

**PROBABILIDAD DE INCUMPLIMIENTO DE UNA CARTERA CASTIGADA, EN  
LA MODALIDAD DE MICROCRÉDITO, PARA UNA ENTIDAD DEL SISTEMA  
FINANCIERO COLOMBIANO EN EL AÑO 2017**



CORPORACION UNIVERSITARIA  
**AUTONOMA**  
DEL CAUCA

**MARÍA CAMILA FERNÁNDEZ BOLAÑOS  
LINA MARCELA BOLAÑOS MARTÍNEZ**

**CORPORACIÓN UNIVERSITARIA AUTÓNOMA DEL CAUCA  
FACULTAD DE CIENCIAS ADMINISTRATIVAS, CONTABLES Y ECONÓMICAS  
TRABAJO DE INVESTIGACIÓN  
2018**

**PROBABILIDAD DE INCUMPLIMIENTO DE UNA CARTERA CASTIGADA, EN  
LA MODALIDAD DE MICROCRÉDITO, PARA UNA ENTIDAD DEL SISTEMA  
FINANCIERO COLOMBIANO EN EL AÑO 2017**



**CORPORACION UNIVERSITARIA  
AUTONOMA  
DEL CAUCA**

**MARÍA CAMILA FERNÁNDEZ BOLAÑOS**

**LINA MARCELA BOLAÑOS MARTÍNEZ**

**FINANZAS Y NEGOCIOS INTERNACIONALES**

**DIEGO FERNANDO FERNÁNDEZ CALAMBAS**

**CORPORACIÓN UNIVERSITARIA AUTÓNOMA DEL CAUCA  
FACULTAD DE CIENCIAS ADMINISTRATIVAS, CONTABLES Y ECONÓMICAS  
LÍNEA DE INVESTIGACIÓN FINANCIERA  
TRABAJO DE INVESTIGACIÓN**

**2018**

## NOTA DE ACEPTACIÓN

Mediante la presente se acepta la sustentación del trabajo “PROBABILIDAD DE INCUMPLIMIENTO DE UNA CARTERA CASTIGADA, EN LA MODALIDAD DE MICROCRÉDITO, PARA UNA ENTIDAD DEL SISTEMA FINANCIERO COLOMBIANO EN EL AÑO 2017”.

---

NORMAN OSWALDO CALDÓN QUIRA

Jurado I c.c.

---

JOHN OLIVER ARCINIEGASHERRERA

Jurado II c.c.

---

MARÍA DEL CARMEN COLLAZOS ASTUDILLO

Presidente del Comité de Investigaciones

c.c.

---

DIEGO FERNANDO FERNANDEZ CALMBÁS

Director de la investigación

c.c.

## DEDICATORIA

*Inicialmente se la dedicamos a Dios, que nos permitió tener la presente formación profesional. A nuestros padres por su constante apoyo, tanto emocional como económico y a todos lo que nos acompañaron en el transcurso de esta etapa.*

## **AGRADECIMIENTOS**

En primera instancia queremos agradecer a Dios, por guiarnos y ayudarnos a derrotar las adversidades con convicción. A nuestro director de tesis: Diego Fernández Calambás, por su compromiso, paciencia, dedicación y aporte de conocimientos para el desarrollo de este trabajo. A la Corporación Universitaria Autónoma del Cauca y en especial a la Facultad de Ciencias Administrativas, Contables y Económicas (FACACE), por todos los conocimientos y calor humano brindados, para nuestro desarrollo como profesionales. A nuestra familia como pilar fundamental de nuestras vidas; y amigos por la motivación y el apoyo incondicional ofrecido, en todo el proceso.

## TABLA DE CONTENIDO

RESUMEN .....	11
INTRODUCCIÓN .....	13
CAPÍTULO I: PROBLEMA.....	16
Planteamiento del problema .....	16
Justificación .....	22
Objetivos .....	25
Objetivo general.....	25
Objetivos específicos .....	25
CAPÍTULO II: MARCO TEÓRICO Y REFERENTES CONCEPTUALES.....	26
Bases teóricas .....	26
Antecedentes .....	33
CAPÍTULO III: METODOLOGÍA.....	40
Proceso.....	41
a. Recopilación de información: .....	41
b. Tratamiento de la información:.....	42
c. Población y Muestra de la investigación: .....	45
d. Variables del modelo.....	46
e. Método ewma:.....	46
f. Aplicación del modelo: .....	47
g. Prueba del modelo mediante backtesting: .....	47
CAPÍTULO IV: RESULTADOS.....	48
Análisis descriptivo.....	49

Cruce de variables .....	59
Modelo Logit .....	65
1. Etapa: Selección de la muestra.....	66
2. Etapa: Variables aplicadas al modelo .....	68
3. Etapa: REGRESIÓN LOGÍSTICA .....	69
Bloque inicial: Bloque 0.....	70
Bloque 1: Método= Avanzar por pasos (Wald).....	71
CURVA COR .....	76
Prueba Backtesting.....	78
Análisis de los resultados del modelo Logit .....	80
a. Analisis de la probabilidad de incumplimiento de los departamentos. ....	80
b. Análisis de la probabilidad de incumplimiento del estado civil. ....	81
c. Análisis de la probabilidad de incumplimiento del nivel educativo .....	82
d. Análisis de la probabilidad de incumplimiento de cada compra de la cartera.....	83
CAPÍTULO V: CONCLUSIONES Y RECOMENDACIONES .....	85
Conclusiones .....	85
Recomendaciones .....	87
Bibliografía .....	88

## ÍNDICE DE FÓRMULAS

Ecuación 1 Formula logit.....	36
Ecuación 2 Formula Wald para variables numéricas .....	37
Ecuación 3 Formula Wald para variables categóricas.....	37
Ecuación 4 R cuadrado de Cox y Snell .....	43
Ecuación 5 R cuadrado Nagelkerke .....	44

## ÍNDICE DE ILUSTRACIONES

Ilustración 1 Curva de ajuste relación perfecta.....	42
Ilustración 2 Curva de ajuste pobremente ajustada .....	43
Ilustración 3 Distribución de clientes que pagan y no pagan.....	49
Ilustración 4 Distribución de la cartera .....	50
Ilustración 5 Departamento .....	52
Ilustración 6 Días de atraso.....	53
Ilustración 7 Estado Civil .....	54
Ilustración 8 Nivel educativo.....	56
Ilustración 9 Cuotas canceladas.....	57
Ilustración 10 Cuotas en mora.....	58
Ilustración 11 Cruce de variables entre departamento y Logit .....	60
Ilustración 12 Cruce de variables entre Cartera y Logit.....	62
Ilustración 13 Cruce de variables entre nivel educativo y Logit .....	63
Ilustración 14 Cruce de variables entre Estado civil y Logit .....	65
Ilustración 15 Curva COR.....	77
Ilustración 16 Probabilidad de incumplimiento por departamento .....	81
Ilustración 17 Probabilidad de incumplimiento por estado civil .....	82
Ilustración 18 Probabilidad de incumplimiento por nivel educativo.....	83
Ilustración 19 Probabilidad de incumplimiento por compra de cartera .....	83

## ÍNDICE DE TABLAS

Tabla 1 Variables independientes .....	46
Tabla 2 Distribución de los clientes que pagan y no pagan .....	48
Tabla 3 Distribución de la Cartera .....	49
Tabla 4 Departamento.....	51
Tabla 5 Días de atraso .....	52
Tabla 6 Estado civil .....	54
Tabla 7 Nivel educativo .....	55
Tabla 8 Cuotas canceladas.....	56
Tabla 9 Cuotas en mora .....	58
Tabla 10 Tabla cruzada Departamento vs Logit.....	59
Tabla 11 Cruce de variables entre cartera y Logit.....	61
Tabla 12 Cruce de variables entre nivel educativo y Logit .....	63
Tabla 13 Cruce de variables entre Estado civil y Logit.....	64
Tabla 14 Prueba T.....	67
Tabla 15 Resumen de procesamiento de casos .....	69
Tabla 16 Codificación de variable dependiente.....	70
Tabla 17 Tabla de clasificación .....	70
Tabla 18 Variables que no están en la ecuación.....	70
Tabla 19 Prueba ómnibus de coeficientes de modelo.....	71
Tabla 20 Resumen del modelo.....	72
Tabla 21 Prueba de Hosmer y Lemeshow .....	72
Tabla 22 Tabla de clasificación .....	73
Tabla 23 Variables en la ecuación .....	74
Tabla 24 Área bajo la curva.....	77
Tabla 25 Cartera año y medio .....	78
Tabla 26 Cartera de un año.....	78
Tabla 27 Cartera de 6 meses .....	79
Tabla 28 Prueba Backtesting .....	79

## RESUMEN

Para el desarrollo del presente trabajo, se tomó información de la base de datos de una cartera de microcrédito, perteneciente a una entidad financiera colombiana, dicha base contiene información de obligaciones desde Julio del 2016 hasta diciembre de 2017. Al tomar esta información se busca determinar la Probabilidad de incumplimiento (PI) de mencionada cartera, apoyados en una estimación econométrica, de forma particular, un modelo de Regresión Logística Binaria. Para encontrar esta PI se tuvieron en cuenta 15 variables predictores y una variable dependiente dicotómica que diferencia entre aquellos clientes que pagan y no pagan, llamada para el estudio logit. Dentro de los primeros pasos para realizar este estudio, se realizó un análisis descriptivo de las variables, para comprender las relaciones entre las predictores y la variable dependiente u objetivo. Seguido se corrió el modelo de regresión logística bajo el método “hacia adelante de Wald” con el programa IBM SPSS; con los datos de un año, es decir de julio de 2016 a junio de 2017, se diseñó el modelo y luego se aplicó una prueba Backtesting entre julio y diciembre de 2017. Se debe tener en cuenta el siguiente comentario: La población objeto de estudio tenía la característica, eran microcréditos castigados, en este sentido el trabajo presentó un reto adicional al tener una población con alta probabilidad de no pagar; pero de acuerdo a la estrategia de recuperación diseñada por la entidad financiera, se observan recuperaciones menores, pero que llevan a evaluar el fenómeno desde una muestra estratificada.

Al utilizar la muestra, aplicando el modelo obtenido con la misma, se obtuvo un porcentaje de acierto en la clasificación del pago e impago de 88,8% y un área bajo la curva (AUROC) del 92,2%, queriendo decir que el modelo posee un aceptable nivel predictivo. Por último, se realizó la prueba backtesting al modelo con las personas que no habían pagado en ese año para predecir su comportamiento de pago en los restantes seis meses, esto dio un buen resultado de clasificación, adicional a eso se realizó un análisis descriptivo de los resultados, que arrojó el

modelo generando así parámetros para el riesgo de crédito, por tanto, se puede asumir que el modelo logit es de gran utilidad para la entidad, al disminuir el riesgo de crédito y aumentar la utilidad de la misma.

**Palabras clave:** regresión logística, modelo logit, credit scoring, default, backtesting, microcrédito.

## INTRODUCCIÓN

El sistema financiero es muy importante para el desarrollo y crecimiento de una economía, por tanto, se debe estar atento a su buen funcionamiento y administración por parte de las entidades financieras, puesto que existen diversos tipos de riesgo financiero, como es, el de liquidez, crédito, operativo, legal y de mercado; entre todos ellos el enfoque principal de este trabajo está en el riesgo de crédito, más precisamente en entidades con cartera de microcrédito. La importancia de este sector radica en, que permite la concesión de financiamiento a una mayor cantidad de personas, debido a requerimientos más flexibles para la obtención del crédito, convirtiéndose en una herramienta de inclusión social para los pequeños emprendedores, “el microcrédito constituye una de las herramientas más significativas para asignar capital...en estos negocios de reducida envergadura, refuerza su productividad y eficiencia, y contribuye a la elevación de los niveles de ingreso y empleo en los sectores de menores recursos” (Calixto Salazar & Casaverde Carranza, 2011, pág. 1).

“Después de las crisis financieras de 1999 y 2008, en Colombia y Estados Unidos respectivamente, se hace evidente que el sistema financiero debe ser regulado de tal manera que corrija las asimetrías de información, las cuales el mercado por sí solo no corrige” (Ochoa P, Galeano M, & Agudelo V, 2010, pág. 192). Entonces, debido a sucesos como los mencionados, que han dejado al descubierto la inestabilidad y la globalización de mercados financieros, se hizo evidente la necesidad de la regulación financiera, por tanto, se creó el Comité de Regulación Bancaria de Basilea, donde se dan lineamientos de regulación internacional que prontamente se convirtieron en pautas generales, seguidas por entidades financieras a nivel mundial. En este orden se creó Basilea I en 1988, para resaltar la importancia de mantener un capital mínimo con el que se cubran los capitales sujetos al riesgo de posibilidad de impago; en 2004 es aprobado el acuerdo conocido como Basilea II, donde se concluye que “las entidades financieras y de

microfinanzas (IMFs) supervisadas requieren adoptar procesos internos que sean capaces de medir el riesgo de crédito” (Cantón, Lara Rubio, & Camino Blasco, 2010, pág. 90), por tanto se vuelve importante disponer de modelos de medición (scorings), con el fin de discriminar a los clientes según su perfil de riesgo; además de esto la calidad de información y las grandes bases de datos se vuelven indispensables para cuantificar el riesgo.

Uno de los antecedentes más importantes en materia de modelos econométricos para la cuantificación del riesgo en microfinanzas, se encuentra en una investigación realizada en Bolivia por Schreiner (2002), para una entidad microfinanciera, en la que se buscaba comprobar si la aplicación de dichos modelos resultaba tan benéfica como en el caso de los créditos de consumo; de esta manera se concluyó lo siguiente: “es beneficioso pues resalta características de los individuos que influyen en el riesgo...opera como un filtro para los créditos que pasaron una evaluación primaria pero siguen teniendo un alto riesgo estimado” (Arenas Díaz, Boccardi Rodríguez, & Piñeyrúa Ibáñez, 2012).

En este sentido, el presente estudio tiene como finalidad determinar la probabilidad de incumplimiento y los factores que influyen en ella, en una cartera de microcrédito perteneciente a una entidad financiera del país, para clasificar los clientes como buenos o malos pagadores. Es importante resaltar que para el desarrollo del trabajo se tomó una base de datos cuya cartera de microcrédito se encuentra castigada, es decir, está conformada por créditos que la entidad considera de difícil cobro. Dicha cartera pertenece a una casa de cobranza con cobertura nacional. El cuerpo del trabajo está dividido en cinco capítulos, en el primero se expone el planteamiento del problema, donde se relatan las complicaciones que ha atravesado el sistema financiero debido al riesgo de crédito, con un caso internacional y otro nacional, se puntualizan las posibles causas del impago, sus consecuencias y se resalta la importancia de la cartera en una entidad financiera como su principal fuente de ingresos. En el capítulo dos se definen conceptos relevantes para la investigación como lo es el riesgo de crédito, la asimetría de la información, selección adversa y

riesgo moral, se señalan las diferentes interpretaciones que ha tenido el riesgo de crédito a través de las teorías económicas. Así mismo se hace referencia a los acuerdos de Basilea I, Basilea II y Basilea III, de manera conjunta con el sistema de Administración del Riesgo Crediticio (SARC) vigente en Colombia. También se explican los modelos de credit scoring más relevantes como son: Análisis discriminante, logit o regresión logística, redes neuronales y árboles de decisión, se destacan antecedentes del modelo scoring y se explica la fórmula de logit.

En el tercer capítulo se enseña la metodología empleada para la aplicación del modelo. En el capítulo cuarto se comparten los resultados del estudio: análisis descriptivo de la probabilidad de incumplimiento, cruce de variables, aplicación del modelo y prueba backtesting. Por último, se encuentra el capítulo quinto con las conclusiones y recomendaciones.

## CAPÍTULO I: PROBLEMA

### Planteamiento del problema

La (Bolsa de Valores de Colombia, 2014, pág. 21) define el sistema financiero como el conjunto de instituciones y operaciones, por medio de las que se transfieren recursos de un sector superavitario a uno deficitario (empresas que requieren financiación). Este sistema comprende el mercado bancario (intermediario) y el de valores (no intermediados). El mercado bancario está compuesto por instituciones que captan recursos del público, principalmente en forma de depósitos y su estructura está formada por establecimientos de crédito y demás entidades que realizan operaciones similares; dentro de la primera categoría mencionada se encuentran los bancos comerciales, corporaciones financieras y las compañías de financiamiento comercial (Bolsa de Valores de Colombia, 2014, pág. 22).

Sin embargo, el sistema financiero a atravesado por varias crisis a través de los años, un ejemplo de ello es la metodología para el cálculo de intereses en créditos hipotecarios, basado en la unidad de poder adquisitivo (UPAC) empleado en Colombia. De acuerdo al (Banco de la República, 2015) fue implementado por el presidente Misael Pastrana en el año 1972, con el objetivo de mantener el poder adquisitivo de la moneda e incentivar el crédito hipotecario para la construcción de vivienda; para ese efecto se cobraba un interés de acuerdo al porcentaje de la inflación. Se creía que este era el mejor sistema ya que incentivaba el ahorro y la obtención de créditos adecuados, pues los ingresos salariales se ajustaban a la inflación, del mismo modo que los intereses crediticios se ajustaban a ella. A pesar de esto, el sistema UPAC cambio su diseño original, “tras una reforma empezó a cambiar con la tasa DTF, la cual cambiaba según el comportamiento del sistema financiero” (Banco de la República, 2015); fue entonces en la década de los 90 que

el DTF subió en porcentajes muy altos que sobrepasaban el IPC. Pero ¿Por qué subieron los porcentajes? De acuerdo a (Urrutia & Namen, 2011, págs. 300-301)

“La burbuja de construcción saturó el mercado, fenómeno que en 1997 y 1998 coincide con la crisis económica y financiera de Asia y la moratoria de Rusia que produjo una reversión de los flujos de capital hacia los países emergentes, entre ellos Colombia, que se encontraba muy vulnerable debido al aumento de deuda pública y privada en ese periodo; la reversión en los flujos de capital aumentó las tasas de interés a razón de las expectativas de devaluación, reducción de liquidez y aumento del riesgo país. Al aumentar la tasa de interés con la crisis internacional en 1997-1998, la inclusión del DTF en el cálculo del UPAC aumentó las cuotas de amortización y el valor de las deudas hipotecarias al tiempo que cayeron los precios de la finca raíz”.

Como consecuencia de la situación mencionada, los créditos resultaron impagables para las familias, por tanto, se afectó todo el sistema financiero, especialmente las corporaciones de Ahorro y Vivienda puesto que surgieron altas carteras morosas.

Más adelante, en el ámbito internacional se presenta la crisis del 2008 en Estados Unidos, donde (Zurita González, Martínez Pérez, & Rodríguez Montoya, 2009, pág. 18), afirman que el preludio de esta situación estuvo dado por la burbuja de acciones de empresas tecnológicas y el ataque terrorista a las Torres Gemelas del 2001, quienes fueron precursoras de la política expansionista tomada para impulsar la economía. Por su parte (Machinea, 2009, pág. 42), afirma que obedeció a una política monetaria excesivamente expansiva de la Reserva Federal desde fines de 2001 hasta diciembre de 2004, que se manifestó en tasas de interés inferiores al 2%. Ello provocó que los bancos pudieran otorgar créditos hipotecarios “baratos”, generando una gran demanda de bienes inmobiliarios, que al mismo tiempo estimuló un incremento impresionante en el precio de estos (Pineda Salido, pág. 143). De esta forma, cuando las tasas llegaron a niveles mínimos, la corriente de

créditos alcanzó a las familias de menores ingresos y escaso o nulo historial crediticio (Osorio, 2008, pág. 2), dando lugar a las hipotecas subprime “créditos hipotecarios otorgados a prestatarios riesgosos” (Romo, 2009, pág. 956). Los bancos también idearon formas para poder prestar a los clientes, un ejemplo de ello fue: “Préstamo de tasa ajustable (AMR) con opción de pago: esta es una hipoteca cuya tasa de interés no es fija, sino que cambia durante la vida útil del préstamo basándose en los movimientos de un índice de interés” (Pineda Salido, pág. 144). Lo cual indica que al incrementar las tasas se incrementara la deuda también.

Más adelante, debido a la guerra de Irak que redujo la oferta de petróleo y la disminución en la producción de alimentos causada por los subsidios a la producción de biocombustibles, los precios de la canasta familiar subieron, entonces los bancos centrales elevaron la tasa de interés del 1% (2004) al 5,25% (2007) para contener la inflación. El aumento de tasas hipotecarias junto a la subida de precios en alimentos y petróleo, hizo impagables las hipotecas (Ferrari, 2008, pág. 58); como consecuencia el valor de los inmuebles comenzó a bajar, las familias perdieron poder adquisitivo y entonces se incrementó la morosidad; pero había un problema adicional ya que las entidades financieras ante las bajas tasas de interés (menor rentabilidad), agruparon las hipotecas en paquetes para luego venderlas en el sistema financiero mundial y para superar la baja calificación de calidad de crédito de los bonos, estos eran agrupados “...entre riesgo alto, medio y bajo” (Osorio, 2008, pág. 3). Finalmente, la inflación de precios, sumada a la de los combustibles, el incremento de las tasas de interés y la pérdida de valor de las viviendas hizo impagables las hipotecas, “...lo que llevo al derrumbe de los derivados financieros” entendido como un activo financiero cuyo valor cambia o se deriva de los cambios en otro activo llamado activo subyacente, (Ferrari, 2008, pág. 62) (hipotecas subprime vendidas por el mundo). De esta manera, se dejó en evidencia a las calificadoras de riesgo, quienes “...tenían conflictos de interés en la calificación de derivados emitidos por los bancos que pagaban la calificación” (Ferrari, 2008, pág. 62), puesto que subestimaron la probabilidad de impago. También condujo a la

iliquidez y a la insolvencia de los bancos; por tanto muchos quebraron, fueron absorbidos o se convirtieron en bancos comerciales (Ferrari, 2008, págs. 58-59 ). Un ejemplo es la liquidación del banco de inversión Lehman Brothers el 15 de septiembre del 2008; se calcula que como consecuencia las pérdidas del sistema financiero alcanzaron los 500.000 millones de dólares (Machinea, 2009, pág. 46).

Los eventos anteriormente descritos dejan al descubierto la insuficiente regulación por parte de los entes de control del sistema financiero, la necesidad de ejercer control y seguimiento por parte de las entidades financieras para el otorgamiento de créditos y el mantenimiento de provisiones que puedan respaldar y/o anticiparse a las pérdidas y la posibilidad de deterioro en la capacidad de pago de los clientes. La afirmación reciente es apoyada por (Hernandez, 2004, pág. 2) quien dice que una entidad financiera debe identificar, monitorear, medir y controlar el riesgo. Del mismo modo la (Superintendencia Financiera de Colombia, 1995) en la circular externa 100 expone que “las entidades vigiladas deben evaluar permanentemente el riesgo incorporado en sus activos crediticios, tanto en el momento de otorgar créditos como a lo largo de la vida de los mismos, incluidos los casos de reestructuraciones”.

Así como las crisis mencionadas, el sistema financiero ha presentado muchas más; que han dado pie, para la creación de instituciones y/o acuerdos sobre la prevención y control del riesgo de crédito, tal como el Comité de Supervisión Bancaria de Basilea que fue creado en 1973 tras la caída del sistema de tasas fijas en materia de intereses, la internacionalización de los mercados financieros y la insolvencia de los bancos Bankhaus Herstatt y Franklin National Bank (Ustáriz González, 2003, pág. 435). Este comité define el impago según los requisitos de capital así:

“Se considera impago cuando uno o más de los eventos siguientes haya tenido lugar:

- Se determina que es improbable que el deudor cumpla por completo con las obligaciones contraídas (principal, intereses u honorarios).
- Un suceso de pérdida asociada al riesgo de crédito asociado a cualquiera de las obligaciones del deudor, tales como la dotación de una provisión específica, o la reestructuración en dificultades implicando la condonación o el aplazamiento del principal, los intereses, o los honorarios.
- El deudor se retrasó más de noventa días en cualquiera de sus obligaciones crediticias.
- El deudor ha solicitado quiebra o una protección similar de sus acreedores (Trigo Martínez, 2009, pág. 70)”.

Pero ¿Cuáles son las causas de que dichos eventos sucedan? Las causas de impago son variadas, entre ellas se puede nombrar algunas variables macroeconómicas que lo dificultan; (Fernández Castaño & Pérez Ramírez, 2005, pág. 64) nombran la tasa de desempleo, explicando que a mayor empleo, menor morosidad; la inflación; PIB y el DTF (menor DTF, menor morosidad). Del mismo modo se encuentra el riesgo moral y la selección adversa, refiriéndose el primero al cambio de motivaciones que puede tener el deudor que lo pueden llevar a preferir no pagar; mientras que la selección adversa se refiere al cliente riesgoso del que no se tiene información y por tanto no se puede elegir correctamente quién representa un buen o mal crédito (Delvasto, 2006, pág. 221). En este sentido se encuentra la ley 1380 del 2010 del régimen de insolvencia, que representa una amenaza para estas entidades pues en ella se pueden acoger las personas deudoras; tiene por objeto permitirle al deudor (persona natural no comerciante), acogerse a un procedimiento legal, que, mediante un trámite de negociación de deudas en audiencia de conciliación extrajudicial se celebre un acuerdo de pago con sus acreedores y cumplir así con sus obligaciones. Además, afirma que para que el deudor se declare insolvente, debe estar en cesación de pagos y esto se da cuando

“implica el pago de dos o más obligaciones a favor de dos o más acreedores, por más de 90 días o cursen en su contra una o más demandas de ejecución o de jurisdicción, exigiendo el pago de alguna de sus obligaciones” (Confederación colombiana de consumidores, 2010).

También es significativo mencionar que la cartera de créditos es el activo más importante de las Entidades de Intermediación Financiera (EIF), debido a que constituye la principal fuente generadora de ingresos, por lo que las operaciones de crédito deben realizarse de acuerdo a estrategias, políticas y procedimientos establecidos por cada EIF (Autoridad De Supervisión Del Sistema Financiero, 2010, pág. 2).

Dada la relevancia de la cartera es significativo considerar que el impago de los clientes y el no tener un sistema de administración del riesgo adecuado, puede traer consecuencias negativas, tanto para la institución como para el sistema financiero en general, como ocurrió en las crisis nombradas (UPAC y Crisis del 2008). (Sagner T, 2012, pág. 28) Señala al riesgo de crédito como una de las principales causas de las crisis bancarias, tanto de carácter sistémico como individual, además se pueden generar pérdidas fiscales elevadas que afectan negativamente el desempeño del sector financiero y su estabilidad. Agregado a esto (Nancy Eugenia Zamudio Gómez , 2007) Afirma que: “Cuando las entidades no cuentan con buenos sistemas de administración y manejo del riesgo, dejan como consecuencia incrementos en el incumplimiento por parte de los clientes, deterioro en la calidad de la cartera y reducción de las utilidades de los bancos”. En el caso particular de la casa de cobranza en la que se realiza este estudio, se puede evidenciar el problema de impago por parte de los clientes de una manera mucho más severa y resaltante, ya que, de una cartera de alrededor 18.000 créditos pertenecientes a un año, solo se paga el 9% de ellos. De allí surge la necesidad de estas entidades por conocer de manera anticipada qué valor podrán recuperar de cada compra de cartera.

Por lo tanto, las entidades deben “evaluar permanentemente el riesgo incorporado en sus activos crediticios, tanto en el momento de otorgar créditos como a lo largo de la vida de los mismos, incluidos los casos de reestructuraciones. Para tal efecto, las entidades deben adoptar un SARC” [Sistema de administración del Riesgo de Crédito] (Superintendencia Financiera de Colombia, 1995, pág. 2). Dentro de los elementos que componen el sistema se encuentran los modelos internos que “deben determinar la pérdida esperada por la exposición al riesgo de crédito (RC)” (Superintendencia Financiera de Colombia, 1995, pág. 7) mediante la medición de la probabilidad de incumplimiento, entonces, se hace pertinente el uso de un modelo como el Logit probit para establecer dicho indicador. Estos modelos son de mucha ayuda para contrarrestar el riesgo de crédito ya que “permiten optimizar la evaluación del desempeño, el rendimiento sobre el capital ajustado por riesgo, las decisiones relativas a la operación, el análisis de rentabilidad así como la toma de decisiones respecto a la estructura de capital” (Helizondo, 2012, pág. 43).

Tomando en cuenta la importancia de la cartera en una entidad financiera al ser esta la principal fuente generadora de ingresos y las consecuencias negativas que puede contraer la falta de control y seguimiento de los créditos, tanto para la entidad en sí misma como para el sistema financiero en general, se considera importante determinar: ¿Cuál es la probabilidad de incumplimiento de una cartera de microcrédito en una entidad del sistema financiero colombiano, en el año 2017?

