

**PRESENTACIÓN PROYECTO DE GRADO  
“SCORING DE CRÉDITO: HERRAMIENTA PARA LA EVALUACIÓN DE RIESGO DE  
CRÉDITO EN ENTIDADES FINANCIERAS”**

MARIA CAMILA CAICEDO CARBONELL

PONTIFICIA UNIVERSIDAD JAVERIANA  
FACULTAD DE CIENCIAS ECONÓMICAS Y ADMINISTRATIVAS  
MAESTRÍA EN ECONOMÍA  
Bogotá D.C, Abril de 2018

**PRESENTACIÓN PROYECTO DE GRADO  
“SCORING DE CRÉDITO: HERRAMIENTA PARA LA EVALUACIÓN DE RIESGO DE  
CRÉDITO EN ENTIDADES FINANCIERAS”**

Trabajo de grado presentado

por

María Camila Caicedo Carbonell

a

A LA DIRECCIÓN DE POSGRADOS EN ECONOMÍA:  
Atención: Dr. FLAVIO JÁCOME LIEVANO

PONTIFICIA UNIVERSIDAD JAVERIANA

MAESTRIA EN ECONOMÍA

BOGOTÁ D.C., ABRIL DE 2018

En cumplimiento parcial de los requisitos  
para optar por el grado Magister (Magistra) en Economía

Bogotá D.C, Abril de 2018

# SCORING DE CRÉDITO: HERRAMIENTA PARA LA EVALUACIÓN DE RIESGO DE CRÉDITO EN ENTIDADES FINANCIERAS

## TABLA DE CONTENIDO

1. Introducción.....	8
2. Marco Teórico.....	16
3. Metodología.....	28
4. Datos y Descripción de Variables .....	33
5. Descripción de la Muestra.....	33
6. Análisis de Resultados.....	35
7. Conclusiones.....	39
8. Referencias y Bibliografía.....	40

Anexos

## **LISTADO DE TABLAS**

1.1 Tipos de clientes y características

1.2 Criterios a la hora de otorgar riesgo de crédito

5.1 Variables empleadas en el modelo

6.1 Estadísticas Descriptivas

6.2 Resultados Econométricos

## **LISTADO DE GRÁFICAS**

- A. Número Morosos VS Inflación
- B. Número Morosos VS Tasa de Desempleo
- C. Número Morosos VS Tasa Repo

# SCORING DE CRÉDITO: HERRAMIENTA PARA LA EVALUACIÓN DE RIESGO DE CRÉDITO EN ENTIDADES FINANCIERAS\*

Trabajo de grado presentado por \*\*:

MARIA CAMILA CAICEDO CARBONELL

## Resumen

El *scoring* de crédito utilizado por algunas entidades financieras, es un modelo estadístico que ordena a los clientes según sus características de perfil para asignarles una puntuación o una probabilidad de mora. Con el presente documento se busca mejorar dicha herramienta teniendo en cuenta variables macroeconómicas; así mismo lograr una herramienta más robusta en el análisis que de cierta manera ayude atenuar contingencias que puedan afectar considerablemente la cartera de la institución financiera. Se emplea un modelo lineal generalizado para datos panel con el fin de determinar que variables macroeconómicas ayudarían a predecir mejor un eventual impago. El trabajo encuentra evidencia empírica a favor de la significancia de algunas variables macroeconómicas (tasa interés (-), inflación (+) y desempleo (+)).

**Palabras clave:** scoring de crédito, herramienta, riesgo, banco, cliente, institución financiera.

---

\*Trabajo presentado para optar por el título de Magister de economía de la Pontificia Universidad Javeriana.

\*\*Agradezco la Dirección del Profesor David Andrés Londoño Bedoya.

# CREDIT SCORING: TOOL USED FOR THE CREDIT RISK EVALUATION IN SOME FINANCIAL INSTITUTIONS \*

Thesis presented by\*\*:

MARIA CAMILA CAICEDO CARBONELL

## Abstract

The credit scoring used by some financial institutions is a statistical model that orders customers according to their profile characteristics to assign a moratorium score or probability. This document seeks to improve the tool taking in mind the macroeconomic variables, and also to achieve a stronger tool in the analysis, that somehow helps to mitigate contingencies that significantly affect the financial institution portfolio. A generalized linear model for panel data is used to determine which macroeconomic variables would help to better predict a possible default. The paper finds empirical evidence in favor of some macroeconomic variables significance (interest rate (-), inflation (+) and unemployment (+)).

**Key words:** credit scoring, tool, risk, bank, customer, financial institution.

---

\*Paper presented to qualify for the Master's degree in economics from the Pontificia Universidad Javeriana.

\*\*I am grateful to the Professor David Andrés Londoño Bedoya guidance.

## 1. INTRODUCCIÓN

En Colombia el sector financiero se encuentra regulado por la Superintendencia Financiera de Colombia que tiene la misión de preservar la confianza pública y la estabilidad del sistema financiero. Sin embargo, existen altas probabilidades de que ocurran eventos financieros inesperados como una crisis económica que puede generar insolvencia económica en los países. Por lo anterior, es que en tiempos de crisis es necesario tener un *Scoring* de Crédito<sup>1</sup> robusto que permita determinar qué impacto tiene un sector determinado que entró en crisis, o la inflación presentada por el Departamento Administrativo Nacional de Estadística (DANE), el desempleo, o la tasa de interés u otros factores sobre la solvencia del cliente para el pago de las obligaciones adquiridas con las entidades financieras, tales como: los bancos.

Los modelos tradicionales de *scoring* de crédito evalúan aspectos básicos acerca de los individuos que están solicitando determinado préstamo, y tienen en cuenta conceptos como pérdida esperada y pérdida no esperada (Elizondo y López, 1999). Se encuentran dos tipos de modelos, uno que se basa en la ponderación de algunos factores que hacen referencia a determinantes de incumplimiento de las obligaciones adquiridas por un individuo y el otro que se basa en conceptos de tipo fundamental. Estos modelos involucran el criterio subjetivo de un analista que tiene la experiencia en la evaluación para otorgar un riesgo de crédito.

Los modelos anteriormente nombrados, han perdido de manera significativa su uso debido a los avances en las herramientas de análisis de riesgo de crédito y los cambios financieros, de esta manera, existen en la actualidad modelos más sofisticados que utilizan técnicas estadísticas y probabilísticas más avanzadas; sin embargo, dichos modelos se siguen utilizando como complemento de los más sofisticados.

Según Sepúlveda, C., Reina, W. y Gutiérrez, J. (2012), uno de los factores más importantes para las entidades financieras es el análisis de riesgo o volatilidad, que está presente en todas las decisiones que afrontan las organizaciones, por lo que es importante para dichas

---

<sup>1</sup> Un *scoring* es un sistema automático de ayuda a la toma de decisiones crediticias. O dicho de un modo más sencillo aún, un programa informático que, partiendo de una información dada, recomendará la aprobación o no de una operación de financiación. En definitiva, es una analista de riesgos con corazón de silicio, una suerte de sistema de inteligencia artificial que busca, en la medida de lo posible, eliminar el sesgo humano y homogeneizar las respuestas a las solicitudes de crédito.

entidades, saberlo, identificarlo y medirlo de manera eficiente, con el fin que se puedan implementar estrategias y planes que permitan minimizarlo.

Por su parte, Domínguez, J., Miranda, F., Pallas, J. y Peraza, C., definen el riesgo como la posibilidad de incurrir en pérdidas como consecuencia del incumplimiento por parte del deudor de las obligaciones en las operaciones de intermediación crediticia. A su vez, señalan que el riesgo de crédito se puede dividir en dos tipos: el riesgo de insolvencia y el riesgo-país; el primero hace referencia a la situación económica financiera del deudor y su incapacidad para cumplir con el pago de las obligaciones adquiridas y el segundo hace referencia al grado de solvencia (o insolvencia) del total de contrapartidas que pertenecen a un área geopolítica legalmente definida como Estado.

En Colombia, el sistema financiero ha sido muy diverso debido a que se tienen diferentes medios que proporcionan el acceso al comienzo de la vida crediticia y de la misma manera permiten conocer el perfil de determinado cliente y ayudan a calcular así su probabilidad de incumplimiento, teniendo en cuenta una serie de variables que se seleccionan según un estudio específico. Además el sistema financiero ofrece una gran variedad de servicios para los agentes económicos tanto a nivel macroeconómico y microeconómico como de carácter público y privado. La estructura que involucra a este sistema hace que se asignen los recursos monetarios de aquellos agentes ahorradores hacia los agentes deficitarios que demandan dinero para satisfacer algunas necesidades, pero también se debe tener en cuenta el riesgo que tienen dichas entidades al momento de prestar dinero independientemente del monto o tipo de producto.

Como ya se mencionó, para medir el riesgo de crédito de los diferentes clientes se emplea una herramienta llamada *Scoring* de Crédito, con el cual se evalúan una serie de variables microeconómicas, que en conjunto arrojan un dictamen para la operación y se complementa el otorgamiento de Riesgo de Crédito por medio de un analista especializado en el tema que valida la información suministrada y toma una decisión final acorde al perfil y la capacidad de pago que tenga el cliente para poder cumplir con la obligación.

Según la Superintendencia Financiera de Colombia, “*el riesgo crediticio es la posibilidad de que una entidad incurra en pérdidas y se disminuya el valor de sus activos, como consecuencia de que sus deudores fallen en el cumplimiento oportuno o cumplan imperfectamente los términos acordados en los contratos de crédito*”. Es por esto, que se debe seguir una metodología con determinados parámetros que fija la Superintendencia y otros por autonomía de la entidad financiera que va a otorgar el riesgo de crédito.

La metodología propuesta, debe contemplar algunos de estos dos elementos:

-Probabilidad de deterioro o probabilidad de no pago.

-La estimación o cuantificación de la pérdida esperada en la cual incurriría la entidad, durante un período de tiempo determinado.

Cabe resaltar que las pérdidas esperadas aumentan en función del monto del crédito solicitado por el individuo y de la probabilidad de deterioro asignada a cada activo. Entre más alta es la tasa de recuperación esperada, las pérdidas son menores. Las provisiones se calculan con respecto a estas pérdidas esperadas.

Estudiar el comportamiento de los individuos es algo que no se puede cuantificar, por lo que al momento de medir el riesgo no se sabe si el cliente efectivamente va a pagar la obligación que acaba de adquirir con determinada entidad financiera, o si por el contrario, decide no hacerlo. Es allí, donde se busca una forma de hacer más robusto el modelo que se tiene actualmente en una Institución Financiera Bancaria en Colombia (IFBEC) para otorgar Riesgo de Crédito a clientes particulares, es decir, personas naturales, teniendo en cuenta variables macroeconómicas que se focalicen en el sector económico en el cual se desempeña el cliente tales como: tasa de desempleo, precios de materias primas, producto interno bruto (PIB), tasa de interés e inflación.

Es importante estudiar y evaluar los efectos del riesgo sobre los créditos a través de herramientas o programas, ya que en el sistema financiero se encuentran diferentes perfiles de clientes, como se detalla en la tabla siguiente, con determinadas características y comportamiento de pago de sus obligaciones.

