

**Universidad Andina Simón Bolívar**

**Sede Ecuador**

**Área de Gestión**

Maestría en Gestión Financiera y Administración de Riesgos Financieros

**Elaboración de un modelo predictivo de deserción de clientes de tarjeta de crédito para una entidad bancaria del Ecuador**

Pastora Fernanda Martínez Tatamués

Tutor: Diego Oriol Vargas Lara

Quito, 2021





## Cláusula de cesión de derecho de publicación

Yo, Pastora Fernanda Martínez Tatamués, autor de la tesis intitulada “Elaboración de un modelo predictivo de deserción de clientes de tarjeta de crédito para una entidad bancaria del Ecuador”, mediante el presente documento dejo constancia de que la obra es de mi exclusiva autoría y producción, que la he elaborado para cumplir con uno de los requisitos previos para la obtención del título de Magíster en Gestión Financiera y Administración de Riesgo Financiero en la Universidad Andina Simón Bolívar, Sede Ecuador.

1. Cedo a la Universidad Andina Simón Bolívar, Sede Ecuador, los derechos exclusivos de reproducción, comunicación pública, distribución y divulgación, durante 36 meses a partir de mi graduación, pudiendo por lo tanto la Universidad, utilizar y usar esta obra por cualquier medio conocido o por conocer, siempre y cuando no se lo haga para obtener beneficio económico. Esta autorización incluye la reproducción total o parcial en los formatos virtual, electrónico, digital, óptico, como usos en red local y en internet.
2. Declaro que en caso de presentarse cualquier reclamación de parte de terceros respecto de los derechos de autor/a de la obra antes referida, yo asumiré toda responsabilidad frente a terceros y a la Universidad.
3. En esta fecha entrego a la Secretaría General, el ejemplar respectivo y sus anexos en formato impreso y digital o electrónico.

18 de octubre de 2021

Firma: \_\_\_\_\_



## Resumen

El propósito de este trabajo es elaborar un modelo matemático que permita identificar la probabilidad de que un cliente de tarjeta de crédito pueda ser un posible desertor.

Esta investigación inicia con una breve explicación del Customer Relationship Management, ya que, al hablar de deserción es preciso contar con estrategias claras de como fidelizarlos creando la mejor relación entre cliente y entidad financiera.

A continuación, se estudia diferentes tipos de modelos de Scoring que permiten predecir la variable dependiente, para este caso se considera la regresión logística.

Un factor importante que se debe revisar es la situación financiera del Ecuador basándose en indicadores económicos que son variables que influye en el consumo de tarjeta de crédito. Luego se presenta el detalle de facturación de tarjeta de crédito de los once principales bancos y su evolución tanto en cantidad de tarjetas colocadas como en el incremento del saldo. Finalmente se presenta una comparación de las tarjetas colocadas versus las canceladas por los clientes donde se verifica que a partir del segundo semestre del 2017 presenta la misma tendencia creciente dentro de la institución de estudio.

Finalmente se presenta el modelo definiendo la variable dependiente dicotómica, es decir, que es igual a cero (cliente con tarjetas de crédito con estado activa) o a uno (cliente que en el periodo de observación cancela su tarjeta de crédito o, que tuvo consumos menores a los gastos de mantenimiento), y las variables independientes que fueron creadas considerando el periodo de observación (seis meses), además, se han tomado en cuenta criterios estadísticos para identificar las variables más predictivas.

El periodo de estudio considera doce meses para cada identificación de forma escalonada llegando a tomar datos entre enero 2018 y diciembre 2019.

Se utilizo el modelo de regresión logística para determinar la probabilidad de deserción que tienen los clientes de tarjeta de crédito y tomando en cuenta los estadísticos: ROC, KS y GINI se verifica el ajuste del mismo.

Palabras clave: deserción de tarjeta de crédito, modelos de score, banca, sistema financiero.



## Agradecimientos

Muchas personas han sido mi apoyo incondicional para culminar este reto de mi vida y es por esto estoy muy agradecida con:

- Dios, por el conocimiento que me brinda día a día y por permitir que se abran las puertas necesarias para conseguir este y más objetivos.
- Mis padres y hermanos por siempre creer en mí y darme ánimo en cada paso para seguir adelante.
- Fernando, mi compañero de vida, por su amor y apoyo incondicional.
- A mis amigos y seres más cercanos que siempre me apoyaron con muestras de cariño pese a cualquier dificultad.
- Finalmente, a mis maestros que supieron guiarme impartiendo sus conocimientos en esta maestría.



## Tabla de contenidos

<b>Introducción .....</b>	<b>15</b>
<b>Capítulo primero .....</b>	<b>17</b>
Planteamiento del problema .....	17
1.1. Línea de investigación .....	17
1.2. Diagnóstico .....	17
1.3. Pronóstico .....	18
1.4. Objetivo General de la Investigación .....	18
1.4.1. Objetivos Específicos .....	18
1.5. Justificación Práctica .....	18
1.6. Delimitación del problema .....	20
<b>Capítulo segundo Marco Teórico .....</b>	<b>21</b>
2.1. Antecedentes de los modelos de Score .....	21
2.2. Modelos Econométricos y sus estadísticos .....	22
2.2.1. Análisis discriminante .....	22
2.2.2. Regresión Logística (Trejo 2016): .....	23
2.2.2.1. Estimación de máxima verosimilitud (Flórez y Rincón. 2012).....	25
2.2.3. Árboles de decisión: .....	26
2.2.4. Redes Neuronales: .....	29
2.2.5. Máquinas Vectoriales de Soporte (SVM): .....	31
2.3. Estrategias de fidelización .....	32
2.3.1. Clasificación de CRM (Customer Relationship Management).....	33
2.3.1.1. El CRM en la banca .....	33
2.4. Conceptos del proceso de modelización .....	34
2.4.1. Factor de inflación de la varianza generalizado .....	34
2.4.2. Prueba Kolmogorov-Smirnov .....	35
2.4.3. Curvas ROC y coeficiente de GINI.....	36
2.4.4. Tablas de performance .....	37
2.4.5. Weight of Evidence (WOE) .....	38
2.4.6. Prueba Chi Cuadrado .....	38

### **Capítulo tercero Estudio de la Economía y el uso de Tarjeta de Crédito en el Ecuador 39**

3.1.	Análisis de la Economía Ecuatoriana .....	39
3.2.	Vista de la Economía Ecuatoriana .....	39
3.2.1.	PIB (Producto Interno Bruto).....	39
3.2.2.	ICC (Índice de Confianza del Consumidor).....	40
3.2.3.	Desempleo.....	41
3.3.	Sistema Financiero Ecuatoriano .....	42
3.4.	Tarjetas de Crédito.....	43
3.4.1.	Tarjeta de crédito en la banca Ecuatoriana.....	44
3.4.2.	Tarjeta de crédito en el banco objeto de estudio .....	46

### **Capítulo cuarto Aplicación de la metodología en un banco ecuatoriano, para tarjetas de crédito 49**

4.1.	Aplicación de la metodología en un banco ecuatoriano, para tarjetas de crédito. 49	
4.2.	Selección de la muestra.....	49
4.3.	Definición de la variable dependiente.....	50
4.4.	Análisis de variables independientes .....	52
4.5.	Modelo Logit .....	55
4.5.1.	Muestra de Validación y Modelamiento .....	55
4.5.2.	Aplicación del modelo Logit.....	55
4.6.	Interpretación de resultados del modelo .....	57
4.7.	Validación del Modelo.....	58
4.7.1.	Curvas ROC y GINI.....	58
4.7.2.	Tabla de performance y estadísticos .....	60
4.7.3.	Multicolinealidad .....	62
4.8.	Especificaciones para aplicación del CRM.....	63

### **Conclusiones y Recomendaciones ..... 65**

### **Obras Citadas ..... 67**

## Lista de Tablas

Tabla 1	Matriz de confusión	36
Tabla 2	Distribución de la muestra	50
Tabla 3	Distribución de personas por punto de observación	50
Tabla 4	Distribución de personas por punto de observación y segmento	51
Tabla 5	Variables iniciales	53
Tabla 6	Variables más significativas	55
Tabla 7	Distribución de muestra de training y testeo	55
Tabla 8	Resultados modelo	56
Tabla 9	Variación en base de entrenamiento y testeo	60
Tabla 10	Performance del modelo -Base Train	60
Tabla 11	Performance del modelo -Base Test	61
Tabla 12	Performance del modelo -Base Total	61
Tabla 13	Factor de inflación de la varianza generalizada	62
Tabla 14	Matriz de correlación en variables del modelo	62



## Lista de Figuras

Figura 1	Modelo Logit	24
Figura 2	Ejemplo de árbol de decisión de 5 hojas	27
Figura 3	Ejemplo de partición del espacio R <sup>2</sup> , de acuerdo a la Figura 2	28
Figura 4	Ejemplo de variable creada usando árboles de decisión.	28
Figura 5	Ejemplo de hiperplano y sus distancias	32
Figura 6	Bancos que aplican CRM	34
Figura 7	Ejemplo de curvas ROC	37
Figura 8	Evolución del PIB 2010-2018	40
Figura 9	Evolución del precio del petróleo y deuda externa 2010-2018	40
Figura 10	Evolución del ICC 2010-2018	41
Figura 11	Evolución de la tasa de desempleo	42
Figura 12	Utilidades de bancos privados (millones de dólares)	42
Figura 13	Activos en bancos privados (millones de dólares)	43
Figura 14	Pasivos en bancos privados (millones de dólares)	43
Figura 15	Facturación de tarjeta de crédito (millones de dólares)	45
Figura 16	Evolución de la Facturación de tarjeta de crédito (millones de dólares)	45
Figura 17	Saldo de tarjeta de crédito (millones de dólares)	46
Figura 18	Evolución de # TC	46
Figura 19	Definición de período de observación y comportamiento	49
Figura 20	Árbol de decisión monto transacciones y continuidad laboral	57
Figura 21	Curva ROC en muestra de entrenamiento	59
Figura 22	Curva ROC en muestra de testeo	59
Figura 23	Curva ROC en muestra de entrenamiento y testeo	60
Figura 24	Definición de puntos de corte de probabilidad	63



## Introducción

En la actualidad el modelamiento del comportamiento de los clientes en el sistema financiero, es una de las herramientas más utilizadas para identificar patrones de comportamiento que permiten tomar estrategias que mitigan los posibles riesgos inmersos en este mercado.

Las metodologías utilizadas para el desarrollo de los modelos de scoring son amplias, ya que mediante la investigación constante, estas se han incrementado y se adaptan a las necesidades de cada estudio.

Por otro lado, el riesgo crediticio es uno de los riesgos principales en las entidades financieras, y es por esto que a más de contar con modelos que determinen si un cliente es o no un buen pagador, se analiza la deserción del mismo, y en este caso en particular, se considera la deserción del uso de tarjeta de crédito (la cual genera cartera), debido a que hoy por hoy es considerado un medio de pago muy utilizado.

Dentro de la entidad financiera sujeto de este estudio, se identificó que existe la deserción de clientes de tarjeta de crédito para lo cual se generó esta necesidad de identificar posibles desertores y de esta manera aplicar diferentes estrategias para fidelizarlos.

Este estudio plantea los siguientes capítulos:

Capítulo primero: Donde se plantea la problemática a la que la entidad financiera se enfrenta y se especifica los objetivos de esta investigación.

Capítulo segundo: Especifica los lineamientos teóricos que se utilizan para el desarrollo del modelo de deserción.

Capítulo tercero: Presenta un enfoque de la economía ecuatoriana y el uso de la tarjeta de crédito como un medio de pago que genera cartera de crédito en las entidades financieras.

Capítulo cuarto: Desarrollo del modelo de score para la deserción de clientes de tarjeta de crédito.



## **Capítulo primero**

### **Planteamiento del problema**

#### **1.1.Línea de investigación**

La línea de investigación con la que se identificó este tema es la estructura de los mercados e instituciones financieras en el Ecuador: Las instituciones bancarias ecuatorianas, manejan un alto porcentaje de los activos de los ecuatorianos ya que el peso del sector bancario en función de PIB nominal en el Ecuador fue del 3.6% en el año 2017 mostrando un constante crecimiento, por lo tanto, constituye una parte importante en la dinámica de la estructura financiera. (Gorospe 2018, 2). Por otro lado, se encarga del estudio de algunos métodos y técnicas que promuevan el mejor funcionamiento de las entidades bancarias.

#### **1.2.Diagnóstico**

Considerando los sectores económicos en la cartera de Bancos Privados de acuerdo a la Asociación de Bancos del Ecuador (ASOBANCA), se mostró la información con corte a diciembre 2019, donde el consumo en tarjeta de crédito presentó el 24.76% de participación, seguida por comercio al por mayor y menor con el 15.72%, consumo no productivo el 14.29%, el 9.40% en industria manufacturera y otros sectores acumulan el 35.83% restante (Cóndor 2019, 4). Por esta razón el estudio del comportamiento en el consumo de tarjeta de crédito es muy importante en la banca privada.

El mercado financiero de tarjeta de crédito presenta mucha competencia en la actualidad, en el que cada entidad, luego de realizar el análisis del comportamiento de sus clientes actuales, buscan conseguir nuevos clientes, inclusive aquellos de la competencia, para lo que realizan estrategias de mercado competitivas, anticipándose a los clientes que están próximos a dejar de utilizar este producto en otras entidades financieras.

La entidad bancaria que es el objeto de estudio (entidad X) tiene aproximadamente el 6% del total de tarjetas de crédito y saldo de 7.7% del sistema ecuatoriano a diciembre 2019; ha presentado algunos problemas en el uso de tarjetas de crédito, mismos que se han identificado cuando se verificó que los clientes dejan de utilizarla o han cambiado de proveedor de tarjeta a cambio de otros beneficios que no se los puede visualizar con facilidad, llegando a representar una pérdida de al menos un 5% de sus tarjetas de crédito (cifras de la entidad en estudio). Por otro lado, al perder un cliente se refleja la pérdida

del costo que se ha incurrido inicialmente al reclutar a estos clientes y el gasto de mantenerlos hasta el momento de la deserción (abandono).

### **1.3.Pronóstico**

La entidad X al no contar con una metodología que permita identificar los clientes que podrían desertar, ha provocado la pérdida de los mismos, sin poder aplicar estrategias que permitan que su fidelización se incremente y así cubrir sus necesidades en el servicio ofertado. Además, se enfrenta a pérdidas económicas ya que al perder un cliente presenta costos tanto de reclutamiento como de mantenimiento hasta el momento de desertar.

### **1.4.Objetivo General de la Investigación**

Elaborar un modelo de SCORING que permita la detección oportuna de deserción de clientes de tarjetas de crédito para una entidad financiera ecuatoriana.

#### **1.4.1. Objetivos Específicos**

- Especificar estrategias de fidelización de clientes aplicado a tarjetas de crédito.
- Establecer una metodología para selección de variables necesarias para el modelo de predicción e identificar el punto de default para determinar un cliente desertor y un cliente fiel.
- Identificar los modelos de SCORIG existentes.
- Elaborar un modelo que permita predecir los clientes que puedan desertar del uso de la tarjeta de crédito.

### **1.5.Justificación Práctica**

Algunos autores han realizado el estudio de deserción de los clientes, para lo que se ha citado alguno de dichos estudios:

Ning, Dong (2009) menciona que es de alto riesgo el negocio de tarjetas de crédito, pese a que su rentabilidad es muy representativa y por esto el controlar la deserción de los clientes de tarjeta de crédito (cuando el cliente cancela su tarjeta de crédito) se ha tornado en un punto importante en las entidades financieras. Tomando en cuenta este nivel de importancia, han visto la necesidad de predecir la deserción de los clientes con lo que optimizan procesos de marketing con la finalidad de disminuir las

pérdidas de clientes. Por otro lado, mencionan que: *“Es más rentable para el banco mantener y satisfacer a los clientes existentes que atraer constantemente nuevos clientes, y se ha demostrado que una pequeña disminución de la tasa de rotación puede dar lugar a cambios significativos en la contribución”*.

De acuerdo a Herrera (2016) las entidades financieras presentan problemas para retener a sus mejores clientes, pese a estar siempre atentos en satisfacer las necesidades de los mismos. Además, la competencia obliga a las entidades financieras a investigar las causas que hacen que un cliente esté satisfecho o no con el servicio recibido, inclusive, según Kaura (2013) indica que estas entidades dependen en su totalidad de dicha satisfacción. Reiccheld y Sasser (1990) concluye que al reducir el 5% de la deserción de una entidad puede generar entre el 25% al 85% de ganancias. Cobeña (2016) menciona que las entidades financieras que trabajan con tarjetas de crédito elaboran estrategias para generar nuevos clientes y mantener los que utilizan este producto, debido a que los réditos generados por el mismo son importantes y es necesario que se rentabilice esta relación.

