



**INSTITUTO TECNOLÓGICO DE COSTA RICA**

**ESCUELA DE ADMINISTRACIÓN DE EMPRESAS**

**PROGRAMA DE LICENCIATURA**

**SEMINARIO GERENCIAL DE FINANZAS**

“Credit Scoring en Costa Rica y la probabilidad de clasificación de créditos personales basados en un modelo estadístico-matemático para aprobar o rechazar”

**Integrantes:**

Hernández Cerdas Patricia 2004-45081

Montoya Leiva José Pablo. 2004-50558

Villarreal Espinoza Allan. 2001-33466

**Profesor:**

Garro Martínez Víctor.

San José, Costa Rica.

I Semestre, 2013

## INDICE

I. INTRODUCCIÓN.....	1
II. DISEÑO DE LA INVESTIGACIÓN.....	2
A. PROBLEMA DE INVESTIGACIÓN.....	2
B. INTERROGANTES.....	2
C. OBJETIVOS .....	3
D. ESTRATEGIAS METODOLÓGICAS.....	4
E. LIMITACIONES.....	4
F. REVISIÓN DE LA LITERATURA .....	5
F. JURISPRUDENCIA.....	15
III. METODOLOGÍA .....	18
IV. ANÁLISIS DE LOS DATOS .....	24
A. MODELO LOGIT.....	24
B. COMPARACIÓN MODELO LOGIT Y MODELO UTILIZADO POR EL BANCO .....	41
V. CONCLUSIONES .....	44
VI. RECOMENDACIONES.....	46
VII. ANEXOS .....	48
VIII. BIBLIOGRAFÍA .....	49

## I. INTRODUCCIÓN

Hoy en día el ser humano vive en un mundo rodeado de consumismo, donde las comodidades y lujos son parte importante de supervivencia, sin embargo, también existen necesidades que se deben satisfacer como lo es la vivienda, salud, educación, entre otras.

Estas necesidades y consumismo ha llevado a las personas a solicitar a las instituciones financieras créditos que les permita lograr satisfacer dichos propósitos, lo que obliga a las entidades a realizar una serie de análisis o estudios para evaluar si dicha persona es confiable para brindarle tal crédito, ya que en los tiempos actuales, el ser humano se ha endeudado de tal forma que no le es posible cancelar sus deudas, o bien simplemente no desean hacerlo, logrando así que las instituciones pierdan importantes sumas de dinero, tiempo y recursos en capital humano.

Dado lo anterior, se puede evidenciar la importancia de contar con un sistema o un modelo que le permita determinar si se puede o no brindar el crédito a la persona solicitante.

Un modelo muy conocido y utilizado no solo a nivel de Costa Rica sino también a nivel mundial, es lo que conocemos como “Modelos de Credit Scoring”, el cual busca hacer más eficiente y eficaz la utilización de los recursos.

Es por esto que el presente trabajo consiste en realizar una revisión de la literatura con el fin de ampliar nuestros conocimientos en el tema, identificando los modelos existentes y de esta forma plantear una propuesta de modelo que le permita mejorar la asignación de recursos y disminuir en un porcentaje el riesgo crediticio.

Es importante aclarar que es posible disminuir dicho riesgo, sin embargo, es difícil, casi imposible lograr eliminarlo, ya que también existe la probabilidad de otorgar un crédito a un mal deudor o bien, negarle el préstamo a un buen pagador.

## II. DISEÑO DE LA INVESTIGACIÓN

### A. PROBLEMA DE INVESTIGACIÓN

¿Cuál de las metodologías del modelo para Credit Scoring es el más adecuado para otorgar o rechazar los créditos de consumo en instituciones bancarias en Costa Rica a personas físicas?

### B. INTERROGANTES

- 1) ¿Cuáles son las ventajas y desventajas del uso de un Credit Scoring para la resolución de solicitudes de crédito de consumo en la Institución que se está analizando?
- 2) ¿Cuáles son los criterios que un banco debe considerar para seleccionar la modalidad de Credit Scoring más adecuado?
- 3) ¿Es el modelo Logit el adecuado para analizar un crédito a individuos en una entidad financiera?
- 4) ¿Es el modelo usado por el Banco, efectivo y eficaz en la toma de decisión en lo que se refiere a otorgarle o no un crédito a una persona física, bajo la perspectiva de los criterios identificados?
- 5) ¿Tiene algún poder predictivo el modelo utilizado con la morosidad de la cartera?
- 6) ¿Qué variables son las que toma en cuenta la entidad bancaria para hacer el análisis de crédito?
- 7) ¿Cuál es la política de la entidad financiera, para otorgar o rechazar un crédito?
- 8) ¿Existe alguna normativa a nivel nacional o internacional aplicada a los modelos Credit Scoring utilizados en Costa Rica?

## **C. OBJETIVOS**

### **1. OBJETIVO GENERAL**

Desarrollar y comparar un modelo de Credit Scoring aplicable en instituciones financieras para la medición y otorgamiento de consumo en instituciones bancarias en Costa Rica.

### **2. OBJETIVOS ESPECÍFICOS**

- 2.1 Enunciar las ventajas y desventajas del uso de un modelo de Credit Scoring.
- 2.2 Desarrollar el Modelo Logit con un caso para obtener un resultado del mismo y así analizar los criterios más adecuados para comparar las principales variables.
- 2.3 Comparar el resultado del modelo con la base de datos de morosidad para corroborar la relación.
- 2.4 Analizar con respecto a los resultados obtenidos si utilizando el modelo se hubiese podido predecir el resultado final.
- 2.5 Determinar las variables que la entidad bancaria considera para realizar el análisis de crédito.
- 2.6 Verificar la política que la entidad financiera utiliza para otorgar o rechazar un crédito.
- 2.7 Consultar mediante SUGEF y/o entidad bancaria si existe alguna normativa que regule el uso de Credit Scoring en Costa Rica.

## D. ESTRATEGIAS METODOLÓGICAS

- Por medio de la revisión de la literatura y de otros estudios realizados se identificó cuales son los modelos más utilizados en otros países, y determinar de esta forma si es eficiente o eficaz el modelo actual, de acuerdo a la base de datos que proporcionaron.
- Se obtuvo por medio de una entrevista al experto del banco información sobre el Credit Scoring que utiliza la institución.
- Utilizando las bases de datos se logro ubicar gran cantidad de información con la cual se plasmo una idea sintetizada de los modelos y las estrategias planteadas.
- Una vez obtenidos los datos de los casos, se realizó una comparación con los resultados reales de las bases de morosidad para validar el modelo. De esta forma se obtuvieron los resultados necesarios para poder realizar las mediciones correspondientes y así tener con datos reales cuales serian los resultados a analizar y con los cuales tomar las decisiones correspondientes.
- Con todos esos datos se realizó un análisis del modelo Logit.

## E. LIMITACIONES

1. Dado la dificultad al acceso de la información y la cantidad de la misma, se limitó el estudio solo al sector de créditos personales para consumo, por lo que préstamos personales fue la categoría seleccionada para estudiar y analizar.
2. Respetar el secreto bancario dado que los datos que se analiza se ha tomado de una entidad financiera, de forma tal que el acceso a la información analizada (en lo que interesa de acuerdo a este trabajo de índole académica) no permite identificar el acceso a ningún dato personal de los usuarios de dicha entidad.
3. Uno de los objetivos planteados no pudo llevarse a cabo debido a que la entidad financiera no facilito las políticas para otorgar o rechazar un crédito.

4. No fue posible conocer a detalle el modelo actual que utiliza el banco, razón por la cual la comparación entre dicho modelo y los realizados en este estudio no pueden profundizarse.

## F. REVISIÓN DE LA LITERATURA

La síntesis de los principales modelos, se hizo para determinar los posibles modelos a considerar en el desarrollo de la investigación y determinar cuál sería el más apropiado.

(Armando, 2012), indican que la implementación de un modelo de medición numérica puede dar datos de mucha importancia en la toma de decisiones basados en estudios de compañías, administradores y de empleados. A través de algoritmos financieros se propusieron minimizar las probabilidades de errores que pudiesen poner en riesgo a la entidad.

Por otro lado (Eva, 2009) busca explicar mediante sus investigaciones, un modelo de Credit Scoring donde no se tome en cuenta una función de coste del error, sino que busca que la clasificación en la selección de los clientes analice de manera más íntegra al posible sujeto que se le vaya a aprobar un crédito. Ellos implementan un sistema de medición por distancias como modelo de medición de datos reales que proporcionen información de mayor confiabilidad para la toma de decisiones a la hora de otorgar los diferentes créditos, además, de demostrar que al implementar una serie de datos estadísticos pueden obtener gran información que una vez trasladada a las fórmulas probabilísticas puedan dar datos suficientes para determinar si una persona es apta o no para un crédito.

Para lograr esto utilizaron las bases de datos de dos entidades financieras para así, determinar los diferentes escenarios en que se comportaba el mercado y de esta forma ubicar de una manera más confiable cuales serían los objetivos a implementar.

El modelo de medición en distancias fue enfocado en métodos no paramétricos, estimación de núcleo, la densidad y el árbol de clasificación, que ayudaron a obtener una solvencia en los datos que se iban obteniendo por parte de los autores.

Los autores citados definen la importancia del modelo por los diferentes escenarios donde se ubican, por ejemplo, determinan la clasificación de la población en un modelo donde identifican si el sujeto es bueno o malo en la

clasificación final, de esta forma se puede determinar si efectivamente el mismo podrá ser tomado en cuenta para un crédito.

No solo es dar un crédito por darlo sino que hay que determinar la capacidad de pago de los sujetos y de esta forma no poner en riesgo la empresa con clientes que al final estén morosos y el riesgo de recuperar el capital sea lento o nulo.

Igual los autores recomiendan no otorgar todos los créditos a los clientes que los datos les proporcione un estatus adecuado de crédito, ya que hay que tomar en cuenta factores de riesgo a futuro. Siempre hay que enfocarse en la estructura que se esté generando en el mercado, evaluar y determinar las posibles situaciones próximas, trabajar con una cartera de crédito ayudara mucho a que la entidad financiera no corra el riesgo de otorgar una mayor cantidad de créditos de los que puede recuperar.

Un punto importante en los autores y que se considera una propuesta que es interesante, es que puede ser utilizada por las empresas sin dejar de seguir innovando y analizando los movimientos del mercado en ese momento.

Por otro lado, los modelos estadísticos tienen una dimensión individual ya que se enfocan en el riesgo de incumplimiento del individuo o empresa independientemente de lo que ocurra con el resto de la cartera de préstamos, permitiendo diferenciarlo de otras herramientas de medición del riesgo de crédito como lo son los modelos de cartera y los VaR marginales, que tienen en cuenta la correlación de la calidad crediticia de los deudores de una cartera de préstamos. (Girault, Modelos de Credit Scoring (Qué, Cómo, Cuándo y Para qué), 2007).

En este punto es importante mencionar que los modelos de Credit Scoring han tomado un verdadero auge en los últimos años, ya que la toma de decisiones crediticias no es solo para las entidades financieras sino que hoy en día su utilización se ha generalizado en actividades de mercadeo y de cobranza, esto se puede observar en casos de telefonía celular, servicios públicos, entre otros.

Como se ha mencionado en ocasiones anteriores, el trabajo desarrollado es con el fin de determinar si una persona es buena o mala pagadora, es decir, la probabilidad de entrar o no en incumplimiento. Para esto es importante que definamos que es incumplimiento.

*Según Basilea I:* “Es la falta de pago de un préstamo o cualquier otro tipo de violación de la condición de un contrato de préstamo.”

Según Basilea II: “Se considera incumplimiento del deudor cuando acontece al menos una de las siguientes circunstancias:

- La EIF considera poco probable que el deudor cancele la totalidad de sus obligaciones crediticias, sin que la EIF recurra a acciones tales como la realización de Fianzas (si existiera).
- El deudor se encuentra en situación de mora por más de 90 días con respecto a cualquier obligación crediticia frente a la EIF.”

El comité de Basilea busca regular a las entidades reglamentando sus actividades y cotejando mediante técnicas de verificación a cada institución financiera a nivel mundial, un resumen de las principales características de ambos modelos de Basilea lo determinan (Arguedas, 2008). Los autores identifican cada elemento mediante la siguiente tabla explicativa:

<b>Cuadro Comparativo</b>		
	<b><u>BASILEA I</u></b>	<b><u>BASILEA II</u></b>
<b>OBJETIVOS</b>	Fortalecer la solidez y estabilidad del sistema bancario internacional a través de requerimientos de capital para disminuir las desigualdades competitivas entre los bancos internacionales que participan en el mercado.	-Promover la seguridad y la solvencia de los sistemas financieros. -Mejorar la igualdad competitiva. -Enfoque mas comprensivo de riesgos, para eliminar las criticas al acuerdo del 88. - Enfocarse en los bancos internacionalmente activos
<b>APLICACION</b>	Quedaba a discreción del ente supervisor.	Se aplicará a todos los bancos internacionalmente activos.
<b>ESTRUCTURA</b>	Aborda únicamente el requisito de capital mínimo, que es un 8%.	Contempla tres “pilares”. 1- Requisitos de capital mínimo, se mantiene en 8%, - 2- Examen del supervisor 3- Disciplina de mercado.
<b>RIESGOS</b>	Requerimiento mínimo de capital dirigido a riesgo de crédito.	Dirigido a Riesgo crediticio, riesgo de mercado y riesgo operativo;
<b>METODOS DE PONDERACION DE RIESGOS</b>	Utiliza únicamente el método estándar.	Utiliza los métodos estándar y métodos de calificación interna (básico y avanzado) para riesgo de crédito y el método básico, estándar y métodos de medición avanzada para riesgo operativo.

Para poder realizar un análisis o modelo de Credit Scoring, existen varias metodologías las cuales deben ser seleccionadas de acuerdo a la información de la que se dispone o se vaya a analizar.

Cada una de ellas arroja datos interesantes para ser evaluados por parte de las instituciones financieras que así lo deseen. A continuación explicaremos algunos de los modelos investigados.

Figura N°1  
MODELOS DE MEDICIÓN



#### A. ANÁLISIS DISCRIMINANTE

Para los autores (Joaquina Laffarge Briones) el análisis discriminante es una técnica estadística multivariable que permite estudiar simultáneamente el comportamiento de un grupo de variables independientes con la intención de clasificar una serie de casos en grupos previamente definidos y excluyentes entre sí, es decir, clasifica diversas observaciones en determinados grupos, por lo que se puede decir que su principal característica es la clasificación de los individuos que se estudian a partir de variables discriminantes que ya se conocen previamente.

