



**1. Datos del Autor:**

Apellidos y Nombres: RODRÍGUEZ SABINO ROCIO NORKA

Código de alumno: 05.1343.9.AR

Teléfono: 942812268

Correo electrónico: [rocio.nrs@gmail.com](mailto:rocio.nrs@gmail.com)

DNI: 44395919

**2. Modalidad de trabajo de investigación:**

- Trabajo de Investigación                       Trabajo académico  
 Trabajo de suficiencia personal               Tesis

**3. Título profesional o grado académico**

- Bachiller                       Título                       Segunda especialidad  
 Licenciado                       Magister                       Doctor

**4. Título del trabajo de investigación:**

Modelo Credit Scoring para obtener la Probabilidad de Impago de Créditos Personales Futuros en la Empresa Servicios Financieros Grupo Buro, Huaraz 2018.

**5. Facultad de Ciencias**

**6. Escuela, Carrera o programa: Escuela Profesional de Estadística e informática**

**7. Asesor:**

Apellidos y Nombres: M.Sc. VARELA ROJAS WALTER ALEJANDRO

Teléfono: 943333089

Correo electrónico: [varow\\_9@hotmail.com](mailto:varow_9@hotmail.com)

DNI: 18117012

A través de este medio autorizo a la Universidad Nacional Santiago Antúnez de Mayolo, publicar el trabajo de investigación en formato digital en el Repositorio Institucional Digital, Repositorio Nacional Digital de Acceso Libre (ALICIA) y el Registro Nacional de Trabajos de Investigación (RENATI).

Asimismo, por la presente dejo constancia que los documentos entregados a la UNASAM, versión impresión y digital, son las versiones finales del trabajo sustentado y aprobado por el jurado y son de autoría del suscrito en estricto respeto a la legislación en materia de propiedad intelectual

FIRMA 

DNI: 44395919

Fecha: 14 de Noviembre de 2018

**UNIVERSIDAD NACIONAL  
SANTIAGO ANTÚNEZ DE MAYÓLO  
FACULTAD DE CIENCIAS  
ESCUELA PROFESIONAL DE ESTADÍSTICA E INFORMÁTICA**



**“MODELO CREDIT SCORING PARA OBTENER LA PROBABILIDAD DE  
IMPAGO DE CRÉDITOS PERSONALES FUTUROS EN LA EMPRESA  
SERVICIOS FINANCIEROS GRUPO BURO, HUARAZ 2018”**

**TESIS GUIADA**

**PARA OPTAR EL TÍTULO PROFESIONAL DE  
LICENCIADA EN ESTADÍSTICA E INFORMÁTICA**

**AUTOR:**

**Bach. Rocío Norka, RODRÍGUEZ SABINO**

**ASESOR**

**Lic. Walter Alejandro, VARELA ROJAS**

**HUARAZ – PERÚ**

**2018**

**PROGRAMA GUIADO DE TITULACIÓN PROFESIONAL CON TESIS**

**N° Registro: T001**

## **HOJA DE VISTO BUENO**

---

Dr. Ángel Deciderio Mendoza López  
Presidente

---

M.Sc. Jorge Luis Llanos Tiznado  
Secretario

---

Lic Hugo Walter Maldonado Leyva  
Vocal

**MODELO CREDIT SCORING PARA OBTENER LA  
PROBABILIDAD DE IMPAGO DE CRÉDITOS PERSONALES  
FUTUROS EN LA EMPRESA SERVICIOS FINANCIEROS GRUPO  
BURO, HUARAZ 2018**

## **ASESORES Y MIEMBROS DEL JURADO**

### **Asesor:**

M.Cs. Varela Rojas Walter Alejandro

### **Miembros del Jurado:**

Dr. Ángel Deciderio Mendoza López

M.Sc. Jorge Luis Llanos Tiznado

Lic. Hugo Walter Maldonado Leyva

## DEDICATORIA

*A mi amado esposo Joffre Cruz, por su sacrificio y esfuerzo, por apoyarme en mi carrera profesional para nuestro futuro y por creer en mi capacidad.*

*A mi adorada hija Sofía Cruz por ser mi fuente de motivación e inspiración para superarme cada día.*

*A mis queridos padres Luciano Rodríguez y Norka Sabino, por haberme apoyado en todas las etapas de mi vida, por sus consejos, sus valores, por la motivación constante que me ha permitido ser una persona de bien, pero más que nada, por su amor.*

*A mi hermano Vladimir Rodríguez, por ser un pilar fundamental en todo lo que soy, en toda mi educación, tanto académica, como de la vida, por su incondicional apoyo a través del tiempo.*

## **AGRADECIMIENTOS**

El presente trabajo de tesis primeramente me gustaría agradecer a ti Dios por bendecirme para llegar hasta donde he llegado, porque hiciste realidad este sueño anhelado.

A la UNIVERSIDAD SANTIAGO ANTÚNEZ DE MÁYALO por darme la oportunidad de estudiar y ser un profesional.

Al Ing. Esteban Julio Medina Rafael Julio Director y comisión del PTCT-FC-UNASAM-2018 por su esfuerzo y dedicación, quien, con sus conocimientos, su experiencia y su motivación ha logrado llevar con éxito el programa.

También me gustaría agradecer a mis profesores durante toda mi carrera profesional porque todos han aportado con un granito de arena a mi formación.

# ÍNDICE

<b>RESUMEN</b> .....	<b>3</b>
<b>INTRODUCCIÓN</b> .....	<b>4</b>
<b>CAPÍTULO I</b>	
<b>PLANTEAMIENTO DEL PROBLEMA</b> .....	<b>5</b>
<b>1.1. DESCRIPCIÓN DE LA REALIDAD PROBLEMÁTICA</b> .....	<b>5</b>
<b>1.2. FORMULACIÓN DEL PROBLEMA</b> .....	<b>6</b>
<b>1.3. OBJETIVO DE LA INVESTIGACIÓN</b> .....	<b>7</b>
<b>1.3.1. Objetivo General.</b> .....	<b>7</b>
<b>1.3.2. Objetivo Específicos.</b> .....	<b>7</b>
<b>1.4. JUSTIFICACIÓN DE LA INVESTIGACIÓN</b> .....	<b>7</b>
<b>1.4.1. Justificación Social.</b> .....	<b>7</b>
<b>1.4.2. Justificación Institucional.</b> .....	<b>8</b>
<b>1.4.3. Justificación Práctica.</b> .....	<b>8</b>
<b>1.5. LIMITACIÓN DEL ESTUDIO</b> .....	<b>8</b>
<b>1.6. VIABILIDAD DEL ESTUDIO</b> .....	<b>9</b>
<b>CAPÍTULO II</b>	
<b>BASES TEÓRICAS</b> .....	<b>10</b>
<b>2.1. ANTECEDENTES DE LA INVESTIGACIÓN</b> .....	<b>10</b>
<b>2.2. BASES TEÓRICAS</b> .....	<b>12</b>
<b>2.2.1. Credit Scoring.</b> .....	<b>12</b>
<b>2.2.2. Metodologías aplicadas en los modelos de Credit Scoring.</b> .....	<b>13</b>
<b>2.2.3. Tipo de Score.</b> .....	<b>13</b>
<b>2.2.4. Pesos de Evidencia (WOE).</b> .....	<b>14</b>
<b>2.2.5. Pruebas sobre la Clasificación de Atributos.</b> .....	<b>15</b>
<b>2.2.6. Regresión Logística.</b> .....	<b>17</b>
<b>2.2.7. Regresión logística binaria.</b> .....	<b>18</b>
<b>2.2.8. Pruebas de significación.</b> .....	<b>20</b>
<b>2.2.9. Bondad del Modelo.</b> .....	<b>23</b>
<b>2.2.10. Pruebas de coeficientes.</b> .....	<b>24</b>
<b>2.2.11. Scorecard.</b> .....	<b>25</b>
<b>2.2.12. Construcción de la Scorecard.</b> .....	<b>26</b>
<b>2.2.13. Determinación del punto de corte o Cut Off.</b> .....	<b>28</b>



2.3.	DEFINICIONES CONCEPTUALES .....	29
2.3.1.	<i>Crédito</i> .....	29
2.3.2.	<i>Riesgo de crédito</i> .....	30
2.3.3.	<i>Score</i> .....	30
2.3.4.	<i>La Superintendencia de Banca y Seguros</i> .....	30
2.4.	FORMULACIÓN DE LA HIPÓTESIS.....	32
<b>CAPITULO III</b>		
<b>METODOLOGÍA</b>		
3.1.	DISEÑO METODOLÓGICO .....	34
3.1.1.	<i>Tipo de estudio</i> .....	34
3.1.2.	<i>Procedimiento de la investigación</i> .....	34
3.2.	POBLACIÓN .....	35
3.3.	OPERACIONALIZACIÓN DE VARIABLES .....	35
3.4.	TÉCNICA DE RECOLECCIÓN DE DATOS .....	37
3.5.	TÉCNICAS PARA EL PROCESAMIENTO DE LA INFORMACIÓN Y ANÁLISIS DE LA INFORMACIÓN .....	37
3.6.	ASPECTOS ÉTICOS .....	38
<b>CAPÍTULO IV</b>		
<b>RESULTADOS</b>		
4.1.	CONFORMACIÓN LA BASE DE DATOS. ....	39
4.2.	DEFINICIÓN DE ATRIBUTOS DE LAS VARIABLES.....	39
4.3.	OBTENER LOS PESOS DE EVIDENCIA (WOE) DE LOS ATRIBUTOS.....	40
4.4.	TRATAMIENTO DE DATOS. ....	48
4.5.	MODELO OBTENIDO POR LA REGRESIÓN EN EL SOFTWARE SPSS. ....	48
4.6.	CONSTRUCCIÓN DE LA SCORECARD .....	54
4.7.	PUNTO DE CORTE O CUT OFF.....	57
<b>CAPÍTULO V .....</b>		<b>61</b>
<b>DISCUSIÓN .....</b>		<b>61</b>
<b>CONCLUSIONES.....</b>		<b>64</b>
<b>REFERENCIAS BIBLIOGRÁFICAS .....</b>		<b>66</b>

## **RESUMEN**

En esta investigación se presenta un modelo credit scoring utilizado como una herramienta complementaria para la decisión final en el otorgamiento o rechazo de la solicitud de crédito. Se trata de una metodología que pronostica el riesgo de incumplimiento de pagos futuros, a través de la designación de un score o puntaje. Basado en datos crediticios, financiero, laboral y sociodemográfico referente a los clientes.

## **ABSTRACT**

In this research a model is presented that is used as a complementary tool for the final decision in the granting or rejection of the credit application. It is a methodology that predicts the risk of default of future payments, through the designation of a score or score. Based on credit, financial, labor and socio-demographic data referring to clients.

## INTRODUCCIÓN

La calificación crediticia es una de las aplicaciones más exitosas de estadísticas en las finanzas y la industria bancaria hoy. Reduce el costo y el tiempo de aplicación procesamiento y le da flexibilidad para compensar el riesgo y las ventas por financiamiento institución. Dado que un punto muy importante de las entidades crediticias es contar con herramientas confiables que midan el riesgo crediticio para determinar si debe o no otorgarse un crédito. El scoring experto o estadístico es una herramienta que pronostica el riesgo que un cliente caiga en incumplimiento de pagos y a su vez explica el por qué; basado en la experiencia propia de la entidad financiera.

