



**DISEÑO DE UN MODELO DE SCORING PARA EL OTORGAMIENTO DE
CRÉDITO DE CONSUMO EN UNA COMPAÑÍA DE FINANCIAMIENTO
COLOMBIANA**

LAURA ARANGO DUQUE*

DANIEL RESTREPO BAENA**

UNIVERSIDAD EAFIT

ESCUELA DE ECONOMÍA Y FINANZAS

MAESTRÍA EN ADMINISTRACIÓN FINANCIERA*

ESCUELA DE ADMINISTRACIÓN

MAESTRÍA EN ADMINISTRACIÓN (MBA)**

MEDELLÍN

2017

**DISEÑO DE UN MODELO DE SCORING PARA EL OTORGAMIENTO DE
CRÉDITO DE CONSUMO EN UNA COMPAÑÍA DE FINANCIAMIENTO
COLOMBIANA**

**Trabajo presentado como requisito parcial para optar al título de magíster en
Administración Financiera (*) y al de magíster en Administración (MBA, **)**

LAURA ARANGO DUQUE^{*1}

DANIEL RESTREPO BAENA^{2}**

Asesor: Rodrigo Henao Jassan, MAF

**UNIVERSIDAD EAFIT
ESCUELA DE ECONOMÍA Y FINANZAS
MAESTRÍA EN ADMINISTRACIÓN FINANCIERA*
ESCUELA DE ADMINISTRACIÓN
MAESTRÍA EN ADMINISTRACIÓN (MBA)****

MEDELLÍN

2017

¹ lauraarango02@hotmail.com

² drestrepob@yahoo.com

Contenido

RESUMEN.....	10
Introducción.....	11
Marco teórico.....	14
1.1. Riesgo de crédito	14
1.1.1. Factores determinantes del riesgo de crédito	15
1.1.2. Tipos de riesgo de crédito	15
1.2. Comité de Basilea	16
1.2.1. Basilea I.....	17
1.2.2. Basilea II.....	18
1.2.3. Basilea III.....	21
1.3. Superintendencia Financiera de Colombia	22
1.4. Sistema de Administración de Riesgo Crediticio (SARC).....	24
1.4.1. Contenido del SARC	25
1.4.2. Etapa de otorgamiento	26
1.5. Crédito de consumo en Colombia.....	30
1.6. Modelos de scoring	33
1.6.1. Análisis discriminante	34
1.6.2. Modelo probabilístico.....	40
1.6.3. Modelo logístico.....	42
1.6.4. Redes neuronales artificiales	43
Aspectos metodológicos	46
1.7. Descripción de la base de datos	47
1.8. Metodología de modelación	54
Presentación y análisis de resultados.....	56
1.9. Elección del mejor modelo	57
1.9.1. Regresión logística	57
1.10. Coeficientes estimados y modelo de scoring.....	60

1.11. Importancia de las variables	63
1.12. Validación cruzada.....	66
1.13. Prueba de Kolmogorov-Smirnov.....	68
Conclusiones y recomendaciones.....	71
Referencias	74

Lista de tablas

	Pág.
Tabla 1. Variables incluidas en el modelo	48
Tabla 2. Matriz de confusión	68
Tabla 3. Matriz de confusión en porcentaje	68

Lista de gráficas

	Pág.
Gráfica 1. Cartera de consumo en Colombia	32
Gráfica 2. Porcentaje de adultos colombianos con algún producto en todas las entidades	33
Gráfica 3. Histogramas de cada grupo y centroides representados en la función discriminante	36
Gráfica 4. Utilización de un punto de corte equidistante de ambos centroides	38
Gráfica 5. Esquema de una red neuronal	45
Gráfica 6. Regresión logística	60
Gráfica 7. Resultado de la prueba de Kolmogorov-Smirnov	70

Lista de figuras

	Pág.
Figura 1. Proceso metodológico	54

Lista de anexos

	Pág.
Anexo 1. Coeficientes estimados	79
Anexo 2. Intervalos de confianza	83

Glosario

CAPACIDAD DE PAGO: indicador que evalúa la suficiencia que tiene un posible deudor para atender una obligación de manera adecuada.

CAPITAL: aporte de los socios o accionistas propietarios de una empresa; se expresa como una cuenta en el patrimonio.

COMPAÑÍA DE FINANCIAMIENTO: entidad cuya función principal es captar recursos a término con el objetivo de realizar operaciones activas de crédito.

CRÉDITO DE CONSUMO: préstamo otorgado a personas naturales con el fin de financiar la adquisición de bienes y servicios no generadores de renta.

INCUMPLIMIENTO: estado en el que entra un deudor cuando deja de realizar los pagos de una deuda, y la entidad financiera considera que el recaudo de los recursos no se va a realizar por parte del primero.

LIQUIDEZ: capacidad de una persona, natural o jurídica, para atender sus obligaciones financieras.

SCORING DE CRÉDITO: sistema de evaluación que permite, a partir de un conjunto de variables cualitativas y cuantitativas, predecir la probabilidad de pago de un deudor.

