posibles mejoras a la implementación del modelo de recomendaciones. 	El sistema permite la extracción de información relevante para retroalimentar y mejorar el modelo de recomendaciones.

Historia de Usuario HU-03-Campañas

Yo Como:	Ejecutivo de Banca Personas Ejecutivo de Banca Empresas
Requiero:	<p>Contar con información de recomendaciones de productos para construir campañas de mercadeo en Oracle Eloqua, dirigidas a posibles clientes de nuevos productos. Entre las columnas de información requeridas se tiene calificación de la recomendación, producto recomendado, monto estimado, estado.</p> <p>La información disponible en Oracle Eloqua debe estar actualizada con respecto a la información disponible en Oracle CX, específicamente el estado de cada recomendación de venta debe estar sincronizado hacia Oracle Eloqua a fin de no enviar campañas a clientes que ya han sido ofertados previamente en un producto dado.</p> <p>Oracle Eloqua debe permitir crear segmentos de clientes con base en las diferentes dimensiones de información disponibles con el fin de poder crear campañas más focalizadas y relevantes para los clientes.</p>
Para:	Ofrecer productos de interés a potenciales clientes de acuerdo con recomendaciones de venta, utilizando un mecanismo automatizado y escalable considerando los posibles grandes volúmenes de información que puede generar el modelo de recomendaciones, asegurando que la información utilizada para campañas de mercadeo sea veraz y actual.

Criterios de aceptación:

Número	Descripción de la solicitud	Evento	Resultado
1	Contar con información de recomendaciones de venta para campañas de mercadeo	<ul style="list-style-type: none"> - Eloqua recibe la información de recomendaciones de venta previamente cargada en Oracle CX. - La información de Eloqua se refresca con alrededor de 15 minutos de diferencia entre plataformas. 	Oracle Eloqua cuenta con información actualizada sobre recomendaciones de productos.
2	Crear segmentos de clientes con base en la información de recomendaciones disponible en Eloqua	<ul style="list-style-type: none"> - Utilizando las columnas de información disponibles en Oracle CX, es posible crear segmentos de clientes en Eloqua para crear campañas de mercadeo. - Los segmentos de clientes de Eloqua son refrescados antes de su uso en una campaña de mercadeo, y son incluidos en las campañas de mercadeo necesarias. 	El sistema permite segmentar clientes con base en las recomendaciones de productos generadas, y crear campañas de mercadeo dirigidas.

Propuesta de solución de minería de datos basada en la metodología CRISP-DM utilizando los datos almacenados en Oracle Engagement Cloud, para la recomendación personalizada de productos a clientes, durante el primer semestre de 2022

10.2. Anexo II. Cotización de *database* Oracle

The screenshot displays the OCI My Estimate interface. At the top, it shows 'My Estimate' with a search icon and a menu icon. Below this, there's a navigation bar with 'Start for Free', 'USD - US Dollar', and 'Estimated Monthly Cost \$116.06'. The main section is titled 'Add Configuration' and contains a 'My Configuration' summary bar with details: SHAPE: VM.OPTIMIZED3.FLEX, PROCESSOR: INTEL, OCPU: 2, MEMORY: 32GB, CAPACITY TYPE: ON-DEMAND, and Estimated Monthly Cost \$116.06.

The configuration is broken down into two main sections:

- Compute - Virtual Machine:**
 - SERVICE:** Compute - Virtual Machine
 - UTILIZATION:** 1 instance x 744 hrs/month
 - HOURLY EQUIVALENT PRICE:** \$0.1560
 - ESTIMATED MONTHLY COST:** \$116.06
 - SHAPE:** Processor: Intel, Shape: VM.Optimized3.Flex, OCPU: 2, Memory [GB]: 32, OS Image: Autonomous Linux.
 - PRICING OPTIONS:** Capacity type: On-Demand.
- Storage - Block Volumes:**
 - SERVICE:** Storage - Block Volumes
 - ESTIMATED MONTHLY COST:** \$0.00
 - Storage capacity [GB]:** 100
 - Performance level:** Balanced
 - VPU:** 10
 - Max IOPS:** 25000
 - Max Throughput (MBps):** 480