El estudio presenta las diferentes etapas que forman el proceso del credit scoring, como es la transformación de las variables en los pesos de evidencia (WOE) de sus respectivos atributos o características, dado este paso podremos entonces estimar la probabilidad de impago con la regresión logística y a partir de las probabilidades se genera un puntaje (score), a través del scorecard, donde se asignan puntajes en diferentes rangos de una variable predictiva. Las empresas crediticias deciden el mínimo puntaje (cut off) para aceptar un nuevo cliente, que suele ser en base a sus metas corporativas.

El modelo requiere de monitoreo conforme pasa el tiempo, para comprobar que se encuentra calibrado para que siga siendo efectivo. De tal modo que puedan ver variaciones para hacer los ajustes pertinentes.

## Capítulo I

### Planteamiento Del Problema

#### 1.1. Descripción de la Realidad Problemática

El sistema financiero siempre ha sido un pilar fundamental en la economía de un país, porque incluye las diversas formas de actividades financieras existentes. “El sistema financiero tiene cómo actividad central transferir los fondos de las personas que los tienen, a quienes tienen un déficit, por lo que el sistema financiero promueve una mayor eficiencia, ya que hace rentable el dinero de quien no lo necesita llevándolo a quien si lo hace producir. Este escritor también menciona que el buen funcionamiento del sistema financiero es un factor clave para el crecimiento de un país o el desempeño deficiente de éste”. (Mishkin, 2008, p. 123).

Las necesidades de liquidez de las personas que son cubiertas por las diversas ofertas de las entidades financieras han mostrado un impacto positivo en la economía. Sin embargo, es importante recalcar que cada vez que se concede u otorga un crédito la entidad financiera asume un riesgo de pérdida económica ya que existe la probabilidad de que el cliente incurra en un impago parcial y total.

En las instituciones que conceden créditos personales conllevan un trabajo mucho más intensivo, no solo para evaluar la concesión de un crédito sino también para encargarse de la cobranza y no caer en pérdida, a ello se suma el hecho de que los créditos personales por lo general son el producto que más se

producen.

De allí, la importancia de gestionar el riesgo para obtener rentabilidad de acorde con el nivel de riesgo asumido, proteger el capital financiero y conceder más tiempo en la asistencia de clientes más óptimos.

## **1.2. Formulación del problema**

Lo anterior impulsa a que se debe hacer un análisis para optimizar el riesgo crediticio o la probabilidad de impago en la concesión de créditos personales, la investigación se realizara en la empresa servicios financieros Grupo Buro, quienes a pesar que en la actualidad cuenta con un modelo estándar, siendo esta una herramienta para que puedan valorar el riesgo crediticio, el cual contiene normas internacionales y responde a la legislación financiera contemplada por la Superintendencia Financiera; sin embargo, se observa que esta herramienta internacional es muy generalizada ya que no siempre refleja la realidad local donde actúan, debido a que las empresas financieras varían de contexto por ejemplo por sus condiciones económicas, políticas, sociales, etc. Por ende, es importante llevar a cabo la investigación interna de dichos aspectos en la empresa servicios financieros Grupo Buro, para elaborar un modelo credit scoring más cercano a la realidad donde se encuentran operando.

En cuanto a la cartera de clientes de la empresa servicios financieros Grupo Buro, el área de mayor aplicación son los créditos personales y en consecuencia es donde se centra la mayor dificultad para evaluar el riesgo crediticio.

Este hecho es el que nos conduce a preguntarnos: ¿Será posible obtener la probabilidad de impago de los créditos personales futuros aplicando el Modelo Credit Scoring en la empresa servicios financieros Grupo Buro, Huaraz?

### **1.3. Objetivo de la Investigación**

#### **1.3.1. Objetivo General.**

Obtener la probabilidad de impago de los créditos personales futuros aplicando el Modelo Credit Scoring en la empresa servicios financieros Grupo Buro, Huaraz 2018.

#### **1.3.2. Objetivo Específicos.**

- Obtener un puntaje o score de decisión capaz de determinar la concesión de un crédito.
- Analizar los beneficios del Modelo Credit Scoring en la entidad financiera.

### **1.4. Justificación de la Investigación**

#### **1.4.1. Justificación Social.**

El trabajo en la empresa servicios financieros Grupo Buro contará con una herramienta que le permitirá tomar mejores decisiones al conceder créditos personales y disminuir la probabilidad de impago que no solo podrían en riesgo la rentabilidad de la empresa financiera sino también de otorgar el crédito a

clientes que no podrían solventar el pago del préstamo, afectándose económicamente.

#### **1.4.2. Justificación Institucional.**

El Acuerdo de Capital del Comité de Basilea propone a las entidades financieras hacer más énfasis en los modelos internos de medición de riesgo de crédito, por lo que un modelo Credit Scoring basado en la cartera de clientes de la misma empresa servicios financieros Grupo Buro, hace que la administración de sus recursos financieros sea más óptima ante la concesión de créditos.

#### **1.4.3. Justificación Práctica.**

El modelo Scoring es una herramienta de trabajo para ayudar a los analistas de crédito de la empresa servicios financieros Grupo Buro, ya que les permitirá prevenir y ayudar a la decisión de aprobar un préstamo o no dejando a un lado la subjetividad y los juicios personales, automatizando las decisiones de crédito.

#### **1.5. Limitación del estudio**

- La cartera de clientes de la Empresa servicios financieros Grupo Buro Huaraz, está incompleta.
- En la investigación no se encontraron limitaciones a nivel de recursos humanos porque no se necesitarán colaboradores, ni recursos económicos ya que se solventarán los gastos.

## **1.6. Viabilidad del estudio**

- El tema de investigación cuenta con suficiente acceso de información secundaria tanto en internet, libros, etc.
- La investigación tendrá como base la cartera de clientes de los últimos 3 años, de la entidad servicios financieros Grupo Buro, Huaraz.



## Capítulo II

### Bases Teóricas

#### 2.1. Antecedentes de la Investigación

- Banco Central de la República Argentina (2007) En su ensayo Anatomía de los modelos de Credit Scoring, se tiene que:  
*El trabajo revisó y buscó clarificar distintos aspectos asociados a los modelos de Credit Scoring. Aunque nada impide que estas técnicas se apliquen a grandes empresas, y de hecho existen algunos desarrollos para ellas, los modelos de Credit Scoring se emplean mayormente para evaluar individuos y pequeñas y medianas empresas, mientras que las grandes empresas se analizan a través de sistemas de rating... las entidades financieras emplean estas herramientas también con otros propósitos, como por ejemplo en el diseño de estrategias de marketing para ofrecer productos de manera proactiva y masiva... la revisión de la literatura especializada muestra que en su construcción predominan los enfoques econométricos (regresión logística, modelos probit) y desarrollos de inteligencia artificial, como los árboles de decisión. Los motivos para su predominio son básicamente dos: en general las metodologías relevadas muestran resultados similares, por lo*

*que tienden a emplearse aquellas cuyo funcionamiento e interpretación son*

*más sencillos, en contraposición a enfoques más sofisticados y de difícil interpretación, como ser las redes neuronales.*

- Arenas, Boccardi y Piñeyrúa (2012), en su tesis Evaluación del Riesgo crediticio de la Cartera de Microcréditos de una institución financiera en Uruguay, afirman que:

*La discusión acerca del papel que cumple la calificación estadística de los candidatos a prestatarios en las microfinanzas, haciendo especial hincapié en la utilidad que los modelos de Credit Scoring pueden tener a la hora de discriminar correctamente entre “buenos” y “malos” prestatarios, así como al diseñar políticas crediticias efectivas que contribuyan a mitigar el riesgo de crédito... Si la institución decide no aplicar el modelo de Credit Scoring, de todas formas, los resultados obtenidos podrían ser útiles a la hora del diseño de las políticas crediticias... Es importante señalar que el modelo fue confeccionado con una muestra pequeña y que sólo incluye información de créditos aprobados, por lo que el mismo sólo podría aplicarse a microempresarios que cumplan ciertos requisitos mínimos de aprobación establecidos por la institución.*

- Sarco (2015) En su tesis de Factores que determinan el otorgamiento de crédito de la financiera Credinka en la ciudad de Ayaviri. Afirma que:

*Las conclusiones de su investigación fueron: Los factores influyentes para la determinación de préstamos financieros en la financiera Credinka son: el Ingreso económico, Gasto, número de Hijos, Edad del cliente, Seguro de salud, tipo de Vivienda y Material de construcción de la vivienda. Los factores con mayor influencia son la edad con 0.037 y el ingreso con 0.062 de significancia en vista de que ambos se encuentran en el rango entre 0 y 1 más cercanos al cero... Mediante la curva ROC se obtuvo que el ingreso mínimo recomendado es de S/.825.00 y el gasto máximo recomendado es de S/.580.00 nuevos soles mensuales. El ahorro para cubrir prestamos asciende a S/. 245.00 nuevos soles mensuales.*

## **2.2. Bases Teóricas**

### **2.2.1. Credit Scoring.**

El Credit Scoring consta de algoritmos que evalúan de manera automática el riesgo de crédito de un solicitante de financiamiento o de alguien que ya es cliente de la entidad. El Credit Scoring emplea principalmente información del evaluado contenida en las solicitudes de crédito en fuentes internas o externas de información, de este modo la evaluación refleja una asignación que permite

comparar y ordenar en función del riesgo y cuantificarlo de manera individual (Banco Central de la República Argentina, 2007, p. 3).

Si bien la escala del score puede oscilar entre 0 y 1. Generalmente el Credit Scoring presenta una relación inversa entre el score y el riesgo, es decir, con una mayor calificación existirá una caída marginal en el riesgo; sin embargo, esto depende de la construcción del modelo. (Escalona, 2011, p. 22).

Es de suma importancia destacar que el Credit Scoring en micro finanzas no elimina el uso de asesores y directores de crédito, ya que solo es una herramienta que ayuda al personal en la toma de decisiones acerca de otorgar o no un microcrédito, tal como lo dice Schreiner (2002, p. 239).

### **2.2.2. Metodologías aplicadas en los modelos de Credit Scoring.**

Existen diversos métodos estadísticos y no estadísticos empleados para la construcción de modelos de Credit Scoring como el análisis discriminante, redes neuronales, Scoring basado en el juicio, regresión lineal, regresión logística, métodos de programación matemática, arboles de decisión, sistemas expertos, entre otros. (Morales, 2014, p. 55 - 56).