Abbasimehr, Setak y Tarokh (2011) han descrito que la saturación de los mercados y la competencia constante a nivel empresarial genera que los clientes deserten de un lugar para instalarse en otro, llegando a ser una preocupación a nivel general. El poder detallar la deserción de clientes, ha pasado a ser una herramienta importante para predecir si un cliente es de tipo churn (abandono) o no churn (clientes que continúan tranzando con la empresa), y con esto las empresas utilizan de forma óptima los recursos de marketing mediante campañas para retención (Coussement, Benoit y Van Den Poel, 2010. 2132-2143).

Por otro lado, Torkzadeh, Chang y Hansen (2006) concluyen que el ganar un nuevo cliente le cuesta hasta 12 veces más que mantener a un cliente existente en la empresa, es por esto que el generar estrategias de fidelización de clientes identificados mediante algún modelo predictivo puede generar un alto beneficio para las empresas. (Poel, Larivière, 2004. 196-217).

Además, Hee-Su, Yoon (2004) indican que es necesario construir un modelo de deserción de clientes. Las metodologías para determinar la probabilidad de deserción de los clientes, que exponen los autores son: Teoría de conjuntos aproximados (RTS) y las máquinas de vectores de soporte de mínimos cuadrados (LS-SVM), con la finalidad de obtener una guía para realizar el marketing de clientes de tarjeta de crédito de las instituciones financieras.

Existen más metodologías que ayudan a identificar la probabilidad de deserción entre ellas se encuentran: árboles de decisión (Seol, Choi y Park, 2007. 432-440) donde han verificado que mientras más se amplía el árbol puede ocasionar que algunas de las hojas del árbol de decisión presenten probabilidades similares entre ellas, algoritmos genéricos (Sarimveis y Bafas, 2003. 59-80), mismos que no logran identificar con mayor certeza la probabilidad asociada, clasificador bayesiano (Shulian, 2005. 109-111), el cual requiere hallar máxima verosimilitud por medio de gran búsqueda, redes neuronales (Bentz y Merunka, 200. 177-200), mismas que no muestran un patrón descubierto de forma simbólica que sea entendible para todos, sobre todo para entes de control, regresión logística (Lemmens y Croux, 2006. 276-286), entre otros modelos. Pese a la diversidad de modelos para determinar la probabilidad de deserción de los clientes es necesario identificar el que sea más preciso y comprensivo para los usuarios.

Una herramienta importante que se utiliza para identificar información importante de una empresa es la minería de datos (Iyakutti y Umayalvathi, 2012). Por otro lado, Ma, Qin y Wang (2009) muestran que las técnicas de minería de datos se dividen en la verificación orientado a que el sistema verifique la hipótesis y el descubrimiento orientado, donde el sistema encuentra reglas y patrones por sí mismo.

Finalmente, como punto adicional, los autores también recomiendan que las empresas deben manejar su sistema de gestión de la relación con sus clientes (CRM, Customer Relationship Management) con ayuda de minería de datos y modelos de deserción de clientes.

## **1.6.Delimitación del problema**

Para este estudio se va a utilizar un modelo de score enfocado en los clientes de tarjeta de crédito de una institución financiera del Ecuador mediante el uso de regresión logística. Este modelo se encaminará en la identificación de la probabilidad de deserción de clientes de tarjeta de crédito.

La información para elaborar este modelo fue entregada por la entidad financiera X, misma que tiene presencia a nivel nacional con un total de 117 sucursales y agencias bancarias en el año 2019 y con activos de alrededor de 5.000 millones de dólares.

## Capítulo segundo

### Marco Teórico

#### 2.1. Antecedentes de los modelos de Score

En esta sección se revisará los modelos de score que se pueden aplicar para identificar la probabilidad de deserción de clientes de tarjeta de crédito, con la finalidad de notar el funcionamiento de cada uno, finalmente se revisa una metodología de marketing para retener a los clientes que estén próximos a desertar.

Las instituciones financieras se encargan de realizar la intermediación financiera donde las mismas contactan o son contactadas por ahorradores o inversionistas y éste último genera rendimientos y por otro lado, el intermediario atrae mayores capitales. En esta actividad se encuentra inmerso un riesgo latente producido por la falta de pago lo que se traduce en pérdidas para la entidad financiera o en el caso de no cubrir expectativas de los clientes se tiene el riesgo de perderlos generando pérdidas al requerir nuevos clientes.

Para poder contrastar este tipo de riesgos es necesario la gestión de los mismos, llegando a ser considerada como una variable importante a la cual, las entidades financieras deben tomarla en cuenta con la finalidad principal de disminuir las pérdidas, además de dar cumplimiento a la normativa vigente (Lizarzaburu, Berggrun y Quispe, 2012. 96-103).

Una de las técnicas que más se utiliza para gestionar los riesgos financieros, específicamente de crédito, son los modelos de score, mediante los que se puede calcular la probabilidad de incumplimiento del cliente. Desde la década de los 30 se inicia el desarrollo de modelos de gestión de riesgo de crédito por medio del uso de información contable financiera (Smith y Winakor, 1935). En la década de los 60 algunos autores complementaron estos estudios por medio de ratios financieros como medio de prever la quiebra (Tamary, 1966, 15-21), análisis discriminante y predicción de quiebras corporativas (Altman, 1968. 589-609). En los 70's se implementan nuevos modelos con mayor énfasis en análisis discriminante, como por ejemplo el análisis discriminante múltiple escalonado con restricción en la correlación (Edmister, 1972. 1477-1493). En la década de los 80's y 90's los modelos se fortalecen en el componente estocástico por medio de modelos de máximo verosimilitud y redes neuronales (Romani, 2002).

Anderson, Sweeney y Williams (2008), concluyen que el riesgo de impago corresponde a un problema multifactorial para lo que proponen el uso de modelos multicriterio con la finalidad de incluir tanto variables cualitativas como cuantitativas y entre los principales modelos que subrayan son de análisis jerárquicos y los scoring. Hand y Henley (1997) enfatizan que los credit scoring se los utiliza cuando no se tiene información para procesos de decisión lógica de crédito que puedan ser parte de la evaluación y clasificación de los clientes, por otro lado, García y Caballero (2013) aportan diciendo que los modelos de credit scoring estiman la probabilidad de impago de los clientes que tienen características similares y por su parte, Puertas y Marti (2013) exponen que estos modelos se traducen en un problema de clasificación entre los clientes que solicitan un crédito y con la ayuda de características de cada persona, dentro de estas se encuentran variables de comportamiento de pago en el sistema financiero o no financiero, características sociodemográficas y financieras que permitan identificar el grupo al que pertenezca, esto es, los que tengan alta probabilidad de impago frente a las obligaciones financieras y aquellos a los cuales su probabilidad de no cumplir dichas obligaciones sea muy baja.

A continuación, se exponen las especificaciones de ciertos modelos de credit scoring:

## **2.2. Modelos Econométricos y sus estadísticos**

### **2.2.1. Análisis discriminante**

De acuerdo a Huysmans (2006), corresponde a una técnica de clasificación o discriminación de datos que siguen un cierto patrón.

El objetivo principal del análisis discriminante es clasificar a los individuos en n grupos que permitan hallar la mejor predicción de su comportamiento objetivo mediante una combinación lineal de las variables independientes, es decir que sean grupos homogéneos y con una varianza inter-grupo máxima comparada con la varianza dentro de cada grupo. (Lee, Chiu y Chen, 2002)

Algunos autores coinciden en que la ecuación del análisis discriminante es una función lineal:

$$Z = \alpha + \omega_1 x_1 + \omega_2 x_2 + \dots + \omega_n x_n \quad (2.1)$$

Donde:

$x_i$ : vector de entrada X

$\omega_n$ : coeficientes o pesos

$\alpha$ : punto de intersección (constante)

$Z$ : variable a discriminar

Desai, Crook y Overstreet (1996), concluyen que el vector  $W$  de pesos debe maximizar la siguiente expresión:

$$(W' \mu_1 - W' \mu_2) / W' \Sigma W \quad (2.2)$$

Donde:

$\mu_1$  y  $\mu_2$ : media

$\Sigma$ : matriz de covarianza

Esta maximización tiene como objetivo, si esta expresión se maximiza, verificar si el riesgo de mal clasificar sea bajo.

### 2.2.2. Regresión Logística (Trejo 2016):

Los modelos logísticos se aplican para predecir resultados dicotómicos, como por ejemplo si un cliente es bueno o malo. Estos modelos consisten en un tipo de análisis de regresión, donde la variable independiente corresponde a una variable dummy.

De acuerdo a Trejo (2016), son buenos discriminantes y dan una probabilidad de impago del cliente. Estos modelos son los mejores estimadores debido a que se concentra en alguna de las colas de la distribución condicional de la variable dependiente misma que toma valores 0 o 1.

$$p = P(y = 1/x) = 1 - P(y = 0/x) \quad (2.3)$$

Donde:

$x$ : variable independiente.

$y$ : variable dependiente, sigue una distribución de Bernoulli.

$$y_i \sim B(\pi_i) \quad (2.4)$$

Donde:  $\pi_i$  :es la probabilidad de ver si un cliente tiene un buen hábito de pagar o no.

La relación entre  $\pi_i$  y la variable independiente está dada por:

$$\log \left( \frac{\pi_i}{1-\pi_i} \right) = \alpha + \beta x \quad (2.5)$$

Donde:

$\log \left( \frac{\pi_i}{1-\pi_i} \right)$  es el logaritmo de las razones de probabilidad para  $y_i$ .

$\beta$ : coeficiente de  $x$

Finalmente, la función logística queda expresada de la siguiente forma:

$$\pi_i = \frac{e^{\alpha+\beta x}}{1+e^{\alpha+\beta x}} \quad (2.6)$$

Con  $0 < \pi_i < 1$

Donde:

$x$  corresponde al vector de variables explicatorias.

$\beta$ : vector de parámetros que también se puede expresar como:

$$\pi_i = \frac{1}{1+e^{-(\alpha+\beta x)}} \quad (2.7)$$

Con la transformación logit se llega a:

$$\frac{\pi_i}{1-\pi_i} = \frac{1+e^{\alpha+\beta x}}{1+e^{-(\alpha+\beta x)}} = e^{\alpha+\beta x} \quad (2.8)$$

Si se toma el logaritmo natural se obtiene:

$$\text{Ln}\left(\frac{\pi_i}{1-\pi_i}\right) = (\alpha + \beta x) \quad (2.9)$$

Como resultado se obtiene el logaritmo de la razón de probabilidad, misma que es lineal, considerando las variables y los parámetros. Para poder estimarlos se utiliza la máxima verosimilitud (Green, 2001).

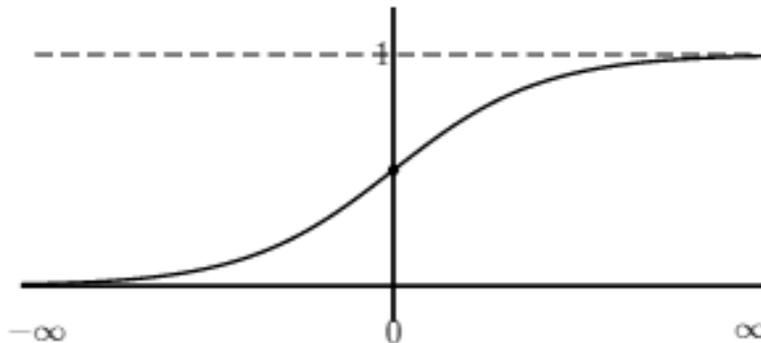


Figura 1: Modelo Logit.  
Fuente y elaboración propia

Flórez y Rincón (2012) señalan las siguientes características de los modelos logit:

- Las probabilidades no son lineales pese a que el modelo transformado sea lineal en sus variables.
- Una de las suposiciones del modelo logit es que el logaritmo de la razón de probabilidades se relaciona linealmente con la variable explicativa, y es por este camino que se puede tener una interpretación de los parámetros.
- Los coeficientes de la regresión logística muestran el cambio en el logaritmo de las probabilidades, considerando que una de sus variables explicativas

cambia en una unidad, tomando en cuenta que el resto de variables se mantengan constantes.

### 2.2.2.1. Estimación de máxima verosimilitud (Flórez y Rincón. 2012)

El modelo logístico se lo puede expresar de forma general como:

$$y_i = E(y_i) + \varepsilon_i \quad (2.10)$$

Donde:

$y_i$ : variable aleatoria independiente que sigue una distribución Bernoulli:

$$E(y_i) = \pi_i = \frac{e^{x'\beta}}{1+e^{x'\beta}} \quad (2.11)$$

Tomando en cuenta que cada observación sigue una distribución de Bernoulli, su distribución se puede expresar por:

$$f_i(y_i) = \pi_i^{y_i} (1 - \pi_i)^{(1-y_i)}, i = 1, 2, 3, \dots, n \quad (2.12)$$

Tomando en cuenta que las observaciones son independientes, su función de verosimilitud es:

$$L(y_1, y_2, \dots, y_n, \beta) = \prod_{i=1}^n f_i(y_i) = \prod_{i=1}^n \pi_i^{y_i} (1 - \pi_i)^{1-y_i} \quad (2.13)$$

Al considerar el logaritmo natural se obtiene:

$$\ln L(y_1, y_2, \dots, y_n, \beta) = \ln \prod_{i=1}^n f_i(y_i) = \sum_{i=1}^n \left[ y_i \ln \left( \frac{\pi_i}{1-\pi_i} \right) \right] + \sum_{i=1}^n \ln(1 - \pi_i) \quad (2.14)$$

Pero:

$$1 - \pi_i = \frac{1}{1+e^{x'\beta}} \quad \text{y,} \quad \ln \left( \frac{\pi_i}{1-\pi_i} \right) = x'\beta$$

Entonces el logaritmo de la verosimilitud se puede expresar para el modelo de regresión logístico:

$$\ln L(y, \beta) = \sum_{i=1}^n y_i x_i' \beta - \sum_{i=1}^n \ln[1 + e^{x_i' \beta}] \quad (2.15)$$

Suponiendo que  $y_i$  sea la cantidad de un observado para la  $i$ -ésima observación y  $n_i$  la cantidad de ensayos en cada observación, por lo tanto, el logaritmo de la verosimilitud es:

$$\ln L(y, \beta) = \sum_{i=1}^n y_i \pi_i + \sum_{i=1}^n n_i \ln(1 - \pi_i) - \sum_{i=1}^n y_i \ln(1 - \pi_i) \quad (2.16)$$

El utilizar un algoritmo de mínimos cuadrados iterativamente re ponderados puede generar estimadores de máxima verosimilitud.

Considerando que  $\hat{\beta}$  sea el estimador objetivo, por medio del método iterativo, se puede demostrar que:

$$E(\hat{\beta}) = \beta \quad \text{y} \quad V(\hat{\beta}) = (X'V^{-1}X)^{-1} \quad (2.17)$$

Entonces el valor estimado del predictor lineal es  $\hat{\eta}_i = x_i \hat{\beta}$ , y el valor esperado de la regresión logística es:

$$\hat{y}_i = \hat{\pi}_i = \frac{1}{1+e^{(-x_i \hat{\beta})}} \quad (2.18)$$

### 2.2.3. Árboles de decisión:

Los árboles de decisión son generalmente utilizados para minería de datos y diversos análisis estadísticos.

Linoff y Berry (2011), concluyen que los árboles de decisión son una herramienta para explorar los datos y encontrar relaciones entre los mismos; otro de los usos es para modelamiento. Por otro lado, describen que un árbol permite dividir conjuntos de datos que presentan similitudes entre sí, en grupos más pequeños mediante la aplicación de reglas de decisión en cada nodo, siendo un nodo una condición sobre un atributo y las ramas son el resultado de haber evaluado la condición y finalmente las hojas corresponden a las variables dependientes.

Por otro lado, Han, Pei y Kamber (2011), mencionan que una de las técnicas utilizadas para la creación de los árboles de decisión es la Inducción de Árbol de Decisión, que empieza buscando una variable que pueda dividir los datos hasta tener los nodos finales o variables dependientes.

Por su parte, Linoff y Berry (2011) identifican la pureza y diversidad como una técnica para clasificación de la información, mismo que considera las siguientes: Índice de Gini, Índice de ganancia de información, Entropía y Prueba Chi-cuadrado. Ante esto se requiere considerar el tipo de variable que se tiene, es decir, si es nominal, numérico, discreto, continuo, entre otros.

Entre los principales algoritmos para elaborar los árboles de decisión se encuentran: ID3, C4.5, C5 y CART (Classification And Regresión Tree).

Breiman, Friedman, Olshen y Stone (1984) estudiaron los árboles de clasificación y regresión (CART), para lo cual partieron de una muestra de entrenamiento:

$$(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n) \quad iid \sim \gamma \quad (2.19)$$

Donde:  $X_i = (X_i^1, \dots, X_i^p)$  vector con  $p$  variables aleatorias que pueden ser continuas, discretas o una combinación de éstas.

