

# Medición de Valor en Riesgo en Cartera de Clientes a Través de Modelos Logísticos y Simulación de Montecarlo

David Esteban Rodríguez Guevara<sup>1</sup>

*drodri15@eafit.edu.co*

Alfredo Trespalcios Carrasquilla<sup>2</sup>

*alfredo.trespalcios@gmail.com*

## Resumen

Los credit score son análisis discriminatorios que proporcionan herramientas de decisión para evaluar los riesgos de crédito en una entidad financiera. El establecer los parámetros de riesgo de impago de los clientes ayuda a mitigar las pérdidas monetarias que afectan directamente en activos y patrimonio. Estos parámetros calculados con modelos logísticos, combinados con simulaciones Montecarlo basados en distribuciones Bernoulli y niveles de confianza (VaR) aportan una herramienta dinámica que puede estructurar nuevas políticas de productos financieros y mejoramiento de análisis de pérdida. Para este trabajo se toma la información de los clientes de una cooperativa de ahorro y crédito ubicada en Armenia – Quindío y de la cual se obtuvo un modelo establecido en 10 variables socioeconómicas las cuales arrojaron un modelo discriminatorio logístico, en el que se la cartera de clientes que aún tienen sus créditos en pago, mostrando el riesgo de impago de la siguiente cuota y su interpretación.

## Palabras clave

Credit score, Regresión Logística, Logit, análisis discriminante, riesgo de cartera, riesgo cliente, simulación de Montecarlo, probabilidad binomial, distribución Bernoulli, Value at Risk (VaR)

## Abstract

The credit score discriminatory analysis provide a decision tools about the risk credit in a financial institution. Set the default risk in customers helps to relieve the losses in assets and patrimony. Those parameters calculated with logistic models, combined with Montecarlo simulations based in Bernoulli distributions and confidence level (VaR) provide a dynamic tool to take decisions in new financial policy and enhance the risk to loss analysis; For this paper, we are working with a data base of customers by a credit union institution located in Armenia – Quindío – Colombia; the results of this paper was: 10 socioeconomic variables which threw a discriminatory logit model, evaluated in a customers portfolio with a Montecarlo, who still have their credits in payment, showing the default risk on the next share and their interpretation.

## Key words

Credit score, Logistic Regression, Logit, Discriminant analysis, default risk, customer risk, Montecarlo simulation, binomial probability, Bernoulli distribution, Value at Risk (VaR)

---

<sup>1</sup> Economista de la Universidad La Gran Colombia seccional Armenia, Especialista en Finanzas de la universidad EAFIT, candidato a Magister en Administración financiera de la Universidad EAFIT.

<sup>2</sup> Ingeniero Electricista de la Universidad Nacional de Colombia. MSc Finanzas de la Universidad EAFIT. Profesional Mercados EPM, Profesor de la Maestría en Administración Financiera de la Universidad EAFIT, de la Especialización en Estadística Aplicada en la Escuela de Ingeniería de Antioquia e Ingeniería Financiera del Instituto Tecnológico Metropolitano.

## Introducción

La actividad principal de las entidades de crédito es prestar capital a sus clientes con el fin de financiar sus necesidades de gasto o de inversión, dichos préstamos pueden exponer a las entidades financieras a la pérdida parcial del capital y el rendimiento del valor del dinero impactando directamente en sus resultados financieros. A lo anterior, se le conoce como riesgo de crédito, y se entiende como la pérdida monetaria generada por la posibilidad de impago en la cartera de los clientes (Superfinanciera, 2002). Este riesgo de crédito se clasifica en: riesgo de impago y de calificación; el primero es aquel en el que el deudor no realiza sus pagos a tiempo; y el segundo asume la calidad crediticia de la persona o un grupo de personas. Como factor de análisis los *scoring* y *rating's* califican el nivel de riesgo de impago de los clientes y reportan las características personales que pueden influenciar dicho impago, sea de un cliente o de una cartera de clientes (Rayo *et al*; 2010).

Los estudios de *credit score* inician en los años 60 con Myers y Forge (1963) y Altman (1968) identificando la bancarrota que resultaba de los impagos de los clientes (empresas) analizados; en los años 70, Orgler (1970) y Apilado *et al* (1974) buscaron la implicación de impagos de los riesgos financieros de los clientes; Wiginton (1980) examinó la eficacia de los modelos *Logit* encontrando un mejor poder de predicción y eficientes procesos de sensibilidad y estimación de los análisis discriminantes; para los años 90, Glorfléd (1990), Crook *et al* (1992), Zhai y Russell (1999) enfocaron sus estudios en la comparación de metodologías matemáticas y econométricas que ofrecieran resultados óptimos de discriminación; para el nuevo milenio los modelos *Logit* en *credit score* añadieron nuevas líneas de modelos de contrastación como: CART (*Classification And Regression Trees*), Probit y Redes neuronales, los cuales son opciones alternativas al modelo *logit* con resultados igualmente eficientes, sean los casos de Mures *et al* (2005), Kocenda y Vojtek (2009), Rayo *et al* (2010) y Medita *et al* (2012).

Basilea II recomienda poner en práctica los parámetros de análisis para la administración del riesgo de crédito, en los cuales se analizan: liquidez, rendimientos, deuda, servicio a la deuda, activos e IRB (*Internal Rating-Based approach*), que permiten medir los riesgos

impago de las entidades financieras, aunque también proponen el uso de *credit score*, este no es muy profundo, por tanto, cabe la posibilidad de estructurar análisis metódicos basándose en medidas econométricas para clientes; otros autores como Fernández, Pérez (2005); Pérez, Fernández (2007), Ochoa *et al* (2010), Ayus *et al* (2010), Peña *et al* (2010), Villano (2013) y Moreno (2014) usan modelos econométricos de análisis al riesgo de impago, pero estos, como los anteriores autores, no realizan una medición del riesgo final que infiere el portafolio de clientes debido a las políticas de créditos definidas, y basan sus predicciones en los modelos econométricos o en la aprobación de márgenes propuestos por Basilea II.

Es por esto, que una alternativa de medición de riesgo (VaR) simulado, muestra los niveles de confianza de pérdida de los elementos que se desean analizar, para este caso, una aplicación de medición que este determinada por la caracterización de los clientes y su influencia en el impago de un portafolio de crédito es fácilmente extensible a valoración de riesgo por modificación a las políticas de asignación de créditos

Para ilustrar la metodología se estima un modelo econométrico *Logit* que permita medir el riesgo de impago de la siguiente cuota de una cartera de crédito, que será corroborada con los parámetros de bondad de ajuste, posteriormente dinamizada con una simulación Montecarlo serán que identifique la pérdida esperada en pesos, el análisis de niveles de confianza (VaR), y gráficas de distribución acumulada del impago; como caso de estudio se usa la base de datos de una Cooperativa de Ahorro y Crédito establecida en la ciudad de Armenia – Quindío – Colombia.

Para este artículo, se analiza inicialmente la metodología del *credit score*, los modelos *logit* y su uso, y la simulación Montecarlo basada en la distribución Bernoulli; después, se opera un caso de estudio en el que se aplican los conceptos metodológicos anteriores y finalmente las conclusiones del estudio.

## **Riesgo de crédito y medición**

Basilea II indica que el riesgo de crédito corresponde a la pérdida monetaria generada por la posibilidad de impago en la cartera de los clientes, para identificar los actores de dichos procesos se crean los análisis de *credit score*, estos comprenden un estudio de las variables socioeconómicas de los clientes para identificar los factores de riesgo de crédito; esta metodología inició en los años 60 con Myers y Forge (1963) y Beaver (1967) en empresas pequeñas y medianas midiendo a sus clientes para determinar las potencialidades de bancarrota por los créditos con modelos lineales discriminatorios (Kocenda y Vojtek, 2009); Altman (1968) de manera similar estudiaba la calidad de la predicción de bancarrota de una compañía para obtener una razón de predicción de la misma.

Para 1970, Yair Orgler introduce un enfoque en el préstamo, orientado en la naturaleza de los créditos usando calificaciones de liquidez, rentabilidad, apalancamiento y actividad con modelos de regresión multivariada. En 1976, Apilado, Warner y Dauten realizan un estudio de variables socioeconómicas con modelos lineales discriminatorios para revisar su implicación en un préstamo de alto nivel, resultando así un estudio cercano al *credit score* actual.

Para 1980, Jhon C. Wiginton 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; un mes después Altman publica un documento sobre el proceso de los préstamos, su naturaleza, y los criterios de evaluación de los créditos con procesos logísticos.

Para 1990, los autores iniciaron procesos de comparación para encontrar un modelo óptimo de estudio de *credit score*, Glorfléd (1990) estudió un modelo de discriminación basado en *Least - Absolute - Value* en la cual buscaba demostrar una alternativa del modelo discriminatorio clásico, Crook *et al* (1992) comparó los modelos discriminatorios clásicos con los modelos de performance de crédito de características multivariadas, Zhai y Russell

(1999) en cambio buscaron una alternativa con las redes neuronales para explicar los factores de riesgo de clientes como alternativa análisis.

Para el nuevo siglo, los estudios de las variables discriminantes han sido combinados con análisis de variables macroeconómicas, buscando robustecer los modelos con indicadores como tasas de interés, tasas de mercado, PIB, indicadores financieros, tasas de empleo, tal como lo demuestran autores como Belloti y Crook (2007) y Rayo *Et al* (2010), señalando que el análisis de estas otras variables tienen una influencia directa en los clientes y su comportamiento como consumidor financiero.

Hay varios tipos de metodología que pueden ser comparadas con los modelos *Logit*; como lo muestran los trabajos de Pérez y Fernández (2005) en el que usan redes neuronales, Belloti y Crook (2007) con el análisis de supervivencia; Kocenda y Vojtek (2009) que comparan un modelo *Logit* con un Modelo CART; Peña *et al* (2011) realizan un *credit score* con ANOVA (ANálisis of VAriance) y finalmente, Ochoa *et al* (2010) y Moreno (2014) comparan los modelos *Logit* con *Probit*, Redes neuronales y *Logit* Mixtos, llegando a un resultado similar de los modelos logísticos para este tipo de estudios.