### **Justificación**

(Crespo G, 2011, pág. 28) Afirma que las entidades financieras deben identificar y medir con precisión los riesgos para posteriormente controlarlos y evaluarlos de forma apropiada, por lo que la administración financiera del riesgo se ha convertido en una herramienta esencial para la supervivencia de estos negocios. Debido a la alta responsabilidad que asumen dichas entidades, al captar y colocar recursos

económicos, estas se preocupan por llevar el control del riesgo de crédito, para tener una certera precisión de su actividad.

En este sentido, se plantea en el presente trabajo de investigación, suministrar una base teórica, la cual argumentará acerca del mejor modelo Credit Scoring, con el fin de calcular la probabilidad de incumplimiento o impago de los créditos (riesgo crediticio), para una casa de cobranza con una cartera nacional. (Salazar Villano, 2013, pág. 417) Afirma en su artículo de investigación que:

“Existe un punto de encuentro teórico-práctico que hace del riesgo crediticio una temática relevante de estudio y más aún en un contexto municipal en Colombia en donde, a pesar de la creciente participación del sector financiero en la actividad productiva, no se ha abordado en ningún documento de investigación”.

Bajo dicha afirmación esta propuesta de trabajo cobra importancia ya que aporta una base teórica y brinda apoyo a las entidades financieras que deseen identificar, los principales componentes a tener en cuenta en el cálculo del riesgo de crédito de cartera. La creación e implementación de modelos internos para el cálculo del riesgo de crédito en las entidades financieras, conlleva ciertos beneficios expresados por (Gutiérrez Girault, 2007, pág. 5) “son más eficientes a la vez que sus predicciones más objetivas y consistentes, por lo que pueden analizar y tomar decisiones sobre una gran cantidad de solicitudes de crédito en poco tiempo y a un bajo costo”. De este modo se puede decir que, prestan un apoyo importante al analista financiero en el momento de identificar el riesgo, para de esta forma tomar decisiones más acertadas a menor costo y en poco tiempo. Es importante resaltar también la transición de la NIC 39 a la NIIF 9 sobre instrumentos financieros, de acuerdo al (Instituto Nacional de Contadores Públicos , 2015) en Colombia, el modelo de “perdida incurrida” (NIC 39) perdió credibilidad a raíz de la crisis financiera del 2008, ya que el conocimiento de la pérdida presente no era suficiente, se necesitaba

conocer la “pérdida esperada” ya que de esta manera se podrían predecir las pérdidas futuras de forma temprana y provisionarlas. Por lo anterior cobra relevancia un modelo interno como el logit, ya que este le permite a la entidad acercarse a las cifras de probabilidad de incumplimiento de cada cliente y poder mitigar o prevenirse ante futuros defaults; en el caso particular de la casa de cobranza le va a permitir pronosticar cuánto dinero va a poder recuperar de dicha cartera y de esta forma dimensionar su rentabilidad. Adicional a eso, (Támara Ayús, Aristizábal, & Velásquez, 2012, pág. 107) añade que seleccionar un modelo que permita predecir el riesgo crediticio, y reduzca la provisión de la pérdida esperada sin necesidad de aumentar el riesgo, aumenta la rentabilidad a través de la liberación de fondos que pueden ser colocados en el mercado financiero. En ese orden de ideas la implementación de dicho modelo busca contrarrestar el riesgo y la optimización de los recursos.

También servirá como un referente adicional para quienes deseen implementar el modelo para la cuantificación del riesgo crediticio. Por otro lado, el desarrollo de esta investigación traerá beneficios en cuanto al trabajo de grado, que se debe presentar para finalizar la carrera de Finanzas y Negocios Internacionales. Adicional a eso, es pertinente rescatar que contraer posibles beneficios a la Corporación Universitaria Autónoma del Cauca, ya que esta tiene planeado hacer un artículo de investigación con dicha información, para ser postulado a una revista científica indexada; que en cuanto se publique, representará un gran reconocimiento tanto para la universidad, como para las personas involucradas.

## **Objetivos**

### **Objetivo general**

Establecer la probabilidad de incumplimiento de una cartera de microcrédito, en una entidad del sistema financiero colombiano en el año 2017.

### **Objetivos específicos**

- Desarrollar un análisis descriptivo de la cartera de microcrédito de la entidad financiera bancaria, para determinar sus características principales.
- Generar un modelo de regresión logística, para calcular la probabilidad de incumplimiento con el mejor grado de ajuste a las observaciones históricas.
- Conocer el desempeño del modelo en el comportamiento real del incumplimiento de la entidad, mediante pruebas de backtesting.
- Identificar parámetros para el riesgo de crédito, teniendo en cuenta la probabilidad de incumplimiento obtenida.

## CAPÍTULO II: MARCO TEÓRICO Y REFERENTES CONCEPTUALES

### Bases teóricas

En finanzas el riesgo es “la posibilidad de que se sufra un perjuicio de carácter económico, ocasionado por la incertidumbre en el comportamiento de variables económicas a futuro” (Avila Bustos, 2005, pág. 6). De acuerdo a la cartilla No.2 coordinada por el (Autorregulador del Mercado de Valores de Colombia, 2014, pág. 6) se encuentra clasificado en cuantificable y no cuantificable; dentro de los cuantificables se haya el riesgo de mercado, liquidez y crédito mientras que en los no cuantificables está el operacional, legal, reputacional y estratégico. Ahora, el enfoque de este trabajo va dirigido al riesgo de crédito que se define como “la posibilidad de que una entidad incurra en pérdidas y se disminuya el valor de sus activos, como consecuencia de que un deudor o contraparte incumpla sus obligaciones” (Superintendencia Financiera de Colombia, 1995, pág. 2). De igual modo, en este tipo de riesgo se considera muy importante la información, todos los negocios requieren de ella pero de forma especial los establecimientos de crédito, y esto se debe a la naturaleza propia de estos agentes, en el que transfieren recursos de un sector superavitario a uno deficitario; en el manejo de dicha información se presenta un problema denominado “asimetría de la información” (Delvasto, 2006, pág. 218). Esta se muestra cuando los participantes del negocio bancario no cuentan con información o es insuficiente, impidiendo que se tomen decisiones acertadas y por ello se incurra en costos de transacción (Delvasto, 2006, pág. 220). Estas fallas traen los denominados efectos de selección adversa y riesgo moral; la selección adversa, según (Rodríguez Sánchez, 2001, pág. 51) ocurre “cuando debido a la asimetría de información no logran diferenciar un buen proyecto de uno malo” por tanto, la falta de información puede llevar a optar por la opción equivocada. El riesgo moral por otro lado, surge una vez desembolsados o entregados los recursos, en aquel momento hay cambio de incentivos en el

prestatario, por ello podrá tomar diferentes actitudes dependiendo de cómo sea su comportamiento con respecto al riesgo (Delvasto, 2006, pág. 221), es decir que la persona puede decidir no pagar.

Desde el siglo XVIII el riesgo de crédito ha tenido diferentes interpretaciones y se hace necesario nombrar algunas teorías que se enfocaban en ello a partir de la disciplina económica. Desde el punto de vista de la teoría neoclásica, en el estudio de (Salazar Villano, 2013, pág. 417) se expresa que no existe el riesgo de crédito pues no deberían haber pérdidas si se asume el supuesto de competencia perfecta, las pérdidas serían generadas por la falta de flexibilidad en su estructura de costos, "...por otro lado, si se habla de un escenario de competencia no perfecta, las pérdidas bancarias surgirían de las estrategias de juego de las firmas (bancos)". De acuerdo a Adam Smith (1776), citado por (Salazar Villano, 2013, pág. 417), "el tipo de interés corriente más bajo debe ser algo más que suficiente para compensar las pérdidas ocasionales a las que los préstamos están expuestos". En la teoría general de Keynes (1936), mediante el concepto de "eficiencia marginal del capital" se explica el riesgo de crédito de acuerdo a (Salazar Villano, 2013, pág. 418), teniendo en cuenta las expectativas frente a ingresos futuros, también da lugar al azar moral, es decir que el prestatario no posea la voluntad de pagar o el incumplimiento involuntario que dejaría en evidencia el fallo de las expectativas.

Por otro lado, de la perspectiva neoinstitucionalista se rescata la teoría de la información de Stigler (1961), quien la define como un "bien económico" que lleva inmersos costos de producción que en el caso de las instituciones bancarias se ven reflejados en todos los métodos usados, para llevar al cliente la información sobre la adquisición del crédito, "...en este sentido todo desajuste de la información generaría riesgo en el mercado de crédito" (Salazar Villano, 2013, pág. 418). Para concluir esta teoría, el neoinstitucionalismo considera que las entidades financieras

son canalizadoras de riesgo que deben determinar qué cliente se considera un buen crédito; (Salazar Villano, 2013, pág. 419) señala que para ello, se debe tener en cuenta dos tipos de costos: los “ex ante” que son aquellos en que se incurre al realizar el estudio de crédito; y los ex post, referentes a los costos dados por las transacciones realizadas que divergen las especificaciones originales del acuerdo.

Una vez vistas algunas teorías económicas que estudian el riesgo de crédito, se hace necesario conocer las instituciones y/o acuerdos más relevantes, que realizan sugerencias a todas las entidades crediticias para el manejo y control del riesgo, para tal efecto se tiene en cuenta lo siguiente. El Banco Internacional de Pagos ha sido una institución que, desde los años treinta, por medio de comités conformados por expertos internacionales, formula recomendaciones en temas financieros y bancarios, uno de esos comités es el Comité de Supervisión Bancaria de Basilea fundado en 1974, cuyo objetivo es formular estándares de supervisión y recomendaciones de práctica de buen gobierno en las entidades financieras (Pérez Ramirez & Fernández Castaño, 2007, pág. 79). El comité de supervisión bancaria de Basilea consciente de que las instituciones se enfrentan al riesgo de crédito, hace público el Acuerdo de Capitales de Basilea en 1988, el cual es conocido como Basilea I, en donde se hacen las recomendaciones necesarias, dada la importancia de asegurar la estabilidad del sistema y mantener un capital mínimo con el que se cubran los capitales sujetos al riesgo de posibilidad de impago, el cual se llamó capital mínimo regulatorio (Ochoa P, Galeano M, & Agudelo V, 2010, pág. 195), sin embargo, Basilea I no tomaba en cuenta diferentes niveles de calidad del activo, diferentes calidades del pagador ni diversificación de la cartera, en ese sentido se crea Basilea II donde se pretende entre otros, tener en cuenta los puntos mencionados y generar fortalezas frente a la medición e identificación del riesgo. Basilea II fue hecho público en 2004 y obtuvo una versión más completa en 2006; este acuerdo recomienda la gestión del sistema financiero basado en tres pilares fundamentales:

1. Requisitos de capital mínimo: cubrimiento de capital en riesgo
2. Proceso de examen supervisor: donde el ente supervisor cumple un papel primordial en la vigilancia y supervisión de la administración por parte de las entidades financieras.
3. La disciplina de mercado: acceso y transparencia de la información suministrada por las entidades financieras. (Ochoa P, Galeano M, & Agudelo V, 2010, pág. 195)

Pero de estos el más destacable para fines del estudio es “Requisitos de capital mínimo”, ya que establece cómo definir los recursos propios del banco en función del riesgo de crédito, con el objetivo de generar provisiones ante el deterioro de la cartera; para este fin se tienen dos opciones: la primera es realizar evaluaciones por parte de entidades externas y la segunda en la que se hace más énfasis, son los indicadores internos (Salazar Villano, 2013, pág. 419). En el requerimiento de provisiones también se defienden los modelos de seguimiento, en donde se pueden encontrar las matrices de transición donde básicamente, se permite determinar que créditos (personas) se están deteriorando y por tal motivo se deben provisionar más; “La aplicación Creditmetrics de JP Morgan fue desarrollada en 1997 y utiliza las matrices de transición para medir el riesgo de crédito. Para nuestro caso se define  $p_{ij}$  como la probabilidad de que un deudor con calificación crediticia  $i$  pueda “migrar” o moverse a otra calificación crediticia  $j$  en un horizonte de tiempo dado” (Támara Ayús, Aristizábal, & Velásquez, 2012, pág. 110). En el año 2010 se planteó la normatividad más reciente del Acuerdo Basilea III, donde se exige:

capital de calidad (al menos la mitad del capital debe ser nivel 1), requerimiento mínimo de capital de nivel 1 del 6%, mantener un “colchón de conservación” del 2,5% del capital ordinario y un colchón de capital anticíclico calibrado entre 0 y 2,5% de los recursos propios; esto último con el fin de que las entidades posean una reserva en la fase alta del ciclo económico que permita contrarrestar las pérdidas en periodos de tensión (Caruana, 2010, págs. 2-4).

Pasando del ámbito internacional al nacional, en Colombia la Superintendencia Financiera de Colombia (SFC) mediante la Carta Circular 31 y la Circular Externa 11 de 2002, exige que todas las instituciones financieras implementen un sistema de Administración del Riesgo Crediticio (SARC); teniendo en cuenta la volatilidad de las variables financieras gracias a la globalización del mercado. De acuerdo a (Ochoa P., Galeano M., & Agudelo V., 2010, pág. 195) el Riesgo crediticio (SARC) se define así:

Un conjunto de políticas, procedimientos, metodologías, herramientas informáticas y capital físico y humano de las entidades financieras, dirigidas hacia la adquisición de conocimientos, medición y control de riesgos crediticios, en los que incurre una entidad dentro del giro normal de su negocio.

La circular externa 100 de 1995, capítulo 2 de la (Superintendencia Financiera de Colombia, 1995, pág. 3) contempla los siguientes elementos básicos del SARC “Políticas de administración del RC, procesos de administración del RC, modelos internos o de referencia para la estimación o cuantificación de pérdidas esperadas, sistema de provisiones para cubrir el RC, procesos de control interno”.

Dentro de la definición anterior se resalta la especificación de modelos internos o de referencia para la estimación o cuantificación de pérdidas esperadas, en el caso del SARC enfocados al riesgo individual que en este asunto es importante contemplar ya que el estudio va dirigido al riesgo de crédito de cartera; y aunque “en el análisis de riesgo de crédito es en el que menos se han desarrollado metodologías para medirlo, debido a las dificultades para la medición e identificación de los factores que intervienen” (Támara Ayús, Aristizábal, & Velásquez, 2012, pág. 110), los avances que se han realizado relacionados a la estadística, la econometría y las finanzas han permitido crear diferentes modelos de credit scoring, estos modelos son denominados como “algoritmos que de manera automática evalúan el

riesgo de crédito de un solicitante de financiamiento o de alguien que ya es cliente de la entidad” (Gutiérrez Girault, 2007, pág. 3).

De acuerdo con (Gutiérrez Girault, 2007, pág. 1) la utilización de modelos de credit scoring para estimar la probabilidad de default de un cliente, empezó en los años 70 y se volvió popular en los 90's, debido a la necesidad de las entidades financieras de conservar un activo (cartera) optimo, acompañado de un nivel de rentabilidad bueno. En seguida se nombrarán algunos de los modelos más utilizados e importantes, a tener en cuenta a través de la historia: en primer lugar se encuentra el Análisis discriminante; que de acuerdo a (Cantón, Lara Rubio, & Camino Blasco, 2010, pág. 91) clasifica buenos y malos pagadores en el momento de realizar un crédito, por medio de la diferenciación de características que definen a cada grupo. Entre sus inconvenientes se encuentra la rigidez para cumplir las hipótesis de linealidad, normalidad, homocedasticidad e independencia. Los autores (Cantón, Lara Rubio, & Camino Blasco, 2010, pág. 92) afirman que, Altman (1968) desarrolló la metodología Z –Score que determinaba la probabilidad de quiebra empresarial, aplicando estas variables: ingresos netos/ventas, ganancias retenidas/activos, EBIT/activos, valor de mercado del patrimonio neto/valor libros de la deuda y ventas/activos.

En seguida se encuentra el modelo logit o de regresión logística, este permite calcular la probabilidad de que el individuo pertenezca a un grupo (cumplido o incumplido), esta clasificación se hace de acuerdo a las variables independientes que explican el resultado de “Y” (variable dependiente), en este modelo “...no es necesario plantear hipótesis de partida, como por ejemplo la normalidad de la distribución de las variables... presenta la ventaja de...mantener la variable explicada siempre dentro de un rango de variación entre cero y uno” (Cantón, Lara Rubio, & Camino Blasco, 2010, pág. 92). Por otro lado, se encuentra, el modelo de

Redes neuronales, que emula la estructura y el comportamiento del cerebro; la red está formada por procesadores simples, llamados nodos, que están interconectados entre sí. "...La finalidad de cada nodo es dar respuesta a una determinada señal de entrada..." sin embargo, "... el uso de esta técnica resulta complicado pues el proceso interno de aprendizaje funciona como caja negra, donde la comprensión de lo que ocurre dentro demanda de conocimientos especializados" (Cantón, Lara Rubio, & Camino Blasco, 2010, pág. 92). Por último, (Flórez López, 2007, pág. 77) establece que los modelos de árboles de decisión son de clasificación jerárquica y secuencial que dividen un conjunto de datos (N), cada árbol integra tres componentes básicos: los nodos de decisión, las ramas y los nodos terminales (hojas); el orden en que van surgiendo los nodos es debido a la importancia de las variables explicativas. Un punto en contra del modelo es que llevan una difícil comprensión interna (Cantón, Lara Rubio, & Camino Blasco, 2010, pág. 93).

Después de dar una breve introducción de los modelos de credit scoring más destacados, es importante resaltar que el modelo Logit es considerado como el más adecuado y superior en cuanto a algunas ventajas que aportan sus características para el análisis del riesgo de crédito, frente a los otros. Esta idea es corroborada por (Támara Ayús, Aristizábal, & Velásquez, 2012, pág. 107) quienes plantean que Lennox compara la capacidad de predicción de los modelos probit y logit frente al análisis discriminante, para demostrar que los dos primeros poseen mayor capacidad predictiva en cuanto a la probabilidad de incumplimiento. Del mismo modo, (Salazar Villano, 2013, pág. 421) afirma que el enfoque logit se considera adecuado ya que se adapta a la información de fuentes primarias y secundarias disponibles en el quehacer bancario. Por otro lado, (Moscote Flórez & Rincón, 2012, pág. 127) resaltan las ventajas que posee el modelo, aseverando que "Es una función flexible y fácil de utilizar, tiene una interpretación relativamente sencilla, la evidencia empírica ha demostrado, que este modelo es adecuado en la mayoría de los casos en los cuales la respuesta es binaria".

## Antecedentes

Los avances que se han realizado en estadística, econometría y finanzas han permitido crear diferentes modelos de credit scoring, estos modelos son denominados como “algoritmos que de manera automática evalúan el riesgo de crédito de un solicitante de financiamiento o de alguien que ya es cliente de la entidad” (Gutiérrez Girault, 2007, pág. 3).

De acuerdo con (Gutiérrez Girault, 2007, pág. 1) la utilización de modelos de credit scoring para estimar la probabilidad de default de un cliente, empezó en los años 70 y se volvió popular en los 90's, debido a la necesidad de las entidades financieras de conservar un activo (cartera) optimo.

Para empezar en 1941 Durand “utilizo características del solicitante para generar un índice de calificación que permitiese definir si otorgar o negar un crédito de consumo” (Salazar Villano, 2013, pág. 5), por otra parte en 1968 Altman “aplico el modelo Z-score<sup>15</sup> en créditos comerciales; sin embargo tal enfoque sigue siendo empleado para medir el riesgo en créditos a persona natural” (Salazar Villano, 2013). Después en 1970, Orgler utilizó un análisis de regresión lineal para créditos comerciales y sólo revisó los créditos vigentes (Moreno Valencia , 2013, pág. 15). Se debe agregar que en 1971, el mismo autor empleo un enfoque de regresión para evaluar los créditos de consumo especiales, consigo trajo como resultado que la información no incluida en el formulario de solicitud tenía una mejor capacidad de predicción que la información proporcionada inicialmente por el solicitante, al evaluar la calidad de los préstamos futuros (Moreno Valencia , 2013, pág. 15). Por otra parte, En 1979, Chandler y Coffman “hicieron un análisis comparativo entre el scoring de crédito y el juicio de un evaluador de créditos, donde ambos sistemas indicaban, en promedio, cuál de los dos grupos en que se clasificaban los individuos, tenía mejor comportamiento” (Moreno Valencia , 2013, pág. 15). Así mismo en 1980 Ohlson expresa que “el modelo logit se convirtió en una metodología estándar para pronosticar la probabilidad de quiebra o el estado de Default de los créditos”; en

este mismo año Wiginton “examinó la eficacia de los modelos Logit encontrando un mejor poder de predicción y eficientes procesos de sensibilidad y estimación de los análisis discriminantes” ( Rodríguez Guevara & Trespalacios Carrasquilla, pág. 2). Así mismo Jhon C.

“crea un estudio de contrastación de los modelos logísticos frente a los análisis discriminantes, en el cual, demuestra que el ajuste de un modelo Logit tiene mejores repercusiones en la predicción por su mejor comprensión de los resultados binomiales, mejores ajustes en las variables, y opciones de medir sensibilidad y puntos de corte más precisos” ( Rodríguez Guevara & Trespalacios Carrasquilla, pág. 4).

En 1984 (Adrich & Nelson, 1984) examinan tres técnicas (modelos de probabilidad lineal, probit y logit) que son adecuadas para variables dependientes dicotómicas, usando ejemplos detallados, señalan las diferencias entre los modelos lineales, logit y probit. Es preciso decir también que Altman “publica un documento sobre el proceso de los préstamos, su naturaleza, y los criterios de evaluación de los créditos con procesos logísticos” ( Rodríguez Guevara & Trespalacios Carrasquilla, pág. 4). Para los noventa algunos autores como, Glorflod (1990) “estudió un modelo de discriminación basado en Least - Absolute – Value, en la cual buscaba demostrar una alternativa del modelo discriminatorio clásico” ( Rodríguez Guevara & Trespalacios Carrasquilla, pág. 4), dos años después Crook et al “comparó los modelos discriminatorios clásicos con los modelos de performance de crédito de características multivariadas” ( Rodríguez Guevara & Trespalacios Carrasquilla, pág. 4) y en 1999 Zhai y Russell “busco una alternativa mediante redes neuronales para explicar los factores de riesgo de clientes” ( Rodríguez Guevara & Trespalacios Carrasquilla, págs. 4-5) .

Más tarde Jones y Hensher ( 2004 ) Y Hensher ( 2007 ) “Se encuentran en contextos ordinales y multinomial y se centran en las proporciones de miembros del grupo de predicción en lugar de probabilidades individuales” ( Kukuk & Ronnberg, 2012, pág. 2). Adicionalmente Abdou (2009) y Greene (1998) manifiestan que “utilizaron

análisis discriminante y regresiones logística y probit, para investigar la eficiencia, en términos de la clasificación correcta de los deudores (buenos y malos), de los modelos tipo scoring para créditos de consumo y comercial, respectivamente” (Moreno Valencia , 2013, pág. 15). En el 2009 Kocenda y Vojtek “comparan un modelo Logit con un Modelo CART” ( Rodríguez Guevara & Trespalcios Carrasquilla, pág. 5) . Para el 2010 Lara “realizó una comparación de las técnicas de credit scoring clasificándolas en modelos paramétricos y no paramétricos” ( Laguna Tubón, 2015, pág. 24) el cual tiene en cuenta análisis discriminante, modelos de probabilidad lineal, modelos de logit y probit, programación lineal, los árboles de decisión y las redes neurales están enmarcadas dentro de los métodos no paramétricos ( Laguna Tubón, 2015, pág. 24). Para el año 2010 (Cantón, Lara Rubio, & Camino Blasco, 2010, pág. 115) emplearon un modelo de regresión logística destinado a una entidad de microfinanzas, donde pudieron identificar que la incorporación de variables explicativas macroeconómicas mejoran significativamente los resultados del modelo. Ochoa et al (2010) y Moreno (2014) “compararon los modelos Logit con Probit, Redes neuronales y Logit Mixtos, llegando a un resultado similar de los modelos logísticos para este tipo de estudios” ( Rodríguez Guevara & Trespalcios Carrasquilla, pág. 5). Sin embargo y a pesar de todos los estudios mencionados, en el campo de los microcréditos se ha podido evidenciar que hay pocas investigaciones sobre credit scoring; en este sentido en el estudio desarrollado por (Cantón, Lara Rubio, & Camino Blasco, 2010, pág. 114) se concluye que a través de la bibliografía consultada sobre credit scoring, para instituciones de microfinanzas se permite comprobar que existen pocos trabajos sobre el tema.

El modelo logit o de regresión logística permite calcular la probabilidad de que el individuo pertenezca a un grupo (cumplido o incumplido), esta clasificación se hace de acuerdo a las variables independientes que explican el resultado de “Y” (variable dependiente), en este modelo “...no es necesario plantear hipótesis de partida, como por ejemplo la normalidad de la distribución de las variables... presenta la

ventaja de...mantener la variable explicada siempre dentro de un rango de variación entre cero y uno” (Cantón, Lara Rubio, & Camino Blasco, 2010, pág. 92). Lo que distingue la regresión logística de los modelos de regresión lineal, es fundamentalmente, que la variable de respuesta es dicotómica, es decir, toma uno de dos valores posibles. “Esta diferencia entre ambos modelos, se refleja tanto en la selección de los parámetros del modelo, como en los supuestos para su construcción” (Llaugel & Fernández, 2011, pág. 598).

La forma específica del modelo de regresión logística es:

*Ecuación 1 Formula logit*

$$\pi(x) = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}} = \frac{e^k}{1 + e^k} \quad \text{donde } k = \beta_0 + \beta_1 x$$

Fuente: (Llaugel & Fernández, 2011, pág. 599)

“El modelo logit permite además de estimar la probabilidad de incumplimiento, identificar los factores de riesgo que determinan las probabilidades, así como el peso relativo de ellos sobre las probabilidades” (Echeverri Valdes, 2006, pág. 83)

En el análisis de regresión se reflejan las siguientes premisas, cuando se cuenta con una variable de respuesta dicotómica:

- a) La media condicional de la ecuación de regresión debe ser formulada para estar contenida entre 0 y 1.
- b) La distribución binomial y no la normal, describe la distribución de los errores y será la distribución en la que estará sustentado el análisis.
- c) Los principios que guían un análisis de regresión lineal, son los mismos para el análisis de regresión logística. (Llaugel & Fernández, 2011, pág. 600)

Para el modelo se deben filtrar las variables de modo que queden solamente las que tengan un verdadero grado de significancia, (Velandia Velandia, 2013, pág. 56) afirma que para ello “es preciso calcular la estadística de Wald, para lo cual se debe observar la escala de medición de la variable ya que esta puede ser categórica o no categórica”. Para las variables no categóricas se cuenta con la siguiente formula:

*Ecuación 2 Formula Wald para variables numéricas*

$$Wald_i = \frac{\hat{\beta}_i^2}{\hat{\sigma}_{\beta_i}^2}$$

Que tiene distribución chi cuadrado con un grado de libertad

Fuente: (Velandia Velandia, 2013, pág. 56)

Si es categórica:

*Ecuación 3 Formula Wald para variables categóricas*

$$Wald_i = \hat{\beta}_i^T \mathbf{C}^{-1} \hat{\beta}_i$$

Que sigue una distribución chi cuadrado, con grados de libertad igual al número de parámetros estimados

Fuente: (Velandia Velandia, 2013, pág. 56)

“La importancia del estadístico de Wald radica en que a través de él se puede determinar la significancia del parámetro en el modelo y el peso que tiene” (Echeverri Valdes, 2006, pág. 45). Esto se obtiene con la siguiente hipótesis:  $H_0: \beta_i = 0$  significa que la variable  $X_i$  no es importante para establecer variaciones en el modelo, mientras  $H_1: \beta_i \neq 0$  significa que sí lo es. También, de acuerdo al valor se da el peso de la variable, con este criterio:

95% Poco significativa

97.5% significativa

99% muy significativa

99.9% altamente significativa

Por otro lado, es importante resaltar que el modelo de regresión logística presenta varias fortalezas frente a otros modelos, dichas fortalezas son evidenciadas por diferentes estudios: Para empezar ( Kukuk & Ronnberg, 2012, pág. 2) expresa que “modelos logit se han convertido en la metodología prevalente para académicos y profesionales para pronosticar las probabilidades de quiebra o de crédito por defecto”. desde luego, el mismo autor elaboro una comparación entre la aplicación del modelo logit tradicional y el logit mixto para un modelo de default de crédito empresarial, el cual llega a la conclusión que el modelo logit mixto parece tener propiedades deseables desde el punto de vista cualitativo y cuantitativo, dado que produjo mejores pronósticos de las probabilidades individuales de Default (Moreno Valencia , 2013, pág. 13), para ( Rodríguez Guevara & Trespalacios Carrasquilla):

“los modelos Logit como modelos no paramétricos discretos y binomiales, acatan a cabalidad los supuestos matemáticos de relación binomial, los cuales son más adecuados que los modelos discriminatorios clásicos basados en modelos lineales probabilísticos, siendo más ajustados, sencillos de realizar y fáciles de interpretar que los modelos de regresión lineal probabilísticos, modelos Probit, CART, Logit Mixtos (que son similares en su método pero más complejos en su tratamiento), modelos híbridos, redes neuronales, máquinas de vectores de soporte entre otros” ( Rodríguez Guevara & Trespalacios Carrasquilla, págs. 8-9).