$Y_i$ : variable unidimensional discreta o continua.

Con la muestra de la ecuación número 2.19 se construyó una etapa denominada árbol maximal y una segunda etapa que corresponde a una poda.

Los autores crean la primera etapa iniciando con la totalidad de los datos en el nodo raíz y partiendo de este se llega a los nodos interiores mediante repetidas particiones donde se involucra una pregunta basada en  $p$  atributos. Hasta este punto se generan árboles binarios, por lo que de cada nodo parten dos nodos hijos; adicional se selecciona un criterio de parada para que cada nodo no tenga más particiones y llegue a constituir los nodos finales que se denominan hojas.

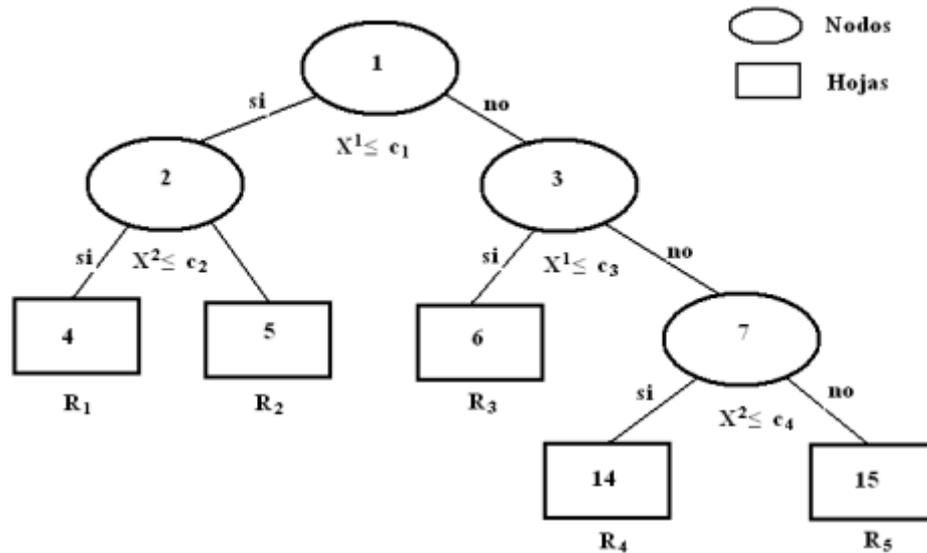


Figura 2: Ejemplo de árbol de decisión de 5 hojas

Fuente y elaboración: Breiman, Friedman, Olshen, y Stone (1984).

Luego de que se construya el árbol maximal, el predictor determina para cada hoja un valor definido por:

$$E[Y|(X^1, X^2)] = \sum_{j=1}^5 f_j (X^1, X^2) 1_{\{(X^1, X^2) \in R_j\}} \quad (2.20)$$

Donde, si  $Y$  es continua, corresponde a un árbol de regresión y se estima  $f_j$  por el promedio de los  $Y_i$  en la región  $R_j$ :

$$\hat{f}_j = \frac{\sum_{\{i:(X^1, X^2) \in R_j\}} Y_i}{\text{card}\{i:(X^1, X^2) \in R_j\}} \quad (2.21)$$

Pero, si  $Y$  es discreta, corresponde a un árbol de clasificación, donde:

$$\hat{f}_j = \text{la clase más frecuente en } R_j \quad (2.22)$$

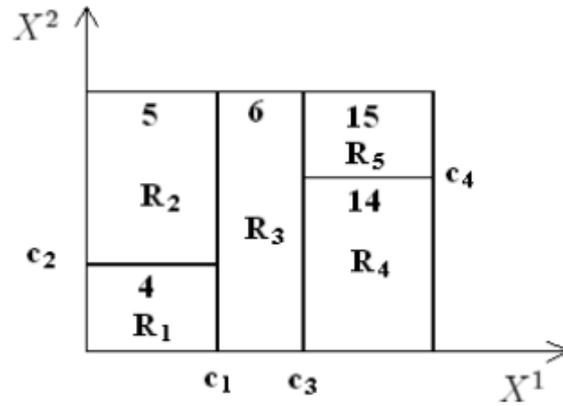


Figura 3: Ejemplo de partición del espacio R2, de acuerdo a la Figura 2.  
Fuente y elaboración: Breiman, Friedman, Olshen, y Stone (1984).

Para el presente estudio se utilizan los árboles de decisión para crear variables basadas en criterios de decisión que permiten generar segmentos que sean homogéneos, estas variables se crearon mediante la asignación de probabilidades de acuerdo a cada nodo terminal, para mayor claridad, se detalla el proceso de construcción de dichas variables:

En la figura 4 se muestra cómo se construye una variable mediante el uso de árboles de decisión para la variable: cantidad de compras máxima en los últimos seis meses, donde podemos observar que el Nodo 0 se denomina nodo padre y el resto de nodos son los nodos hijos; Identificando que en el Nodo 1 acumula el 39% de casos buenos, en el Nodo 2 el 80.3% de buenos, en el Nodo 3 el 92.7% de buenos, en el Nodo 4 el 95.2% de buenos, en el Nodo 5 el 97% y finalmente en el Nodo 6 el 98% de buenos:

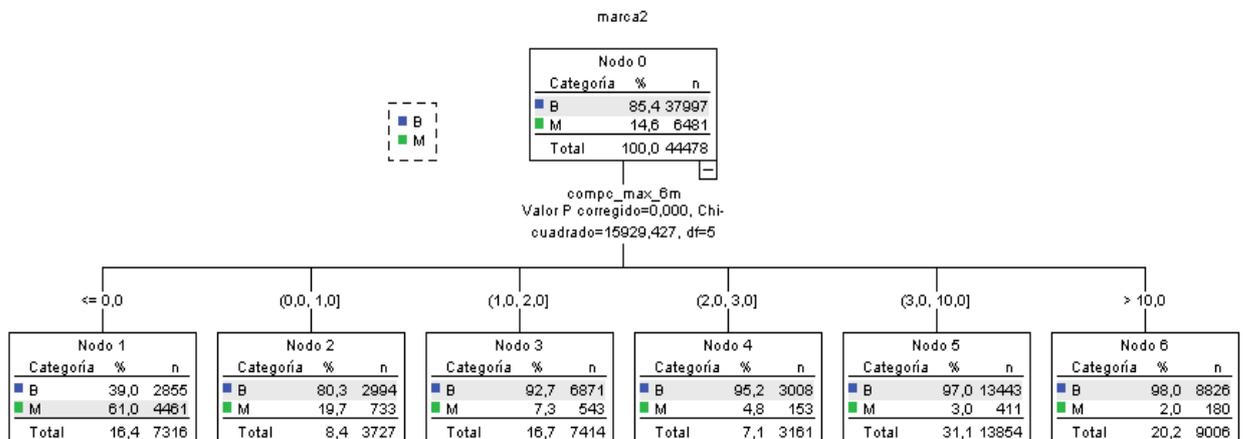


Figura 4: Ejemplo de variable creada usando árboles de decisión.

Fuente y elaboración propias

Esta variable se resume en la siguiente expresión:

$$prob\_compc\_max6m = \begin{cases} 0.39, & \text{si } compc\_max\_6m \leq 0 \\ 0.803, & \text{si } 0 < compc\_max\_6m \leq 1 \\ 0.927, & \text{si } 1 < compc\_max\_6m \leq 2 \\ 0.952, & \text{si } 2 < compc\_max\_6m \leq 3 \\ 0.97, & \text{si } 3 < compc\_max\_6m \leq 10 \\ 0.98, & \text{si } compc\_max\_6m > 10 \end{cases} \quad (2.23)$$

#### 2.2.4. Redes Neuronales:

Tan, P. N. (2006), en su investigación comentó que las Redes Neuronales constituyen estructuras computacionales que son un resultado de una aproximación al comportamiento de las neuronas del sistema biológico. En esta metodología cada nodo es conocido como una neurona, las que se conectan entre sí. La información inicial corresponde a valores numéricos normalizados para luego de transitar por las neuronas o nodos resulta en valores numéricos finales. Las conexiones de los nodos determinan el comportamiento del modelo, es decir que cada nodo determina la división de los datos y su respectiva asignación al grupo que corresponda.

Por su parte, Lee y Chen (2005) y Hsieh (2005), concuerdan en que dentro de cada red neuronal se encuentran muchas capas y gran cantidad de neuronas ocultas, pero para poder modelar se puede realizar con una sola capa oculta y obtener una gran precisión en los resultados.

Por otro lado, Kim y Sohn (2004), indican que se debe cumplir las siguientes etapas para poner en funcionamiento una red neuronal:

1. Elaboración de la arquitectura de la red neuronal, iniciando con la identificación de las variables dependiente e independientes y la metodología topológica de la red. Como se mencionó con anterioridad, es importante que las variables de entrada se encuentren normalizadas para no tener descompensación de las mismas.
2. Entrenamiento y validación del modelo por medio del ajuste de los pesos de las neuronas hasta que el error final sea mínimo.
3. Clasificación y/o predicción. Cuando ya se tiene el modelo final, se le provee nuevos registros al modelo para verificar su nivel de predicción, mismo que debe ser similar que los resultados obtenidos en el paso anterior.

Pero, Linoff, y Berry, (2011), identificaron 7 etapas para crear el modelo y que funcione con nueva información, pero las mismas se resumen en las tres etapas citadas por Kin y Sohn (2004).

Finalmente, Chen y Huang (2003), definieron claramente los componentes de una red neuronal en un modelo de scoring:

- Cantidad de neuronas proporcionales a la cantidad de atributos del solicitante de un crédito (variables numéricas)
- Una neurona que clasifique a los clientes como buenos o malos.
- Neuronas y capas ocultas.

El funcionamiento de cada capa consiste en que toma un conjunto de pesos multiplicados por la entrada, en este punto inicialmente se tiene pesos aleatorios, es decir:

$$I = \sum \omega_i x_i + \theta \quad (2.24)$$

Donde:

$I$ : Entrada de la neurona

$\theta$ : peso adicional

$\omega_i$ : coeficientes de la variable dependiente  $x$

La entrada es procesada mediante una función de activación que dan como resultado valores entre 0 y 1 o entre -1 y 1, con lo que se puede verificar que la información entrante fue previamente normalizada.

De acuerdo a Han, Kamber, Tan, Steinbach (2011) identificaron las principales funciones de activación: “función sigmoideal, función logística y la función tangente hiperbólica”, mismas que son no lineales.

Un paso final consiste en un proceso de ajuste iterativo considerando diversos criterios apropiados, como es el minimizar el error cuadrático medio:

$$SEE(W) = \frac{1}{2} \sum (y_i - \hat{y}_i)^2 \quad (2.25)$$

Mismo que depende de los pesos  $W$ , adicionalmente, es necesario utilizar la tasa de aprendizaje, siendo un paso muy importante la selección de la misma, porque si su valor es muy pequeño provoca que el proceso de aprendizaje de la neurona se vuelva lento, con un nivel de convergencia demorado, y dado el caso contrario en el que se inicia con una tasa de aprendizaje muy alta provocando que no se dé la convergencia, de acuerdo a Lee y Chen (2005).

Como complemento, West (2000), menciona las siguientes arquitecturas de redes neuronales que se aplican a los modelos de scoring:

- “Multi – Layer Perceptron (MLP)
- Mixture of Experts, MOE.
- Radial Basis Function, RBF.
- Learning Vector Quantization, LVQ.
- Fuzzy Adaptive Resonance, FAR.”

### 2.2.5. Máquinas Vectoriales de Soporte (SVM):

De acuerdo a Tan. (2006), las máquinas vectoriales de soporte necesitan encontrar hiperplanos que puedan separar de forma óptima un conjunto de datos en una cantidad específica de grupos. Se debe considerar que muchos datos no se los puede separar de forma lineal, por lo que se requiere cambios de dimensionalidad, esto significa que se cambia de dimensión de acuerdo a la cantidad de atributos que se tenga y se identifican los hiperplanos que separen de mejor forma los datos en grupos de manera correcta.

Una vez identificados los hiperplanos que pudieron separar a los datos en grupos, se requiere identificar un hiperplano óptimo, considerando la menor distancia entre el hiperplano y los datos. El hiperplano que separa los datos esta dado por:

$$WX + b = 0 \quad (2.26)$$

Donde:

$W$ : pesos

$X$ : vector de valores de entrada

$b$ : intersección del eje.

La distancia que se utiliza para encontrar el hiperplano óptimo está dada por:

$$d = \frac{2}{\|w\|} \quad (2.27)$$

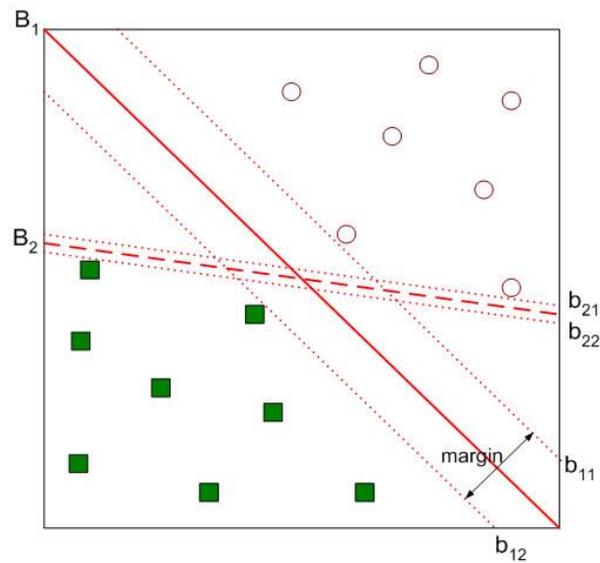


Figura 5: Ejemplo de hiperplano y sus distancias  
 Fuente: Ramírez, A. Y. (2007). Elaboración propia

El problema para optimizar el hiperplano se reduce a:

$$\min \frac{\|w\|}{2} \quad (2.28)$$

Sujeto a restricciones:  $y_i(WX_i + b) \geq 1, i = 1, 2, \dots, N$

Donde:

$y_i$ : el valor que toma la clase para el registro  $i$ .

$N$ : número de registros a examinar.

### 2.3. Estrategias de fidelización

La obtención de agrupaciones de los clientes permite que se pueda generar estrategias de fidelización (Bambino Contreras, C. 2005).

El marketing relacional de acuerdo a Berry (1983, 61) lo define como la “atracción, mantenimiento e intensificación de las relaciones con los clientes”, por otro lado Parvatiyar y Sheth (2001, 5) considera que el CRM involucra los procesos y estrategias de retención, adquisición y asociación del negocio con ciertos clientes para crear un beneficio para la empresa como también para el cliente o usuario, todo esto mediante la interacción entre las áreas internas de la empresa para encontrar las necesidades tanto de la empresa como del cliente, en cambio la Asociación Española de Marketing Relacional (AEMR, 2002, 9) considera al CRM como las estrategias de negocio, marketing, infraestructura tecnológica y comunicación, desarrolladas con la

finalidad de tener una mejor conexión con los clientes, y una relación de fidelidad con la compañía mediante la identificación, comprensión de sus necesidades para poder satisfacerlas de la mejor manera y llegar al objetivo propuesto. Se enfoca básicamente, en identificar de forma amplia las preferencias y necesidades del cliente, que no busque una relación temporal, sino más bien de largo plazo convirtiéndose en el cliente fiel que llene sus necesidades y no busque nuevas opciones

De acuerdo a Berry (1983) define al CRM (Customer Relationship Management) como la “atracción, mantenimiento e intensificación de las relaciones con los clientes”.

### **2.3.1. Clasificación de CRM (Customer Relationship Management)**

Según Bhattacharya (1998) y Payne (2012) consideran la siguiente clasificación:

**CRM Operacional:** Prioriza la automatización de procedimientos, con el fin de generar información para que la empresa le facilite el entendimiento al cliente. De otro lado, automatiza los procesos front office que comprende ventas, servicio al cliente y marketing, mediante el uso de tecnología permitiéndole llegar al objetivo de la empresa.

**CRM Analítico:** Se enfoca en los datos y como se van analizar los mismos; es decir, integración de los datos obtenidos de los clientes mediante procesos automatizados tecnológicamente, lo que nos permitirá conocer al cliente y sus preferencias en consumos.

**CRM Colaborativo:** Relación entre el cliente y la empresa. Este tipo de CRM se enfoca en los canales de comunicación que permita una información óptima y de esta manera poder ofertar los productos requeridos por el cliente.

**CRM Estratégico:** Enfoque en la estrategia del negocio, mismo que puede ser focalizado en la sistematización de la fuerza de ventas.

**e-CRM:** Uso de canales y medios electrónicos con el fin de obtener una buena relación con el cliente.

**PRM (Partner Relationship Management):** corresponde a la creación de información del cliente para beneficio de la empresa.