La aplicación de un modelo de *credit score* para la cartera de clientes se ha aplicado en su gran mayoría en entidades enfocadas al microcrédito o banca de capital semilla, el estudio de los factores de riesgo intrínsecos en una persona ha demostrado favorables resultados en la predicción de los pagos de los créditos, permitiendo la aplicación de estrategias que eviten el impago de estos y reestructuren la visión para fortalecer los productos y sus respectivas ganancias. (Fernández, Pérez; 2005) (Rayo *et al*; 2010)

En la siguiente tabla se presenta algunos de los exponentes de las metodologías de *Credit score* en los últimos 52 años:

**Tabla 1. Autores, Metodología y Aplicación de los modelos de *Credit score***

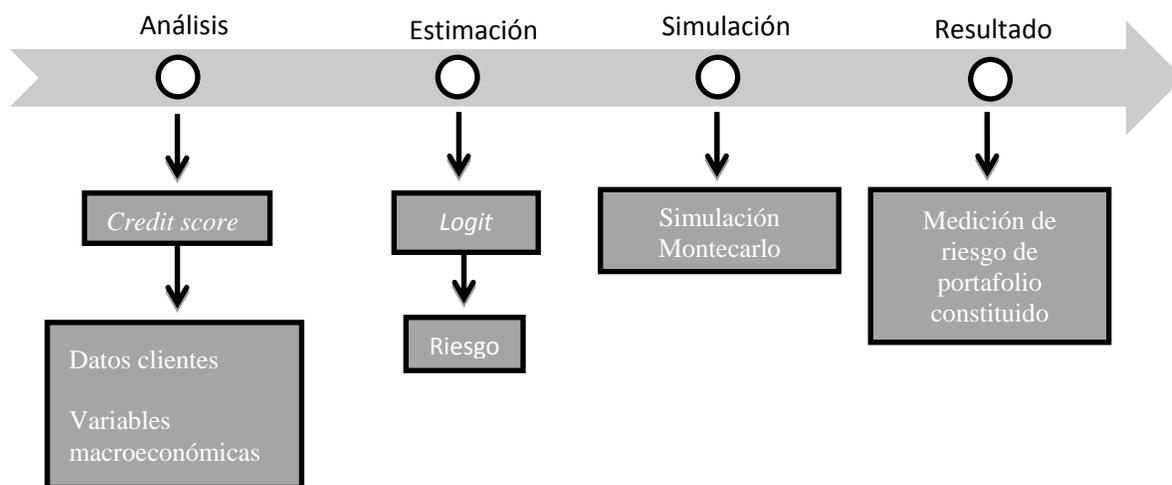
<b>Autor</b>	<b>Metodología</b>	<b>Aplicación en</b>
Myers, Forgy (1963)	Modelo discriminatorio	U.S.A; Universal Finance Co.
Altman (1968)	Modelo discriminatorio	U.S.A, Entidades bancarias
Orgler (1970)	Regresión multivariada	U.S.A, Entidades bancarias
Apilado, <i>et al</i> (1974)	Modelo discriminatorio	U.S.A, Entidades bancarias
Anderson, Narasimhan (1979)	Modelo discriminante	U.S.A, Mineapolis
Wiginton (1980)	<i>Logit</i> , Modelo discriminante	U.S.A; Mayor Oil Company
Altman (1980)	Modelo discriminatorio	U.S.A, Entidades bancarias
Reichert, Cho, Wagner (1983)	Modelo discriminatorio	U.S.A, Entidades bancarias
Kolesar, Showers (1985)	Modelo discriminatorio lineal	U.S.A, Entidades bancarias
Glorfeld (1990)	<i>Least – Absolute - Value</i>	U.S.A, Entidades bancarias
Crook <i>et al</i> (1992)	Modelo Performance	U.S.A, Entidades bancarias
Zhai, Russell (1999)	Redes Neuronales, Markov	U.S.A, Entidades bancarias
Copas (1999)	<i>Logit</i>	Reino Unido
Lipovetsky, Conklin (2004)	<i>Logit</i> , Modelo discriminatorio	U.S.A, Entidades bancarias
Roszbach (2004)	Modelo Tobit bivariado	Suecia, Institución financiera
Mures, <i>et al</i> (2005)	<i>Logit</i> , Modelo discriminatorio	España, Entidad Financiera
Fernández, Pérez (2005)	<i>Logit, Probit</i>	Colombia
Smith, <i>et al</i> (2005)	Modelos Híbridos	Taiwan, Citibank
Alessie <i>et al</i> (2005)	<i>Logit</i>	Italia, Findomestic
Verstraeten, Van den Poel (2005)	Sample Bias	Belgica, Entidad Bancaria
Majer (2006)	<i>Logit</i> , Método de divergencia	Polonia, Entidad bancaria
Deng, Gabriel (2006)	Insured Mortgage Performance	U.S.A, Entidades bancarias
Bellotti, Crook (2007)	<i>Survival analisis</i>	Reino Unido, UK bank
Pérez, Fernández (2007)	Redes neuronales	Colombia
Esteve (2007)	Modelo discriminatorio	Sector Bancario Hipotecario
Gonçalves, Braga (2008)	<i>Logit</i> multinomial	Brasil, Cooperativa de ahorro
Zhou, <i>et al</i> (2009)	Support Vector Machine	Alemania, Entidad Bancaria
Kocenda, Vojtek (2009)	<i>Logit, CART</i>	Europa, Entidad Académica
Bolton (2010)	<i>Logit</i> , Redes neuronales	Sudáfrica, Entidad Bancaria
Ochoa, <i>et al</i> (2010)	Redes neuronales, <i>Logit, Probit</i>	Colombia, Cooperativa
Rayo, <i>et al</i> (2010)	<i>Logit</i>	Perú, Entidad Micro bancaria
Arbelaez, Mendoza (2010)	Bootstrapping	Colombia
Ayús <i>et al</i> (2010)	Redes neuronales, <i>Logit</i>	Colombia, Entidad financiera
Martens <i>et al</i> (2010)	Colonia de Hormigas	Bélgica
Mileris (2010)	Modelo Bayesiano	Lituania, Entidad Bancaria
Han (2011)	Probit	U.S.A, Entidades bancarias
Malik, Thomas (2010)	<i>Logit</i>	Reino Unido, Entidad Bancaria
Selau, Ribeiro (2010)	CRF (Conditional Random Field)	Brasil
Matoussi, Krichene (2010)	<i>Logit</i> - Redes Neuronales	Túnez, Entidad Bancaria
Řezáč, Řezáč (2011)	Smirnov- Kolmogorov, Lorentz	Republica Checa
Constangioara (2011)	<i>Logit</i> , Redes neuronales, Bootstrap	Hungría, Entidad financiera
Gutierrez, Velandia (2011)	Support Vector Machine	Colombia, Banco de la Rep.
Pope, Sydnor (2011)	Modelo discriminatorio	U.S.A, Entidad Bancaria
Webster (2011)	Regresión logística bayesiana	Dinamarca
Peña, <i>et al</i> (2011)	ANOVA	Colombia, Cooperativa
Chaudhuri, Cheral (2012)	<i>Logit</i>	India, Entidad bancaria
Medina, Marti Selva (2012)	<i>Logit, CART</i>	España, Entidad Financiera
Sousa <i>et al</i> (2013)	Modelos discriminantes de Fisher	Brasil, Institución financiera
Espin-, Rodríguez- (2013)	<i>Logit</i>	Mexico, Institución Bancaria
Brevoort <i>et al</i> (2013)	Probit,	U.S.A, Reserva Federal
Villano (2013)	<i>Logit</i> , Modelo log-log	Colombia, Entidad Bancaria
Moreno (2014)	<i>Logit, Probit, Logit Mixto</i>	Colombia, Cooperativa
Baklouti(2014)	Fuzzy Clasifiers, CART	Tunez, Banco Microfinanciero

*Fuente: elaboración propia*

## Descripción de la metodología aplicada

Tomando en cuenta la construcción del *credit score* por parte de los anteriores autores, se toma la base de datos de créditos de los clientes y se une con las variables macroeconómicas por estudiar según su año de establecimiento del préstamo, a continuación, se refina dicha base de datos con tratamientos estadísticos de depuración de datos posteriormente descritos; acto seguido, se calibra el modelo con procesos de bondad de ajuste, estos garantizan la calibración de un modelo que determine la probabilidad de riesgo de impago de los clientes; finalmente, bajo una simulación Montecarlo con distribución Bernoulli se obtiene la medición del riesgo esperado de un portafolio de clientes, complementado con un análisis de niveles de confianza (Value at Risk) que determinen el máximo riesgo de impago.

**Gráfica 1. Diagrama de Flujo de la metodología aplicada**



*Fuente: elaboración propia*

### ***Credit score***

Según Orgler (1970), Apilado *et al* (1970), Wiginton (1980), Kolesar y Showers (1985), Mures *et al* (2005), Kocenda y Vojten (2009), Rayo *et al* (2010), Abdou y Pointon (2011), Palacio *et al* (2011), Medina, Marti Selva (2012), los *credit score* son procedimientos de

análisis discriminante estadísticos multivariados que explican dos grupos de análisis “los no riesgosos” y “los riesgosos”, siendo la variable independiente la explicación de los resultados binomiales conocidas como **variables discriminantes**, y las dependientes son las variables socioeconómicas que alimentan al modelo; este tipo de modelos muestran las relaciones entre el prestamista, el préstamo y el prestatario y la reducción del tiempo gastado en cobranza (Schreiner, 2002).

Bellotti y Crook (2007) y Rayo *et al* (2010), manifiestan que el uso de las variables macroeconómicas tienen un impacto directo en la decisión de inversión; variables como tasa de desempleo, tasas bancarias, precio finca raíz, ingresos, producción nacional, índices del consumidor, son recomendadas por los autores. Estas proporcionan el valor de riesgo del mercado tal como lo expresa Mures *et al* (2005). Para ello se muestra un ejemplo de la base de datos consultada en el anexo 1, que identifica las variables usadas y quienes las usaron.

Para determinar el uso adecuado de la información que alimente al modelo *logit* se clasificará la base de datos y el acoplamiento de las variables socioeconómicas en una sola matriz de datos:

Primero, los factores socioeconómicos determinados por la información de la entidad se trabajarán con variables cuantitativas o cualitativas; las variables cuantitativas como: los ingresos, los egresos, la edad, los saldos y las tasas de los créditos, y las cualitativas que como el género, el estrato social, el estado civil o la residencia clasificados en variables dicotómicas; segundo se acomodaran las variables macroeconómicas como el PIB, la tasa de interés, la tasa de usura y la tasa de desempleo.

### **Modelos *Logit***

Los estudios de Wiginton (1980), Orgler (1980), Gujarati (2004), Rayo *et al* (2010), Abdou y Pointon (2011), Gutiérrez y Velandia (2011) y Martí Selva (2012), manifiestan que 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.

Para identificar la probabilidad de impago de las variables discriminantes de un cliente se debe identificar un modelo tal que:

$$P(\text{Impago}) = \frac{1}{1 + e^{-(\alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_k X_k)}} [1]$$

Dónde:

$P(\text{Impago})$ : Probabilidad de impago de crédito por parte de un cliente

$X_{i_k}$ : Variables explicativas

$\alpha$ : Término independiente o constante del modelo

$\beta_{i_k}$ : Coeficiente de regresión asociado a la variable explicativa

El cual se interpreta como la probabilidad de impago de un cliente de la inversa de la razón de probabilidad negativa de las variables exógenas.