Además, incluir cambios en este tipo de análisis, podrían implicar mejoras estructurales sobre las decisiones de préstamo de las Entidades Financieras, garantizando menores pérdidas y mayores controles por parte de las autoridades competentes.

### 1.1 Tipos de clientes y características.

TIPOS DE CLIENTES Y CARACTERÍSTICAS		
TIPO DE CLIENTE	CONTRATO	DESCRIPCIÓN
AUTÓNOMO	INDEPENDIENTE	Aquellos individuos que trabajan en una actividad independiente de la cual obtienen un rendimiento, algunos de estos individuos se ven obligados a declarar renta que va acorde a sus ingresos operacionales anuales, patrimonio bruto y consumos mediante tarjetas de crédito; en cambio hay otros que no se ven obligados debido a que no cumplen con los topes exigidos por la ley para hacerlo, pero para estos se dificulta otorgar riesgo de crédito ya que solo tienen como soporte de ingresos sus estados financieros, y esto para algunas entidades financieras no es un perfil admisible. Además, los que realizan como tal una actividad de comercio o tienen algún establecimiento se ven obligados a registrar en la Cámara de Comercio para asegurar que su actividad está activa y renovada a determinado año y contar con el Registro Único Tributario (RUT) para validar la antigüedad laboral.
TEMPORALES	POR OBRA O LABOR (CONTRATO TÉRMINO FIJO)	Son aquellos individuos que tienen un contrato laboral a un término fijo que puede ser renovable o no a un determinado plazo que está estipulado en el contrato, de igual manera se establece el valor a pagar mensualmente y se debe tener en cuenta que en cualquier momento el contrato puede terminar antes de lo que está determinado.
PRESTADORES DE SERVICIOS	PRESTACIÓN DE SERVICIOS (CONTRATO TÉRMINO FIJO)	Son aquellos individuos que desarrollan su actividad en un plazo determinado para determinada compañía y en su contrato se estipula el tiempo de duración, el valor total del mismo y cuál va a ser el valor a pagar mensualmente.
ASALARIADOS	INDEFINIDO	Son aquellos individuos que tienen un contrato laboral con una entidad determinada y con un tipo de contrato a término indefinido, lo que significa que no hay un tiempo estipulado de terminación del mismo a menos que el empleado decida renunciar o incurra en una falta grave que lleve a terminación del contrato. Estos individuos ganan un salario fijo estipulado en el contrato, en donde algunas empresas pagan a sus empleados más de 12 salarios anuales, las que no, pagan un salario integral. Este tipo de clientes pueden tener domiciliada la nómina al banco al cual están solicitando el préstamo, con lo que de una u otra forma se facilita más a la hora de otorgar un préstamo ya que la cuota del crédito que adquirió se le descuenta de sus abonos de nómina y así el cliente mensualmente sabe que le están descontando y el banco tiene la certeza de que se le va a pagar el dinero.
RENTISTA	INDEPENDIENTE	Son aquellos que se dedican a alquilar propiedades con bienes propios o arrendados, mediante un contrato que firman tanto el arrendador como el arrendatario y en el cual se estipulan una serie de cláusulas entre las cuales se disponen la duración y renovación del contrato, así mismo como la posible terminación anticipada y sus sanciones; para este tipo de individuos sus ingresos se validan mediante la declaración de renta o contratos de arrendamiento. Es relevante para este segmento de clientes solicitar el certificado de tradición y libertad de vivienda para corroborar que la propiedad se encuentra libre de embargos judiciales y que está efectivamente a nombre del cliente o puede que tenga más propietarios con lo cual se deben dividir los ingresos del contrato en el número de propietarios, ya que no se pueden tomar al 100%.

Elaboración: propia.

Todos estos perfiles de clientes se evalúan mediante un modelo de *credit scoring*, ya que según la Superintendencia Financiera de Colombia, mediante Carta Circular 31 y circular externa 11 de 2002, todas las instituciones financieras del país deben implementar un Sistema de Administración del Riesgo Crediticio (SARC) por el hecho de enfrentarse a pérdidas debido a que están prestando dinero y no saben si el cliente va a decidir pagarlo o no, por lo cual las entidades realizan provisiones para determinar la cantidad de capital que le están prestando a los clientes y que quizás pueden perder. Este sistema permite identificar, medir y controlar el riesgo.

Siguiendo la circular, la medida de riesgo de crédito (pérdida por incumplimiento en el contrato original) debe hacerse con la metodología que incorpore los siguientes elementos: la probabilidad de incumplimiento del deudor respectivo y el cálculo de la pérdida en el momento que se dé incumplimiento. Esto es lo que se denomina como pérdida esperada de un crédito y la cuantificación de riesgo de crédito. Así mismo las entidades de crédito deberán provisionar desde el momento que se otorga el crédito y mantener la provisión durante el tiempo que dure el crédito y el riesgo que represente.

Adicional la Superintendencia Financiera de Colombia, exige a las entidades financieras tener en cuenta los siguientes criterios a la hora de otorgar un crédito:

## **1.2 Criterios a la hora de otorgar riesgo de crédito.**

<b>Criterios a la hora de otorgar riesgo de crédito</b>
<ul style="list-style-type: none"><li>● Capacidad pago del deudor.</li><li>● Flujo ingresos o egresos, flujo caja del deudor y del proyecto financiado o a financiar.</li><li>● La solvencia del deudor (Balance de caja del individuo para poder cumplir con la obligación que va a adquirir).</li><li>● Información de las centrales de riesgo sobre el cumplimiento actual y pasado de las obligaciones que tuvo determinado individuo.</li><li>● Número de veces que el crédito ha sido reestructurado y la naturaleza de su reestructuración.</li><li>● Conocer los posibles efectos de los riesgos financieros a los que se encuentra expuesto el balance de caja.</li></ul>

Fuente: Superintendencia Financiera de Colombia, Elaboración: propia

Otro factor importante es cuándo y cómo se va a devolver ese capital prestado, ya que en muchas ocasiones al momento de desembolsar el crédito, el cliente hace una serie de cálculos y se organiza acorde a los ingresos disponibles que tiene mensualmente para así poder pagar la cuota del crédito adquirido; sin embargo, a veces el plazo que se pacta desde un comienzo no es suficiente para el cliente, por lo que debe optar por reestructurar la obligación y esto no es una buena señal de su salud financiera y debilita su calificación en las centrales de riesgo.

Es importante aclarar, que a pesar de que las entidades financieras utilizan herramientas como un *scoring* de crédito para ayudar a determinar si se otorga o no el crédito, esto es una herramienta que no es el veredicto final sino que también es importante el análisis que se hace por parte del área de riesgos, en donde se cuenta con profesionales especializados en los diferentes productos que ofrece determinada entidad financiera y que gracias al bagaje que tienen en el tema, hacen que los tiempos de respuesta sean más eficientes para el cliente y unos menores costos para la entidad financiera.

Según Gutiérrez, M (2007) hay que tener en cuenta que esta herramienta hace una evaluación individual independientemente de lo que pueda ocurrir con el resto de la cartera de préstamos. Sin embargo, hay otros modelos tales como los VaR marginales que si tienen en cuenta la correlación que tiene la calidad crediticia de los diferentes deudores de una cartera de préstamos.

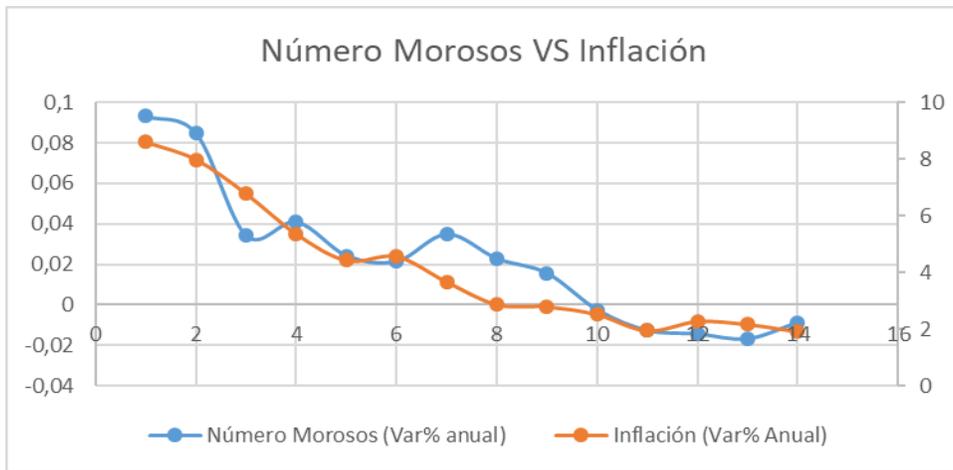
Se hace necesario recalcar que, en este análisis y como tal la herramienta, no tienen en cuenta explícitamente los factores macroeconómicos e institucionales que puedan ayudar a dar más robustez al modelo tales como: tasa de desempleo, tasa de interés, precios de materias primas, producto interno bruto (PIB) e inflación.

Así mismo, aparte de incluir variables macroeconómicas, también se debería exigir una garantía real para así tener una mayor seguridad sobre el crédito, de esta manera, si en algún momento el individuo presenta insolvencia y no puede seguir pagando la deuda adquirida con determinada entidad financiera, este tiene la operación atada a la garantía y de cierta manera sería el respaldo de la misma, es decir, un colateral.

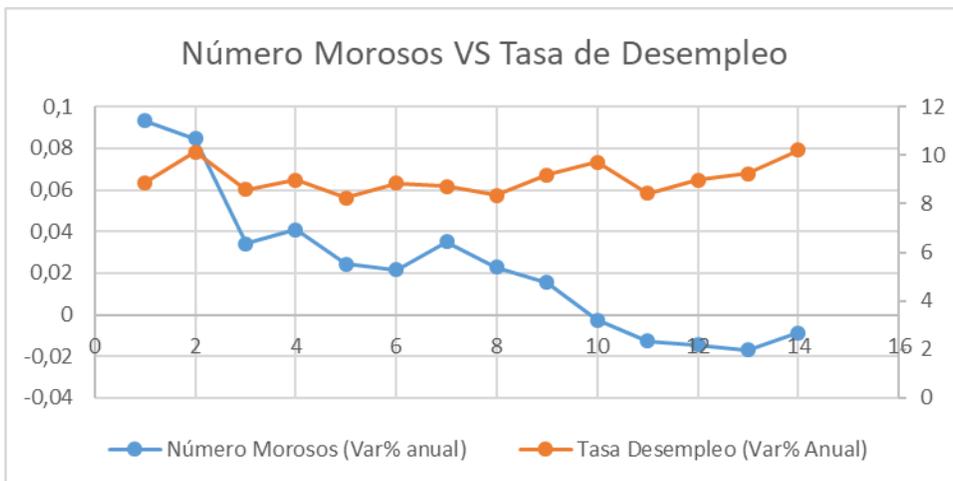
Un ejemplo de esto son los créditos de vivienda donde la garantía es la hipoteca, generalmente los clientes se comportan de una manera formidable con este tipo de créditos ya que para los individuos tiene más peso pagar mensualmente las cuotas de la vivienda, que a futuro será parte de su patrimonio, y no permitir que el banco disponga de la garantía.