#### **2.3.1.1.El CRM en la banca**

A nivel mundial el CRM se ha desarrollado en diversos bancos, Han – Yuh (2007) menciona algunos:



Figura 6: Bancos que aplican CRM  
Fuente: Han – Yuh 2007. Elaboración propia

En diversos países la implementación del sistema de CRM genera disminución de la deserción de los clientes en la banca y mejora la relación entre los bancos y sus clientes mediante el uso de diferentes herramientas implementadas con el uso de la tecnología o estrategias que implementaron los bancos.

## 2.4. Conceptos del proceso de modelización

Para el desarrollo de este proyecto es necesario definir algunos conceptos de pruebas estadísticas que se van a utilizar:

### 2.4.1. Factor de inflación de la varianza generalizado

Neter, Wasserman y Kutner (1990) concluyeron que la existencia de multicolinealidad infla la varianza y para esto sugieren la verificación del valor de inflación de la varianza, mismo que se define como:

$$VIF_j = \frac{1}{1+R_j^2} \quad (2.29)$$

Donde:

$R_j^2$ : coeficiente de determinación de la regresión del j-ésimo regresor sobre el resto de variables explicativas.

Con  $Y 1 \leq VIF \leq \infty$ ,

Fox (2003) define al factor de inflación de la varianza generalizado como:

$$GVIF = \frac{\det R_i - \det R_{-i}}{\det R} \quad (2.30)$$

Donde:

$R_i$ : Determinante de la matriz de correlación sobre las columnas de la variable  $i$ .

$R_{-i}$ : Determinante de la matriz de correlación sobre las columnas de variables distintas de  $i$ .

$R$ : Determinante de toda la matriz de correlación.

Además, los autores consideran que existe problemas de multicolinealidad cuando este valor es superior a 10 (Yoo et al. 2014, 3).

#### 2.4.2. Prueba Kolmogorov-Smirnov

Prueba K-S: Consiste en una prueba no paramétrica que permite probar la independencia de muestras que siguen la misma distribución, considerando la siguiente prueba de hipótesis: (Moreno y Meléndez 2011)

$H_0$ : La distribución de score es la misma para buenos y malos

$H_1$ : La distribución de score no es la misma para buenos y malos

Rezác y Řezáč (2011) comenta que una vez definida la variable objetivo como:

$$D_k = \begin{cases} 1, & \text{cliente es bueno} \\ 0, & \text{caso contrario} \end{cases} \quad (2.31)$$

Se puede definir las funciones de distribución acumulada de los clientes buenos y de clientes malos, mismas que dependen de su score:

$$F_{n,Buenos}(a) = \frac{1}{n} \sum_{i=1}^n I(s_i \leq a \wedge D_k = 1), \quad (2.32)$$

$$F_{m,Malos}(a) = \frac{1}{m} \sum_{i=1}^m I(s_i \leq a \wedge D_k = 0), \quad a \in [L, H], \quad (2.33)$$

Donde:

$s_i$ : score del cliente  $i$ -ésimo

$n$ : número de clientes buenos

$m$ : número de clientes malos

$I$ : Función de indicador donde  $I(\text{verdad})=1$ .

$L$ : mínimo valor de score

$H$ : máximo valor de score

Para identificar la proporción de clientes malos se puede calcular mediante:

$$p_{Malos} = \frac{m}{m+n}, \quad y \quad p_{Buenos} = \frac{n}{m+n} \quad (2.34)$$

Además, Rezac y Řezáč (2011) muestra la definición del estadístico KS como:

$$KS = \max_{a \in [L,H]} |F_{m,Malos}(a) - F_{n,Buenos}(a)| \quad (2.35)$$

### 2.4.3. Curvas ROC y coeficiente de GINI

*Curvas ROC:*

Franco Nicolás (2007) define a la curva ROC como la técnica para comparar el poder predictivo en modelos de clasificación. Para tener una mejor comprensión de este análisis es necesario entender conceptos básicos:

Partiendo de una matriz de confusión que corresponde a una herramienta que resume el desempeño de un modelo, comparando los valores reales versus los proyectados mediante el modelo que se desea evaluar:

Tabla 1  
Matriz de confusión

		Resultado Proyectado	
		Positivo	Negativo
Valores Reales	Positivo	Verdaderos Positivos (VP)	Falsos Negativos (FN)
	Negativo	Falsos Positivos (FP)	Verdaderos Negativos (VN)

Fuente: Franco Nicolás (2007)  
Elaboración propia

Se define la sensibilidad como la razón entre los individuos positivos que fueron clasificados correctamente versus el total de positivos, es decir mide al grupo que son verdaderamente positivos.

$$S = \frac{VP}{FP+VP} \quad (2.36)$$

La especificidad mide los verdaderos negativos, y corresponde a la razón entre los verdaderos negativos y el total de negativos.

$$E = \frac{VN}{VN+FN} \quad (2.37)$$

Las curvas ROC detalla la sensibilidad y el complemento de especificidad, es decir que mide la capacidad e incapacidad de clasificar correctamente el modelo que se evalúa.

El área bajo la curva ROC, permite evaluar el nivel de discriminación del modelo y mientras este valor sea más cercano a 1 existe una mayor capacidad de discriminar. Tal como se observa en la Figura 7 se nota que la curva 1 tiene menor poder predictivo que la curva 2, ya que es más cercana a los ejes superior e izquierdo.

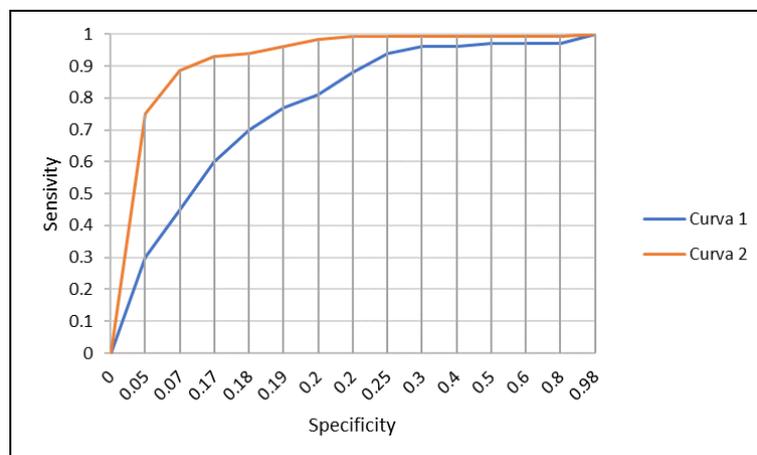


Figura 7: Ejemplo de curvas ROC  
Elaboración propia

#### *Coefficiente GINI:*

Permite medir la desigualdad entre dos poblaciones y para su definición se requiere entender la curva de Lorenz que es similar a la curva ROC.

La curva de Lorenz representa la distribución de los casos malos y casos totales por deciles en todos sus rangos de puntuación, es decir que determina el poder de clasificación de buenos y malos en deciles que se seleccionen y se calcula como:

$$Gini = 2AUC - 1 \quad (2.38)$$

#### **2.4.4. Tablas de performance**

Estas tablas también son conocidas como tablas de rendimiento o de desempeño, mismas que permiten visualizar el nivel de discriminación de los modelos, por medio del uso de percentiles de la probabilidad objeto del estudio. Por lo general se segmenta por deciles identificando la cantidad de casos que se agrupan en cada uno y permite analizar el nivel de discriminación del modelo que se está evaluando. (Pérez, 2014. 65).

### 2.4.5. Weight of Evidence (WOE)

Este indicador es una medida del nivel de predicción que tiene la variable independiente en función de la variable dependiente. En este trabajo se lo utiliza para transformar una variable cualitativa en cuantitativas (Bhalla. 2016).

$$WOE = \left[ \ln \left( \frac{Db_{ij}}{Dm_{ij}} \right) \right] * 100 \quad (2.39)$$

Donde:

$$Db_{ij} = \frac{\text{Número de buenos en el atributo j de la característica i}}{\text{Número de buenos para la característica i}}$$

$$Dm_{ij} = \frac{\text{Número de malos en el atributo j de la característica i}}{\text{Número de malos para la característica i}}$$

### 2.4.6. Prueba Chi Cuadrado

Como herramienta para identificar la diferencia del comportamiento de las distribuciones de frecuencia esperadas y observadas, se utiliza la prueba  $\chi^2$  (Chi-Cuadrado), misma que de acuerdo a Mendivelso y Rodríguez (2018) corresponde una prueba no paramétrica que toma valores mayores a 1 y se la considera para tres pruebas diferentes: bondad de ajuste, independencia y homogeneidad.

Prueba de hipótesis:

$H_0$ : No hay asociación entre la distribución de buenos y malos.

$H_1$ : Si hay asociación entre la distribución de buenos y malos.

Su representación estadística es:

$$\chi^2 = \sum_{i=1}^k \left[ \frac{(Obs_i - Esp_i)^2}{Esp_i} \right] \quad (2.40)$$

Donde:

$Obs_i$ : Valor observado

$Esp_i$ : Valor esperado

## **Capítulo tercero**

### **Estudio de la Economía y el uso de Tarjeta de Crédito en el Ecuador**

#### **3.1. Análisis de la Economía Ecuatoriana**

En este capítulo se trata la importancia de la Economía Ecuatoriana, misma que se la considera un punto importante de partida para entender el comportamiento del sistema bancario y en específico del producto “Tarjetas de Crédito”.

Se inicia mediante un estudio de indicadores financieros ecuatorianos de los años 2009 y 2018, luego se adentra en un estudio del Sistema Financiero y posterior se analiza el comportamiento de un banco grande en el Ecuador y específicamente en la evolución de tarjetas de crédito.

#### **3.2. Vista de la Economía Ecuatoriana**

Se ha elegido algunos indicadores que permitan visualizar la evolución de la economía ecuatoriana y su coyuntura, entre ellos se inicia con el PIB, mismo que muestra la evolución de la producción de bienes y servicios del Ecuador, además, se analiza el comportamiento del precio del petróleo y la deuda externa. El siguiente indicador importante que nos permite entender la economía desde los hogares es el índice de confianza del consumidor (ICC), mismo que muestra el comportamiento que presentan los hogares ecuatorianos en función de su seguridad en la economía del país. Por otro lado, el nivel de desempleo permite tener una visión más clara de la estabilidad económica del país, así como del nivel de consumo que puede generar en dicha economía.

##### **3.2.1. PIB (Producto Interno Bruto)**

De acuerdo a la figura 8, se puede visualizar que entre los años 2010 y 2018 el PIB crece en USD 38.843 millones. El PIB muestra mejores valores entre el año 2011 y 2014, esto causado por la depreciación del dólar y el precio del petróleo, mismo que se ubicó sobre los USD 100 en este período, siendo estos factores determinantes para que el Ecuador mantenga niveles superiores a 3.5% de variación interanual.

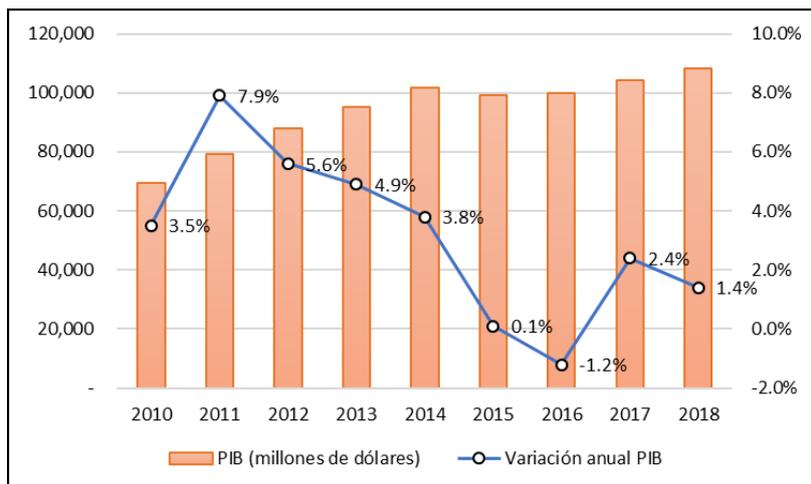


Figura 8: Evolución del PIB 2010-2018

Fuente: Banco Central del Ecuador. Elaboración propia

A partir del año 2015 se observa que la tasa de variación del PIB disminuye debido al comportamiento del precio del petróleo, llegando a su pico mínimo del -1.2% en el año 2016, reflejando así una recesión económica. Desde el año 2013 al 2017 el gobierno del Ecuador (presidente Rafael Correa) plantea nuevas estrategias que permitan el cambio de la matriz productiva, enfocado en la diversificación de la producción y dejando de depender de bienes primarios como lo son la agricultura y la minería junto con la recuperación del precio del petróleo permite que la economía empiece a mejorar.

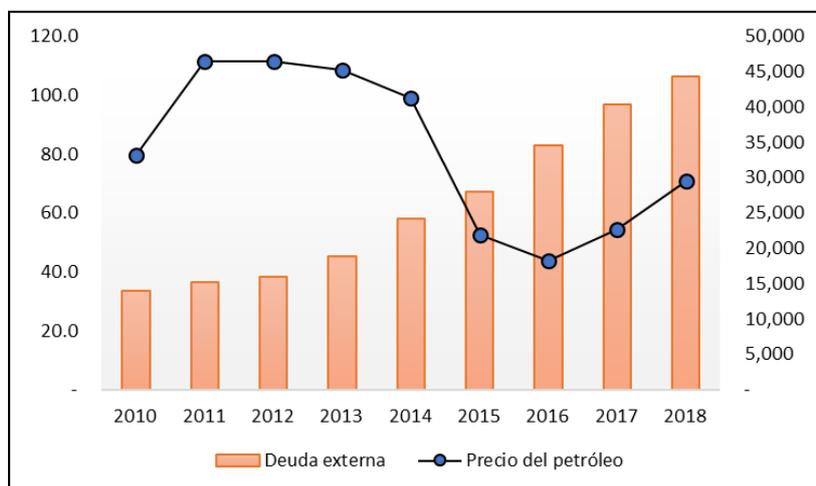


Figura 9: Evolución del precio del petróleo y deuda externa 2010-2018

Fuente: Banco Central del Ecuador. Elaboración propia

### 3.2.2. ICC (Índice de Confianza del Consumidor)

Este indicador muestra la confianza que tienen los hogares sobre la permanencia de sus ingresos, mismos que son fuente para el consumo, que es parte principal para la

dinámica económica. Para su construcción se utiliza los resultados de la encuesta ENEMDU que el INEC la realiza de forma mensual.

De acuerdo a la figura 10 se puede observar cómo la confianza de los hogares ecuatorianos se incrementaba desde el año 2010 hasta llegar a su pico más alto en el año 2013, esto además se refleja en el consumo de los hogares permitiendo que haya mayor expansión económica, pero durante los años 2014 y 2015 se nota que existe desconfianza de parte de los hogares, por lo que prefieren ahorrar en lugar de gastar, generando así una contracción en la economía por miedo a que se llegue a una recesión. (Morales, 2008).

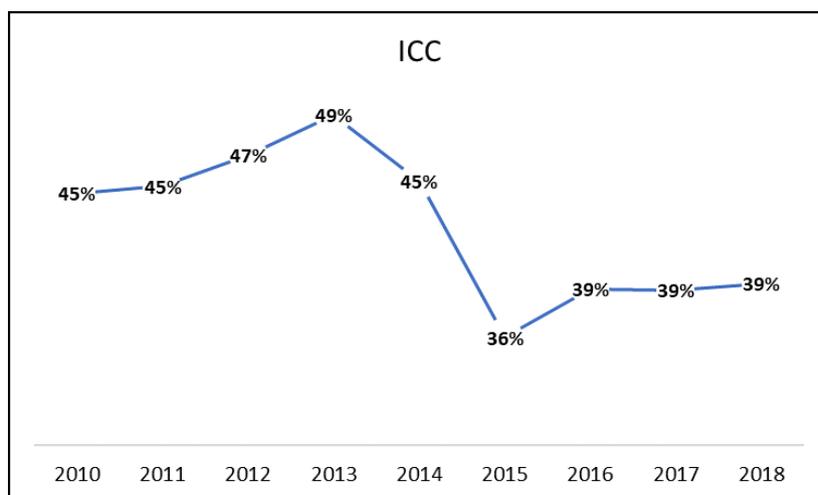


Figura 10: Evolución del ICC 2010-2018

Fuente: Banco Central del Ecuador. Elaboración propia

### 3.2.3. Desempleo

El empleo es una fuente importante de la evolución de la economía de cualquier país, siendo así la fuente de ingresos para cada hogar ecuatoriano, pero debido a la coyuntura de los años 2010 y 2016 (año de contracción económica), las fuentes generadoras de trabajo se han visto obligadas a reducir su personal, incrementando así el nivel de desempleo hasta llegar a 5.0% y 5.2% respectivamente, por otro lado, en los años 2014 y 2018 presentaron las tasas más bajas de nivel de desempleo, siendo estas de 3.8% y 3.7% correspondientemente.