10.3. Anexo III. Cotización de almacenamiento en AWS

The screenshot shows the AWS Pricing Calculator interface for Redshift. It includes a header with 'aws pricing calculator', 'Feedback', 'English', and 'Contact Sales'. The main content area lists various cost components for Redshift:

- Redshift instance cost (upfront): 0.00 USD
- Additional backup storage
- Redshift Spectrum
- Redshift Managed Storage
- Redshift Data Transfer

At the bottom, the 'Amazon Redshift estimate' section provides a summary:

Redshift instance cost (monthly)	182.50 USD
Data transfer cost (monthly)	0.30 USD
Total monthly cost:	182.80 USD

Buttons for 'Cancel' and 'Save' are visible at the bottom right. An acknowledgment note is present at the very bottom.

Propuesta de solución de minería de datos basada en la metodología CRISP-DM utilizando los datos almacenados en Oracle Engagement Cloud, para la recomendación personalizada de productos a clientes, durante el primer semestre de 2022

10.4. Anexo IV. Cotización de almacenamiento en Microsoft

The screenshot displays the Azure pricing calculator interface. At the top, there are buttons for 'Contact Sales' and 'Free account'. The main section is titled 'Your Estimate' and includes a '+', a close button 'x', and a notification: 'You can now rearrange your services by clicking on the rearrange icon below.' Below this, the service 'Azure Synapse Analytics' is listed with a tier of 'Compute Optimized Gen2, Dedicated SQL Pool...' and costs of '\$0.00 upfront' and '\$1,913.84 monthly'. The configuration includes 'REGION: East US' and 'PERFORMANCE TIER: Compute Optimized Gen2'. Under 'Data Exploration & Data Warehousing', a 'Dedicated SQL Pool' is configured with '100 DWU blocks' and '730 Hours' at a rate of '\$1.510 Per hour'. 'Savings Options' are listed: 'Pay as you go' (selected), '1 year reserved (~37% savings)', and '3 year reserved (~65% savings)'. The total cost is shown as '\$1,102.30 Average per month (\$0.00 charged upfront)'. A green bar at the bottom right indicates the total cost: '\$1,102.30 Average per month (\$0.00 charged upfront)'.

10.5. Anexo V. Cotización de herramienta IBM Watson Studio

IBM Cloud | Catalog | Cost Estimator | Docs | Cookie Preferences | Log in | Sign up

Updated on 04/01/2022

Category: AI / Machine Learning

Compliance: HIPAA Enabled, IAM-enabled

Location: Frankfurt, London, Tokyo, Dallas

Related links: Docs, Terms

Plan	Collaborators	Price
Lite	1 authorized user 10 capacity unit-hours monthly limit Environment = # of capacity units required per hour • 1 vCPU + 4 GB RAM = 0.5 • 2 vCPU + 8 GB RAM = 1 • 4 vCPU + 16 GB RAM = 2 • Decision Optimization = Environment + 5 • Watson NLP (beta) = Environment + 5	Free
Professional	Unlimited collaborators Unlimited elastic compute environments Environment = # of capacity units required per hour • 1 vCPU + 4 GB RAM = 0.5 • 2 vCPU + 8 GB RAM = 1 • 4 vCPU + 16 GB RAM = 2 • 8 vCPU + 32 GB RAM = 4 • 16 vCPU + 64 GB RAM = 8 • 4 vCPU + 24 GB RAM + 0.5 NVIDIA Tesla K80 (1 GPU) = 6 • 8 vCPU + 48 GB RAM + 1 NVIDIA Tesla K80 (2 GPU) = 12 • 16 vCPU + 96 GB RAM + 2 NVIDIA Tesla K80 (4 GPU) = 24 • Decision Optimization = Environment + 5 • Watson NLP (beta) = Environment + 5 NVIDIA K80 GPU environments available only in Dallas on IBM Cloud HIPAA readiness option available only in Dallas on IBM Cloud	\$0.98 USD/Capacity Unit-Hour