Por otro lado, se compara el número de individuos morosos de cada uno de los trimestres desde el año 2012q1 hasta el año 2016q2 con respecto a las variables macroeconómicas: inflación, tasa de desempleo y tasa de interés, teniendo en cuenta una variación porcentual:

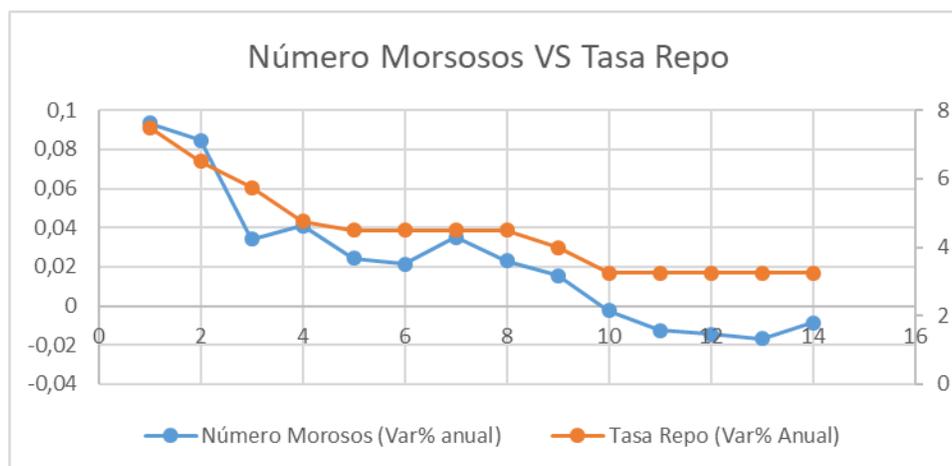
**A. Número Morosos VS Inflación**



**B. Número Morosos VS Tasa de Desempleo**



### C. Número Morosos VS Tasa Repo



Con las anteriores gráficas se evidencia que con respecto a la tasa de inflación cuando esta aumenta, los individuos pierden poder adquisitivo y su coste de vida es mucho más elevado por cual se vuelve más difícil poder pagar un crédito y de la misma manera aumenta el número de individuos morosos; con respecto a la tasa repo cuando aumenta ocasiona que los créditos se vuelvan mucho más costosos y por ende los individuos no solicitan créditos. Así mismo esta medida del Banco de la República de subir tasas se conoce como política contractiva, para que la economía se contraiga, la gente consuma menos y la demanda interna sea menor.

Por otro lado, si la tasa repo disminuye los créditos se vuelven más baratos y los individuos deciden endeudarse; por último, al aumentar la tasa de desempleo esto hace que se aumente el número de morosos debido a que ya no tienen la fuente de recursos económicos para poder seguir pagando el crédito adquirido en determinado momento del tiempo.

Esto evidencia que las variables macroeconómicas tenidas en cuenta en el modelo, ayudan a predecir de alguna manera la probabilidad de impago de los individuos y tienen una correlación con el número de individuos morosos, por lo cual deberían ser tenidas en cuenta para la evaluación de riesgo de crédito.

Por lo anteriormente nombrado, de acuerdo al modelo empleado por la IFBEC y teniendo la información disponible de una entidad bancaria<sup>2</sup>, la principal pregunta que intenta responder este trabajo es: ¿Es posible mejorar los modelos de *scoring* mediante variables macroeconómicas que ayuden a predecir la probabilidad de incumplimiento teniendo en cuenta los distintos perfiles de clientes?

## 2. MARCO TEÓRICO

Los modelos de *credit scoring* comenzaron en los años 70's pero se generalizaron ya en los 90's, según Gutiérrez, M (2007), gracias a contribuciones en sistemas computacionales y estadísticos que son de relevancia para desarrollar esta herramienta. En la actualidad, la mayoría de entidades financieras utilizan un *credit scoring* para otorgar los créditos, además una de las ventajas de esta herramienta son sus bajos costos y la eficacia al momento de dar una respuesta a la solicitud de crédito por parte de personas naturales.

Siguiendo a Gutiérrez, M (2007) esta herramienta sirve para estimar las probabilidades de default<sup>3</sup> y también catalogar a los deudores o solicitantes de alguno de los productos de crédito en función del riesgo que corre la entidad ante el incumplimiento de estos individuos. Adicional estos modelos están basados en el *data mining* (minería de datos), que es útil al momento de estudiar patrones de comportamiento acorde a los diferentes perfiles de clientes que se tienen en un sistema financiero.

Así mismo, para determinar si debe o no otorgarse determinado riesgo de crédito se utiliza el concepto humano (analista), ya que la respuesta al crédito se da de una manera más eficaz y este debe emplear las políticas que determinada entidad financiera tiene establecidas para los diferentes productos que ofrece. Lo anterior es del autor

---

<sup>2</sup> Por cuestiones de seguridad y protección de los datos no se puede mencionar el nombre de la Entidad Financiera.

<sup>3</sup> Default o suspensión de pagos es un término habitualmente utilizado en finanzas para hacer referencia a una situación en la que el prestatario no hace frente a las obligaciones legales que tiene con sus acreedores en la forma establecida en el contrato de reconocimiento de la deuda. El default puede producirse con distintos tipos de deuda, ya sean bonos, hipotecas o préstamos.

Siguiendo al autor, al momento de evaluar un crédito se debe tener en cuenta el historial crediticio del cliente lo cual se puede obtener mediante las centrales de riesgo que suministran esta información, o también se evalúa si el individuo al que se le está haciendo el estudio de crédito ya fue cliente del banco y como se comportó en el pago de las obligaciones durante el período pactado; esto sirve para determinar si el cliente ha tenido una buena o mala salud crediticia. Según la Superintendencia Financiera de Colombia las políticas de administración del Riesgo Crediticio deben contemplar como mínimo con los siguientes aspectos: estructura organizacional, límites de exposición crediticia y de pérdida tolerada, otorgamiento de crédito, garantías, seguimiento y control, constitución de provisiones, capital económico y recuperación de cartera.

Esto es lo que se conoce como análisis comportamental, según Schreiner (2002) el modelo de *Scoring* utiliza la misma lógica que el analista de crédito, pues se basa en experiencias y seguimientos de créditos otorgados en el pasado, mediante un análisis de las características de los nuevos solicitantes, con el fin de calificar o descalificar los perfiles. Adicionalmente, según Ochoa J., Galeano W. y Agudelo L. (2010) el *scoring* estadístico está basado en historiales de pago, toma comportamientos pasados para pronosticar comportamientos futuros de los créditos.

El único problema que tiene este método es que sólo se sabe el comportamiento si es un cliente antiguo del banco, pero sino lo es, entonces se procede a observar el comportamiento con el resto de entidades tanto del sector financiero y solidario como del sector real.

Algunos autores han dado su perspectiva acerca de que trata un modelo de *scoring* y los modelos que han desarrollado, a continuación se presentan algunos:

Sun, J., Fujita, P., Chen y Li, H. (2017), hablan sobre modelos de predicción financiera dinámica distress, sus siglas en inglés DFDP, son importantes para la gestión del riesgo financiero corporativo; el estudio propone dos nuevos enfoques llamados DEVE-AT y ADAVSM-TW, basado en la ponderación de tiempo combinado con Adaboost SVM ensemble, con el fin de enriquecer el sistema de metodología de DFDP.

DEBE-AT hace DFDP decisión combinando ambos resultados de Adaboost-SVM y Timeboost-SVM. El Adaboost-SVM tradicional está basado puramente en un error de decisión experta, que considera muestras clasificadas incorrectamente más importantes que las muestras que se encuentran correctamente clasificadas.

Harris, T. (2015), el paper investiga la práctica del credit scoring e introduce el uso de clustered support vector machine (CSVM) para el desarrollo de la scorecard de crédito. Este algoritmo diseñado recientemente aborda algunas de las limitaciones señaladas en la literatura que está asociada con los métodos basados en support vector machine (SVM) no lineal tradicional para la clasificación. Este paper investiga la idoneidad para el credit scoring de un reciente desarrollo de SVM basado en un algoritmo que ha sido propuesto por Gu y Han (2013).

Su máquina de vector de soporte agrupada ha demostrado ofrecer un rendimiento comparable a los enfoques basados en kernel mientras que permanece más económico en términos de tiempo computacional. Este paper introduce el uso del CSVM para el credit scoring. El CSVM representa una posible solución de las limitaciones de la actual cosecha de clasificadores utilizados en la práctica. Los resultados de este paper sugieren que el CSVM comparado con SVM no lineal basado en técnicas en términos de AUC, mientras que los supera en términos de tiempo de entrenamiento. Esto se debe a que CSVM divide los datos en varios clústeres antes de entrenar un clasificador de máquina de vector de soporte lineal para cada clúster. Esto permite la ponderación local del clasificador para una clasificación rápida.

Carling, K., Jacobson, T., Lindé, J. y Roszbach, K. (2007), estiman un modelo de duración para explicar el tiempo de supervivencia al incumplimiento de los prestatarios en la cartera de préstamos comerciales de un importante banco sueco durante el período 1994-2000. El modelo toma ambas características específicas de la firma, tales como ratios contables y comportamiento de pago, información relacionada con préstamos y teniendo en cuenta las condiciones macroeconómicas vigentes. La brecha del producto, la curva de rendimientos y las expectativas de los consumidores del futuro desarrollo económico tiene un significativo poder explicativo para el riesgo de incumplimiento de las empresas. Ellos comparan su

modelo con un uso frecuentemente del riesgo de incumplimiento de las empresas que está condicionado solamente en información específica de la empresa.

La comparación muestra que si bien este último modelo puede hacer una clasificación razonablemente precisa de las empresas en función del riesgo de incumplimiento, este modelo, teniendo en cuenta las condiciones macroeconómicas, también es capaz de explicar el nivel absoluto de riesgo. En esencia cuatro enfoques han evolucionado: “modelos estructurales”, que sigue el enfoque de Merton (1974), modelos de factor de riesgo econométrico, que aplican un método ascendente y calculan las tasas de incumplimiento, ya sea en el nivel de la empresa individual o en el nivel de la subcartera, modelos actuariales descendentes, que no hacen suposiciones respecto a la causalidad, y métodos no paramétricos.

Un número de estudios empíricos han considerado el impacto de las condiciones macroeconómicas de no pago o el riesgo de crédito. Wilson (1997a,b) describe algunos principios generales detrás del modelo de riesgo de portafolio de McKinsey's. Nickell et al. (2000) usan indicadores dummy para integrar cambios en la postura macroeconómica en un modelo de transiciones de calificación de bonos corporativos. Pesaran et al. (2006) emplean un modelo macroeconómico global y lo vinculan con los retornos específicos de la empresa en un modelo de tipo Merton. Ellos hacen contingencia del riesgo de crédito de la empresa en variables macroeconómicas tales como cambios en índices de equidad, tasas de interés y rendimiento.

Este trabajo contribuye al campo mediante la estimación de un modelo de duración para el tiempo de supervivencia de las líneas de crédito de las empresas en un banco. Este modelo puede utilizarse no sólo para analizar los incumplimientos de las empresas, sino también como insumo en un modelo de riesgo de crédito de la cartera. El modelo incluye variables explicativas macroeconómicas y un set de las variables más comunes de una empresa específica, tales como ratios contables y observaciones de pago.