### **2.2.3. Tipo de Score.**

La colección de técnicas que conforman el scoring tiene como propósito principal generar un puntaje de riesgo a las solicitudes de crédito o a cuentas ya existentes. (Nieto, 2010, p. 47 a)

#### **2.2.3.1. Modelo experto.**

Es cuando el modelo se construye con información de otras instituciones; esto es, está listo para usar. Se trata de modelos de crédito genéricos que se compran a consultores externos, que son adaptaciones de modelos hechos en otras matrices, etc. Esto es común en instituciones que apenas se inician como prestamistas dado que no tiene historial de sus clientes. Estos modelos con similitudes de población se utilizan en primera instancia, aunque no es lo más conveniente. (Nieto, 2010, p 49 b)

#### **2.2.3.2. *Modelo estadístico.***

Son modelos que se construyen con información propia, son conocidos como modelos in-house. Tiene como beneficio que se pueden construir modelos específicos para distintos segmentos de la población. Se maneja información propia flexible y manejable. Se adquiere conocimiento y experiencia propia sobre la población, y también habilidad en el diseño e interpretación de los resultados. Se conserva la confidencialidad de la información. (Nieto, 2010, p 49 c)

#### **2.2.4. Pesos de Evidencia (WOE).**

El poder de predicción en cada atributo o grupo de atributos se calcula con los Pesos de Evidencia (Weigth of Evidence) que se denotan como WOE, que es una medida entre la diferencia de las proporciones de buenos y malos en cada atributo. (Nieto 2010). La definición del WOE es:

$$\begin{aligned}
WOE_{ij} &= 100 \cdot \ln \left( \frac{\text{Distribucion de buenos en el atributo } j \text{ de la característica } i}{\text{Distribucion de malos en el atributo } j \text{ de la característica } i} \right) \\
&= 100 \cdot \ln \left( \frac{Pb_{ij}}{Pm_{ij}} \right) = 100 \cdot \ln \left( \frac{b_{ij} \cdot m_i}{m_{ij} \cdot b_i} \right)
\end{aligned}$$

Cuando se obtienen valores negativos del WOE significa que se tienen proporciones altas de malos sobre los buenos. De esta medida se toman en cuenta algunas consideraciones:

- El  $WOE_{ij}$  varía dependiendo de la forma en que se agrupan los atributos. Se acostumbra a ordenar de manera creciente al  $WOE$ .
- Para que el  $WOE_{ij}$  este definido, ninguna de las clases debe estar formada únicamente por buenos o por malos.
- Se sugiere no tener más de 8 clases y para que cada clase sea significativa debe contener al menos un 5% de los datos analizados.

Esto permite identificar datos atípicos y clases raras, además de identificar comportamientos y adquirir conocimiento acerca del portafolio.

### 2.2.5. Pruebas sobre la Clasificación de Atributos.

### 2.2.5.1. Estadístico $X^2$ .

Suponiendo que la proporción de cuentas buenas y cuentas malas en el atributo j de la característica i coincide con la proporción de cuentas buenas y cuentas malas en la característica i, entonces el estimador del valor esperado de las cuentas buenas y malas en el atributo j es igual a:

$$\hat{b}_{ij} = \frac{(b_{ij} + m_{ij})b_i}{b_i + m_i} \text{ y } \hat{m}_{ij} = \frac{(b_{ij} + m_{ij})m_i}{b_i + m_i}$$

y el estadístico  $X_c^2$  para la característica i está dado por:

$$X_c^2 = \sum_{j=1}^{n_i} \frac{(b_{ij} - \hat{b}_{ij})^2}{\hat{b}_{ij}} + \frac{(m_{ij} - \hat{m}_{ij})^2}{\hat{m}_{ij}} \sim X_{k-1}^2$$

Este valor será "pequeño" si  $b_{ij} \approx \hat{b}_{ij}$  y  $m_{ij} \approx \hat{m}_{ij}$  será "grande" en caso contrario. De esta manera  $X_c^2$  se usa para determinar cuan diferentes son los odds en cada clase. Entre más grande resulte el estadístico  $X_c^2$  refleja mayores diferencias en los odds, entonces al comparar dos agrupaciones se prefiere la de mayor valor de  $X_c^2$ , ya que de esta manera la probabilidad de equivocarse al rechazar que los atributos de la característica i predicen mal es pequeña  $P(X^2 > X_c^2)$ .

### 2.2.5.2. Valor de Información (IV).

El Valor de Información (IV) es una medida de entropía que aparece en la teoría de Siddiqi (2006) y es una medida del poder de predicción global de la característica, y se define como:

$$IV = \sum \left( \frac{b_{ij}}{b_i} - \frac{m_{ij}}{m_i} \right) \ln \left( \frac{b_{ij} \cdot m_i}{m_{ij} \cdot b_i} \right)$$

Los valores que puede tomar el estadístico IV son no negativos, y es cero cuando  $\frac{b_{ij}}{b_i} = \frac{m_{ij}}{m_i}$  lo que equivale, directamente de la definición, que  $b_{ij} = \hat{b}_{ij}$  y  $m_{ij} = \hat{m}_{ij}$ .

Siddiqi (2006) considera que una característica con un IV

- menor a 0.02 es tiene nulo valor predictivo
- entre 0.02 y 0.1 es de predicción débil
- entre 0.1 y 0.3 es de predicción media
- más de 0.3 es de predicción fuerte. (IV mayor a 0.5 sobre predice)

Siddiqi (2006) aconseja que las características con  $IV < 2\%$  deben excluirse del modelo.

### 2.2.6. Regresión Logística.



La regresión logística forma parte del conjunto de métodos estadísticos que caen bajo tal denominación y es la variante que corresponde al caso en que se valora la contribución de diferentes factores en la ocurrencia de un evento simple. (Santiago, 2011)

En general, la regresión logística es adecuada cuando la variable de respuesta  $Y$  es politómica (admite varias categorías de respuesta, tales como mejora mucho, empeora, se mantiene, mejora, mejora mucho), pero es especialmente útil en particular cuando solo hay dos posibles respuestas (cuando la variable de respuesta es dicotómica). (Santiago, 2011).

### **2.2.7. Regresión logística binaria.**

La regresión logística es un grupo de técnicas estadísticas que tienen como objetivo comprobar hipótesis o relaciones causales cuando la variable dependiente es una variable binaria, es decir, de solo dos categorías, también conocidas como dummy o dicotómica. Por ejemplo: 1-Sí, 0-No o por ejemplo 1-Soltero, 0-No Soltero. (Cárdenas 2014).

**Definición:** Sea  $Y$  una variable dependiente binaria que toma dos valores posibles (0 y 1). Sean  $x_1 \dots x_k$  un conjunto de variables independientes observadas con el fin de explicar y/o predecir el valor de  $Y$ .

Se construye un modelo:

$$P[Y = 1/x_1 \dots x_k] = p(x_1 \dots x_k; \beta)$$

Donde  $p(x_1 \dots x_k; \beta)$  es una función que recibe el nombre de función de enlace (función de probabilidad) cuyo valor depende de un vector de parámetros.

$$\beta = (\beta_1, \dots, \beta_k)$$

El modelo de regresión logística parte de la hipótesis de que los datos siguen el siguiente modelo:

$$\ln\left(\frac{p}{1-p}\right) = b_0 + b_1x_1 + \dots + b_kx_k + u = xb + u$$

Con el fin de simplificar la notación, definimos Z:

$$z = b_0 + b_1x_1 + \dots + b_kx_k$$

Por lo tanto, el modelo se puede representar como:

$$\ln\left(\frac{p}{1-p}\right) = z + u$$

Como la función de distribución logística es:

$$F_{(x)} = \frac{e^x}{1 + e^x}$$

Por tanto, podemos reescribir el modelo de forma mucho más compacta:

$$p = \frac{e^{z+b}}{1 + e^{z+b}}$$

$$x = F_{(z)} = F(x * b)$$

Recordemos las dos formas más importantes de expresar el modelo de regresión logística:

$$\ln p - \ln(1 - p) = b_0 + b_1x_1 + \dots + b_kx_k$$

La primera expresión se llama logit y a la segunda Odds ratio o cociente de probabilidades.

### 2.2.8. Pruebas de significación.

Una vez estimado los coeficientes del modelo, se tiene que verificar si el modelo predice de manera adecuada a la variable dependiente. Para evaluar la bondad del modelo se utiliza el logaritmo del cociente de verosimilitud y la prueba de Hosmer-Lemeshow.

#### 2.2.8.1. Verosimilitud.

Se prueba la hipótesis de que el modelo encontrado es el que mejor se ajusta a través de la razón de verosimilitud.

$$D = -2\ln \frac{\text{verosimilitud del modelo sin la variable (modelo analizado)}}{\text{verosimilitud del modelo sin la variable (modelo saturado)}}$$

La diferencia entre estos dos valores de  $-2\ln$  se llama Devianza y prueba si la o las variables  $X_i$  son significativas.

$$H_0: b_0 = b_1 = \dots = b_k$$

$H_1$ : Algún  $\beta_i$  distinto de 0

Estadístico de prueba  $D \sim \chi^2$  con  $n-k-1$  grados de libertad. Región de Rechazo: si  $D > \chi^2_{\alpha, (n-k-1)}$  si se rechaza  $H_0$ , significa que al menos uno de los coeficientes es diferente de cero y la variable correspondiente es significativa con respecto a la probabilidad de que ocurra o no el suceso en estudio.

#### 2.2.8.2. *Hosmer-Lemeshow.*

Cuando el número de variables del modelo es grande y/o tiene variables independientes cuantitativas, se debe utilizar la prueba de Hosmer-Lemeshow, como consecuencia de la gran cantidad de variables, el número de pautas de variación existente entre ellas es tan elevado que puede invalidar la utilidad de los estadísticos de bondad de ajuste clásicos. Evalúa la bondad del modelo construyendo una tabla de contingencia, divide la muestra en 10 grupos a partir de los deciles de las probabilidades estimadas (a los 10 grupos resultantes se les denomina deciles de riesgo).

En cada decil de riesgo se calcula el número de casos que pertenecen a cada categoría de la variable dependiente (número observado) y el número de casos que el modelo pronostica que pertenecen a cada categoría de la variable dependiente (número esperado). A continuación, se compara el número de casos observado con el número de casos esperado (esta comparación se

realiza en cada una de las 20 casillas definidas por la combinación de las 2 categorías de la variable dependiente con los 10 deciles de riesgo).

Este estadístico permite contrastar la hipótesis nula de igualdad de distribuciones o, que no existen diferencias entre las frecuencias observadas y las esperadas. El estadístico de contraste tiene siempre  $k$  grados de libertad. Las hipótesis que se contrasta es:

$H_0$ : El modelo es adecuado

$H_1$ : El modelo ajustado no es adecuado

Decisión si el estadístico de prueba es mayor o igual que  $X^2_{\alpha, (n^{\circ} \text{ de grupos}-2)}$ , se rechaza  $H_0$  y se concluye que el modelo no es adecuado con un nivel de significación  $\alpha$ .

## **2.2.9. Bondad del Modelo.**

### **2.2.9.1. *Significación de chi-cuadrado del modelo en la prueba ómnibus.***

Si la significación es menor de 0,05 indica que el modelo ayuda a explicar el evento, es decir, las variables independientes explican la variable dependiente.

### **2.2.9.2. *R-cuadrado de Cox y Snell, y R-cuadrado de Nagelkerke.***

Indica la parte de la varianza de la variable dependiente explicada por el modelo. Hay dos R-cuadrados en la regresión logística, y ambas son válidas. Se acostumbra a decir que la parte de la variable dependiente explicada por el modelo oscila entre la R-cuadrado de Cox y Snell y la R-cuadrado de Nagelkerke. Cuanto más alto es la R-cuadrado más explicativo es el modelo, es decir, las variables independientes explican la variable dependiente.

### **2.2.9.3. *Porcentaje global correctamente clasificado.***

Este porcentaje indica el número de casos que el modelo es capaz de predecir correctamente. Me explico. En base a la ecuación de regresión y los datos observados, se realiza una predicción del valor de la variable dependiente (valor pronosticado). Esta predicción se compara con el valor observado. Si acierta, el caso es correctamente clasificado. Si no acierta, el caso no es correctamente clasificado.

Cuanto más casos clasifica correctamente (es decir coincide el valor pronosticado con el valor observado) mejor es el modelo, más explicativo, por tanto, las variables independientes son buenos predictores del evento o variable dependiente. Si el modelo clasifica correctamente más del 50% de los casos, el modelo se acepta. Si no, punto y final y a volver a empezar, y seleccionaríamos nuevas variables independientes. Los siguientes pasos son para evaluar la relación de cada variable independiente con la variable dependiente.