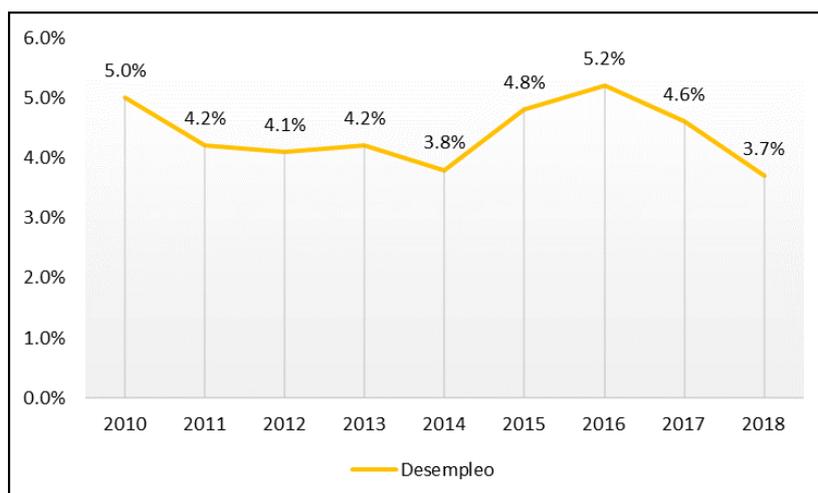


Figura 11: Evolución de la tasa de desempleo  
Fuente: INEC. Elaboración propia

### 3.3.Sistema Financiero Ecuatoriano

Los bancos privados presentaron disminución de sus resultados en el año 2016, pero a partir del año 2017 comienza a crecer llegando a incrementar sus resultados en USD 394 millones en el año 2019, presentando así resultados favorecedores para este sector.

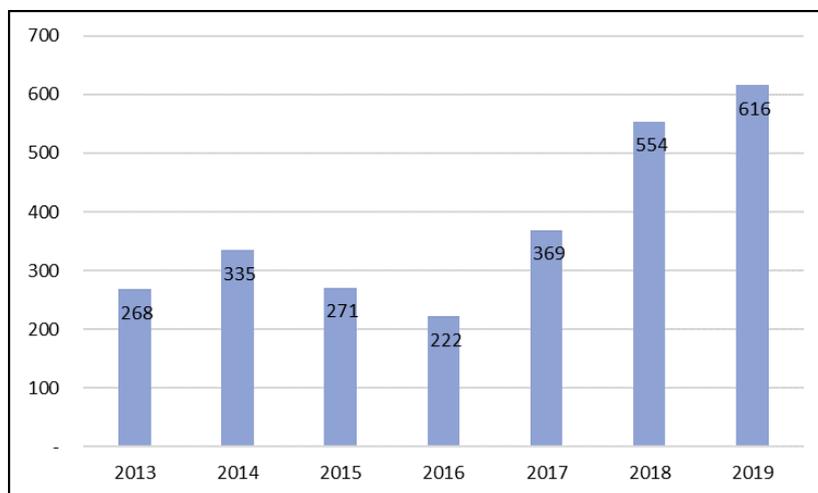


Figura 12: Utilidades de bancos privados (millones de dólares)  
Fuente: Banco Central del Ecuador. Elaboración propia

Al comparar el nivel de activos en el 2019 para los bancos privados en el Ecuador, se refleja un crecimiento del 25% comparado con el año 2016, lo que refleja una recuperación en el sector financiero, esto es debido a que los activos se encuentran conformados en su mayoría por cartera e inversiones.

Las colocaciones en la banca privada para el año 2019 incrementaron en un 48.2% respecto al año 2016, siendo el valor del incremento de USD 9.170 millones en este período.

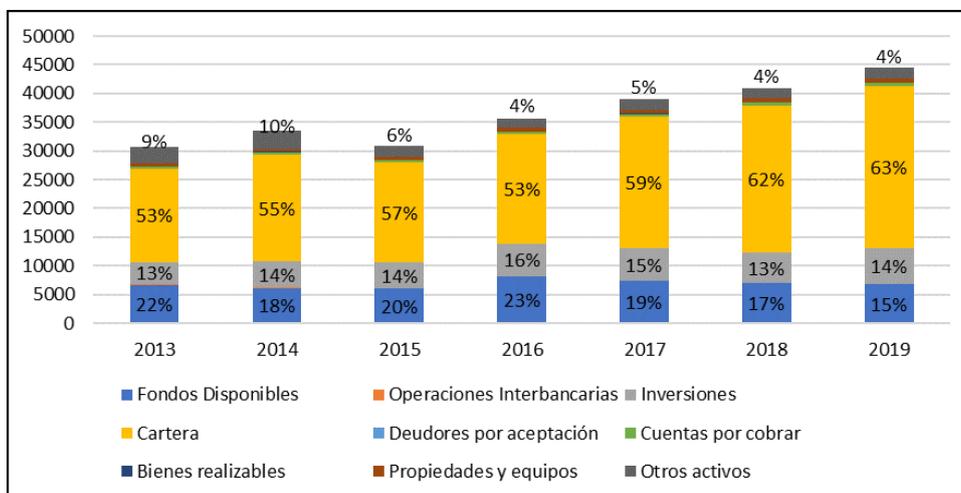


Figura 13: Activos en bancos privados (millones de dólares)

Fuente: Banco Central del Ecuador. Elaboración propia

De forma paralela, los pasivos mostraron un incremento significativo, partiendo del año 2016, llegaron a alcanzar un valor de USD 7.460 millones en el año 2019, conformados en un alto porcentaje por obligaciones con el público, mismo que presentó una tasa de crecimiento del 17.4%.

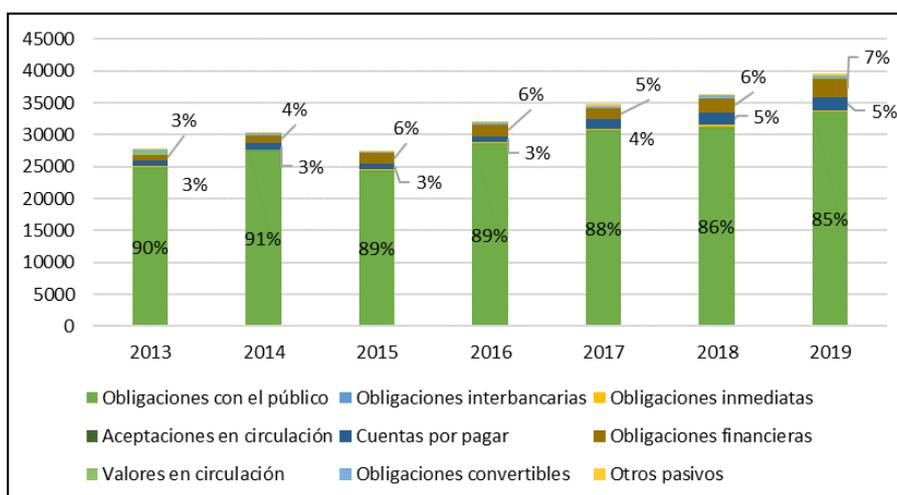


Figura 14: Pasivos en bancos privados (millones de dólares)

Fuente: Banco Central del Ecuador. Elaboración propia

### 3.4. Tarjetas de Crédito

Gutiérrez (1992) recopila la historia de la tarjeta de crédito, partiendo desde sus orígenes, los cuales menciona que fue en 1920 donde Western Union entrega a los

mejores clientes tarjetas de metal con la finalidad de acceder de forma ágil al servicio brindado y con opción de diferir sus consumos. Por esta década, General Petroleum Corporation emite la primera tarjeta de crédito exclusiva para sus clientes, quienes la utilizaba en el consumo de gasolina, pero luego se extendió a sus clientes. Posteriormente los bancos empezaron a utilizar esta forma de pago.

En América Latina data sus inicios en Brasil en 1956 con Dinners Club, posterior lanza su producto City Bank en 1970 la tarjeta con el nombre de City Card, en esta época se realiza una alianza entre City Bank, Itaú y Unibanco y lanzaron al mercado la marca de tarjeta Credicard (Gutiérrez, 1992).

Para el Ecuador, fue Dinners Club quien introdujo la tarjeta de crédito, como acto seguido, el Banco Continental llega al mercado con la tarjeta Visa Internacional, mientras que Filanbanco obtiene los derechos para la tarjeta American Express. Durante el año 1983 Unicredit administra la tarjeta MasterCard y en 1984 la misma empresa incluye en sus productos la tarjeta MasterCard Gold. (Rodríguez y Silva, 2011)

La tarjeta de crédito es un instrumento de pago muy utilizado a nivel mundial, considerando además como un acceso a dinero plástico y con esto la sociedad se ha convertido en consumidora, aunque no tiene la preparación adecuada para hacerlo de forma adecuada. (Amar, 2001).

### **3.4.1. Tarjeta de crédito en la banca Ecuatoriana**

Actualmente, según los datos reportados por la Superintendencia de Bancos, el Ecuador presenta 15 bancos que prestan servicios de tarjeta de crédito, generando a diciembre del 2019 una facturación total de USD 1.309,5 millones.

El banco que presenta mayor participación en el mercado, en cuanto a la facturación es Dinners Club con un total de USD 453 millones, representando el 35% del total de la facturación. Por otro lado, el banco que tiene mayor participación en el mercado en función de la cantidad de tarjetas de crédito emitidas es el Pacífico con 937.130 tarjetas que corresponde al 27% del mercado.

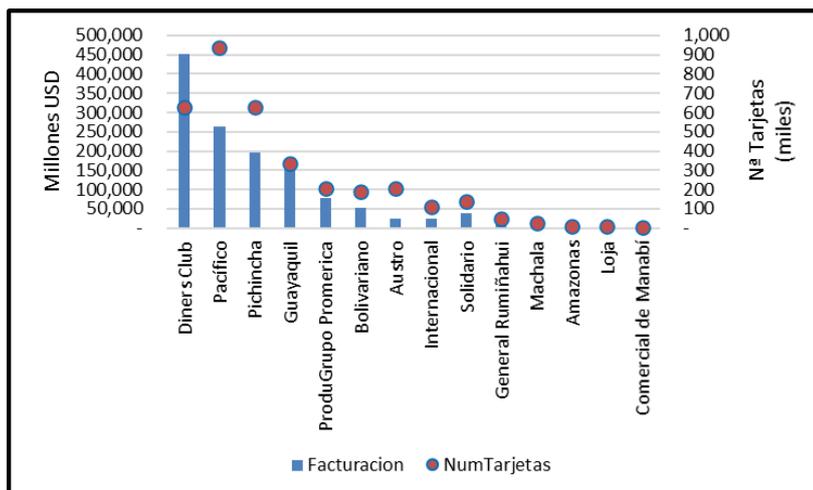


Figura 15: Facturación de tarjeta de crédito (millones de dólares)  
Fuente: Superintendencia de Bancos. Elaboración propia

La facturación por consumo en tarjeta de crédito se incrementó en los últimos 3 años; partiendo desde junio del 2017 donde se facturó un total de USD 857 millones creciendo hasta USD 1.309 millones en diciembre del 2019.

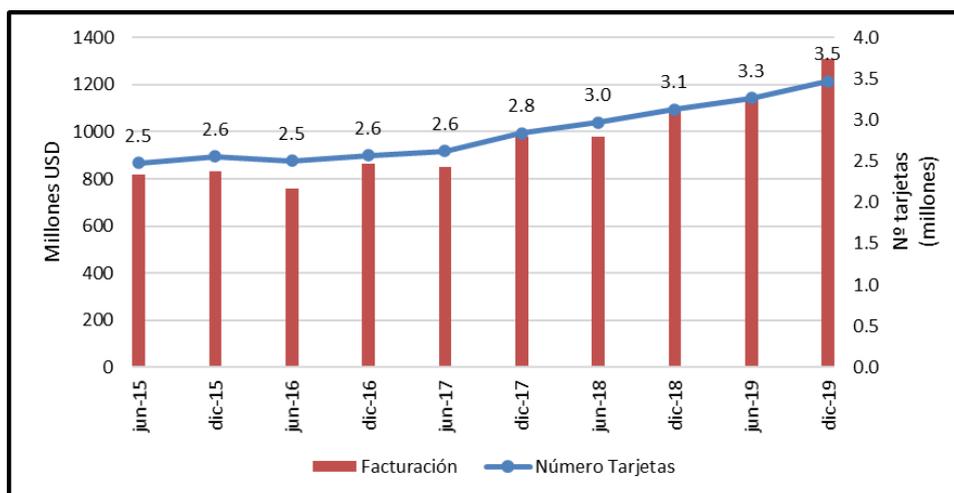


Figura 16: Evolución de la Facturación de tarjeta de crédito (millones de dólares)  
Fuente: Superintendencia de Bancos. Elaboración propia

El crecimiento de la facturación por consumo en tarjeta de crédito se ve reflejada en los últimos 3 años; partiendo desde el año 2017 donde se facturó un total de 491 millones se refleja un crecimiento del 6.4% hasta el año 2019.

El saldo total en tarjetas de crédito se encuentra en constante crecimiento, llegando a diciembre 2019 a un total de USD 7.575 millones, concentrado en 3.5 millones de tarjetas de crédito.

Hasta junio del año 2017, el consumo con tarjeta de crédito se concentra en el saldo rotativo, pero desde diciembre 2017 hasta diciembre 2019 esta tendencia cambió, ya que el consumo se agrupa mayoritariamente en el saldo diferido, es decir que las personas prefieren pagar en cuotas.

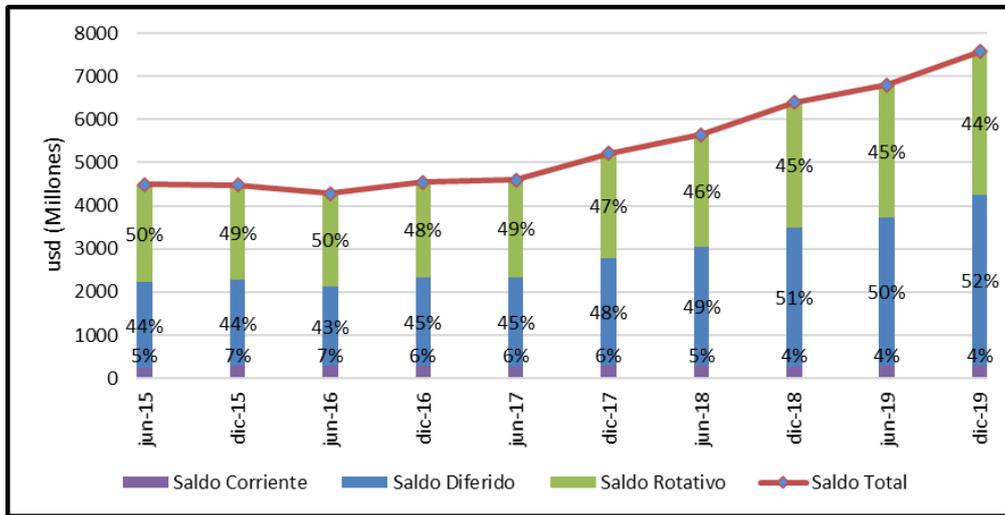


Figura 17: Saldo de tarjeta de crédito (millones de dólares)  
 Fuente: Superintendencia de Bancos. Elaboración propia

### 3.4.2. Tarjeta de crédito en el banco objeto de estudio

Dentro de la entidad objeto de estudio, se observa el crecimiento en la cantidad de tarjetas de crédito activas, esto debido a estrategias utilizadas para ganar clientes. Además, la evolución de tarjetas canceladas por el cliente presenta también una tendencia creciente, siendo similar a la de tarjetas activas a partir del segundo semestre el año 2017.