Por otro lado Altman, E. (1968) el propósito del paper es buscar la calidad del ratio de análisis como técnica analítica. Se ha sugerido que el análisis de proporción tradicional ya no es una técnica analítica importante en el ambiente académico debido a la manera relativamente poco sofisticada en la que se ha presentado.

La teoría es que los ratios, si se analizan dentro de un marco multivariado, tendrán una mayor significancia estadística que la técnica común de comparaciones de ratio secuencial.

Es importante tener en cuenta que una firma con una pobre rentabilidad y/o solvencia puede considerarse como una posible quiebra.

La técnica estadística utilizada escogida fue el análisis discriminante múltiple, sus siglas en inglés (MDA). Es una técnica estadística utilizada para clasificar una observación en uno de varios grupos a priori dependiendo de las características individuales de las observaciones. Se utiliza principalmente para clasificar y/o hacer predicciones en problemas donde la variable dependiente aparece en forma cualitativa, por ejemplo: masculino o femenino, bancarota o no bancarota. Por lo tanto, el primer paso es establecer grupos explícitos de clasificación. El número original de grupos pueden ser dos o más.

Después de que se establezcan los grupos, se recolectan los datos para los objetos en los grupos; los intentos de MDA para derivar una combinación lineal de estas características que "mejor" discrimine entre los grupos. Si un objeto particular, por ejemplo una corporación, tiene características (razones financieras) que pueden cuantificarse para todas las compañías en el análisis, el MDA determina un conjunto de coeficientes discriminantes. Cuando estos coeficientes se aplican al ratio actual, existe una base para la clasificación en uno de los grupos mutuamente excluyentes.

La técnica MDA tiene la ventaja de considerar un perfil completo de características comunes a las empresas relevantes, así como la interacción de estas propiedades. Un estudio univariado, por otro lado, sólo puede considerar las mediciones utilizadas para las asignaciones de grupo, una a la vez.

La principal ventaja del MDA en el tratamiento de problemas de clasificación es el potencial de analizar todo el perfil variable del objeto simultáneamente en lugar de examinar secuencialmente sus características individuales.

Ohlson, J. (1980), presenta en el paper algunos resultados empíricos de un estudio que predice el fracaso corporativo como lo demuestra el evento de quiebra. La metodología empleada en la estimación de máxima verosimilitud o también llamada modelo logit condicional.

Una característica de este estudio es que los datos de las empresas fallidas no se derivaron del manual de Moody's. Los datos se obtuvieron en su lugar a partir de los estados financieros de 10 k según se informó en ese momento. Este procedimiento tiene una ventaja importante: los informes indican en qué momento se publicaron al público, y por lo tanto se puede comprobar si la empresa entró en quiebra antes o después de la fecha de liberación.

Los principales hallazgos del presente estudio fueron: primero fue posible identificar cuatro (4) factores básicos como estadísticamente significativos en afectar la probabilidad de fracaso (dentro de un año). Estos son: (i) el tamaño de la compañía; (ii) la(s) medida(s) de la estructura financiera; (iii) la(s) medida(s) del rendimiento; (iv) la(s) medida(s) de la liquidez corriente (la evidencia demuestra que este factor no es claramente comparado con los casos (i)-(iii)). Segundo emplea predictores derivados de declaraciones que fueron liberadas después de la fecha de la quiebra, entonces la evidencia indica que será más fácil "predecir" la bancarrota.

Sin embargo, incluso si se admite este factor, para la muestra de empresas utilizada en este estudio, la tasa de error de predicción es mayor en comparación con la tasa registrada en el estudio original de Altman (1968), así como en la mayoría de los otros estudios utilizando datos extraídos desde períodos anteriores a 1970. Más importante aún, la tasa de error de predicción es también mayor que la reportada en Altman et al. (1977).

Una de las desventajas del modelo es que no utiliza transacciones de mercado (precio) datos de las firmas. Se puede esperar, por supuesto, que el poder predictivo del modelo podría ser mejorado mediante la incorporación de datos.

En primer lugar, el poder predictivo de cualquier modelo depende de cuándo se asume que la información (informe financiero) está disponible. En segundo lugar, los poderes predictivos de transformaciones lineales de un vector de relaciones parecen ser robustos a través de procedimientos de estimación de gran muestra.

Li, Z., Crook, J. y Andreeva, G. (2017), acreedores tales como lo bancos frecuentemente usan sistemas expertos para soportar sus decisiones cuando emiten créditos y evalúan créditos ha sido un área importante de aplicación de las técnicas de aprendizaje automático durante décadas. En la práctica, a menudo se requiere que los bancos proporcionen la razón detrás de sus decisiones, además de poder predecir el desempeño de las empresas al evaluar a los solicitantes corporativos de préstamos. Una solución es usar el análisis envolvente de datos, sus siglas en inglés (DEA) para evaluar múltiples unidades de toma de decisiones (DMUs o compañías) que se clasifican según las mejores prácticas en su sector industrial. Se emplea un algoritmo de programación lineal para calcular la eficiencia corporativa como una medida para distinguir a las empresas sanas de las que están en dificultades financieras.

Hand y Henley (1997), especifican que un modelo de *credit-scoring* (MCS) tradicionalmente se ha definido como algún método, generalmente estadístico, utilizado para clasificar solicitudes de crédito entre las categorías de “buenas” o “malas”, según la expectativa de pago que se les pueda retribuir.

Según Boyes, Hoffman y Low (1989) y Greene (1992), para los modelos de deuda retail utilizan variables socioeconómicas como edad, cantidad de personas a cargo, estado civil, nivel educativo, antigüedad laboral y residencial, si es propietario de la vivienda, gastos e ingresos mensuales promedio, profesión, si tiene tarjeta crédito cuenta corriente o de ahorros, número de consultas en los credit bureaus y como está calificado en ellos.

Por otro lado, Aguas y Castillo (2002), proponen un modelo en el cual se estiman las pérdidas potenciales en las que puede incurrir cualquier entidad financiera al momento de otorgar y realizar el respectivo seguimiento al crédito por medio de una regresión logística binaria.

Es importante recalcar que en este trabajo se realizan el cálculo de varios tipos de default; si la economía se llegase a encontrar en recesión o no, es cuando se incluiría en el análisis la hipótesis de la inestabilidad mencionada por Minsky (1992).

Un punto relevante a tener en cuenta es que el sistema financiero es inestable, lo cual Minsky (1992) explica con su hipótesis de la inestabilidad financiera; basándose en la Teoría General de Keynes donde enfatiza que la momentánea estabilidad financiera, es la que permite que las instituciones bancarias basadas en cierta confianza promovida por el crecimiento económico, otorguen un mayor número de créditos con un mayor riesgo, lo que a su vez ocasiona una mayor liquidez en la economía e inflación de los precios, y esto aduce a la especulación financiera, ya que finalmente cuando estalla la burbuja se encuentra con una cartera vencida alta y una crisis de grandes proporciones.

Así mismo, Medina y Paniagua (2008) realizaron un estudio en una cooperativa de servicios financieros (Cooperativa de Ahorro y Crédito), donde por medio del sistema de inferencia difuso querían evaluar la solvencia de los clientes que solicitaban un crédito. Se utilizó la base de datos de la cooperativa y se tuvieron en cuenta variables tales como plazo, aportes sociales, garantías, historial crediticio y monto del crédito otorgado con el fin de incluirlas en el desarrollo del modelo y así poder determinar la relación entre las variables de entrada y salida con ayuda de criterio experto.

Concluyen diciendo que cuando se hace una gestión de riesgos se pueden emplear varios mecanismos que ayuden a disminuir la probabilidad de pérdida del capital. Adicionalmente, dentro de los procesos preventivos se tienen en cuenta el análisis previo del cliente, el monitoreo y control de riesgos, los sistemas de análisis financiero, informes comerciales, control de la deuda del cliente y los sistemas de control de límites de pérdidas. Con el sistema difuso lo que se busca es ayudar a los evaluadores a tomar decisiones más acordes al monto y el plazo para cada cliente y así mismo observar las debilidades y fortalezas.

Por otro lado, Pérez y Fernández (2007) utilizan un modelo de redes neuronales para aplicarlos a una base de datos de una cartera comercial, con lo que tuvieron que realizar un análisis de las variables e investigar la correlación que había entre ellas.

Lo que hacen es cruzar variables de cada cliente, del crédito y del respectivo comportamiento contra la variable de default (fallido o no fallido). Esta variable predice un procedimiento de clasificación además de determinar las ponderaciones necesarias y establecer la probabilidad de fallido.

El estudio, concluye que utilizaron primero un perceptrón simple que se encargaba de clasificar los buenos como buenos y los malos como malos con un porcentaje de aciertos del 43.22%, luego utilizaron un perceptrón de dos capas y para este caso el porcentaje de aciertos fue del 94,19%, ya por último trabajaron con una red neuronal probabilística y esta última resulto ser la más adecuada para resolver el problema de clasificación por el hecho de que obtuvo un porcentaje de aciertos más alto.

Adicionalmente, Gutiérrez (2007) se basa en la construcción de un modelo logit-probit teniendo en cuenta la información de la Central de Deudores del Sistema Financiero (CENDEU), que fue suministrada por el Banco Central Argentino y que tiene información sobre todas las deudas y deudores del sistema financiero. Las variables empleadas son socioeconómicas o datos básicos del emprendimiento productivo o de grandes empresas, esto es para el caso de créditos retail.

Por último, Mermelstein, D (2006), hace referencia sobre las faltas de default en carteras hipotecarias, haciendo énfasis en los canales de transmisión que existen entre la evolución del contexto macroeconómico e institucional y el factor microeconómico de no pago.

Hacen énfasis que tanto variables macroeconómicas como factores institucionales son candidatos naturales para entrar como argumento en los esquemas de decisión microeconómicos que pone en práctica el deudor a la hora de incurrir en el no pago de su hipoteca.

Para concluir, sustentan que el incumplimiento de pagos surge como consecuencia de la insolvencia del deudor, ocurridas por shocks macroeconómicos o microeconómicos.

Lo que se buscó es poder determinar, después de un análisis exhaustivo, que variables permiten tener un modelo más robusto a la hora de determinar si se debe otorgar o no un

crédito a los diferentes perfiles de clientes. Para ello es muy importante la significancia de cada una de las variables y la consistencia de los betas.

Es importante resaltar, que los clientes que solicitan un crédito deben tener la capacidad de pago suficiente para así cumplir con la obligación adquirida y del mismo modo que demuestren una estabilidad laboral; aunque teniendo en cuenta que la economía es tan cambiante puede suceder que algunos de los sectores económicos entren en crisis y por esto las empresas toman decisiones de despido, por lo que el cliente se queda sin solvencia para atender la obligación y allí es cuando se debe hacer una reestructuración del plazo del crédito y la forma de pago con el fin de ayudar al cliente y así mismo que el banco no pierda el dinero que prestó o por lo menos que recupere una parte del mismo.

Otro punto importante es que los bancos busquen mecanismos de costos más altos para los clientes, con el fin de fomentar el pago de los créditos otorgados y no entrar en default, ya que no es conveniente para el banco por el hecho de que tiene que provisionar ese crédito que da por perdido y que afecta sus estados financieros y tampoco para el cliente porque se va a ver afectado su historial crediticio.