#### **2.2.10. Pruebas de coeficientes.**

Una vez encontrado el mejor conjunto de variables explicativas que predicen la variable dependiente Y, se debe evaluar cada coeficiente para determinar cuáles ingresan al modelo, este proceso se realiza mediante el estadístico Wald.

##### **2.2.10.1. Estadístico de Wald.**

Para determinar si una variable debe ser incluida en un modelo porque tiene un peso significativo se aplica la prueba de estadístico de Wald.

La prueba resulta de contrastar la hipótesis nula:

$$H_0: \beta_i = 0$$

Contra la alternativa

$$H_1: \beta_i \neq 0$$

Con un estadístico de prueba definido como:

$$w_j = \frac{\hat{\beta}_i}{s(\hat{\beta}_i)}$$

Que bajo el supuesto que  $H_o$  es cierto sigue una distribución  $t$  con  $n - p - 1$  grados de libertad y para muestras grandes se distribuye como una normal estándar. (Santiago, 2011). Se entiende que si  $w_i$  es un valor menor que 0.05, se rechaza esa hipótesis nula que afirma que ese coeficiente es cero, y se entiende entonces que ese coeficiente no es cero y que, por lo tanto, el modelo es útil para representar una determinada relación. Si, por el contrario, el p-valor es mayor que 0.05 eso significa que el valor del coeficiente podría ser perfectamente cero y estar viendo lo que vemos, por lo tanto, esa variable no influye a la hora de determinar la variable dependiente del modelo de regresión. (Jaume Llopis Pérez, 2013).

#### **2.2.11. Scorecard.**

Una scorecard es una tabla que contiene los puntajes asignados a cada atributo, grupo o característica de cada una de las variables usadas (Thomas, 2002). El puntaje determina la probabilidad de pago de la deuda para un cliente cuando se le otorgue una tarjeta de crédito. Así que a mayores puntajes corresponden a una mayor probabilidad de pago. La compañía prestamista es la que define finalmente la probabilidad mínima de pago para determinar cuando un cliente



es considerado bueno; esto es, el puntaje de separación entre clientes buenos y malos; esto es, a partir de que puntaje un cliente se hará acreedor de un crédito.

### 2.2.12. Construcción de la Scorecard.

La scorcard se construye con los estimadores de los parámetros de la regresión logística. Los puntajes del score son resultado de un reescalamiento y una traslación del modelo logístico, dado por la ecuación

$$Score = Offset + Factor \cdot \ln(odds)$$

- Donde offset es un término de traslación (o compensación) y Factor es un término de reescalamiento.
- Offset y Factor deben satisfacer condiciones impuestas por la empresa de crédito. Este procedimiento permite la estandarización del score.
- Los valores de la scorecard son resultado de una transformación de los coeficientes  $\beta_i$  del modelo de regresión logística. La transformación tiene como finalidad obtener valores enteros para cada atributo  $j$  de la característica  $i$  en un rango considerado adecuado por la empresa crediticia.
- Se acostumbra a calibrar la scorecard de tal manera que cada cierto incremento en el puntaje  $P_0$ , se obtenga el doble de la relación good/bad.

La probabilidad  $P_0$ , se puede derivar usando la función exponencial:

$$P_0 = \frac{\exp(\sum b_i \times x_i + b_0)}{1 + \exp(\sum b_i \times x_i + b_0)}$$

Para obtener los valores de Offset y Factor se resuelve el siguiente sistema de ecuaciones:

$$Score = Offset + Factor \cdot \ln(odds)$$

$$Score + P_0 = Offset + Factor \cdot \ln(2 \cdot odds)$$

$$\ln(odds) = - \left( \sum b_i \times x_i + b_0 \right)$$

De aquí obtenemos:

$$Factor = \frac{P_0}{\ln(2)}$$

$$Offset = Score - Factor \cdot \ln(odds)$$

Estos puntajes están en escala lineal y la relación del modelo de regresión logística con los WOE está dada por:

$$Score = Offset + Factor \cdot \ln(odds)$$

$$= Offset + Factor \cdot \ln(b_0 + \sum_{i=1}^p \hat{b}_i \cdot woe_{ij})$$

$$= Offset + Factor \cdot \hat{b}_0 + Factor \cdot \left( \sum_{j=1}^p \hat{b}_j \cdot woe_{ij} \right)$$

De esta última ecuación se puede ver que los puntajes en la scorecard se pueden descomponer en un puntaje inicial dado por:

$$Offset + Factor \cdot \hat{b}_0$$

y el puntaje asociado al atributo j de la característica i dado por:

$$\text{Factor. } \hat{b}_i \cdot \text{woe}_{ij}$$

Es claro que los puntajes de una scorecard dependen de los parámetros de traslación y re escalamiento que se utilicen:

$$\text{Offset} + \text{Factor}$$

### 2.2.13. Determinación del punto de corte o Cut Off.

Cuando se tiene los datos de un nuevo solicitante, se calcula su score y con el resultado se decide si se le otorga o no el crédito. Si  $\text{score} > a$  se otorga el crédito, en caso contrario si  $\text{score} \leq a$  se rechaza la solicitud. El punto “a” se conoce como punto de corte o Cut Off y es importante determinarlo para optimizar la decisión.

El valor de “a” responde a la siguiente ecuación:

$$\frac{e^{\hat{\beta}_0 + \hat{\beta}_1^T x}}{1 + e^{\hat{\beta}_0 + \hat{\beta}_1^T x}} = \hat{p} > \frac{1}{2}$$

De aquí sigue que

$$e^{\hat{\beta}_0 + \hat{\beta}_1^T x} > \frac{1}{2} (1 + e^{\hat{\beta}_0 + \hat{\beta}_1^T x})$$

$$\frac{1}{2} e^{\hat{\beta}_0 + \hat{\beta}_1^T x} > \frac{1}{2} \Rightarrow e^{\hat{\beta}_0 + \hat{\beta}_1^T x} > 1$$

$$\Rightarrow \hat{\beta}_0 + \hat{\beta}_1^T x > 0$$

$$\Rightarrow a = \textit{Offset}$$

Para obtener el punto de corte es calcular el score para todas las cuentas de la base, estos score's se ordenan y el valor de "a" satisface la ecuación:

$$\frac{\textit{score} > a}{\textit{score}} = \textit{porcentaje}$$

El porcentaje es un valor seleccionado por la empresa.

## **2.3. Definiciones conceptuales**

### **2.3.1. Crédito.**

La palabra "crédito" es una derivación del latín credo, "yo creo". El crédito implica la confianza entre un deudor y un acreedor. El crédito es la prestación de un servicio a una persona o entidad, la cual, al cabo del tiempo convenido o de los pactos estipulados deberá devolverlo.

Es decir, el crédito es el acto de confianza del acreedor en el deudor, esperando que éste cumpla todo lo estipulado para dicho fin, existiendo el riesgo del incumplimiento o insolvencia, por lo cual es necesario establecer garantías y expedir documentos de carácter legal que precisen los términos en que se da el crédito.

### **2.3.2. Riesgo de crédito.**

Conocido como insolvencia, hace referencia a la incertidumbre del rendimiento de la inversión debida a la posibilidad de pérdidas por la incapacidad o falta de voluntad de los deudores, contrapartes, o terceros obligados, para cumplir sus obligaciones contractuales registradas dentro o fuera del balance.

### **2.3.3. Score.**

es un puntaje basado en un modelo estadístico que pronostica la probabilidad de cumplimiento de una persona en el pago de sus obligaciones. El resultado es un número, o puntaje, que el mercado analiza para determinar niveles de riesgo crediticio de la persona. Cada entidad otorgadora de crédito fija sus políticas; en consecuencia, un mismo puntaje puede generar diferentes decisiones según la entidad que está analizando el caso.

### **2.3.4. La Superintendencia de Banca y Seguros.**

La Superintendencia de Banca y Seguros (SBS) es el organismo encargado de la regulación y supervisión de los Sistemas Financieros, de Seguros y, a partir del 25 de julio del 2000, del Sistema Privado de Pensiones (SPP) (Ley 27328). La labor de la SBS comprende dos tareas básicas: regulación y supervisión. La regulación establece las reglas a las cuales se someten las empresas supervisadas desde su entrada al sistema, durante su operación y eventual salida del mercado. La SBS tiene como fin generar las condiciones que permitan maximizar el valor de los sistemas financiero, de seguros y privado de pensiones, propiciando una mayor confianza y adecuada protección de los intereses del público usuario, a

través de la transparencia, veracidad y calidad de la información; cuidando la estabilidad y solvencia de las instituciones que conforman los sistemas.

El Comité de Supervisión Bancaria de Basilea con el objetivo de formular recomendaciones para la regulación de instituciones financieras y enfrentar de manera más eficaz las inestabilidades producidas por un mercado financiero mundial.

Debido a los vacíos regulatorios a nivel mundial del sistema y la rápida innovación de los productos financieros, estas recomendaciones fueron adoptadas por los entes reguladores de la mayoría de los países del mundo. Posteriormente, las entidades desarrollaron mejoras en la administración del riesgo, incorporando sus planteamientos estratégicos y operacionales en el manejo de este factor. En el año 1999 el Comité de Basilea se reunió por segunda vez y se creó un acuerdo: Basilea II, hecho público en 2004 y una versión más completa en 2006, donde se amplía el tratamiento de los riesgos a los que se enfrenta el sistema financiero; teniendo en cuenta además del riesgo de crédito, a los riesgos operacionales y los de mercado. El acuerdo Basilea II, consta de tres grandes pilares:

- i. ***El Pilar I aborda los requisitos de capital que originan los tres principales riesgos que enfrentan los bancos.***

Los riesgos de crédito, de mercado y operacional. Este pilar puede considerarse una extensión del Primer Acuerdo de Capital, o Basilea I, que el Comité de Supervisión Bancaria de Basilea emitió en 1988.

ii. *El Pilar II se refiere al proceso supervisor de la suficiencia de capital en los bancos.*

Teniendo en cuenta todos sus riesgos, incluyendo aquellos considerados en el Pilar I. El Pilar II enfatiza el proceso interno en cada banco de evaluación de su suficiencia de capital en relación con todos sus riesgos materiales o importantes. La autoridad supervisora basa su examen supervisor en los resultados de este proceso.

iii. *El Pilar III considera la divulgación de información material o importante sobre riesgos y el capital por parte de los propios bancos.*

Ello, en función de exigentes requisitos de transparencia y oportunidad, a fin de que los depositantes e inversionistas puedan ejercer disciplina de mercado a través de decisiones informadas a partir de la situación financiera de los bancos.

Su fundamento es que los requerimientos de capital sean más sensibles al riesgo, especialmente al riesgo de crédito. La propuesta del nuevo acuerdo hace más énfasis en los modelos internos de medición de riesgo de crédito de cada banco, la revisión del supervisor y la disciplina del mercado; es más flexible, tiene varios enfoques e incentivos, pero es también más complejo.

#### **2.4. Formulación de la hipótesis.**

El Modelo Credit Scoring obtiene la probabilidad de impago de créditos futuros en la empresa servicios financieros Grupo Buro, Huaraz 2018.





## **Capítulo III**

### **Metodología**

#### **3.1. Diseño Metodológico**

##### **3.1.1. Tipo de estudio.**

De acuerdo a los objetivos planteados el estudio a desarrollarse es:

- **Cuantitativa:** la investigación es cuantitativa, ya que los datos con que se trabajara la investigación son numéricos.
- **Retrospectiva:** la ocurrencia de hechos se encuentra en la cartera de clientes de la entidad financiera en mención.
- **Correlación:** porque en la investigación los datos serán usados para predecir, para lo cual se medirá el grado de relación de varias variables.