Al analizar las diferentes variables en una base de datos, se puede determinar de acuerdo a ellas y al comportamiento del individuo en procesos pasados cual sería la proyección para la evaluación que se le va a aplicar, por lo que mediante este análisis se puede obtener resultados para determinar si un individuo es apto o no para aceptarle un crédito personal.

Como se conoce con anterioridad el grupo a estudiar, los resultados que se obtienen son de un criterio de análisis un poco más detallado y con el cual el evaluador puede tomar las decisiones con fundamentos ya determinados.

Esa es la base del modelo de Altman quien desarrollo la metodología más utilizada para pronosticar la insolvencia de una empresa aplicando el modelo Z.

Esta metodología se adaptó para la predicción de morosidad de clientes en entidades bancarias, la cual es conocida como Análisis Discriminante.

$$Z = 1,2 * X1 + 1,4 * X2 + 3,3 * X3 + 0,6 * X4 + 1,0 * X5$$

El modelo tiene 5 variables:

X1: (Working Capital/Total Assets)

X2: (Retained Earnings/Total Assets)

X3: (EBITDA/Total Assets)

X4: (Market Value of Equity/Total Liabilities)

X5: (Net Sales/Total Assets)

Este tipo de análisis puede tener las siguientes ventajas:

- Utiliza toda la información disponible.
- Toma en cuenta la interdependencia que puede existir entre las variables y permite reducir la dimensionalidad del espacio ya que lo transforma de multivariado en una sola función discriminante.

Para estos mismos autores, entre los inconvenientes que presenta el análisis discriminante, está la rigidez para cumplir las hipótesis de partida (lineal, normalidad, e independencia) y sobre todo, la incapacidad para el cálculo de las probabilidades de impago.

## **B. REGRESIÓN LINEAL**

Este es un modelo el cual utiliza el enfoque de una regresión por cuadrados mínimos. La variable dependiente toma el valor de 1 si el cliente es malo y no cumple con su obligación y la variable independiente el valor de 0 si el cliente es bueno y cumple con los pagos que le corresponden. Esta técnica ha sido utilizada en la construcción de modelos de Credit Scoring para préstamos de consumo, haciendo uso de variables sobre comportamiento del cliente, clasificándolas en grupos de liquidez, rentabilidad, apalancamiento y actividad, para obtener los datos que podrán identificar si un cliente es apto o no para un crédito. (Girault, Modelos de Credit Scoring, 2007).

### **C. EL MODELO PROBIT**

El Modelo Probit busca mediante estadísticas generar valores de medición que permitan determinar si un cliente puede ser considerado adecuado para el otorgamiento de un crédito. (Simbaqueba, 2004)

Generando las diferentes variables y codificando los números de 1 ó 0 se logra traer un valor mediante las funciones algorítmicas, las mismas nos ayudan a calificar de una forma un poco más detallada al cliente.

La verosimilitud empleada en este modelo puede generar resultados confiables que ayudan a evaluar y determinar los principales factores que son ingresados, se puede determinar que un valor como un individuo que está divorciado podría devolver un resultado no muy positivo en la base de datos, pero si a ese factor se le suman otros datos, esa calificación podía generar que la evaluación tome una curva positiva al final.

Es por esta razón que la calidad de los factores es importante en la medición final del modelo, ya que nos ayuda a que los datos sean lo más confiables posibles.

### **D. EL MODELO LOGIT**

Este tipo de modelos considera intervalos entre 0 y 1, y es considerado un modelo cualitativo en donde la variable dependiente es dicotómica, es decir, que solamente tiene dos opciones o es A ó es B, dicho de otra forma, o es 1 ó es 0 y lo que busca es clasificar a los individuos en una de las dos opciones. (Eva Boj, 2010)

Este tipo de herramienta nos proporciona gran cantidad de información, y en el caso de esta investigación la metodología busca la probabilidad que un individuo nuevo pertenezca a un grupo específico. Con él se puede verificar el comportamiento que podría tener un sujeto de crédito y con las diferentes variables se puede determinar si este puede ser catalogado de una forma positiva, negativa o pasar a un análisis un poco más exhaustivo para tomar la decisión final. Utilizando las diferentes variables se puede tener con certeza cuál es el comportamiento de los mismos.

La principal ventaja del modelo de regresión Logit radica en que no es necesario plantear hipótesis de partida, como por ejemplo la normalidad de la distribución de

las variables, mejorando el tratamiento de las variables cualitativas o categóricas (Joaquina Laffarge Briones).

Además este modelo presenta la ventaja de medir probabilidades de incumplimiento al mantener la variable explicada siempre dentro de un rango de variación entre cero y uno. Wiginton (1980), fue uno de los primero autores en publicar un estudio comparando entre análisis discriminante y el modelo Logit en el que determino que dicho modelo ofrecía un porcentaje de calificación mejor que el discriminante.

### **E. JUICIO HUMANO**

En este modelo es importante indicar que independientemente de los resultados que arrojan cada modelo, y de sus porcentajes, el analista de crédito puede valorar información que no fue tomada en cuenta en las bases de datos y que puede determinar en un futuro si la persona que potencialmente es un cliente catalogado como “malo”, puede tener factores que determinen a juicio humano que se le puede otorgar un crédito a pesar de la mala calificación que se le pueda dar.

Caso contrario, puede existir que un cliente posea una calificación de “bueno”, pero algunos signos ubicados en el analista puede determinar que no sea conveniente para la institución determinar un crédito.

### **MODELO A APLICAR: LOGIT**

De acuerdo a la investigación realizada, a la información y programas con los que se contaba se determinó efectuar el modelo únicamente por medio del Logit, por tener la limitante de contar con el tiempo muy restringido y la falta de un programa estadístico que nos permitiera generar la información. En el caso del modelo Logit este nos daría los datos para ser analizados, en el principio se tenía previsto crear una comparación entre Discriminante y Logit, pero se decidió solo enfocarse en el segundo debido a que el principal inconveniente que posee el modelo discriminante es que supone implícitamente que las variables siguen una distribución normal multivariable de las dos poblaciones, con matrices de varianza-covarianza iguales, diferenciándose solamente en los valores de los vectores medios. Estas hipótesis son difíciles de verificar en un problema práctico. Así, en la mayoría de los problemas financieros, al menos una de las variables (generalmente la variable dependiente) es cualitativa, eliminando la posibilidad de

normalidad. En tales casos una técnica que trate específicamente con variables cualitativas como puede ser el análisis Logit, podría parecer más oportuna.

Por su parte la ventaja que presenta el modelo Logit es la suavidad de sus hipótesis de partida, que son fácilmente verificables en cualquier problema empírico. (Joaquina Laffarga, 1987)

Por otro lado y de acuerdo con la investigación de (Vasquez, 2013) los bancos podrán optar por dos métodos para la determinación del riesgo de crédito, este aspecto es importante debido a que el desarrollo del modelo se llevará a cabo por medio del análisis Logit desarrollado mediante el programa de EXCEL.

Para este modelo es importante considerar que existe el método estándar y el método basado en calificaciones internas conocido como IRB. Dicho método se divide en dos categorías:

- El IRB Básico: cuando los bancos solo estiman la probabilidad de incumplimiento, y el resto de indicadores es proporcionado por el Comité de Basilea.
- El IRB Avanzado, donde los bancos estiman todos los indicadores que se requieren para el cálculo de las ecuaciones proporcionadas por el Comité de Basilea.



El proceso de Credit Scoring inicia con el tratamiento de los datos, el cual consiste en la depuración de la base de datos histórica, definición de bueno y malo (incumplimiento) y la partición de la muestra, seguido del análisis univariante que consiste en determinar las variables discretas y cuantitativas, además, del análisis a través de indicadores como el IV, KS y el GINI, para determinar cuáles son las variables que se deben tomar en consideración. Una vez realizado este paso se deben optimizar las variables discretas que presenten datos anormales.

Según (Vasquez, 2013) el Information Value conocido como IV se debe medir por una ponderación de los diferentes WOE<sub>i</sub>, con esas variables se puede lograr determinar si una variable discrimina o no entre buenos y malos

$$IV = \sum_{i=1}^k (dist\_buenos_i - dist\_malos_i) * WOE_i$$

donde:  $k$  = número de categorías de la variable

Para (Vasquez, 2013) el Kolmogorov Smirnov es la medición de datos estadísticos que busca generar los resultados que se acerquen más a la realidad y la probabilidad que se busca.

La prueba del Kolmogorov-Smirnov, se le conoce como una prueba no paramétrica que genera los datos necesarios para determinar la bondad en los ajustes que generan las distribuciones de que ocurra alguna situación.

Es importante indicar que este modelo estadístico es más sensible a los datos cercanos a la media de los que están a los extremos de las bases de datos estudiadas.

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n \begin{cases} 1 & \text{si } y_i \leq x, \\ 0 & \text{alternativa.} \end{cases}$$

Para dos colas el estadístico viene dado por

$$D_n^+ = \max(F_n(x) - F(x))$$

$$D_n^- = \max(F(x) - F_n(x))$$

donde  $F(x)$  es la distribución presentada como hipótesis.

Para este mismo autor el GINI es un análisis estadístico comprendido entre 0% y 100% que permite determinar el poder discriminante de cada variable del modelo. Su cálculo se determina por:

$$G = 1 - \sum_{i=1}^{n-1} (AB\%_{0i} + AB\%_{0i-1})(AM\%_{0i} - AM\%_{0i-1})$$

0% <= GINI <= 5%	No predictiva
5% < GINI <= 15%	Predictiva débil pero se incluye en el modelo
GINI > 15%	Predictiva se recomienda usar la variable

El siguiente paso a seguir es el desarrollo del modelo multivariante que es cuando la variable endógena a modelar es una variable discreta con varias alternativas posibles, para lo cual se realiza un análisis de correlaciones entre las variables y para este caso se corre el modelo Logit.

El Scoring de Admisión cuyo objetivo es evaluar las solicitudes de créditos para su aceptación o rechazo, el Scoring de Comportamiento con el objeto de evaluar el incumplimiento de las operaciones otorgadas y el Scoring de Recobro evalúa la probabilidad de que una operación se recupere el incumplimiento.

Para calcular la probabilidad de incumplimiento se hace a través de la siguiente fórmula:

$$PE = PD * LGD * EAD$$

Donde PE es la pérdida esperada y PD es la probabilidad de pérdida porcentual.

La LGD es la pérdida dado el incumplimiento de forma porcentual, y EAD la exposición al momento del incumplimiento en términos monetarios.

El desarrollo de un Credit Scoring tiene sus ventajas pero también sus desventajas, para el autor (Schreiner, 2002), menciona las siguientes:

### **VENTAJAS DEL USO DEL MODELO DE CREDIT SCORING**

- Cuantifica el riesgo como una probabilidad: el riesgo pronosticado con el scoring podría indicar una probabilidad por medio de un porcentaje que un préstamo determinado se atrase y en el caso del scoring subjetivo simplemente expresa que un préstamo tiene un riesgo por debajo del promedio.
- Es consistente: Dos personas con las mismas características tendrán el mismo pronóstico de riesgo.
- Es explícito: Se conoce y puede explicar el proceso exacto que se realizó para pronosticar el riesgo.

- Considera una amplia gama de factores: Puede cuantificar cómo cambiaría el pronóstico de riesgo si un préstamo tuviera una cobertura de garantía de solamente un 80 por ciento o solamente de un 100 por ciento.
- Puede probarse antes de usarlo: Se pueden realizar pruebas y compararse con lo observado en fechas anteriores.

## **DESVENTAJAS**

- El scoring requiere datos sobre muchos préstamos.
- El scoring requiere muchos datos de cada préstamo.
- Requiere información de calidad adecuada.
- Puede denegar solicitudes pero no puede aprobarlas o modificarlas.
- Funciona con probabilidades no con certezas.

## **F. JURISPRUDENCIA**

### JURISPRUDENCIA NACIONAL

Se indagó sobre la jurisprudencia que existe en nuestro país en el tema de Credit Scoring, por lo que se realizó la consulta a la Superintendencia General de Entidades Financieras SUGEF.

Se consultó si ellos tenían alguna normativa o directriz para regular el servicio de préstamos personales en las entidades financieras, la respuesta que indicaron fue que la función de la SUGEF es velar por la estabilidad del Sistema Financiero Nacional como un todo y ésta tiene competencias primordialmente en materia financiero-contable, no así en asuntos relativos a prestación de servicios en las diferentes entidades financieras, todo amparado en la Ley Orgánica del Banco Central de Costa Rica número 7558.

Agregan que la SUGEF no es una instancia judicial y más bien como ente superior carece de potestad para co-administrar los entes que supervisa, adicional a esto los servicios que se prestan en las instituciones financieras corresponden a políticas internas de la misma entidad.

Para poder tener competencia sobre estos asuntos financieros que se presenten entre las entidades y los clientes tendrían que tener la potestad por ley, y no la tienen, en este caso un posible conflicto entre las partes lo tendría que resolver las instancias judiciales correspondientes.

Por último la normativa emitida por la SUGEF, está orientada a las entidades supervisadas y no a sus clientes. Asimismo, dicha normativa tiene el objeto de cuantificar el riesgo de crédito de los deudores y constituir las estimaciones correspondientes con el fin de salvaguardar la estabilidad y solvencia de las entidades y conglomerados financieros.

### JURISPRUDENCIA INTERNACIONAL

La normativa internacional para las entidades financieras es muy amplia y tiene bases para verificar la solvencia de cada uno de los entes que en su entorno económico tiene el préstamo a clientes para inyectar recursos a la empresa.

Basilea es la normativa internacional con la que empresas financieras buscan seguir para obtener veracidad en sus actividades, el acuerdo busca dar las recomendaciones necesarias para que las entidades minimicen los riesgos financieros.

Basilea I, fue el primer pilar para unificar los criterios de los grandes bancos centrales a nivel mundial, Alemania, Bélgica, Canadá, España, EEUU, Francia, Japón, Italia, Reino Unido, Luxemburgo, Suecia, Suiza y Holanda, donde unieron los principales criterios para realizar una serie de recomendaciones para contar con un capital mínimo en función del riesgo a los que se exponen estas entidades.

Se estableció que las entidades contarán con capital regulatorio en dos sectores para ellos, una era si cumplen con requisitos de permanencia de capacidad de absorción en caso de pérdidas y una protección adecuada en caso de quiebra.

En este caso se puso mucho énfasis en el riesgo de crédito, se agrupaba en cinco categorías para poder evaluar a los clientes en una categoría específica.