El objetivo del presente estudio es medir el riesgo de crédito para poder prever las pérdidas potenciales en las que podría incurrir una institución al momento de otorgar créditos, es por ello que las diferentes metodologías para medir el riesgo de crédito buscan calcular la probabilidad de incumplimiento de un deudor frente a determinado acreedor que para este caso es el sistema bancario, una vez se asume la deuda.

Algunos autores tuvieron su primera aproximación a los *Credit Scoring* como “*métodos estadísticos utilizados para clasificar los solicitantes de crédito, o incluso a quienes ya son clientes de la entidad evaluadora, entre las cosas de riesgos ‘bueno’ y ‘malo’*” (Hand y Henley (1997)). Esto se enfoca al sentido en que la entidad bancaria tiene información interna y externa sobre el cliente que le puede ayudar a determinar si es bueno o no otorgarle riesgo de crédito.

Teniendo en cuenta a Gordy (2000), busca comparar los diferentes modelos de cartera de riesgo crediticio empleando un modelo probit para estimar la probabilidad de que un deudor pueda caer en impago en las diferentes carteras.

Además Srinivasan y Kim (1987) analizaron el problema de una empresa comercial con el objetivo de determinar el límite de crédito óptimo a otorgar a cada cliente; lo que hacen es evaluar un problema dinámico teniendo en cuenta el riesgo que tiene cada cliente relacionado con los beneficios que se obtendrían de este.

Para Schreiner, M (2002), la idea central es que el *scoring* de crédito hace referencia a características y desempeño relacionados con préstamos del pasado, para de esta manera pronosticar desempeños de préstamos en el futuro. Adicional, este autor menciona que los beneficios del *scoring* son: la herramienta como tal permite calificar a cada uno de los clientes, es más eficiente en tiempo por lo que reduce costos, cada crédito tiene una puntuación por lo que permite tener un seguimiento estadístico a cada portafolio y que la evaluación de crédito que realiza el score es objetiva y consistente.

Sin embargo, el score presenta desventajas o limitaciones tales como: son procesos que pueden tardar bastante tiempo, son sistemas que con el tiempo se deterioran, es indispensable contar con un área especial y son modelos de predictibilidad limitada.

Existe también la categoría de “*scoring* de comportamiento” que se basa en los clientes existentes que tiene el banco y su historial de pago (Thomas, 2000). Estos modelos de comportamiento ayudan a los encargados de decisiones a determinar estrategias críticas sobre los límites de crédito de los clientes, las dificultades de pago, quiebra, etc (Hsieh, 2004).

Con las contribuciones que han hecho diferentes autores a lo largo del tiempo se ha podido analizar de una mejor manera el *credit scoring*, por lo que se ha vuelto un proceso más eficaz medido en tiempos de respuesta al cliente al momento de solicitar el crédito y significa una reducción de costos para la entidad que otorga el crédito, ya que posee un grupo de profesionales especializados que tienen la experiencia en los diferentes productos que se ofrecen y conocen las políticas establecidas para otorgar los mismos, por lo cual al momento de evaluar el crédito se hace de una manera más integral.

Complementando la información anterior, es importante tener en cuenta los Acuerdos de Basilea, conformados por representantes de los Bancos Centrales y de las autoridades regulatorias del Grupo de los Diez, más Luxemburgo y España, relacionados con la supervisión y las recomendaciones acerca de la regulación bancaria. Lo anterior fue emitido por el Comité de Basilea de Supervisión Bancaria. Estas recomendaciones se encuentran contenidas en los acuerdos de Basilea I, Basilea II y Basilea III.

Cabe resaltar que este comité no tiene la potestad para imponer recomendaciones, pero los países pueden decidir autónomamente implementar las políticas del comité. Lo anterior hace referencia, a que las recomendaciones son aplicadas mediante leyes y regulaciones nacionales o a nivel comunitario en la Unión Europea, antes que como resultado de una recomendación internacional del Comité.

El acuerdo de Basilea I fue publicado en 1988, se trataba de un conjunto de recomendaciones para establecer un capital mínimo, el cual debería tener cada entidad bancaria en función de los riesgos que afrontaba. El acuerdo estableció una definición de capital regulatorio compuesto por elementos que se agrupan en dos categorías y debían cumplir los requisitos de permanencia, capacidad de absorción de pérdidas y protección ante quiebra. El capital mínimo que debía tener la entidad bancaria según el acuerdo debía ser del 8% del total de los activos de riesgo y debía ser suficiente para hacer frente a los riesgos de mercado, riesgo y tipo de cambio.

Este acuerdo tenía una deficiencia y era que consideraba que todos los créditos tenían la misma posibilidad de incumplirse; esta probabilidad de incumplimiento no puede ser constante ya que los prestatarios a los cuales las entidades financieras están prestando el dinero, poseen perfiles diferentes y por lo tanto deben evaluarse acorde al perfil y no de una única manera.

El acuerdo de Basilea II fue publicado en el 2004 y tenía en cuenta la calidad crediticia de los prestatarios, agregando requisitos de capital debido al riesgo operacional, este es el primer pilar de dicho acuerdo.

El segundo pilar hace referencia al proceso de supervisión de la gestión de los fondos propios con el fin de poder hacer frente a una crisis económica, con lo que se puede obligar a la entidades a incrementarlos en función de los resultados y el tercer pilar hace referencia a la disciplina del mercado, es decir, a publicaciones periódicas de información que esté relacionada con la exposición a los diferentes riesgos y la suficiencia de sus fondos propios.

Y por último, está el acuerdo de Basilea III que engloba propuestas sobre reforma de la regulación bancaria, las cuales fueron publicadas desde el 16 de Diciembre del 2010. Incluye una serie de iniciativas promovidas por el Foro de Estabilidad Financiera y el G-20, con el fin de lograr fortalecer el sistema financiero teniendo en cuenta la crisis de las hipotecas subprime; también se enfoca en el riesgo de pánico bancario, exigiendo distintos niveles de capital para los diversos tipos de depósitos que ofrecen las entidades financieras y otros préstamos.

### **3. METODOLOGÍA**

La hipótesis del trabajo es que la inclusión de variables macroeconómicas como la tasa de desempleo, la tasa de interés o la inflación van a permitir no solo establecer de mejor manera la demanda y la cantidad de préstamos solicitados por los clientes de cierta entidad bancaria sino también la forma como se solicitan los mismos.

Los modelos scoring tradicionales utilizan un *scorecard*<sup>4</sup> que consiste en clasificar una población determinado en dos o más grupos empleando determinadas técnicas estadísticas.

Por lo anterior, para estimar el *credit scoring* del presente trabajo se utilizó el modelo lineal generalizado, que asume observaciones independientes o al menos no correlacionadas; este modelo se estima suponiendo que existe un componente aleatorio o término de error único. El modelo lineal generalizado permite incluir un margen de maniobrabilidad entre el predictor lineal y la media de la selección del link o función, y de la selección de las variables relevantes.

---

<sup>4</sup> Scorecard es una tabla de puntajes de calificación para cada variable que resultó relevante en el análisis. Estos puntajes son dados por un modelo estadístico y representa la probabilidad de buen pago del crédito a otorgar.

Para este trabajo, el link o la función de enlace es la función *logit*, dicha variable explicativa es dicótoma: es decir el individuo entra en impago o no. Definirla de esta forma, permite conocer cómo se relaciona la transformación de la variable respuesta con las variables explicativas en términos de probabilidades.

Modelo actual:

$$y_{it} = \alpha_i + (x_{1i,t}\beta_1 + \dots + x_{48i,t}\beta_{48}) + \varepsilon_{it} \quad (1)$$

donde  $X =$  (constantes, edad, sexo, patrimonio, ingresos, situación laboral, vivienda habitual,..., anexos tabla 1)

Con  $i = 1, \dots, N$ ;  $t = 1, \dots, T$ .

Sea  $x_{it}$  la variable  $i$  en el tiempo  $t$ ,  $\mathbf{y}^T = (y_1, \dots, y_n)$  realización de un vector aleatorio y  $\mathbf{Y}^T = (y_1, \dots, y_n)$  cuyas componentes son estadísticamente independientes y distribuidas con medias  $\boldsymbol{\mu}^T = (\mu_1, \dots, \mu_n)$ :

1. La componente aleatoria asume independencia de las componentes  $\mathbf{Y}^T = (y_1, \dots, y_n)$  y pertenencia a distribuciones de la familia exponencial, con  $E[\mathbf{Y}] = \boldsymbol{\mu}$ .
2. La componente sistemática del modelo consiste en la especificación de un vector  $\boldsymbol{\eta}$ , el predictor lineal a partir de un número reducido de parámetros a estimar y regresores; parámetros  $\boldsymbol{\beta}^T = (\beta_1, \dots, \beta_p)$  y regresores  $\mathbf{X} = (x_1, \dots, x_p)$ . Esta especificación responde, en notación matricial a  $\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta}$  donde  $\boldsymbol{\eta}$  es  $n \times 1$ ,  $\mathbf{X}$  es  $n \times p$  y  $\boldsymbol{\beta}$  es  $p \times 1$ .
3. El vector  $\boldsymbol{\mu}$  está funcionalmente relacionado con el predictor lineal  $\boldsymbol{\eta}$  a través de la denominada función link, es notada como  $\mathbf{g}(\cdot)$ , de manera  $\boldsymbol{\eta} = \mathbf{g}(\boldsymbol{\mu})$ ,  $\boldsymbol{\mu}$  es  $n \times 1$ .

Sea  $y$  una variable aleatoria y su función de probabilidad, si es discreta o su función de densidad de probabilidad si es continua; en cualquier caso depende de un único parámetro  $\theta$ .

La distribución pertenece a la familia exponencial si puede escribirse de la siguiente manera:

$$f_Y(y, \theta, \phi) = \exp\left(\frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi)\right) \quad (2)$$

Donde  $a(\cdot)$ ,  $b(\cdot)$  y  $c(\cdot)$  son funciones específicas.

- Si  $\phi$  es conocido entonces el modelo pertenece a la familia exponencial y se denomina al único parámetro  $\theta$ : parámetro canónico.
- Si  $\phi$  no es conocido, el modelo puede pertenecer o no a la familia exponencial (con 2 parámetros).