Del mismo modo, (Llaugel & Fernández, 2011, pág. 598) resaltan que los modelos de regresión logística permiten unos supuestos menos restrictivos e introducir como independientes en el modelo a variables categóricas, en este sentido Cox y Snell (1989) recomiendan la distribución logística, porque “desde un punto de vista matemático, es extremadamente flexible y fácil de usar, y segundo, se llegan a conclusiones fácilmente interpretables” (Llaugel & Fernández, 2011, pág. 598). Mientras tanto el modelo lineal presenta grandes falencias puesto que exhibe inconvenientes referentes a la heterocedasticidad, ya que la variable dependiente

solo debe tomar uno de dos valores (1 o 0) y ello puede producir varianzas negativas, provocando que los resultados caigan por fuera del rango (0,1), generando estimaciones ineficientes. En cambio, el modelo logit supera dichos problemas mediante el uso de “funciones de distribución, clasificando los datos y asociándolos con las variables en diferentes grupos” (Echeverri Valdes, 2006, pág. 82).

El modelo discriminante también demuestra falencias, para la determinación del incumplimiento en el pago de la cartera ya que presenta “la necesidad de incluir en su cálculo todos los elementos de una población, además de considerar en el proceso de cálculo variables de tipo cuantitativo” (Echeverri Valdes, 2006, pág. 83); esto enseña una gran falencia para identificar el riesgo de cartera, ya que en él influyen bastantes variables cualitativas importantes.

### CAPÍTULO III: METODOLOGÍA

La presente investigación se llevó a cabo bajo un enfoque cuantitativo, el cual “se basó en la recopilación de datos o de información, que permitió comprobar hipótesis de forma numérica; (Bunge, 2004, pág. 194) define hipótesis como “hecho no sujeto hasta ahora a experiencia o en general no es sometido a la misma”. Logrando descubrir, explicar y predecir los fenómenos (causalidad) y para generar políticas relacionadas con los resultados (Hernández Sampieri, Fernández Collado, & Baptista Lucio, 2010, pág. 11 y 16). Así como también afirma (Bunge, 2004, págs. 14-15) “las matemáticas sirven para arbitrar experimentos, para llevar a cabo mediciones y la contrición de instrumentos para registrar y elaborar datos con procedimientos exactos que garantiza un informe preciso”. De modo que a través de la recolección de datos y de análisis numéricos se pretende llegar al objetivo propuesto.

Además, se realizó un estudio exploratorio ya que “permite al investigador familiarizarse con el fenómeno que se investiga. Es el punto de partida para la formulación de otras investigaciones con mayor nivel de profundidad” (Méndez Álvarez, 2011, pág. 227). La investigación se desarrolló bajo el uso de fuentes secundarias ya que es “información escrita que ha sido recopilada y transcrita por personas que han recibido tal información a través de otras fuentes escritas o por un participante en un suceso” (Méndez Álbare , 2011, pág. 249), las cuales están relacionadas con modelos cuantitativos de probabilidad de incumplimiento, como son: revistas científicas, libros y documentos de páginas web oficiales .También se tomaran fuentes primarias que según (Méndez Álbare , 2011, pág. 255) “implica utilizar técnicas y procedimientos que suministren información”, en este caso se contó con el suministro de la base de datos por parte de la entidad financiera .

De igual manera, es de estudio descriptivo porque “identifica características del universo de investigación, señala formas de conducta, establece comportamientos y descubre y comprueba asociación de variables” (Méndez Álvarez, 2011, pág. 227). Se realizó una exploración y descripción de los datos encontrados en la base de datos de la entidad financiera (hallar la concentración de los créditos, altura de mora máxima, la mínima, promedio).

También es explicativo ya que estos estudios están “orientados a la comprensión de hipótesis causales... se orienta a comprobar hipótesis de tercer grado; esto es, identificación y análisis de las causales (variables independientes) y sus resultados, los que se expresan en hechos variables (dependientes)” (Méndez Álvarez, 2011, pág. 232). Este punto se desarrolló por medio de la relación de dependencia entre la probabilidad de incumplimiento en el pago (variable dependiente) y las variables independientes relacionadas con la cartera. Tuvo un diseño documental y de campo lo cual será clave en el alcance de los objetivos inicialmente propuestos.

Finalmente, es de método inductivo que según (Méndez Álvarez, 2011, pág. 236) es un “proceso de conocimiento que se inicia por la observación de fenómenos particulares con el propósito de llegar a conclusiones generales”, ya que se observó la composición de la base de datos. En el mismo sentido es de método deductivo, que de acuerdo a (Méndez Álbare , 2011, pág. 236) se refiere a la “observación de fenómenos generales con el propósito de señalar las verdades particulares”; puesto que al obtener información sobre la base se empieza a realizar correlaciones entre las variables.

## **Proceso**

### **a. Recopilación de información:**

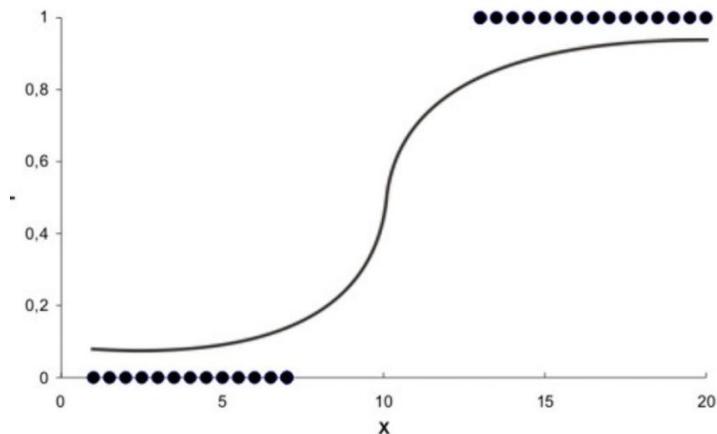
Fueron recopilados 18.753 créditos de la cartera, suministrados por la base de datos de la entidad financiera, correspondientes al periodo de julio de 2016 hasta diciembre de 2017. Es importante mencionar que estos créditos pertenecen a una cartera castigada, es decir, se encuentran catalogados como de difícil cobro.

## b. Tratamiento de la información:

El modelo logístico tiene forma de curva y para estimar el modelo se busca la curva que más se adecue a los datos reales. En los siguientes dos gráficos se puede evidenciar diferentes curvas de ajuste, la primera es de relación perfecta mientras la segunda tiene una relación pobre.

### Relación bien definida

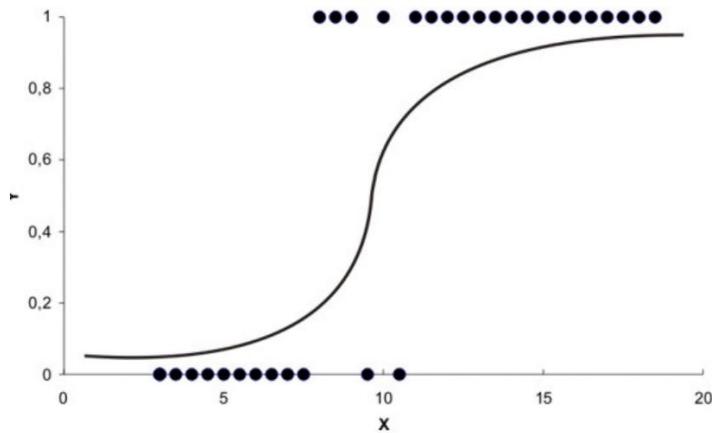
*Ilustración 1 Curva de ajuste relación perfecta*



Fuente: (Alderete, 2006, pág. 56)

### Relación pobremente ajustada

Ilustración 2 Curva de ajuste pobremente ajustada



Fuente: (Alderete, 2006, pág. 57)

Para evaluar la capacidad predictiva del modelo y su ajuste a la realidad, se realizan distintas medidas de desempeño, tales como:

- **La R cuadrado de Cox y Snell:** De acuerdo a (Alderete, 2006, pág. 58) en programas estadísticos como el SPSS se computa de la siguiente manera:

Dónde:  $-2LL$  (nulo) es la desviación del modelo nulo solo o con una constante, sin tener en cuenta variables predictoras,  $-2LL$  (modelo) es la desviación del modelo con las variables predictoras y  $N$  es el tamaño de la muestra.

Ecuación 4 R cuadrado de Cox y Snell

$$R^2_L = 1 - \left[ \frac{-2LL_{(nulo)}}{-2LL_{(modelo)}} \right]^{2/N}$$

- **R cuadrado Nagelkerke:** “versión corregida del R-cuadrado de Cox y Snell con valor máximo igual a 1” (Meza Saldaña, Reyes Cervantes, Pérez Salvador, & Tajonar Sanabria, 2017, pág. 4).

Ecuación 5 R cuadrado Nagelkerke

$$\bar{R}^2_L = \frac{1 - \left[ \frac{-2LL_{(nulo)}}{-2LL_{(modelo)}} \right]^{2/N}}{1 - (2LL_{(modelo)})^{2/N}}$$

Fuente: (Alderete, 2006, pág. 59)

**Bondad de ajuste de Hosmer-Lemeshow:** se construyen tablas para comparar los resultados de estimación del modelo, frente a los reales en la muestra. “Se basa en agrupar los casos en deciles de riesgo y comparar la probabilidad observada con la probabilidad esperada dentro de cada decil” (Velasco S, 2010, pág. 16). Esta medida de desempeño debería dar una cifra superior o igual a 0,05, sin embargo, de acuerdo al artículo de investigación de (Cantón, Lara Rubio, & Camino Blasco, 2010) aún si este indicador tiene un nivel de significancia bajo, “se debe tomar en cuenta que este dato estadístico solamente brinda pistas sobre la calidad del ajuste del modelo”, por tanto se puede decir que el modelo no pierde su validez.

- 
- **Área bajo la curva ROC (Receiver Operating Characteristic):** Es un gráfico en el que se observan los pares sensibilidad/especificidad resultante de la variación continua de los puntos de corte, en todo el rango de resultados observados. En el eje “y” se sitúa la sensibilidad (verdaderos positivos), en el eje “x” los falsos positivos (1-especificidad) (Meza Saldaña, Reyes Cervantes, Pérez Salvador, & Tajonar Sanabria, 2017, pág. 5).  
Cuando exista una proporción correcta de aciertos frente al Default, se dirá que el modelo tiene alta Sensibilidad; y cuando exista una proporción de aciertos correctos de No-default se dirá que el modo goza de alta especificidad (falsas alarmas), que es lo que se busca en un modelo adecuado o correcto.

“Estadísticamente, la predicción de los falsos positivos (1-especificidad) se conoce como Error Tipo I, el cual se define como el error de rechazar la hipótesis siendo ésta verdadera. La predicción de los falsos negativos (1-sensibilidad) es llamado el Error Tipo II, que se comete cuando se acepta la hipótesis nula siendo falsa” (Moreno Valencia , 2013, pág. 9).

### **c. Población y Muestra de la investigación:**

**Población:** 18.753 créditos.

La información de las personas de crédito fue tomada de una casa de cobranza perteneciente a Colombia. La base de datos está clasificada por compras de cartera, desde la primera hasta la cuarta compra, pertenecientes a julio de 2016 hasta diciembre de 2017, con una totalidad de 18.753 créditos.

**Muestra:** De la población se sacaron 4.501 créditos a través de la muestra probabilística estratificada que es aquel “muestreo en el que la población se divide en segmentos y se selecciona una muestra para cada segmento” (Hernández Sampieri, Fernández Collado, & Baptista, 2014, pág. 181), en este caso se desarrolló por medio del programa IBM SPSS, ya que al poseer una población desbalanceada, es decir, severamente sesgada en los clientes que no pagan, se hacía necesario tomar proporciones un poco similares entre los grupos de personas que pagaban y las que no; para la obtención de mejores resultados.

A la muestra se aplicó una prueba llamada prueba T de student. “Es una prueba poderosa, en la que aunque una de las muestras no tenga distribución normal pero la otra sí y la razón de la varianza más grande a la más pequeña sea  $< 2$ , resulta adecuada al comparar dos medias” (Sánchez Turcios, 2015, pág. 59). De esta forma se comprobó la validez de la muestra tomada, comparando la media de la población y la muestra, donde la diferencia de ambos resultados debe ser

pequeña para que se pueda decir que la muestra refleja a la población en sus resultados.

#### d. Variables del modelo

**Variable dependiente categórica dicotómica:** se denotan solo dos opciones como resultado, ya sea 1= buen cliente (no default) o 0=Mal cliente (Default).

#### Variables independientes del modelo:

*Tabla 1 Variables independientes*

Variable categórica	Variable numérica
Cartera	Fecha apertura
Departamento	Operación: número que identifica cada crédito.
Estado civil	Edad mora: días de atraso del crédito.
Nivel educativo	Valor inicial
	Saldo deuda
	Cuota mensual
	Total cuotas
	Cuotas canceladas
	Cuotas mora
	Oficina: número del sitio en que se realizó el trámite de la solicitud de crédito.
	AE: Actividad económica.

**e. Método ewma:** se utilizaron los “promedio móvil ponderado con suavizamiento exponencial”, por medio del método EWMA “Exponentially Weighted Moving Average...basado en un estadístico que asigna menos peso a medida que los datos son más antiguos” (Luzardo Briceño, 2006, pág. 45); para recoger los datos de los últimos años de cada variable (moda,

ingreso, altura de mora máxima-mínima...), es decir, se traen variables históricas con promedios que son agregadas a las variables de corte transversal; con el objetivo de hacer el modelo más robusto y preciso.

- f. Aplicación del modelo:** Una vez superada la etapa anterior, se ingresaron los datos a un modelo LOGIT bajo el método “Wald” que introduce las variables y dependiendo su significancia estadística las deja en el modelo o las retira. Después de esto, el modelo arroja los resultados, y se procede a realizar pruebas estadísticas relacionadas con el modelo para su evaluación. El LOGIT ubica al cliente entre 0 (no paga) o 1 (paga).
  
- g. Prueba del modelo mediante backtesting:** Para probar el modelo se realiza una prueba backtesting “para realizar un backtesting es necesario comparar el valor del riesgo observado con las pérdidas y/o ganancias reales. Lo que se mide es la eficiencia en el modelo, contando las observaciones de pérdidas y/o ganancias que fueron mayores al VaR” (De Lara Haro, 2005, pág. 155). La prueba es aplicada para el periodo de Julio a diciembre del 2017, con el fin de comprobar que tanto nivel de predicción tiene el modelo, es decir a cuantos créditos clasifico de forma correcta. La prueba, se desarrolló por medio de la formula logit para identificar la probabilidad de incumplimiento y el grupo de pertenencia de cada crédito (si incurre a pago o impago); y de esta manera comparar los resultados de la formula con los reales.

## CAPÍTULO IV: RESULTADOS

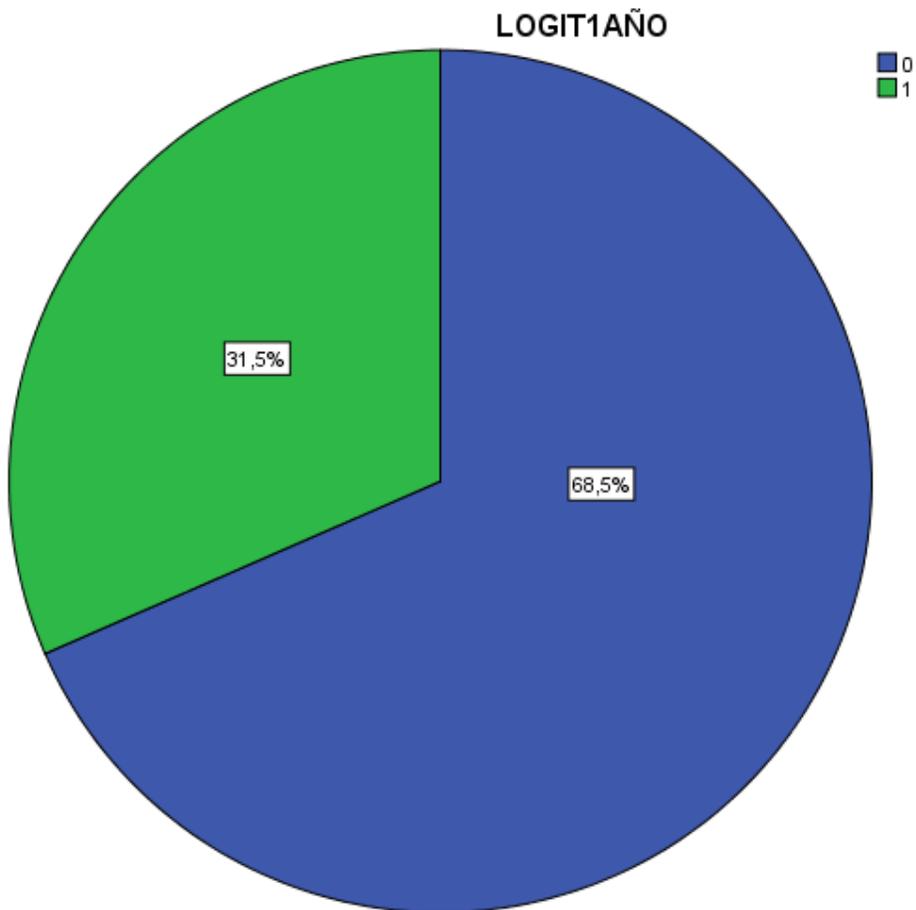
Para la aplicación se cuenta con una muestra de 4.501 clientes de microcréditos, concedidos entre julio de 2016 a junio de 2017 de una entidad financiera colombiana. La población contiene la variable respuesta (dependiente) DEFAULT y 15 variables independientes, que describen las características del cliente y su comportamiento crediticio que permiten explicar el grado de incumplimiento.

La variable dependiente indica si el cliente alcanzó el estado de morosidad o no. Es una variable dicotómica, donde los deudores que cayeron en incumplimiento se codificaron con el valor 0, mientras que los otros tomaron el valor de 1. En la siguiente tabla se evidencia el número de clientes que pagan, los cuales representan el 31,5% del total (tabla 2).

*Tabla 2 Distribución de los clientes que pagan y no pagan*

		LOGIT1AÑO			
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	0	3081	68,5	68,5	68,5
	1	1420	31,5	31,5	100,0
	Total	4501	100,0	100,0	

Ilustración 3 Distribución de clientes que pagan y no pagan



## Análisis descriptivo

A continuación, se presenta un análisis descriptivo de los 4.453 créditos pertenecientes a la entidad financiera colombiana.

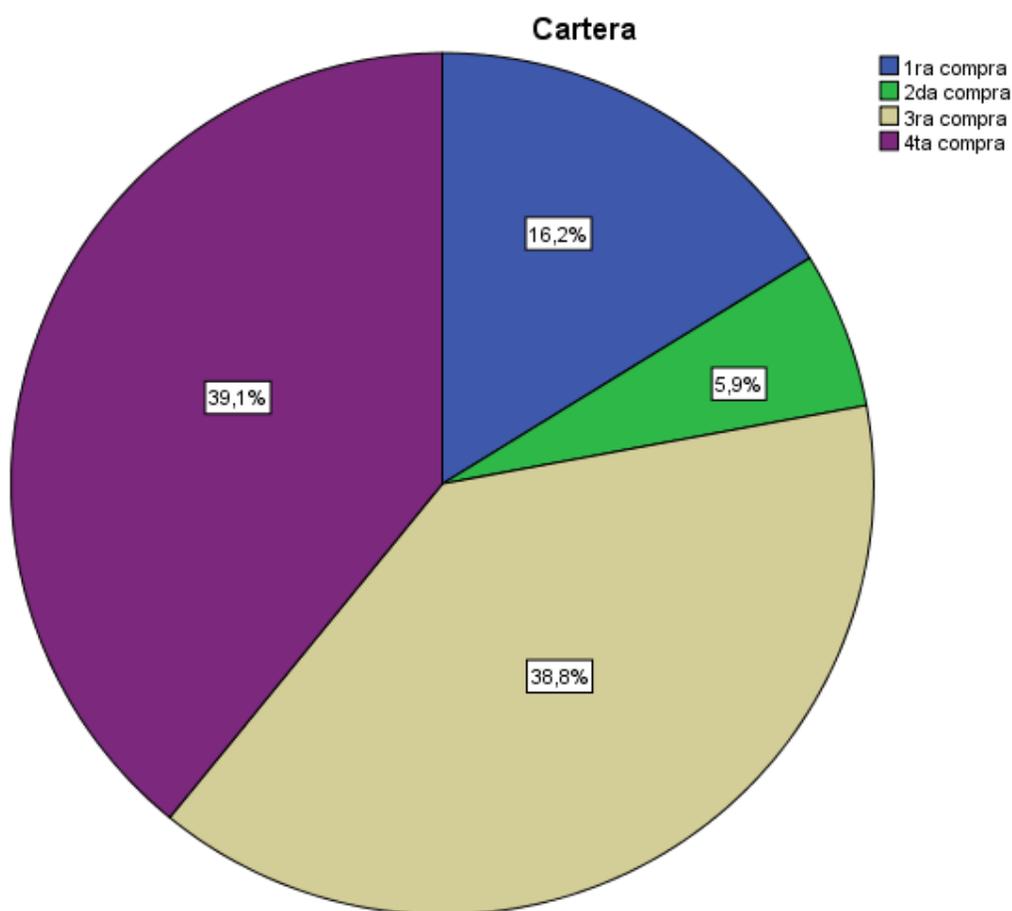
- **Variable Cartera**

Tabla 3 Distribución de la Cartera

**Cartera**

		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	1ra compra	730	16,2	16,2	16,2
	2da compra	264	5,9	5,9	22,1
	3ra compra	1747	38,8	38,8	60,9
	4ta compra	1760	39,1	39,1	100,0
	Total	4501	100,0	100,0	

Ilustración 4 Distribución de la cartera



La presente cartera se encuentra distribuida de la siguiente manera: la cuarta compra obtiene mayor representación con 1.760 créditos que equivalen a 39,1%, seguida de la tercera compra con 1.747 créditos equivalentes a 38,8%, en tercer lugar se encuentra la prima compra con 730 créditos que representan el 16,2%, por

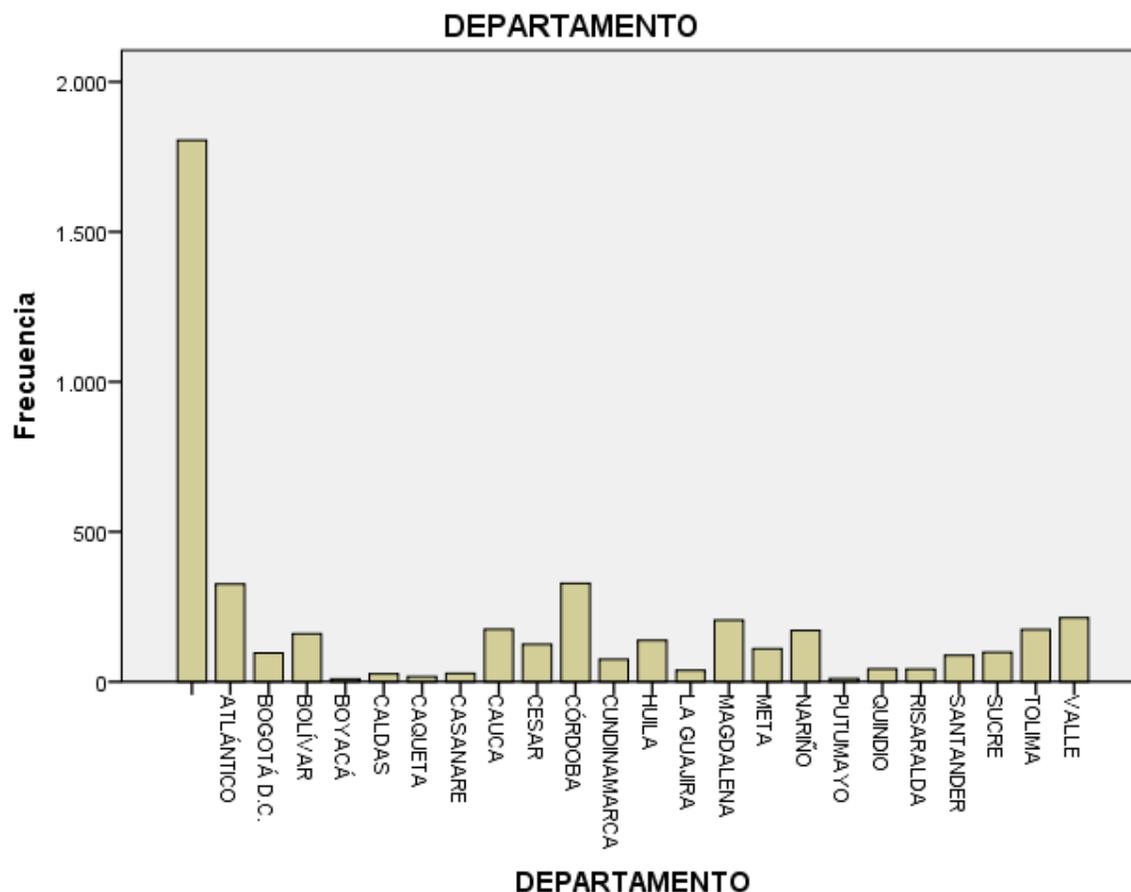
último la de menor representación es la segunda compra con un total de 264 créditos que equivalen a un 5,9%.

### Variable departamento

Tabla 4 Departamento

DEPARTAMENTO				
	Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	1805	40,1	40,1	40,1
<b>ATLÁNTICO</b>	<b>326</b>	<b>7,2</b>	7,2	47,3
BOGOTÁ D.C.	96	2,1	2,1	49,5
BOLÍVAR	160	3,6	3,6	53,0
<b>BOYACÁ</b>	<b>9</b>	<b>,2</b>	,2	53,2
CALDAS	26	,6	,6	53,8
<b>CAQUETA</b>	<b>17</b>	<b>,4</b>	,4	54,2
CASANARE	28	,6	,6	54,8
CAUCA	175	3,9	3,9	58,7
CESAR	125	2,8	2,8	61,5
<b>CÓRDOBA</b>	<b>328</b>	<b>7,3</b>	<b>7,3</b>	68,8
CUNDINAMARCA	75	1,7	1,7	70,4
HUILA	138	3,1	3,1	73,5
LA GUAJIRA	38	,8	,8	74,3
MAGDALENA	205	4,6	4,6	78,9
META	110	2,4	2,4	81,3
NARIÑO	171	3,8	3,8	85,1
<b>PUTUMAYO</b>	<b>11</b>	<b>,2</b>	,2	85,4
QUINDIO	43	1,0	1,0	86,3
RISARALDA	42	,9	,9	87,3
SANTANDER	88	2,0	2,0	89,2
SUCRE	98	2,2	2,2	91,4
TOLIMA	174	3,9	3,9	95,3
<b>VALLE</b>	<b>213</b>	<b>4,7</b>	4,7	100,0
Total	4501	100,0	100,0	

Ilustración 5 Departamento



En esta variable se puede identificar que los departamentos con mayor concentración de créditos en su respectivo orden son: Córdoba con 328 créditos, Atlántico con 326 y Valle con 213. Y los de menor concentración son Boyacá, Putumayo y Caquetá con 9, 11 y 17 créditos respectivamente.