##### **3.1.2. Procedimiento de la investigación.**

El modelo Credit Scoring experto o estadístico es una herramienta que permite predecir la probabilidad de impago. Para ello, se diseñó una metodología que analiza el riesgo de impago en la concesión de créditos. Los datos para tal fin son de carácter socio demográfico, laboral, financiero y crediticio, datos internos es decir de la misma empresa, basados en su experiencia. De esta manera se pudo estimar la probabilidad de casos buenos y malos que representan los atributos o características de cada variable y a partir de estas probabilidades se

genera un puntaje, es decir un score para indicar un nivel de riesgo. La empresa crediticia decide el mínimo puntaje (cut off) para aceptar un nuevo cliente, ya que como empresa Outsourcing está sujeta a parámetros y políticas que el banco le impone.

### **3.2. Población**

La población de la presente investigación está conformada por los clientes de la empresa servicios financieros Grupo Buro, Huaraz, los clientes considerados serán aquellos que tengan de 12 a más cuotas programas de pago, este criterio temporal se fijara ya que recoge una madurez adecuada determinado según los especialistas y antecedentes.

### **3.3. Operacionalización de variables**

Las variables independientes que se seleccionaron para la predicción del modelo, se ha tenido en cuenta los factores que la entidad financiera considera para medir el nivel de riesgo en las operaciones de crédito que concede a sus clientes; se clasifican en cuatro factores:

- Factor Crediticio: Variables relativas a la operación de crédito o préstamo, como son: Promedio de crédito otorgado, garantía, Experiencia crediticia SBS, Plazos a Pagar, Cantidad de tarjeta.
- Factor Financiero: variables relacionadas con la disponibilidad de ingresos

para pagar el crédito, como son: Endeudamiento e Ingreso.

- Factor Laboral: variables que reflejan la situación profesional del cliente, como son: tipo de trabajo, tiempo de empleo.
- Factor Socio-demográfico: variables relativas a las características personales del cliente, como son: sexo, edad, tipo de vivienda, estado civil, nivel educativo, número de hijos, valorización de bienes.

Tabla 1

*Variables de la Investigación*

Variable	Dimensión	Tipo de Variable
<b>Dependiente</b>	1. Riesgo Crediticio	Dicotómica
<b>Independiente</b>	1. Edad	Continua
	2. Sexo	Cualitativa
	3. Estado civil	Cualitativa
	4. Nivel educativo	Cualitativa
	5. Número de hijos	Discreta
	6. Valorización de bienes	Continua
	7. Tipo de vivienda	Cualitativa
	8. Tipo de trabajo	Cualitativa
	9. Tiempo de empleo	Continua
<b>Independiente</b>	10. Ingreso	Continua
	11. Endeudamiento	Continua

	12. Promedio de Crédito Otorgado	Continua
	13. Garantía	Cualitativa
	14. Experiencia crediticia SBS	Continua
	15. Plazos a Pagar	Continua
	16. Cantidad de tarjeta debito/ crédito	Continua

Elaboración: propia

### **3.4. Técnica de recolección de datos**

El instrumento de recolección de datos que se usó en nuestra investigación es la base datos que se extrajo del formulario de solicitud de crédito, información externa proveniente de empresas financieras especializadas y del historial del prestatario de la empresa servicios financieros Grupo Buro, Huaraz. Dicha información es entregada directamente de la empresa, quienes han recopilado información del cliente al momento de la concesión del préstamo.

### **3.5. Técnicas para el procesamiento de la información y análisis de la información**

La construcción del modelo de regresión logística binaria se lleva a cabo con el apoyo del software Statistical Product and Service Solutions (SPSS) Versión 19 ya que es un programa estadístico fácil de utilizar cuando se trabaja con grandes bases de datos.

### **3.6. Aspectos Éticos**

El investigador se compromete a desarrollar la presente investigación con datos reales y sin alteración alguna de los mismos. Respetando la confidencialidad de los datos obtenidos durante la investigación.

## **Capítulo IV**

### **Resultados**

#### **4.1. Conformación la Base de Datos.**

Se inició con el vaciado de la información contenida en los formularios de solicitudes de los clientes, información externa proveniente de empresas financieras especializadas y el registro histórico del prestatario. Los casos que se excluyeron son aquellos que tenían menos de 12 cuotas de pago, porque no presentan suficiente madurez crediticia para analizar el riesgo crediticio.

#### **4.2. Definición de Atributos de las Variables**

Después de tener la base se procedió a formar intervalos de clase o grupos de clase, estas clases también se les llama atributos, para las variables: Edad, Ingreso, Monto otorgado, Plazo a pagar y Margen de endeudamiento. Para hallar la agrupación de atributos que mejor predice se utilizó Valor de Información (IV) que es una función de la proporción de buenos y malos clientes en los atributos de cada atributo, donde la agrupación que tuvo mayor valor en IV es la que se determinó como la mejor agrupación de atributos.

En el caso de la variable: Número de hijos, se observó que menos del 2% de casos no tienen hijos, al ser información insuficiente se agrupó con aquellos que tienen un hijo obteniendo un valor mayor en IV que eliminando el atributo. Las

demás variables que son nominales y/o dicotómicas presentan atributos con más del 2% de casos representando suficiente información para continuar con su análisis, por lo que los atributos de estas variables ya están definidos.

#### 4.3. Obtener los Pesos de Evidencia (WOE) de los Atributos

Para hallar el valor de WOE se registró para cada atributo la cantidad de casos de buenos y malos pagadores que contenga, el siguiente paso es obtener la probabilidad de buenos y malos pagadores de cada atributo para luego aplicar la ecuación WOE.

Tabla 2

*WOE o Pesos de Evidencia para la variable “Edad”*

<b>Atributo/intervalo de edad</b>	<b>Malo</b>	<b>Bueno</b>	<b>Pm</b>	<b>Pb</b>	<b>WOE</b>
< 25	80	110	0.2667	0.1571	-52.8844129
26-35	118	280	0.3933	0.4000	1.68071183
36-45	55	171	0.1833	0.2443	28.7032511
46-55	28	87	0.0933	0.1243	28.6405748
> 56	19	52	0.0633	0.0743	15.9506879
<b>Total</b>	<b>300</b>	<b>700</b>	<b>1</b>	<b>1</b>	

Fuente: Grupo Buro

La variable se agrupo en 5 intervalos o atributos, debido a que con esta agrupación de los valores de la variable se obtuvo una mayor medición en el  $X^2$  y el Valor de Información (IV), que otras formas de agrupación.

Tabla 3

*WOE o Pesos de Evidencia para la variable “Sexo”*

<b>Atributo/ Sexo</b>	<b>Malo</b>	<b>Bueno</b>	<b>Pm</b>	<b>Pb</b>	<b>WOE</b>
M	191	499	0.6367	0.7129	11.3034807
F	109	201	0.3633	0.2871	-23.5340835
<b>Total</b>	<b>300</b>	<b>700</b>	<b>1</b>	<b>1</b>	

Fuente: Grupo Buro

El WOE para el atributo F (femenino) tiene valor negativo lo que significa que se tiene para este atributo proporciones altas de malos sobre los buenos.

Tabla 4

*WOE o Pesos de Evidencia para la variable "Estado Civil"*

<b>Atributo/ Estado Civil</b>	<b>Malo</b>	<b>Bueno</b>	<b>Pm</b>	<b>Pb</b>	<b>WOE</b>
Casado	120	237	0.4000	0.3386	-16.6729462
Separado	34	61	0.1133	0.0871	-26.2784521
Soltera	146	402	0.4867	0.5743	16.5547607
<b>Total</b>	<b>300</b>	<b>700</b>	<b>1</b>	<b>1</b>	

Fuente: Grupo Buro

El WOE para los atributos Casados y Separados se observa que tienen valor negativo lo que significa que se tiene proporciones altas de malos sobre los buenos.

Tabla 5

*WOE o Pesos de Evidencia para la variable "Estado Civil"*

<b>Atributo/Otra Deuda</b>	<b>Malo</b>	<b>Bueno</b>	<b>Pm</b>	<b>Pb</b>	<b>WOE</b>
--------------------------------	-------------	--------------	-----------	-----------	------------



Ninguno	272	635	0.9067	0.9071	0.05250722
Co- solicitante	18	23	0.0600	0.0329	-60.2175402
Garante	10	42	0.0333	0.0600	58.7786665
<b>Total</b>	<b>300</b>	<b>700</b>	<b>1</b>	<b>1</b>	

Fuente: Grupo Buro

El WOE para el atributo Co-solicitante se observa que tiene valor negativo lo que significa que se tiene proporciones altas de malos sobre buenos.

Tabla 6

*WOE o Pesos de Evidencia para la variable "Tipo de Vivienda"*

<b>Atributo/Tipo Vivienda</b>	<b>Malo</b>	<b>Bueno</b>	<b>Pm</b>	<b>Pb</b>	<b>WOE</b>
Alquiler	70	109	0.2333	0.1557	-40.444522
Propio	186	527	0.6200	0.7529	19.4156014
Familiar	44	64	0.1467	0.0914	-47.2604411
<b>Total</b>	<b>300</b>	<b>700</b>	<b>1</b>	<b>1</b>	

Fuente: Grupo Buro

El WOE para los atributos Alquiler y Familiar se observa que tienen valor negativo lo que significa que se tiene para estos atributos proporciones altas de malos sobre los buenos.

Tabla 7

*WOE o Pesos de Evidencia para la variable "Intervalo de Ingreso"*

<b>Atributo/ intervalo de ingreso</b>	<b>Malo</b>	<b>Bueno</b>	<b>Pm</b>	<b>Pb</b>	<b>WOE</b>
---	-------------	--------------	-----------	-----------	------------

< 1500	162	326	0.5400	0.4657	-14.7996814
1500 a 2500	124	325	0.4133	0.4643	11.6245756
>2500	14	49	0.0467	0.0700	40.5465108
<b>Total</b>	<b>300</b>	<b>700</b>	<b>1</b>	<b>1</b>	

Fuente: Grupo Buro

La variable se agrupo en 3 intervalos o atributos, debido a que con esta agrupación de los valores de la variable se obtuvo una mayor medición en el  $X^2$  y el Valor de Información (IV), que otras formas de agrupación.

Tabla 8

*WOE o Pesos de Evidencia para la variable “Cantidad de Cuentas de Ahorro”*

<b>Atributo/ Cantidad de cuentas de ahorro</b>	<b>Malo</b>	<b>Bueno</b>	<b>Pm</b>	<b>Pb</b>	<b>WOE</b>
0	32	151	0.1067	0.2157	70.4246074
1	217	386	0.7233	0.5514	-27.1357844
2	34	69	0.1133	0.0986	-13.955188
3	11	52	0.0367	0.0743	70.6050585
4	6	42	0.0200	0.0600	109.861229
<b>Total</b>	<b>300</b>	<b>700</b>	<b>1</b>	<b>1</b>	

Fuente: Grupo Buro

El WOE para los atributos Cuenta de ahorro 1 y, 2 se observa que tienen valor negativo lo que significa que tiene proporciones altas de malos sobre los buenos.