Basilea I (Salvador Rayo Cantón, 2010) juega un papel importante en el fortalecimiento financiero mundial, ya que regularizó la solvencia entre los países miembros y del cual muchos otros países a nivel mundial lo han implementado.

Después de esto vendría Basilea II, que buscaba cubrir las debilidades que existían en la primera parte del acuerdo y así hacer más eficiente la normativa de las entidades. (Salvador Rayo Cantón, 2010)

Mucho más enfocado en variables de medición para la proyección de incumplimiento por parte de los clientes Basilea II busco que el riesgo minimizara a la hora de otorgar el capital.

Apoyándose en tres pilares fundamentales, se fundamento en el cálculo de los requisitos mínimos de capital, con el cual se obtiene los parámetros necesarios para evaluar a los clientes, en este primer pilar las estadísticas buscan cuatro elementos para que la veracidad de los datos sea lo más cercano a la exactitud a saber:

- PD (Probability Of Default), que se conoce como la probabilidad de incumplimiento que puede tener el cliente.
- LGD (Loss Given Default), que se define como la gravedad que puede generar en caso de incumplimiento.
- EAD (Exposure at Default), exposición en el momento de incumplimiento.
- M (Vencimiento efectivo), que es el plazo a vencer.

Basilea II dispone de opciones para que la institución financiera adopte el método que mejor le convenga, por ejemplo Método Estándar, Método Basado en Calificaciones Internas, entre otros.

El segundo Pilar de Basilea II está enfocado en el proceso de supervisión de la gestión de los fondos propios, según la teoría el acuerdo no solo busca garantizar que las entidades financieras cuenten con el capital necesario para cubrir riesgos en sus actividades de crédito, sino que también se cuente con la capacidad para que se desarrollen técnicas de gestión de riesgo en el cumplimiento de control incrementado el nivel de prudencia en las actividades crediticias que realizan.

Por último el tercer pilar, está enfocado en la disciplina de mercado, que es el conjunto de requisitos que posee el individuo a nivel de entidades, con las cuales se puede evaluar el comportamiento que se ha tenido a lo largo de otros procesos de crédito y de esta forma evaluar el riesgo.

Se definieron objetivos para verificar la veracidad y cumplir con la transparencia, se exigió la publicación periódica de la información acerca de las políticas de cada entidad, indicando la descripción de la gestión de riesgo, técnicas, objetivos, y todos los demás detalles relevantes en la regularización de los estudios de riesgo.

### III. METODOLOGÍA

El desarrollo de la investigación inicio por medio de la revisión literaria, en la cual se ampliaron los conocimientos y se identificaron los diferentes métodos utilizados para el desarrollo de un modelo de Credit Scoring.

Seguidamente se procedió a concretar una entrevista informal con un funcionario de una entidad bancaria donde los créditos personales tienen un alto porcentaje en la cartera de clientes que se maneja, con el fin de obtener la información y bases de datos necesaria para llevar a cabo el desarrollo de una propuesta de modelo.

Con dicha cartera se trabajó de forma detallada para determinar las mejoras a recomendar y las técnicas a implementar.

La metodología está enfocada en analizar cuáles son los principales modelos a estudiar, determinar cuáles son las variables más significativas y cuales podían ser tomadas en cuenta para aplicar en la institución financiera.

La investigación se enfocó en dos segmentos:

**a. Fase Exploratoria:** el equipo de trabajo se enfocó en recolectar toda la información necesaria, mediante la obtención de datos y análisis de los mismos obtenidos en bases de datos de revistas indexadas.

Adicionalmente se consultaron bases de datos históricas que proporcionó la entidad bancaria para llevar a cabo un análisis como ejemplo de casos, con el fin de conocer el comportamiento real de los clientes de crédito personal.

Con respecto a la base de datos obtenida por la institución financiera, fue necesario codificar aquellas variables en la cuales se contaba con datos textuales, para así salvaguardar la confidencialidad de cada uno de los clientes en la institución bancaria.

A continuación realizaremos una breve explicación de cada una de las variables utilizadas:

- Número cliente: Número por el cual la institución identifica a cada una de las personas que adquirieron un producto. Para efectos de este trabajo y por

cuestiones de seguridad bancaria se procedió a asignar a cada código cliente un número consecutivo iniciando con el uno hasta completar el total de registros con la finalidad de resguardar el secreto bancario y cumplir con la solicitud por parte de la institución de no divulgar información confidencial.

- **Agencia:** Se trata de las diferentes agencias que posee la institución a nivel nacional, en esta parte se codifico por sectores, ya que se contaba con gran cantidad de agencias y con esto se buscó simplificar esta variable para la aplicación del modelo.
- **Género:** Identifica si el cliente solicitante es hombre o mujer, con lo cual se evaluara cual es el porcentaje de incumplimiento de ambos.

VARIABLE	CODIFICACIÓN
Hombre	1
Mujer	2

- **Edad:** La edad que tiene cada uno de los clientes solicitantes. Para el desarrollo del modelo fue necesario agruparlo o clasificarlo en rango.

MIN	MAX
19	26
26	29
29	33
33	36
36	40
40	44
44	48
48	52
52	58
58	87

- **Estado Civil:** Corresponde a la evaluación de los individuos tomando en cuenta si son casados, solteros, viudos, divorciados o en unión libre.

VARIABLE	CODIFICACIÓN
Solteros	1
Casado	2
Divorciado	3
Viudo	4
Unión Libre	5
Separado	6

- Escolaridad: Para esta variable utilizaremos primaria, secundaria y universidad dependiendo del grado académico de la persona.

VARIABLE	CODIFICACIÓN
Primaria	1
Secundaria	2
Universidad	3

- Moneda: La institución financiera ha otorgado créditos en tres modalidades:

VARIABLE	CODIFICACIÓN
Colones	1
Dólares	2
UD's	3

Esta última fue una especie de moneda virtual creada por el Banco Central de Costa Rica y está basada en la inflación. A pesar que ya no se utiliza, aún existen clientes pagando su préstamo en este tipo de financiamiento.

- Sector Económico: Muestra el sector al que se encuentra dirigido la persona física o jurídica que está solicitando el crédito. Para efectos de esta investigación únicamente se consideraron personas físicas, las cuales pueden trabajar como asalariadas o propietarios de empresas, por lo que de igual forma estarían dirigidas a un sector económico.

CATEGORIA
ASALARIADO
COMERCIO
PENSIONADO
SERVICIOS
SERVICIOS PROFESIONALES
TRANSPORTISTA
ENSEÑANZA
CONSTRUCCIÓN
INTERMEDIACIÓN FINANCIERA
INDUSTRIA
ACTIVIDADES INMOBILIARIAS, EMPRESARIALES Y DE ALQUILER
AGRICULTURA Y GANADERIA

- Ingreso: Muestra por medio de rangos el nivel de ingreso de cada individuo clasificado de la siguiente forma:

VARIABLE	CODIFICACIÓN
\$1-\$1.000	1
\$1.001-\$2.500	2
\$2.501-\$5.000	3
\$5.001-\$7.500	4
\$7.501- a más	5

- Plazo, tasa nominal, monto solicitado: Son variables características de cada uno de los préstamos otorgados.
- Categoría: Corresponde a la calificación que la SUGEF brinda de una forma individual a los deudores, considerando la categoría de riesgo A1 de menor riesgo de crédito y la categoría E a la de mayor riesgo. De una forma más detallada se puede observar:

CATEGORÍA DE RIESGO	MOROSIDAD	COMPORTAMIENTO DE PAGO HISTÓRICO	CAPACIDAD DE PAGO
<b>A1</b>	Igual o menor a 30 días	Nivel 1	Nivel 1
<b>A2</b>	Igual o menor a 30 días	Nivel 2	Nivel 1
<b>B1</b>	Igual o menor a 60 días	Nivel 1	Nivel 1 o Nivel 2
<b>B2</b>	Igual o menor a 60 días	Nivel 2	Nivel 1 o Nivel 2
<b>C1</b>	Igual o menor a 90 días	Nivel 1	Nivel 1 o Nivel 2 o Nivel 3
<b>C2</b>	Igual o menor a 90 días	Nivel 1 o Nivel 2	Nivel 1 o Nivel 2 o Nivel 3
<b>D</b>	Igual o menor a 12 <sup>o</sup> días	Nivel 1 o Nivel 2	Nivel 1 o Nivel 2 o Nivel 3 o Nivel 4

- **Garantía:** Corresponde al bien que el sujeto deudor esta dejando como respaldo para el pago de la deuda.

VARIABLE	CODIFICACIÓN
Fianza	1
Fideicomiso	2
Hipoteca	3
Pagaré	4
Prenda	5
T Valores	6

- **Incumplimiento:** Esta variable refleja si el sujeto ha cumplido con el pago de sus cuotas o no, por lo que 0 se considera bueno, es decir, que si lleva al día el pago de su deuda, mientras que 1 es incumplimiento.

- b. Fase Descriptiva:** se procedió a realizar una descripción de una forma clara de cada uno de los hallazgos de la fase exploratoria, ubicando los datos más significativos y así explicándolos de una forma concisa, con el fin de obtener respuestas a las interrogantes y a su vez poder remitir algún tipo de recomendación al banco o bien a futuros estudios de investigación.
- c. Técnicas de investigación**

Se siguieron con las siguientes técnicas de investigación para la recolección de datos:

- Obtención de una base de datos con el historial de los clientes que han solicitado créditos y que se les aprobaron, proporcionados por una entidad bancaria.
- Con la anterior base de datos se seleccionaron los individuos que pertenecen a la categoría seleccionada, y se aplicó a cada caso el modelo de Credit Scoring definido.
- Se realizó consultas a expertos, que aportaran información a fin de triangular los resultados.

Para el modelo en estudio se procedió a darle un valor específico a las diferentes variables, con el propósito de buscar identificar las mismas de acuerdo a las necesidades y comportamiento de los individuos, ejemplo, sexo, estado civil, edad, todo cuenta con valores de medición.

Se realizó para los análisis diferentes tipos de medición Logit, por ejemplo para los datos de sexo se utilizó el Logit dicotómico, con la cual utilizamos solo dos respuestas para las variables, que son hombre o mujer, en ese caso se le otorgó un valor específico a cada uno.

Con respecto a datos de varias respuestas se utilizó el Logit de respuesta múltiple con la cual se analizaron más de dos respuestas, por ejemplo en el estado civil de los clientes se contaba con datos como solteros, casados, divorciaos, unión libre, entre otros.

## IV. ANÁLISIS DE LOS DATOS

### A. MODELO LOGIT

#### 1. Obtención y depuración de la Base de datos

Se procedió a solicitar a la entidad bancaria la base de datos de clientes que solicitaron préstamos, así como sus principales variables, una vez que se obtuvo dicha información se realizó la validación de los principales.

De una base de diecisiete mil ciento noventa y nueve datos, fue necesario validar solo los que eran de interés para el modelo, es decir, aquellos datos que representaran un segmento de consumo para personas físicas. Después de realizar un exhaustivo trabajo se minimizó la base a seis mil seiscientos diecinueve personas que si eran de interés para la elaboración del modelo de Credit Scoring.

Al disminuir de esta forma la cantidad de datos, se considero que no era necesario seleccionar una muestra para poder desarrollar el modelo, por lo que se trabajo con el 100% de los datos, además, se consideró que dicha base no contaba con lo que se conoce como Dummies (datos con errores o celdas vacías), es decir, aquellos datos que en su registro cuentan con la numeración 999, lo que significa que no se conoce la información real, por ejemplo, si un cliente cuenta con 999 en su estado civil quiere decir que no se sabe si es soltero, casado, divorciado, etc.

Para el desarrollo del trabajo se escogió un modelo elaborado mediante una hoja de cálculo en Excel, enfocándose en el modelo Logit para obtener el porcentaje de posibles clientes que estadísticamente nos den los resultados como malos pagadores, de esa forma se podría evaluar el riesgo inherente que generaría aceptarle un crédito a un cliente que en un futuro podría no pagar.

#### 2. Análisis de variables para determinar cuáles son significativas

Luego de depurada y definida la base de datos se procedió a desarrollar el Análisis Univariable, método por el cual se determina cuáles son las variables significativas para utilizar en el modelo del Credit Scoring y cuales se deben descartar debido a que no aportan un valor significativo al mismo.

Para esto se procedió a evaluar cada variable y distribuirla en rangos de clase, en algunos casos fue necesario dividir la variable en percentiles para definir el nivel óptimo de clases dentro de la variable, tal es el caso de las variables edad y monto

solicitado en los cuales se utilizaron diez rangos de diez percentiles cada uno o lo que es igual se dividió la variable en deciles, para luego definir la cantidad de clases a utilizar, de igual manera se le aplicó dicho procedimiento a las variables, ingreso, tasa y plazo, determinando así los límites superiores e inferiores de cada rango de clase para cada variable. Para el resto de las variables no fue necesario dicho proceso por cuanto se trata de variables categóricas.

Una vez obtenida dicha información se procedió a calcular el total de clientes que pertenecen a cada rango, así como los clientes clasificados como malos y buenos de la cual deriva la probabilidad default.

Seguidamente se procedió a calcular el WOE en donde cada decisión se basa en la probabilidad de que ocurra algún evento. Se evalúan las circunstancias y se determina el peso de la evidencia. El WOE se utiliza para evaluar el riesgo relativo de las diferentes características de una variable.

La fórmula WOE cuenta con dos componentes: una parte variable que evalúa la probabilidad de pertenecer a un grupo y una parte constante asociados a la población. Un WOE negativo indica que la proporción de valores de default es mayor para esa variable respecto a la proporción global e indica un mayor riesgo.

$$WOE = \ln\left(\frac{N/P}{(\sum N / \sum P)}\right)$$

Donde P es el número de ocurrencias, es decir, los clientes buenos y N es el número de no ocurrencias o clientes malos.

El WOE no considera la proporción de observaciones de cada variable analizada, sólo el riesgo relativo, por lo que se deben aplicar herramientas adicionales para determinar la contribución relativa de cada variable, como Information Value.

El Information Value se utiliza para clasificar las variables de orden en cuanto a su capacidad de predicción. Este concepto se extiende al Credit Scoring observando la no ocurrencia como los no-defaults (buenos) y la incidencia como valores predeterminados (malos). Un alto Information Value indica una alta capacidad de discriminar.