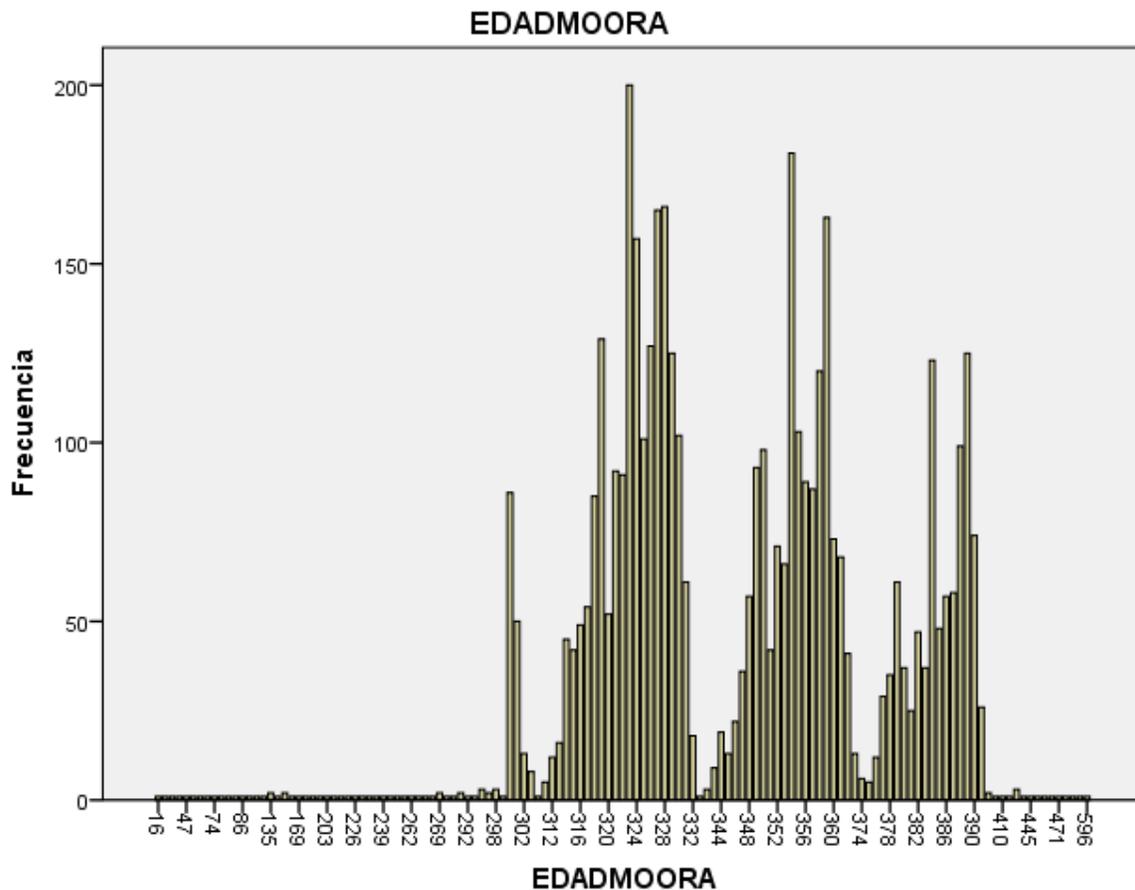
- **Variable Edad de Mora**

Tabla 5 Días de atraso

Estadísticos		
EDAD MORA		
N	Válido	4501
	Perdidos	0

Media	344,00
Moda	323
Mínimo	16
Máximo	596

Ilustración 6 Días de atraso



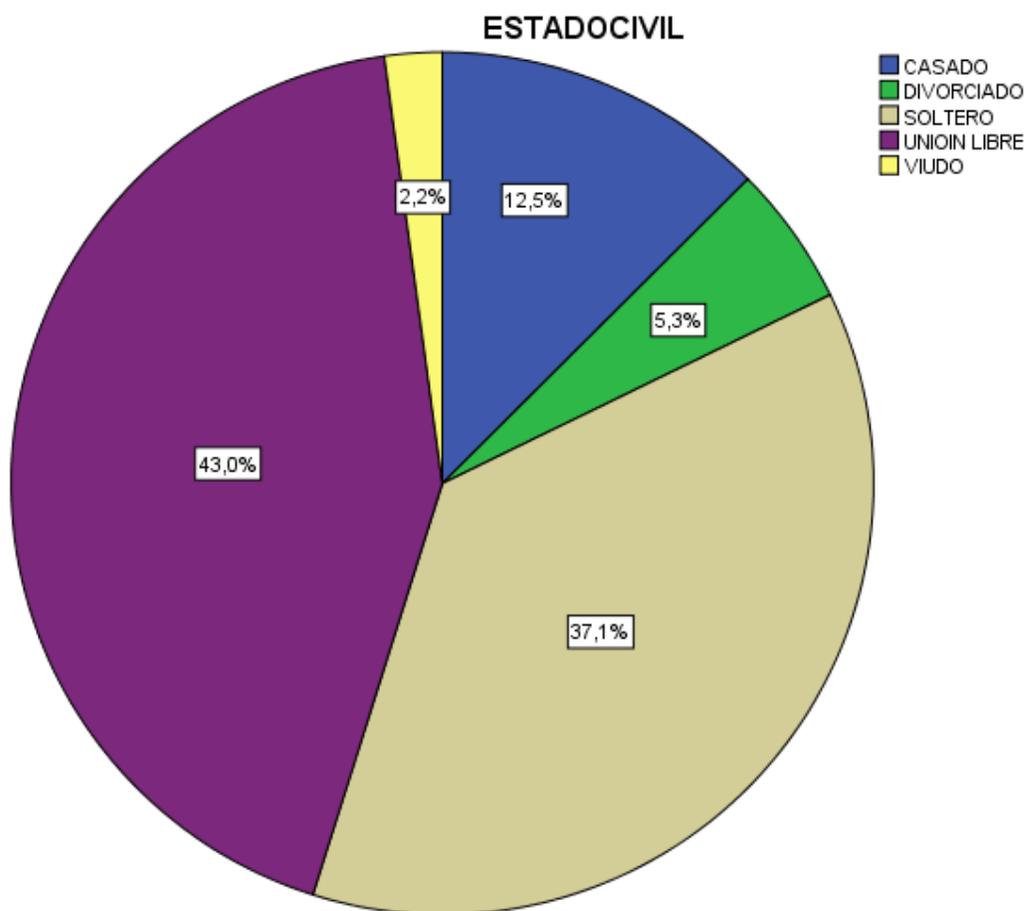
El promedio de días de atraso son 344 días, siendo 596 el máximo de días de atraso y 16 el mínimo. Por su parte, la cantidad de días de mora más recurrentes entre los clientes son 323.

- **Variable Estado Civil**

Tabla 6 Estado civil

		ESTADOCIVIL			
		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	CASADO	564	12,5	12,5	12,5
	DIVORCIADO	237	5,3	5,3	17,8
	SOLTERO	1668	37,1	37,1	54,9
	UNIOIN LIBRE	1935	43,0	43,0	97,8
	VIUDO	97	2,2	2,2	100,0
	Total	4501	100,0	100,0	

Ilustración 7 Estado Civil



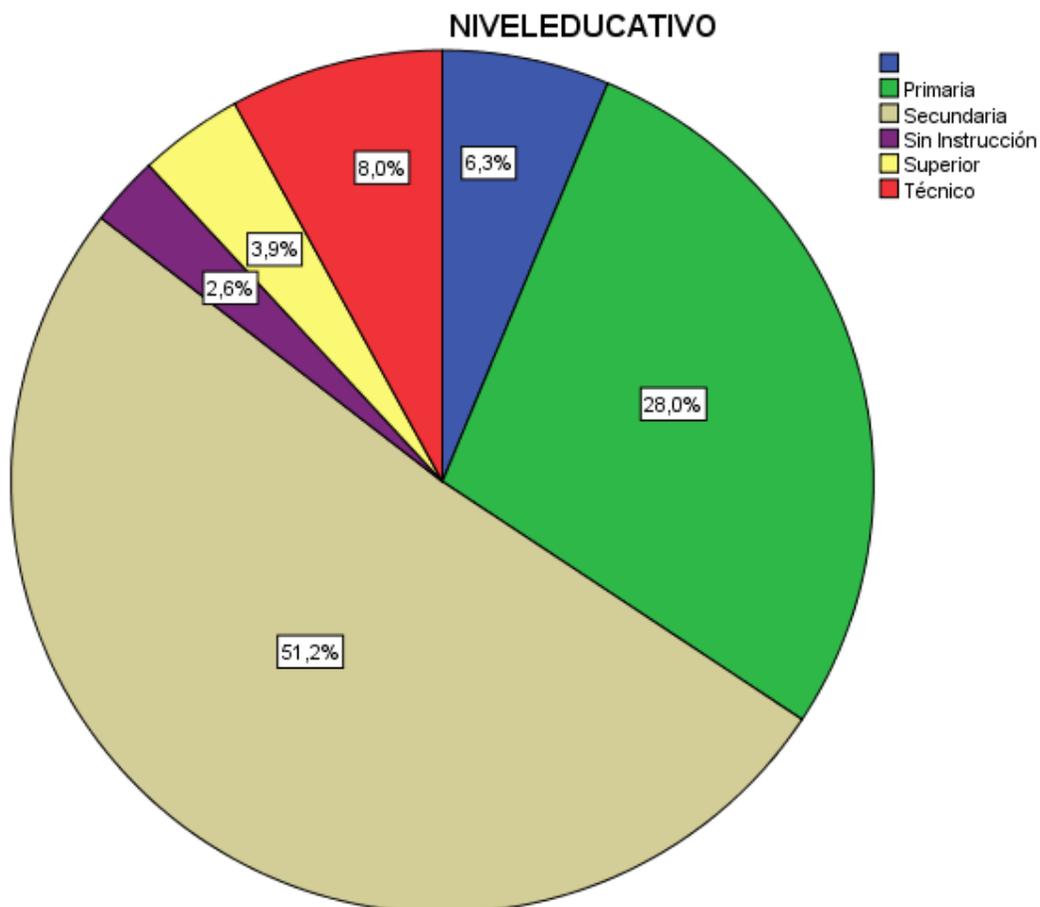
De la muestra las personas que solicitaron más créditos a la entidad bancaria son los de unión libre con 1935 clientes representando el 43%, seguido de los solteros con 1668 equivalentes al 37,1%, después los casados con 564 clientes que equivalen a un 12,5% y por último se encuentran los clientes con menor representación en la cartera que son los divorciados y los viudos con una participación del 5,3% y 2,2% respectivamente.

- **Variable Nivel Educativo**

Tabla 7 Nivel educativo

<b>NIVELEDUCATIVO</b>				
	Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
Válido	282	6,3	6,3	6,3
Primaria	1262	28,0	28,0	34,3
Secundaria	2303	51,2	51,2	85,5
Sin Instrucción	118	2,6	2,6	88,1
Superior	176	3,9	3,9	92,0
Técnico	360	8,0	8,0	100,0
Total	4501	100,0	100,0	

Ilustración 8 Nivel educativo



En cuanto al nivel educativo predominan los clientes con nivel de educación secundaria (2.303 clientes) con una representación del 51,2% del total de los créditos, seguidos de la población con educación primaria que reflejan el 28% con un total de 1262 clientes, por último, se encuentran los de menor participación como: los de educación técnica, superior y sin instrucción con una distribución de 8%, 3,9% y 2,6% respectivamente.

- **Variable cuotas canceladas**

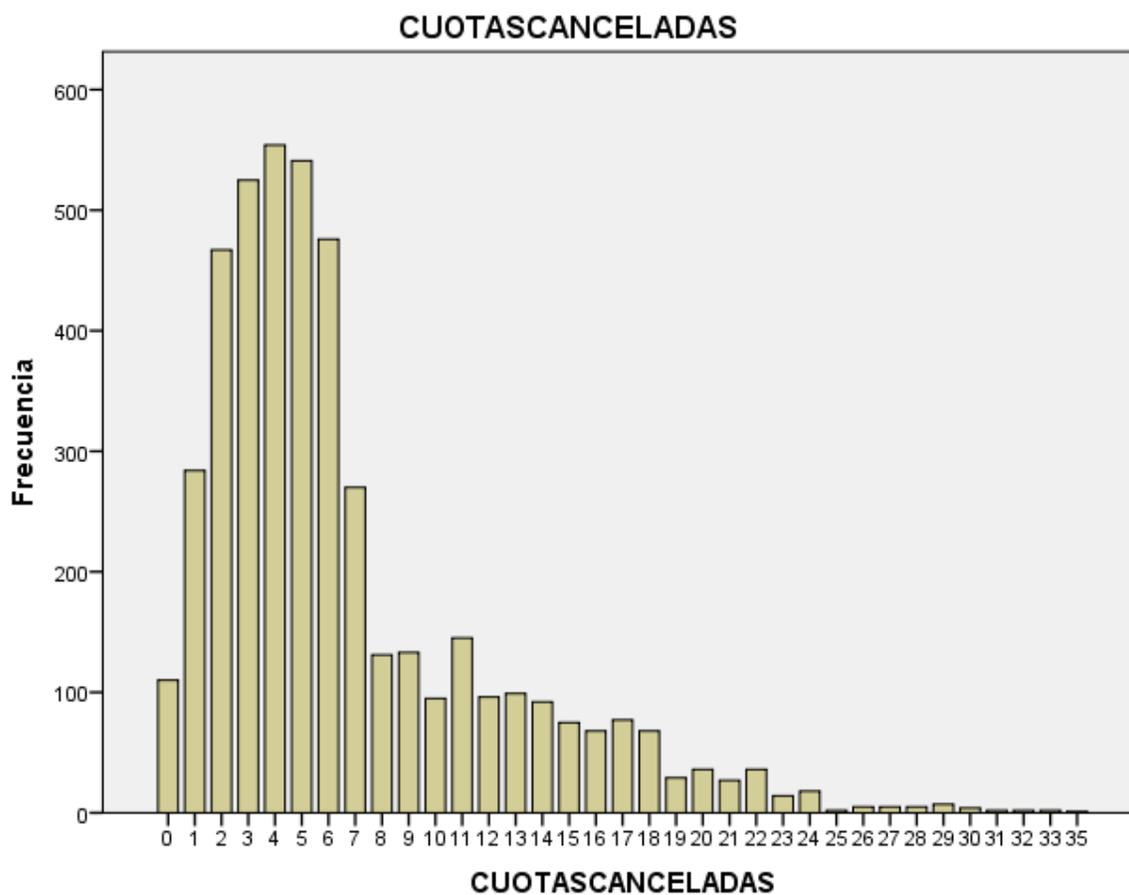
Tabla 8 Cuotas canceladas

### Estadísticos

## CUOTAS CANCELADAS

N	Válido	4501
	Perdidos	0
Media		6,69
Mediana		5,00
Moda		4
Mínimo		0
Máximo		35
Suma		30112

Ilustración 9 Cuotas canceladas



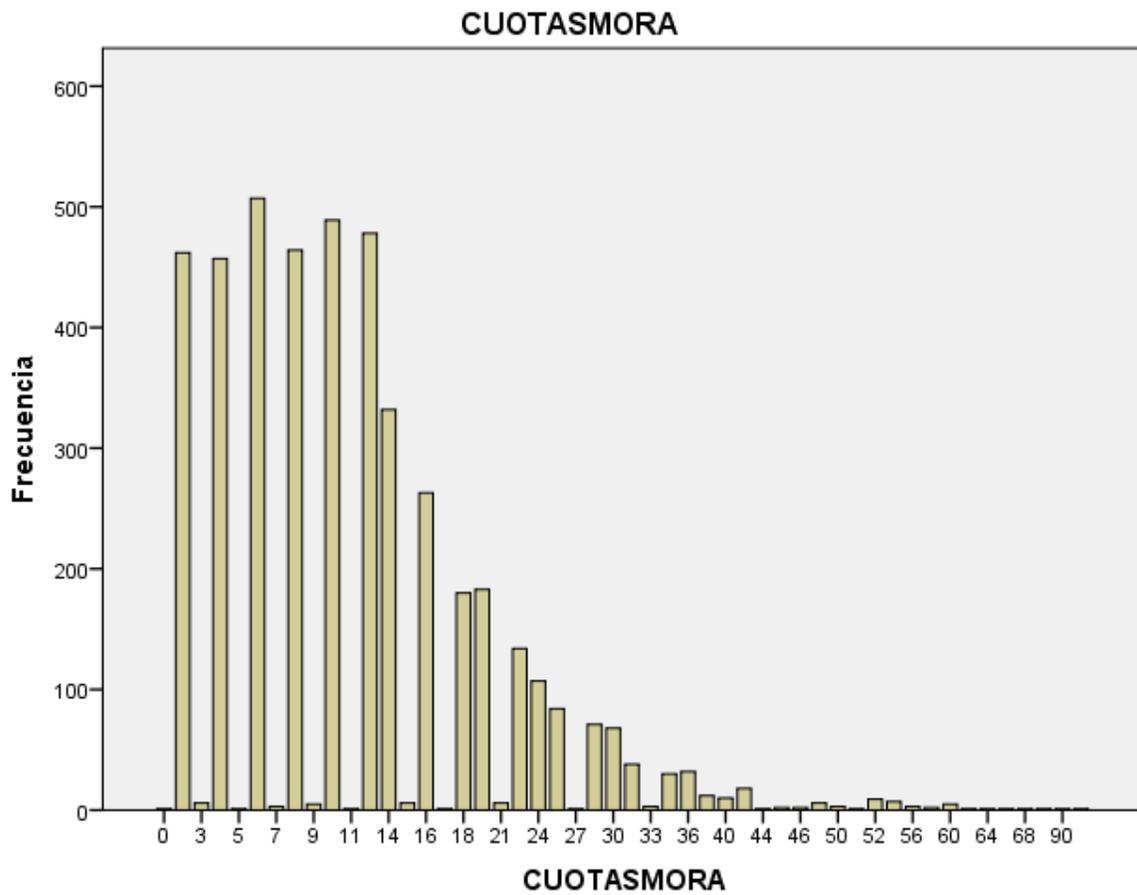
El número de cuotas canceladas más recurrentes es 4, el mínimo es 0 y el máximo es 35. Para un total de 30.112 cuotas canceladas.

- **Variable cuotas en mora**

Tabla 9 Cuotas en mora

<b>Estadísticos</b>		
CUOTASMORA		
N	Válido	4501
	Perdidos	0
Media		12,39
Mediana		10,00
Moda		6
Mínimo		0
Máximo		96
Suma		55754

Ilustración 10 Cuotas en mora



El número de cuotas en mora más recurrentes entre los clientes son 6, siendo 0 el mínimo de cuotas en mora y 96 el máximo. Para un total de 55.754 cuotas en mora y un promedio de 12.

## Cruce de variables

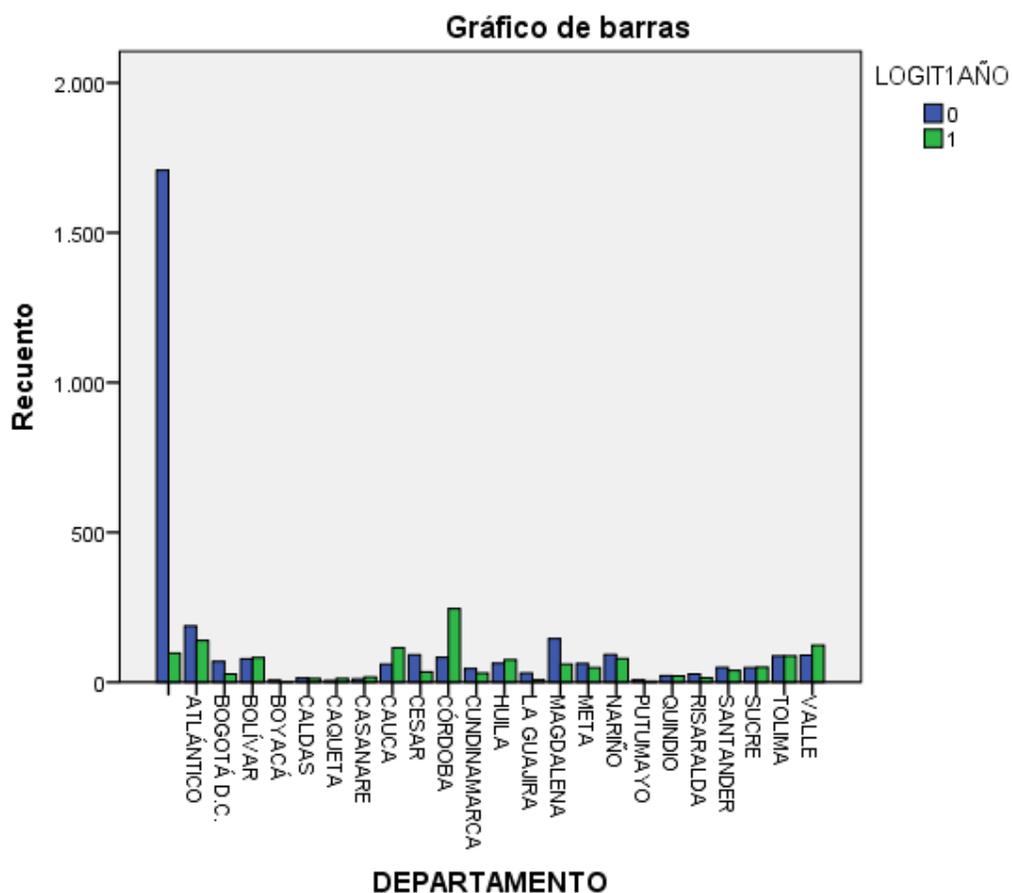
- **Cruce de variables departamento y logit**

Tabla 10 Tabla cruzada Departamento vs Logit

		Recuento		Total	Diferencia
		LOGIT1AÑO			
		0	1		
DEPARTAMENTO		<b>1708</b>	<b>97</b>	<b>1805</b>	<b>1611</b>
	<b>ATLÁNTICO</b>	<b>187</b>	<b>139</b>	<b>326</b>	<b>48</b>
	BOGOTÁ D.C.	69	27	96	42
	BOLÍVAR	78	82	160	-4
	BOYACÁ	7	2	9	5
	CALDAS	14	12	26	2
	CAQUETA	5	12	17	-7
	CASANARE	11	17	28	-6
	<b>CAUCA</b>	<b>60</b>	<b>115</b>	<b>175</b>	<b>-55</b>
	CESAR	91	34	125	57
	<b>CÓRDOBA</b>	<b>83</b>	<b>245</b>	<b>328</b>	<b>-162</b>
	CUNDINAMARCA	45	30	75	15
	HUILA	63	75	138	-12
	LA GUAJIRA	30	8	38	22
	<b>MAGDALENA</b>	<b>145</b>	<b>60</b>	<b>205</b>	<b>85</b>
	META	62	48	110	14
	NARIÑO	92	79	171	13
	PUTUMAYO	8	3	11	5
	QUINDIO	22	21	43	1

RISARALDA	27	15	42	12
SANTANDER	49	39	88	10
SUCRE	48	50	98	-2
TOLIMA	87	87	174	0
<b>VALLE</b>	<b>90</b>	<b>123</b>	<b>213</b>	<b>-33</b>
Total	3081	1420	4501	

Ilustración 11 Cruce de variables entre departamento y Logit



En el presente cruce de variables y para los siguientes cruces, se toma en cuenta la diferencia entre el número de impago y pagos por cada departamento, ya que entre mayor sea la diferencia positiva, mayor será el impago y para el caso de la diferencia negativa significa que, hay mayor cantidad de clientes que incurren a pago, por lo tanto, a mayor numero negativo mayor será la participación no default.

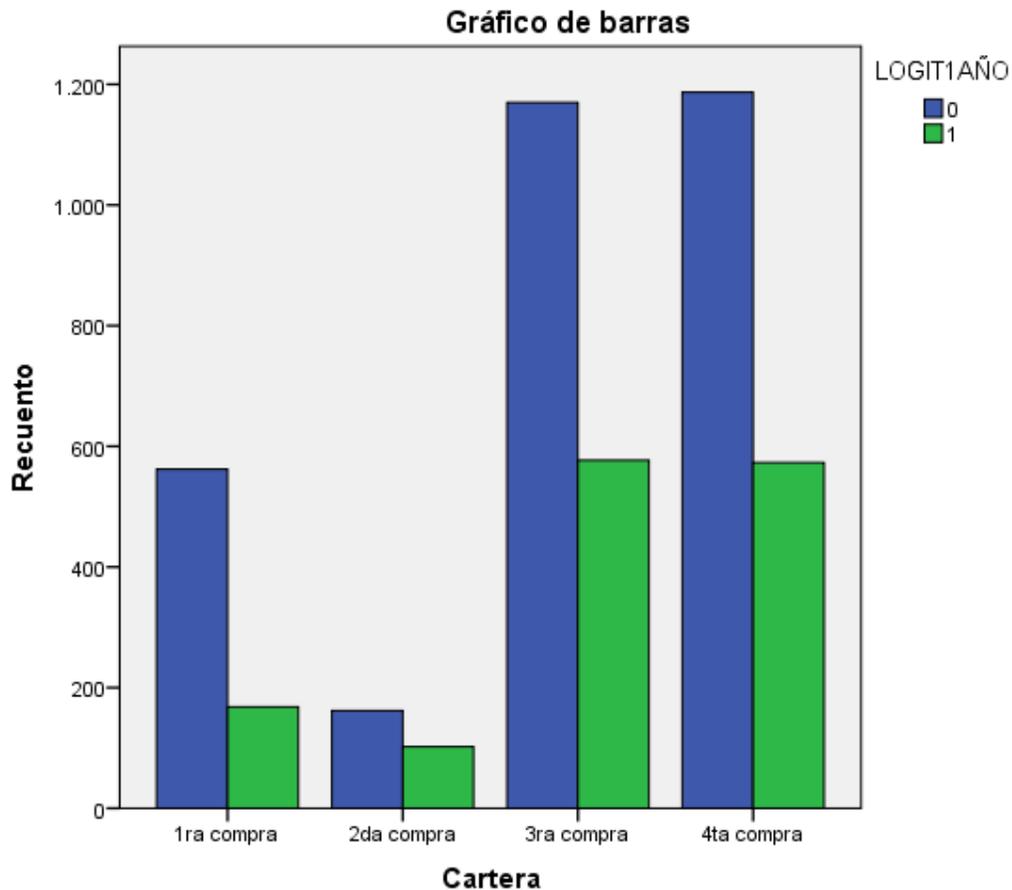
De este modo se puede evidenciar que los departamentos con mayor representación de impago son: los créditos que no cuenta con esta variables (por lo que implica que la información no es muy certera en esta variable, debido a la ausencia de información) con 1708 en impago y con una diferencia de 1611, seguido de Magdalena con 145 en impago y con una diferencia de 85 y por último Atlántico con 187 en impago y con una diferencia de 48. El departamento de mayor participación en pago se encuentra concentrado en, Córdoba con 245 créditos en pago y una diferencia de -162, seguida de Cauca con 115 créditos en pago y una diferencia de -55 y por último Valle con 123 créditos en pago y una diferencia de -33.

### Cruce de variables cartera y logit

Tabla 11 Cruce de variables entre cartera y Logit

		Recuento		Total	Diferencia
		LOGIT1AÑO			
		0	1		
Cartera	1ra compra	562	168	730	394
	2da compra	162	102	264	60
	3ra compra	1170	577	1747	593
	4ta compra	1187	573	1760	614
Total		3081	1420	4501	1661

Ilustración 12 Cruce de variables entre Cartera y Logit



En el presente cruce de variables se puede evidenciar que la mayor participación de impago se presenta en la cuarta compra con una diferencia de 614 créditos de impago, seguida de la tercera compra con una diferencia de 593 de créditos de impago y por último la primera compra con una diferencia de 394 de créditos de impago y en la segunda compra tiene una diferencia de 60 (siendo esta la de menor diferencia quiere decir que es donde mejor pagan de todas las compras), como se evidencia en la gráfica las barras azules tienen mayor proporción que las verdes.