### Tratamiento de los modelos logísticos (Bondad de ajuste)

Algunos de los conceptos revisados corresponden con las propuestas de las siguientes investigaciones:

**Tabla 2: Procesos de bondad de Ajuste**

Análisis	Tratamiento	Autor
<i>Rastreo "back-forward"</i>	Agregar variable a variable en el modelo y revisar su comportamiento.	Salazar (2010) Moreno (2014)
AIC ( <i>Akaike Information criteria</i> ) y BIC ( <i>Bayesian Information criterion</i> )	en los dos test entre menor sea el valor del AIC y el BIC es mejor el modelo	Salazar (2010) Moreno (2014)
test ANOVA de dos modelos	Según test $X^2$ si la variable añadida o eliminada da al modelo valor de $P < 0,05$ es un cambio significativo adecuado.	Salazar (2010)
Devianza y Devianza de Pearson	Entre más se explique la devianza de los errores mejor es la explicación de los errores en el modelo.	Bolton (2009) Webster (2011) Iglesias (2013)

Pseudo-R <sup>2</sup>	valores entre 0,3 y 0,6 son óptimos para ajustar la relación de las variables; pero se toman más como valores de estudio	Gujarati (2004) Bolton (2009) Salazar (2010) Iglesias (2013)
Signo	El signo debe ser coherente con la realidad, sino se debe explicar el cambio.	Salazar (2010) Moreno (2014)
Odds Value	Es el valor exponencial del coeficiente de una variable explicativa, la cual determina la sensibilidad o propensión de dicha variable frente al modelo; para analizar estos valores se puede tomar el valor del coeficiente que es el cambio porcentual de una variable y cotejarla con la sensibilidad de la misma.	Gujarati (2004) Bolton (2009) Salazar (2010) Webster (2011) Iglesias (2013) Moreno (2014)
VIF (Factor de inflación de la Varianza)	Analiza la correlación de las variables explicativas <i>ante y post</i> calibración; si una variable supera un valor de 5 existe alta correlación.	Bolton (2009) Moreno (2014)
Valores atípicos e influénciales ( <i>Leverage vs error estándar de Pearson</i> )	Las distancias de Cook, deben ser menores a 1; los errores estándar pueden ser cercanos a cero, caso contrario algunas variables presentan observaciones atípicas.	Salazar (2010) Moreno (2014)
<i>test Hosmer – Lemeshow</i>	En una distribución X <sup>2</sup> si P > 0,05 el modelo globalmente es ajustado para un modelo logístico y aceptado.	Bolton (2009) Salazar (2010) Webster (2011) Iglesias (2013) Moreno (2014)
ROC test ( <i>Receiver Operating Characteristic test</i> )	Se identifica el área bajo la curva (AUROC); si AUROC > 0,7 se considera un modelo bueno en discriminar las binomiales.	Bolton (2009) Salazar (2010) Cor <i>et al</i> (2012) Moreno (2014)

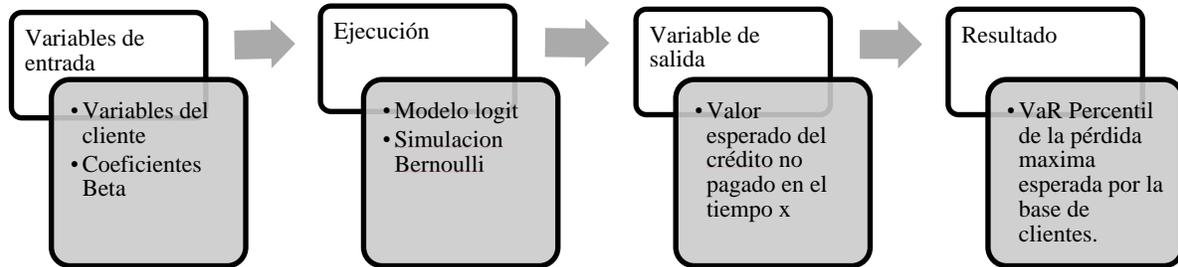
Fuente: elaboración propia

## Simulación de Montecarlo

Ampliando la respuesta de valoración del modelo *logit*, se analiza el riesgo de impago calculado simulado a un millón de escenarios bajo distribución Bernoulli y tomando en cuenta el teorema del límite central en el cual entre más datos menor será la probabilidad de error en la media, en un solo experimento por individuo, donde el éxito de la simulación es el “no pago” del crédito; de este proceso se obtendrá la media esperada que se interpretará como el “valor esperado de impago de la siguiente cuota”; y haciendo uso de los *Value at Risk* (VaR), se analizan los niveles de confianza de pérdida máxima para conocer la distribución acumulada empírica del impago de dicha cuota.

La simulación Montecarlo contempla como variables de entrada las variables del modelo logístico que arrojan las condiciones de riesgo de los clientes, y la variable de salida el saldo que pueda poner en riesgo la cuenta para el pago del mes siguiente, por tanto:

**Gráfica 2. Diagrama de Flujo de la simulación Montecarlo**



*Fuente: elaboración propia*

### Aplicación de la metodología

#### Data

Para este ejercicio se obtuvo la información de clientes de 8664 créditos entre los años 2009 a 2013 para satisfacer las condiciones de los modelos probabilísticos.

**Tabla 3: Data**

	No Riesgo	Riesgo	Total
<b>N° Datos</b>	6692	1972	8664
<b>Porcentaje</b>	77,23%	22,76%	100%

*Fuente: elaboración propia; Datos: Entidad Financiera*

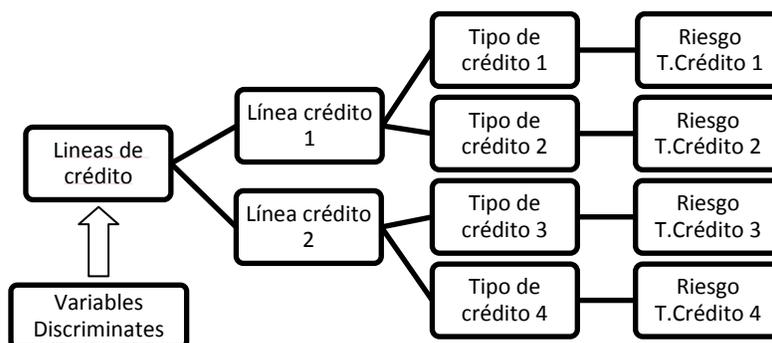
#### Tratamiento de la Data

Las variables suministradas fueron: género, edad, estrato, actividad económica, tipo de contrato, estado civil, salario, estudios, ingresos, egresos, departamento de residencia del cliente, agencia, tasa del crédito, plazo de pago, saldo inicial, saldo final, garantía, código línea del crédito, código destino del crédito, garantía, valor de la garantía, calificación de la entidad, cuotas pagadas, cantidad de atrasos, máximo de días en atraso, promedio de días en atrasos, mínimo de días en atrasos, y calificación propia; las variables macroeconómicas fueron: PIB, PIB de demanda, tasas de desempleo - empleo nacional y departamental, IPC, tasas de intervención, tasas de usura comercial. (Ver Anexo 2 y 3).

Para las variables ingresos y egresos se crean variables dicotómicas de “declaración” debido a información faltante por parte de los clientes.

Como el modelo logístico requiere conocer la composición de los créditos y la influencia de riesgo de los productos, se trabajan las variables de “líneas de crédito” y de “crédito destinos” como cualitativos dicotómicos, dando así la siguiente distribución.

**Gráfica 3: Distribución de los créditos para el análisis de los modelos**



*Fuente: elaboración propia; Gráfica: Propia*

### Calibración del modelo

Como sugiere Moreno (2014), los estudios de crédito pueden encontrar disparidad en la información del portafolio de clientes, encontrándose menor cantidad de clientes riesgosos en proporción a los no riesgosos, para eliminar dicho problema se toma 80% de la información aleatoriamente, y con esta se establece el modelo; por tanto, se trabaja con:

**Tabla 4. 80% de información usada para el modelo**

	No Riesgo	Riesgo	Total
N° Datos	5350	1581	6931
Porcentaje	77,18%	22,81%	100%

*Fuente: elaboración propia*

Realizando los procesos de tratamiento previos a la evaluación, se encuentra que 4 datos presentan problemas en el test Leverage vs Standart deviance of Pearson, las variables

tomadas para el modelo general no presentaron VIF mayor a 5, por lo que se asume la no correlación de los datos y la inflación del error supuesto.

### Parámetros Obtenidos.

El siguiente es el modelo logístico que identifica la probabilidad de impago de los clientes:

Tabla 5. Modelo General				
Variable	B	P. Value (wald)	Signo	Odds Value
(Intercept)	-6,731	< 2e-16 ***	-	0,001
Cantidad de Atrasos	5,378	< 2e-16 ***	+	216,5
Cuotas Pagas	-0,103	2,05e-12 ***	-	0,902
Logaritmo Saldo Inicial (en millones)	-0,980	0,002 **	-	0,375
Plazo del crédito	0,031	0,004 **	+	1,032
Declara egresos	-0,450	0,095 .	-	0,638
Pagaré: Hipoteca	1,098	0,005 **	+	2,998
Crédito: Rotativo 1	0,997	0,070 .	+	2,710
Crédito: Otros créditos de Consumo	-1,451	0,002**	-	0,234
Crédito: Microempresa	-0,609	0,041 *	-	0,544
Crédito: Otros Microcréditos	-2,160	0,001**	-	0,115
Sector Laboral: Agrícola	1,242	0,032 *	+	3,462
Sector Laboral: No indica	0,724	0,003**	+	2,062
Estado civil: separado	1,261	0,013 *	+	3,530
Residencia cliente: Pereira	0,502	0,062 .	+	1,652
Residencia cliente: Caldas	0,961	0,043 *	+	2,614
Residencia cliente: Valle del Cauca	-1,125	0,008 **	-	0,325

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Bondad de ajuste: Ver anexo 4)

Fuente: elaboración propia

### Interpretaciones del modelo

Cantidad de atrasos: Por cada cuota de atraso que tenga el cliente en un crédito, se incrementa el riesgo de impago en 5,37%, lo que hace que el cliente sea 216 veces más propenso a ser riesgoso.

Cuotas pagas: A menor cantidad de cuotas pagas en un crédito, el cliente reduce su riesgo en 0,1%, haciéndolo 0,9 veces menos propenso a ser riesgoso en para el impago.

Log Saldo inicial: Por cada unidad de logaritmo de saldo que se reste a la variable, este reduce el riesgo en 0,9%, haciendo al cliente 0,37 veces menos propenso al riesgo de impago.

Plazo: Por cada incremento de 1 mes en el plazo de pago, el riesgo de impago incrementa en 0,03%, volviendo al cliente 1,03 veces más propenso al riesgo de impago.

Declaración de egresos: Por cada cliente que indique que declaró sus egresos, el riesgo reduce sus riesgos en 0,45%.

Pagaré hipoteca: Por cada cliente que indique que tiene una garantía de hipoteca es 3 veces más propenso al riesgo.

Créditos: Los destinos de créditos analizados, arrojaron que la elección de los créditos rotativos 2 veces más riesgoso de impago al cliente; en los casos de otros microcréditos, los microcréditos c y otros créditos de consumo, su elección reduce los niveles de riesgo.