Tabla 9

*WOE o Pesos de Evidencia para la variable “Monto Otorgado”*

<b>Atributo/Monto otorgado</b>	<b>Malo</b>	<b>Bueno</b>	<b>Pm</b>	<b>Pb</b>	<b>WOE</b>
1	163	436	0.5433	0.6229	13.6594182
2	65	167	0.2167	0.2386	9.63086821

3	34	65	0.1133	0.0929	-19.9271115
4	16	21	0.0533	0.0300	-57.5364145
5	13	7	0.0433	0.0100	-146.633707
6	7	3	0.0233	0.0043	-169.459572
7	2	1	0.0067	0.0014	-154.044504
<b>Total</b>	<b>300</b>	<b>700</b>	<b>1</b>	<b>1</b>	

Fuente: Grupo Buro

La variable se agrupo en 7 atributos, debido a que con esta agrupación se obtuvo una mayor medición en el Valor de Información (IV).

Tabla 10

*WOE o Pesos de Evidencia para la variable "Categoría de la SBS"*

<b>Atributo/ Categoría de la SBS</b>	<b>Malo</b>	<b>Bueno</b>	<b>Pm</b>	<b>Pb</b>	<b>WOE</b>
Cat. Perdida	25	15	0.0833	0.0214	-135.812348
Cat Dudosa	28	21	0.0933	0.0300	-113.497993
Cat. Deficiencia	169	361	0.5633	0.5157	-8.8318617
Cat. Problemas P.	28	60	0.0933	0.0857	-8.51578083
Cat. Normal	50	243	0.1667	0.3471	73.3740578
<b>Total</b>	<b>300</b>	<b>700</b>	<b>1</b>	<b>1</b>	

Fuente: Grupo Buro

El WOE para las Categoría Perdida, Dudosa, Deficiencia y Problemas Potenciales; se observa que tienen proporciones altas de malos sobre los buenos.

Tabla 11

*WOE o Pesos de Evidencia para la variable "Plazo a pagar"*

<b>Atributo/ Plazo a pagar</b>	<b>Malo</b>	<b>Bueno</b>	<b>Pm</b>	<b>Pb</b>	<b>WOE</b>
1	76	283	0.2533	0.4043	46.7415697
<=2	122	289	0.4067	0.4129	1.5107783
<=3	57	86	0.1900	0.1229	-43.6001832

> 3	45	42	0.1500	0.0600	-91.6290732
<b>Total</b>	<b>300</b>	<b>700</b>	<b>1</b>	<b>1</b>	

Fuente: Grupo Buro

La variable se agrupo en 4 intervalos o atributos, debido a que con esta agrupación de los valores de la variable se obtuvo una mayor medición en el Valor de Información (IV), que otras formas de agrupación.

Tabla 12

*WOE o Pesos de Evidencia para la variable “Plazo a pagar”*

<b>Atributo/ Bienes</b>	<b>Malo</b>	<b>Bueno</b>	<b>Pm</b>	<b>Pb</b>	<b>WOE</b>
Mueble	131	383	0.4367	0.5471	22.5539806
Auto	102	230	0.3400	0.3286	-3.41913647
No tiene	67	87	0.2233	0.1243	-58.6082361
<b>Total</b>	<b>300</b>	<b>700</b>	<b>1</b>	<b>1</b>	

Fuente: Grupo Buro

El WOE para los atributos Auto y los que No tienen Bienes se observa que tienen valor negativo lo que significa que se tienen proporciones altas de malos sobre los buenos.

Tabla 13

*WOE o Pesos de Evidencia para la variable “Número de Hijos”*

<b>Atributo/núm. hijos</b>	<b>Malo</b>	<b>Bueno</b>	<b>Pm</b>	<b>Pb</b>	<b>WOE</b>
1	104	252	0.3467	0.3600	3.7740328
2	188	425	0.6267	0.6071	-3.16506543
3	6	21	0.0200	0.0300	40.5465108
4	2	2	0.0067	0.0029	-84.729786

<b>Total</b>	<b>300</b>	<b>700</b>	<b>1</b>	<b>1</b>
--------------	------------	------------	----------	----------

Fuente: Grupo Buro

El WOE para Número de Hijos 2 y 4 se observa que tienen valor negativo lo que significa que se tienen proporciones altas de malos sobre los buenos. Los casos en que no tienen hijos se juntaron con los que tienen “1” hijo, ya que son menos del 2%, y no representa la cantidad mínima que es el 20%, para aplicar WOE.

Tabla 14

*WOE o Pesos de Evidencia para la variable “Margen de Endeudamiento”*

<b>Atributo/margen de endeudamiento</b>	<b>Malo</b>	<b>Bueno</b>	<b>Pm</b>	<b>Pb</b>	<b>WOE</b>
< 8%	96	271	0.3200	0.3871	19.0472769
8% - 16%	161	341	0.5367	0.4871	- 9.68197481
16% - 20%	43	88	0.1433	0.1257	- 13.1161162
<b>Total</b>	<b>300</b>	<b>700</b>	<b>1.00</b>	<b>1.00</b>	

Fuente: Grupo Buro

El WOE para todos los atributos tienen proporciones altas de buenos sobre malos.

Tabla 15

*WOE o Pesos de Evidencia para la variable “Tipo de empleo”*

<b>Atributo/ Tipo de empleo</b>	<b>Malo</b>	<b>Bueno</b>	<b>Pm</b>	<b>Pb</b>	<b>WOE</b>
Cas	93	141	0.3100	0.2014	-43.1137463
Contratado	104	235	0.3467	0.3357	-3.21032454
Nombrado	39	135	0.1300	0.1929	39.4415272

Directivo	64	189	0.2133	0.2700	23.5566071
<b>Total</b>	<b>300</b>	<b>700</b>	<b>1</b>	<b>1</b>	

Fuente: Grupo Buro

El WOE para los atributos CAS y Contratado se observa que tienen valor negativo lo que significa que se tienen proporciones altas de malos sobre los buenos.

Tabla 16

*WOE o Pesos de Evidencia para la variable "Tiempo de empleo"*

Atributo/tiempo de empleo	Malo	Bueno	Pm	Pb	WOE
1	11	18	0.0367	0.0257	-35.4821375
2	82	123	0.2733	0.1757	-44.1832752
3	104	235	0.3467	0.3357	-3.21032454
4	39	135	0.1300	0.1929	39.4415272
5	64	189	0.2133	0.2700	23.5566071
<b>Total</b>	<b>300</b>	<b>700</b>	<b>1</b>	<b>1</b>	

Fuente: Grupo Buro

El WOE para los atributos de tiempo 1 año, 2 años y 3 años se observa que tienen valor negativo lo que significa que se tienen proporciones altas de malos sobre los buenos.

Tabla 17

*WOE o Pesos de Evidencia para la variable "Experiencia Crediticia"*

Atributo/experiencia crediticia	Malo	Bueno	Pm	Pb	WOE
No	43	78	0.1433	0.1114	-25.1789149

Si	257	622	0.8567	0.8886	3.65661475
<b>Total</b>	<b>300</b>	<b>700</b>	<b>1.0000</b>	<b>1.0000</b>	

Fuente: Grupo Buro

El WOE para el atributo No tiene valor negativo lo que significa que se tiene para este atributo proporciones altas de malos sobre los buenos.

#### 4.4. Tratamiento de Datos.

Se genera una nueva base, donde los atributos ingresan como variables:

$$x_i = WOE_{i1}, WOE_{i2}, \dots, WOE_{ini_j}$$

Estas nuevas variables son dicotómicas con los valores 0 si el caso no presenta este atributo y el valor WOE, si es que el caso presenta este atributo o característica. Para el análisis en el software SPSS los valores de WOE se define en la etiqueta a los valores como 1.

#### 4.5. Modelo Obtenido por la Regresión en el Software SPSS.

- i. *Codificación de la variable dependiente.*

*Codificación de variable dependiente*

Valor original	Valor interno
no pago	0
si pago	1

Fuente: Grupo Buro

**ii. Significación de chi-cuadrado del modelo en la prueba ómnibus.**

Tabla 19

*Pruebas ómnibus de coeficientes de modelo*

		Chi-cuadrado	gl	Sig.
Paso 1	Paso	208,986	46	,000
	Bloque	208,986	46	,000
	Modelo	208,986	46	,000

Fuente: Grupo Buro

La significación de los coeficientes del modelo se obtiene valores menores de 0,05 lo que indica que el modelo ayuda a explicar el evento, es decir, las variables independientes explican la variable dependiente.

**iii. Bondad del modelo.**

Tabla 20

*Resumen del modelo*

Paso	Logaritmo de la verosimilitud -2	R cuadrado de Cox y Snell	R cuadrado de Nagelkerke
1	2,743 <sup>a</sup>	,189	,267

Fuente: Grupo Buro



R-cuadrado de Cox y Snell, y R-cuadrado de Nagelkerke nos indica la varianza de la variable dependiente explicada por el modelo. Se observa que la variable dependiente explicada oscila entre la R-cuadrado de Cox con un valor de 0.189 y Snell y la R-cuadrado de Nagelkerke con un valor de 0.267, valores que son mayor a cero. Por tanto, las variables independientes explican significativamente la variable dependiente.

Tabla 21

*Prueba de Hosmer y Lemeshow*

Paso	Chi-cuadrado	gl	Sig.
1	9,257	8	,321

Fuente: Grupo Buro

En la tabla anterior se observa que no hay antecedentes para rechazar  $H_0$ , por lo cual se concluye que el modelo ajustado es significativo.

iv. **Porcentaje global correctamente clasificado:**

Tabla 22

*Porcentaje Global*

Observado		Pronosticado		
		Riesgo_pago		Porcentaje correcto
		no pago	si pago	
Riesgo_pago	no pago	120	180	40,0
	si pago	60	640	91,4
Porcentaje global				76,0

Fuente: Grupo Buro

Este porcentaje indica el número de casos que el modelo es capaz de predecir correctamente. El valor pronosticado de la variable dependiente se compara con el valor observado. Los casos acertados correctamente son el 76%, más del 50% de los casos, por lo que el modelo es significativo y se acepta.

v. **Procedimiento para la selección de Variables**

Tabla 23

*Relación de variables independientes con la dependiente*

Variables	B	Error estándar	Wald	gl	Sig.	Exp(B)
Edad_menor_25(1)	-,517	,553	,874	1	,350	,596
Edad_26_35(1)	,267	,447	,358	1	,550	1,306
Edad_36_45(1)	,054	,435	,016	1	,901	1,056
Edad_46_55(1)	,211	,520	,165	1	,685	1,235
Sexo_M(1)	,035	,250	,019	1	,890	1,035
Casado(1)	-,357	,259	1,907	1	,167	,700
Separado(1)	-,486	,306	2,524	1	,112	,615