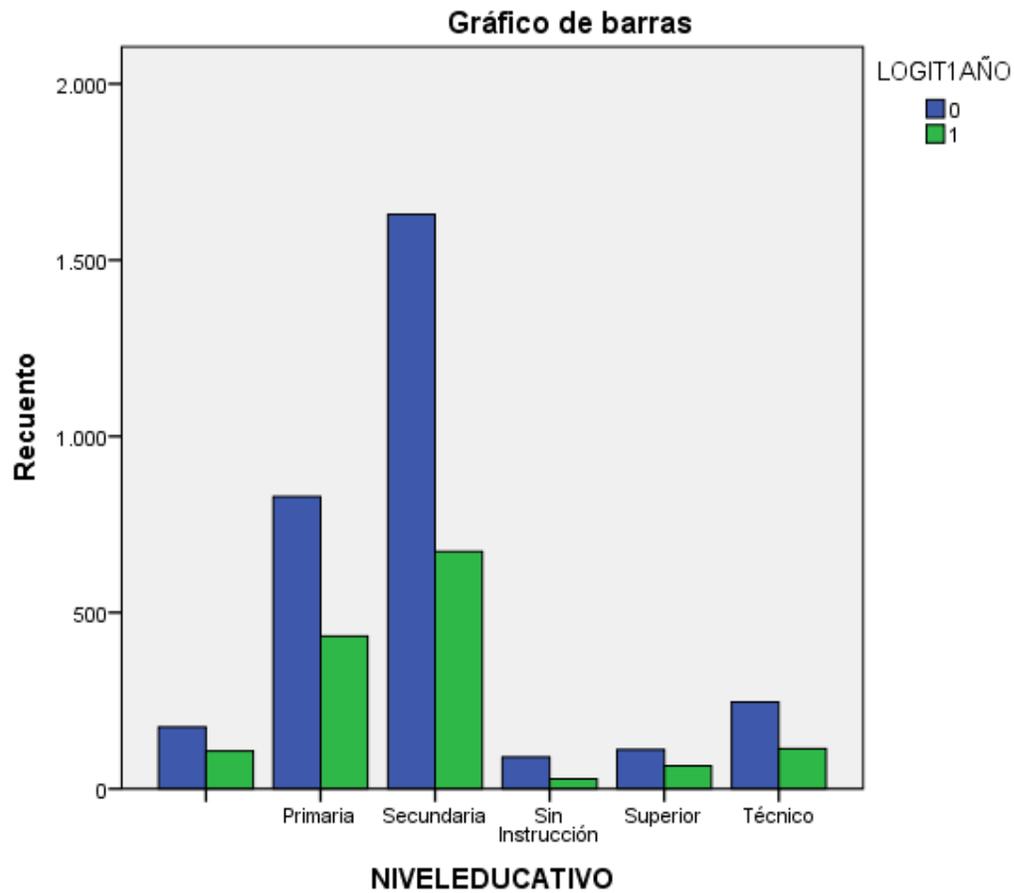
- **Cruce de variables nivel educativo y logit**

Tabla 12 Cruce de variables entre nivel educativo y Logit

**Tabla cruzada NIVELEDUCATIVO\*LOGIT1AÑO**

		Recuento		Total	Diferencia
		LOGIT1AÑO			
		0	1		
NIVELEDUCATIVO		175	107	282	68
	Primaria	829	433	1262	396
	Secundaria	1630	673	2303	957
	Sin Instrucción	90	28	118	62
	Superior	111	65	176	46
	Técnico	246	114	360	132
Total		3081	1420	4501	1661

Ilustración 13 Cruce de variables entre nivel educativo y Logit



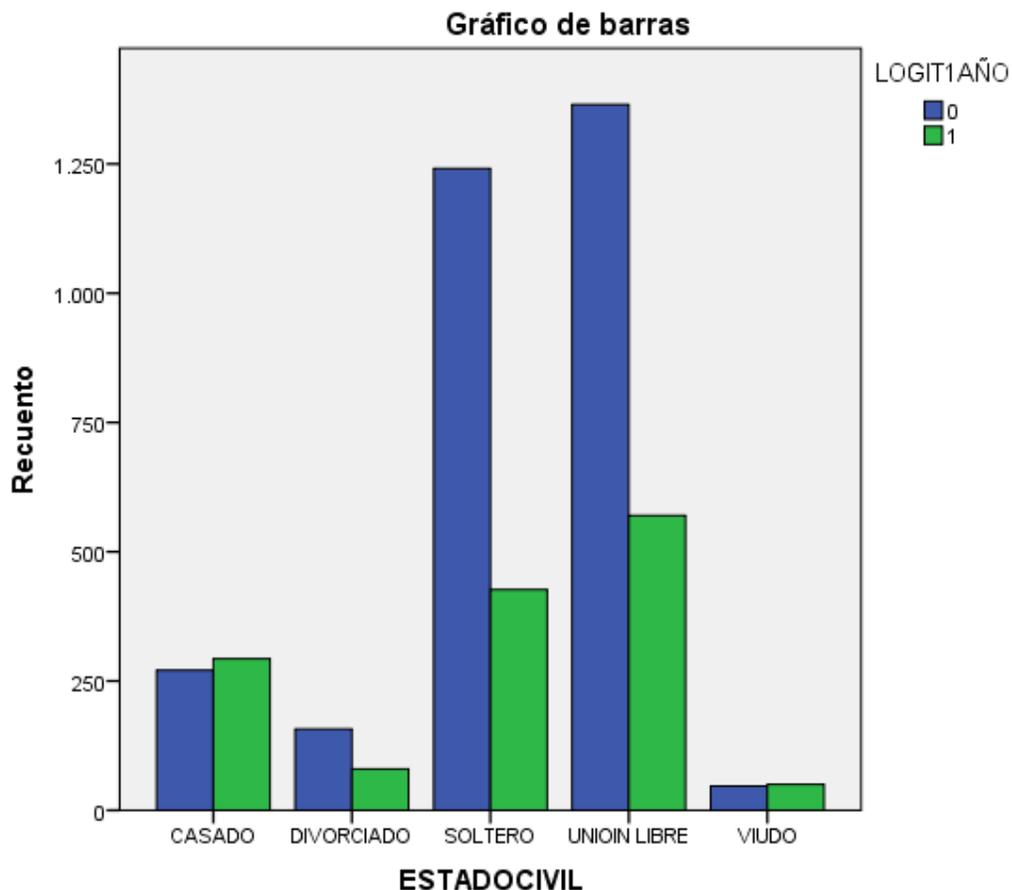
En el presente cruce de variables se puede evidenciar que la mayor participación de impago se presenta en el nivel educativo secundaria con una diferencia de 957, seguida del nivel educativo primaria con una diferencia de 396, seguido del nivel educativo educación técnica con una diferencia de 132, seguida de los créditos que no tiene la información de la variable con una diferencia de 68 y finalmente el nivel educativo sin instrucción y superior con una diferencia de 62 y 46 respectivamente (siendo el nivel superior el de menor diferencia, quiere decir que es el que mejor pagan).

- **Cruce de variables estado civil y logit**

Tabla 13 Cruce de variables entre Estado civil y Logit

		Recuento		Total	Diferencia
		LOGIT1AÑO			
		0	1		
ESTADOCIVIL	CASADO	271	293	564	-22
	DIVORCIADO	157	80	237	77
	SOLTERO	1241	427	1668	814
	UNIOIN LIBRE	1365	570	1935	795
	VIUDO	47	50	97	-3
Total		3081	1420	4501	1661

Ilustración 14 Cruce de variables entre Estado civil y Logit



En el presente cruce de variable se puede evidenciar que, se encuentra con mayor participación de impago el estado civil soltero con una diferencia de 814, seguido de unión libre con una diferencia de 795, seguido de los divorciados con una diferencia de 77 y finalmente el estado civil que incurre al cumplimiento crediticio es casado y viudo con una diferencia de -22 y -3 respectivamente

### Modelo Logit

El desarrollo sobre el modelo planteado se dio entre varias etapas con un orden lógico y secuencial para la obtención apropiada de los resultados, de la siguiente manera:

**1. Etapa: Selección de la muestra:** La base de datos de la cartera de microcréditos de la entidad contiene información del comportamiento de los clientes referente al pago de los créditos solicitados, durante el periodo de julio de 2016 a Julio de 2017 pertenecientes a la compra de cuatro carteras. En la determinación de la muestra implementada en el modelo de credit scoring se tomaron en cuenta los siguientes puntos:

- Se seleccionó la muestra de modo que el número de casos de crédito de impago se aproximara al de pago, ya que se poseía una población donde predominaban enormemente los créditos de impago, al ser una cartera castigada. De esta manera se buscó eliminar el problema de diferencia de tamaños de grupos que “generalmente, provoca una estimación inexacta al no definirse correctamente los factores de clasificación hacia los grupos previamente definidos” Mures, García y Vallejo (citado de Cantón, Lara Rubio, & Camino Blasco, 2010).
- Los créditos deben obtener la mayor cantidad de información completa sobre todas las variables explicativas expuestas en la base de datos, Por tanto, se eliminaron casos en los que había grandes faltantes de información.

El historial de crédito fue obtenido por parte de la casa de cobranza, con información tanto cualitativa como cuantitativa de cada cliente perteneciente al periodo del año ya mencionado. Dicha información fue procesada en Excel y se descartaron créditos con faltantes en variables explicativas. Finalmente, a través del programa SPSS y teniendo en cuenta que se requería una muestra balanceada (porcentaje de créditos de impago cercanos al de pagos), por medio de la aplicación de muestras complejas, se determinó un total de 5.501 créditos a insertar en el modelo.

- **PRUEBA T:**

Tabla 14 Prueba T

PRUEBA T			
Variables	muestra 4501	población 18753	Diferencia
	Media	Media	Media
Operación	2667669	2483854	183815
LOGIT1AÑO	0,32	0,09	0,22
FECHAAPERTURA	20144617	20142172	2444
EDADMOORA	344,00	346,20	-2,20
VALORINICIAL	1651915	2200572	-548657
SALDODEUDA	1111685	1393743	-282058
CUOTAMENSUAL	160598	188252	-27654
TOTALCUOTAS	12,7	15,7	-3,0
CUOTASCANCELADAS	6,7	9,4	-2,7
CUOTASMORA	12,4	13,2	-0,9
OFICINA	318,00	414,12	-96
actividadeconomica2	149,67	148,28	1
EC	3,1	3,0	0,0
@NE	1,9	1,7	0,2
C	2,9	2,7	0,2
D	7,1	9,1	-2,0

En la presente tabla se puede evidenciar que la prueba T refleja, que la muestra se comporta de casi igual manera que la población, en las once siguientes variables: LOGIT1AÑO, EC (estado civil), NE (nivel educativo), C (compra de cartera), D (departamento), cuotas moras, cuotas canceladas, edad de mora, total cuotas, oficina y actividad económica; ya que su diferencia es muy cercana a cero y tan solo en las siguiente cinco variables operación, fecha de apertura, valor inicial, saldo de la deuda y cuota mensual se alejan de cero, pero sin embargo se puede justificar ya que estas variables son de grandes cantidades y diferentes para cada crédito. Por tanto la prueba es acertada en gran proporción.

**2. Etapa: Variables aplicadas al modelo:** La variable dependiente es dicotómica, que hace referencia a los clientes que si pagan con valor de “1” y a los incumplidos con valor de “0”. Por su parte las variables independientes que se dividen en categóricas/cualitativas y numéricas/cuantitativas, se convirtieron a numéricas solamente ya que de esta forma se obtuvo un mejor ajuste del modelo. En seguida su interpretación:

- Cuota mensual: variable numérica que hace referencia al valor de la cuota que paga cada cliente mensualmente.
- Cuotas moras: variable numérica que hace alusión al número de cuotas que no se han pagado por crédito.
- Edad mora: variable numérica que indica el número de días de atraso que posee cada cliente.
- Oficina: variable numérica que expresa la ubicación donde el cliente solicitó el crédito.
- Total cuotas: variable numérica que refleja la cantidad total de cuotas que posee cada crédito.
- Valor inicial: variable numérica.
- Fecha apertura: variable numérica.
- Cartera
- Saldo de deuda: variable numérica que muestra el valor que resta para terminar de cancelar el crédito.
- Cuotas canceladas: variable numérica que hace referencia a las cuotas que han sido pagadas por parte del cliente.
- Departamento: variable categórica que enseña la ubicación geográfica desde donde se expidió el crédito. Los departamentos se encuentran numerados del 1 al 23 para su codificación e implementación en el modelo.

- Nivel educativo: variable categórica que se encuentra codificada de la siguiente manera:
  - 1=Primaria
  - 2=Secundaria
  - 3=Sin instrucción
  - 4=Superior
  - 5=Técnico
- AE: variable numérica que enseña los tipos de ocupación que tienen los clientes.

**3. Etapa: REGRESIÓN LOGÍSTICA:** En esta etapa se ingresaron las variables mencionadas al modelo de regresión logística binaria, por medio del programa IBM SPSS para de esta manera poder determinar el nivel de significancia de estas en la explicación de la variable dependiente.

Primero se evidencia un cuadro que contiene el número de casos incluidos en el análisis, casos perdidos y casos no seleccionados; como los dos últimos son cero entonces el total de casos seleccionados es 4.501.

*Tabla 15 Resumen de procesamiento de casos*

Resumen de procesamiento de casos		N	Porcentaje
Casos sin ponderar <sup>a</sup>			
Casos seleccionados	Incluido en el análisis	4501	100,0
	Casos perdidos	0	,0
	Total	4501	100,0
Casos no seleccionados		0	,0
Total		4501	100,0

En seguida, se presenta la tabla de codificación de la variable dependiente dicotómica, siendo 1 “pago” y 0 “no pago”.

Tabla 16 Codificación de variable dependiente

Codificación de variable dependiente	
Valor original	Valor interno
0	0
1	1

### Bloque inicial: Bloque 0

Tabla 17 Tabla de clasificación

Tabla de clasificación					
		Pronosticado			
		LOGIT1AÑO		Porcentaje correcto	
Observado		0	1		
Paso 0	LOGIT1AÑO	0	3081	0	100,0
		1	1420	0	,0
	Porcentaje global				
a. La constante se incluye en el modelo.					
b. El valor de corte es ,500					

Para el análisis de regresión logística el bloque 0 indica que hay un 68,5 % de probabilidad de acierto en el resultado de la variable dependiente, asumiendo que todas las personas no pagan el crédito. Este resultado se da debido a que no hay variables independientes, así que el programa sencillamente se basa en la frecuencia de los resultados de la variable dependiente para sustentarse.

Tabla 18 Variables que no están en la ecuación

Las variables no están en la ecuación			Puntuación	gl	Sig.
Paso 0	Variables	EDADMOORA	,587	1	,443
		VALORINICIAL	834,556	1	,000
		SALDODEUDA	47,180	1	,000

TOTALCUOTAS	1434,123	1	,000
CUOTASCANCELADAS	2028,268	1	,000
CUOTASMORA	,236	1	,627
OFICINA	706,261	1	,000
C	4,916	1	,027

Esta tabla enseña las variables que se van a aplicar de ahora en adelante para el desarrollo del modelo. De esta manera se procede al siguiente bloque:

### Bloque 1: Método= Avanzar por pasos (Wald)

Tabla 19 Prueba ómnibus de coeficientes de modelo

		Chi-cuadrado	gl	Sig.
Paso 1	Paso	2410,129	1	,000
	Bloque	2410,129	1	,000
	Modelo	2410,129	1	,000
Paso 2	Paso	69,138	1	,000
	Bloque	2479,266	2	,000
	Modelo	2479,266	2	,000
Paso 3	Paso	163,261	1	,000
	Bloque	2642,528	3	,000
	Modelo	2642,528	3	,000
Paso 4	Paso	87,400	1	,000
	Bloque	2729,928	4	,000
	Modelo	2729,928	4	,000
Paso 5	Paso	23,379	1	,000
	Bloque	2753,307	5	,000
	Modelo	2753,307	5	,000
Paso 6	Paso	40,345	1	,000
	Bloque	2793,652	6	,000
	Modelo	2793,652	6	,000
Paso 7	Paso	8,563	1	,003
	Bloque	2802,214	7	,000
	Modelo	2802,214	7	,000
Paso 8	Paso	6,697	1	,010

Bloque	2808,911	8	,000
Modelo	2808,911	8	,000

Para el bloque 1 del modelo, la puntuación de eficiencia estadística de ROA (Chi cuadrado) indica que hay una mejora significativa en la predicción de la probabilidad de las variables de la variable dependiente (VD) (Chi cuadrado: 2808,911; gl: 8;  $p < ,001$ ). Es decir, al introducir las variables independientes está mejorando la predicción de la variable dependiente.

Tabla 20 Resumen del modelo

Resumen del modelo			
Paso	Logaritmo de la verosimilitud -2	R cuadrado de Cox y Snell	R cuadrado de Nagelkerke
1	3201,894 <sup>a</sup>	,415	,582
2	3132,756 <sup>a</sup>	,424	,594
3	2969,495 <sup>a</sup>	,444	,623
4	2882,095 <sup>a</sup>	,455	,638
5	2858,716 <sup>a</sup>	,458	,642
6	2818,371 <sup>a</sup>	,462	,649
7	2809,809 <sup>a</sup>	,463	,650
8	2803,112 <sup>a</sup>	,464	,651

a. La estimación ha terminado en el número de iteración 6 porque las estimaciones de parámetro han cambiado en menos de ,001.

El valor de R cuadrado de Nagelkerke indica que el modelo propuesto explica el 65,1% de la varianza de la variable dependiente (,651), por tanto se posee un nivel de ajuste aceptable en regresión logística.

Tabla 21 Prueba de Hosmer y Lemeshow

Prueba de Hosmer y Lemeshow			
Paso	Chi-cuadrado	gl	Sig.
1	210,686	7	,000
2	182,919	8	,000
3	58,933	8	,000

4	39,898	8	,000
5	26,341	8	,001
6	36,381	8	,000
7	53,323	8	,000
8	51,048	8	,000

Esta prueba indica que tan significativa es la varianza explicada del R2 para la variable dependiente. Para este modelo la prueba indica que la varianza explicada por el modelo no es satisfactoria pues el grado de significancia es bajo. De acuerdo al artículo de investigación de (Cantón, Lara Rubio, & Camino Blasco, 2010) a pesar de su nivel de significación bajo, “se debe tomar en cuenta que este dato estadístico solamente brinda pistas sobre la calidad del ajuste del modelo”.

Tabla 22 Tabla de clasificación

Tabla de clasificación					
	Observado		Pronosticado		
			LOGIT1AÑO		Porcentaje correcto
			0	1	
Paso 1	LOGIT1AÑO	0	2961	120	96,1
		1	397	1023	72,0
	Porcentaje global				88,5
Paso 2	LOGIT1AÑO	0	2953	128	95,8
		1	387	1033	72,7
	Porcentaje global				88,6
Paso 3	LOGIT1AÑO	0	2935	146	95,3
		1	372	1048	73,8
	Porcentaje global				88,5
Paso 4	LOGIT1AÑO	0	2935	146	95,3
		1	371	1049	73,9
	Porcentaje global				88,5
Paso 5	LOGIT1AÑO	0	2937	144	95,3
		1	372	1048	73,8
	Porcentaje global				88,5
Paso 6	LOGIT1AÑO	0	2939	142	95,4
		1	373	1047	73,7

	Porcentaje global				88,6
Paso 7	LOGIT1AÑO	0	2947	134	95,7
		1	366	1054	74,2
	Porcentaje global				88,9
Paso 8	LOGIT1AÑO	0	2944	137	95,6
		1	367	1053	74,2
	Porcentaje global				88,8
a. El valor de corte es ,500					

Para el análisis de regresión logística el bloque 1 indica que hay un 88,8% de probabilidad de acierto en el resultado de la variable dependiente. Por tanto, se puede deducir que, al agregar las variables independientes, el grado de acierto pasa de ser del 68,5% identificado en el bloque 0 a 88%; lo que quiere decir que hay una mejora significativa en la predicción del modelo, recatando el 74% que corresponde a la predicción de los clientes que pagan.

Tabla 23 Variables en la ecuación

		Variables en la ecuación					
		B	Error estándar	Wald	gl	Sig.	Exp(B)
Paso 1 <sup>a</sup>	CUOTASCANCELADAS	,449	,014	1029,070	1	,000	1,567
	Constante	-3,923	,105	1407,985	1	,000	,020
Paso 2 <sup>b</sup>	CUOTASCANCELADAS	,409	,014	833,464	1	,000	1,505
	OFICINA	,001	,000	69,346	1	,000	1,001
	Constante	-4,150	,112	1378,272	1	,000	,016
Paso 3 <sup>c</sup>	CUOTASCANCELADAS	,420	,015	800,850	1	,000	1,521
	OFICINA	,002	,000	157,157	1	,000	1,002
	C	,550	,045	147,481	1	,000	1,734
	Constante	-6,158	,216	812,342	1	,000	,002
Paso 4 <sup>d</sup>	CUOTASCANCELADAS	,440	,015	816,259	1	,000	1,553
	CUOTASMORA	,047	,005	83,065	1	,000	1,048
	OFICINA	,002	,000	160,951	1	,000	1,002
	C	,621	,047	175,739	1	,000	1,860
	Constante	-7,074	,250	801,128	1	,000	,001
Paso 5 <sup>e</sup>	SALDODEUDA	,000	,000	24,744	1	,000	1,000

	CUOTASCANCELADAS	,456	,016	812,632	1	,000	1,578
	CUOTASMORA	,080	,009	87,849	1	,000	1,083
	OFICINA	,002	,000	164,974	1	,000	1,002
	C	,632	,047	180,157	1	,000	1,882
	Constante	-7,350	,260	802,027	1	,000	,001
Paso 6 <sup>f</sup>	VALORINICIAL	,000	,000	30,203	1	,000	1,000
	SALDODEUDA	,000	,000	47,803	1	,000	1,000
	CUOTASCANCELADAS	,322	,028	130,022	1	,000	1,380
	CUOTASMORA	,098	,009	107,642	1	,000	1,103
	OFICINA	,002	,000	172,434	1	,000	1,002
	C	,637	,048	178,247	1	,000	1,890
	Constante	-7,164	,264	738,217	1	,000	,001
Paso 7 <sup>g</sup>	EDADMOORA	,005	,002	8,696	1	,003	1,005
	VALORINICIAL	,000	,000	32,362	1	,000	1,000
	SALDODEUDA	,000	,000	51,289	1	,000	1,000
	CUOTASCANCELADAS	,320	,029	125,505	1	,000	1,377
	CUOTASMORA	,101	,010	111,623	1	,000	1,106
	OFICINA	,002	,000	157,229	1	,000	1,002
	C	,617	,048	162,877	1	,000	1,854
	Constante	-8,875	,645	189,264	1	,000	,000
Paso 8 <sup>h</sup>	EDADMOORA	,005	,002	9,727	1	,002	1,005
	VALORINICIAL	,000	,000	35,796	1	,000	1,000
	SALDODEUDA	,000	,000	55,510	1	,000	1,000
	TOTALCUOTAS	,113	,044	6,558	1	,010	1,119
	CUOTASCANCELADAS	,206	,052	15,454	1	,000	1,229
	CUOTASMORA	,059	,019	9,718	1	,002	1,060
	OFICINA	,002	,000	158,694	1	,000	1,002
	C	,616	,048	161,924	1	,000	1,852
	Constante	-9,095	,653	194,201	1	,000	,000

a. Variables especificadas en el paso 1: CUOTASCANCELADAS.

b. Variables especificadas en el paso 2: OFICINA.

c. Variables especificadas en el paso 3: C.

d. Variables especificadas en el paso 4: CUOTASMORA.

e. Variables especificadas en el paso 5: SALDODEUDA.

f. Variables especificadas en el paso 6: VALORINICIAL.

g. Variables especificadas en el paso 7: EDADMOORA.

h. Variables especificadas en el paso 8: TOTALCUOTAS.

En el octavo paso de esta tabla se pueden ver las variables que el programa encontró como mayormente explicativas y por lo tanto las dejará en la ecuación. Aquí se denotan las betas, error estándar, valor estadístico de Wald, la significancia estadística, la constante y el valor de Exp (B). Las variables que dejó son: cuotas canceladas, oficina, C (Cartera), cuotas mora, saldo de la deuda, valor inicial, edad de mora y total cuotas.

La puntuación de Wald para el modelo probado indica que la variable independiente aporta significativamente a la predicción de la variable dependiente y los resultados obtenidos se pueden generalizar a la población:

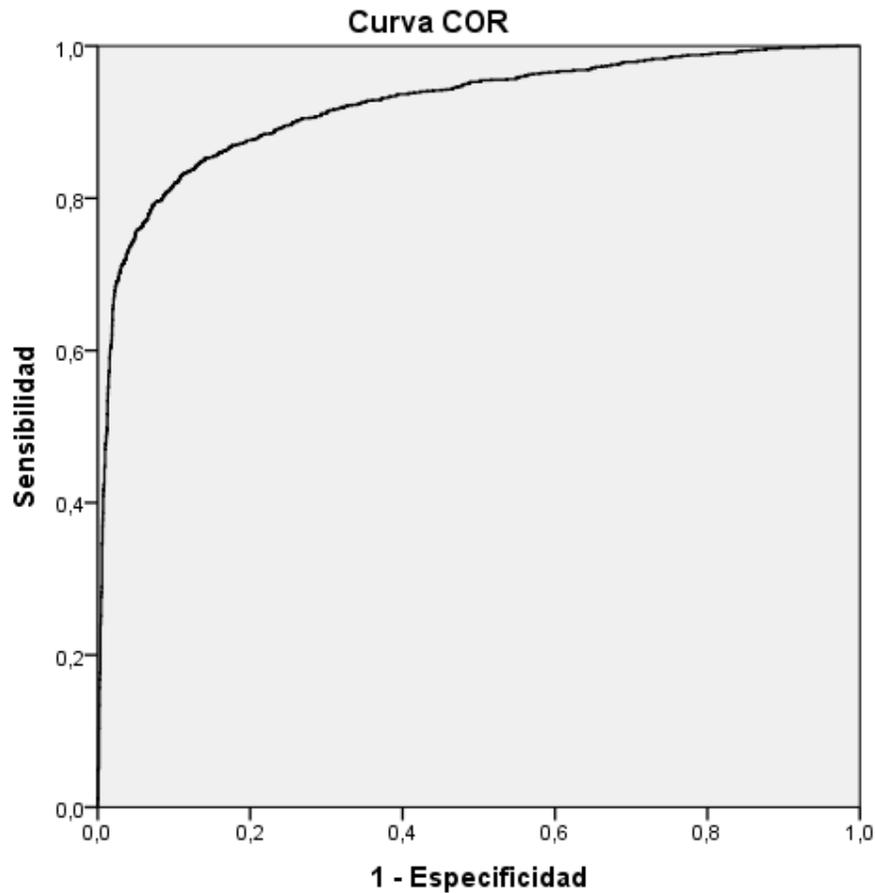
- **Matriz de correlaciones:** En el anexo A se muestran las correlaciones entre las variables para percibir el grado de relación que poseen entre ellas. De acuerdo con la matriz en el octavo paso no se encuentran variables relacionadas o superiores a 0,5, todas tienen un valor inferior e incluso negativo, lo que quiere decir que estas variables tienen relación inversa. El hecho de que no se encuentren correlacionadas es positivo para el modelo ya que no se quieren variables que digan lo mismo entre sí.

## **CURVA COR**

La curva COR o ROC identifica la capacidad de diagnóstico que tiene el modelo de regresión logística para clasificar entre los clientes que pagan y los que no. A la proporción de predicciones correctas de “no pago” (Default) se le dice sensibilidad (aciertos) y a la proporción de predicciones acertadas de “pago” (no default) se les llama especificidad (falsas alarmas).

De esta manera el eje Y representa el grupo de créditos con probabilidad de no pago y en el eje X se encuentra los falsos positivos (1- Especificidad), es decir el grupo de créditos con probabilidad de pago mal clasificados.

Ilustración 15 Curva COR



La curva ROC para el modelo logit indica que este modelo tiene un buen poder discriminatorio, puesto que el trazo es inclinado a la línea horizontal, de hecho, el área bajo la curva (AUROC) es 0,922. Entonces, se confirma que el modelo dispone de una correcta capacidad de discriminación entre los dos grupos de clientes.

Tabla 24 Área bajo la curva

<b>Área bajo la curva</b>
---------------------------

VARIABLES DE RESULTADO DE PRUEBA: Probabilidad pronosticada
Área
,922
Las variables de resultado de prueba: Probabilidad pronosticada tienen, como mínimo, un empate entre el grupo de estado real positivo y el grupo de estado real negativo. Las estadísticas podrían estar sesgadas.

## Prueba Backtesting

Tabla 25 Cartera año y medio

CARTERA AÑO Y MEDIO		
NOMBRE	VALOR	PORCENTAJE
PAGO	2074	11,1%
IMPAGO	16679	88,9%
TOTAL CARTERA AÑO Y MEDIO	18753	100,0%

En el presente cuadro se puede observar la cartera completa, la cual pertenece al periodo de julio del 2016 hasta diciembre 2017 y está compuesta así: con 2.074 créditos pagos que equivalen a un 11,1% y con 16.679 créditos que incurren a impago que equivalen a un 88,9% para un total del 100%.

Tabla 26 Cartera de un año

CARTERA UN AÑO		
PAGO	1722	9,2%
IMPAGO	17031	90,8%
TOTAL CARTERA AÑO	18753	100,0%

En el presente cuadro se muestra la cartera de crédito desde julio 2016 hasta junio 2017, la cual corresponde al año en que se corrió el modelo, dejando así, el semestre restante (julio 2017- diciembre 2017) por fuera, para hacer la prueba backtesting. Esta cartera está compuesta por 1.722 créditos pagos que equivalen a un 9,4% y con 17.031 créditos que incurren al impago, para un total del 100%.

Tabla 27 Cartera de 6 meses

CARTERA EN 6 MESES		
PAGO	352	1,9%
IMPAGO	16679	88,9%
PAGO DEL PRIMER AÑO	1722	9,2%
TOTAL CARTERA SEIS MESES	18753	100,0%

En la tabla número 27 se muestra la cartera de créditos correspondiente a los 6 meses siguientes (a partir de julio 2017 hasta diciembre 2017), que se distribuye así: 352 créditos en pagos que corresponden a un 2,1%, seguido de 16.679 créditos que incurren a impago, equivalentes a 88,9% y finalmente, los 1.722 créditos que fueron pagados en el año (julio 2016 hasta junio 2017); para un total del 100%.

Tabla 28 Prueba Backtesting

PRUEBA BACKTESTING				
ACIERTOS DEL MODELO			NO ACIERTOS DEL MODELO	
PAGO	185	52,6%	167	47,4%
IMPAGO	10218	61,3%	6461	38,7%
TOTAL ACIERTOS	10403		6628	

Teniendo en cuenta que se asume que la muestra selecciona va a comportarse de la misma manera que la población, en el presente cuadro se muestran los resultados de la prueba backtesting realizada para toda la cartera. De los 352 créditos (como se menciona en el cuadro anterior) que pagan en los seis meses, el modelo acierta a 185 que corresponden a un 52,6% por lo tanto, no acertó en el 47,4% del total de pagos.