Los valores para el Information Value siempre serán positivos y puede tomar valores superiores a 3, lo que implica que es una variable altamente predictiva. Las variables con Information Value menores a 0,10 se consideran generalmente como débiles, mientras que valores superiores a 0,30 deben ser considerados en la utilización del modelo, sin embargo, este estadístico es muy extremista al

asumir que todo es bueno o todo es malo, por lo que se debe combinar con otros estadísticos como el Kolmogorov Smirnov y el GINI.

El estadístico KS se utiliza para comparar una variable desconocida contra la distribución observada y la distribución teórica. La distancia máxima entre las distribuciones acumulativas se calcula y se mide contra un valor crítico. Si la distancia máxima es menor que el valor crítico, hay una buena probabilidad de que las distribuciones sean iguales obteniendo una buena aproximación al pronóstico. En el Scoring se utiliza como una herramienta de visualización de datos para ilustrar la eficacia del modelo. Las distribuciones acumulativas de buenos y malos se representan frente a la puntuación.

El GINI es un método gráfico que ilustra el grado de desigualdad en la distribución de la variable en cuestión. Sin embargo, Gini también se puede utilizar para determinar la potencia de una variable individual. Las características de la variable se agrupan de la mejor a la peor situación de riesgo, mediante el establecimiento de un punto de corte, según el modelo que se construirá, y las variables se pueden elegir para su inclusión en el proceso de modelado en base a este punto de corte.

Las técnicas utilizadas para el análisis bivariado determinarán los criterios de inclusión para el análisis multivariado. Cualquier variable cuyo bivariado tenga un valor menor al 0,25 es un candidato para el modelo multivariado, junto con todas las variables de importancia. No hay un límite "estadístico" para el índice de Gini, sin embargo, un índice inferior al 10%, es considerado débil, mientras que uno por arriba del 15% indica que la variable debe ser considerada en el modelo.

Este procedimiento se utilizó con cada una de las variables obteniendo los siguientes resultados:

**Edad:** para la definición de los rangos de clase de esta variable se realizó una división de la variable en diez rangos utilizando la técnica de percentiles, para posteriormente darle normalidad a la serie de datos mediante la optimización de los rangos.

### EDAD

CONSIDERANDO PERCENTIL 10% PARA CADA TRAMO

$$IV = \sum_{i=1}^n \left( \frac{[BUENOS]_{Ti} - [MALOS]_{Ti}}{[BUENOS]_{Ti} + [MALOS]_{Ti}} \right)^n$$

$$G = \left| 1 - \sum_{k=1}^{n-1} (X_{k+1} - X_k)(Y_{k+1} + Y_k) \right|$$

TRAMO	MIN	MAX	TOTAL	MALOS	BUENOS	PD	WOE ESPAÑOL	WOE AMERICANO	% MALOS	% BUENOS	IV	ACUMULADO MALOS	ACUMULADO BUENOS	ACUMULADO MALOS %	ACUMULADO BUENOS %	KS	Abi%+ Abi%-1	AMi%- AMi%-1	GINI
1	19	26	566	95	471	16,78%	0,701603891	-0,701603889	15,10%	7,49%	0,05	95	471	15,10%	7,49%	7,62%	7,49%	15,10%	1,13%
2	26	29	606	74	532	12,21%	0,330006697	-0,33000667	11,76%	8,46%	0,01	169	1003	26,87%	15,95%	10,92%	23,43%	11,76%	2,76%
3	29	33	816	91	725	11,15%	0,227272945	-0,22727294	14,47%	11,53%	0,01	260	1728	41,34%	27,47%	13,86%	43,42%	14,47%	6,28%
4	33	36	614	74	540	12,05%	0,315081047	-0,31508105	11,76%	8,59%	0,01	334	2268	53,10%	36,06%	17,04%	63,53%	11,76%	7,47%
5	36	40	696	59	637	8,48%	-0,07664712	0,076647119	9,38%	10,13%	0,00	393	2905	62,48%	46,18%	16,30%	82,24%	9,38%	7,71%
6	40	44	712	66	646	9,27%	0,021440331	-0,02144033	10,49%	10,27%	0,00	459	3551	72,97%	56,45%	16,52%	102,64%	10,49%	10,77%
7	44	48	732	41	691	5,60%	-0,52198266	0,521982664	6,52%	10,99%	0,02	500	4242	79,49%	67,44%	12,05%	123,90%	6,52%	8,08%
8	48	52	661	43	618	6,51%	-0,36270325	0,362703249	6,84%	9,83%	0,01	543	4860	86,33%	77,27%	9,06%	144,71%	6,84%	9,89%
9	52	58	766	56	710	7,31%	-0,23732819	0,237328186	8,90%	11,29%	0,01	599	5570	95,23%	88,55%	6,68%	165,82%	8,90%	14,76%
10	58	87	750	30	720	4,00%	-0,87546874	0,875468737	4,77%	11,45%	0,06	629	6290	100,00%	100,00%	0,00%	188,55%	4,77%	8,99%
TOTAL			6919	629	6290						IV	18%			KS	17,04%		GINI	22,15%

### EDAD TRAMAO OPTIMIZADO

TRAMO	MIN	MAX	TOTAL	MALOS	BUENOS	PD	WOE ESPAÑOL	WOE AMERICANO	% MALOS	% BUENOS	IV	ACUMULADO MALOS	ACUMULADO BUENOS	ACUMULADO MALOS %	ACUMULADO BUENOS %	KS	Abi%+ Abi%-1	AMi%- AMi%-1	GINI	
1	19	26	566	95	471	16,78%	0,701285875	-0,70128587	15,10%	7,49%	0,05	95	471	15,10%	7,49%	7,61%	7,49%	15,10%	1,13%	
2	26	29	606	74	532	12,21%	0,329688681	-0,32968868	11,76%	8,46%	0,01	169	1003	26,87%	15,95%	10,92%	23,44%	11,76%	2,76%	
3	29	36	1430	165	1265	11,54%	0,265385150	-0,265385150	26,23%	20,12%	0,02	334	2268	53,10%	36,07%	17,03%	52,02%	26,23%	13,65%	
4	36	44	1408	125	1283	8,88%	-0,026375550	0,026375550	19,87%	20,40%	0,00	459	3551	72,97%	56,47%	16,50%	92,54%	19,87%	18,39%	
5	44	58	2159	140	2019	6,48%	-0,36644812	0,366448118	22,26%	32,11%	0,04	599	5570	95,23%	88,58%	6,65%	145,05%	22,26%	32,29%	
6	58	87	750	30	718	4,00%	-0,873005110	0,873005110	4,77%	11,42%	0,06	629	6288	100,00%	100,00%	0,00%				
TOTAL			6919	629	6288							IV	12%			KS	17,03%		GINI	31,79%

Se obtuvo la cantidad y los porcentajes de clientes buenos y malos sobre el cual se calcula la probabilidad de default, así como el riesgo de cada uno que sería evaluado posteriormente. Por tratarse de datos continuos se procedió a “suavizar” los mismos de tal manera que generaran mediante optimización datos más certeros.

Una vez optimizada la variable, los datos nos revelan que el mayor riesgo lo generan los clientes que están en el primer tramo que va de los 19 años hasta los 26.

Contrario a lo anterior, se encuentra el sexto tramo el cual nos indica que los clientes que van de 58 a 87 años poseen un porcentaje de riesgo menor en comparación a los otros tramos.

El WOE español nos indica el peso que posee cada rango de clase dentro de la variable y el signo de la misma, en el caso de los pesos positivos indica que cuanto menor edad mayor es el riesgo y viceversa.

Para definir la significancia estadística se basó el análisis en el teorema de GINI, el cual plantea que cuando la variable cuenta con un índice superior al 15% debe ser considerada dentro del modelo, para el caso de la variable edad esta arrojó un resultado de 31.79% por lo que bajo el análisis univariable sería considerada dentro del modelo logit.

**Estado Civil:** los resultados obtenidos en este apartado de medición generaron que la mayor cantidad de clientes corresponde a clientes que son casados o solteros, los otros poseen un porcentaje menor de participación dentro de la población en estudio.

#### ESTADO CIVIL

TRAMO	CATEGORIA	TOTAL	MALOS	BUENOS	PD	WOESPANOL	WOE AMERICANO	% MALOS	% BUENOS	IV	ACUMULADO MALOS	ACUMULADO BUENOS	ACUMULADO MALOS %	ACUMULADO BUENOS %	KS	Abi%+ Abi%-1	AMi%- AMi%-1	GINI	
1	SOLTERO	2247	268	1979	11,93%	0,303225128	-0,30322513	42,61%	31,47%	3,37%	268	1979	42,61%	31,47%	11,13%	31,47%	42,61%	13,41%	
2	CASADO	3752	273	3479	7,28%	-0,24244329	0,242443287	43,40%	55,33%	2,90%	273	3479	43,40%	55,33%	11,93%	55,33%	43,40%	24,01%	
3	DIVORCIADO	457	36	421	7,88%	-0,1565288	0,156528802	5,72%	6,70%	0,15%	36	421	5,72%	6,70%	0,97%	6,70%	5,72%	0,38%	
4	VIUDO	85	5	80	5,88%	-0,47000363	0,470003629	0,79%	1,27%	0,22%	5	80	0,79%	1,27%	0,48%	1,27%	0,79%	0,01%	
5	U. LIBRE	269	35	234	13,01%	0,402612039	-0,40261204	5,56%	3,72%	0,74%	35	234	5,56%	3,72%	1,84%	3,72%	5,56%	0,21%	
6	SEPARADO	109	12	97	11,01%	0,212780764	-0,21278076	1,91%	1,54%	0,08%	12	97	1,91%	1,54%	0,37%	1,54%	1,91%	0,03%	
TOTAL		6919	629	6290						IV	7,46%				KS	11,93%		GINI	61,95%

Los datos generados indican que el mayor riesgo está asociado a los clientes que tienen la condición de unión libre, según el modelo scoring estos clientes tienen un riesgo más alto de que incumplan en sus obligaciones en comparación con los casados, por ejemplo, es importante indicar que los separados y solteros comparten también un riesgo alto que en el momento de aplicar el modelo en la etapa multivariable generaría cambios.

El GINI mostró un índice de 61.95% siendo una de las variables más altas que se generaron en el modelo univariable, lo que implica alta correlación entre el estado civil y la variable default, por lo que el estado civil es una variable significativa a considerar en el modelo.

**Escolaridad:** de acuerdo a la codificación de la variable se tienen tres elementos de medición, los cuales muestran la particularidad de contar con una distribución en cantidad de personas muy similar, permitiendo un rico análisis respaldado por la cantidad de unidades evaluadas y basados en los WOES se desprende que el sector de mayor riesgo asociado son lo que cuentan con un nivel de educación de primaria, caso contrario a los universitarios que tienen mejor comportamiento de pago.

### GRADO DE ESCOLARIDAD

TRAMO	CATEGORIA	TOTAL	MALOS	BUENOS	PD	WOE ESPANOL	WOE AMERICANO	% MALOS	% BUENOS	IV	ACUMULADO MALOS	ACUMULADO BUENOS	ACUMULADO MALOS %	ACUMULADO BUENOS %	KS	Abi%+ Abi%-1	AMi%-1 AMi%+1	GINI
1	PRIMARIA	1757	193	1564	10,98%	0,210273361	-0,20995535	30,68%	24,87%	1,22%	193	1564	30,68%	24,87%	5,81%	24,87%	30,68%	7,63%
2	SECUNDARIA	2436	292	2144	11,99%	0,308910373	-0,30859236	46,42%	34,10%	3,80%	485	3708	77,11%	58,97%	18,14%	83,84%	46,42%	38,92%
3	UNIVERSIDAD	2726	144	2582	5,28%	-0,58392118	0,584239194	22,89%	41,06%	10,61%	629	6290	100,00%	100,03%	0,03%	159,00%	22,89%	36,40%
TOTAL		6919	629	6290						IV					KS			GINI
										15,64%					18,14%			17,05%

Para definir la significancia estadística, se basó el análisis en el teorema de GINI, el cual cuenta con un índice del 17.05%, por lo que la variable debe ser considerada dentro del modelo logit.

**Género:** con el estudio de esta variable se obtuvo varios datos de relevancia, uno de ellos es que según la base hay un porcentaje mayor de hombres que solicitan préstamos personales, sin embargo, para efectos del modelo lo que interesa es que las mujeres muestran un mejor comportamiento de pago respecto a los hombres, lo que evidencia que según el scoring el riesgo de prestarle a un hombre es mayor que el de una mujer.

### GENERO

TRAMO	CATEGORIA	TOTAL	MALOS	BUENOS	PD	WOE ESPANOL	WOE AMERICANO	% MALOS	% BUENOS	IV	ACUMULADO MALOS	ACUMULADO BUENOS	ACUMULADO MALOS %	ACUMULADO BUENOS %	KS	Abi%+ Abi%-1	AMi%-1 AMi%+1	GINI
1	HOMBRE	4725	453	4272	9,59%	0,058639838	-0,05863984	72,02%	67,92%	0,24%	453	4272	72,02%	67,94%	4,08%	67,94%	72,02%	48,93%
2	MUJER	2194	176	2018	8,02%	-0,13679311	0,136793113	27,98%	32,08%	0,56%	629	2018	100,00%	32,09%	67,91%	32,09%	27,98%	8,98%
TOTAL		6919	629	6290						IV					KS			GINI
										0,80%					67,91%			42,09%

Para esta variable el índice GINI presenta un porcentaje de 42.09%, lo que hace que la variable de género sea considerada dentro del modelo.

**Moneda:** la moneda de preferencia por parte de los clientes que solicitan créditos personales para consumo en la institución financiera es la de colones, que representan un alto porcentaje del préstamo, después en menor cantidad se reflejan los préstamos en dólares y unidades de desarrollo, esto evidencia la preferencia de utilizar moneda local en las transacciones.

### MONEDA

TRAMO	CATEGORIA	TOTAL	MALOS	BUENOS	PD	WOE ESPANOL	WOE AMERICANO	% MALOS	% BUENOS	IV	ACUMULADO MALOS	ACUMULADO BUENOS	ACUMULADO MALOS %	ACUMULADO BUENOS %	KS	Abi%+ Abi%-1	AMi%-1 AMi%+1	GINI
1	COLONES	6663	621	6042	9,32%	0,027425813	-0,02742581	98,73%	96,06%	0,07%	621	6042	98,73%	96,09%	2,64%	96,09%	98,73%	94,87%
2	DOLARES	224	5	219	2,23%	-1,47704872	1,477048724	0,79%	3,48%	-3,97%	626	6261	99,52%	99,57%	0,05%	195,66%	0,79%	1,56%
3	UD'S	32	3	29	9,38%	0,033901552	-0,03390155	0,48%	0,46%	0,00%	629	6290	100,00%	100,03%	0,03%	199,60%	0,48%	0,95%
TOTAL		6919	629	6290						IV					KS			GINI
										4,04%					2,64%			2,63%

Un dato a rescatar de esta variable es el buen comportamiento de pago que tienen los solicitantes de crédito en moneda extranjera, lo que hace que cuenten con un riesgo menor en el modelo después de generar los datos, sin embargo, la preferencia de los clientes es mayor en colones, estando estos créditos asociados a un mayor riesgo.