No_garante(1)	,847	,403	4,419	1	,036	2,333
Co_solicitante(1)	-1,219	,550	4,907	1	,027	,296
Viv_alquiler(1)	-,582	,435	1,796	1	,180	,559
Viv_propio(1)	,147	,407	,131	1	,717	1,159
Ing_menos_1500(1)	-,488	,349	1,959	1	,162	,614
Ing_1500_2500(1)	,130	,354	,134	1	,714	1,138
Cuentas_0(1)	,218	,516	,179	1	,673	1,243
Cuentas_1(1)	-1,394	,480	8,423	1	,004	,248
Cuentas_2(1)	-1,036	,524	3,903	1	,048	,355
Cuentas_3(1)	,345	,591	,341	1	,559	1,413
Monto_menor_2846(1)	,345	,374	,850	1	,357	1,412
Monto_2847_5444(1)	,054	,403	,018	1	,893	1,056
Monto_5445_8041(1)	-,323	,463	,487	1	,485	,724
Monto_8042_10638(1)	,587	,570	1,061	1	,303	1,798
Monto_10639_13235(1)	-1,552	,678	5,237	1	,022	,212
Monto_13236_15833(1)	-2,006	,948	4,483	1	,034	,134
Monto_15834_18430(1)	-,012	1,430	,000	1	,993	,988
Cat_Normal(1)	-1,841	,403	20,838	1	,000	,159
Cat_Problem_Potenciales(1)	1,129	,442	6,514	1	,011	3,091
Cat_Deficiencia(1)	,997	,386	6,673	1	,010	2,710
Cat_Dudosa(1)	-,219	,494	,196	1	,658	,803
Plazo_1_año(1)	-1,385	,359	14,885	1	,000	,250
Plazo_menor_2_años(1)	-,896	,325	7,615	1	,006	,408
Plazo_menor_3_años(1)	,360	,332	1,171	1	,279	1,433
Mueble(1)	-,732	,359	4,151	1	,042	,481
Auto(1)	,670	,363	3,418	1	,064	1,955
Hijo_1(1)	-1,807	1,161	2,423	1	,120	,164
Hijo_2(1)	1,603	1,157	1,920	1	,166	4,967
Hijo_3(1)	-1,439	1,274	1,276	1	,259	,237
Endeudamiento_8(1)	-,668	,265	6,344	1	,012	,513
Endeudamiento_16(1)	,224	,238	,885	1	,347	1,251

Trabajo_CAS(1)	-,039	,597	,004	1	,947	,961
Trabajo_Contratado(1)	-,356	,299	1,413	1	,235	,701
Trabajo_Nombrado(1)	,218	,248	,773	1	,379	1,243
Tiempo_empleo_1(1)	-,324	,498	,424	1	,515	,723
Tiempo_empleo_2(1)	-,375	,259	2,104	1	,147	,687
Tiempo_empleo_3(1)	-,121	,231	,275	1	,600	,886
Tiempo_empleo_4(1)	-,403	,278	2,104	1	,147	,668
Experiencia_No(1)	-,301	,235	1,638	1	,201	,740
Constante	5,379	1,961	7,526	1	,006	216,777

Fuente: Grupo Buro

- Columna B: valores de las pendientes  $\beta_i$ . El signo indica la dirección de la relación.
- Columnas Wald, gl y Sig.: Los resultados del test  $H_0: \beta_i = 0$  para cada variable. Las variables que rechazan la hipótesis nula son: No\_garante, Co\_solicitante, Cuentas\_1, Cuentas\_2, Monto\_10639\_13235, Monto\_13236\_15833, Cat\_Normal, Cat\_Problem\_Potenciales, Cat\_Deficiencia, Plazo\_1\_año, Plazo\_menor\_2\_años, Mueble, Auto, Endeudamiento\_8.
- Exp(b): Nos indica la fortaleza de la relación. Cuanto más alejada de 1 es más fuerte la relación entre la variable independiente y la variable dependiente.

El modelo mantendrá todas las variables, aunque varias de estas no son significativas a un nivel de significancia del 5% se consideró la lógica de las variables, y la estimación de los estadísticos globales del modelo.

#### vi. Modelo Final.

Siendo la mejor ecuación de regresión logística para predecir el

otorgamiento de crédito la siguiente ecuación:

$$\begin{aligned} P(\text{credito}) = & 1/(1 + e^{-(5.379 - 0.517\text{Edad}<25 + 0.267\text{Edad}_{26\_35} \\ & + 0.054\text{Edad}_{36\_45} + 0.211\text{Edad}_{46\_55} + 0.035\text{Sexo}_M \\ & - 0.357\text{Casado} - 0.486\text{Separado} + 0.847\text{No\_garante} \\ & - 1.219\text{Co\_solicitante} - 0.582\text{viv\_alquiler} + 0.147\text{viv\_propio} \\ & - 0.488\text{Ing}<1500 + 0.130\text{Ing}_{1500\_2500} + 0.218\text{Cuentas}_0 \\ & - 1.394\text{Cuentas}_1 - 1.036\text{Cuentas}_2 + 0.345\text{Cuentas}_3 \\ & + 0.345\text{Monto}<2846 + 0.054\text{Monto}_{2847\_5444} \\ & - 0.323\text{Monto}_{54445\_8041} + 0.587\text{Monto}_{8042\_10638} \\ & - 1.552\text{Monto}_{10639\_13235} - 2.006\text{Monto}_{13236\_15833} \\ & - 0.12\text{Monto}_{15834\_18430} - 1.841\text{Cat}_{Normal} + 1.129\text{Cat}_{P\_P} \\ & + 0.997\text{Cat}_{Deficiencia} - 0.219\text{Cat}_{Dudosa} - 1.385\text{Plazo}_1 \\ & - 0.896\text{Plazo\_menor}_2 + 0.360\text{Plazo\_menor}_3 - 0.732\text{Mueble} \\ & + 0.670\text{Auto} - 1.807\text{Hijo}_1 + 1.603\text{Hijo}_2 - 1.439\text{Hijo}_3 \\ & - 0.668\text{Endeudamiento}_8 + 0.224\text{Endeudamiento}_{16} \\ & - 0.039\text{Trabajo}_{CAS} - 0.356\text{Trabajo}_{Contratado} \\ & + 0.218\text{Trabajo}_{Nombrado} - 0.324\text{Tiempo\_empleo1} \\ & - 0.375\text{Tiempo\_empleo2} - 0.121\text{Tiempo\_empleo3} \\ & - 0.403\text{Tiempo\_empleo4} - 0.301\text{Experiencia}_{No} \end{aligned}$$

#### 4.6. Construcción de la Scorecard

Una vez que los WOE's han sido calculados y las betas del modelo han sido obtenidos, la variable dependiente en credit scoring, se debe tratar en puntos. Los puntajes para la construcción del scorecard deben ser resultado de un reescalamiento y una traslación del modelo logístico, dado por la ecuación:

$$\text{Score} = \text{Offset} + \text{Factor} \cdot \ln(\text{odds})$$

De aquí se obtuvo que:

$$\text{Factor} = \frac{PDO}{\ln(2)} = \frac{20}{\ln(2)} = 28.85$$

Este es un factor de escala común utilizado para los modelos de calificación de crédito contruidos con regresión logística. Una convención ampliamente utilizada en la calificación crediticia es el valor predeterminado común para PDO es 20, porque produce rangos de puntaje de crédito aceptables. Entonces, la interpretación del  $20 / \ln(2)$  es que, para un aumento de 20 puntos en el puntaje, las probabilidades se duplican.

$$\text{Offset} = \text{Score} - \text{Factor} \cdot \ln(\text{odds})$$

$$\text{Offset} = 600 - 28.85(511.32)$$

$$\text{Offset} = 420.04$$

Por lo general, en la calificación de crédito se elegiría para Score una puntuación de referencia, que es 600.

La ecuación final se deriva de la siguiente manera:

$$\text{Score} = \text{Offset} + \text{Factor} \cdot \ln(\text{odds})$$

$$= Offset + Factor \cdot \ln(b_0 + \sum_{i=1}^p \hat{b}_i \cdot woe_{ij})$$

$$= Offset + Factor \cdot \hat{b}_0 + Factor \cdot (\sum_{j=1}^p \hat{b}_j \cdot woe_{ij})$$

De esta última ecuación se desglosa:

El puntaje inicial dado por:

$$Offset + Factor \cdot \hat{b}_0 = 575.24$$

y el puntaje asociado al atributo j de la característica i dado por:

$$Factor \cdot \hat{b}_i \cdot woe_{ij}$$

Obteniendo como resultado:

Tabla 24

Scorecard

Edad	< 25 años	-27
	26 a 35 años	-0.4
	36 a 45 años	-2
	46 a 55 años	-6
Sexo	M	-0.4
	F	0
Estado Civil	Casado	-6
	Separado	13
	Soltero	0
Garante	No	0
	Co-solicitante	-73
	Si	0
Vivienda	Alquiler	-24
	Propio	-3
	Familiar	0
Ingresos S/.	Menor a 2500	-7
	1500 a 2500	-2
	Mayor a 2500	0
Cuentas de Ahorro	Cuentas 0	-15
	Cuentas 1	-38
	Cuentas 2	-14
	Cuentas 3	-24
	Cuentas 4	0
Monto Otorgado	< 2846	-5
	2847 a 5444	-1
	5445 a 8041	-6
	8042 a 10638	34
	10639 a 13235	-228
	13236 a 15833	-340
	15834 a 18430	-2

Categoría de la SBS	Normal	135
	Problema Potencial	10
	Deficiencia	9
	Dudosa	-25
Plazo a Pagar	Perdida	0
	1 año	65
	< 2 años	1
	< 3 años	16
Bienes	> 3 años	0
	Mueble	17
	Auto	2
Núm. de Hijo	No tiene	0
	1 ó 0	7
	2	5
	3	58
Margen de Endeudamiento	4	0
	8%	13
	16%	2
Trabajo	20%	0
	CAS	-2
	Contratado	-1
	Nombrado	-9
Tiempo de Empleo	Directivo	0
	1 año	-11
	2 años	-17
	3 años	0
	4 años	16
Experiencia	5 años	0
	No	-8
Constante	Si	0
		506

Fuente: Grupo Buro

4.7. Punto de corte o Cut Off.



El punto de corte fue seleccionado por la empresa, debido a que la empresa brinda servicios financieros de tercerización u outsourcing, por este motivo el punto de corte está sujeto a los términos de la contratación, este aspecto se tuvo en cuenta al momento de determinar el punto de corte. El punto de corte debe ser de tal manera que si el score es  $> "c"$  se otorga el crédito, en caso contrario si  $\text{score} \leq "c"$  se rechaza la solicitud. Donde  $"c"$  será el punto de corte, teniendo en cuenta que existe una restricción de una tasa de interés mínima de 12%.

Para hallar el punto de corte se calculó primero el score para todas las cuentas de la base, y se ordenó de menor a mayor.

Tabla 24

*Cut Off*

Caso	Score	Mora	%
454	420	0.01	21
455	421	0.00	16
456	421	0.00	18
457	422	0.00	14
458	422	0.00	21
459	422	0.01	21
460	422	0.00	21
461	422	0.00	16
462	424	0.00	20
463	424	0.00	14
464	425	0.03	15
465	425	0.00	14
466	426	0.00	20
467	426	0.00	22
468	427	0.02	19

469	427	0.00	13
470	427	0.00	21
471	427	0.01	15
472	428	0.00	16
473	428	0.00	22
474	429	0.00	20
475	429	0.00	19
476	429	0.01	14
477	430	0.00	17
478	430	0.00	21
479	430	0.00	20
480	430	0.00	21
481	431	0.00	18
482	431	0.00	13
483	431	0.00	22
484	431	0.00	19

485	431	0.00	13
486	431	0.00	21
487	431	0.00	15
488	431	0.00	16
489	431	0.00	22
490	432	0.00	20
491	432	0.00	19
492	432	0.00	14
493	432	0.00	17
494	432	0.00	21
495	433	0.00	20
496	433	0.00	21
497	433	0.00	18
498	434	0.00	12
499	767	0.00	12
500	768	0.00	14

Fuente: Grupo Buro

Luego se determina el valor de “c” satisfaciendo la ecuación:

$$\frac{\text{score} > c}{\text{score}} = \text{porcentaje}$$

Dicho punto “a” será el Cut Off según la siguiente ecuación:

$$\frac{e^{\hat{\beta}_0 + \hat{\beta}_1^T x}}{1 + e^{\hat{\beta}_0 + \hat{\beta}_1^T x}} = \hat{p} > \frac{1}{2}$$

$$\frac{1}{2} e^{\hat{\beta}_0 + \hat{\beta}_1^T x} > \frac{1}{2} \Rightarrow e^{\hat{\beta}_0 + \hat{\beta}_1^T x} > 1$$

$$\Rightarrow c = \text{Offset}$$

$$c = 420$$

De aquí se desprende que el valor mínimo que puede tomar el punto “c” es de 420.