Para el caso de impagos, de un total de 16.679 créditos que incurren a default en los seis meses, el modelo acertó en 10.218 créditos que corresponden a un 61,3%, por lo tanto, el modelo no predijo correctamente el 38,7% del total de impagos. Aunque los porcentajes de acierto puedan parecer pequeños, es necesario reconocer que esto es debido a la población que desde un principio se encontraba sesgada, en la gran superioridad de los créditos en estado de impago frente al no

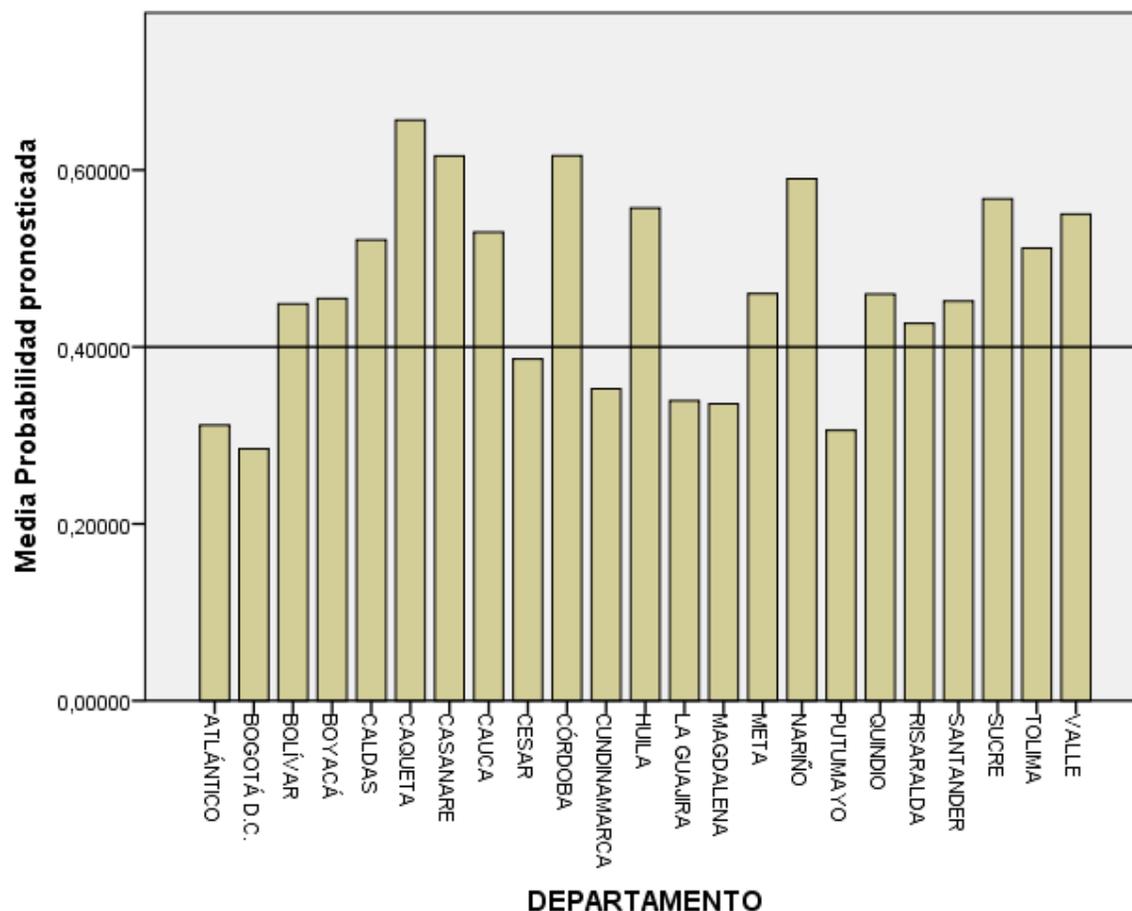
default, debido al estado de deterioro de la cartera. Por tanto, al equilibrar la muestra, se presentaba dificultad para que los resultados del modelo se reflejaran a toda la población.

### **Análisis de los resultados del modelo Logit**

En esta etapa se enseña de acuerdo a los resultados de probabilidad de pago arrojados por el modelo, las personas con mayor o menor propensión a pago, según las siguientes variables categoricas:

#### **a. Analisis de la probabilidad de incumplimiento de los departamentos.**

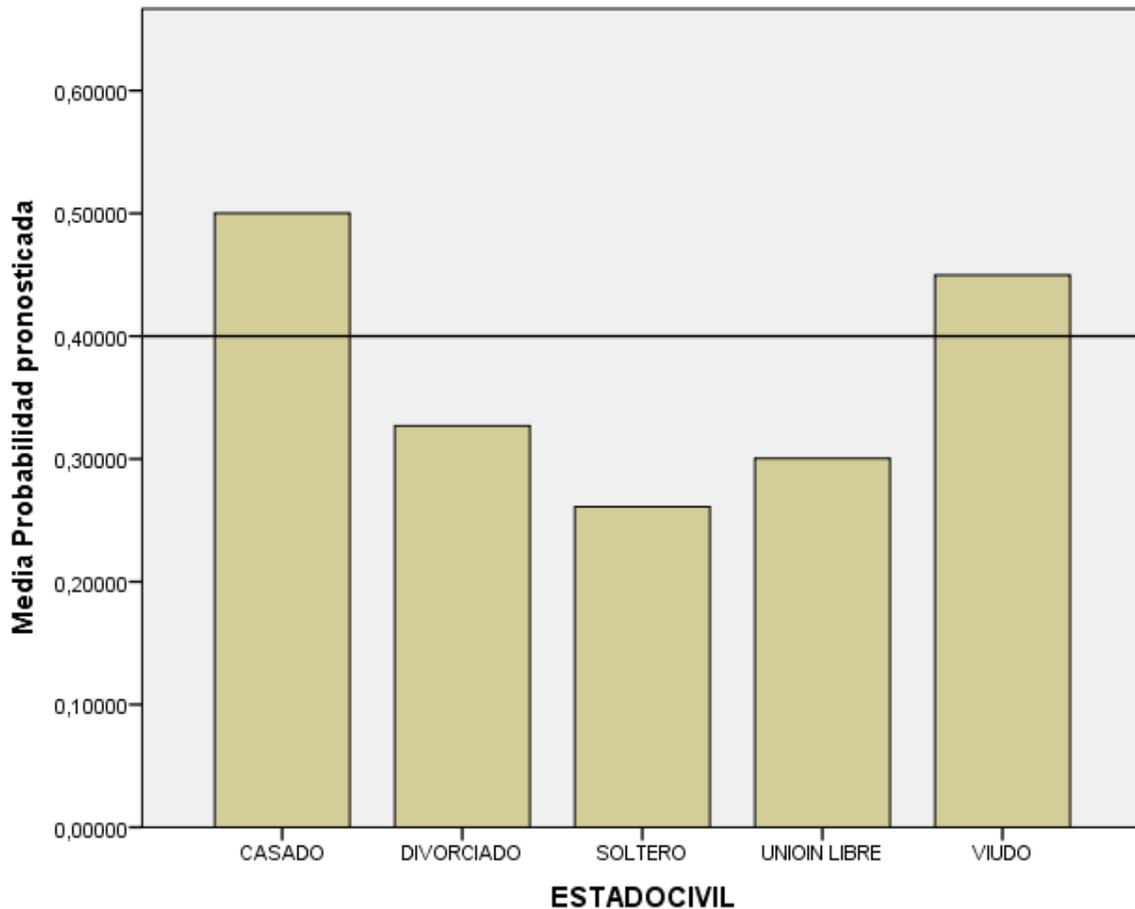
Ilustración 16 Probabilidad de incumplimiento por departamento



En presente grafico se puede evidenciar que los departamentos que incurren mayormente a cumplimiento de las obligaciones de crédito son: Caquetá, Casanare, Córdoba, Nariño y Sucre, entre otros y los departamentos que incurren al incumplimiento de las obligaciones de crédito son: Bogotá, Putumayo, Magdalena y la Guajira, entre otros.

#### **b. Análisis de la probabilidad de incumplimiento del estado civil.**

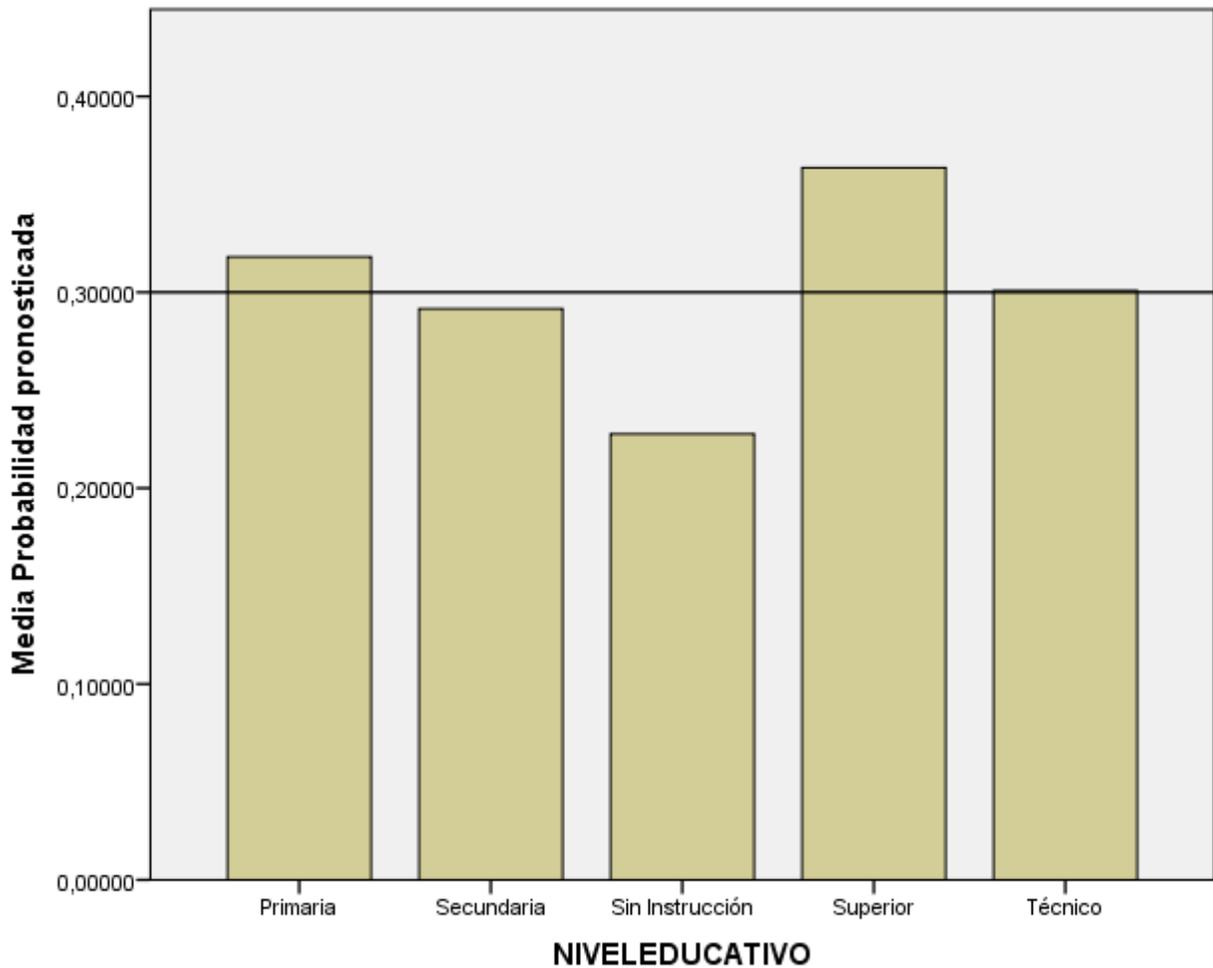
Ilustración 17 Probabilidad de incumplimiento por estado civil



En el presente grafico se puede evidenciar que las personas con estado civil casados y viudos son los que más incurren a cumplimiento de las obligaciones crediticias y por lo contrario los de estado civil divorciados, unión libre y solteros son mayormente propensos al incumplimiento de las obligaciones crediticias.

### c. Análisis de la probabilidad de incumplimiento del nivel educativo

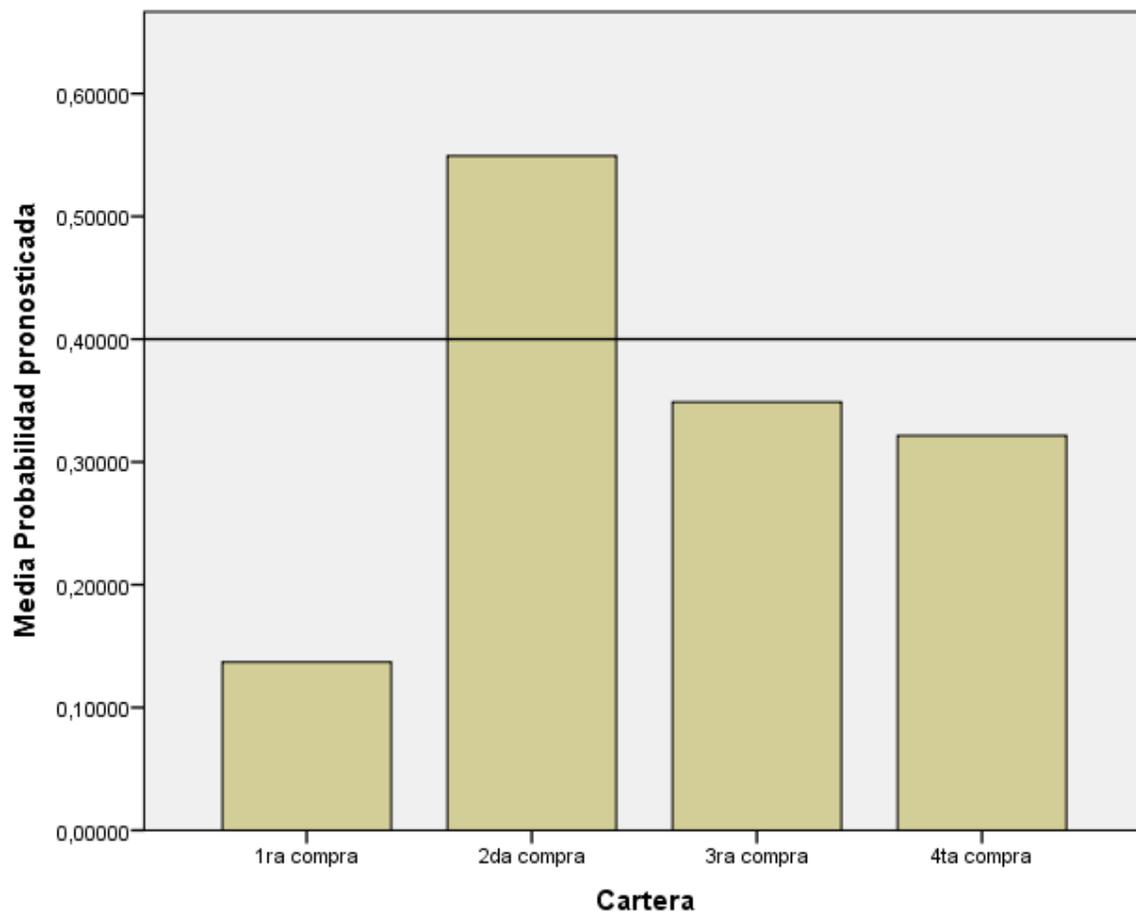
Ilustración 18 Probabilidad de incumplimiento por nivel educativo



En el presente grafico se puede evidenciar que las personas con nivel educativo superior y primaria son los que más incurren al cumplimiento de las obligaciones de crédito y por lo contrario las personas que más incurren al incumplimiento son los de nivel educativo técnico, secundaria y sin instrucción

#### **d. Análisis de la probabilidad de incumplimiento de cada compra de la cartera**

Ilustración 19 Probabilidad de incumplimiento por compra de cartera



En el presente grafico se puede evidenciar que la segunda compra es la que tiene mayor tendencia al cumplimiento de las obligaciones crediticias, seguida de la tercera compra. Por lo contrario, las compras 1ra y 4ta son las que más incurren al incumplimiento de las obligaciones crediticias.

## CAPÍTULO V: CONCLUSIONES Y RECOMENDACIONES

### Conclusiones

A lo largo del desarrollo del trabajo se ha podido identificar que las instituciones se enfrentan de forma inminente al riesgo de crédito y es por esto que la Superintendencia Financiera de Colombia, crea el SARC, buscando adaptarlo en cada entidad para gestionar conjuntamente el riesgo. Esto se da debido a que se reconoce que el activo más representativo de toda entidad financiera es la cartera, por tanto debe cuidarse, administrarse y desarrollar seguimiento de la mejor manera posible. Para tal efecto existen los modelos de credit scoring como el logit que en este caso permite reconocer los créditos con mayor propensión a pago e impago, para así reconocer qué suma se puede recuperar de ellos y obtener una buena rentabilidad para el establecimiento.

En el análisis descriptivo de las variables, antes de ser ingresadas al modelo; se pudo detectar que la mayor proporción de impago se presenta en personas con un nivel educativo de primaria y secundaria. Así mismo, en la variable estado civil los solteros y en unión libre. Finalmente, en el caso de los departamentos se encuentran Magdalena, Atlántico y Cesar con mayor default.

Las variables con mayor poder explicativo según el modelo son: cuotas canceladas, oficina, cartera, cuotas en mora, saldo de la deuda, valor inicial, edad de mora y total cuotas.

El modelo de regresión logística demostró ser considerablemente predictivo en cuanto a la clasificación en la probabilidad de impago con un porcentaje global del 88,8%, con una probabilidad de clasificar bien a los buenos pagadores del 74,2%, y una curva ROC para el modelo logit cuyo trazo es cercano a la línea horizontal que indica que este modelo tiene un buen poder discriminatorio (sensibilidad-especificidad), de hecho, el área bajo la curva (AUROC) es 0,922.

Uno de los puntos difíciles para poder correr el modelo generalmente es la información incompleta que yace en las bases de datos y en esta ocasión, también se encontraron variables con celdas vacías como es el caso de la variable departamento. También la población tan desbalanceada al pertenecer a una cartera castigada, dificultó que los resultados de la muestra se reflejaran completamente en la población.

La prueba backtesting realizada de seis meses, respalda positivamente los resultados del modelo, a pesar de acertar el 52,6% del pago de los créditos realizados en el semestre. Pues este resultado ya estaba condicionado por el caso particular de la población sesgada en el default y desbalanceada.

De acuerdo a los resultados de la probabilidad de incumplimiento arrojados por el modelo Logit, las personas con mayor propensión a pago, pertenecen a los siguientes departamentos: Caquetá, Casanare, Córdoba y Nariño. Así mismo el estado civil: soltero y viudo. Los de nivel educativo superior y primaria, también tienen mayor probabilidad de pago. Finalmente, las carteras 2 y 3 están más próximas al pago.

Una desventaja del modelo es que esta basa sus predicciones en acontecimientos pasados, así que supone que lo que paso antes sucederá de igual forma en el futuro. El inconveniente radica en que los factores externos pueden alterar los resultados, por ejemplo: no se puede asumir que el comportamiento de créditos en un entorno económico estable será el mismo que en un entorno de crisis económica.

## **Recomendaciones**

El anterior modelo ayuda a construir y a acercarse a la probabilidad de incumplimiento de los créditos en la cartera a pesar de la falta de información, por tanto es recomendable que en otra investigación se mejore el modelo con inclusión de nueva información y quizás variables macroeconómicas que permitan mejorar la robustez del mismo.

Es importante propender por un sistema de información más efectivo, ya que la información de los clientes es el principal insumo del modelo para poder catalogar los créditos de acuerdo a su perfil de riesgo como buenos o malos pagadores.

## Bibliografía

(s.f.).

Kukuk, M., & Ronnberg, M. (2012). Corporate credit default models: a mixed logit approach. *Springer Science+Business Media*, 2. Obtenido de <https://link.springer.com/article/10.1007/s11156-012-0281-4>

Lagua Tubón, V. (2015). *ponteficia universidad catolica del ecuador*. Obtenido de <http://repositorio.pucesa.edu.ec/bitstream/123456789/1524/1/76061.pdf>

Rodríguez Guevara, D. E., & Trespalcios Carrasquilla, A. (s.f.). *universida EAFIT*. Obtenido de <http://www.eafit.edu.co/escuelas/economiayfinanzas/laboratorio-financiero/burkenroad/Paginas/Medici%C3%B3n%20de%20Valor%20en%20Riesgo%20en%20Cartera%20de%20Clientes%20a%20Trav%C3%A9s.pdf>

Adrich, J., & Nelson, F. (1984). *Linear Probability, Logit, and Probit Models*. California: SAGE PUBLICATIONS.

Alderete, A. M. (2006). Fundamentos del Análisis de Regresión Logística en la Investigación Psicológica. *Revista Evaluar*, 52-67.

Arenas Díaz, M. A., Boccardi Rodriguez, P., & Piñeyrúa Ibáñez, A. (Mayo de 2012). *Observatorio de Inclusión Financiera*. Obtenido de Observatorio de Inclusión Financiera: [http://oif.ccee.edu.uy/wp-content/uploads/2015/03/Tesis-Arenas-Mar%C3%ADa-Alejandra\\_FCEyA-2012.pdf](http://oif.ccee.edu.uy/wp-content/uploads/2015/03/Tesis-Arenas-Mar%C3%ADa-Alejandra_FCEyA-2012.pdf)

Autoridad De Supervisión Del Sistema Financiero. (2010). *RECOPILACIÓN DE NORMAS PARA BANCOS Y ENTIDADES FINANCIERAS*. Obtenido de <http://servdmzw.asfi.gob.bo/circular/Anexos/T05/C01/T05C01A01.pdf>

Autorregulador del Mercado de Valores de Colombia. (Mayo de 2014). *Conozca los riesgos del mercado de valores*. Obtenido de Conozca los riesgos del mercado de valores:

<http://www.amvcolombia.org.co/attachments/data/20120605120911.pdf>

Avila Bustos, J. C. (Agosto de 2005). *Pontificia Universidad Javeriana Bogotá*. Obtenido de Pontificia Universidad Javeriana Bogotá:

<http://www.javeriana.edu.co/biblos/tesis/economia/tesis01.pdf>

Banco de la República. (2015). *Banco de la República*. Obtenido de Banco de la República:

[http://www.banrepcultural.org/blaavirtual/ayudadetareas/economia/el\\_upac\\_y\\_la\\_uvr#Documentos\\_relacionados\\_en\\_la\\_Biblioteca\\_Virtual](http://www.banrepcultural.org/blaavirtual/ayudadetareas/economia/el_upac_y_la_uvr#Documentos_relacionados_en_la_Biblioteca_Virtual)

Banco de la República de Colombia. (s.f.). Obtenido de

[http://www.banrepcultural.org/blaavirtual/ayudadetareas/economia/riesgo\\_bancario](http://www.banrepcultural.org/blaavirtual/ayudadetareas/economia/riesgo_bancario)

Bolsa de Valores de Colombia. (2014). *Bolsa de Valores de Colombia*. Obtenido de Bolsa de Valores de Colombia:

[http://www.bvc.com.co/pps/tibco/portalbvc/Home/Empresas/Guia\\_Mercado\\_Valores?com.tibco.ps.pagesvc.action=updateRenderState&rp.currentDocumentID=-7ca0c036\\_147b6b20b27\\_5e970a0a600b&rp.revisionNumber=1&rp.attachmentPropertyName=Attachment&com.tibco.ps.pagesv](http://www.bvc.com.co/pps/tibco/portalbvc/Home/Empresas/Guia_Mercado_Valores?com.tibco.ps.pagesvc.action=updateRenderState&rp.currentDocumentID=-7ca0c036_147b6b20b27_5e970a0a600b&rp.revisionNumber=1&rp.attachmentPropertyName=Attachment&com.tibco.ps.pagesv)

Bunge, M. (2004). *La Investigación Científica* (tercera ed.). Mexico: Siglo XXI editoriales. Obtenido de

<https://books.google.es/books?hl=es&lr=&id=iDjRhR82JHYC&oi=fnd&pg=PA1&dq=investigacion+mario+bunge&ots=cL6kTat2Uf&sig=WmDdv2fsNMXU-dxt19i1Ow26FNo#v=onepage&q=investigacion%20mario%20bunge&f=false>

- Calixto Salazar, M. M., & Casaverde Carranza, L. F. (2011). *Repositorio de la Universidad del Pacífico*. Obtenido de Repositorio de la Universidad del Pacífico:  
[http://repositorio.up.edu.pe/bitstream/handle/11354/1056/Mar%C3%ADa\\_Tesis\\_maestria\\_2011.pdf?sequence=1&isAllowed=y](http://repositorio.up.edu.pe/bitstream/handle/11354/1056/Mar%C3%ADa_Tesis_maestria_2011.pdf?sequence=1&isAllowed=y)
- Cantón, S. R., Lara Rubio, J., & Camino Blasco, D. (Junio de 2010). *Un Modelo de Credit Scoring para instituciones de microfinanzas en el marco de Basilea II*. Obtenido de Un Modelo de Credit Scoring para instituciones de microfinanzas en el marco de Basilea II:  
<file:///C:/Users/USER/Downloads/RIESGO%20DE%20CREDITO%20CARTERA%20CLIENTES%202.pdf>
- Caruana, J. (2010). Basilea III: hacia un sistema financiero más seguro. *Basilea III: hacia un sistema financiero más seguro*, (págs. 1-8). Madrid.
- Chorafas. (2000). *Modelos para medir el riesgo de crédito de la banca* . Obtenido de Modelos para medir el riesgo de crédito de la banca :  
<file:///C:/Users/USER/Downloads/modelos%20para%20riesgo%20de%20credito%20de%20la%20cartera.pdf>
- Congreso De La Republica . (1998). *Decreto 663 de 1993*. Obtenido de  
[http://www.secretariasenado.gov.co/senado/basedoc/estatuto\\_organico\\_sistema\\_financiero.html](http://www.secretariasenado.gov.co/senado/basedoc/estatuto_organico_sistema_financiero.html)
- Crespo G, J. Y. (2011). CAMEL vs. discriminante, un análisis de riesgo al sistema financiero venezolano. *Ecos de Economía*, 25-47.
- De Lara Haro, A. (2005). *Medición y control de riesgos financieros*. México: LIMUSA.
- Delvasto, C. A. (2006). *Vitela Pontificia Universidad Javeriana Cali*. Obtenido de Vitela Pontificia Universidad Javeriana Cali:

[http://vitela.javerianacali.edu.co/bitstream/handle/11522/2319/Asimetrías\\_información\\_fallas.pdf?sequence=1](http://vitela.javerianacali.edu.co/bitstream/handle/11522/2319/Asimetrías_información_fallas.pdf?sequence=1)

Echeverri Valdes, F. (2006). *Universidad Nacional de Colombia*. Obtenido de Universidad Nacional de Colombia:

<http://www.bdigital.unal.edu.co/1083/1/faneryecheverrivaldes.2006.pdf>

Fernández Castaño, H., & Pérez Ramírez, F. O. (2005). El modelo logístico: Una herramienta estadística para evaluar el riesgo de crédito. *Revista Ingenierías Universidad de Medellín*, 55-75.

Ferrari, C. (2008). Tiempos de incertidumbre. Causas y consecuencias de la crisis mundial. *Revista de economía institucional, VOL. 10*, 55-78.

Flórez López, R. (2007). Análisis de los determinantes de crédito en presencia de carteras de bajo incumplimiento. Una nueva propuesta de aplicación. *Revista Europea de Dirección y Economía de la empresa*, 71-92.

Gutiérrez Girault, M. A. (Octubre de 2007). *Banco Central de la República Argentina*. Obtenido de Banco Central de la República Argentina: <http://www2.bcra.gov.ar/Pdfs/Publicaciones/CreditScoring.pdf>

Helizondo, A. (2012). *Medición integral del riesgo*. México: Limusa.

Hernández Sampieri, Fernández Collado, & Baptista, L. (2014). Selección de la muestra. En Hernández Sampieri, Fernández Collado, & L. Baptista, *Metodología de la Investigación* (págs. 170-191). Ciudad de México: McGraw-Hill.

Hernández Sampieri, R., Fernández Collado, C., & Baptista Lucio, P. (2010). *Metodología de la Investigación*. Mexico D.F: The McGraw-Hill Companies, Inc.