Por último el GINI nos indica un porcentaje de 2.63%, el cual no se ajusta al teorema, por lo que la dicha variable se descarta para la definición del modelo final.

**Sector Económico:** como se detalló anteriormente el sector económico fue codificado para agrupar los sectores similares o cuya razón social es la misma, para un mejor tratamiento de los datos los resultados que se obtuvieron demuestran que el sector de clientes asalariados representan un porcentaje de riesgo mayor, por lo que se debería tener mayor cuidado a la hora de otorgar préstamos.

#### SECTOR ECONÓMICO

TRAMO	CATEGORIA	TOTAL	MALOS	BUENOS	PD	WOESPANOL	WOE AMERICANO	% MALOS	% BUENOS	IV	ACUMULADO MALOS	ACUMULADO BUENOS	ACUMULADO MALOS %	ACUMULADO BUENOS %	KS	Abi%+ Abi%-1	AMi%- AMi%-1	GINI
1	ASALARIADO	256	45	211	17,58%	0,757389449	-0,75738945	7,15%	3,35%	2,88%	45	211	7,15%	3,35%	3,80%	3,35%	7,15%	0,24%
2	COMERCIO	1660	194	1466	11,69%	0,28015037	-0,28015037	30,84%	23,31%	2,11%	239	1677	38,00%	26,66%	11,34%	30,02%	30,84%	9,26%
3	PENSIONADO	116	3	113	2,59%	-1,32619044	1,326190437	0,48%	1,80%	1,75%	242	1790	38,47%	28,46%	10,02%	55,12%	0,48%	0,26%
4	SERVICIOS	2175	215	1960	9,89%	0,092523369	-0,09252337	34,18%	31,16%	0,28%	457	3750	72,66%	59,62%	13,04%	88,08%	34,18%	30,11%
5	SERVICIOS PROFESIONALES	425	32	393	7,53%	-0,20548862	0,205488616	5,09%	6,25%	0,24%	489	4143	77,74%	65,87%	11,88%	125,48%	5,09%	6,38%
6	TRANSPORTISTA	375	41	334	10,93%	0,205016167	-0,20501617	6,52%	5,31%	0,25%	530	4477	84,26%	71,18%	13,08%	137,04%	6,52%	8,93%
7	ENSEÑANZA	180	18	162	10,00%	0,105360516	-0,10536052	2,86%	2,58%	0,03%	548	4639	87,12%	73,75%	13,37%	144,93%	2,86%	4,15%
8	CONSTRUCCIÓN	212	12	200	5,66%	-0,51082562	0,510825624	1,91%	3,18%	0,65%	560	4839	89,03%	76,93%	12,10%	150,68%	1,91%	2,87%
9	INTERMEDIACIÓN FINANCIERA	798	9	789	1,13%	-2,17095665	2,170956651	1,43%	12,54%	24,13%	569	5628	90,46%	89,48%	0,99%	166,41%	1,43%	2,38%
10	INDUSTRIA	319	30	289	9,40%	0,037355787	-0,03735579	4,77%	4,59%	0,01%	599	5917	95,23%	94,07%	1,16%	183,55%	4,77%	8,75%
11	ACTIVIDADES INMOBILIARIAS, EMPRESARIALES Y DE SERVICIOS	109	11	98	10,09%	0,115512887	-0,11551289	1,75%	1,56%	0,02%	610	6015	96,98%	95,63%	1,35%	189,70%	1,75%	3,32%
12	AGRICULTURA Y GANADERIA	294	19	275	6,46%	-0,36974703	0,369747026	3,02%	4,37%	0,50%	629	6280	100,00%	100,00%	0,00%	195,63%	3,02%	5,91%
TOTAL		6919	629	6290						IV	32,84%			KS	13,37%		GINI	17,43%

Para definir la significancia estadística, se basó el análisis en el teorema de GINI, el cual plantea que cuando la variable cuenta con un índice superior al 15% la variable debe ser considerada dentro del modelo, para el caso de la variable sector económico arrojó un resultado de 17,43% por lo que bajo el análisis univariable sería considerada dentro del modelo logit.

**Ingreso:** evaluado por rangos de clase para mayor factibilidad de estudio el ingreso determina una variable importante. Se puede observar que entre menor ingreso mayor riesgo, lo que genera un factor determinante en la medición ya que dependiendo del ingreso y el monto solicitado se genera la relación de riesgo en el modelo.

**INGRESO**

TRAMO	MIN	MAX	TOTAL	MALOS	BUENOS	PD	WOESPANOL	WOE AMERICANO	% MALOS	% BUENOS	IV	ACUMULADO MALOS	ACUMULADO BUENOS	ACUMULADO MALOS %	ACUMULADO BUENOS %	KS	Abi%+ Abi%-1	AMi%- AMi%-1	GINI	
1	\$ 1,00	\$ 1.000,00	3680	445	3235	12,09%	0,318875168	-0,31887517	70,75%	51,43%	6,16%	445	3235	70,75%	51,43%	19,32%	51,43%	70,75%	36,39%	
2	\$ 1.001,00	\$ 2.500,00	2225	128	2097	5,75%	-0,49364767	0,493647674	20,35%	33,34%	6,41%	573	5332	91,10%	84,77%	6,33%	136,20%	20,35%	27,72%	
3	\$ 2.501,00	\$ 5.000,00	679	33	646	4,86%	-0,67170685	0,671706849	5,25%	10,27%	3,37%	606	5978	96,34%	95,04%	1,30%	179,81%	5,25%	9,43%	
4	\$ 5.001,00	\$ 7.500,00	152	7	145	4,61%	-0,7282385	0,7282385	1,11%	2,31%	0,87%	613	6123	97,46%	97,34%	0,11%	192,38%	1,11%	2,14%	
5	\$ 7.500,00	más	183	16	167	8,74%	-0,04282	0,042819997	2,54%	2,66%	0,00%	629	6290	100,00%	100,00%	0,00%	197,34%	2,54%	5,02%	
TOTAL			6919	629	6290						IV	16,82%				KS	19,32%		GINI	19,30%

La variable de ingreso una vez aplicada en el modelo univariable nos genera un GINI de 19.30%, la misma fue generada como factible para ser aplicada en la segunda parte del estudio.

**Tasas:** la variable de tasa nominal es uno de los elementos considerados en la evaluación del Credit Scoring en el modelo Logit, ya que es una variable que puede generar que el cliente sea bueno o malo por el tipo de porcentaje y en combinación con los demás factores.

**TASAS**

TRAMO	MIN	MAX	TOTAL	MALOS	BUENOS	PD	WOESPANOL	WOE AMERICANO	% MALOS	% BUENOS	IV	ACUMULADO MALOS	ACUMULADO BUENOS	ACUMULADO MALOS %	ACUMULADO BUENOS %	KS	Abi%+ Abi%-1	AMi%- AMi%-1	GINI	
1	2,27	10	545	8	537	1,47%	-1,90397146	1,90397146	1,27%	8,54%	0,14	8	537	1,27%	8,54%	7,27%	8,54%	1,27%	0,11%	
2	10	15	2459	92	2367	3,74%	-0,94500494	0,945004939	14,63%	37,63%	0,22	100	2904	15,90%	46,17%	30,27%	54,71%	14,63%	8,00%	
3	15	18	514	48	466	9,34%	0,02960047	-0,02960047	7,63%	7,41%	0,00	148	3370	23,53%	53,58%	30,05%	99,75%	7,63%	7,61%	
4	18	32	3401	481	2920	14,14%	0,499113468	-0,49911347	76,47%	46,42%	0,15	629	6290	100,00%	100,00%	0,00%	153,58%	76,47%	117,44%	
TOTAL			6919	629	6290						IV	51%				KS	30,27%		GINI	33,16%

En la primera etapa de la medición se determinó que entre mayor porcentaje tenga la tasa nominal, mayor será el riesgo de que los clientes no cumplan con las obligaciones.

En este apartado el GINI obtuvo un 33.16%, lo cual indica que es una variable que si debe ser tomada en cuenta en el modelo.

**Monto Solicitado:** En este caso se verifico que entre menor sea el préstamo solicitado por parte del cliente el riesgo de morosidad es mayor, esto se podría explicar de forma sencilla que al ser montos bajos los clientes no le toman la importancia.

**MONTO SOLICITADO**

TRAMO	MIN	MAX	TOTAL	MALOS	BUENOS	PD	WOE ESPANOL	WOE AMERICANO	% MALOS	% BUENOS	IV	ACUMULADO MALOS	ACUMULADO BUENOS	ACUMULADO MALOS %	ACUMULADO BUENOS %	KS	Abi%+ Abi%-1	AMi%- AMi%-1	GINI
1	110.000,0	750.000,00	664	81	583	12,20%	0,328847061	-0,32884706	12,88%	9,27%	0,01	81	583	12,88%	9,27%	3,61%	9,27%	12,88%	1,19%
2	750.000,00	990.000,00	717	126	591	17,57%	0,757050983	-0,75705098	20,03%	9,40%	0,08	207	1174	32,91%	18,66%	14,24%	27,93%	20,03%	5,60%
3	990.000,00	1.250.000,00	681	108	573	15,86%	0,633830603	-0,6338306	17,17%	9,11%	0,05	315	1747	50,08%	27,77%	22,31%	46,44%	17,17%	7,97%
4	1.250.000,00	2.000.000,00	702	79	623	11,25%	0,237486427	-0,23748643	12,56%	9,90%	0,01	394	2370	62,64%	37,68%	24,96%	65,45%	12,56%	8,22%
5	2.000.000,00	3.000.000,00	676	39	637	5,77%	-0,49062292	0,490622916	6,20%	10,13%	0,02	433	3007	68,84%	47,81%	21,03%	85,48%	6,20%	5,30%
6	3.000.000,00	4.500.000,00	683	34	649	4,98%	-0,6464871	0,646487099	5,41%	10,32%	0,03	467	3656	74,24%	58,12%	16,12%	105,93%	5,41%	5,73%
7	4.500.000,00	6.516.400,00	720	26	694	3,61%	-0,98179033	0,981790329	4,13%	11,03%	0,07	493	4350	78,38%	69,16%	9,22%	127,28%	4,13%	5,26%
8	6.516.400,00	10.000.000,00	574	21	553	3,66%	-0,96825047	0,968250471	3,34%	8,79%	0,05	514	4903	81,72%	77,95%	3,77%	147,11%	3,34%	4,91%
9	10.000.000,00	16.721.635,00	810	54	756	6,67%	-0,33647224	0,336472237	8,59%	12,02%	0,01	568	5659	90,30%	89,97%	0,33%	167,92%	8,59%	14,42%
10	16.721.635,00	777.900.000,00	692	61	631	8,82%	-0,03384691	0,033846905	9,70%	10,03%	0,00	629	6290	100,00%	100,00%	0,00%	189,97%	9,70%	18,42%
TOTAL			6919	629	6290						IV				KS	24,96%		GINI	22,98%

**MONTO SOLICITADO TRAMADO OPTIMIZADO**

TRAMO	MIN	MAX	TOTAL	MALOS	BUENOS	PD	WOE ESPANOL	WOE AMERICANO	% MALOS	% BUENOS	IV	ACUMULADO MALOS	ACUMULADO BUENOS	ACUMULADO MALOS %	ACUMULADO BUENOS %	KS	Abi%+ Abi%-1	AMi%- AMi%-1	GINI
1	110.000,00	1.250.000,00	2062	315	1747	15,28%	0,589184406	-0,58918441	50,08%	27,78%	0,13	315	1747	50,08%	27,78%	22,30%	27,78%	50,08%	13,91%
2	1.250.000,00	2.000.000,00	702	79	623	11,25%	0,237168411	-0,23716841	12,56%	9,91%	0,01	394	2370	62,64%	37,69%	24,95%	65,47%	12,56%	8,22%
3	2.000.000,00	3.000.000,00	676	39	637	5,77%	-0,49094093	0,490940932	6,20%	10,13%	0,02	433	3007	68,84%	47,82%	21,02%	85,51%	6,20%	5,30%
4	3.000.000,00	777.900.000,00	3479	196	3282	5,63%	-0,51582654	0,515826535	31,16%	52,19%	0,11	629	6289	100,00%	100,02%	0,02%	147,84%	31,16%	46,07%
TOTAL			6919	629	6289						IV				KS	24,95%		GINI	26,49%

El teorema de GINI indica que hay significancia estadística para esta variable, al contar con un índice superior al 15% por lo que debe ser considerada dentro del modelo, el resultado es de 26.49%.

**Garantía:** los resultados que se obtuvieron en cuanto a garantía después de aplicar el modelo univariable es que el fideicomiso es el más riesgoso, lo que genera que los clientes que utilizan este medio sean potencialmente propensos a fallar en sus pagos en algún momento.

**GARANTIA**

TRAMO	CATEGORIA	TOTAL	MALOS	BUENOS	PD	WOE ESPANOL	WOE AMERICANO	% MALOS	% BUENOS	IV	ACUMULADO MALOS	ACUMULADO BUENOS	ACUMULADO MALOS %	ACUMULADO BUENOS %	KS	Abi%+ Abi%-1	AMi%- AMi%-1	GINI	
1	FIANZA	1258	84	1174	6,68%	-0,33477011	0,334770109	13,35%	18,66%	1,78%	84	1174	13,35%	18,66%	5,31%	18,66%	13,35%	2,49%	
2	FIDEICOMISO	46	17	29	36,96%	1,768502607	-1,76850261	2,70%	0,46%	3,96%	101	1203	16,06%	19,13%	3,07%	37,79%	2,70%	1,02%	
3	HIPOTECA	3124	183	2941	5,86%	-0,47443369	0,474433693	29,09%	46,76%	0,38%	284	4144	45,15%	65,88%	20,73%	85,01%	29,09%	24,73%	
4	PAGARE	2278	339	1939	14,88%	0,558657545	-0,55865755	53,90%	30,83%	12,89%	623	6083	99,05%	96,71%	2,34%	162,59%	53,90%	87,63%	
5	PRENDA	76	4	72	5,26%	-0,58778666	0,587786665	0,64%	1,14%	0,30%	627	6155	99,68%	97,85%	1,83%	194,56%	0,64%	1,24%	
6	T VALORES	137	2	135	1,46%	-1,9095425	1,909542505	0,32%	2,15%	3,49%	629	6290	100,00%	100,00%	0,00%	197,85%	0,32%	0,63%	
TOTAL			6919	629	6290						IV				KS	20,73%		GINI	17,74%

Para verificar si era factible o no utilizar esta variable en el modelo se determinó el GINI con resultado satisfactorio con un índice de 17.74%, y basado en el teorema se toma la decisión de incluirla dentro del modelo.