Sea  $Y$  de distribución binomial  $B = (m, \pi)$  con parámetros  $m$  y  $\pi$ , donde la observación es  $y$ , y

$$E[Y] = \mu = m\pi$$

$$\begin{aligned} f_y(y, \theta, \phi) &= \binom{m}{y} \pi^y (1-\pi)^{m-y} \quad (3) \\ &= \exp(y \log(\pi) - (y-m) \log(1-\pi) + \log \binom{m}{y}) \\ &= \exp\left(y \log\left(\frac{\pi}{1-\pi}\right) - m \log\left(\frac{1}{1-\pi}\right) + \log \binom{m}{y}\right) \\ &= \exp\left(y\theta - m \log(1 + e^\theta) + \log \binom{m}{y}\right) \quad (4) \end{aligned}$$

$$\text{donde } a(\phi) = 1, b(\theta) = m \log(1 + e^\theta), \theta = \text{logit}(\pi) = \log\left(\frac{\pi}{1-\pi}\right),$$

$$c(y, \phi) = \log \binom{m}{y}, \mu(\theta) = m\pi = m \frac{e^\theta}{1 + e^\theta}. \text{ Por lo tanto,}$$

$$\ell(\theta, \phi, y) = \log f_Y(y, \theta, \phi) = \frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi) = y\theta - m \log(1 + e^\theta) + \log \binom{m}{y} \quad (5)$$

En la descripción de las componentes de un modelo lineal generalizado, el vector  $E[Y] = \mu$ , valor esperado de los datos  $y$ , está funcionalmente relacionado con el predictor lineal  $\eta$ , a través de la denominada función de link, notada como  $g(\cdot)$ , de manera  $\eta = g(\mu)$ , donde  $\eta$  y  $\mu$  son de tamaño  $n \times 1$ . Para en este caso en especial se tomará la distribución binomial  $B(m, \pi)$ , la media debe tomar valores entre 0 y 1,  $0 \leq E[Y_i] = \mu_i \leq 1$  y como el predictor lineal puede

tomar cualquier valor real, la identidad otra vez no parece razonable, se requiere mapear el predictor lineal en el intervalo 0,1.

Adicional se considera la función de link logit:

$$\boldsymbol{\eta}^T = \boldsymbol{g}^T(\boldsymbol{\mu}) = \left( \log\left(\frac{\mu_1}{1-\mu_1}\right), \dots, \log\left(\frac{\mu_n}{1-\mu_n}\right) \right) \quad (6)$$

Las funciones de link de la familia de potencias son importantes para observaciones con media positiva y se definen para cada observación como  $\eta = g(\mu) = \frac{\mu^\lambda - 1}{\lambda}$  (7)

El proceso de estimación de los parámetros en los modelos estadísticos, permite reemplazar los valores observados  $\boldsymbol{y}^T = (y_1, \dots, y_n)$  por los valores ajustados (fitted values).

El estimador usado es un estimador máximo verosímil que hace uso de Mínimos Cuadrados Generalizados.

Se debe tener en cuenta que no se hace necesario verificar normalidad ya que se está utilizando una Binomial, adicional tampoco se debe verificar homocedasticidad ya que se está trabajando métodos numéricos y por último no se verifica endogeneidad debido a que Efectos Fijos es robusto.

El modelo tiene la siguiente especificación:

Sea

$$y_i = \beta_0 + \beta_1 x_{1i} + \dots + \beta_k x_{ki} + \varepsilon_i \quad (8)$$

$$i = 1, \dots, n$$

$$\text{Si } \mathbf{Y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$$

$$\mathbf{X} = \begin{bmatrix} 1 & x_{11} & \dots & x_{k1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1n} & \dots & x_{kn} \end{bmatrix}$$

$$\boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}; \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}$$

$$Y_{it} = \alpha_i + \mathbf{X}_{it}\boldsymbol{\beta}_i + (x_{49i,t}\beta_{49} + x_{50i,t}\beta_{50} + x_{51i,t}\beta_{51} + x_{52i,t}\beta_{52} + x_{53i,t}\beta_{53} + x_{54i,t}\beta_{54}) + \varepsilon_{it}$$

→ *Modelo Nuevo* (9)

Donde  $\mathbf{X}_i$  = (constantes, edad, sexo, patrimonio, ingresos, situación laboral, vivienda habitual,..., anexos tabla 1)

$x_{49}$  = tasa de desempleo

$x_{50}$  = tasa de interés

$x_{51}$  = tasa de inflación

$x_{52}$  = producto interno bruto

$x_{53}$  = tasa de cambio

$x_{54}$  = precio del petróleo

Con  $i=1,\dots,N$ ;  $t=1,\dots,T$ .

El cual modela los log odds de probabilidad de sucesos como una función de variables explicativas.

- Componente aleatorio: la distribución de  $Y$  se asume que es Binomial  $(n, \pi)$ , donde  $\pi$  es la probabilidad de suceso.
- Componente sistemático: las  $X$  son variables explicativas (pueden ser continuas, discretas, o ambas) y son lineales en los parámetros.
- Función link: link logit:

$$\eta = \text{logit}(\pi) = \log\left(\frac{\pi}{1-\pi}\right) \quad (10)$$

Más general, el link logit modela los log odds de la media, y la media para este caso es  $\pi$ .

#### 4. DATOS Y DESCRIPCIÓN DE VARIABLES

Para la estimación del modelo lineal se utilizarán los datos del scoring de crédito de una Entidad Financiera en Colombia, la muestra se compone de 334908 registros y 6 variables.

Para los fines de esta investigación y a partir de las ecuaciones mencionadas, se utilizaron las siguientes variables:

**Variables cualitativas del individuo:** estado civil, sexo, tipo de vivienda, domiciliación de nómina, profesión, nacionalidad, situación laboral, tipo de contrato, actividad que desempeña y ciudad en la que reside.

**Variables cuantitativas del individuo:** estrato, edad, ingresos fijos, ingresos variables, patrimonio declarado, número de personas a cargo, información crediticia obtenida por las centrales de riesgo, antigüedad en el empleo y antigüedad como cliente.

#### 5. DESCRIPCIÓN DE LA MUESTRA

El proceso de estimación se realizó teniendo en cuenta los individuos para cada momento del tiempo (t) con el fin de poder evaluar como se comportaba determinado individuo en cada t al incluir las variables macroeconómicas. En la tabla N° 5.1 se muestran las variables macroeconómicas empleadas y sus respectivas definiciones.

## 5.1 Variables empleadas en el modelo.

VARIABLES EMPLEADAS EN EL MODELO		
VARIABLE	DESCRIPCIÓN	FUENTES
Tasa de desempleo	Porcentaje de la población activa desempleada: Tasa de desempleo = ( número de desempleados / población activa ) X 100	Departamento Administrativo Nacional de Estadística (DANE)
Tasa interés	Es el precio del dinero en el mercado financiero. Se emplea la tasa repo del Banco de la República.	Banco de la República
Inflación	Aumento generalizado y sostenido en el nivel de precios tanto de los bienes como de los servicios que existen en el mercado durante un período de tiempo determinado. Se emplea la inflación del consumidor.	Banco de la República
Mora_30	Esta variable hace referencia a una ventana de tiempo de 3 meses donde el cliente ha alcanzado una altura de mora de 30 días, al menos 1 vez. Con lo cual 1= entró en mora y 0= no entró en mora; una vez se haya marcado con un 1, el cliente seguirá marcado.	Bases de una Institución Financiera Bancaria en Colombia (IFBEC)
Mora_60	Esta variable hace referencia a una ventana de tiempo de 6 meses donde el cliente ha alcanzado una altura de mora de 60 días, al menos 1 vez. Con lo cual 1= entró en mora y 0= no entró en mora; una vez se haya marcado con un 1, el cliente seguirá marcado.	IFBEC
Mora_90	Esta variable hace referencia a una ventana de tiempo de 12 meses donde el cliente ha alcanzado una altura de mora de 90 días, al menos 1 vez. Con lo cual 1= entró en mora y 0= no entró en mora; una vez se haya marcado con un 1, el cliente seguirá marcado.	IFBEC
D_SITUA_LABORAL	Se clasifica en: fijo (contrato término indefinido), temporal, independiente (actividades como comerciantes, agricultores, más que todo clientes autónomos), otros hace referencia a pensionados, rentista de capital.	IFBEC
Tasa_mora_30	Desde el desembolso del crédito hasta hoy, determina si el cliente ha tenido incumplimiento de 30	IFBEC
Tasa_mora_60	Desde el desembolso del crédito hasta hoy, determina si el cliente ha tenido incumplimiento de 60	IFBEC
Tasa_mora_90	Desde el desembolso del crédito hasta hoy, determina si el cliente ha tenido incumplimiento de 90	IFBEC
Estado_inversion	Hace referencia al estado en el cual se encuentra la operación: al día, vigente, vencido o en mora	IFBEC
EDAD	Edad del cliente, medida en años	IFBEC
Estrato	Estrato donde vive el cliente, va desde el 1 hasta el 6	IFBEC
D_PROFESION	La profesión que informa el cliente y es soportada con documentos	IFBEC

Fuente: IFBEC, Elaboración: propia

Tener en cuenta que las demás variables utilizadas en el modelo se encuentran especificadas en el anexo.

Adicional a las variables anteriormente nombradas, para comprobar también se tuvieron en cuenta otras variables macroeconómicas tales como: Producto Interno Bruto (PIB), tasa de cambio y precio del petróleo; estas variables no tienen significancia estadística dentro del modelo.

## 6. ANÁLISIS RESULTADOS

### 6.1 ESTADÍSTICAS DESCRIPTIVAS

Variable	Media	Desviación Estándar
Mora	0,9016518	0,2977853
Tasa de Desempleo	10,38333	1,053171
Tasa de Interés	4,513889	1,057839
Inflación	3,788889	1,851846
Mora_30	0,0083018	0,0907364
Mora_60	0,0059918	0,077176
Mora_90	0,0115297	0,1067576

Según los resultados obtenidos en las estadísticas descriptivas entre los años 2012q1-2016q2, con respecto a la variable dependiente (mora) se tiene que el 90,1% de la muestra entró en mora al menos una vez. Con respecto a las demás variables, se tiene que del total de la muestra en promedio la tasa de desempleo es 10,3%, se tiene que en promedio la tasa de interés es 4,5%, en promedio la inflación es 3,7%; se evidencia que en promedio el 0,83% de los individuos de la muestra ha alcanzado una altura de mora de 30 días, al menos 1 vez; así mismo se evidencia que en promedio el 0,59% de los individuos de la muestra ha alcanzado una altura de mora de 60 días, al menos 1 vez y por último se evidencia que en promedio el 1,15% de los individuos de la muestra ha alcanzado una altura de mora de 90 días, al menos una vez.

Los datos están dispersos alrededor de la tendencia central de la mora en un 29,7%, alrededor de la tendencia central de la tasa de interés y tasa de desempleo en un 1,05%, alrededor de la tendencia central de la inflación en un 1,85%, alrededor de la tendencia central de mora\_30 en un 9,07%, alrededor de la tendencia central de mora\_60 en un 7,77% y por último alrededor de la tendencia central de mora\_90 en un 10,6%.