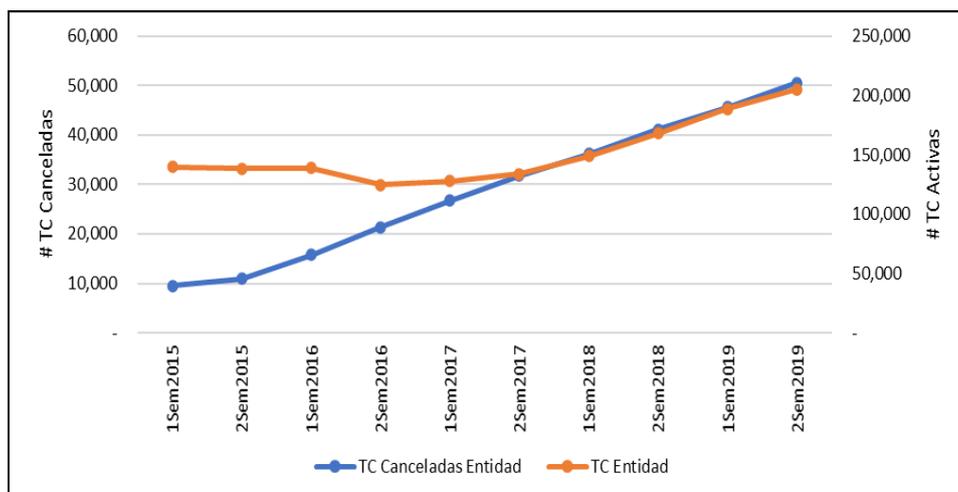


Figura 18: Evolución de # TC  
 Fuente: Superintendencia de Bancos. Elaboración propia

Lo último, llegó a ser un punto de preocupación para la institución y por lo que se requirió la elaboración de un modelo para predecir los clientes que vayan a desertar (abandonar) del uso de tarjeta de crédito y poder enfocarse con distintas estrategias de marketing para que la tarjeta de crédito de esta institución llegue a tener mayor usabilidad.



## Capítulo cuarto

### Aplicación de la metodología en un banco ecuatoriano, para tarjetas de crédito

#### 4.1. Aplicación de la metodología en un banco ecuatoriano, para tarjetas de crédito.

Este capítulo describe la creación del modelo logístico que permita identificar la deserción de clientes de tarjeta de crédito, con información financiera y de sus clientes, misma que fue entregada por un banco con alta presencia en el Ecuador.

Este modelo permite obtener la probabilidad de que un cliente deje de utilizar la tarjeta de crédito, con lo cual el banco puede tomar estrategias de marketing para poder fidelizarlos mediante la selección de puntos de corte la probabilidad para establecer regiones de aplicación del CRM diferenciado en cada una.

#### 4.2. Selección de la muestra

Inicialmente se ha definido un período de observación y de comportamiento, tomando en cuenta seis meses para cada uno, de la siguiente manera:

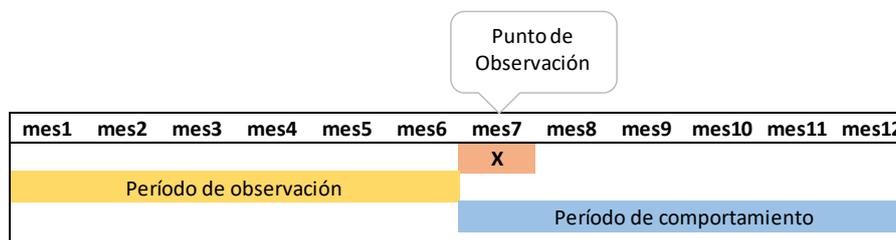


Figura 19: Definición de período de observación y comportamiento  
Fuente: Superintendencia de Bancos. Elaboración propia

Además, con la finalidad de recoger un año de comportamiento en el uso de tarjeta de crédito, y, por otro lado, eliminar los efectos de estacionalidad en consumos de tarjeta de crédito, se han tomado 5 puntos de observación: julio 2018, octubre 2018, enero 2019, abril 2019 y julio 2019. Cada identificación y número de tarjeta se presentará por una sola ocasión en la muestra y para tener este comportamiento se consideró el último dígito de la cédula, cada uno representa el 10% de las identificaciones, quedando así la muestra de la siguiente manera:

Tabla 2  
Distribución de la muestra

Dígito Cédula	ene 18	feb 18	mar 18	abr 18	may 18	jun 18	jul 18	ago 18	sep 18	oct 18	nov 18	dic 18	ene 19	feb 19	mar 19	abr 19	may 19	jun 19	jul 19	ago 19	sep 19	oct 19	nov 19	dic 19
0-1	X	X	X	X	X	X	X	X	X	X	X	X												
2-3				X	X	X	X	X	X	X	X	X	X	X	X									
4-5							X	X	X	X	X	X	X	X	X	X	X	X						
6-7										X	X	X	X	X	X	X	X	X	X	X	X			
8-9													X	X	X	X	X	X	X	X	X	X	X	X

Fuente y elaboración propias

Las condiciones que se verificaron considerando la experticia del banco son:

a) Tener al menos tres meses de uso de tarjeta de crédito en el periodo de estudio, esto debido a que se requiere que el cliente de tarjeta de crédito genere un comportamiento de uso de tarjeta de crédito. Y de acuerdo a la experticia del banco, se puede establecer un patrón de comportamiento a partir del cuarto mes de uso.

b) Tarjetas que en el punto de observación estuvieron activas, esto con la finalidad de considerar tarjetas que registraron consumos.

c) Tener al menos 6 meses de vida de la tarjeta. En los históricos del banco se han verificado que pueden pasar entre tres o cuatro meses desde la entrega de la tarjeta de crédito hasta su primer uso.

Por otro lado, la institución ha identificado diferente comportamiento entre los clientes de tarjeta de crédito VISA y MASTERCARD, y es por esto que este estudio considera las tarjetas VISA.

Una vez aplicadas estas consideraciones se tiene el siguiente resumen de la base a estudiar:

Tabla 3  
Distribución de personas por punto de observación

Punto Observación	Personas
31-jul-18	7,625
31-oct-18	8,491
31-ene-19	8,850
30-abr-19	10,183
31-jul-19	10,935
<b>Total</b>	<b>45,994</b>

Fuente y elaboración propias

#### 4.3. Definición de la variable dependiente

Para definir la variable dependiente se realizaron diferentes análisis que relacione la cancelación de la tarjeta por el cliente y la variación de los consumos tomando en cuenta

los últimos tres meses, en los que no se logró identificar una segmentación que permita explicar esta relación.

Con este antecedente, se recopiló información sobre el costo de mantenimiento y gastos de cobranza promedio anual de la tarjeta de crédito. Se verificó que en promedio el mantener una tarjeta de crédito en un año es de USD 40 (valor mensual de USD 3.33 al dividirlo por 12 meses) pero si se considera además los gastos de cobranza y administrativos se estima un valor total de hasta USD 100 al año (valor mensual de USD 8.33 al dividirlo por 12 meses).

Una vez determinado estos valores, se considera la siguiente segmentación:

1. Aquellas tarjetas que tienen consumos mensuales menores o iguales USD3.33 en los 6 meses de comportamiento se le asigna la marca de inactivo (MALO).
2. Las tarjetas que tienen consumos mensuales mayores o iguales a USD 8.33 se los considerará como activos (BUENOS).
3. Las que consumen mensualmente entre los USD 3.33 y USD 8.33, se los asigna como indeterminados
4. Como punto adicional, se considera el estado cancelado de la tarjeta en el periodo de comportamiento como inactivos observado (MALO).

Con esta segmentación se define la variable dependiente como:

$$Marca = \begin{cases} 1 & \text{Si la tarjeta es activa} \\ 0 & \text{Si la tarjeta es inactiva observada o inactiva} \end{cases} \quad (4.1)$$

La distribución de los datos de acuerdo a la segmentación definida queda de la siguiente manera:

Tabla 4  
**Distribución de personas por punto de observación y segmento**

<b>Punto Observación</b>	<b>Activas</b>	<b>Inactivas</b>	<b>Indeterminadas</b>	<b>Total</b>
31-jul-18	6,243	919	463	7,625
31-oct-18	6,861	977	653	8,491
31-ene-19	7,209	1,058	583	8,850
30-abr-19	8,313	1,129	741	10,183
31-jul-19	8,816	1,407	712	10,935
<b>Total</b>	<b>37,442</b>	<b>5,490</b>	<b>3,062</b>	<b>45,994</b>
Porcentaje	81%	12%	7%	100%

Fuente y elaboración propias

Cabe aclarar que las tarjetas que se catalogaron como indeterminadas, no se las incluirá en la construcción del modelo, ya que de las mismas no se tiene certeza si van a convertirse en tarjetas activas o inactivas, y el porcentaje de estos casos es relativamente bajo.

#### **4.4. Análisis de variables independientes**

Dentro de la información que se va a utilizar se tiene la transaccional que corresponde al detalle de cada transacción realizada con la tarjeta de crédito donde incluye a todos los consumos, el monto, la fecha de transacción, tipo de transacción, las que permiten identificar el comportamiento del uso de tarjeta de crédito.

Por otro lado, el banco facilitó la información financiera que se reporta al ente de control (Superintendencia de Bancos), tal como saldo vencido, saldo total, días de morosidad, pago mínimo realizado, pago mínimo sugerido, adicionalmente la información de pólizas y operaciones de cartera. Como punto importante, se debe tomar en cuenta que dentro de los saldos reportados a fin de mes puede no incluir todos los consumos realizados por cliente con su tarjeta de crédito, ya que pudo haber realizado el pago del consumo antes del corte de su tarjeta, es por esto que la información transaccional es parte fundamental en este modelo.

Finalmente, está la información sociodemográfica que corresponde a variables propias del cliente como fecha de nacimiento, género, estado civil, provincia de residencia, entre otras.

Las variables que se van a utilizar son cualitativas y cuantitativas, y de acuerdo a Rey (2000) las define como:

- a. Cualitativas: expresan cualidad esto es que permite tener atributos y generar conteos definidas en categorías fija; estas variables no permiten operaciones aritméticas y requiere el uso de estadística no paramétrica.
- b. Cuantitativas: Para estos datos se utiliza generalmente la estadística paramétrica, permite operaciones aritméticas.

En esta sección se realiza el análisis de las variables crudas y transformadas mediante operaciones específicas o por medio de árboles de decisión, luego analizar el aporte de dichas variables a la variable dependiente para finalmente incluirlas en el modelamiento.

Se inicia con un total de 118 variables divididas en 11 categóricas y 97 numéricas:

Tabla 5  
Variables iniciales

Nombre de la variable	Descripción de la variable	Tipo variable
Cupof	Cupo Total	numérica
Cupo_max	Cupo Máximo	numérica
Cupo_min	Cupo Mínimo	numérica
Cupo_prom	Cupo Promedio	numérica
min_mesi	pago mínimo mes i	numérica
mor_mesi	mora mes i	numérica
pag_mesi	pago mes i	numérica
sal_mesi	saldo mes i	numérica
ven_mesi	saldo vencido mes i	numérica
avm_mesi	monto de avances mes i	numérica
avc_mesi	cantidad de avances mes i	numérica
compm_mesi	monto en compras mes i	numérica
compc_mesi	cantidad de compras mes i	numérica
difm_mesi	monto en diferidos mes i	numérica
difc_mesi	cantidad de diferidos en el mes i	numérica
avdifm_mesi	monto en avances diferidos mes i	numérica
avdifc_mesi	cantidad de avances diferidos mes i	numérica
trantotm_mesi	monto total de transacciones mes i	numérica
trantotc_mesi	cantidad de transacciones totales en el mes i	numérica
antig	antigüedad de la tarjeta	numérica
Edad	Edad	numérica
AntigüedadLaboral	Antigüedad laboral del cliente	numérica
C_mesi	Tiene operaciones de cartera mes i	Categórica
P_mesi	Tiene pólizas en el mes i	Categórica
RelacionClienteId	Relación que tiene el cliente	Categórica
GeneroId	Genero del cliente	Categórica
EstadoCivilId	Estado Civil	Categórica
CargasFamiliares	Cantidad de cargas familiares	Categórica
NivelInstruccion	Nivel de Instrucción	Categórica
RelacionDependenciaId	Relación dependencia	Categórica
ProvinciaResidencia	Provincia Residencia	Categórica
TipoEmpleadoId	Tipo de empleado público o privado	Categórica
RelacionDependenciaEfx	Relación dependencia 2 categorías	Categórica

Fuente y elaboración propias

Buckinx et al (2006) concluye que el uso de información transaccional permite generar nuevas variables que tienen alto poder de predicción.

Para construir variables nuevas que sean más predictivas, de acuerdo a Ordoñez y Zapata (2016) utiliza los promedios de los últimos 6 meses, proporciones y valores totales de las variables financieras. Finalmente, Pérez (2014) indica que se puede utilizar diferentes tipos de transformaciones entre las principales están: frecuencias de consumos,

valores mínimos o máximos de consumos en los últimos 3 o 6 meses, valores promedios en consumos, etc.

Para las variables cuantitativas se utilizaron los siguientes criterios:

- a. Relaciones entre las variables, por ejemplo,

$$\text{Porcentaje de uso del mes } i = \frac{\text{Saldo mes}_i}{\text{Cupo}}, i = 1, 2, \dots, 6 \quad (4.2)$$

- b. Promedios, Medianas, Desviaciones estándar, por ejemplo:

$$\text{Promedio de avances diferidos de los últimos 3 meses} = \text{prom}(av\_dif_4, av\_dif_5, av\_dif_6) \quad (4.3)$$

- c. Logaritmos, máximos, mínimos

$$\text{Máximo consumo en los últ. 3 meses} = \max(cons_4, cons_5, cons_6) \quad (4.4)$$

- d. Árboles de decisión

Por otro lado, las variables cualitativas se modificó las variables con el uso del indicador WOE.

Con estas transformaciones se han creado un total de 338 variables adicionales, con las que se inicia el análisis para identificar las variables que tienen mayor poder predictivo en el modelamiento.

Una vez obtenidas todas las variables candidatas a ser utilizadas en el modelo, se realizan los siguientes filtros:

a) Es necesario realizar un análisis de correlación entre las variables, para que no haya duplicidad de explicación entre la variable independiente y la variable dependiente. Se elimina variables con correlación mayor al 65%.

b) El indicador KS se utiliza para obtener las variables numéricas más significativas que pronostiquen la variable dependiente, donde se considera aquellos casos con mayor valor.

c) Otro indicador que nos ayuda afinando la selección de las variables con mayor poder de discriminación es la prueba chi-cuadrado, donde toma las variables que son significativas.

d) Tomando en cuenta las variables donde la prueba chi-cuadrado es significativa, se utiliza arboles de decisión para crear nuevas variables y por otro lado poder ver si por sí misma, estas variables pueden mostrar ordenamiento.

Tabla 6  
Variables más significativas

Variable	Descripción	Chi cuadrado
trantotc_max_6m	Cantidad máxima de transacciones últimos 6 meses	17745.496
compm_max_6m	Monto máximo de compras últimos 6 meses	12603.294
can_mov6m	Cantidad de movimientos en últimos 6 meses	8979.368
AnalisisCrediticio	Análisis crediticio	8509.029
trantotm_min_6m	Cantidad mínima de transacciones últimos 6 meses	4292.443
ContinuidadLaboralId	Mantiene continuidad laboral	3834.067
difc_max_6m	Cantidad máxima de diferidos últimos 6 meses	3453.302

Fuente y elaboración propias

#### 4.5. Modelo Logit

##### 4.5.1. Muestra de Validación y Modelamiento

Siddiqui (2006) considera que para la construcción del modelo se puede considerar entre el 70% al 80% de la muestra total, mientras que para realizar el testeo se utiliza la muestra restante que sería entre el 20% y 30%. Para este ejercicio se consideró tomar el 70% para elaborar el modelo y el 30% para el testeo del modelo y verificar que tenga precisión similar que la obtenida con la base de modelamiento. Se consideró una selección aleatoria, de tal manera que las dos tengan la misma distribución en función a la variable dependiente:

Tabla 7  
Distribución de muestra de training y testeo

		Activas	Inactivas	Total	% Casos
Base de Training	Casos	26210	3843	30053	70%
	Proporción	87%	13%	100%	
Base Testeo	Casos	11,232	1,647	12,879	30%
	Proporción	87%	13%	100%	
Total		37,442	5,490	42,932	100%

Fuente y elaboración propias

##### 4.5.2. Aplicación del modelo Logit

Luego de probar algunos modelos de regresión considerando los principales indicadores como KS, ROC y GINI se obtiene el siguiente modelo:

$$\text{Modelo} = \text{glm}(\text{marca} = \text{avdifc\_max\_6m} + \text{Arbtrantotm\_min\_6mContLab} + \text{difc\_max\_6m} + \text{can\_mov6m} + \text{compc\_max\_6m} + \text{trantotm\_min\_3m} + \text{Pol\_6m} + \text{Genero}) \quad (4.5)$$

Con los siguientes pesos:

Tabla 8  
Resultados modelo

Variable	Coefficiente	Error Estándar	Valor Z	Probabilidad	Odds
(Intercept)	- 0.09441	0.04522	-2.088	0.0368 ***	0.91
avdifc_max_6m	0.65037	0.02206	29.476	<2e-16 ***	1.92
Arbtrantotm_min_6mContLab	0.04678	0.00095	49.454	<2e-16 ***	1.05
difc_max_6m	- 0.13855	0.02601	-5.326	1.00E-07 ***	0.87
can_mov6m	- 0.04870	0.00072	-67.985	<2e-16 ***	0.95
compc_max_6m	0.02688	0.00298	9.033	<2e-16 ***	1.03
trantotm_min_3m	0.00035	0.00002	15.051	<2e-16 ***	1.00
Pol_6m	1.23036	0.05052	24.353	<2e-16 ***	3.42
GeneroIdMAS	0.36397	0.03571	10.193	<2e-16 ***	1.44

Fuente y elaboración propias

De acuerdo a los resultados, se observa que con el 99% de confianza, los coeficientes son estadísticamente significativos.