**Agencia:** la agencia según se determinó no representa una variable importante en la medición ya que no es significativa para determinar si en una agencia en

especifico existiría morosidad, por esa razón se determinó mediante el porcentaje del GINI 0.95% que no era una variable que debía ser utilizada.

#### AGENCIAS

CATEGORÍA	TOTAL	MALOS	BUENOS	PD	WOE ESPAÑOL	WOE AMERICANO	% MALOS	% BUENOS	IV	ACUMULADO MALOS	ACUMULADO BUENOS	ACUMULADO MALOS %	ACUMULADO BUENOS %	KS	Abi%+ Abi%-1	AMI%-1	GINI
CREDITO CN CARTAGO	1029	47	982	4,57%	-0,73685861	0,736858614	7,47%	15,61%	6,00%	47	982	7,47%	15,61%	8,14%	15,61%	7,47%	1,17%
PLATAFORMA CREDITO SAN JOSE	998	122	876	12,22%	0,331240047	-0,33124005	19,40%	13,93%	1,81%	169	1858	26,87%	29,54%	2,67%	45,15%	19,40%	8,76%
PLATAFORMA SERVICIOS ALAJUELA	354	36	318	10,17%	0,124052649	-0,12405265	5,72%	5,06%	0,08%	205	2176	32,59%	34,59%	2,00%	64,13%	5,72%	3,67%
PLATAFORMA SERVICIOS HEREDIA	487	93	394	19,10%	0,858833677	-0,85883368	14,79%	6,26%	7,32%	298	2570	47,38%	40,86%	6,52%	75,45%	14,79%	11,16%
PLATAFORMA SERVICIOS LIBERIA	177	21	156	11,86%	0,297251523	-0,29725152	3,34%	2,48%	0,26%	319	2726	50,72%	43,34%	7,38%	84,20%	3,34%	2,81%
PLATAFORMA SERVICIOS LIMON	96	19	77	19,79%	0,90321865	-0,90321865	3,02%	1,22%	1,62%	338	2803	53,74%	44,56%	9,17%	87,90%	3,02%	2,66%
PLATAFORMA SERVICIOS PUNTARENA	35	1	35	2,86%	-1,25276297	1,252762968	0,16%	0,56%	0,50%	339	2838	53,90%	45,12%	8,78%	89,68%	0,16%	0,14%
PLATAFORMA SERVICIOS TRES RIOS	350	16	334	4,57%	-0,73596718	0,735967178	2,54%	5,31%	2,04%	355	3172	56,44%	50,43%	6,01%	95,55%	2,54%	2,43%
PLATAFORMA SERVICIOS TURRIALBA	506	28	478	5,53%	-0,53482113	0,534821129	4,45%	7,60%	1,68%	383	3650	60,89%	58,03%	2,86%	108,46%	4,45%	4,83%
PTO COM- ALAJUELA AGCIA CTRAL	152	14	138	9,21%	0,014388737	-0,01438874	2,23%	2,19%	0,00%	397	3788	63,12%	60,22%	2,89%	118,25%	2,23%	2,63%
PTO COM- CURRIDABAT	84	9	75	10,71%	0,182321557	-0,18232156	1,43%	1,19%	0,04%	406	3863	64,55%	61,41%	3,13%	121,64%	1,43%	1,74%
PTO COM- DESAMPARADOS	171	15	156	8,77%	-0,03922071	0,039220713	2,38%	2,48%	0,00%	421	4019	66,93%	63,90%	3,04%	125,31%	2,38%	2,99%
PTO COM- ESCAZU	145	16	129	11,03%	0,215361411	-0,21536141	2,54%	2,05%	0,11%	437	4148	69,48%	65,95%	3,53%	129,84%	2,54%	3,30%
PTO COM- GUADALUPE SAN JOSE	244	31	213	12,70%	0,375280132	-0,37528013	4,93%	3,39%	0,58%	468	4961	74,40%	69,33%	5,07%	135,28%	4,93%	6,67%
PTO COM- GUAPILES	131	11	120	8,40%	-0,08701138	0,087011377	1,75%	1,91%	0,01%	479	4481	76,15%	71,24%	4,91%	140,57%	1,75%	2,46%
PTO COM- JESUS JIMENEZ	103	6	97	5,83%	-0,48036642	0,480366416	0,95%	1,54%	0,28%	485	4578	77,11%	72,78%	4,32%	144,02%	0,95%	1,37%
PTO COM- LA URUCA	71	6	65	8,45%	-0,08004271	0,080042708	0,95%	1,03%	0,01%	491	4643	78,06%	73,82%	4,24%	146,60%	0,95%	1,40%
PTO COM- LOS ANGELES	163	6	157	3,68%	-0,96190124	0,961901243	0,95%	2,50%	1,48%	497	4800	79,01%	76,31%	2,70%	150,13%	0,95%	1,43%
PTO COM- METROCENRO	74	7	67	9,46%	0,043802623	-0,04380262	1,11%	1,07%	0,00%	504	4867	80,13%	77,38%	2,75%	153,69%	1,11%	1,71%
PTO COM- OREAMUNO	248	19	229	7,66%	-0,18669793	0,186697931	3,02%	3,64%	0,12%	523	5096	83,15%	81,02%	2,13%	158,39%	3,02%	4,78%
PTO COM- PACAYAS	131	8	123	6,11%	-0,43015772	0,430157721	1,27%	1,96%	0,29%	531	5219	84,42%	82,97%	1,45%	163,99%	1,27%	2,09%
PTO COM- PARAISO	285	19	266	6,67%	-0,32647224	0,326472237	3,02%	4,23%	0,41%	550	5485	87,44%	87,20%	0,24%	170,17%	3,02%	5,14%
PTO COM- PARQUE INDUSTRIAL	163	14	149	8,59%	-0,06230388	0,062303883	2,23%	2,37%	0,01%	564	5634	89,67%	89,57%	0,10%	176,77%	2,23%	3,93%
PTO COM- PEREZ ZELEDON	93	3	90	3,23%	-1,09861229	1,098612289	0,48%	1,43%	1,05%	567	5724	90,14%	91,00%	0,86%	180,57%	0,48%	0,86%
PTO COM- SAN CARLOS	98	18	80	18,37%	0,810930216	-0,81093022	2,86%	1,27%	1,29%	585	5804	93,00%	92,27%	0,73%	183,28%	2,86%	5,24%
PTO COM- TEJAR DEL GUARCO	231	27	204	11,69%	0,280301965	-0,28030197	4,29%	3,24%	0,29%	612	6008	97,30%	95,52%	1,78%	187,79%	4,29%	8,06%
PTO COM- TERRAMALL	32	1	31	3,13%	-1,131402111	1,131402111	0,16%	0,49%	0,38%	613	6039	97,46%	96,01%	1,45%	191,53%	0,16%	0,30%
PTO COM- TIERRA BLANCA	119	7	112	5,88%	-0,47000363	0,470003629	1,11%	1,78%	0,31%	620	6151	98,57%	97,79%	0,78%	198,80%	1,11%	2,16%

**Calificación SUGEF:** Se determina que la calificación de SUGEF evidencia que entre menor sea la calificación mayor es el riesgo por parte del banco de otorgar un préstamo.

#### CALIFICACION SUGEF

TRAMO	CATEGORIA	TOTAL	MALOS	BUENOS	PD	WOE ESPAÑOL	WOE AMERICANO	% MALOS	% BUENOS	IV	ACUMULADO MALOS	ACUMULADO BUENOS	ACUMULADO MALOS %	ACUMULADO BUENOS %	KS	Abi%+ Abi%-1	AMI%-1	GINI
1	A1	5792	0,1	5792	0,00%	-8,66423293	8,664232934	0,02%	92,08%	797,69%	0,1	5792	0,02%	92,08%	92,07%	92,08%	0,02%	0,01%
2	A2	149	0,1	149	0,07%	-5,00394631	5,003946306	0,02%	2,37%	11,77%	0,2	5941	0,03%	94,45%	94,42%	186,53%	0,02%	0,03%
3	B1	141	103	38	73,05%	3,299727921	-3,29972792	16,38%	0,60%	52,04%	103,2	5979	16,41%	95,06%	78,65%	189,51%	16,38%	31,03%
4	B2	47	39	8	82,98%	3,886705197	-3,8867052	6,20%	0,13%	23,60%	142,2	5987	22,61%	95,18%	72,58%	190,24%	6,20%	11,80%
5	C1	92	47	45	51,09%	2,346070205	-2,3460702	7,47%	0,72%	15,85%	189,2	6032	30,08%	95,90%	65,82%	191,08%	7,47%	14,28%
6	C2	22	21	1	95,45%	5,347107531	-5,34710753	3,34%	0,02%	17,77%	210,2	6033	33,42%	95,91%	62,50%	191,81%	3,34%	6,40%
7	D	27	16	11	59,26%	2,677278542	-2,67727854	2,54%	0,17%	6,34%	226,2	6044	35,96%	96,09%	60,13%	192,00%	2,54%	4,88%
8	E	649	403	246	62,10%	2,796190119	-2,79619012	64,07%	3,91%	168,22%	629,2	6290	100,03%	100,00%	0,03%	196,09%	64,07%	125,63%
		TOTAL	6919	629,2	6290				IV	1093,28%				KS	94,42%		GINI	94,07%

El GINI da un 94.07%, el mismo es casi perfecto existiendo una correlación casi de 1, por lo que la variable dependiente y la calificación SUGEF, prácticamente son la misma, al presentar una correlación tan elevada se descarta dentro del análisis del Score Crediticio.

**Plazo:** Si bien es cierto que en el plazo el GINI nos da un porcentaje de 26,42% no la podemos considerar como una variable significativa, debido a que no presenta una distribución normal en los datos, esto a pesar de que los estadísticos indicaban que la variable podía ser utilizada dentro del modelo.

PLAZO																				
TRAMO	MIN	MAX	TOTAL	MALOS	BUENOS	PD	WOESPANOL	WOE AMERICANO	% MALOS	% BUENOS	IV	ACUMULADO MALOS	ACUMULADO BUENOS	ACUMULADO MALOS %	ACUMULADO BUENOS %	KS	Abi%+ Abi%-1	AMi% - AMi%-1	GINI	
1	-	5,00	1853	338	1515	18,24%	0,802460271	-0,80246027	53,74%	24,08%	0,24	338	1515	53,74%	24,08%	29,65%	24,08%	53,74%	12,94%	
2	5,00	6,00	1371	88	1283	6,42%	-0,37703446	0,377034457	13,99%	20,40%	0,02	426	2798	67,73%	44,48%	23,24%	68,57%	13,99%	9,59%	
3	6,00	10,00	675	8	667	1,19%	-2,12076341	2,120763411	1,27%	10,60%	0,20	434	3465	69,00%	55,09%	13,91%	99,57%	1,27%	1,27%	
4	10,00	15,00	412	14	398	3,40%	-1,04480958	1,044809583	2,23%	6,33%	0,04	448	3863	71,22%	61,41%	9,81%	116,50%	2,23%	2,59%	
5	15,00	20,00	767	42	725	5,48%	-0,54591694	0,545916944	6,68%	11,53%	0,03	490	4588	77,90%	72,94%	4,96%	134,36%	6,68%	8,97%	
6	20,00	30,00	1832	139	1702	7,59%	-0,20250028	0,202500283	22,10%	27,06%	0,01	629	6290	100,00%	100,00%	0,00%	172,94%	22,10%	38,22%	
TOTAL			6910	629	6290						IV	54%				KS	29,65%		GINI	26,42%

Posteriormente se realizó una optimización de manera que la variable presentara una tendencia normal, una vez hecho este ejercicio los estadísticos ya no recomendaban utilizar la variable.

Con respecto a la determinación de la significancia de las variables se presentaron casos donde fue necesario realizar una optimización en algunas variables, esto debido a que la información que mostraba en algunos tramos, no presentaba normalidad.

### 3. Pesos

Una vez evaluada cada variable de forma independiente se procede a unir en un solo archivo toda la información de cada unidad bajo estudio con cada una de sus características según las variables consideradas como relevantes para el análisis, de esta forma la base de datos queda transformada en términos de WOE's, tal y como se muestra en la tabla siguiente, para proceder luego con el análisis Logit.

DEFAULT	EDAD	ESTADO CIVIL	ESCOLARIDAD	GENERO	SECTOR ECONOMICO	INGRESO	TASA NOMINAL	MONTO SOLICITADO	GARANTIA
1	0,701286	0,402612	0,308910373	0,05864	0,757389449	0,3188752	0,4970323	0,589184406	-0,33477011
1	0,265385	0,3032251	0,210273361	0,05864	0,092523369	0,3188752	0,4970323	0,589184406	0,558657545
1	-0,36645	-0,242443	0,210273361	0,05864	-0,369747026	0,3188752	0,4970323	0,589184406	-0,33477011
0	0,265385	0,3032251	0,308910373	0,05864	0,092523369	0,3188752	0,4970323	0,589184406	0,558657545
0	-0,36645	-0,242443	0,210273361	0,05864	-0,205488616	-0,493648	0,0296005	-0,515826535	-0,47443369
0	0,265385	-0,242443	0,308910373	0,05864	0,092523369	0,3188752	0,4970323	0,589184406	0,558657545
0	-0,02638	-0,156529	0,210273361	0,05864	0,28015037	-0,04282	-0,945005	-0,515826535	-0,47443369
0	0,265385	0,3032251	-0,583921179	-0,13679	0,28015037	-0,671707	0,4970323	-0,515826535	0,558657545

#### 4. Desarrollo de Logit

Una vez transformada la base de datos de las nueve variables categóricas y continuas, y expresadas en términos de WOE,s (peso de la evidencia), se tienen los pesos analizados de forma individual, por lo que se procede a realizar la regresión logística a través de un análisis multivariable, determinando los valores  $\beta$  de cada una de las nueve variables que van a componer el modelo, para esto se realiza un primer paso que es asumir la normalidad sobre la serie y se le da un peso de 2 a cada variable desde la constante, hasta la novena variable tal y como se muestra en el cuadro siguiente:

Pesos del Logit	$\beta_0$
c	2
Edad	2
Estado Civil	2
Grado de escolaridad	2
Género	2
Sector Económico	2
Ingreso	2
Tasa Nominal	2
Monto Solicitado	2
Garantía	2

Recordando la función que nos dará el score (Z Score de Altman):

$$Z = C + \beta_1 * X_1 + \beta_2 * X_2 + \beta_3 * X_3 + \beta_n * X_n$$

se procede a calcular un primer score, en donde se multiplica cada WOE por su respectivo  $\beta$  y se suman todos los resultados incluyendo la constante, en este momento todas las  $\beta_s$  están fijadas en 2, asumiendo normalidad en la serie.