## 6.2 RESULTADOS ECONÓMICOS

Después de correr el panel, los resultados obtenidos fueron los siguientes:

Regresión por Max Verosimilitud No. Observaciones = 334908. Grupos = 18606

Variable	Coefficiente	P-Value	Significancia
Tasa de Interés	-0,1952525*** (0.0200067)	0	99% : *** 95%: ** 90%: *
Inflación	0,4249541*** (0.0130757)	0	
Tasa de Desempleo	0,0429982*** (0.007033)	0	
GDP	-0,042107 (0.017028)	0.013	
Precio del petróleo	-0,0027676 (0.001202)	0.021	
Tasa de cambio	9.62e-11 (1.41e-11)	0	
Constante	1.531099*** (0.1886888)	0	

VARIABLES	MODELO 1	MODELO 2	MODELO 3	MODELO 4	MODELO 5	MODELO 6
vr_desembolso	-2.20e-09 (0.471)	-3.22e-09 (0.321)	-7.66e-10 (0.808)	-1.07e-09 (0.730)	-1.40e-09 (0.656)	-1.41e-09 (0.655)
plazo_mes	.0042596 (0.096)	.0048197 (0.061)	.003058 (0.243)	.0032356 (0.221)	.0032603 (0.217)	.0032554 (0.218)
edad	.0173538 (0.186)	.0162214 (0.219)	.012109 (0.369)	.011956 (0.377)	.0119247 (0.378)	.0118996 (0.379)
estrato	-.2438978 (0.030)	-.2263534 (0.050)	-.2393746 (0.038)	-.2322632 (0.045)	-.2318818 (0.045)	-.2320717 (0.045)
antiguo_clie	.0225804 (0.198)	.0229872 (0.195)	.0225958 (0.210)	.0222865 (0.216)	.0225084 (0.211)	.0225383 (0.211)
antiguo_labo	-.0123377 (0.422)	-.0103414 (0.503)	-.0070724 (0.651)	-.007103 (0.651)	-.0070718 (0.652)	-.007055 (0.653)
vr_patrimonio	2.49e-15 (0.002)	2.54e-15 (0.001)	2.44e-15 (0.002)	2.43e-15 (0.003)	2.46e-15 (0.003)	2.46e-15 (0.002)
vr_vivienda	-4.47e-10 (0.533)	-7.75e-10 (0.335)	-7.27e-10 (0.353)	-7.16e-10 (0.363)	-7.35e-10 (0.354)	-7.33e-10 (0.356)
saldo_actual	3.99e-09 (0.842)	4.13e-09 (0.853)	3.05e-09 (0.890)	1.63e-09 (0.939)	1.56e-09 (0.941)	1.58e-09 (0.941)
saldo_m_ant	1.99e-09 (0.922)	6.04e-09 (0.788)	3.31e-09 (0.882)	4.48e-09 (0.834)	5.15e-09 (0.810)	5.15e-09 (0.810)
saldo_t1	-1.36e-09 (0.699)	1.77e-09 (0.646)	4.79e-09 (0.145)	4.48e-09 (0.186)	4.85e-09 0.156	4.85e-09 (0.155)
saldo_t2	1.69e-10 (0.932)	2.33e-09 (0.407)	-1.23e-09 (0.597)	-1.05e-10 (0.960)	-6.48e-10 (0.768)	-6.11e-10 (0.786)
saldo_dic_t1	-5.26e-09 (0.039)	-1.14e-08 (0.000)	-8.21e-09 (0.006)	-7.32e-09 (0.014)	-7.91e-09 (0.010)	-7.92e-09 (0.010)
saldo_dic_t2	6.53e-09 (0.061)	3.79e-09 (0.351)	1.75e-09 (0.533)	1.93e-09 (0.511)	2.19e-09 (0.450)	2.15e-09 (0.465)
saldo_dic_t3	-3.40e-09 (0.320)	-1.78e-09 (0.553)	-2.99e-09 (0.400)	-7.44e-09 (0.101)	-6.90e-09 (0.128)	-6.84e-09 (0.136)
cat_linea	-.6967708 (0.007)	-.7264148 (0.004)	-.7629044 (0.003)	-.8012641 (0.002)	-.7963793 (0.002)	-.7958688 (0.002)
cat_linea_subpr	.0774189 (0.183)	.0779162 (0.184)	.081839 (0.168)	.0847607 (0.157)	.0846475 (0.159)	.0845759 (0.159)
cat_d_estado_ci	.1520289 (0.018)	.1469523 (0.023)	.139262 (0.033)	.1376395 (0.035)	.1378973 (0.035)	.1378963 (0.035)
cat_d_situa_labr	-.0205935 (0.861)	-.0382791 (0.747)	-.0249221 (0.838)	-0.219554 (0.858)	-.0221863 (0.856)	-.0221341 (0.856)
cat_d_estado_bi	-.2241406 (0.113)	-.208226 (0.144)	-.1641287 (0.259)	-.1483896 (0.311)	-.1513137 (0.302)	-.1513814 (0.302)
cat_d_situa_vivi	.1295918 (0.215)	.1362709 (0.210)	.1326537 (0.222)	.1308119 (0.228)	.1321741 (0.225)	.131971 (0.225)
cat_d_profesion	-0.615708 (0.033)	-.0645537 (0.027)	-.0663522 (0.026)	-.0668349 (0.025)	-.066944 (0.025)	-.0669642 (0.025)
cat_d_actividad	-0.223795 (0.239)	-.0239074 (0.212)	-.0259786 (0.178)	-.0255548 (0.186)	-.0256282 (0.185)	-.025643 (0.184)
CPI	.5397047 (0.000)	.4477206 (0.000)	1.062494 (0.000)	.7869168 (0.000)	.9580248 (0.000)	.9426575 (0.002)
Unempl		-.7926825 (0.000)	-1.210221 (0.000)	-1.342124 (0.000)	-1.3388 (0.000)	-1.339742 (0.000)
Brent			.047236 (0.000)	-0.184404 (0.416)	.0073521 (0.829)	.0083415 (0.818)
trade				7.97e-10 (0.001)	5.21e-10 (0.154)	5.03e-10 (0.239)
GDP					.41787 (0.320)	.4369211 (0.367)
Interest						.0339861 (0.937)
_cons	-7.090373 (0.000)	1.216885 (0.488)	-.5695327 (0.763)	8.460425 (0.015)	4.86369 (0.324)	4.657071 (0.404)

Teniendo en cuenta que el P-value es 0 para cada una de las variables, tales como: tasa interés, tasa de desempleo e inflación, son significativas al 99% de confianza, que es equivalente a un alfa de 1% con lo cual es mayor a todos los P-values; entonces se rechaza la hipótesis nula de que los betas sean cero.

Y (variable respuesta): El cliente entra en impago o no

Interest: Tasa de Interes REPO BANREP -0,1952525

CPI: Inflación del consumidor 0,4249541

Unempl: Tasa de desempleo 0,0429982

Pvalue todos 0

En relación con las variables estimadas, se encuentra que, tasa de interés, tasa de inflación y desempleo, son estadísticamente significativas y sus signos son relevantes económicamente. Los resultados encontrados evidencian que el aumento de la tasa de interés disminuye la probabilidad de que el cliente entre en impago en un 19,52%. Esto tiene relación que el hecho de que las tasas de interés altas disminuyen la demanda por crédito y por ende los créditos que caen en default, adicionalmente el aumento de la tasa repo del Banco de la República hace que se reduzca la inflación y por ende la probabilidad de impago. Es importante tener en cuenta como precedente que los créditos analizados ya están contratados y el banco les aplica una tasa fija, por lo que su valor no cambia en el tiempo.

Por otro lado, los efectos de un aumento en la inflación y en la tasa de desempleo, aumentan la probabilidad de que el cliente entre en impago en un 42,49% y 4,29% respectivamente. La inflación refleja la disminución del poder adquisitivo de la moneda, por lo que genera desahorro e incertidumbre sobre la inversión y el valor futuro del dinero. Es decir, cada unidad monetaria permite comprar menos bienes y servicios por lo que aumenta la probabilidad de que el cliente entre en impago, teniendo en cuenta que disminuye el poder adquisitivo del mismo. Económicamente el incremento de las tasas tiene un efecto indeseable: puede frenar la producción y el empleo. De manera similar, los efectos del aumento de la tasa de desempleo significan aumentos en la probabilidad de impago, esto se debe al hecho de que

hay deterioro en el poder adquisitivo del cliente, disminución en el consumo y dado que no se reciben salarios es más probable que no se asuman las responsabilidades contraídas con los bancos. Mayor tasa de desempleo significa disminuciones en el PIB de la economía.

## **7. CONCLUSIONES**

En el presente trabajo se utilizó un modelo lineal generalizado que asume observaciones independientes o al menos no correlacionadas, con el que se busca determinar si la inclusión de variables macroeconómicas como la tasa desempleo, la tasa de interés o la inflación permiten ayudar a predecir la probabilidad de incumplimiento teniendo en cuenta los distintos perfiles de clientes. El modelo fue concluyente y permitió determinar que la inclusión de variables macroeconómicas si ayudan a determinar de mejor manera el impago o no de un determinado cliente.

Después del análisis realizado se puede concluir que las variables macroeconómicas propuestas en la herramienta credit scoring modelada si tienen efectos sobre el incremento o disminución de la solicitud de créditos. Adicional hay que resaltar la importancia de como cambia la probabilidad de impago cuando una variable macroeconómica cambia como en el caso de Colombia, ya que este contexto permite predecir de una mejor manera la probabilidad de impago de un cliente ante dichas situaciones que afectan el ambiente económico de manera importante; además al tener en cuenta variables macroeconómicas dentro del análisis permite dar una mayor robustez al modelo.

Por ejemplo, si la tasa de interés se incrementa, los créditos se vuelven más costosos y al serlo hay menor endeudamiento, por lo que los préstamos solicitados a las entidades financieras disminuyen en general.

El desempleo por su parte afecta la forma de pago de los clientes, debido a que si en determinado momento los individuos se quedan sin trabajo o el sector en el cual trabajan como independientes se ve afectado por factores internos o externos, esto causa que no tengan los recursos para cumplir con sus obligaciones financieras lo cual afecta tanto al banco que prestó el dinero como al cliente debido a los intereses que siguen corriendo sobre el préstamo y la calificación que se genera en las centrales de riesgo.

Adicionalmente, si los clientes no cuentan con un empleo estable, no pueden demandar créditos para determinadas necesidades que tengan, por lo cual la manera de buscar recursos se disminuye y esto tendrá un impacto en la reducción de solicitud de créditos a las entidades financieras.

Todos los escenarios anteriormente mencionados, tienen un gran impacto sobre la probabilidad de que el cliente entre en impago o no, de esta forma, incluir variables macroeconómicas en la estimación del scoring de crédito va a permitir predecir de mejor forma el comportamiento de los individuos sobre la solicitud de préstamos en determinada entidad bancaria.

## **8. REFERENCIAS Y BIBLIOGRAFÍA**

- Aguas, D. y Castillo, M. (2002). “Modelo de administración del riesgo crediticio para la cartera comercial de una entidad financiera colombiana”, Apuntes de Banca y Finanzas, No. 6, pgs. 1-8.
- Altman, E. I. (1968). “Financial ratios, discriminant analysis and the prediction of the corporate bankruptcy”. *Journal of Finance* 43, pgs. 589-609.
- Bonfim, D. (2009) “Credit risk drivers: evaluating the contribution of firm level information and of macroeconomic dynamics”. *Journal of Banking & Finance* 33 (2), pgs. 281-299.
- Boyes, W. J., Hoffman D. L. and Low, A. S. (1989). “An Econometric Analysis of the Bank Credit Scoring Problem”. *Journal of Econometrics*, 40, pgs. 3-14
- Carling, K, Jacobson, T., Lindé, R. K. (2007). “Corporate credit risk modeling and the macroeconomy” *Journal of Banking & Finance* 31 (3), pgs. 845-868.
- Domínguez, J., Miranda, F., Pallas, J. y Peraza, C. (2016). “La medición del riesgo de crédito y el nuevo acuerdo de capital del comité de Basilea”, [en línea], disponible en: <http://webcache.googleusercontent.com/search?q=cache:at2XXUAS9FkJ:www.uv.es/as/epuma/XI/31.pdf+&cd=1&hl=es&ct=clnk&gl=co>, recuperado: 1 de mayo de 2016.
- Elizondo, A. (2003). “Medición integral del riesgo de crédito”. México, Editorial Limusa.