Descripción de las variables finales:

- Avdifc\_max\_6m.- Cantidad máxima de avances diferidos en los últimos 6 meses.

$$\text{Avdifc\_max\_6m} = \max_i(\text{Cantidad de avances diferidos mes } i), \quad i = 1, 2, \dots, 6 \quad (4.6)$$

- Arbtrantotm\_min\_6mContLab. - Árbol de decisión entre el mínimo monto de transacciones y la continuidad laboral del cliente.

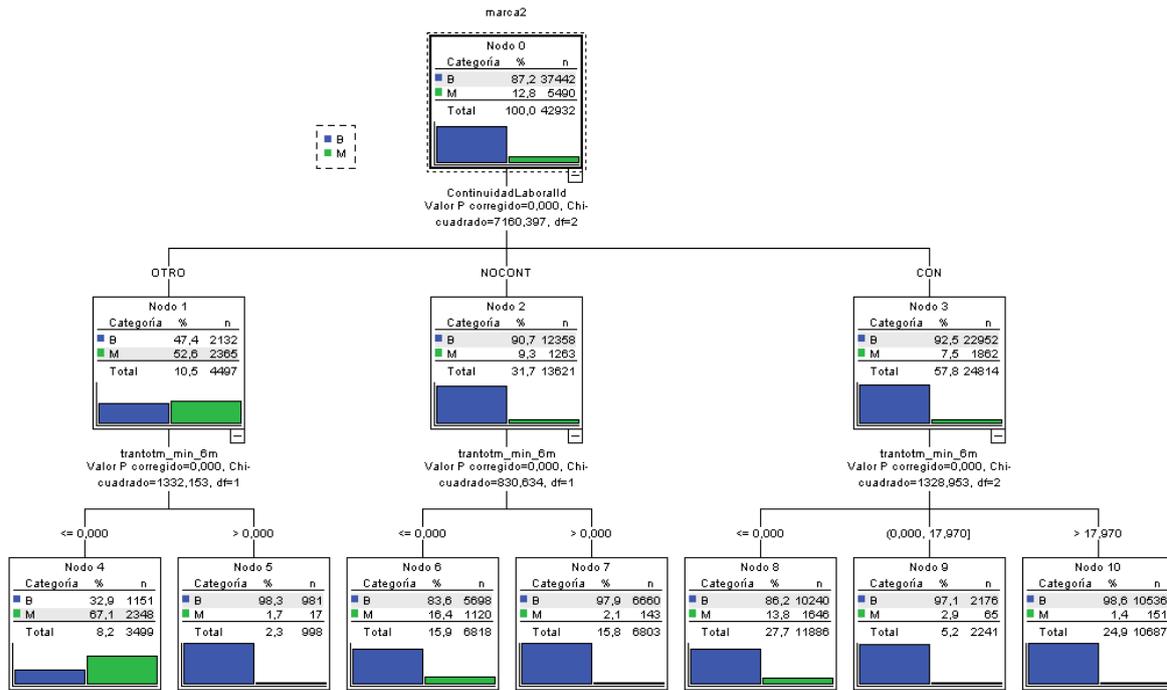


Figura 20: Árbol de decisión monto transacciones y continuidad laboral  
Elaboración propia

- Difc\_max\_6m.- Máxima cantidad de diferidos en los últimos 6 meses.

$$Difc\_max\_6m = \max_i(Cantidad\ de\ diferidos\ mes\ i), \quad i = 1,2, \dots, 6 \quad (4.7)$$

- Can\_mov6m. – Cantidad de movimientos financieros en los últimos 6 meses.

$$Can\_mov6m = \sum_{i=1}^6 cantidad\ movimientos\ mes\ i, \quad i = 1,2, \dots, 6 \quad (4.8)$$

- Compc\_max\_6m. – Cantidad máxima de compras en los últimos 6 meses.

$$Compc\_max\_6m = \max_i(Cantidad\ de\ compras\ mes\ i), \quad i = 1,2, \dots, 6 \quad (4.9)$$

- Trantotm\_min\_3m.- Monto mínimo de transacciones totales en los últimos 3 meses.

$$Trantotm\_min\_3m = \min_i(Monto\ total\ de\ transacción\ mes\ i), \quad i = 1,2, \dots, 6 \quad (4.10)$$

- Pol\_6m.- Marca de si tiene o no póliza en los últimos 6 meses que corresponde a una variable dicotómica.

$$Pol\_6m = \begin{cases} 0, & \text{no tiene póliza en los últimos 6 meses} \\ 1, & \text{tiene póliza en cualquiera de los últimos 6 meses} \end{cases} \quad (4.11)$$

- Género: Género del cliente, que toma dos valores: masculino y femenino.

#### 4.6. Interpretación de resultados del modelo

En esta sección se verificará la consistencia de las variables finales en la interpretación de su signo que permite identificar si cada una aporta como castigo en la probabilidad final.

1. Avdific\_max\_6m: Su odds ratio es de 1.92 lo que indica si una persona tiene máximo un avance diferido en los últimos 6 meses es un 92% más propenso a ser desertor.
2. Dific\_max\_6m: Su odds ratio es de 0.87 lo que indica si una persona tiene máximo una transacción diferida en los últimos 6 meses es un 13% menos propenso a ser desertor.
3. Can\_mov6m: El odds ratio es de 0.95, lo que muestra que, si una persona tiene un movimiento en los últimos 6 meses, es en un 5% menos propenso de ser desertor.
4. Compc\_max\_6m: Su odds ratio es de 1.03, lo que indica que, si una persona presenta una compra en los últimos 6 meses, es en un 3% más propenso a ser desertor.
5. Trantotm\_min\_3m: El odds ratio es de 1.0004, lo que muestra que, si una persona tiene al menos un dólar de consumo en los últimos 3 meses, es en un 0.04% más propenso a ser desertor.
6. Pol\_6m: Su odds ratio es de 3.42, esto significa que si una persona ha tenido al menos una póliza en los últimos 6 meses es en un 242% más propenso de ser desertor.
7. Género: El odds es de 1.44, lo que se resume en que, si un cliente tiene género masculino, tiene un 44% de probabilidad de ser más propenso un desertor.

#### **4.7. Validación del Modelo**

##### **4.7.1. Curvas ROC y GINI**

Se puede verificar el nivel de ajuste del modelo en la base de entrenamiento de acuerdo a la curva ROC y los indicadores GINI y KS.

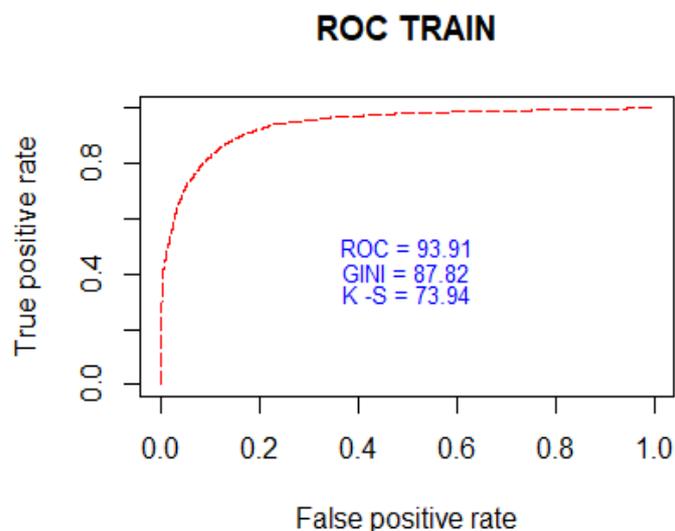


Figura 21: Curva ROC en muestra de entrenamiento  
Elaboración propia

Como validación del modelo se prueba con la muestra de testeo. En esta base se incorporan los casos de la base de testeo con marca de indeterminados, considerando los casos como malos ya que no se tiene certeza si consumirán con la tarjeta de tal manera que generen rentabilidad para el banco o la cancelarán. Pese a que no formaron parte de la muestra de modelamiento, pertenecen a la población total y al incluirlos en el testeo se esperaría obtener comportamiento similar en el performance (Pérez, 2014). Dentro de esta base se tiene el siguiente resultado:

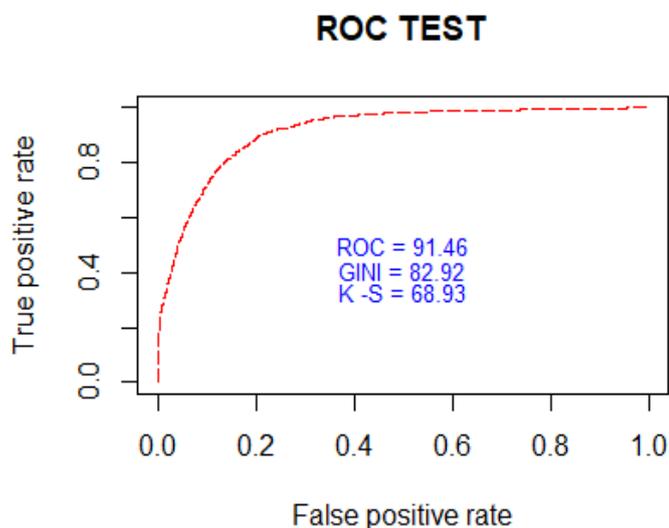


Figura 22: Curva ROC en muestra de testeo  
Elaboración propia

Al unificar las bases de entrenamiento y testeo se tiene los siguientes indicadores:

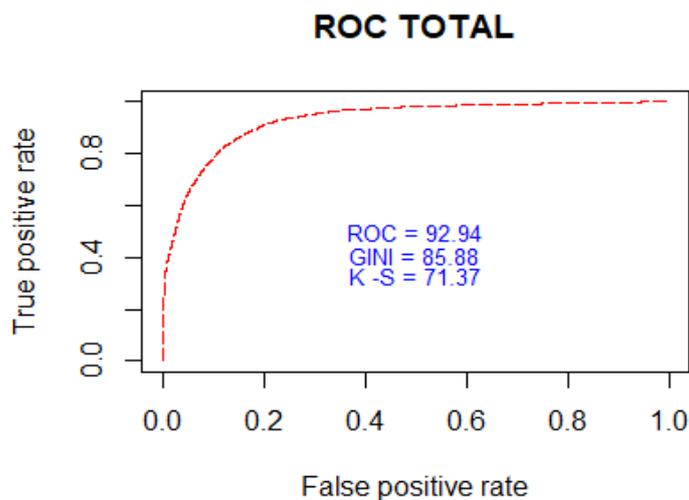


Figura 23: Curva ROC en muestra de entrenamiento y testeo  
Elaboración propia

Al comparar los indicadores obtenidos se observa que se tiene comportamiento similar en otra base diferente a la de entrenamiento:

Tabla 9  
**Variación en base de entrenamiento y testeo**

	Entrenamiento	Testeo	Total
ROC	93.91	91.46	92.94
GINI	87.82	82.92	85.88
KS	73.94	68.93	71.37

Fuente y elaboración propias

#### 4.7.2. Tabla de performance y estadísticos

Para poder verificar el performance del modelo se tiene el siguiente resultado que refleja el ordenamiento en función de los deciles de probabilidad, además muestra discriminación, ubicando un mayor porcentaje de casos malos en los deciles de peor probabilidad como se muestra en la tabla:

Tabla 10  
**Performance del modelo -Base Train**

Prob. deserción		Total		Buenos			Malos			
Mín	Max	Núm.	%	Núm.	%	Proporción	Núm.	%	% Decum	Proporción
0.0%	0.1%	3,006	10%	2,997	11.3%	99.7%	9	0.2%	100%	0.3%
0.1%	0.5%	3,005	10%	2,991	11.2%	99.5%	14	0.4%	99.8%	0.5%
0.5%	1.8%	3,005	10%	2,988	11.2%	99.4%	17	0.4%	99.4%	0.6%
1.8%	4.1%	3,005	10%	2,984	11.2%	99.3%	21	0.5%	99.0%	0.7%
4.1%	7.9%	3,006	10%	2,973	11.2%	98.9%	33	0.9%	98.5%	1.1%
7.9%	13.5%	3,005	10%	2,949	11.1%	98.1%	56	1.5%	97.6%	1.9%
13.5%	21.5%	3,005	10%	2,880	10.8%	95.8%	125	3.3%	96.1%	4.2%
21.5%	34.5%	3,015	10%	2,654	10.0%	88.0%	361	9.4%	92.8%	12.0%
34.5%	61.3%	2,995	10%	2,072	7.8%	69.2%	923	24.0%	<b>83.4%</b>	30.8%
61.3%	100.0%	3,006	10%	722	2.7%	24.0%	2284	59.4%	59.4%	76.0%
Total		30,053	100%	26,210	100%	87.2%	3,843	100.00%		12.8%

Fuente y elaboración propias

De acuerdo a la tabla 10, se verifica el nivel de ordenamiento de los casos malos donde la proporción de malos es creciente conforme se incrementa la probabilidad, por otro lado, el porcentaje de casos malos, tomando en cuenta los dos deciles de mayor probabilidad suman el 83.4%, por lo que se verifica el nivel de discriminación.

Tabla 11  
Performance del modelo -Base Test

Prob. deserción		Total		Buenos			Malos		
Mín	Max	Núm.	%	Núm.	%	Proporción	Núm.	%	Proporción
0.0%	0.1%	1,380	10%	1,372	12.2%	99.4%	8	0.3%	0.6%
0.1%	0.6%	1,379	10%	1,371	12.2%	99.4%	8	0.3%	0.6%
0.6%	2.1%	1,380	10%	1,368	12.2%	99.1%	12	0.5%	0.9%
2.1%	5.0%	1,379	10%	1,364	12.1%	98.9%	15	0.6%	1.1%
5.0%	9.4%	1,380	10%	1,349	12.0%	97.8%	31	1.2%	2.2%
9.4%	15.9%	1,380	10%	1,285	11.4%	93.1%	95	3.7%	6.9%
15.9%	25.1%	1,379	10%	1,173	10.4%	85.1%	206	8.0%	14.9%
25.1%	38.4%	1,380	10%	943	8.4%	68.3%	437	17.0%	31.7%
38.4%	63.2%	1,379	10%	680	6.1%	49.3%	699	27.3%	50.7%
63.2%	99.7%	1,380	10%	327	2.9%	23.7%	1053	41.1%	76.3%
Total		13,796	100%	11,232	100%	81.4%	2,564	100.00%	18.6%

Fuente y elaboración propias

Para la muestra de testeo se incluyó el 20% de casos indeterminados, y se tiene similar comportamiento que en la base de entrenamiento tanto para el nivel de ordenamiento como para la discriminación.

Tabla 12  
Performance del modelo -Base Total

Prob. deserción		Total		Buenos			Malos			
Mín	Max	Núm.	%	Núm.	%	Proporción	Núm.	%	% Decum	Proporción
0.0%	0.1%	4,385	10%	4,369	11.7%	99.6%	16	0.2%	100.0%	0.4%
0.1%	0.5%	4,385	10%	4,362	11.7%	99.5%	23	0.4%	99.8%	0.5%
0.5%	1.9%	4,385	10%	4,355	11.6%	99.3%	30	0.5%	99.4%	0.7%
1.9%	4.4%	4,385	10%	4,350	11.6%	99.2%	35	0.5%	98.9%	0.8%
4.4%	8.4%	4,384	10%	4,326	11.6%	98.7%	58	0.9%	98.4%	1.3%
8.4%	14.2%	4,385	10%	4,252	11.4%	97.0%	133	2.1%	97.5%	3.0%
14.2%	22.6%	4,385	10%	4,078	10.9%	93.0%	307	4.8%	95.4%	7.0%
22.6%	35.9%	4,385	10%	3,584	9.6%	81.7%	801	12.5%	90.6%	18.3%
35.9%	61.9%	4,381	10%	2,716	7.3%	62.0%	1665	26.0%	78.1%	38.0%
61.9%	100.0%	4,389	10%	1,050	2.8%	23.9%	3339	52.1%	52.1%	76.1%
Total		43,849	100%	37,442	100%	85.4%	6,407	100.00%		14.6%

Fuente y elaboración propias

Al unificar la base de testeo y entrenamiento se puede constatar que, en los últimos dos deciles de probabilidad de ser malo, acumula aproximadamente el 78% de casos malos y además el ordenamiento muestra un crecimiento razonable.