El paso siguiente es aplicar la fórmula de la regresión logística la cual está definida de la siguiente manera:

$$Logit = \frac{1}{1 + e^x}$$

dónde:

X= Score

e= 2,71828182845904523536028747135266249775724709369995

Una vez realizado este paso es donde estamos en presencia del modelo Logit y al mismo tiempo estamos definiendo la probabilidad de default de la variable, sin

embargo, este paso no es suficiente para determinar si un cliente es bueno o malo.

Paso siguiente es definir la verosimilitud, que es un método habitual para ajustar un modelo y encontrar sus parámetros, la elección de cuál es el modelo que se utiliza en última instancia. En cualquier situación debe considerar que los rangos de probabilidad son la información para resolver el problema, en otras palabras es una fórmula que nos ayudara a definir cuál es la secuencia de pesos que nos dará un modelo mejor ajustado.

Una vez programada en el modelo la función de la verosimilitud individual:

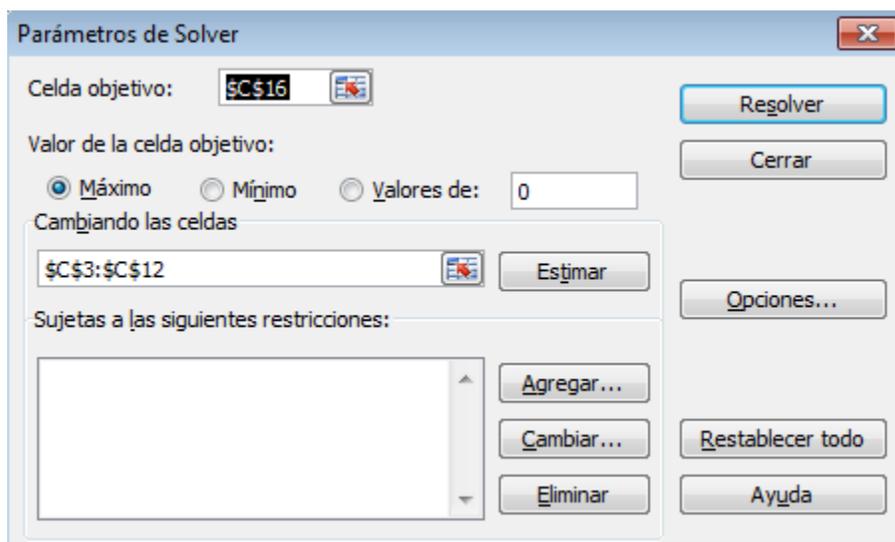
$$\text{Verosimilitud individual} = \text{Default}_i * \text{LN}(\text{PD}_i) + (1 - \text{Default}_i) * \text{LN}(1 - \text{PD}_i)$$

Se procede a sumar todas verosimilitudes individuales y dentro del teorema tenemos que cuanto más cercano a cero sea esta menor será el error del pronóstico.

Para el ejercicio realizado la sumatoria de las verosimilitudes individuales previo a la calibración de los betas era de:

<b>Suma</b>	-13.791,5068
-------------	--------------

La calibración de las  $\beta_s$  se realizó a través de la herramienta solver, definido la celda objetivo (la suma de las verosimilitudes individuales) y como el resultado es negativo se procedió a maximizarlo, tal y como se muestra en la imagen siguiente:



Luego de ejecutado el Solver, quedaron definidos los pesos de las nueve variables y la constante tal y como se muestra en el cuadro siguiente:

Pesos del Logit	$\beta_0$
c	-2.3075682
Edad	0.42953442
Estado Civil	-0.04558874
Grado de escolaridad	0.41895163
Género	0.6135450
Sector Económico	0.81886842
Ingreso	0.2024480
Tasa Nominal	0.49834829
Monto Solicitado	-0.22232124
Garantía	0.77534385

Así la función de verosimilitud paso de un valor de -13.791,5068 a uno ajustado de:

<b>Suma</b>	-1890,9668
-------------	------------

Siendo un valor más cercano a cero, cumpliendo con el teorema de la función de la máxima verosimilitud.

Una vez realizado el ejercicio se procede a definir el punto de corte el cual se definió inicialmente de forma arbitraria en 50% con los siguientes resultados:

PUNTO DE CORTE	50%
SUMA	-1890,9668
PRECISIÓN TOTAL	91,30%
PRECISIÓN BUENOS	99,89%
PRECISIÓN MALOS	0,00%

Con un punto de corte en 50% estaría aceptando casi el 100% de los buenos, pero no está identificando a ninguno de los malos, por lo que el mismo debe ser calibrado tratando de minimizar los errores tipo I y Tipo II de la estadística.

Luego de una serie de pruebas se definió que el punto de corte de mejor ajuste es un 11%, dado que paso aceptar el 83.70% de los buenos y a rechazar el 52.33% de los malos, obteniendo una precisión total del pronóstico de 81.00%.

<b>Punto de corte</b>	11%
<b>Suma</b>	-1890,9668
<b>Precisión Total</b>	81,00%
<b>Precisión Buenos</b>	83,70%
<b>Precisión Malos</b>	52,33%

### 5. Punto de corte

Si vemos los resultados del score del ejercicio planteado, el mismo oscila entre -7.0001 y 0.1084. Entonces generalmente esto en los modelos score no es interpretable ni muy manejado, por lo que hacemos es transformar los valores o resultados del modelo Logit en un **score alineado**, con el fin de poder interpretarlos de una mejor manera y de esta forma obtener un score definitivo, para lo cual desarrollamos la siguiente fórmula:

$$SCOREALINEADO = -SCORE * FACTOR + OFFSET$$

Siendo el cálculo del score igual a:

$$Score_{variablen} = -\left(\frac{\beta_0}{n} + \beta_n * woe_n\right) * factor + \frac{offset}{n}$$

Donde obtenemos que:

$\beta_0$  = a la constante

n = cantidad de variables

$\beta_n$  = es el peso de la variable

Para calcular el factor y el offset:

$$600 = \text{Ln}(50) * \text{FACTOR} + \text{OFFSET}$$

$$\text{FACTOR} = 20 / \text{Ln}(2)$$

$$\text{OFFSET} = 600 - \text{FACTOR} * \text{Ln}(50)$$

Lo anterior significa que a un deudor que se le otorga 600 puntos es 50 veces más probable que sea bueno a que sea malo y cada vez que se dobla esta proporción en los odds, se incrementa la puntuación en 20.

Dado lo anterior y una vez aplicado el cálculo anterior obtenemos los siguientes resultados:

PARÁMETROS Y PESOS	
<b>FACTOR</b>	28.8539008
<b>OFFSET</b>	487.122876

Estos resultados nos permiten determinar el score definitivo de cada variable y de esta manera proceder a sumar cada una para obtener el score total, tal y como se muestra en el siguiente cuadro.

EDAD	ESTADO CIVIL	ESCOLARIDAD	GENERO	SECTOR ECONÓMICO	INGRESO	TASA NOMINAL	MONTO SOLICITADO	GARANTIA	SCORE TOTAL
52.83124495	62.05240332	57.78857356	60.48469137	43.62754662	59.6601199	54.34590387	65.30232308	69.0121769	525.10
58.23368675	61.92166849	58.98093611	60.48469137	59.3367019	59.6601199	54.34590387	65.30232308	49.0246863	527.29
66.06446625	61.20388921	58.98093611	60.48469137	70.25901786	59.6601199	54.34590387	65.30232308	69.0121769	565.31
58.23368675	61.92166849	57.78857356	60.48469137	59.3367019	59.6601199	54.34590387	65.30232308	49.0246863	526.10
66.06446625	61.20388921	58.98093611	60.48469137	66.3779944	73.18648639	61.09716844	58.21386006	72.1366879	577.75
58.23368675	61.20388921	57.78857356	60.48469137	59.3367019	59.6601199	54.34590387	65.30232308	49.0246863	525.38
61.84969405	61.31690224	58.98093611	60.48469137	54.90353592	62.53453379	75.11130437	58.21386006	72.1366879	565.53

Para definir el punto de corte procedemos a desarrollar la última tabla en la cual vamos a identificar del score total cual es el dato mínimo y por medio de percentiles identificar cada uno de los rangos a utilizar, de esta forma obtenemos un total de nueve rangos de clase o tramos, donde la calificación mínima será de 483,96 puntos y la punto máximo es de 689.10. De igual forma podemos observar que todos los rangos contiene prácticamente la misma cantidad de clientes la cual oscila entre 760 y 784 clientes.

Adicionalmente podemos observar que conforme avancemos de rango, la cantidad de clientes malos disminuye hasta llegar a encontrar o definir 6 clientes malos de setecientos sesenta y nueve clientes en el tramo número nueve.

En el caso de la probabilidad default (PD) podemos observar como la misma disminuye el porcentaje según avancemos de tramo y esto se debe a que va muy de la malo con los datos anteriores, es decir con la cantidad de clientes buenos y malos.

Con respecto al WOE Español nos brinda información positiva en los primeros tres tramos, por lo que podemos decir que todos aquellos clientes que cuenten con una puntuación inferior a los 547,96 son de alto riesgo, o dicho en otras palabras existe una alta probabilidad de no pago. De igual forma en el WOE Americano observamos que los mismos tres primeros datos son negativos, por lo que reafirma la alta probabilidad de impago.

En el caso de los estadísticos podemos observar que el Information Value (IV) nos da un total de 0,82, lo cual está muy por encima del 0,3 por lo que de acuerdo a la teoría es un indicador muy predictivo, por lo que podemos decir que la variable discrimina altamente entre los clientes buenos y malos, por lo que definitivamente se debe utilizar.

TRAMO	MIN	MAX	TOTAL	MALOS	BUENOS	PD	WOE ESPAÑOL	WOE AMERICAN O	% MALOS	% BUENOS	IV	ACUMULADO MALOS	ACUMULADO BUENOS	ACUMULADO MALOS %	ACUMULADO BUENOS %	KS	Abi% Abi%-1	+ AMI% AMI%-1	- GINI
1	483.963019	527.2907177	763	197	566	25.82%	1.24719474	-1.24719474	31.32%	9.00%	0.28	197	566	31.32%	9.00%	22.32%	9.00%	31.32%	2.82%
2	527.2907177	537.6944828	775	123	652	15.87%	0.63472489	-0.63472489	19.55%	10.37%	0.06	320	1218	50.87%	19.36%	31.51%	28.36%	19.55%	5.55%
3	537.6944828	547.9610468	768	93	675	12.11%	0.3204719	-0.3204719	14.79%	10.73%	0.01	413	1893	65.66%	30.10%	35.56%	49.46%	14.79%	7.31%
4	547.9610468	557.4634186	769	56	713	7.28%	-0.24154464	0.24154464	8.90%	11.34%	0.01	469	2606	74.56%	41.43%	33.13%	71.53%	8.90%	6.37%
5	557.4634186	566.8422742	769	53	716	6.89%	-0.30080316	0.30080316	8.43%	11.38%	0.01	522	3322	82.99%	52.81%	30.17%	94.24%	8.43%	7.94%
6	566.8422742	575.9389406	760	45	715	5.92%	-0.46303496	0.46303496	7.15%	11.37%	0.02	567	4037	90.14%	64.18%	25.96%	117.00%	7.15%	8.37%
7	575.9389406	586.5795906	762	31	731	4.07%	-0.85784116	0.85784116	4.93%	11.62%	0.06	598	4768	95.07%	75.80%	19.27%	139.98%	4.93%	6.90%
8	586.5795906	609.9456122	784	25	759	3.19%	-1.11054086	1.11054086	3.97%	12.07%	0.09	623	5527	99.05%	87.87%	11.18%	163.67%	3.97%	6.51%
9	609.9456122	689.1033933	769	6	763	0.78%	-2.54291347	2.54291347	0.95%	12.13%	0.28	629	6290	100.00%	100.00%	0.00%	187.87%	0.95%	1.79%
<b>TOTAL</b>			<b>6919</b>	<b>629</b>	<b>6290</b>						<b>IV</b>					<b>KS</b>			<b>GINI</b>
											<b>0.82</b>					<b>35.56%</b>			<b>46.45%</b>

Esta información nos va a permitir definir cuál es el punto de corte que vamos a utilizar para decidir o considerar si un cliente es bueno o malo, en otras palabras para decir si debemos o no otorgarle el crédito.

## 6. Resultados

Una vez desarrollado la propuesta del modelo Credit Scoring, podemos observar que del total de clientes que utilizamos para el desarrollo del modelo, el 79,26% coincide con la clasificación generada por el banco, es decir, que de los 6.919 clientes, 5.484 es igual al resultado de la institución, en otras palabras si de acuerdo al modelo un cliente en específico debe ser aprobado, esta coincidencia se muestra en el Banco donde efectivamente se le aprobó el crédito al cliente y a la fecha no ha incumplido en su pago, y vrs en aquellos casos que el modelo muestra que se debe rechazar el cliente y en el registro de la institución el mismo se encuentra como mal pagador por incumplimiento.

Si comparamos el 79,26% de predictibilidad que nos da la comparación de resultados obtenidos junto con la información que nos presenta la institución, podemos observar que hay una diferencia de 1,74%, esto se debe a los errores estadísticos tipo 1 y tipo 2, el cual consiste en la probabilidad de otorgar un crédito a un cliente mal pagador o bien denegarle el préstamo a un cliente buen pagador.

Esta diferencia difícilmente la podemos eliminar ya que son resultados típicos de las probabilidades en los temas de la estadística.

## **B. COMPARACIÓN MODELO LOGIT Y MODELO UTILIZADO POR EL BANCO**

El modelo actual con el que cuenta la institución es utilizado por los encargados de crédito para determinar las estadísticas necesarias para la validación de los factores.