The Professional plan for Watson Studio enables your team to solve problems and gain a competitive advantage with enterprise-scale data science and AI. For HIPAA support, you must enable it in your IBM Cloud account settings. For instructions, go to <https://dataplatform.cloud.ibm.com/docs/content/getting-started/security.html#hipaa>

Estimate | Costa Rica | x | x

Watson Studio
Enter your expected usage to view cost details.

Capacity Unit-Hour: 100

Estimated cost per month | Estimated cost doesn't include tax or discounts. | **\$98.00**

Save

10.6. Anexo VI. Cotización de herramienta AWS Sage Maker

aws pricing calculator | Feedback | English | Contact Sales

SageMaker Studio Notebooks

Number of data scientist(s): 1

Number of Studio Notebook instances per data scientist: 1

Studio Notebook hours per day: 5

Studio Notebook days per month: 20

Instance: ml.c5.12xlarge

Selected Instance: ml.c5.12xlarge
Compute Type: Compute Optimized Instances | V CPU: 48 | Memory: 96 GiB | Clock Speed: 3 GHz | GPU: 0 | Network Performance: N/A | Storage: EBS only | GPU Memory: 0

Show calculations

SageMaker Studio Notebooks estimate

Total monthly cost: 244.80 USD

Amazon SageMaker estimate

Total monthly cost: 244.80 USD

Cancel | Add to my estimate

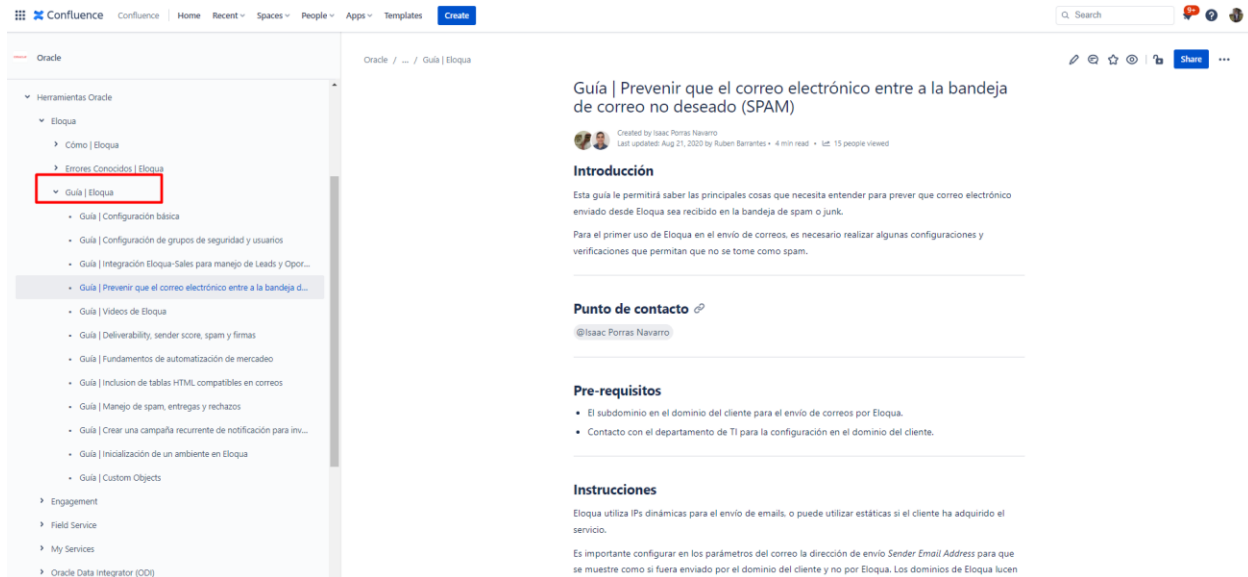
10.7. Anexo VII. Versiones disponibles de la herramienta Anaconda