Luego se identificó en la Tabla 18, el valor de “c” y el puntaje que se mayor a este y cuya tasa mínima debe ser del 12%. El porcentaje es un valor que también fue seleccionado por la empresa, de la siguiente manera:

$$\frac{420}{434} = 97\%$$

$$\frac{430}{434} = 99\%$$

Por lo que la empresa decidió definir como Punto de corte a 430 puntos.

## Capítulo V

### DISCUSIÓN

El modelo Credit Scoring elaborado tiene el 76% de predicción, todas las variables incluidas presentan signo coherente de influencia en la variable explicativa, las pruebas de Significación de chi-cuadrado del modelo en la prueba ómnibus cuyos valores son menores a 0,05 lo cual indica que el modelo ayuda a explicar el evento, además de R-cuadrado de Cox y Snell, y R-cuadrado de Nagelkerke indican que las variables independientes explican significativamente la variable dependiente. El elemento estadístico utilizado para valorar la bondad de ajuste del modelo es la prueba de Hosmer y Lemeshow, que determina el grado de semejanza que existe entre los valores observados y los pronosticados por el modelo, se observó que no hay antecedentes para rechazar  $H_0$ , por lo que el modelo es significativo. Estos son los indicativos de un ajuste aceptable en regresión logística por lo que el modelo es significativo y se aceptó. La presentación del Modelo Credit Scoring a través del scorecard, permite obtener la probabilidad de impago reflejado en un score que va de 0 a 1000 cuyo punto óptimo de corte para la empresa es de 430 puntos, es decir que si el score es  $> 430$  se otorga el crédito caso contrario se rechaza, y adicionalmente muestra el por qué se está dando el score o puntaje a un cliente determinado ya que el scorecard muestra los atributos o características del solicitante de crédito y de qué manera está influyendo en el score. De manera similar el ensayo realizado por el banco Central de Reserva de Argentina, concluye que el uso predominante de los modelos de Credit Scoring es por su funcionamiento e interpretación más sencilla, en contraposición a enfoques más sofisticados y de

difícil interpretación.

- Esta herramienta que nos permite obtener la probabilidad de impago de clientes futuros está aportando a la evaluación de las acciones o estrategias tomadas en la empresa para alcanzar las metas corporativas. Cabe destacar que la base utilizada para la elaboración de Modelo solo incluye información de créditos aprobados, es decir datos internos y solo podría utilizarse en empresas similares o iguales políticas y/o requisitos de aprobación de créditos. Del análisis de los resultados obtenidos se desglosa que el modelo seleccionado destaca más al predecir si el cliente si pagara (error tipo I) porque estos resultarían efectivamente como buenos pagadores al 91% de los casos. En caso contrario al predecir que un cliente no pagara se acierta en un 40%, por lo que estos casos tendrán que ser analizados por el encargado de la aprobación de créditos. Por lo que se induce una vez más que el modelo tiene una función complementaria en el proceso de concesión de un crédito. Encontrando concordancia con los resultados y conclusiones de de la tesis elaborada por Arenas, Boccardi y Piñeyrúa (2012), que en ella menciona la utilidad que los modelos de Credit Scoring pueden tener a la hora de discriminar correctamente entre “buenos” y “malos” prestatarios, así como al diseñar políticas crediticias efectivas que contribuyan a mitigar el riesgo de crédito, así mismo recalca que si la institución decide no aplicar el modelo de Credit Scoring, de todas formas, los resultados obtenidos podrían ser útiles a la hora del diseño de las políticas crediticias.
- Se determinó que el modelo incluya todas las variables, aunque varias de estas

no son significativas a un nivel de significancia del 5% ya que se revisó la lógica de todas las variables, es decir que el signo de cada una de las variables tiene un sentido lógico respecto a la variable dependiente así mismo se tuvo en cuenta la estimación de los estadísticos globales del modelo, que indican que el modelo es aceptable y significativo, por consiguiente, todas las variables presentan cierto grado de influencia del modelo. Esto nos llevó a un resultado final más real, por ello la realización de ajustes no solamente es recomendable, sino que es obligatoria. Además, se debe tener en cuenta que la aplicación es parte de la construcción del modelo credit scoring que finalmente dio resultados óptimos, así mismo las variables ingresadas en la regresión son datos previamente tratados de acuerdo a su probabilidad de casos buenos sobre malos (woe) de las variables originarias. Un análisis similar es el de Sarco (2015) cuyas conclusiones de su investigación fueron: Los factores influyentes para la determinación de préstamos financieros en la financiera Credinka son: el Ingreso económico, Gasto, número de Hijos, Edad del cliente, Seguro de salud, tipo de Vivienda y Material de construcción de la vivienda. De los cuales solo el factor edad tiene una significancia menor de 0.05, con 0.037. Una vez obtenido su modelo Sarco, procede a utilizar la Curva de ROC para establecer el punto de corte que va a estar dado por el cruce en el gráfico entre la especificidad y la sensibilidad. Para nuestra investigación el Punto de Corte está sujeto a políticas que la empresa está sujeta con el banco, ya que la empresa es prestataria de servicios crediticios. Por lo que la empresa decidió definir como Punto de corte a 430 puntos a un 99%.

-

## CONCLUSIONES

- La probabilidad de impago obtenida a través de la regresión logística se transformó en una escala de 0 a 1000 puntos o score, para crear una herramienta cuantitativa que nos indique si se debe conceder o no el crédito, para ello la entidad financiera determino que el punto de corte es 430, es decir si el score es  $> 430$  se otorga el crédito, en caso contrario si  $\text{score} \leq 430$  se rechaza la solicitud.
- El reescalamiento elaborado con el scorecard muestra 60 atributos o características que favorecen o no al solicitante de crédito en ese momento, este análisis se toma del puntaje positivo o negativo que tiene asociado cada atributo, que se muestra en la tabla 24; por tanto, el scorecard tiene una doble funcionalidad: es predictiva y por otra parte es explicativa. Esta herramienta no reemplaza el criterio experto que existe en el banco, sino la complementa para apoyar en la gestión diaria y para tomar una decisión más objetiva
- El Modelo credit scoring permite obtener la probabilidad de impago para comprobar que el modelo tiene validez con los datos reales, se hicieron pruebas de backtesting de los créditos personales siendo capaz de predecir correctamente a un 76%, para los casos en que predice que el cliente si pagara acierta en un 91.4% y en el caso de predecir que el cliente no pagara acierta en un 40% por lo que es necesario revisar el caso para no incurrir en un error.

## RECOMENDACIONES

- Para la eficiencia futura del modelo scoring, debe tenerse en cuenta factores como la actualización de los datos, el examen y ajuste periódico del modelo para incorporar los cambios del contexto económico, cambios que puedan reflejar en los atributos y el comportamiento de pago.
- Se sugiere a la empresa servicios financieros Grupo Buro, Huaraz, redefinición de las estrategias para alcanzar las metas corporativas así mismo mostrar a los clientes más sus fortalezas o debilidades ante la concesión de un préstamo creando una buena imagen a partir de un trabajo responsable.
- La regresión logística binaria es una técnica estadística muy útil para sacar e identificar perfiles así mismo identificar las causas que lo producen y en qué medida estas influyen, algo importantísimo si queremos incidir o intervenir sobre la realidad social.



## REFERENCIAS BIBLIOGRÁFICAS

- Albarran, Montes y Meza (2015). *El riesgo de no pago en una institución micro financiera del municipio de Tenancingo estado de México, 2011-2014*. Recuperado en: [http://ri.uaemex.mx/bitstream/handle/20.500.11799/40683/TESIS-IMF\\_Password\\_Removed.pdf?sequence=1](http://ri.uaemex.mx/bitstream/handle/20.500.11799/40683/TESIS-IMF_Password_Removed.pdf?sequence=1)
- Banco Central de la República Argentina: Gutiérrez Girault M. A. (2007). *Anatomía de los modelos de Credit Scoring*. Recuperado en: <http://www2.bcra.gob.ar/Pdfs/Publicaciones/CreditScoring.pdf>
- Escalona A. (2011). *Uso de los Modelos Credit Scoring en Microfinanzas*. Recuperado en: <http://docplayer.es/1515368-Colegio-de-postgraduados.html>
- Mishkin, Frederic. (2008). *Política Monetaria*, Quinta Edición. Estados Unidos: McGraw Hill.
- Santiago F. (2011). *Regresión Logística*. <http://www.estadistica.net/ECONOMETRIA/CUALITATIVAS/LOGISTICA/regresion-logistica.pdf>
- Sarco M. (2015). *Factores que determinan el otorgamiento de crédito de la financiera Credinka en la ciudad de Ayaviri*. Recuperado en: <https://es.scribd.com/document/332372102/tesis5>
- Schreiner, M. (2002). *Scoring: The Next Breakthrough in Microcredit, CGAP*. Recuperado en: <https://www.cgap.org/sites/default/files/CGAP-Occasional-Paper-Scoring-The-Next-Breakthrough-in-Microcredit-Jan-2003.pdf>

(Jaume Llopis Pérez, 2013) *Factores que determinan el otorgamiento de crédito*

<https://estadisticaorquestainstrumento.wordpress.com/2013/04/30/test-de-wald/>

(Cardenas, 2004) un Modelo de Credit scoring para instituciones de microfinanzas en el

marco de Basilea II. <http://www.scielo.org.pe/pdf/jefas/v15n28/a05v15n28.pdf>



UNIVERSIDAD NACIONAL  
SANTIAGO ANTÚNEZ DE MAYOLO

"Una Nueva Universidad para el Desarrollo"



ESCUELA PROFESIONAL DE ESTADÍSTICA E INFORMÁTICA

"Año del Diálogo y la Reconciliación Nacional"

**CONSTANCIA DE PRIMERA MATRICULA**  
**N°019-2018-UNASAM-EPEI/D.**

EL QUE SUSCRIBE, DIRECTOR DE LA ESCUELA PROFESIONAL DE ESTADÍSTICA E INFORMÁTICA DE LA FACULTAD DE CIENCIAS DE LA UNIVERSIDAD NACIONAL "SANTIAGO ANTÚNEZ DE MAYOLO".

**HACE CONSTAR:**

Que, la egresada **RODRIGUEZ SABINO ROCIO NORKA** con código N° 05.1343.9.AR de la Escuela Profesional de Estadística e Informática de la Facultad de Ciencias, de la Universidad Nacional "Santiago Antúnez de Mayolo", ha registrado su **PRIMERA MATRICULA** en el **SEMESTRE 2005 - I, CON FECHA 05 DE MAYO DE 2005.**

Se expide la presente a solicitud de la interesada, para los fines que estime conveniente.

Huaraz, "Capital de la Amistad Internacional", 08 de noviembre de 2018.



*Carlos Antonio Reyes Pareja*  
Dr. Carlos Antonio Reyes Pareja  
DECANO  
FACULTAD DE CIENCIAS  
- UNASAM -



*Angel D. Mendoza López*  
Dr. Angel D. Mendoza López  
Director (e) Escuela Profesional  
de Estadística e Informática  
FACULTAD DE CIENCIAS - UNASAM