- López, C. (1999). "El riesgo de crédito en México: una evaluación de modelos recientes para cuantificarlo". *Gaceta de economía*, 4(8), pgs. 51-74.
- Fernández, H. y Pérez, F. (2005). "El modelo logístico: una herramienta estadística para evaluar el riesgo de crédito", *Revista de Ingenierías Universidad de Medellín*, Vol. 4, No. 006, pgs. 55-75.
- Gordy, M. B. (2000). "A Comparative Anatomy of Credit Risk Models". *Journal of Banking & Finance* 24, pgs. 119-149.
- Greene, W. H. (1992). "A Statistical Model for Credit Scoring". Mimeo
- Gutierrez, M. (2007). "Modelos de credit scoring: qué, cómo, cuándo y para qué", MPRA Paper, No. 16377, pgs.1-30
- Hand, D. J. and Henley, W. E. (1997). "Statistical Classification Methods in Consumer Credit Scoring: a Review". *Royal Statistical Society*, 160 (1997), Part 3, pgs. 523-541
- Harris, T. (2015). "Credit scoring using the clustered support vector machine". *Expert Systems with Applications* 42 (2), pgs. 741-750.
- Li, Z., Crook, J., Andreeva, G. (2017). "Dynamic prediction of financial distress using malmquist DEA. *Expert Systems with Applications* 80, pgs. 94-106.
- Medina, S. y Paniagua, G. (2008). "Modelo de Inferencia Difuso para estudio de crédito" *Dyna*, Año 75, No. 154, pgs. 215-229
- Mermelstein, D. (2006). "Mortgage defaults, macroeconomics, and institucional arrangements: Beyond the standar Credit Scoring".
- Minsky, H. (1992). "The financial instability hypothesis", The Jerome Levy Economics Institute, Working paper N° 74.
- Ochoa J., Galeano W., Agudelo L. (2010). "Construcción de un modelo de scoring para el otorgamiento de crédito en una entidad financiera". *Perfil de Coyuntura Económica*, No. 16, pgs. 191-222, Universidad de Antioquia.
- Ohlson, J. (1980). "Financial ratios and the probabilistic prediction of bankruptcy". *Journal of Accounting Research* 18, pgs. 109-131.
- Pérez, F. Y Fernández, H. (2007). "Las Redes Neuronales y la Evaluación del Riesgo de Crédito", *Revista Ingenierías Universidad de Medellín*, Vol. 6, No. 10, pgs. 77-91.

- Schreiner, M. (2002). “Ventajas y desventajas del scoring estadístico para las microfinanzas”, *microfinancerisk Management*, Washington University in St. Louis, pgs. 1-40.
- Sepúlveda, C., Reina, W. y Gutiérrez, J. (2012). “Estimación del riesgo de crédito en empresas del sector real en Colombia”. *Estudios gerenciales*, VOL 28, NO. 124, pgs. 169-190.
- Srinivasan, V. and Kim, Y. H., (1987). “Credit Granting: A Comparative Analysis of Classification Procedures”. *The Journal of Finance*, vol. XLII, N° 3.
- Sun J., Fujita H., Chen P., Li H. (2017). “Dynamic financial distress prediction with concept drift base on time weighting combined with Adaboost support vector machine ensemble. *Knowledge- Based Systems* 120, pgs. 4-14.
- Superintendencia Financiera de Colombia. (1995). Circular básica contable y financiera 100 de 1995, Colombia
- Thomas, L. C. (2000). “A Survey of credit and behavioral scoring: forecasting financial risk of lending to consumers”. *International Journal of Forecasting* 16, pgs. 149-172.

## ANEXOS

VARIABLES UTILIZADAS EN EL MODELO TRADICIONAL POR LA IFBEC	
VARIABLE	DESCRIPCIÓN
NObs (Observación)	Indica el individuo Xi, en determinado periodo de tiempo ti
Fecha_eval (fecha evaluación)	Fecha en que el crédito fue evaluado por el analista del área de riesgos
Fecha_contable ( fecha contable)	Fecha en la que se desembolsó el crédito
VR_DESEMBOLSO (valor del desembolso)	Valor desembolsado del crédito
Plazo_mes ( plazo en meses )	Plazo del crédito otorgado, medido en meses
LINEA ( línea a la que corresponde el crédito)	En la base de datos se tienen en cuenta tres productos: consumo, tarjeta de crédito e hipotecario
LINEA_subproducto ( especificación del producto solicitado por el cliente )	hipotecaria ahí especifica si es VIS ( interés social) NO VIS ( no interés social), para consumo se encuentra libre inversión, libranza y vehículo y para TDC se encuentra cupo rotativo o tarjeta de crédito
SEXO	Si el cliente es Hombre (M), si es Mujer (F)
EDAD	Edad del cliente, medida en años
SI_NO_CLIENTE	Si es cliente -> SI, si no es cliente ->NO
SI_NO_NOMINA	Si el cliente domicilia la nómina ->SI, sino domicilia la nómina ->NO
Estrato	Estrato donde vive el cliente, va desde el 1 hasta el 6
ANTIGUO_CLIE	Antigüedad del cliente con la entidad financiera, medida en años.
ANTIGUO_LABO	Antigüedad laboral del cliente
VR_PATRIMONIO	Patrimonio declarado por parte del cliente
VR_VIVIENDA	Valor de la vivienda, informado por el cliente
D_ESTADO_CIVIL	Estado civil del civil catalogado así: casado bienes mancomunados, casado bienes separados, divorciado, separado, soltero, unión libre y viudo
D_BIN	Hace referencia a las tarjetas de crédito de acuerdo a la franquicia que pertenecen : VISA o MASTER CARD. Cada franquicia tiene tarjetas para los diferentes perfiles de clientes
D_ESTADO_BIEN	Hace referencia al estado del bien a comprar esto solo aplica para hipotecario y vehículo que puede ser nuevo o usado
D_SITUA_LABORAL	Se clasifica en: fijo (contrato término indefinido) , temporal, independiente ( actividades como comerciantes, agricultores, más que todo clientes autónomos ), otros hace referencia a pensionados, rentista de capital.
D_SITUA_VIVIENDA	Tipo de vivienda que posee el cliente puede ser hipotecada, alquilada, propia, otros, familiar.
D_NACIONALIDAD	Nacionalidad del cliente: colombiano o extranjero
D_PROFESION	La profesión que informa el cliente y es soportada con documentos
D_ACTIVIDAD	Esta hace referencia a la actividad que desempeña el cliente por ejemplo puede tener como profesión administrador y trabajar en una empresa petrolera entonces la actividad saldría petróleo.

MORA_ACTUAL	Todas las variables que tienen ACTUAL es a corte de Julio, y en esta variable las que se encuentren con signo positivo son los días que el cliente tiene de mora en la obligación, y las que tienen signo negativo son los días con anterioridad que el cliente
MORA_M_ANT	Hace referencia a la mora en el mes anterior, un cliente entra en mora cuando supera los 30 días.
MORA_t1	Mora hace 1 año
MORA_t2	Mora hace 2 años
Estado_inversion	Hace referencia al estado en el cual se encuentra la operación : al día, vigente, vencida o en mora
SALDO_ACTUAL	Saldo actual de la obligación adquirida
SALDO_M_ANT	Saldo mes anterior
SALDO_t1	Saldo hace 1 años de la obligación adquirida
SALDO_t2	Saldo hace 2 años de la obligación adquirida
SALDO_DIC_t1	
SALDO_DIC_t2	
SALDO_DIC_t3	
Mora_30	Esta variable hace referencia a una ventana de tiempo de 3 meses donde el cliente ha alcanzado una altura de mora de 30 días, al menos 1 vez. Con lo cual 1= entró en mora y 0= no entró en mora; una vez se haya marcado con un 1, el cliente seguirá marcado.
Mora_60	Esta variable hace referencia a una ventana de tiempo de 6 meses donde el cliente ha alcanzado una altura de mora de 60 días, al menos 1 vez. Con lo cual 1= entró en mora y 0= no entró en mora; una vez se haya marcado con un 1, el cliente seguirá marcado.
Mora_90	Esta variable hace referencia a una ventana de tiempo de 12 meses donde el cliente ha alcanzado una altura de mora de 90 días, al menos 1 vez. Con lo cual 1= entró en mora y 0= no entró en mora; una vez se haya marcado con un 1, el cliente seguirá marcado
Tasa_mora_30	Desde el desembolso del crédito hasta hoy, determina si el cliente ha tenido incumplimiento de 30
Tasa_mora_60	Desde el desembolso del crédito hasta hoy, determina si el cliente ha tenido incumplimiento de 60
Tasa_mora_90	Desde el desembolso del crédito hasta hoy, determina si el cliente ha tenido incumplimiento de 90

Inflación (Variación Anual)		Tasa Desempleo		Tasa Repo (Corte Trimestral)	
COCPYOY Index		COUNTOTN Index		CORRRMIN Index	
Date	PX_LAST	Date	PX_LAST	Date	PX_LAST
3/30/2012	3,4	3/30/2012	10,36	3/30/2012	5,25
6/29/2012	3,2	6/29/2012	10,03	6/29/2012	5,25
9/28/2012	3,08	9/28/2012	9,94	9/28/2012	4,75
12/31/2012	2,44	12/31/2012	9,55	12/31/2012	4,25
3/29/2013	1,91	3/29/2013	10,21	3/29/2013	3,25
6/28/2013	2,16	6/28/2013	9,24	6/28/2013	3,25
9/30/2013	2,27	9/30/2013	8,98	9/30/2013	3,25
12/31/2013	1,94	12/31/2013	8,44	12/31/2013	3,25
3/31/2014	2,51	3/31/2014	9,73	3/31/2014	3,25
6/30/2014	2,79	6/30/2014	9,19	6/30/2014	4
9/30/2014	2,86	9/30/2014	8,35	9/30/2014	4,5
12/31/2014	3,66	12/31/2014	8,72	12/31/2014	4,5
3/31/2015	4,56	3/31/2015	8,86	3/31/2015	4,5
6/30/2015	4,42	6/30/2015	8,25	6/30/2015	4,5
9/30/2015	5,35	9/30/2015	8,98	9/30/2015	4,75
12/31/2015	6,77	12/31/2015	8,59	12/31/2015	5,75
3/31/2016	7,98	3/31/2016	10,14	3/31/2016	6,5
6/30/2016	8,6	6/30/2016	8,88	6/30/2016	7,5

Fuente: Bloomberg

\*Se deben tener en cuenta algunas propiedades de los logaritmos, tales como:

$$\exp \ln(a) = a$$

$$\log a^y = y \log a$$

$$\log(ab) = \log a + \log b$$

$$\log(a) - \log(b) = \log\left(\frac{a}{b}\right)$$