ANACONDA PRODUCTS

Pricing

Free	Pro	Business	Enterprise
<p>Learn more ↗</p> <p>Students, academics, and hobbyists</p> <p>Get Started</p>	<p>Learn more ↗</p> <p>Professionals</p> <p>Buy Now</p>	<p>Learn more ↗</p> <p>Professionals with security needs</p> <p>Contact Us</p>	<p>Learn more ↗</p> <p>On-prem and private cloud users</p> <p>Contact Us</p>
FREE	Intro price \$14.95 / (\$149/yr)**	Custom	Custom
<p>Anaconda Distribution:</p> <ul style="list-style-type: none"> More than 8,000 DM/ML packages* Conda pkgg/env management system Anaconda & Miniconda installers Anaconda Navigator desktop GUI OS: Windows, macOS, Linux Architecture: x86, ARM, IBM/z, M1 Desktop integrations: Jupyter, RStudio, VSCode, PyCharm, Spyder <p>Nucleus Community Membership:</p> <ul style="list-style-type: none"> Community content & support Environment backup on Nucleus <p>* Mirroring rights not included</p>	<p>All the features of Free, plus:</p> <ul style="list-style-type: none"> Compliant for commercial use Professional-grade repository Conda signature verification Tokenized user access control Basic package usage reporting <p>Available add-ons: Contact us</p> <ul style="list-style-type: none"> Site license Custom private mirroring Support services Kickstart services Long-term Support (LTS) Jupyter notebook services <p>** Get started now at our introductory rate, available for a limited time</p>	<p>All the features of Pro, plus:</p> <ul style="list-style-type: none"> Curated vulnerability matching Security policy & license filtering Channel management Role-based access control Standard package usage reporting <p>Available add-ons:</p> <ul style="list-style-type: none"> Custom installers Custom private mirroring Support services Kickstart services Long-term Support (LTS) Jupyter notebook services 	<p>All the features of Business, plus:</p> <ul style="list-style-type: none"> On-prem repository & server Air-gap environment support Mirror CRAN, PyPI, & conda-forge Administer & track artifact history Support services <p>Available add-ons:</p> <ul style="list-style-type: none"> Site license Custom installers Implementation services Premium support services Kickstart services Enterprise DS platform Long-term Support (LTS)

10.8. Anexo VIII. Documentación interna de Eloqua



The screenshot shows a Confluence page within the Oracle ecosystem. The left sidebar displays a navigation menu under 'Oracle' with a sub-section 'Herramientas Oracle' containing 'Eloqua'. Under 'Eloqua', there is a sub-section 'Errores Conocidos | Eloqua' where the item 'Guía | Eloqua' is highlighted with a red box. The main content area shows the title 'Guía | Prevenir que el correo electrónico entre a la bandeja de correo no deseado (SPAM)' and the author 'Created by Isaac Porras Navarro'. The page content includes an introduction, a contact point for Isaac Porras Navarro, pre-requisites, and instructions for configuring email delivery.

Guía | Prevenir que el correo electrónico entre a la bandeja de correo no deseado (SPAM)

Created by Isaac Porras Navarro
Last updated: Aug 21, 2020 by Ruben Barrios • 4 min read • 13 people viewed

Introducción

Esta guía le permitirá saber las principales cosas que necesita entender para prevenir que correo electrónico enviado desde Eloqua sea recibido en la bandeja de spam o junk.

Para el primer uso de Eloqua en el envío de correos, es necesario realizar algunas configuraciones y verificaciones que permitan que no se tome como spam.