Hernandez, P. A. (diciembre de 2004). Obtenido de [https://www.emis.de/journals/RCE/V27/V27\\_2\\_139Cardona.pdf](https://www.emis.de/journals/RCE/V27/V27_2_139Cardona.pdf)

Instituto Nacional de Contadores Públicos . (31 de Mayo de 2015). *Instituto Nacional de Contadores Públicos* . Obtenido de Instituto Nacional de Contadores Públicos : <https://www.incp.org.co/niif-9-perdida-esperada-vs-perdida-incurrida/>

Laffaye, S. (Diciembre de 2008). *La crisis financiera: origen y perspectivas*. Obtenido de La crisis financiera: origen y perspectivas: <http://www.cei.gob.ar/userfiles/13%20La%20crisis%20financiera%20origen%20y%20perspectivas%20.pdf>

Llaugel, F., & Fernández, A. (2011). Evaluación del uso de modelos de regresión logística para el diagnóstico de instituciones financieras. *Red de revistas Científicas de América Latina y el Caribe, España y Portugal*.

Luzardo Briceño, M. (2006). Aplicación de un modelo EWMA para el análisis de control de calidad en la producción de aluminio primario: caso CVG-Venalum. *Economía, XXXI*, 41-55.

Machinea, J. L. (2009). La crisis Financiera internacional: su naturaleza y los desafíos de política económica. *Revista CEPAL 97*, 33-56.

Méndez Álbarez , C. E. (2011). *Metodología*. Mexico: LIMUSA S.A. Obtenido de <https://es.scribd.com/document/324262554/METODOLOGIA-DE-LA-INVESTIGACION-CARLOS-MENDEZ-1-pdf>

Méndez Álvarez, C. E. (2011). *Metodología Diseño y Desarrollo del proceso de investigación con énfasis en ciencias empresariales*. México: Limusa.

Meza Saldaña, E., Reyes Cervantes, H., Pérez Salvador, B. R., & Tajonar Sanabria, F. (2017). *BUAP (Universidad Autónoma de Puebla) FCFM* . Obtenido de BUAP (Universidad Autónoma de Puebla) FCFM : <https://www.fcfm.buap.mx/SIEP2017/Memorias/Extensos/Carteles/7.pdf>

Moreno Valencia , S. (2013). *Universidad nacional de colombia* . Obtenido de <http://www.bdigital.unal.edu.co/39466/1/43596322.2014.pdf>

- Moscote Flórez, O., & Rincón, W. (2012). Modelo Logit y Probit: un caso de aplicación. *Comunicaciones en estadística*, 123-133.
- Nancy Eugenia Zamudio Gómez . (2007). *Banco de la República*. Obtenido de Banco de la República:  
<http://www.banrep.gov.co/sites/default/files/publicaciones/pdfs/borra466.pdf>
- Ochoa P, J. C., Galeano M, W., & Agudelo V, L. G. (2010). Construcción de un modelo de scoring para el otorgamiento de crédito en una entidad financiera. *Perfil de Coyuntura económica*, 191-222.
- Ochoa P., J. C., Galeano M., W., & Agudelo V., L. G. (25 de octubre de 2010). Obtenido de  
<https://mail.google.com/mail/u/0/?tab=wm#search/diego+fernando+fern%C3%A1ndez/15b62d3d7d378ef3?projector=1>
- Osorio, J. A. (30 de Mayo de 2008). *Banco Central de Reserva de El Salvador*. Obtenido de Banco Central de Reserva de El Salvador:  
<http://www.bcr.gob.sv/bcrsite/uploaded/content/category/1910288814.pdf>
- Pérez Ramirez, F. O., & Fernández Castaño, H. (2007). Las redes neuronales y la evaluación del riesgo de crédito . *Revista Ingenierías Universidad de Medellín*, 78-91.
- Pineda Salido, L. (s.f.). La crisis financiera de los Estados Unidos y la respuesta regulatoria internacional. *Revista Aequias* , 129.214.
- Rodríguez Sánchez, M. T. (2001). Premio nobel de economía 2001: El libre mercado no funciona. *Momento Económico*, 51.
- Romo, H. G. (2009). La crisis financiera de los préstamos subprime. *Comercio Exterior*, Vol 59, NÚM.12, 947-969.

Saavedra García, M. L., & Saavedra García, M. J. (21 de mayo de 2010). Obtenido de  
<https://drive.google.com/drive/u/1/folders/0B1S92cziHt7IVDB4am5kcTIIOTQ>

Sagner T, A. (2012). El influjo de cartera vencida como medida de riesgo de credito: Análisis y aplicación al caso de Chile. *Revista de análisis económico*, 27-54.

Salazar Villano, F. E. (2013). cuantificación del riesgo de incumplimiento en créditos de libre inversión: un ejercicio econométrico para una entidad bancaria del municipio de Popayán, Colombia. *Estudios Gerenciales*, 416-427.

Sánchez Turcios, R. A. (2015). T-Student. Usos y abusos. *Revista Mexicana de Cardiología*, 59-61.

Superintendencia Financiera de Colombia. (1995). *Circular Básica Contable y Financiera (Circular externa 100 de 1995)*. Obtenido de Circular Básica Contable y Financiera (Circular externa 100 de 1995):  
<https://www.superfinanciera.gov.co/jsp/loader.jsf?IServicio=Publicaciones&ITipo=publicaciones&IFuncion=loadContenidoPublicacion&id=15466>

Támara Ayús, A., Aristizábal, R., & Velásquez, E. (2012). Matrices de transición en el análisis del riesgo crediticio como elemento fundamental en el cálculo de la pérdida esperada en una institución financiera colombiana. *Revista Ingenierías Universidad de Medellín*, 105-114.

Trigo Martínez, E. (2009). *Biblioteca universitaria Universidad de Malaga*. Obtenido de Biblioteca universitaria Universidad de Malaga:  
<http://www.biblioteca.uma.es/bbldoc/tesisuma/17968872.pdf>

Urrutia , M., & Namen, O. M. (2011). Historia del crédito hipotecario en Colombia. *Ensayos sobre política económica vol.30,Edición especial historia de la banca central*, 280-306.

Ustáriz González, L. H. (2003). El comité de Basilea y la supervisión bancaria. *VNIVERSITAS*, 431-462.

Velandia Velandia, N. (2013). *Universidad Nacional de Colombia*. Obtenido de Universidad Nacional de Colombia:  
<http://www.bdigital.unal.edu.co/10202/1/7709584.2013.pdf>

Velasco S, C. (08 de Marzo de 2010). *Biblioteca digital Universidad del Valle*. Obtenido de Biblioteca digital Universidad del Valle:  
<http://bibliotecadigital.univalle.edu.co/bitstream/10893/3685/4/CB-0449603.pdf>

Villano, F. E. (29 de noviembre de 2013). Obtenido de  
<https://drive.google.com/drive/u/1/folders/0B1S92cziHt7IVDB4am5kcTIIOTQ>

Zurita González, J., Martínez Pérez, J. F., & Rodríguez Montoya, F. (2009). La crisis financiera y económica del 2008. Origen y consecuencias en los Estados Unidos y México. *El Cotidiano*, 17-27.

## ANEXOS

## ANEXO A

Matriz de correlaciones

		Constante	CUOTASCANCELADA	OFICINA	C	CUOTASMORA	SALDODEUDA	VALORINICIAL	EDADMOORARA	TOTALCUCOTAS
Paso 1	Constante	1,000	-,902							
	CUOTASCANCELADAS	-,902	1,000							
Paso 2	Constante	1,000	-,725	-,355						
	CUOTASCANCELADAS	-,725	1,000	-,251						
	OFICINA	-,355	-,251	1,000						
Paso 3	Constante	1,000	-,585	-,475	-,852					
	CUOTASCANCELADAS	-,585	1,000	-,153	,244					
	OFICINA	-,475	-,153	1,000	,387					
	C	-,852	,244	,387	1,000					
Paso 4	Constante	1,000	-,614	-,460	-,837	-,493				
	CUOTASCANCELADAS	-,614	1,000	-,128	,300	,239				
	CUOTASMORA	-,493	,239	,097	,222	1,000				
	OFICINA	-,460	-,128	1,000	,401	,097				
	C	-,837	,300	,401	1,000	,222				
	Constante	1,000	-,639	-,464	-,825	-,495	,271			

Paso	SALDODEUDA	,271	-,264	-,066	-,084	-,812	1,000			
5	CUOTASCANCEL ADAS	-,639	1,000	-,105	,309	,347	-,264			
	CUOTASMORA	-,495	,347	,114	,196	1,000	-,812			
	OFICINA	-,464	-,105	1,000	,407	,114	-,066			
	C	-,825	,309	,407	1,000	,196	-,084			
Paso	Constante	1,000	-,442	-,454	-,814	-,460	,004	,099		
6	VALORINICIAL	,099	-,818	,086	,039	,250	-,940	1,000		
	SALDODEUDA	,004	,721	-,111	-,071	-,508	1,000	-,940		
	CUOTASCANCEL ADAS	-,442	1,000	-,136	,140	-,016	,721	-,818		
	CUOTASMORA	-,460	-,016	,150	,204	1,000	-,508	,250		
	OFICINA	-,454	-,136	1,000	,412	,150	-,111	,086		
	C	-,814	,140	,412	1,000	,204	-,071	,039		
Paso	Constante	1,000	-,170	-,053	-,220	-,304	,129	-,058	-,912	
7	EDADMOORA	-,912	-,012	-,141	-,123	,126	-,140	,110	1,000	
	VALORINICIAL	-,058	-,817	,072	,025	,254	-,941	1,000	,110	
	SALDODEUDA	,129	,717	-,093	-,054	-,512	1,000	-,941	-,140	
	CUOTASCANCEL ADAS	-,170	1,000	-,138	,142	-,014	,717	-,817	-,012	
	CUOTASMORA	-,304	-,014	,131	,187	1,000	-,512	,254	,126	
	OFICINA	-,053	-,138	1,000	,414	,131	-,093	,072	-,141	
	C	-,220	,142	,414	1,000	,187	-,054	,025	-,123	
Paso	Constante	1,000	,029	-,062	-,218	-,026	,155	-,076	-,911	-,141
8	EDADMOORA	-,911	-,066	-,135	-,122	,000	-,147	,115	1,000	,072
	VALORINICIAL	-,076	-,542	,075	,023	,030	-,932	1,000	,115	,129

SALDODEUDA	,155	,592	-,097	-,046	-,037	1,000	-,932	-,147	-,265
TOTALCUOTAS	-,141	-,844	,050	,004	-,860	-,265	,129	,072	1,000
CUOTASCANCEL ADAS	,029	1,000	-,117	,074	,715	,592	-,542	-,066	-,844
CUOTASMORA	-,026	,715	,022	,086	1,000	-,037	,030	,000	-,860
OFICINA	-,062	-,117	1,000	,416	,022	-,097	,075	-,135	,050
C	-,218	,074	,416	1,000	,086	-,046	,023	-,122	,004

**PROBABILIDAD DE INCUMPLIMIENTO DE UNA CARTERA CASTIGADA, EN LA MODALIDAD DE MICROCRÉDITO, PARA UNA ENTIDAD DEL SISTEMA FINANCIERO COLOMBIANO EN EL AÑO 2017**

*Lina Marcela Bolaños Martínez*

*María Camila Fernández Bolaños*

## **RESUMEN**

Para el desarrollo del presente artículo se tomó información de la base de datos de una cartera de microcrédito perteneciente a una entidad financiera colombiana, dicha base contiene información de obligaciones. Al tomar esta información se busca determinar la Probabilidad de incumplimiento (PI) de mencionada cartera, apoyados en una estimación econométrica, de forma particular, un modelo de Regresión Logística Binaria (Logit). Dentro de los primeros pasos para realizar este estudio, se realizó un análisis descriptivo de las variables, para comprender las relaciones entre las predictores y la variable dependiente, seguido se corrió el modelo de regresión logística bajo el método “hacia adelante de Wald” con el programa IBM SPSS, se diseñó el modelo, se aplicó una prueba Backtesting y finalmente se realizó un análisis descriptivo de los resultados que arrojó el modelo.

Se debe tener en cuenta el siguiente comentario: La población objeto de estudio tenía la característica, eran microcréditos castigados, en este sentido el trabajo presentó un reto adicional al tener una población con alta probabilidad de no pagar.

Palabras clave: regresión logística, modelo logit, credit scoring, default, backtesting, microcrédito.

## **INTRODUCCIÓN**

El sistema financiero es muy importante para el desarrollo y crecimiento de una economía, por tanto, se debe estar atento a su buen funcionamiento y administración por parte de las entidades financieras, puesto que existen diversos tipos de riesgo financiero como es el de liquidez, crédito, operativo, legal y de mercado; entre todos ellos el enfoque principal de este trabajo está en el riesgo de crédito, más precisamente en entidades con cartera de microcrédito. La importancia de este sector radica en que permite la concesión de financiamiento a una mayor cantidad de personas debido a requerimientos más flexibles para la obtención del crédito, convirtiéndose en una herramienta de inclusión social para los pequeños emprendedores, “el microcrédito constituye una de las herramientas más significativas para asignar capital...en estos negocios de reducida envergadura, refuerza su productividad y eficiencia, y contribuye a la elevación de los niveles de ingreso y empleo en los sectores de menores recursos” (Calixto Salazar & Casaverde Carranza, 2011, pág. 1).

“Después de las crisis financieras de 1999 y 2008, en Colombia y Estados Unidos respectivamente, se hace evidente que el sistema financiero debe ser regulado de tal manera que corrija las asimetrías de información, las cuales el mercado por sí

solo no corrige” (Ochoa P, Galeano M, & Agudelo V, 2010, pág. 192). Entonces, debido a sucesos como los mencionados, que han dejado al descubierto la inestabilidad y la globalización de mercados financieros, se hizo evidente la necesidad de la regulación financiera, por tanto, se creó el Comité de Regulación Bancaria de Basilea donde se dan lineamientos de regulación internacional que prontamente se convirtieron en pautas generales seguidas por entidades financieras a nivel mundial. En este orden se creó Basilea I en 1988 para resaltar la importancia de mantener un capital mínimo con el que se cubran los capitales sujetos al riesgo de posibilidad de impago; en 2004 es aprobado el acuerdo conocido como Basilea II, donde se concluye que “las entidades financieras y de microfinanzas (IMFs) supervisadas requieren adoptar procesos internos que sean capaces de medir el riesgo de crédito” (Cantón, Lara Rubio, & Camino Blasco, 2010, pág. 90), por tanto se vuelve importante disponer de modelos de medición (scorings) con el fin de discriminar a los clientes según su perfil de riesgo; además de esto la calidad de información y las grandes bases de datos se vuelven indispensables para cuantificar el riesgo.

En este sentido, el presente estudio tiene como finalidad determinar la probabilidad de incumplimiento y los factores que influyen en ella, en una cartera de microcrédito perteneciente a una entidad financiera del país, para clasificar los clientes como buenos o malos pagadores. Es importante resaltar que para el desarrollo del trabajo se tomó una base de datos cuya cartera de microcrédito se encuentra castigada, es decir, está conformada por créditos que la entidad

considera de difícil cobro. Dicha cartera pertenece a una casa de cobranza con cobertura nacional.

El cuerpo del trabajo está dividido en cuatro capítulos, en el primero se definen conceptos relevantes para la investigación como lo es el riesgo de crédito, la asimetría de la información, selección adversa y riesgo moral, se señalan las diferentes interpretaciones que ha tenido el riesgo de crédito a través de las teorías económicas. Así mismo se hace referencia a los acuerdos de Basilea I y Basilea II de manera conjunta con el sistema de Administración del Riesgo Crediticio (SARC) vigente en Colombia. También se explican los modelos de credit scoring más relevantes como son: Análisis discriminante, logit o regresión logística, redes neuronales y árboles de decisión, se destacan antecedentes del modelo scoring y se explica la fórmula de logit.

En el segundo capítulo se enseña la metodología empleada para la aplicación del modelo. En el capítulo tercero se comparten los resultados del estudio: análisis descriptivo de la probabilidad de incumplimiento, cruce de variables, aplicación del modelo y prueba backtesting. Por último, se encuentra el capítulo cuarto con las conclusiones y recomendaciones.

## **MARCO TEÓRICO Y REFERENTES CONCEPTUALES.**

### **Bases teóricas**

En finanzas el riesgo es “la posibilidad de que se sufra un perjuicio de carácter económico, ocasionado por la incertidumbre en el comportamiento de variables económicas a futuro” (Avila Bustos, 2005, pág. 6). De acuerdo a la

cartilla No.2 coordinada por el (Autorregulador del Mercado de Valores de Colombia, 2014, pág. 6) se encuentra clasificado en cuantificable y no cuantificable; dentro de los cuantificables se haya el riesgo de mercado, liquidez y crédito mientras que en los no cuantificables está el operacional, legal, reputacional y estratégico. Ahora, el enfoque de este trabajo va dirigido al riesgo de crédito que se define como “la posibilidad de que una entidad incurra en pérdidas y se disminuya el valor de sus activos, como consecuencia de que un deudor o contraparte incumpla sus obligaciones” (Superintendencia Financiera de Colombia, 1995, pág. 2). De igual modo, en este tipo de riesgo se considera muy importante la información, todos los negocios requieren de ella pero de forma especial los establecimientos de crédito, y esto se debe a la naturaleza propia de estos agentes en el que transfieren recursos de un sector superavitario a uno deficitario; en el manejo de dicha información se presenta un problema denominado “asimetría de la información” (Delvasto, 2006, pág. 218). Esta se muestra cuando los participantes del negocio bancario no cuentan con información o es insuficiente, impidiendo que se tomen decisiones acertadas y por ello se incurra en costos de transacción (Delvasto, 2006, pág. 220). Estas fallas traen los denominados efectos de selección adversa y riesgo moral; la selección adversa, según (Rodríguez Sánchez, 2001, pág. 51) ocurre “cuando debido a la asimetría de información no logran diferenciar un buen proyecto de uno malo” por tanto, la falta de información puede llevar a optar por la opción equivocada. El riesgo moral por otro lado, surge una vez desembolsados

o entregados los recursos, en aquel momento hay cambio de incentivos en el prestatario, por ello podrá tomar diferentes actitudes dependiendo de cómo sea su comportamiento con respecto al riesgo (Delvasto, 2006, pág. 221), es decir que la persona puede decidir no pagar.

### **Teorías económicas**

Desde el siglo XVIII el riesgo de crédito ha tenido diferentes interpretaciones y se hace necesario nombrar algunas teorías que se enfocaban en ello a partir de la disciplina económica. Desde el punto de vista de la teoría neoclásica, en el estudio de (Salazar Villano, 2013, pág. 417) se expresa que no existe el riesgo de crédito pues no deberían haber pérdidas si se asume el supuesto de competencia perfecta, las pérdidas serían generadas por la falta de flexibilidad en su estructura de costos, “...por otro lado, si se habla de un escenario de competencia no perfecta, las pérdidas bancarias surgirían de las estrategias de juego de las firmas (bancos)”. De acuerdo a Adam Smith (1776), citado por (Salazar Villano, 2013, pág. 417), “el tipo de interés corriente más bajo debe ser algo más que suficiente para compensar las pérdidas ocasionales a las que los prestamos están expuestos”. En la teoría general de Keynes (1936), mediante el concepto de “eficiencia marginal del capital” se explica el riesgo de crédito de acuerdo a (Salazar Villano, 2013, pág. 418) teniendo en cuenta las expectativas frente a ingresos futuros, también da lugar al azar moral, es decir que el prestatario no posea la voluntad de pagar o el incumplimiento involuntario que dejaría en evidencia el fallo de las expectativas.

Por otro lado, de la perspectiva neoinstitucionalista se rescata la teoría de la información de Stigler (1961), quien la define como un “bien económico” que lleva inmersos costos de producción que en el caso de las instituciones bancarias se ven reflejados en todos los métodos usados para llevar al cliente la información sobre la adquisición del crédito, “...en este sentido todo desajuste de la información generaría riesgo en el mercado de crédito” (Salazar Villano, 2013, pág. 418)

### **Regulación bancaria.**

Una vez vistas algunas teorías económicas que estudian el riesgo de crédito, se hace necesario conocer las instituciones y/o acuerdos más relevantes que realizan sugerencias a todas las entidades crediticias para el manejo y control del riesgo. Uno de ellos es el Comité de Supervisión Bancaria de Basilea fundado en 1974, cuyo objetivo es formular estándares de supervisión y recomendaciones de práctica de buen gobierno en las entidades financieras (Pérez Ramirez & Fernández Castaño, 2007, pág. 79). Este comité público el Acuerdo de Capitales de Basilea en 1988, conocido como Basilea I, en donde se hacen las recomendaciones necesarias para mantener un capital mínimo con el que se cubran los capitales sujetos al riesgo de posibilidad de impago, el cual se llamó capital mínimo regulatorio (Ochoa P, Galeano M, & Agudelo V, 2010, pág. 195); más adelante se creó Basilea II fue hecho público en 2004 y obtuvo una versión más completa en 2006; este acuerdo recomienda la gestión del sistema financiero basado en tres pilares:

4. Requisitos de capital mínimo:

cubrimiento de capital en riesgo

5. Proceso de examen supervisor: donde el ente supervisor cumple un papel primordial en la vigilancia y supervisión de la administración por parte de las entidades financieras.

6. La disciplina de mercado: acceso y transparencia de la información suministrada por las entidades financieras. (Ochoa P, Galeano M, & Agudelo V, 2010, pág. 195)

Pasando del ámbito internacional al nacional, en Colombia la Superintendencia Financiera de Colombia (SFC) mediante la Carta Circular 31 y la Circular Externa 11 de 2002, exige que todas las instituciones financieras implementen un sistema de Administración del Riesgo Crediticio (SARC); así mismo, la circular externa 100 de 1995, capítulo 2 de la (Superintendencia Financiera de Colombia, 1995, pág. 3) contempla los siguientes elementos básicos del SARC “Políticas de administración del RC, procesos de administración del RC, modelos internos o de referencia para la estimación o cuantificación de pérdidas esperadas, sistema de provisiones para cubrir el RC, procesos de control interno”. Dentro de la definición anterior se resalta la especificación de modelos internos ya que son ellos los que clasifican a los clientes en buenos y malos de acuerdo al pago o impago.

## Modelos para establecer la probabilidad de impago.

En seguida se nombrarán algunos de los modelos de credit scoring entendidos como “algoritmos que de manera automática evalúan el riesgo de crédito de un solicitante de financiamiento o de alguien que ya es cliente de la entidad” (Gutiérrez Girault, 2007, pág. 3), más utilizados e importantes, a tener en cuenta a través de la historia:

En primer lugar se encuentra el *Análisis discriminante*; que de acuerdo a (Cantón, Lara Rubio, & Camino Blasco, 2010, pág. 91) clasifica buenos y malos pagadores en el momento de realizar un crédito, por medio de la diferenciación de características que definen a cada grupo. Entre sus inconvenientes se encuentra la rigidez para cumplir las hipótesis de linealidad, normalidad, homocedasticidad e independencia.

En seguida se encuentra el modelo *logit o de regresión logística*, este permite calcular la probabilidad de que el individuo pertenezca a un grupo (cumplido o incumplido), esta clasificación se hace de acuerdo a las variables independientes que explican el resultado de “Y” (variable dependiente), en este modelo “...no es necesario plantear hipótesis de partida, como por ejemplo la normalidad de la distribución de las variables... presenta la ventaja de...mantener la variable explicada siempre dentro de un rango de variación entre cero y uno” (Cantón, Lara Rubio, & Camino Blasco, 2010, pág. 92).

Por otro lado, se encuentra, el modelo de *Redes neuronales*, que emula la estructura y el comportamiento del cerebro; la red está formada por procesadores simples, llamados nodos,

que están interconectados entre sí. “...La finalidad de cada nodo es dar respuesta a una determinada señal de entrada...” sin embargo, “... el uso de esta técnica resulta complicado pues el proceso interno de aprendizaje funciona como caja negra, donde la comprensión de lo que ocurre dentro demanda de conocimientos especializados” (Cantón, Lara Rubio, & Camino Blasco, 2010, pág. 92).

Por último, (Flórez López, 2007, pág. 77) establece que los modelos de *árboles de decisión* son de clasificación jerárquica y secuencial que dividen un conjunto de datos (N), cada árbol integra tres componentes básicos: los nodos de decisión, las ramas y los nodos terminales (hojas); el orden en que van surgiendo los nodos es debido a la importancia de las variables explicativas. Un punto en contra del modelo es que llevan una difícil comprensión interna (Cantón, Lara Rubio, & Camino Blasco, 2010, pág. 93).

Por otro lado, (Moscote Flórez & Rincón, 2012, pág. 127) resaltan las ventajas que posee el modelo, aseverando que “Es una función flexible y fácil de utilizar, tiene una interpretación relativamente sencilla, la evidencia empírica ha demostrado que este modelo es adecuado en la mayoría de los casos en los cuales la respuesta es binaria”.

### Formula Logit

La forma específica del modelo de regresión logística es:

*Ecuación 6 Formula logit*

$$\pi(x) = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}} = \frac{e^k}{1 + e^k} \quad \text{donde } k = \beta_0 + \beta_1 x$$

Fuente: (Llaugel & Fernández, 2011, pág. 599)

“El modelo logit permite además de estimar la probabilidad de incumplimiento, identificar los factores de riesgo que determinan las probabilidades, así como el peso relativo de ellos sobre las probabilidades” (Echeverri Valdes, 2006, pág. 83)

En el análisis de regresión se reflejan las siguientes premisas, cuando se cuenta con una variable de respuesta dicotómica:

a) La media condicional de la ecuación de regresión debe ser formulada para estar contenida entre 0 y 1.

b) La distribución binomial y no la normal, describe la distribución de los errores y será la distribución en la que estará sustentado el análisis.

c) Los principios que guían un análisis de regresión lineal, son los mismos para el análisis de regresión logística. (Llaugel & Fernández, 2011, pág. 600)

Para el modelo se deben filtrar las variables de modo que queden solamente las que tengan un verdadero grado de significancia, (Velandia Velandia, 2013, pág. 56) afirma que para ello “es preciso calcular la estadística de Wald, para lo cual se debe observar la escala de medición de la variable ya que esta puede ser categórica o no categórica”.

## METODOLOGÍA

La base de datos que se utilizará para el trabajo, fue facilitada por una entidad financiera, de la cual por razones de privacidad y seguridad no se revelará el nombre. Para la aplicación del modelo y la inserción de variables se utilizara el programa estadístico SPSS. Esta base de datos contiene un total de 18.753 créditos pertenecientes a cuatro carteras

castigadas, es decir, obligaciones de difícil cobro. La información pertenece al periodo entre julio de 2016 a diciembre de 2017, con un total de 16 variables, entre la dependiente y las predictores.

Para evaluar la capacidad predictiva del modelo y su ajuste a la realidad, se realizan distintas medidas de desempeño, tales como:

- **La R cuadrado de Cox y Snell:** De acuerdo a (Alderete, 2006, pág. 58) en programas estadísticos como el SPSS se computa de la siguiente manera:

Dónde:  $-2LL$  (nulo) es la desviación del modelo nulo solo o con una constante, sin tener en cuenta variables predictoras,  $-2LL$  (modelo) es la desviación del modelo con las variables predictoras y  $N$  es el tamaño de la muestra.

*Ecuación 2 R cuadrado de Cox y Snell*

$$R^2_L = 1 - \left[ \frac{-2LL_{(nulo)}}{-2LL_{(modelo)}} \right]^{2/N}$$

- **R cuadrado Nagelkerke:** “versión corregida del R-cuadrado de Cox y Snell con valor máximo igual a 1” (Meza Saldaña, Reyes Cervantes, Pérez Salvador, & Tajonar Sanabria, 2017, pág. 4).

*Ecuación 3 R cuadrado Nagelkerke*

$$\bar{R}^2_L = \frac{1 - \left[ \frac{-2LL_{(nulo)}}{-2LL_{(modelo)}} \right]^{2/N}}{1 - (2LL_{((modelo))}^{2/N}}$$

Fuente: (Alderete, 2006, pág. 59)

- **Bondad de ajuste de Hosmer-Lemeshow:** se construyen tablas para comparar los resultados de

estimación del modelo frente a los reales en la muestra. “Se basa en agrupar los casos en deciles de riesgo y comparar la probabilidad observada con la probabilidad esperada dentro de cada decil” (Velasco S, 2010, pág. 16). Su valor debe ser mayor a 0,05 de significancia, sin embargo, de acuerdo al artículo de investigación de (Cantón, Lara Rubio, & Camino Blasco, 2010) aún si este indicador tiene un nivel de significancia bajo, “se debe tomar en cuenta que este dato estadístico solamente brinda pistas sobre la calidad del ajuste del modelo”, por tanto se puede decir que el modelo no pierde su validez.

- 
- **Área bajo la curva ROC (Receiver Operating Characteristic):** Es un gráfico en el que se observan los pares sensibilidad/especificidad resultante de la variación continúa de los puntos de corte en todo el rango de resultados observados. En el eje “y” se sitúa la sensibilidad (verdaderos positivos), en el eje “x” los falsos positivos (1-especificidad) (Meza Saldaña, Reyes Cervantes, Pérez Salvador, & Tajonar Sanabria, 2017, pág. 5).