El banco utiliza una herramienta establecida en el año 2007, con la cual de acuerdo a una serie de variables obtiene la información necesaria para aplicar en el estudio respectivo, en lo que respecta a préstamos de consumo.

Dichas variables se utilizan para determinar el grado de riesgo que podría generar un determinado cliente, las mismas son aplicadas en parámetros como:

VARIABLES	
Código de cliente	Valor de la cuota
Año	Monto del principal
Tipo de préstamo	Valor actualizado
Identificación del núm. De operación	Valor intereses
Fecha de apertura de la operación	Código amortización
Fecha de vencimiento	Código tipo
Codificación del tipo de actividad	Código frecuencia capital
Codificación del tipo de moneda	Código frecuencia interés
Tasas de interés	

Dichas variables pueden según sea el caso influir directa o indirectamente en los resultados finales.

Mediante tablas dinámicas con fórmulas debidamente aplicadas se generan resultados para ser evaluados, entre más exactos y veraces los datos se podrá hablar de que la base es más fiable a la hora de obtener los resultados.

Al contar con gran cantidad de datos se buscó generar los códigos necesarios para ser aplicados, además de codificar por códigos que por su naturaleza era muy complicado asignarles un solo código, por ejemplo las profesiones, que en

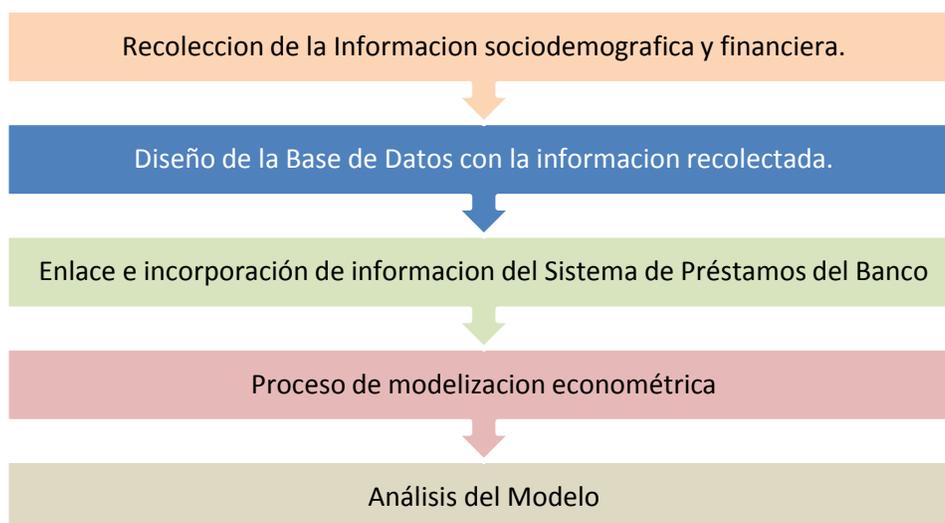
lugar de darles una numeración determinada se les asigno un código por segmento.

Una vez validados los datos se generan los resultados que pueden ser colocados en categorías que van de A a E, las mismas permiten evaluar la capacidad de pago que podría tener los clientes.

Con los datos que se generan se logra obtener estadísticas de cuáles son los clientes con potencial de riesgo más alto, de acuerdo a las variables se verifica si un casado puede ser menos riesgoso que un divorciado, pero las otras variables pueden modificar la estadística a favor o en contra del cliente, es por esta razón que el banco utiliza una base de datos amplia con factores de medición.

El banco utiliza un proceso detallado para la obtención de los datos mediante un sistema de recolección, evaluación y determinación final de los análisis.

La institución obtiene la información que es cargada a la base de datos con las diferentes variables, se realiza un muestreo de los datos significativos, se codifica la información para ser transformada a parámetros de medición más apropiados, de esa forma se realizan las pruebas en las herramientas donde se generan los resultados de predicción a futuro de la posible morosidad.



Si buscamos una comparación del modelo que mantiene el banco con el modelo que se ha desarrollado en esta investigación, podemos decir que hay una diferencia significativa en las variables que utiliza la institución ya que ellos no consideran algunas instituciones tales como como características propias del ser

humano, lo cual pueden brindarnos mucha información de si un posible cliente puede ser buen o mal pagador.

Lamentablemente y como lo hemos mencionado en ocasiones anteriores no fue posible visualizar el modelo que obtiene el banco ni se logró obtener mayor información de la situación y variables de los clientes, por lo que llevar a cabo una comparación más avanzada y profunda es un poco complicado.

## V. CONCLUSIONES

- 1) El desarrollo de Modelos Credit Scoring depende de cada institución financiera y de los resultados que se deseen obtener, ya que de acuerdo a la investigación realizada existen varios modelos a utilizar, así como ventajas y desventajas de los mismo, ya que si bien es cierto, permiten cuantificar el riesgo como una probabilidad además de ser consistente y explícito, debe requerir una serie de datos los cuales deben ser de alta calidad y una muestra lo suficientemente amplia para poder ser más certeros.
- 2) Entre más variables se puedan obtener y codificar para el desarrollo del modelo, mejor es el resultado que se puede obtener, por lo que los criterios a seleccionar dependen de la información con la que cuente la institución, y es aquí donde se considera el modelo desarrollado en Excel como el más óptimo, ya que no es suficiente solo con contar con muchas variables, sino que se debe definir si las mismas ofrecerán un valor agregado al estudio, tal y como se mostró con el caso del GINI, donde logramos identificar cuales variables debíamos seleccionar para el modelo.
- 3) Del estudio realizado inferimos que el mejor modelo a utilizar es el Logit, ya que como se indicó anteriormente, el análisis discriminante muestra el inconveniente de que las variables muestran una distribución normal en cada una de las poblaciones con matrices de varianza y covarianza iguales diferenciándose solamente en los valores de los vectores medios. Adicionalmente existe la incapacidad para el cálculo de las probabilidades de impago.
- 4) Dado que no se logró obtener mayor información del modelo que utiliza la institución para la toma de decisión en lo que se refiere a otorgarle o no un crédito a una persona física bajo los criterios que utilizaron, únicamente se desarrollaron los modelos planteados en el proyecto y utilizarlo en la comparación de los clientes de la base de datos que fue facilitada. Esto arrojó una predicción del modelo de un 79,26%, es decir que el 79,26% de la categoría en la que se encuentran los clientes ya sean en incumplimiento o no es la misma que muestra uno de los modelos utilizados en esta investigación, específicamente el Logit.
- 5) Dado el inconveniente de no tener más información del modelo credit score utilizado por el Banco, no es posible indicar si dicho modelo es o no factible en comparación a la propuesta realizada en este proyecto.

- 6) Las variables que utiliza actualmente la institución para hacer el análisis de crédito están referenciadas directamente al crédito tales como cuota, monto del principal, interés, amortización entre otros y no considera características propias del individuo como estado civil, escolaridad, edad y género, características que nosotros consideramos importantes y relevantes para un Credit Scoring.
- 7) Con respecto a la política de la entidad financiera, no fue posible obtener dicha información debido a las restricciones de confiabilidad bancaria con la que cuentan, por lo que no podemos emitir un criterio si dichas políticas son las adecuadas para un modelo de Credit Scoring o bien, sugerir alguna otra, sin embargo, el plazo recomendado para la actualización de los modelos Credit Score es de seis meses a un año, por lo que actualmente el modelo utilizado por la institución se encuentra desactualizado.
- 8) Con respecto a la normativa a nivel nacional o internacional, determinamos que la SUGEF como ente regular de las instituciones financieras no emite ningún criterio sobre la aplicación o no de los Credit Scoring, mientras que a nivel internacional existe Basilea I y II, pero tampoco profundizan en el desarrollo en el uso de este tipo de modelos.
- 9) A nivel general, se concluye que el uso del Credit Scoring genera beneficios a una entidad financiera, siempre y cuando se cuente con la información necesaria y confiable para el desarrollo del mismo. En el caso del modelo de Excel logramos determinar de 13 variables que presentaba la base de datos 9 variables significativas, es decir, que se debería considerar para el desarrollo del modelo.
- 10) Con un 81% de predictibilidad o bien con un 79,26 considerando los errores estadísticos 1 y 2, consideramos que el modelo puede ser eficiente y recomendable para su uso.
- 11) Con este mismo modelo se logró determinar una calificación del score total que permitiera evaluar si un cliente puede ser considerado como buen o mal pagador.

## VI. RECOMENDACIONES

Una vez realizada la investigación del tema del Credit Scoring en una entidad financiera se procede a realizar las siguientes recomendaciones:

- 1) Mantener una evaluación periódica del modelo de Credit Scoring (cada seis meses preferiblemente) que se esté utilizando y que de esta forma se proceda a realizar las mejoras para que el mismo tenga un grado de predicción acorde con los cambios generacionales y ajustado a la realidad del momento.
- 2) Encomendar al comité de crédito la evaluación y planificación periódica del modelo, con el propósito de estar más familiarizado con los procesos internos. De esta forma se podrá analizar cada una de las variables actuales y discutir las variables a futuro que podrían generarse en este mundo cambiante.
- 3) Verificar el manual aquí descrito para recomendar como una opción la implementación del modelo logit en los procesos de análisis de clientes que solicitan prestamos en la entidad.
- 4) Enfocar a la institución con un horizonte donde se determine que el análisis de riesgo se vea como parte de los activos de la empresa, ya que un proceso de análisis eficiente minimizaría el riesgo de otorgar créditos a personas con capacidades deficientes de pago, con esto se recomienda generar modelos de medición actualizados.
- 5) Continuar actualizándose mediante normativa internacional, nacional, y experiencias pasadas, que cada uno de los integrantes del equipo de estudio de crédito cuente con capacitaciones periódicas, para que de esta forma se unifiquen criterios a nivel departamental y de esta forma no existan elementos de juicio que puedan generar diferentes comentarios a la hora de otorgar un crédito.
- 6) Verificar la veracidad de los requisitos aportados por los clientes, de esta forma podemos certificar que los datos que se entregan no son falsos o alterados y de esta forma minimizar el riesgo dentro del proceso de otorgamiento de crédito.

- 7) Fortalecer las políticas internas en el manejo de Credit Scoring con el afán de lograr un proceso óptimo de préstamos personales, recordemos que el incumplimiento de un pago le genera a la institución pérdidas monetarias significativas.
- 8) Verificar el presupuesto asignado a la investigación y mejoramiento del Credit Scoring con el cual se faciliten los procesos de gestión de riesgo para que la columna financiera de la institución siga siendo la más importante dentro de la organización.
- 9) Dentro de la estructura del equipo evaluador que se recomienda en el punto dos, se invita a definir la estructura del mismo, ubicando las funciones y responsabilidades que tendrá cada integrante del equipo con el fin de sacar el máximo provecho a las mismas.

## VII. ANEXOS

### ANEXO N°1

#### Criterio Emitido por la SUGEF vía correo electrónico

-----Mensaje original-----

**De:** sugefcr@sugef.fi.cr [mailto:sugefcr@sugef.fi.cr]

**Enviado el:** lunes, 29 de abril de 2013 09:11 a.m.

**Para:** avillareal@vivicon.co.cr

**Asunto:** Caso 2013-2951-CRM

Estimado Sr. Villareal:

En relación con su consulta, nos permitimos indicarle que la función de la SUGEF, según se define en la Ley Orgánica del Banco Central de Costa Rica (Ley No. 7558), es velar por la estabilidad del Sistema Financiero Nacional como un todo y ésta tiene competencias primordialmente en materia financiero-contable, no así en asuntos relativos a prestación de servicios. Asimismo, la SUGEF no es una instancia judicial y como ente supervisor carece de potestad para coadministrar los entes que supervisa. Consecuentemente, los servicios y requisitos gestionados en una entidad supervisada para un trámite de financiamiento, corresponde a políticas crediticias internas de dicha entidad y la SUGEF no tiene injerencia sobre las mismas.

Asimismo, nos permitimos indicarle, que esta Superintendencia carece de atribuciones legales para arbitrar o resolver conflictos que se presenten entre las entidades supervisadas y sus clientes, como consecuencia de los contratos privados suscritos entre ambos. En ese sentido, en caso de no dirimirse por los procedimientos administrativos contemplados en nuestro ordenamiento jurídico, dichos conflictos necesariamente deberán resolverse en las instancias judiciales correspondientes.

Por otra parte, la normativa emitida por la SUGEF, está orientada a las entidades supervisadas y no a sus clientes. Asimismo, dicha normativa tiene el objeto de cuantificar el riesgo de crédito de los deudores y constituir las estimaciones correspondientes con el fin de salvaguardar la estabilidad y solvencia de las entidades y conglomerados financieros.

Atentamente,

Superintendencia General de Entidades Financieras (SUGEF)

## VIII. BIBLIOGRAFÍA

(s.f.).

Arguedas, J. A. (2008). *Acuerdo de Capitales Basilea II" y sus efectos sobre las entidades Bancarias-Financieras y los organismos de supervisión de Costa Rica*. Costa Rica.

Armando, B. N. (2012). Credit Scoring for SME Using a Manifold Supervised Learning Algorithm. *Intelligent Data Engineering and Automated*, 763-764.

Eva Boj, M. M. (2010). *CRITERIOS DE SELECCIÓN DE MODELO EN CREDIT SCORING. APLICACIÓN DEL ANÁLISIS DISCRIMINANTE BASADO EN DISTANCIAS*.

Eva, M. A. (2009). Criterios de selección de modelo en el Credit Scoring. *Anales del Instituto de Actuarios Españoles*, 209-230.

Girault, M. A. (2007). *Modelos de Credit Scoring*.

Girault, M. A. (2007). Modelos de Credit Scoring (Qué, Cómo, Cuándo y Para qué).

Joaquina Laffarga, J. L. (1987). *Predicción de la crisis bancaria en España comparacion entre el analisis logit y el discriminante*. Sevilla.

Joaquina Laffarge Briones, J. L. (s.f.). *Predicción de la crisis bancaria en España, comparación entre el análisis logit y el analisis discriminante*. Sevilla.

Salvador Rayo Cantón, J. L. (2010). *Un Modelo de Credit Scoring para instituciones de microfinanzas en el marco de Basilea II*.

Schreiner, M. (2002). *Ventajas y Desventajas del Scoring Estadístico para microfinanzas*.

Simbaqueba, L. (2004). *Que es Scoring. Una visión práctica de la gestión del riesgo de crédito*.

Vasquez, D. W. (2013). *Metodología para la construcción de modelos scoring*. San Jose.