Sector laboral: El cliente que indique que trabaja en el sector agrícola representa una propensión al riesgo 3 veces mayor que las personas que trabajan en el sector comercial, para el dato no indicado esta propensión es 2 veces mayor.

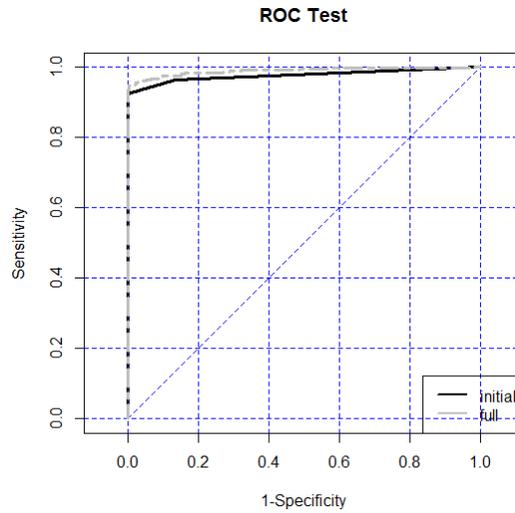
Estado civil: Un cliente que sea separado es 3,5 veces más riesgoso que uno que sea soltero.

Residencia del cliente: Las personas que residen en Pereira y Caldas pueden ser 1,6 y 2,6 veces más riesgosas que las que viven en el Valle del Cauca.

### **Tabla de confusión. (ROC test)**

Para la comparación del ROC test se utilizó como predicción el 20% restante de la información total para confirmar la viabilidad de los predictores y la sensibilidad del modelo:

**Gráfica 4: ROC Test y resultados de sensibilidad y especificidad.**



*Fuente: elaboración propia*

**Tabla 6. Resultados de la Tabla de Confusión.**

	No Riesgo	Riesgo	Total	%	Analisis
No Riesgo	1342	83	1425	94,18%	Especificidad
Riesgo	0	308	308	99,62%	Sensibilidad
Total	1342	391	1733	95,21%	Predicciones

*Fuente: elaboración propia*

El ROC test y la tabla de confusión, permiten inferir que el modelo tiene una eficiencia de reconocimiento del 99,62% en las variables discriminatorias, aunque, hay que tener en cuenta que esta sensibilidad se debe confrontar con los datos que si fueron discriminados en riesgo (83/308), por tanto, el modelo identifica el riesgo de impago de 7 de cada 10 clientes con 99% de certeza. Por otro lado este mismo puede predecir 95% de la información entre riesgosos y no riesgosos.

## Simulación de Montecarlo

**Aplicación a un cliente:** Se explica la metodología propuesta en el trabajo para un cliente de la base de datos seleccionado al azar. La información corresponde a la de un cliente que está pagando un crédito de consumo por valor de 2,98 millones de pesos a 36 meses.

**Tabla 7. Información del Cliente 1**

Cantidad atrasos	0	Crédito Rotativo	No	E. Civil: Separado	No
Cuotas pagas anteriores	12	Otros Créditos de consumo	Si	Residencia: Pereira	No
Log saldo del crédito	0,47	Microcrédito C	No	Residencia: Caldas	No
plazo	36	Otros Microcréditos	No	Residencia: Valle	No
Declara egresos	Si	Sector Agrícola	No		
Garantía: hipoteca	No	No Indica Sector	No		

*Fuente: elaboración propia*

Cada una de las variables exógenas se multiplica por los coeficientes obtenidos en la calibración, como sigue:

**Tabla 8. Ejemplo de uso del modelo para un cliente**

	C	atrasos	Cuotas pagas	Log saldo	D.Egreso	Plazo	OtroConsu	Z
<b>Variable<sup>1</sup></b>	-6,73	5,38	-0,1	-0,98	-0,45	0,03	-1,45	
<b>Condiciones<sup>2</sup></b>		0	12	0,47	1	36	1	1
<b>Resultado<sup>(1*2)</sup></b>	-6,73	0	-1,2	-0,188	-0,45	1,08	-1,45	-8,938

*Fuente: elaboración propia*

$$P(\text{Impago}) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 \text{Atraso} + \beta_2 \text{Cuotas} + \beta_3 \text{LogSaldo} + \beta_4 \text{Plazo} + \beta_5 \text{Egreso} + \beta_6 \text{Otroconsu})}} = 0,01\% [2]$$

Por tanto, al cliente se le estima una probabilidad de 0,01% de no pagar una cuota de \$118.794,3 pesos de su crédito para el próximo mes. Para determinar el monto de la pérdida, se realiza un millón de ensayos en una distribución Bernoulli, donde éxito corresponde al no pago. De la muestra generada se extrae la media que resulta ser una pérdida esperada de \$11,04 pesos, y demás estadísticos necesarios para el reporte; el valor de riesgo de pago de la primera cuota (VaR) de dicha pérdida se establece en:

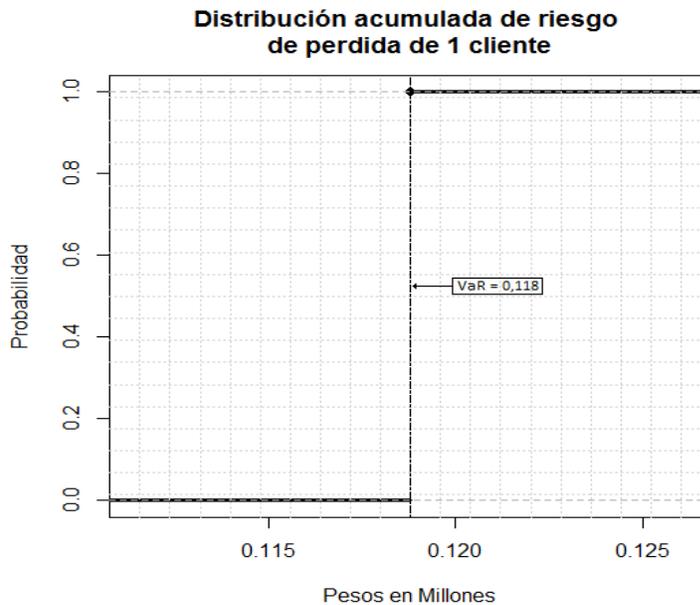
**Tabla 9. Niveles de confianza de Pérdida máxima posible del cliente**

Promedio	VaR 25%	VaR 20%	VaR 10%	VaR 5%	VaR 1%	VaR 0.5%	VaR 0.1%
\$ 11,04	\$118.794,3	118.794,3	\$118.794,3	\$118.794,3	\$118.794,3	\$118.794,3	\$118.794,3

*Fuente: elaboración propia*

De los cuales la pérdida total esperada en todos los niveles de confianza es de la cuota por pagar en el siguiente mes del crédito en el cual el cliente no decida pagar a pesar que su probabilidad de impago es relativamente nula. En este caso los niveles de confianza, asumen el mismo valor porque asumen la constante de la cuota por un cliente.

**Gráfica 5: Distribución acumulada de pérdida de 1 cliente**



*Fuente: elaboración propia*

**Aplicación a cinco clientes:** Para determinar la eficacia del modelo frente a varios clientes, se muestra el resultado simulado de la probabilidad de impago de 5 clientes a la vez escogidos al azar.

**Tabla 10. Información de 5 clientes escogidos al azar**

	C 1	C2	C 3	C 4	C 5
Atrasos	3	7	0	2	1
Cuotas	7	18	42	3	12
log saldo	0,30	0,69	0,54	1,65	0,30
plazo	18	48	48	59	18
Declara egresos	1	1	1	1	1
Pagaré: Hipoteca	0	0	0	1	0
Microcrédito C	0	0	0	0	1
Actividad: no indica	0	1	1	0	1
<b>Riesgo de Impago</b>	<b>99 %</b>	<b>100%</b>	<b>71%</b>	<b>0,2 %</b>	<b>6 %</b>

*Fuente: elaboración propia*

Para este ejemplo, se observa que aplicando el modelo *Logit* de *credit score* obtenemos que los riesgos de impago de los clientes 1, 2 y 3 son altos, y los clientes 4 y 5 tienen riesgos muy bajos, aun así, como se demostró anteriormente, ser calificado bajo riesgo no implica que exista probabilidad de pérdida; por tanto aplicando la simulación binomial en un millón de iteraciones vemos que:

$$E(Deuda) = \mu \left( \sum_i^n Rbinom(Cliente_n) \right) = \$ 385.889 \text{ pesos}$$

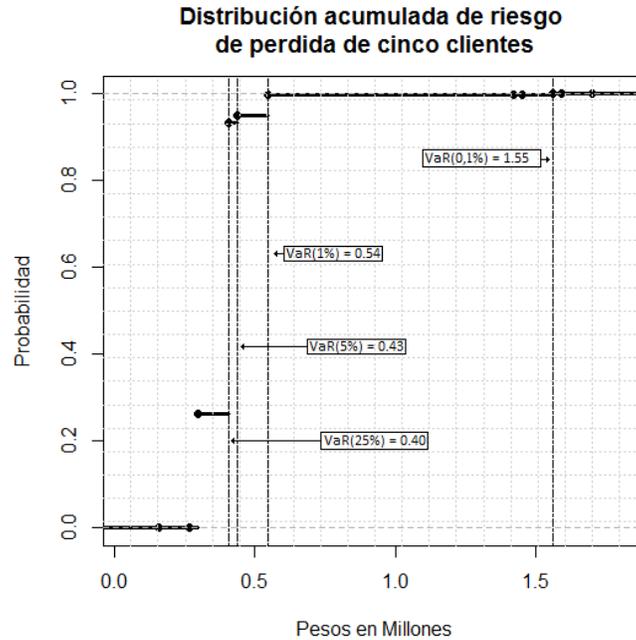
Para determinar entonces la pérdida máxima posible en estos 5 créditos se usa VaR en 6 niveles de confianza, donde vemos que la máxima pérdida de 5 créditos en el siguiente pago es de \$ 385.889 pesos, lo que representa un 22% de pérdida del valor de las cuotas esperadas para estos 5 clientes.

**Tabla 11. Niveles de confianza de Pérdida máxima posible de los cinco clientes**

VaR 25%	VaR 20%	VaR 10%	VaR 5%	VaR 1%	VaR 0.5%	VaR 0.1%
\$405.314,7	\$405.314,7	\$405.314,7	\$435.153,9	\$544.663,4	\$544.663,4	\$1.559.982,7

*Fuente: elaboración propia*

**Gráfica 6: Distribución acumulada de pérdida de 5 clientes**



*Fuente: elaboración propia*

Para estos 5 clientes se espera una pérdida máxima de \$1.559.982,7 pesos en la siguiente cuota en caso de impago de todos los clientes.