Cuando exista una proporción correcta de aciertos frente al Default se dirá que el modelo tiene alta Sensibilidad; y cuando exista una proporción de aciertos correctos de No-default se dirá que el modo goza de alta especificidad (falsas alarmas), que es lo que se busca en un modelo adecuado o correcto.

#### **Variables del modelo**

*Variable dependiente categórica dicotómica:* se denotan solo dos opciones como resultado, ya sea 1= buen cliente (no default) o 0=Mal cliente (Default).

*Variables independientes del modelo:*

- Variable categórica: Cartera, departamento, estado civil, nivel educativo.
- Variable numérica: fecha apertura, operación, edad mora, valor inicial, saldo deuda, cuota mensual, total cuotas, cuotas canceladas, cuotas mora, oficina, actividad económica.

#### **Selección de la muestra**

En la determinación de la muestra implementada en el modelo de credit scoring se tomaron en cuenta los siguientes puntos:

Se seleccionó la muestra de modo que el número de casos de crédito de impago se aproximara al de pago, ya que se poseía una población donde predominaban enormemente los créditos de impago, al ser una cartera castigada. De esta manera se buscó eliminar el problema de diferencia de tamaños de grupos que “generalmente, provoca una estimación inexacta al no definirse correctamente los factores de clasificación hacia los grupos previamente definidos” Mures, García y Vallejo (citado de Cantón, Lara Rubio, & Camino Blasco, 2010). De la población se sacaron 4.501 créditos a través de la muestra probabilística estratificada que es aquel “muestreo en el que la población se divide en segmentos y se selecciona una muestra para cada segmento” (Hernández Sampieri, Fernández Collado, & Baptista, 2014, pág. 181), en

este caso se desarrolló por medio del programa IBM SPSS, para tomar proporciones un poco similares entre los grupos de personas que pagaban y las que no; y así obtener mejores resultados.

Los créditos deben obtener la mayor cantidad de información completa sobre todas las variables explicativas expuestas. Por tanto, se eliminaron casos en los que había grandes faltantes de información, aunque quedaron algunos representativos en la variable Departamento.

## RESULTADOS DEL MODELO

Para la aplicación se cuenta con un total de 4.501 clientes de microcréditos concedidos entre julio de 2016 a junio de 2017 de una entidad financiera colombiana. La población contiene la variable respuesta (dependiente) DEFAULT y 16 variables independientes que describen las características del cliente y su comportamiento crediticio que permiten explicar el grado de incumplimiento.

La variable dependiente indica si el cliente alcanza el estado de morosidad o no. Es una variable dicotómica, donde los deudores que cayeron en incumplimiento se codificaron con el valor 0 mientras que los otros tomaron el valor de 1.

### Análisis descriptivo

La cartera se encuentra distribuida de la siguiente manera: la cuarta compra obtiene mayor representación con 1.760 créditos que equivalen a 39,1%, seguida de la tercera compra con 1.747 créditos equivalentes a 38,8%, en tercer lugar se encuentra la prima compra con 730 créditos que representan el 16,2%, por último la de menor representación es la

segunda compra con un total de 264 créditos que equivalen a un 5,9%.

En la variable departamento se puede identificar que los de mayor concentración de créditos en su respectivo orden son: Córdoba con 328 créditos, Atlántico con 326 y Valle con 213. Y los de menor concentración son Boyacá, Putumayo y Caquetá con 9, 11 y 17 créditos respectivamente.

El promedio de días de atraso son 344 días, siendo 596 el máximo de días de atraso y 16 el mínimo. Por su parte, la cantidad de días de mora más recurrentes entre los clientes son 323.

En cuanto al nivel educativo predominan los clientes con nivel de educación secundaria (2.303 clientes) con una representación del 51,2% del total de los créditos, seguidos de la población con educación primaria que reflejan el 28% con un total de 1262 clientes, por último, se encuentran los de menor participación como: los de educación técnica, superior y sin instrucción con una distribución de 8%, 3,9% y 2,6% respectivamente.

Finalmente, el estado civil de las personas que solicitaron más créditos a la entidad bancaria son los de unión libre con 1935 clientes representando el 43%, seguido de los solteros con 1668 equivalentes al 37,1%, después los casados con 564 clientes que equivalen a un 12,5% y por último se encuentran los clientes con menor representación en la cartera que son los divorciados y los viudos con una participación del 5,3% y 2,2% respectivamente.

### Modelo de Regresión logística (Logit)

*Tabla 29: resumen del modelo*

Resumen del modelo			
Paso	Logaritmo de la verosimilitud	R cuadrado de Cox y Snell	R cuadrado de Nagelkerke
8	-2	,464	,651
	2803,112 <sup>a</sup>		

El valor de R cuadrado de Nagelkerke indica que el modelo propuesto explica el 65,1% de la varianza de la variable dependiente (,651), por tanto se posee un nivel de ajuste aceptable en regresión logística.

Tabla 30: prueba del Hosmer Lemeshow

Prueba de Hosmer y Lemeshow			
Paso	Chi-cuadrado	gl	Sig.
5	26,341	8	,001
6	36,381	8	,000
7	53,323	8	,000
8	51,048	8	,000

Esta prueba indica que tan significativa es la varianza explicada del R2 para la variable dependiente. Para este modelo la prueba indica que la varianza explicada por el modelo no es satisfactoria pues el grado de significancia es bajo. De acuerdo al artículo de investigación de (Cantón, Lara Rubio, & Camino Blasco, 2010) a pesar de su nivel de significación bajo, “se debe tomar en cuenta que este dato estadístico solamente brinda pistas sobre la calidad del ajuste del modelo”.

Tabla de clasificación: En el anexo A se puede ver la tabla completa.

Tabla 31: tabla de clasificación

Paso	LOGIT1	0	2944	137	95,6
8	AÑO	1	367	1053	74,2
	Porcentaje global				88,8

Para el análisis de regresión logística el bloque 1 indica que hay un 88,8% de probabilidad de acierto en el resultado de la variable dependiente.

Variabes en la ecuación: en el anexo B se pueden ver las variables que el programa encontró como mayormente explicativas y por lo tanto las dejará en la ecuación, por su mayor grado de significancia estadística y por tener valores de Wald que se consideraban realmente explicativos para el resultado de la variable dependiente. Allí se denotan las betas, error estándar, valor estadístico de Wald, la significancia estadística y el valor de Exp (B). Las variables que dejó son: cuotas canceladas, oficina, C (Cartera), cuotas mora, saldo de la deuda, valor inicial, edad de mora y total cuotas.

Matriz de correlaciones: En el anexo C se muestran las correlaciones entre las variables para percibir el grado de relación que poseen entre ellas. De acuerdo con la matriz en el quinto paso no se encuentran variables relacionadas o superiores a 0,5, todas tienen un valor inferior e incluso negativo, lo que quiere decir que estas variables tienen relación inversa. El hecho de que no se encuentren correlacionadas es positivo para el modelo ya que no se quieren variables que digan lo mismo entre sí.

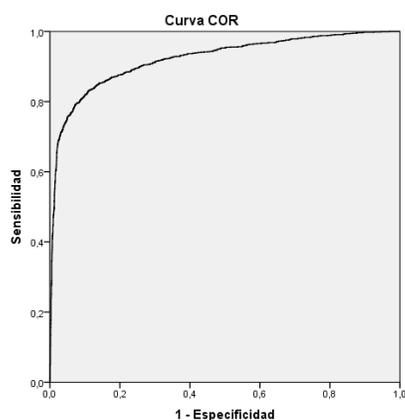
### Curva COR

La curva COR o ROC identifica la capacidad de diagnóstico que tiene el

modelo de regresión logística para clasificar entre los clientes que pagan y los que no. A la proporción de predicciones correctas de “no pago” (Default) se le dice sensibilidad (aciertos) y a la proporción de predicciones acertadas de “pago” (no default) se les llama especificidad (falsas alarmas).

La curva ROC para el modelo logit indica que este modelo tiene un buen poder discriminatorio, puesto que el trazo es inclinado a la línea horizontal, de hecho, el área bajo la curva (AUROC) es 0,922.

Ilustración 20: curva ROC



## Prueba Backtesting

Ilustración 21: prueba backtesting

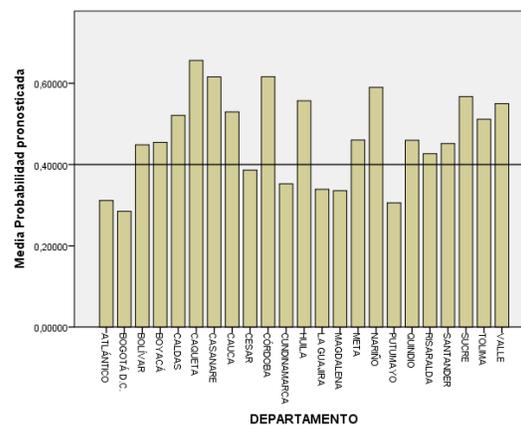
PRUEBA BACKTESTING				
ACIERTOS DEL MODELO			NO ACIERTOS DEL MODELO	
PAGO	185	52,6%	167	47,4%
IMPAGO	10218	61,3%	6461	38,7%
TOTAL ACIERTOS	10403		6628	

Teniendo en cuenta que se asume que la muestra selecciona va a comportarse de la misma manera que la población, en el presente cuadro se muestran los resultados de la prueba backtesting realizada para toda la cartera. De los 352 créditos (como se menciona en el cuadro anterior) que pagan en los seis meses, el modelo acierta a 185 que corresponden a un 52,6% por lo tanto, no acertó en el 47,4% del total de pagos.

Aunque los porcentajes de acierto puedan parecer pequeños, es necesario reconocer que esto es debido a la población que desde un principio se encontraba sesgada, en la gran superioridad de los créditos en estado de impago frente al no default, debido al estado de deterioro de la cartera. Por tanto, al equilibrar la muestra, se presentaba dificultad para que los resultados del modelo se reflejaran a toda la población.

## Análisis de los resultados del modelo

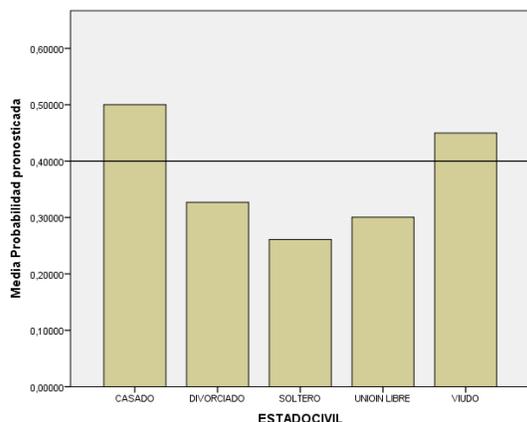
Ilustración 22: departamento y probabilidad de incumplimiento



En presente grafico se puede evidenciar que los departamentos que incurren mayormente a cumplimiento de las obligaciones de crédito son: Caquetá, Casanare, Córdoba, Nariño y Sucre, entre

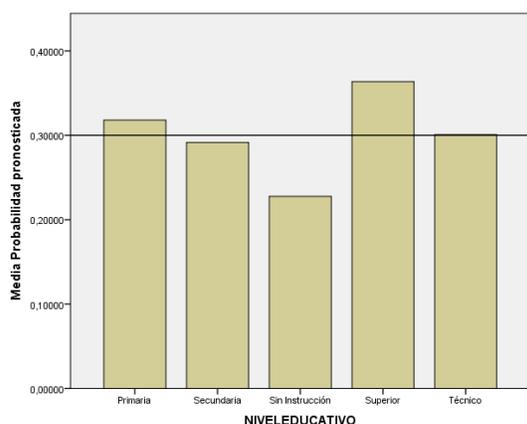
otros y los departamentos que incurren al incumplimiento de las obligaciones de crédito son: Bogotá, Putumayo, Magdalena y la Guajira, entre otros.

*Ilustración 23: estado civil y probabilidad de incumplimiento*



En el presente grafico se puede evidenciar que las personas con estado civil casados y viudos son los que más incurren a cumplimiento de las obligaciones crediticias y por lo contrario los de estado civil divorciados, unión libre y solteros son mayormente propensos al incumplimiento de las obligaciones crediticias.

*Ilustración 24: nivel educativo y probabilidad de incumplimiento*



Las personas con nivel educativo superior y primaria son los que más incurren al

cumplimiento de las obligaciones de crédito y por lo contrario las personas que más incurren al incumplimiento son los de nivel educativo técnico, secundaria y sin instrucción.

## CONCLUSIONES Y RECOMENDACIONES

### Conclusiones

A lo largo del desarrollo del trabajo se ha podido identificar que las instituciones se enfrentan de forma inminente al riesgo de crédito y es por esto que la Superintendencia Financiera de Colombia, crea el SARC, buscando adaptarlo en cada entidad para gestionar conjuntamente el riesgo. Esto se da debido a que se reconoce que el activo más representativo de toda entidad financiera es la cartera, por tanto debe cuidarse, administrarse y desarrollar seguimiento de la mejor manera posible. Para tal efecto existen los modelos de credit scoring como el logit que en este caso permite reconocer los créditos con mayor propensión a pago e impago, para así reconocer qué suma se puede recuperar de ellos y obtener una buena rentabilidad para el establecimiento.

Las variables con mayor poder explicativo según el modelo son: Departamento, total cuotas, nivel educativo, cartera y edad mora (días de atraso).

Las variables con mayor poder explicativo según el modelo son: cuotas canceladas, oficina, cartera, cuotas en mora, saldo de la deuda, valor inicial, edad de mora y total cuotas.

El modelo de regresión logística demostró ser considerablemente predictivo en cuanto a la clasificación en la probabilidad de impago con un porcentaje global del

88,8%, con una probabilidad de clasificar bien a los buenos pagadores del 74,2%, y una curva ROC para el modelo logit cuyo trazo es cercano a la línea horizontal que indica que este modelo tiene un buen poder discriminatorio (sensibilidad-especificidad), de hecho, el área bajo la curva (AUROC) es 0,922.

La prueba backtesting realizada de seis meses, respalda positivamente los resultados del modelo, a pesar de acertar el 52,6% del pago de los créditos realizados en el semestre. Pues este resultado ya estaba condicionado por el caso particular de la población sesgada en el default y desbalanceada.

De acuerdo a los resultados de la probabilidad de incumplimiento arrojados por el modelo Logit, las personas con mayor propensión a pago, pertenecen a los siguientes departamentos: Caquetá, Casanare, Córdoba y Nariño. Así mismo el estado civil: soltero y viudo. Los de nivel educativo superior y primaria, también tienen mayor probabilidad de pago. Finalmente, las carteras 2 y 3 están más próximas al pago.

### Recomendaciones

El anterior modelo ayuda a construir y a acercarse a la probabilidad de incumplimiento de los créditos en la cartera a pesar de la falta de información, por tanto es recomendable que en otra investigación se mejore el modelo con inclusión de nueva información y quizás variables macroeconómicas que permitan mejorar la robustez del mismo.

Es importante propender por un sistema de información más efectivo, ya que la información de los clientes es el principal

insumo del modelo para poder catalogar los créditos de acuerdo a su perfil de riesgo como buenos o malos pagadores.

### Bibliografía

Alderete, A. M. (2006). Fundamentos del Análisis de Regresión Logística en la Investigación Psicológica. *Revista Evaluar*, 52-67.

Autorregulador del Mercado de Valores de Colombia. (Mayo de 2014). *Conozca los riesgos del mercado de valores*. Obtenido de Conozca los riesgos del mercado de valores: <http://www.amvcolombia.org.co/attachments/data/20120605120911.pdf>

Avila Bustos, J. C. (Agosto de 2005). *Pontificia Universidad Javeriana Bogotá*. Obtenido de Pontificia Universidad Javeriana Bogotá: <http://www.javeriana.edu.co/biblos/tesis/economia/tesis01.pdf>

Calixto Salazar, M. M., & Casaverde Carranza, L. F. (2011). *Repositorio de la Universidad del Pacífico*. Obtenido de Repositorio de la Universidad del Pacífico: [http://repositorio.up.edu.pe/bitstream/handle/11354/1056/Mar%C3%ADa\\_Tesis\\_maestria\\_2011.pdf?sequence=1&isAllowed=y](http://repositorio.up.edu.pe/bitstream/handle/11354/1056/Mar%C3%ADa_Tesis_maestria_2011.pdf?sequence=1&isAllowed=y)

Cantón, S. R., Lara Rubio, J., & Camino Blasco, D. (Junio de 2010). *Un Modelo de Credit Scoring para instituciones de microfinanzas en el marco de Basilea II*. Obtenido de Un Modelo de Credit Scoring para instituciones de microfinanzas en el marco de Basilea II: <file:///C:/Users/USER/Downloads/RIESGO%20DE%20CREDITO%20CARTERA%20CLIENTES%202.pdf>

- Delvasto, C. A. (2006). *Vitela Pontificia Universidad Javeriana Cali*. Obtenido de Vitela Pontificia Universidad Javeriana Cali:  
[http://vitela.javerianacali.edu.co/bitstream/handle/11522/2319/Asimetrías\\_información\\_fallas.pdf?sequence=1](http://vitela.javerianacali.edu.co/bitstream/handle/11522/2319/Asimetrías_información_fallas.pdf?sequence=1)
- Echeverri Valdes, F. (2006). *Universidad Nacional de Colombia*. Obtenido de Universidad Nacional de Colombia:  
<http://www.bdigital.unal.edu.co/1083/1/faneryecheverrivaldes.2006.pdf>
- Flórez López, R. (2007). Análisis de los determinantes de crédito en presencia de carteras de bajo incumplimiento. Una nueva propuesta de aplicación. *Revista Europea de Dirección y Economía de la empresa*, 71-92.
- Gutiérrez Girault, M. A. (Octubre de 2007). *Banco Central de la República Argentina*. Obtenido de Banco Central de la República Argentina:  
<http://www2.bcra.gob.ar/Pdfs/Publicaciones/CreditScoring.pdf>
- Llaugel, F., & Fernández, A. (2011). Evaluación del uso de modelos de regresión logística para el diagnóstico de instituciones financieras. *Red de revistas Científicas de América Latina y el Caribe, España y Portugal*.
- Meza Saldaña, E., Reyes Cervantes, H., Pérez Salvador, B. R., & Tajonar Sanabria, F. (2017). *BUAP (Universidad Autónoma de Puebla) FCFM*. Obtenido de BUAP (Universidad Autónoma de Puebla) FCFM :  
<https://www.fcfm.buap.mx/SIEP2017/Memorias/Extensos/Carteles/7.pdf>
- Moscote Flórez, O., & Rincón, W. (2012). Modelo Logit y Probit: un caso de aplicación. *Comunicaciones en estadística*, 123-133.
- Ochoa P, J. C., Galeano M, W., & Agudelo V, L. G. (2010). Construcción de un modelo de scoring para el otorgamiento de crédito en una entidad financiera. *Perfil de Coyuntura económica*, 191-222.
- Pérez Ramirez, F. O., & Fernández Castaño, H. (2007). Las redes neuronales y la evaluación del riesgo de crédito . *Revista Ingenierías Universidad de Medellín*, 78-91.
- Rodríguez Sánchez, M. T. (2001). Premio nobel de economía 2001: El libre mercado no funciona. *Momento Económico*, 51.
- Salazar Villano, F. E. (2013). cuantificación del riesgo de incumplimiento en créditos de libre inversión: un ejercicio econométrico para una entidad bancaria del municipio de Popayán, Colombia. *Estudios Gerenciales*, 416-427.
- Superintendencia Financiera de Colombia. (1995). *Circular Básica Contable y Financiera (Circular externa 100 de 1995)*. Obtenido de Circular Básica Contable y Financiera (Circular externa 100 de 1995):  
<https://www.superfinanciera.gov.co/jsp/loader.jsf?IServicio=Publicaciones&ITipo=publicaciones&IFuncion=loadContenidoPublicacion&id=15466>
- Velandia Velandia, N. (2013). *Universidad Nacional de Colombia*. Obtenido de Universidad Nacional de Colombia:  
<http://www.bdigital.unal.edu.co/10202/1/709584.2013.pdf>
- Velasco S, C. (08 de Marzo de 2010). *Biblioteca digital Universidad del Valle*. Obtenido de Biblioteca digital Universidad del Valle:  
<http://bibliotecadigital.univalle.edu.co/bitstream/10893/3685/4/CB-0449603.pdf>

## ANEXO A

Tabla de clasificación					
	Observado	LOGIT1AÑO	Pronosticado		
			LOGIT1AÑO		Porcentaje correcto
			0	1	
Paso 1		0	2961	120	96,1
		1	397	1023	72,0
		Porcentaje global			88,5
Paso 2		0	2953	128	95,8
		1	387	1033	72,7
		Porcentaje global			88,6
Paso 3		0	2935	146	95,3
		1	372	1048	73,8
		Porcentaje global			88,5
Paso 4		0	2935	146	95,3
		1	371	1049	73,9
		Porcentaje global			88,5
Paso 5		0	2937	144	95,3
		1	372	1048	73,8
		Porcentaje global			88,5
Paso 6		0	2939	142	95,4
		1	373	1047	73,7
		Porcentaje global			88,6
Paso 7		0	2947	134	95,7
		1	366	1054	74,2
		Porcentaje global			88,9
Paso 8		0	2944	137	95,6
		1	367	1053	74,2
		Porcentaje global			88,8

a. El valor de corte es ,500

## ANEXO B

## Variables en la ecuación

		B	Error estándar	Wald	gl	Sig.	Exp(B)
Paso 1 <sup>a</sup>	CUOTASCANCELADAS	,449	,014	1029,070	1	,000	1,567
	Constante	-3,923	,105	1407,985	1	,000	,020
Paso 2 <sup>b</sup>	CUOTASCANCELADAS	,409	,014	833,464	1	,000	1,505
	OFICINA	,001	,000	69,346	1	,000	1,001
	Constante	-4,150	,112	1378,272	1	,000	,016
Paso 3 <sup>c</sup>	CUOTASCANCELADAS	,420	,015	800,850	1	,000	1,521
	OFICINA	,002	,000	157,157	1	,000	1,002
	C	,550	,045	147,481	1	,000	1,734
	Constante	-6,158	,216	812,342	1	,000	,002
Paso 4 <sup>d</sup>	CUOTASCANCELADAS	,440	,015	816,259	1	,000	1,553
	CUOTASMORA	,047	,005	83,065	1	,000	1,048
	OFICINA	,002	,000	160,951	1	,000	1,002
	C	,621	,047	175,739	1	,000	1,860
	Constante	-7,074	,250	801,128	1	,000	,001
Paso 5 <sup>e</sup>	SALDODEUDA	,000	,000	24,744	1	,000	1,000
	CUOTASCANCELADAS	,456	,016	812,632	1	,000	1,578
	CUOTASMORA	,080	,009	87,849	1	,000	1,083
	OFICINA	,002	,000	164,974	1	,000	1,002
	C	,632	,047	180,157	1	,000	1,882
	Constante	-7,350	,260	802,027	1	,000	,001
Paso 6 <sup>f</sup>	VALORINICIAL	,000	,000	30,203	1	,000	1,000
	SALDODEUDA	,000	,000	47,803	1	,000	1,000
	CUOTASCANCELADAS	,322	,028	130,022	1	,000	1,380
	CUOTASMORA	,098	,009	107,642	1	,000	1,103
	OFICINA	,002	,000	172,434	1	,000	1,002
	C	,637	,048	178,247	1	,000	1,890
	Constante	-7,164	,264	738,217	1	,000	,001
Paso 7 <sup>g</sup>	EDADMOORA	,005	,002	8,696	1	,003	1,005
	VALORINICIAL	,000	,000	32,362	1	,000	1,000
	SALDODEUDA	,000	,000	51,289	1	,000	1,000
	CUOTASCANCELADAS	,320	,029	125,505	1	,000	1,377
	CUOTASMORA	,101	,010	111,623	1	,000	1,106

	OFICINA	,002	,000	157,229	1	,000	1,002
	C	,617	,048	162,877	1	,000	1,854
	Constante	-8,875	,645	189,264	1	,000	,000
Paso 8 <sup>h</sup>	EDADMOORA	,005	,002	9,727	1	,002	1,005
	VALORINICIAL	,000	,000	35,796	1	,000	1,000
	SALDODEUDA	,000	,000	55,510	1	,000	1,000
	TOTALCUOTAS	,113	,044	6,558	1	,010	1,119
	CUOTASCANCELADAS	,206	,052	15,454	1	,000	1,229
	CUOTASMORA	,059	,019	9,718	1	,002	1,060
	OFICINA	,002	,000	158,694	1	,000	1,002
	C	,616	,048	161,924	1	,000	1,852
	Constante	-9,095	,653	194,201	1	,000	,000

a. Variables especificadas en el paso 1: CUOTASCANCELADAS.

b. Variables especificadas en el paso 2: OFICINA.

c. Variables especificadas en el paso 3: C.

d. Variables especificadas en el paso 4: CUOTASMORA.

e. Variables especificadas en el paso 5: SALDODEUDA.

f. Variables especificadas en el paso 6: VALORINICIAL.

g. Variables especificadas en el paso 7: EDADMOORA.

h. Variables especificadas en el paso 8: TOTALCUOTAS.

## ANEXO C

## Matriz de correlaciones

	Constante	CUOTASCANCELADAS	OFICINA	C	CUOTASMORA	SALDODEUDA	VALORINICIAL	EDADMOORRA	TOTALCUOTAS
Paso 1	Constante	1,000	-.902						
	CUOTASCANCELADAS	-.902	1,000						
Paso 2	Constante	1,000	-.725	-.355					
	CUOTASCANCELADAS	-.725	1,000	-.251					
	OFICINA	-.355	-.251	1,000					
Paso 3	Constante	1,000	-.585	-.475	-.852				
	CUOTASCANCELADAS	-.585	1,000	-.153	,244				
	OFICINA	-.475	-.153	1,000	,387				
	C	-.852	,244	,387	1,000				
Paso 4	Constante	1,000	-.614	-.460	-.837	-.493			
	CUOTASCANCELADAS	-.614	1,000	-.128	,300	,239			
	CUOTASMORA	-.493	,239	,097	,222	1,000			
	OFICINA	-.460	-.128	1,000	,401	,097			
	C	-.837	,300	,401	1,000	,222			
Paso 5	Constante	1,000	-.639	-.464	-.825	-.495	,271		
	SALDODEUDA	,271	-.264	-.066	-.084	-.812	1,000		
	CUOTASCANCELADAS	-.639	1,000	-.105	,309	,347	-.264		

	CUOTASMORA	-,495	,347	,114	,196	1,000	-,812			
	OFICINA	-,464	-,105	1,000	,407	,114	-,066			
	C	-,825	,309	,407	1,000	,196	-,084			
Paso 6	Constante	1,000	-,442	-,454	-,814	-,460	,004	,099		
	VALORINICIAL	,099	-,818	,086	,039	,250	-,940	1,000		
	SALDODEUDA	,004	,721	-,111	-,071	-,508	1,000	-,940		
	CUOTASCANCELA DAS	-,442	1,000	-,136	,140	-,016	,721	-,818		
	CUOTASMORA	-,460	-,016	,150	,204	1,000	-,508	,250		
	OFICINA	-,454	-,136	1,000	,412	,150	-,111	,086		
	C	-,814	,140	,412	1,000	,204	-,071	,039		
Paso 7	Constante	1,000	-,170	-,053	-,220	-,304	,129	-,058	-,912	
	EDADMOORA	-,912	-,012	-,141	-,123	,126	-,140	,110	1,000	
	VALORINICIAL	-,058	-,817	,072	,025	,254	-,941	1,000	,110	
	SALDODEUDA	,129	,717	-,093	-,054	-,512	1,000	-,941	-,140	
	CUOTASCANCELA DAS	-,170	1,000	-,138	,142	-,014	,717	-,817	-,012	
	CUOTASMORA	-,304	-,014	,131	,187	1,000	-,512	,254	,126	
	OFICINA	-,053	-,138	1,000	,414	,131	-,093	,072	-,141	
	C	-,220	,142	,414	1,000	,187	-,054	,025	-,123	
Paso 8	Constante	1,000	,029	-,062	-,218	-,026	,155	-,076	-,911	-,141
	EDADMOORA	-,911	-,066	-,135	-,122	,000	-,147	,115	1,000	,072
	VALORINICIAL	-,076	-,542	,075	,023	,030	-,932	1,000	,115	,129
	SALDODEUDA	,155	,592	-,097	-,046	-,037	1,000	-,932	-,147	-,265
	TOTALCUOTAS	-,141	-,844	,050	,004	-,860	-,265	,129	,072	1,000

CUOTASCANCELA DAS	,029	1,000	-,117	,074	,715	,592	-,542	-,066	-,844
CUOTASMORA	-,026	,715	,022	,086	1,000	-,037	,030	,000	-,860
OFICINA	-,062	-,117	1,000	,416	,022	-,097	,075	-,135	,050
C	-,218	,074	,416	1,000	,086	-,046	,023	-,122	,004