Punto de contacto

@Isaac Porras Navarro

Pre-requisitos





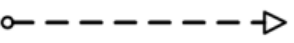
- El subdominio en el dominio del cliente para el envío de correos por Eloqua.
- Contacto con el departamento de TI para la configuración en el dominio del cliente.

Instrucciones

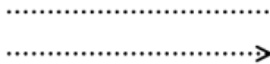




Eloqua utiliza IPs dinámicas para el envío de emails, o puede utilizar estáticas si el cliente ha adquirido el servicio.

Es importante configurar en los parámetros del correo la dirección de envío *Sender Email Address* para que se muestre como si fuera enviado por el dominio del cliente y no por Eloqua. Los dominios de Eloqua lucen

10.9. Anexo IX. Notación BPMN

Elemento	Descripción	Notación
Evento	Un evento es algo que sucede durante un proceso. Gráficamente, son círculos con centros abiertos y existen tres tipos que pueden afectar el flujo: inicio, intermedio y final.	
Actividad	Una actividad se refiere a subprocesos o tareas que ejecuta la organización y se representa con rectángulos redondeados.	
Compuerta	Una compuerta se utiliza para enlazar y controlar la divergencia y convergencia de flujos de secuencia en un proceso y se representa por medio de un rumbo.	
Flujo de secuencia	Un flujo de secuencia se utiliza para mostrar el orden en que las actividades realizan el proceso y se representa por medio de una flecha.	
Flujo de mensajes	Un flujo de mensajes se utiliza para mostrar la comunicación entre los participantes del proceso, se representa por medio de una flecha punteada.	

Propuesta de solución de minería de datos basada en la metodología CRISP-DM utilizando los datos almacenados en Oracle Engagement Cloud, para la recomendación personalizada de productos a clientes, durante el primer semestre de 2022

Elemento	Descripción	Notación
Asociación	Una asociación se utiliza para vincular información entre los elementos gráficos del proceso. Se representa por medio de una línea punteada y cuando se incorpora la fecha es para brindar dirección.	
Contenedor	Un contenedor es la representación gráfica de un participante en el proceso. Se representa mediante un rectángulo horizontal.	
Carril	Un carril es una subpartición dentro de un proceso, se utilizan para organizar y clasificar actividades específicas. Se representa por medio de un rectángulo horizontal particionado en dos.	
Mensaje	Un mensaje se utiliza para representar el contenido de una comunicación entre dos participantes del proceso.	
Anotación de texto	Las anotaciones de texto son un mecanismo para que el modelador proporcione información adicional al lector del diagrama.	

10.10. Anexo X. Evaluaciones de la empresa

Rubro	Puntaje obtenido		
	I Evaluación	II Evaluación	III Evaluación
A. HABILIDADES ESTRATÉGICAS DEL ESTUDIANTE	3	3	3
a. Responsabilidad y puntualidad en las reuniones y entregas.	3	3	3
b. Comunicación asertiva y facilidad de expresión. (41696)	3	3	3
c. Proactividad.	3	3	3

Rubro	Puntaje obtenido		
	I Evaluación	II Evaluación	III Evaluación
d. Trabajo colaborativo y capacidad organizativa.	3	3	3
B. ACERCA DEL TRABAJO REALIZADO A LA FECHA	3	3	3
a. Disposición autodidacta.	3	3	3
b. Seguimiento a recomendaciones que se le dan.	3	3	3
c. Cumplimiento del cronograma de su trabajo.	3	3	3
d. Pensamiento sistemático o estratégico.	3	3	3
C. SOBRE LOS ENTREGABLES DEL ESTUDIANTE	3	3	3
a. Estructura lógica de los informes, minutas, correos que elabora, entre otros. (41704)	3	3	3
b. Claridad en la secuencia de ideas que expone.	3	3	3
c. Las minutas reflejan los acuerdos tomados en las reuniones.	3	3	3
d. Uso correcto de idioma oficial de la compañía.	3	3	3
e. Profundidad del contenido desarrollado dentro de sus documentos o propuestas.	3	3	3
D. ÉTICA PROFESIONAL DEL ESTUDIANTE	3	3	3
a. Compromiso con la calidad de su trabajo.	3	3	3
b. Respeto a la confidencialidad de la información brindada por la organización.	3	3	3
c. Honestidad en su actuar diario.	3	3	3
d. Tolerancia y aceptación a todo tipo de diversidad.	3	3	3