**Aplicación a los créditos que siguen en vigencia:**

Para la aplicación de la simulación con las cuentas que siguen en pago, se toman 6772 clientes, se analizara su riesgo de impago para el mes siguiente según lo acordado por el modelo. Para el caso en 1 millón de iteraciones por cada cliente, se encontró que la deuda esperada para el siguiente mes es de:

$$E(Deuda) = \mu \left( \sum_i^n Rbinom(Cliente_n) \right) = \$376.288.617 \text{ pesos}$$

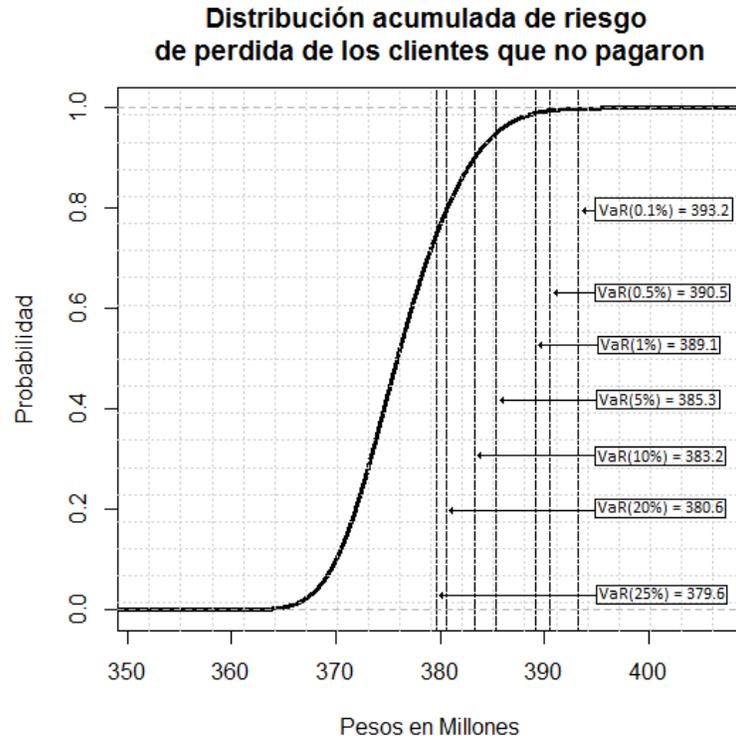
Según la información obtenida, de los datos, esto representa una pérdida esperada corresponde con el 21% de las cuotas del mes siguiente.

**Tabla 12. Niveles de confianza de Pérdida máxima de la base de clientes estudiados**

VaR 25%	VaR 20%	VaR 10%	VaR 5%	VaR 1%	VaR 0.5%	VaR 0.1%
\$379.658.222	\$380.644.240	\$383.256.652	\$385.367.720	\$389.177.440	\$390.523.433	\$393.282.058

*Fuente: elaboración propia*

**Gráfica 7: Distribución acumulada de pérdida de los clientes que no pagaron**

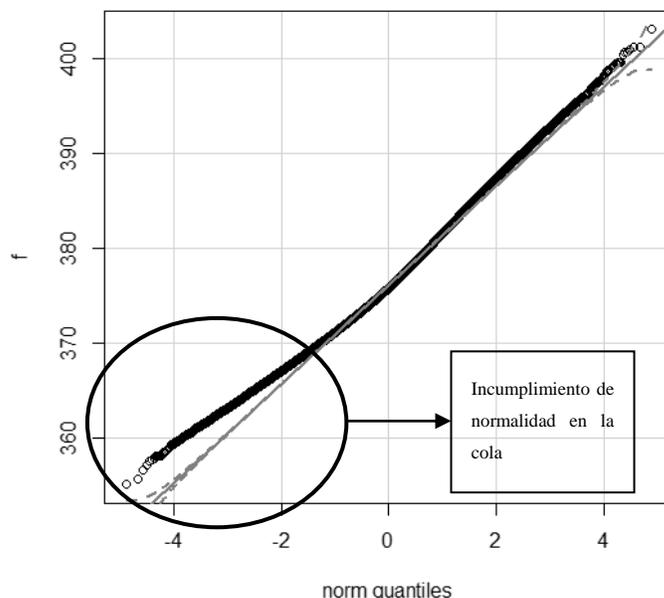


*Fuente: elaboración propia*

Como se puede apreciar, los niveles de confianza de pérdida de la siguiente cuota puede ascender entre 379 millones a un máximo de 393 millones de pesos para los índices de confianza entre el 25% al 0,1%.

## Normalidad de los resultados

**Gráfica 8: Normalidad de los resultados**



*Fuente: elaboración propia*

Como análisis final resultante de la simulación Montecarlo, se expone la presunción de normalidad de la información respecto al valor de dinero en pérdida de los datos analizado, mostrando que la distribución del riesgo de los clientes para determinar su impago de la siguiente cuota no es normal; el test Jarque Bera arroja un resultado de 19.356,27, rechazando la hipótesis de presunción de normalidad.

## CONCLUSIONES

Se ha presentado una metodología para la medición de riesgo de impago de una cartera de clientes, la cual establece los parámetros de análisis logístico y de simulación Montecarlo para su aplicación en el estudio de *credit score*. Cabe resaltar, que estos estudios pueden tener metodologías y análisis variopintos de acuerdo a sus objetivos, pero, los modelos logísticos son más adecuados por su fácil manejo, interpretación y uso.

Los autores consultados para esta investigación usaron una amplia serie de variables socioeconómicas (variables exógenas) para la determinación del análisis discriminante, esta investigación toma las variables más importantes debido a la naturaleza de la base de datos

de la entidad estudiada. Con respecto a las variables macroeconómicas analizadas, estas no tuvieron influencia alguna en el modelo *logit*, pero no se descarta, ni se menosprecia su análisis y uso en modelos de *credit score*.

Para los resultados del modelo, se estableció que 10 variables discriminantes fueron determinantes de riesgo; estas fueron: cuotas vencidas, cuotas pagas, logaritmo del saldo inicial, plazo del crédito, declaración de egresos, pagaré: hipoteca, tipos de crédito, sector laboral, estado civil: separado y residencia del cliente que informan una sensibilidad de análisis de casi 100%, afirmando que 7 de cada 10 clientes pueden ser riesgosos. Respecto al signo esperado, algunas variables presentan signo negativo, confirmando que dichas variables se ajustan mejor interpretando la reducción del riesgo.

La combinación de los modelos *logit* con simulaciones Montecarlo revelan una herramienta más dinámica de análisis financiero; no se conforma con la mera interpretación del modelo y la predicción de riesgo, esta ofrece simulaciones de escenarios posibles, mostrando la pérdida esperada en pesos de cualquier momento de tiempo, en este caso la siguiente cuota; además el análisis de niveles de confianza (VaR), y gráficas de distribución acumulada del impago, crean herramientas de interpretación más adecuada a las políticas de productos crediticios ofrecidos.

Respecto a las distribuciones de probabilidad, tanto los datos para el modelo *logit*, como la simulación dichas distribuciones no son normales, por lo que en este estudio no se puede asumir una distribución normal de la medición de riesgo de crédito, revelando que la normalidad no es factor condicional de análisis estadístico previo.

Sobre el desarrollo de la investigación se puede indicar que: las bases de datos deben ser de información completa, ordenada y bien ajustada para tratamiento estadístico; el tener bases de datos deficientes puede incurrir en una inadecuada estimación de los modelos y problemas de interpretación de variables; como solución propuesta por los autores, el usar herramientas estadísticas de transformación ayudan a depurar y armonizar la base de datos. Al momento de calibrar el modelo es imperativo que se cumplan las bondades de ajuste del modelo, esto garantiza una discriminación de variables más exacta.

Hay que tener especial cuidado con la información crediticia de los clientes, la realidad muestra que un cliente puede tener una cantidad “X” de créditos otorgados generando errores de contradicción en la estimación en el modelo discriminante, para confrontar este problema es bueno cotejar las políticas de calificación de la entidad y revisar el comportamiento del cliente para establecer un único valor discriminante.

## Referencias

Abdou, H. A., & Pointon, J. (2011). Credit scoring, statistical techniques and evaluation criteria: A review of the literature - *Intelligent Systems in Accounting, Finance and Management*, 18(2-3), 59-88.

Altman, E. I. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The journal of finance*, 23(4), 589-609.

Altman, E. I. (1980). Commercial bank lending: process, credit scoring, and costs of errors in lending. *Journal of Financial and Quantitative Analysis*, 15(04), 813-832.

Anderson, J., & Narasimhan, R. (1979). Assessing project implementation risk: a methodological approach. *Management Science*, 25(6), 512-521.

Anderson, S. Williams (2009) Métodos Cuantitativos para los negocios. *Cengage Learning, México*.

Apilado, V. P., Warner, D. C., & Dauten, J. J. (1974). Evaluative Techniques in Consumer Finance—Experimental Results and Policy Implications for Financial Institutions. *Journal of financial and Quantitative Analysis*, 9(02), 275-283.

Arbeláez, A. G., & Mendoza, J. C. *Análisis comparativo del riesgo crediticio: una aproximación no paramétrica* (No. 050). Banco de la Republica de Colombia.

Ayús, A. L. T., Velasquez, R. E. A., & Ceballos, H. V. (2010). Estimación de las provisiones esperadas en una institución financiera utilizando modelos *Logit* y *Probit*. *Revista Ciencias Estratégicas*, 18(24), 259-270.

- Baklouti, I. (2014). A Credit Scoring Model for Microfinance Bank Based on Fuzzy Classifier Optimized by a Differential Evolution Algorithm. *IUP Journal of Financial Risk Management*, 11(2).
- Bellotti, T., & Crook, J. (2009). Credit scoring with macroeconomic variables using survival analysis. *Journal of the Operational Research Society*, 60(12), 1699-1707.
- Bolton, C. (2010). Logistic regression and its application in credit scoring.
- Brevoort, K. P., Avery, R. B., & Canner, G. B. (2013). Credit Where None Is Due? Authorized-User Account Status and Piggybacking Credit. *Journal of Consumer Affairs*, 47(3), 518-547.
- Castaño, H. F., & Ramírez, F. O. P. (2005). El modelo logístico: una herramienta estadística para evaluar el riesgo de crédito. *Revista Ingenierías Universidad de Medellín*, 4(6), 55-75.
- Chaudhuri, K., & Cheral, M. M. (2012). Credit rationing in rural credit markets of India. *Applied Economics*, 44(7), 803-812.
- Constangioara, A. (2011). Consumer credit scoring. *Romanian Journal of Economic Forecasting*, 3, 162-177.
- Copas, J. (1999). The effectiveness of risk scores: the *Logit* rank plot. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 48(2), 165-183.
- Cor, T. C., del Val, E. B., & Gregori, J. F. (2012). Bondad de ajuste y elección del punto de corte en regresión logística basada en distancias. Aplicación al problema de "credit scoring". In *Anales del Instituto de Actuarios Españoles*(No. 18, pp. 19-40). Instituto de Actuarios Españoles.
- Crook, J. N., Hamilton, R., & Thomas, L. C. (1992). A comparison of a credit scoring model with a credit performance model. *Service Industries Journal*, 12(4), 558-579.
- Deng, Y., & Gabriel, S. (2006). Risk-based pricing and the enhancement of mortgage credit availability among underserved and higher credit-risk populations. *Journal of money, Credit and Banking*, 1431-1460.