### 4.7.3. Multicolinealidad

Un punto adicional es verificar que no exista multicolinealidad entre las variables finales del modelo utilizando el factor de inflación de la varianza generalizada:

Tabla 13  
**Factor de inflación de la varianza generalizada**

Variable	VIF
avdific_max_6m	1.260205
Arbtrantotm_min_6mContLab	1.115518
dific_max_6m	1.106191
can_mov6m	1.345767
compc_max_6m	1.216114
trantotm_min_3m	1.130464
Pol_6m	1.059655
GeneroIdMAS	1.034757

Fuente y elaboración propias

De acuerdo a la tabla anterior se verifica que no existe multicolinealidad ya todos los valores son cercanos a uno.

Por otro lado, para reforzar la verificación de la multicolinealidad, se comprueba la correlación entre las variables, donde se tiene:

Tabla 14  
**Matriz de correlación en variables del modelo**

	avdific_max_6m	Arbtrantotm_min_6mContLab	dific_max_6m	can_mov_6m	compc_max_6m	Trantotm_min_3m	Pol_6m
avdific_max_6m	1.000	-0.052	0.085	0.231	-0.113	0.082	-0.081
Arbtrantotm_min_6mContLab		1.000	-0.217	-0.305	-0.287	-0.126	-0.080
dific_max_6m			1.000	0.260	0.180	0.199	0.029
can_mov_6m				1.000	0.364	0.366	0.023
compc_max_6m					1.000	0.285	0.092
trantotm_min_3m						1.000	0.018
Pol_6m							1.000

Fuente y elaboración propias

El nivel de correlación para todas las variables del modelo es aceptable, ya que ninguna sobrepasa el 50% del nivel de correlación.

#### 4.8. Especificaciones para aplicación del CRM

Con la finalidad de poder identificar las estrategias que se deben aplicar en función de la tabla de performance se definen puntos de corte:

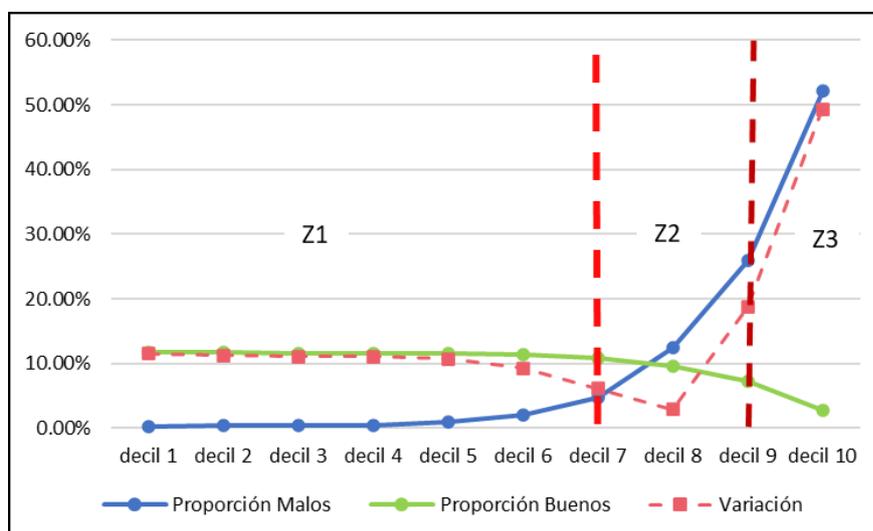


Figura 24: Definición de puntos de corte de probabilidad  
Elaboración propia

En función de la tendencia en el crecimiento de la proporción de los casos malos se divide en tres zonas:

1. Zona 1 (Z1): presenta menor probabilidad de que un cliente pueda ser desertor, no se requiere utilizar estrategia de fidelización focalizada,
2. Zona 2: corresponde a los deciles 7 y 8 donde la proporción de casos desertores empieza a incrementar. En este grupo se puede aplicar estrategias focalizadas mediante el CRM analítico donde prima el uso de los datos de los clientes identificando las preferencias en sus consumos y paralelamente el e-CRM para reforzar la relación con el cliente mediante el uso de canales y medios electrónicos.
3. Zona 3: a partir del decil 9 donde el nivel de proporción de casos desertores se incrementa de forma casi exponencial, se puede aplicar el CRM estratégico, donde por medio de técnicas de negocio se puede focalizar la automatización de la fuerza de ventas y llegar a estos clientes con propuestas agresivas que convencen al cliente del uso de tarjeta de crédito de este banco.

Una vez obtenido el modelo se puede clasificar a los clientes considerando su probabilidad de default y con esto se puede aplicar la técnica de CRM analítico y de esta manera proponer estrategias de fidelización de los clientes tomando en cuenta mayor gestión en los que presenten mayor probabilidad de ser desertores de tarjeta de crédito, e inclusive a quienes se encuentren en una probabilidad que no identifica si un cliente va a desertar o ser fiel usuario de tarjeta se puede aplicar una estrategia para que se incremente el uso de tarjeta de crédito con estrategias segmentadas.

## Conclusiones y Recomendaciones

### 1. Conclusiones

Para reducir la cantidad de variables independientes se utilizaron diferentes estrategias, entre ellas tenemos el análisis de correlación identificando casos con valor mayor a 65%, estadístico KS para cada variable y finalmente la prueba chi cuadrado que Mendivelso y Rodríguez (2018) mencionaron ser necesaria para identificar la independencia entre estas variables.

La variable dependiente se determinó mediante la identificación de los costos de mantenimiento (USD40) y los gastos de cobranza (USD 60) anuales para identificar un corte específico que distinga los casos de clientes desertores (inactivos) o no desertores (activos), en este caso se consideraron aquellas tarjetas que en los 6 meses de desempeño han tenido hasta USD 3.33 de consumos mensual como desertores, y aquellas que tengan consumos de mayores a USD 8.33 (valor de mantenimiento + gastos de cobranza mensual) en los 6 meses tendrán marca de no desertores.

Los casos indeterminados, corresponden a quienes tienen costo de mantenimiento y cobranza mensual entre USD 3.33 y USD 8.33 en los 6 meses de desempeño y estos casos no se los incluirá en el proceso de modelamiento porque no se tiene certeza si serán desertores o no.

El modelo que permite identificar si un cliente es un posible desertor (inactivo) o es no desertor en el uso de tarjeta de crédito es la regresión logística, ya que permite identificar claramente como cada variable influye en la probabilidad final de ser o no desertor, tal como lo sugirió Lemmens y Croux, (2006) e inclusive sería transparente ante la supervisión de los entes reguladores.

Las principales variables que indican la tendencia de desertar de un cliente son: avances diferidos, monto de transacciones, continuidad laboral, diferidos, cantidad de movimientos, cantidad de compras y si tiene pólizas en la institución financiera.

Se puede aplicar estrategias de CRM en función a la probabilidad de deserción para fidelizar a los clientes de manera focalizada, como lo especificó Bambino Contreras (2005) y con la finalidad de mantener y fidelizar a sus clientes (Berry. 1983).

## **2. Recomendaciones**

Se recomienda el uso del modelo logístico ya que permite tener una clara interpretación de cada variable incluida y puede ser de fácil entendimiento tanto para el usuario final del modelo como para el ente regulador ante una supervisión al banco.

Se recomienda el uso del CRM analítico ya que este se enfoca en procesos automatizados tecnológicamente y pueden ser aplicados en función de su valor de probabilidad de deserción con la finalidad de que la entidad financiera X ponga en práctica las estrategias para que sus clientes se fidelicen en el uso de la tarjeta de crédito.

Se recomienda seleccionar dos puntos de corte en función de la probabilidad de ser desertor y considerar tres zonas, donde la primera no requiere una estrategia de fidelización sino más bien las dos restantes (con mayor probabilidad de deserción), para esto se sugiere aplicar técnicas de CRM específico para cada una enfocándose en los CRM estratégico, e-CRM y analítico para poder generar estrategias agresivas que permitan fidelizar a clientes que estén próximos a desertar.

## Obras Citadas

- Abbasimehr, H, Mostafa, S y M. J. Tarokh. 2011. *A Neuro-Fuzzy Classifier for Customer Churn Prediction*, International Journal of Computer Applications (0975 – 8887), Volume 19– No.8.
- AEMR. 2002. *Asociación Española de Marketing Relacional*
- Altman, E. 1968. *Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy*, Journal of Finance, 23(4), 589-609.
- Amar, J., Abello, R., Denegri, M., y Llanos, M. 2001. *La comprensión del funcionamiento bancario en la adolescencia: ¿Crónica de un sobreendeudamiento anunciado?* Barranquilla, Colombia.
- Anderson, D., Sweeney, D. y Williams, T. 2008. *Statistics for Business and Economics*, Mason, Ohio, Estados Unidos: Thomson
- Assael, H., y Roscoe Jr, A. M. 1976. *Approaches to market segmentation analysis*. Journal of Marketing, Vol 40.
- Bhalla, D. 2016. *Weight of Evidence (WOE) and Information Value Explained*. URL: <https://www.listendata.com/2015/03/weight-of-evidence-woe-and-information>.
- Berry, L. L. 2002. *Relationship marketing of services perspectives from 1983 and 2000*. Journal of relationship marketing, Vol 1.
- Berry, M. J., y Linoff, G. S. 2004. *Data mining techniques: for marketing, sales, and customer relationship management*. John Wiley & Sons.
- Bentz, Y, Merunka, D. 2000. *Neural networks and the multinomial legit for brand choice modeling: a hybrid approach*, Journal of Forecasting, vol. 19, pp.177–200.
- Breiman, L., Friedman, J., Olshen, R., y Stone, C. 1984. *Classification and Regression Trees*. Belmont, CA: Chapman & Hall.
- Buckinx et al., 2006 Buckinx W, Verstraeten G,y Van del Poel, D. 2006. *Predicter customer loyalty using the internal transactional database. Expert Sistem with application*.
- Cobeña Terán, B. D. 2016. *Modelo de Predicción de Deserción de Clientes de Tarjetas de Crédito (Bachelor's thesis, Espol)*.
- Cóndor, J. 2018. “Reporte comportamiento crediticio sectorial”. Subdirección de estadísticas y estudios: SBS.

- Coussement, K. F. Benoit, D. Van den Poel, D. 2010. *Improved marketing decision making in a customer churn prediction context using generalized additive models*, Expert Systems with Applications 37, 2132–2143
- Chen, I.J.; Popovich, K. 2003: *Understanding Customer Relationship Management- People, Process and Technology*, Business Process Management Journal, Vol 9.
- EC. 2014. *Código Orgánico Monetario y Financiero*. Registro Oficial 332, Segundo Suplemento, 12 de septiembre.
- Edmister, R. 1972, *An Empirical Test of Financial Ratio Analysis for Small Business Failure Prediction* Journal of Financial and Quantitative Analysis, pp. 1477-1493
- Espin-García, O., y Rodríguez-Caballero, C. V. 2013. *Metodología para un scoring de clientes sin referencias crediticias*. Cuadernos de economía, 32(59), 137-162.
- Flórez, O., y Rincón, W. 2012. *Modelo Logit y Probit: Un caso de aplicación*. Comunicaciones en Estadística, 5(2).
- Fox, J. 2003. *Linear Models, Problems*. Canada: McMaster University.
- Gorospe, I. 2018. “Sector Financiero en Ecuador”. Oficina Económica y Comercial de la Embajada de España en Quito.
- Gutiérrez, R. 1992. *Mercadeo para banqueros*. Buenos Aires: Fondo editorial FELABAN.
- Han – Yuh, L. 2007. *Development of a framework for customer relationship management (CRM) in the banking industry*. International Journal of Management, Vol. 24.
- Hand D. y Henley W. 1997. *Statistical Classification Methods in Costumer Credit Scoring: A review*, Journal of the Royal Statistical Association Series A. Vol. 160
- Hangxia, Ma, Min Qin, Jianxia W. 2009. *Analysis of the Business Customer Churn Based on Decision Tree Method*, The Ninth International Conference on Control and Automation, Guangzhou, China.
- Hee-Su KimyChoong-Han Yoon. 2004. *Determinants of subscriber churn and customer loyalty in the Korean mobile telephony market*. Telecommunications Policy, vol. 28, pp.751-765.
- Herrera L, 2016. *Diseño de un programa de retención de usuarios de tarjeta de crédito*. Ingeniería Administrativa.
- Huertas, R. y Martí, M., 2013. *Análisis del credit scoring*, Revista de Administração de Empresas
- Kaura, V. 2013. *Antecedents of customer satisfaction: A study of indian public and private sector banks*. International Journal of Bank Marketing, 167 - 186.

- Lemmens, A., and Croux, C. 2006. *Bagging and boosting classification trees to predict churn*, Journal of Marketing Research, vol. 43, no. 2, pp. 276-286.
- Lizarzaburu, E. R., Berggrun, L., y Quispe, J. 2012. *Gestión de riesgos financieros. Experiencia en un banco latinoamericano I*. Estudios Gerenciales, 28(125), 96-103.
- Mendivelso, F., y Rodríguez, M. 2018. *Prueba Chi-cuadrado de independencia aplicada a tablas 2xN*. Revista Médica Sanitas, 21(2), 92-95.
- Morales, H., 2008. *Índice de confianza del consumidor: Barómetro de opinión de los hogares estadounidenses*. El Salvador: Banco Central de Reserva de El Salvador
- Neter, J., Wasserman, W., y Kutner, M. H. 1989. *Applied linear regression models*.
- Ning Wang, Dong-xiao Niu. 2009., *Credit Card Customer Churn Prediction Based on the RST and LS-SVM*, 6th International Conference on Service Systems and Service Management pp. 275-279, 2009.
- Ordoñez C, R. A. D. C., y Pastor Z., M. D. 2016. *Sistema de predicción de clientes desertores de tarjetas de crédito para la banca peruana usando Support Vector Machine*.
- Parvatiyar, A., y Sheth, J. N. 2001. *Customer relationship management: Emerging practice, process, and discipline*. Journal of Economic and Social Research, Vol. 3.
- Payne, A. 2012. *Handbook of CRM: Achieving Excellence in Customer Management* (Vol. 1). Routledge.
- Pérez Tatamués, A. E. 2014. *Modelo de activación de tarjetas de crédito en el mercado crediticio ecuatoriano a través de una metodología analítica y automatizada en R* (Bachelor's thesis, Quito, 2014.).
- Reichheld, F. y Sasser, W. 1990. *Zero defection: quality comes to services*. Harvard Business Review, 105–111.
- Rey, F. L. G. 2000. *Lo cualitativo y lo cuantitativo en la investigación de la psicología social*. Revista cubana de Psicología, 17(1).
- Rodríguez, E., y Silva, E. 2011. *Análisis del costo y el crédito generado por el uso de la tarjeta de crédito como medio de pago transaccional en el Ecuador (2000-2010)*. Tesis. Quito, Pichincha, Ecuador.
- Romani, G., Aroca, P., Aguirre, N., Leiton, P. y Muñoz, J. 2002. *Modelos de clasificación y predicción de quiebra de empresas: Una aplicación a empresas chilenas*, Forum Empresarial, Vol. 7, nº 001

- Sarimveis, H. y Bafas, G. 2003. *Fuzzy model predictive control of non-linear processes using genetic algorithms*. Fuzzy Sets Syst., vol.139, pp.59–80, 2003.
- Siddiqi, N. 2006. *Credit risk scorecards: developing and implementing intelligent credit scoring*. Hoboken, N.J: Wiley.
- Smith, R. y Winakor, A. 1935, *Changes in the Financial Structure of Unsuccessful Industrial Corporations, Bulletin No. 51*, University of Illinois, Bureau of Business Research, Urbana.
- Tamari, M. 1966. *Financial ratios as a means of forecasting bankruptcy*”, Management International Review, 4, 15-21.
- Torkzadeh, G., Chang, J. C.-J., y Hansen, G. W. 2006. *Identifying issues in customer relationship management at Merck-Medco*. *Decision Support Systems*, 42(2).
- Umayaparvathi, V., Iyakutti, K. 2012. *Applications of Data Mining Techniques in Telecom Churn Prediction*, International Journal of Computer Applications (0975 – 8887), Volume 42– No.20
- Van den Poel, D., y Larivière, B. 2004. *Customer attrition analysis for financial services using proportional hazard models*. European Journal of Operational Research, 108 157(1), 196–217.
- Wonsuk, Y., Mayberry, R., Bae, S, Singh, K., Qinghua (Peter) He, y James W. Lillard. 2014. *A Study of Effects of MultiCollinearity in the Multivariable Analysis*. International journal of applied science and technology 4 (5): 9–19.
- Yang S. 2005. *Application of data mining in analysis of customers to leave company in telecom*, Computer and Modernization, vol.2, pp.109-111.
- Yap, B., Ong, S., Husain, Nor Huselina Mohamed. 2011. *Using data mining to improve assessment of credit worthiness via credit scoring models*. *Expert Systems with Applications*, vol. 38, no 10, p. 13274-13283.