10.11. Anexo XI. Reporte de minutas firmado

NOTA ACLARATORIA

Aprobación de minutas TFG

Ante la situación que afronta el país por la pandemia mundial, se aprueban por parte de la Ing. María José Artavia Jiménez, las siguientes minutas correspondientes al proyecto de graduación: Propuesta de solución de minería de datos basada en la metodología CRISP-DM utilizando los datos almacenados en Oracle Engagement Cloud, para la recomendación personalizada de productos a clientes, durante el primer semestre de 2022. Realizado por el estudiante Emanuel Alvarado Pérez, carné 2017135013, cédula 117600746.

A continuación, se muestra la lista de las minutas aprobadas por la Ing. María José Artavia Jiménez:

- Minuta 001. Reunión inicial
- Minuta 002. Sesión inicial de TFG con la organización
- Minuta 003. Sesión semanal de seguimiento
- Minuta 005. Sesión acuerdos iniciales con la organización
- Minuta 006. Sesión semanal de seguimiento
- Minuta 007. Sesión semanal de seguimiento
- Minuta 008. Sesión semanal de seguimiento
- Minuta 009. Sesión semanal de seguimiento
- Minuta 010. Sesión semanal de seguimiento
- Minuta 011. Segunda sesión de seguimiento de TFG con la organización
- Minuta 012. Sesión semanal de seguimiento
- Minuta 013. Sesión semanal de seguimiento
- Minuta 014. Sesión semanal de seguimiento
- Minuta 015. Sesión de seguimiento con la coordinación de TFG
- Minuta 016. Sesión semanal de seguimiento
- Minuta 017. Sesión semanal de seguimiento
- Minuta 019. Sesión de cierre con la organización

Firmado digitalmente por
MARIA JOSE ARTAVIA
JIMENEZ (FIRMA)
Fecha: 2022.05.28 21:44:07
-06'00'

Ing. María José Artavia Jiménez

Propuesta de solución de minería de datos basada en la metodología CRISP-DM utilizando los datos almacenados en Oracle Engagement Cloud, para la recomendación personalizada de productos a clientes, durante el primer semestre de 2022

10.12. Anexo XII. Carta de la filóloga

San José, Costa Rica, 29 de mayo de 2022

Ingeniero
Luis Javier Chavarría Sánchez
Director
Licenciatura en Administración de Tecnología de Información
Área Académica de Administración de Tecnologías de Información
Instituto Tecnológico de Costa Rica

Estimado señor:

Por este medio hago constar que he revisado el Trabajo Final de Graduación, titulado *Propuesta de solución de minería de datos basada en la metodología CRISP-DM utilizando los datos almacenados en Oracle Engagement Cloud, para la recomendación personalizada de productos a clientes, durante el primer semestre de 2022*, elaborado como requisito para obtener el grado de Licenciatura en Administración de Tecnología de Información, por el estudiante Emanuel Alvarado Pérez, cédula 1-1760-0746.

Al respecto, indico que he revisado y corregido errores gramaticales, de puntuación y ortografía, construcción de párrafos y vicios del lenguaje, empleo adecuado del formato APA séptima edición, así como otros aspectos relacionados con el campo filológico, que se manifestaron en el documento escrito.

Desde ese punto de vista, considero que, con las correcciones realizadas en el documento, está listo para ser presentado como Trabajo Final de Graduación.

Atentamente,



Licda. Marjorie Martínez Castro
Cédula 5-0339-0587
Carné de colegiatura: 94974