Espin-García, O., & Rodríguez-Caballero, C. V. (2013). Metodología para un scoring de clientes sin referencias crediticias. *Cuadernos de Economía*, 32(59), 137-162.

Esteve, E. M. (2007). Un modelo de credit scoring basado en el conocimiento de la aplicación de basilea ii y su papel innovador en el sector bancario. In *Decisiones basadas en el conocimiento y en el papel social de la empresa: XX Congreso anual de AEDEM* (p. 52). Asociación Española de Dirección y Economía de la Empresa (AEDEM)

Garcia, S (S.f). Docente de UDI Econometría. Modelos de regresión con respuesta cualitativa [En línea]. Universidad Autonoma de Madrid

Glorfeld, L. W. (1990). A robust methodology for discriminant analysis based on least-absolute-value estimation. *Managerial and Decision Economics*, 11(4), 267-277.

Gonçalves, R. M. L., & Braga, M. J. (2008). Determinantes de risco de liquidez em cooperativas de crédito: uma abordagem a partir do modelo *Logit* multinomial. *Revista de Administração Contemporânea, Curitiba*, 12(4), 1-019.

Gujarati, d. N. (2004). *Econometria* (4a. Ed.). Mexico: mcgraw-hill interamericana.

Gutiérrez, J. F. M., & Velandia, L. F. M. (2011). *Pronóstico de incumplimientos de pago mediante máquinas de vectores de soporte: una aproximación inicial a la gestión del riesgo de crédito* (no. 009079). Banco de la república.

Han, S. (2011). Créditor learning and discrimination in lending. *Journal of Financial Services Research*, 40(1-2), 1-27.

Hochguertel, S., Weber, G., & Alessie, R. (2005). Consumer credit: evidence from Italian micro data. *Journal of the European Economic Association*, 3(1), 144-178.

Kocenda, E., & Vojtek, M. (2009). *Default predictors and credit scoring models for retail banking* (No. 2862). CESifo working paper.

Kolesar, P., & Showers, J. L. (1985). A robust credit screening model using categorical data. *Management Science*, 31(2), 123-133.

- Lipovetsky, S., & Conklin, M. (2004). Decision making by variable contribution in discriminant, *Logit*, and regression analyses. *International Journal of Information Technology & Decision Making*, 3(02), 265-279.
- Long, M. S. (1976). Credit screening system selection. *Journal of Financial and Quantitative Analysis*, 11(02), 313-328.
- Majer, I. (2006). *Application scoring: Logit model approach and the divergence method compared* (Doctoral dissertation, Warsaw School of Economics).
- Malik, M., & Thomas, L. C. (2010). Modelling credit risk of portfolio of consumer loans. *Journal of the Operational Research Society*, 61(3), 411-420.
- Martens, D., Van Gestel, T., De Backer, M., Haesen, R., Vanthienen, J., & Baesens, B. (2010). Credit rating prediction using ant colony optimization. *Journal of the Operational Research Society*, 61(4), 561-573.
- Mileris, R. (2010). Estimation of loan applicants default probability applying discriminant analysis and simple Bayesian classifier. *Economics and management*, 15.
- Moreno Valencia, S. (2014). *El Modelo Logit Mixto para la construcción de un Scoring de Crédito* (Doctoral dissertation, Universidad Nacional de Colombia Sede Medellín). Costa et al (2012)
- Mures Quintana, M. J., García Gallego, A., & Vallejo Pascual, E. (2011). Aplicación del análisis discriminante y regresión logística en el estudio de la morosidad en las entidades financieras: comparación de resultados.
- Myers, J. H., & Forgy, E. W. (1963). The development of numerical credit evaluation systems. *Journal of the American Statistical Association*, 58(303), 799-806.
- Ochoa, J. C., Galeano, W., & Agudelo, L. G. (2011). Construcción de un modelo de scoring para el otorgamiento de crédito en una entidad financiera. *Perfil de Coyuntura Económica*, (16), 191-222.
- Orgler, Y. E. (1970). A Credit Scoring Model for Commercial Loans. *Journal of Money, Credit & Banking* (Ohio State University Press), 2(4).

- Palacio, A. P., Lochmüller, C., Murillo, J. G., Pérez, M. A., & Vélez, C. A. (2011). Modelo cualitativo para la asignación de créditos de consumo y ordinario-el caso de una cooperativa de crédito. *Revista Ingenierías Universidad de Medellín*, 10(19), 101-111.
- Pope, D. G., & Sydnor, J. R. (2011). What's in a Picture? Evidence of Discrimination from Prosper. com. *Journal of Human Resources*, 46(1), 53-92.
- Ramírez, F. O. P., & Castaño, H. F. (2007). Las redes neuronales y la evaluación del riesgo de crédito. *R. Ingenierías Universidad de Medellin*, 6(10), 77-91.
- Rayo Cantón, S., Lara Rubio, J., & Camino Blasco, D. (2010). Un Modelo de Credit Scoring para instituciones de microfinanzas en el marco de Basilea II. *Journal of Economics, Finance and Administrative Science*, 15(28), 89-124.
- Reichert, A. K., Cho, C. C., & Wagner, G. M. (1983). An examination of the conceptual issues involved in developing credit-scoring models. *Journal of Business & Economic Statistics*, 1(2), 101-114.
- Řezáč, M., & Řezáč, F. (2011). How to measure the quality of credit scoring models. *Finance a úvěr: Czech Journal of Economics and Finance*, 61(5), 486-507.
- Riera, M. Y., & Gando, P. (2005). El Valor En Riesgo Aplicado A Fondos De Inversion.
- Roszbach, K. (2004). Bank lending policy, credit scoring, and the survival of loans. *Review of Economics and Statistics*, 86(4), 946-958.
- Schreiner, M. (2002). Ventajas y desventajas del scoring estadístico para las microfinanzas. *microfinancerisk Management*, 1-40.
- Selau, L. P. R., & Ribeiro, J. L. D. (2011). A systematic approach to construct credit risk forecast models. *Pesquisa Operacional*, 31(1), 41-56.
- Smith, L. D., Bilir, C., Huang, V. W., Hung, K. Y., & Kaplan, M. (2005). Citibank models credit risk on hybrid mortgage loans in Taiwan. *Interfaces*, 35(3), 215-229.
- Sousa, M. R., Gama, J., & Brandão, E. (2013). *Introducing time-changing economics into credit scoring* (No. 513). Universidade do Porto, Faculdade de Economia do Porto.

Superintendencia Bancaria de Colombia (2002).Circular externa 11 de 2002. Gestión de riesgo de crédito

Verstraeten, G., & Van den Poel, D. (2005). The impact of sample bias on consumer credit scoring performance and profitability. *Journal of the operational research society*, 56(8), 981-992.

Villano, F. E. S. (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*, 29(129), 416-427.

Webster, G. R. (2011). *Bayesian logistic regression models for credit scoring*(Doctoral dissertation, Rhodes University).

Wiginton, J. C. (1980). A note on the comparison of *Logit* and discriminant models of consumer credit behavior. *Journal of Financial and Quantitative Analysis*, 15(03), 757-770

Zhai, H., & Russell, J. S. (1999). Stochastic modelling and prediction of contractor default risk. *Construction Management & Economics*, 17(5), 563-576.

Zhou, L., Lai, K. K., & Yen, J. (2009). Credit scoring models with AUC maximization based on weighted SVM. *International journal of information technology & decision making*, 8(04), 677-696.

## ANEXOS

### ANEXO 1: VARIABLES DESCRITAS POR AUTORES PARA *CREDIT SCORE*

Variable	Cuanti	Cuali	Autor
Activos	x		Ayús <i>et al</i> (2010);Crook, Hamilton, Thomas (1992);Chaudhuri, Cheral (2012);Gonçalves, Braga (2008)
Agencia de solicitud		x	Ochoa, Galeano, Agudelo (2010);Pope, Sydnor (2011)
Antigüedad entidad	x		Medina, Marti Selva (2012);Kocenda, Vojtek (2009);Moreno (2013);Apilado, Warner, Dauten (1974);Peña <i>et al</i> (2011);Rayo, Lara, Camino (2010);Crook, Hamilton, Thomas (1992)
Antigüedad laboral	x		Mures, Garcia, Vallejo (2005);Myers, Forgy (1963);Ochoa, Galeano, Agudelo (2010);Kocenda, Vojtek (2009);Moreno (2013);Peña <i>et al</i> (2011);Wiginton (1980);Crook, Hamilton, Thomas (1992); Baklouti(2014);Reichert, Cho, Wagner(1983);Webster (2011);Sousa, Gama, Brandão (2013);Espin-García, Rodríguez-Caballero (2013)
Antigüedad crédito	x		Malik, Thomas (2010);Verstraeten, Van den Poel (2005)
Apalancamiento	x		Orgler (1970);Peña <i>et al</i> (2011);Rayo, Lara, Camino (2010);Gonçalves, Braga (2008)
Bancarrotas		x	Myers, Forgy (1963)
Calificación central de riesgo		x	Moreno (2013);Rayo, Lara, Camino (2010);Roszbach (2004); Brevoort, Avery, Canner (2013);Malik, Thomas (2010);Verstraeten, Van den Poel (2005); Pope, Sydnor (2011)
Cambio residencial reciente		x	Wiginton (1980);Myers, Forgy (1963);Roszbach (2004)
Capacidad de pago	x		Ochoa, Galeano, Agudelo (2010);Espin-García, Rodríguez-Caballero (2013)
Capital de trabajo	x		Orgler (1970);Gonçalves, Braga (2008)
Capital de trabajo/Activos corrientes	x		Orgler (1970);Altman (1968);Gonçalves, Braga (2008)
Contrato de anticipo	x		Myers, Forgy (1963)
Créditos anteriores (cantidad)	x		Rayo, Lara, Camino (2010);Salazar (2013);Myers, Forgy (1963);Moreno (2013);Mures, Garcia, Vallejo (2005);Baklouti(2014); Roszbach (2004); Brevoort, Avery, Canner (2013); Sousa, Gama, Brandão (2013)
Créditos último año	x		Rayo, Lara, Camino (2010);Moreno (2013);Sousa, Gama, Brandão (2013)
Créditos denegados	x		Rayo, Lara, Camino (2010);Roszbach (2004); Brevoort, Avery, Canner (2013)
Cuotas en Mora	x		Rayo, Lara, Camino (2010);Baklouti(2014); Han (2011); Brevoort, Avery, Canner (2013); Moreno (2013); Roszbach (2004)
Cuotas pagadas en historial	x		Rayo, Lara, Camino (2010);Sousa, Gama, Brandão (2013)
Destino del crédito	x		Medina, Marti Selva (2012);Mures, Garcia, Vallejo (2005);Kocenda, Vojtek (2009);Apilado, Warner, Dauten (1974);Peña <i>et al</i> (2011);Rayo, Lara, Camino (2010);Ochoa, Galeano, Agudelo (2010);Myers, Forgy (1963);Baklouti(2014); Webster (2011); Han (2011); Pope, Sydnor (2011)
DTF	x		Fernandez, Perez (2005)
Edad	x		Medina, Marti Selva (2012);Mures, Garcia, Vallejo (2005);Myers, Forgy (1963);Ochoa, Galeano, Agudelo (2010);Kocenda, Vojtek (2009);Moreno (2013);Apilado, Warner, Dauten (1974)Peña <i>et al</i> (2011);Esteve (2007);Rayo, Lara, Camino (2010); Ayús <i>et al</i> (2010);Crook, Hamilton, Thomas (1992); Baklouti(2014); Roszbach (2004); Alessie, Hochguertel, Weber (2005); Constangioara (2011); Brevoort, Avery, Canner (2013); Chaudhuri, Cheral (2012); Sousa, Gama, Brandão (2013); Espin-García, Rodríguez-Caballero (2013); Pope, Sydnor (2011)
Egresos	x		Peña <i>et al</i> (2011);Crook, Hamilton, Thomas (1992)
Estado civil		x	Medina, Marti Selva (2012);Mures, Garcia, Vallejo (2005);Myers, Forgy (1963);Ochoa, Galeano, Agudelo (2010);Kocenda, Vojtek (2009);Moreno (2013);Apilado, Warner, Dauten (1974);Peña <i>et al</i> (2011);Rayo, Lara, Camino

			(2010);Baklouti(2014); Roszbach (2004); Alessie, Hochguertel, Weber (2005); Han (2011); Brevoort, Avery, Canner (2013); Sousa, Gama, Brandão (2013); Espin-García, Rodríguez-Caballero (2013); Pope, Sydnor (2011)
Estrato social		x	Ochoa, Galeano, Agudelo (2010);Kocenda, Vojtek (2009);Moreno (2013);Peña <i>et al</i> (2011);Wiginton (1980)
Ganancia neta	x		Orgler (1970);Bellotti, Crook (2007);Altman (1968)
Garantías		x	Mures, García, Vallejo (2005);Kocenda, Vojtek (2009);Moreno (2013);Apilado, Warner, Dauten (1974);Peña <i>et al</i> (2011);Rayo, Lara, Camino (2010)
Gastos en compra a crédito	x		Esteve (2007)
Gastos en tarjeta de crédito	x		Esteve (2007);Crook, Hamilton, Thomas (1992)
Gastos hipoteca	x		Esteve (2007)
Gastos préstamo	x		Esteve (2007)
Genero		x	Ochoa, Galeano, Agudelo (2010);Salazar (2013);Kocenda, Vojtek (2009);Moreno (2013);Apilado, Warner, Dauten (1974);Peña <i>et al</i> (2011);Rayo, Lara, Camino (2010);Baklouti(2014); Roszbach (2004); Han (2011); Brevoort, Avery, Canner (2013); Chaudhuri, Cheral (2012); Sousa, Gama, Brandão (2013); Espin-García, Rodríguez-Caballero (2013); Pope, Sydnor (2011)
Hijos	x		Myers, Forgy (1963);Peña <i>et al</i> (2011);Esteve (2007); Crook, Hamilton, Thomas (1992);Alessie, Hochguertel, Weber (2005); Chaudhuri, Cheral (2012)
Índice bolsa (IGBC)	x		Rayo, Lara, Camino (2010);Bellotti, Crook (2007)
Índice de vivienda	x		Bellotti, Crook (2007)
Ingresos	x		Medina, Marti Selva (2012);Mures, García, Vallejo (2005);Myers, Forgy (1963);Ochoa, Galeano, Agudelo (2010);Kocenda, Vojtek (2009);Moreno (2013);Apilado, Warner, Dauten (1974)Esteve (2007); Ayús <i>et al</i> (2010);Crook, Hamilton, Thomas (1992); Roszbach (2004); Alessie, Hochguertel, Weber (2005); Constangioara (2011); Han (2011); Chaudhuri, Cheral (2012); Sousa, Gama, Brandão (2013); Espin-García, Rodríguez-Caballero (2013);
Inventarios/activos corrientes	x		Orgler (1970);Gonçalves, Braga (2008)
IPC	x		Fernandez, Perez (2005);Rayo, Lara, Camino (2010);Bellotti, Crook (2007);Malik, Thomas (2010)
Liquidez	x		Orgler (1970);Rayo, Lara, Camino (2010);Salazar (2013); Ayús <i>et al</i> (2010)
Margen de ganancia	x		Orgler (1970);Mileris (2010)
Máximo Mora	x		Moreno (2013);Rayo, Lara, Camino (2010)
Monto		x	Medina, Marti Selva (2012);Mures, García, Vallejo (2005);Ochoa, Galeano, Agudelo (2010);Salazar (2013);Kocenda, Vojtek (2009);Moreno (2013);Apilado, Warner, Dauten (1974);Medina, Marti Selva (2012);Peña <i>et al</i> (2011);Crook, Hamilton, Thomas (1992); Baklouti(2014); Reichert, Cho, Wagner(1983); Webster (2011); Verstraeten, Van den Poel (2005); Pope, Sydnor (2011)
Monto rechazado	x		Rayo, Lara, Camino (2010);Webster (2011)
Nivel educativo		x	Ochoa, Galeano, Agudelo (2010);Kocenda, Vojtek (2009);Moreno (2013);Peña <i>et al</i> (2011);Baklouti(2014); Constangioara (2011); Espin-García, Rodríguez-Caballero (2013)
Numero de codeudores	x		Kocenda, Vojtek (2009);Roszbach (2004)
Ocupación	x		Mures, García, Vallejo (2005);Myers, Forgy (1963);Ochoa, Galeano, Agudelo (2010);Moreno (2013);Peña <i>et al</i> (2011);Kocenda, Vojtek (2009);Rayo, Lara, Camino (2010)
Otros ingresos	x		Myers, Forgy (1963);Esteve (2007);Roszbach (2004); Alessie, Hochguertel, Weber (2005); Constangioara (2011);
Personas dependientes	x		Wiginton (1980);Myers, Forgy (1963);Ochoa, Galeano, Agudelo (2010);Kocenda, Vojtek (2009);Moreno (2013);Apilado, Warner, Dauten (1974);Peña <i>et al</i> (2011);Esteve (2007);Crook, Hamilton, Thomas (1992); Han (2011); Sousa, Gama, Brandão (2013); Espin-García, Rodríguez-Caballero (2013)
PEA en casa	x		Moreno (2013)

PIB	x		Fernandez, Perez (2005);Rayo, Lara, Camino (2010);Bellotti, Crook (2007);Malik, Thomas (2010)
Plazo	x		Medina, Marti Selva (2012);Mures, Garcia, Vallejo (2005);Ochoa, Galeano, Agudelo (2010);Salazar (2013);Kocenda, Vojtek (2009);Moreno (2013);Apilado, Warner, Dauten (1974);Rayo, Lara, Camino (2010);Rayo, Lara, Camino (2010);Reichert, Cho, Wagner(1983); Brevoort, Avery, Canner (2013); Pope, Sydnor (2011)
Productividad bruta	x		Rayo, Lara, Camino (2010)
Promedio Mora	x		Moreno (2013);Rayo, Lara, Camino (2010)
Pronostico del analista		x	Rayo, Lara, Camino (2010)
Prueba acida	x		Orgler (1970);Mileris (2010)
Razón corriente	x		Orgler (1970);Peña <i>et al</i> (2011); Ayús <i>et al</i> (2010);Mileris (2010); Deng, Gabriel (2006)
Razón de endeudamiento	x		Moreno (2013);Rayo, Lara, Camino (2010);Peña <i>et al</i> (2011); Ayús <i>et al</i> (2010);Mileris (2010); Deng, Gabriel (2006); Verstraeten, Van den Poel (2005); Pope, Sydnor (2011);
Razón de patrimonio neto	x		Orgler (1970); Ayús <i>et al</i> (2010)
Lugar residencia del cliente		x	Kocenda, Vojtek (2009);Moreno (2013);Peña <i>et al</i> (2011);Rayo, Lara, Camino (2010);Peña <i>et al</i> (2011);Alessie, Hochguertel, Weber (2005); Chaudhuri, Cheral (2012); Sousa, Gama, Brandão (2013); Espin-García, Rodríguez-Caballero (2013);
ROA (Rentabilidad de activos)	x		Orgler (1970);Altman (1968);Rayo, Lara, Camino (2010);Webster (2011);
ROE (Rentabilidad del patrimonio)	x		Rayo, Lara, Camino (2010)
Rotación de activos	x		Orgler (1970);Rayo, Lara, Camino (2010)
Rotación de cartera	x		Orgler (1970)
Rotación inventario	x		Orgler (1970)
Salario	x		Peña <i>et al</i> (2011)
Actividad/Sector económico		x	Salazar (2013);Moreno (2013);Rayo, Lara, Camino (2010);Wiginton (1980);Peña <i>et al</i> (2011); Ayús <i>et al</i> (2010);Baklouti(2014); Roszbach (2004); Chaudhuri, Cheral (2012); Verstraeten, Van den Poel (2005);
Servicio Agua		x	Rayo, Lara, Camino (2010)
Servicio de Teléfono		x	Esteve (2007);Rayo, Lara, Camino (2010);Crook, Hamilton, Thomas (1992); Sousa, Gama, Brandão (2013)
Servicio Energía		x	Rayo, Lara, Camino (2010)
Tasa de cambio	x		Rayo, Lara, Camino (2010);Malik, Thomas (2010)
Tasa de desempleo	x		Bellotti, Crook (2007);Fernandez, Perez (2005);Rayo, Lara, Camino (2010);Han (2011); Malik, Thomas (2010)
Tasa de interés	x		Salazar (2013);Rayo, Lara, Camino (2010);Bellotti, Crook (2007);Roszbach (2004); Han (2011); Sousa, Gama, Brandão (2013); Malik, Thomas (2010)
Tasa del crédito	x		Han (2011);Salazar (2013); ;Rayo, Lara, Camino (2010); ;Bellotti, Crook (2007); Roszbach (2004); Brevoort, Avery, Canner (2013); Sousa, Gama, Brandão (2013)
Tenencia del préstamo		x	Medina, Marti Selva (2012);Myers, Forgy (1963);;Han (2011)
Tipo de compañía		x	Wiginton (1980);Han (2011)
Tipo de contrato		x	Ochoa, Galeano, Agudelo (2010);Kocenda, Vojtek (2009);Peña <i>et al</i> (2011); Crook, Hamilton, Thomas (1992);Roszbach (2004); Webster (2011); Constangioara (2011); Chaudhuri, Cheral (2012); Sousa, Gama, Brandão (2013)
Tipo de vivienda		x	Ochoa, Galeano, Agudelo (2010);Moreno (2013);Apilado, Warner, Dauten (1974);Myers, Forgy (1963);Crook, Hamilton, Thomas (1992); Alessie, Hochguertel, Weber (2005); Chaudhuri, Cheral (2012); Sousa, Gama, Brandão (2013); Verstraeten, Van den Poel (2005); Espin-García, Rodríguez-Caballero (2013)
Valor Hogar	x		Esteve (2007);Crook, Hamilton, Thomas (1992); Roszbach (2004)
Vehículo		x	Wiginton (1980);Myers, Forgy (1963);Peña <i>et al</i> (2011);Espin-García, Rodríguez-Caballero (2013)

Balance de amortización	x		Crook, Hamilton, Thomas (1992);Roszbach (2004); Webster (2011); Han (2011)
Tiempo en el Hogar	x		Crook, Hamilton, Thomas (1992);Reichert, Cho, Wagner(1983); Sousa, Gama, Brandão (2013); Deng, Gabriel (2006); Espin-García, Rodríguez-Caballero (2013)
Nacionalidad		x	Roszbach (2004)
Código postal	x		Roszbach (2004);Sousa, Gama, Brandão (2013)
Profesión		x	Alessie, Hochguertel, Weber (2005);Espin-García, Rodríguez-Caballero (2013)
Raza		x	Brevoort, Avery, Canner (2013);Han (2011); Deng, Gabriel (2006); Pope, Sydnor (2011)
Otros créditos		x	Sousa, Gama, Brandão (2013);Deng, Gabriel (2006); Verstraeten, Van den Poel (2005)

## ANEXO 2: ESTADISTICA DESCRIPTIVA DE LAS VARIABLES CUANTITATIVAS DE LOS CLIENTES

Tabla A2: ESTADISTICA DE LAS VARIABLES ANALIZADAS EN CASOS DE NO RIESGO: TOTAL: 6692 clientes							
Variable	Media	Desv.Est.	Q1	Mediana	Q3	IQR	Kurtosis
saldoin(millones)	5,664	11,08	1,5	2,6	5	3,5	52,26
logsaldo	0,428	0,501	0,176	0,415	0,699	0,523	0,77
saldofin(millones)	3,063	8,728	0	0,81	2,668	2,668	76,62
plazo	27,27	14,08	15	24	36	21	-0,75
tipagomeses	10,19	8,77	3,584	7,89	13,447	9,863	1,85
cuotaspagas	9,498	8,573	3	7	13	10	1,83
cantiatra	0,135	0,342	0	0	0	0	2,58
maxdias	1,354	4,3939	0	0	0	0	15,14
salario(millones)	0,513	1,335	0	0	0,630	0,630	103,32
ingresos(millones)	2,8	4,9964	0,98	1,6	3	2,02	118,33
edad	43	13	33	43	53	20	-0,8
tientidadmeses	31,758	50,644	5,257	10,513	33,478	28,222	8,11
tasae fec	0,266	0,056	0,234	0,265	0,297	0,063	-0,480
ipc	0,151	0,246	0,040	0,140	0,290	0,250	0,470
tasadesocupnal	0,101	0,014	0,092	0,099	0,111	0,019	-0,230
tasadesocupdep	0,155	0,025	0,138	0,149	0,168	0,030	0,220
tasainterv	0,040	0,008	0,038	0,038	0,045	0,008	11,680
tasausura	0,365	0,105	0,298	0,312	0,502	0,204	-1,160
pibvar	1,152	0,667	0,804	0,902	1,368	0,564	0,500
pibdemcons	0,067	0,086	0,001	0,076	0,148	0,147	-0,980
pibdemhogar	0,021	0,016	0,009	0,016	0,027	0,018	0,020
pibdem	0,022	0,025	0,012	0,022	0,026	0,014	0,670
pibofer	0,018	0,018	0,007	0,013	0,027	0,020	0,840
pibbiendur	0,014	0,013	0,006	0,008	0,032	0,026	-0,780
pibbiensemidur	0,0281	0,0376	0,005	0,015	0,051	0,046	2,1
pibserv	0,0192	0,0155	0,011	0,011	0,024	0,013	-0,16
pibbiennodur	0,0673	0,0859	0,001	0,076	0,148	0,147	-0,98

<b>Tabla A3: ESTADISTICA DE LAS VARIABLES ANALIZADAS EN CASOS DE RIESGO</b>							
<b>TOTAL: 1972 clientes</b>							
Variable	Media	Desv.Est.	Q1	Mediana	Q3	IQR	Kurtosis
saldoin(millones)	5,163	9,053	1,3	2,5	5	3,700	26,940
logsaldo	0,397	0,510	0,114	0,398	0,699	0,585	0,680
saldofin(millones)	1,661	5,339	0	0	1,111	1,111	45,95
plazo	29,12	14,11	18	24	36	18	-0,91
tipagomeses	18,58	10,67	10,69	16,11	24,99	14,30	0,24
cuotaspagas	15,29	10,20	8	12	22	14	0,33
cantiatra	6,521	5,692	3	5	9	6	5,25
maxdias	75	103	15	40	100	85	24
salario(millones)	0,559	1,620	0	0	0,6	0,6	140
ingresos(millones)	2,852	3,959	0,95	1,7	3,2	2,25	58,14
edad	40,60	12,501	30	40	50	20	-0,78
tientidadmeses	28,53	49,040	5,55	10,53	26,68	21,13	10,91
tasaefec	0,259	0,052	0,224	0,265	0,293	0,069	0,07
ipc	0,214	0,252	0,040	0,160	0,310	0,270	0,28
tasadesocupnal	0,109	0,013	0,099	0,109	0,118	0,019	-0,12
tasadesocupdep	0,169	0,025	0,153	0,166	0,188	0,035	0,15
tasainterv	0,042	0,011	0,035	0,040	0,048	0,013	5,77
tasasura	0,343	0,095	0,283	0,312	0,369	0,086	-0,39
pibvar	1,186	0,717	0,804	1,134	1,470	0,666	0,31
pibdemcons	0,057	0,093	-0,027	0,044	0,094	0,121	-0,79
pibdemhogar	0,017	0,016	0,008	0,011	0,018	0,010	1,67
pibdem	0,020	0,023	0,004	0,019	0,024	0,019	1,92
pibofer	0,017	0,016	0,009	0,011	0,017	0,008	3,34
pibbiendur	0,011	0,013	0,005	0,007	0,013	0,008	-0,05
pibbiensemibur	0,028	0,043	0,005	0,019	0,051	0,046	0,85
pibserv	0,016	0,014	0,007	0,011	0,015	0,008	2,17
pibbiennodur	0,057	0,093	-0,027	0,044	0,094	0,121	-0,79

### ANEXO 3: ESTADISTICA DESCRIPTIVA DE LAS VARIABLES CUALITATIVAS DE LOS CLIENTES

<b>Tabla A4: ESTADISTICA DE LAS VARIABLES CUALITATIVAS ANALIZADAS</b>				
	No Riesgo	6692	Riesgo	1972
<b>Pagaré</b>				
Codeudor	1567	23,42%	679	34,43%
Hipoteca	379	5,66%	146	7,40%
Pagaré	281	4,20%	136	6,90%
Pagaré sin codeudor	3637	54,35%	800	40,57%
Prenda	828	12,37%	211	10,70%
<b>Código Línea</b>				
Comercial	92	1,37%	16	0,81%
Consumo	4171	62,33%	1209	61,31%
Microcrédito	2429	36,30%	747	37,88%
<b>Código Destino</b>				
Microcrédito C	926	13,84%	606	30,73%
Ordinarios	3352	50,09%	991	50,25%
Otros créditos comerciales	47	0,70%	13	0,66%
Otros créditos de consumo	506	7,56%	107	5,43%
Otros Microcréditos	104	1,55%	27	1,37%
Rotativo I	313	4,68%	111	5,63%
Vehicular	45	0,67%	3	0,15%

Woccu	1399	20,91%	114	5,78%
Agencia				
Chinchiná	329	4,92%	49	2,48%
Fundadores	267	3,99%	15	0,76%
La tebaida	216	3,23%	14	0,71%
Minorista	765	11,43%	326	16,53%
Pereira centro	826	12,34%	536	27,18%
Principal	2653	39,64%	753	38,18%
Quimbaya	621	9,28%	72	3,65%
Santa Rosa	179	2,67%	14	0,71%
Tuluá	836	12,49%	193	9,79%
Declara Ingresos				
No	891	13,31%	131	6,64%
Si	5801	86,69%	1841	93,36%
Declara Egresos				
No	1028	15,36%	316	16,02%
Si	5664	84,64%	1656	83,98%
Actividad Económica				
agrícola	208	3,11%	27	1,37%
comercial	2645	39,52%	366	18,56%
industrial	242	3,62%	34	1,72%
No indica	3597	53,75%	1545	78,35%
Tipo de Contrato				
asociado	50	0,75%	10	0,51%
Contrato indefinido	1401	20,94%	328	16,63%
independiente	2332	34,85%	806	40,87%
otro	2576	38,49%	759	38,49%
Termino fijo	333	4,98%	69	3,50%
Genero				
femenino	3284	49,07%	903	45,79%
masculino	3408	50,93%	1069	54,21%
Estado Civil				
casado	2535	37,88%	656	33,27%
divorciado	255	3,81%	93	4,72%
otro	24	0,36%	4	0,20%
separado	170	2,54%	66	3,35%
soltero	2182	32,61%	679	34,43%
Unión libre	1365	20,40%	437	22,16%
viudo	161	2,41%	37	1,88%
Residencia del Cliente				
Armenia	2502	37,39%	626	31,74%
Caldas	337	5,04%	57	2,89%
Otros	43	0,64%	21	1,06%
Pereira	1259	18,81%	700	35,50%
Quindío	1140	17,04%	177	8,98%
Risaralda	244	3,65%	59	2,99%
Tuluá	680	10,16%	152	7,71%
Valle del Cauca	487	7,28%	180	9,13%
Escolaridad				
Ninguno	15	0,22%	1	0,05%
Otros	434	6,49%	122	6,19%
Primaria	1328	19,84%	340	17,24%
Secundaria	3021	45,14%	874	44,32%
Técnico	757	11,31%	229	11,61%
Universitario	1137	16,99%	406	20,59%

## ANEXO 4: BONDAD DE AJUSTE DEL MODELO ECONOMETRICO

Las siguientes son los resultados de ajuste del modelo:

**Tabla A5: Resultados Generales y Bondad de Ajuste**

Log(Pseudo - R <sup>2</sup> )	0,941
Deviance difference	90%
Hosmer-Lemeshow C statistic	1,161e-05
Hosmer-Lemeshow H statistic	0,05565
le Cessie-van Houwelingen-Copas-Hosmer	5,999e-08
AUROC	0,99
Cut-off ROC	0,98

*Fuente: elaboración propia*

Respecto a la *deviance*, el modelo explica el 90% de los errores totales; los test de Hosmer se encuentra que el test Hosmer H, tiene significancia superior al 0,05 en  $X^2$ , haciendo al modelo significativo en globalidad.