RESUMEN

Los scoring de crédito son herramientas estadísticas usadas desde los años setenta para asignar una probabilidad de incumplimiento al solicitante de un crédito, al realizar una ponderación de sus variables cualitativas y cuantitativas y, a su vez, con inclusión de las condiciones del crédito solicitado. El uso de scoring se ha potenciado desde los años noventa con el fin de mitigar el riesgo de crédito de las solicitudes y buscar los clientes que se ajusten al perfil deseado por la entidad, con el fin de dar cumplimiento a lo establecido por el ente regulador en cuanto a la implementación de un sistema de riesgo que disminuya las posibles pérdidas económicas por medio del cuidado de cada uno de las etapas que comprenden el ciclo de crédito, en este caso, de manera específica, el otorgamiento. El objetivo de este trabajo es diseñar un modelo de scoring basado en la metodología de regresión logística para una entidad financiera del valle de Aburrá, vigilada por la Superintendencia Financiera de Colombia, con base en la información entregada por dicha entidad, que comprende variables cualitativas y cuantitativas de un grupo de clientes en un período establecido. Para alcanzar el objetivo se llevó a cabo la modelación de una base de datos suministrada por la entidad financiera mencionada en el programa estadístico *R*, de lo que se obtuvo el diseño de un modelo de scoring de 23 variables cualitativas y cuantitativas capaz de predecir la probabilidad de incumplimiento del deudor.

Palabras clave: scoring, riesgo de crédito, otorgamiento de crédito, crédito de consumo, modelo logístico.

Abstract

Credit scorings are tools used since the 1970s to grant a probability of default to the applicant of a credit, to weigh their qualitative and quantitative variables; and in turn, including the conditions of the requested credit. The use of scoring has been boosted since the 1990s, in order to mitigate the credit risk of the requests and to

find the clients that fit the profile desired by the entity, complying with what was established by the regulator regarding the implementation of a risk system that reduces economic losses taking care of each stage of the credit cycle, in this case specifically, the granting. Therefore, the objective of this work is to design a scoring model based on the methodology of logistic regression for a financial institution of the valle de Aburrá, supervised by the Superintendencia Financiera de Colombia, based on information provided by the entity, which comprises qualitative and quantitative variables of a group of clients in a set period of time.

Key words: *scoring, credit risk, credit granting, consumer credit, logit model.*

Introducción

Las instituciones financieras en Colombia que tienen entre sus funciones principales el otorgamiento de créditos a empresas y personas se denominan Establecimientos de Crédito (Presidencia de la República, 1993). Esta actividad de prestar dinero, a la espera de que sea devuelto en un tiempo posterior por un valor mayor –como consecuencia del cobro de intereses– conlleva un riesgo de pérdidas para las entidades, en caso de que los clientes a los que se les otorgaron los recursos no los devuelvan en las condiciones pactadas, es decir, cuando los clientes incumplen sus compromisos (Superintendencia Financiera de Colombia, 1995).

Por lo anterior resulta necesario para dichas instituciones establecer mecanismos que permitan identificar la probabilidad de incumplimiento de los clientes desde el momento mismo en que solicitan los préstamos, de forma que la decisión de aprobación se pueda tomar por medio de la medición de la exposición de riesgo inherente a cada uno de los créditos por otorgar. Según Saavedra García y Saavedra García (2009, p. 299), “los modelos para estimar la probabilidad de

incumplimiento surgieron de manera formal durante la década de los setenta; sin embargo, desde los años treinta ya se habían iniciado estudios basados en el análisis tradicional de razones financieras”.

Así mismo, la evolución que ha tenido el sector financiero a lo largo de la historia muestra que el riesgo de pérdida antes referido (denominado riesgo de crédito) se comporta de manera diferente según el tipo de cliente y el destino que se les da a los recursos (Samaniego Medina, 2008). No es lo mismo otorgar un préstamo a una empresa para invertir en su actividad económica que prestarle dinero a una persona para un viaje de vacaciones. En consecuencia, los enfoques para la gestión del riesgo de crédito se fueron especializando según esos y muchos otros criterios.

Los préstamos que las instituciones financieras ofrecen a las personas del común para que utilicen el dinero en lo que deseen se denominan de consumo y son una modalidad de alta participación en dicha industria: de acuerdo con la información publicada por la Superintendencia Financiera de Colombia al corte de septiembre de 2016, las operaciones de crédito de consumo representaban el 28% del total de las operaciones activas de todos los establecimientos de crédito, solo superadas por la cartera comercial (préstamos a empresas), con una participación del 56%. De ahí que resulte relevante abordar la problemática relacionada con la necesidad de esquemas de medición de riesgo para el otorgamiento de créditos de consumo.

Por otro lado, la compañía de financiamiento sobre la que se pretende desarrollar el presente trabajo por tradición ha sido poco activa en la colocación de créditos de consumo, que representan escasamente el 1% del total de su cartera. Su plan de negocios ha estado más enfocado hacia el ofrecimiento de soluciones de financiación para las pequeñas y las medianas empresas en las principales

ciudades de Colombia mediante un portafolio de productos de créditos comerciales. Sin embargo, por una decisión estratégica del negocio, motivada por los drásticos cambios del entorno económico que se han presentado en los últimos dos años, así como por la necesidad manifiesta de tener una composición de cartera más diversificada, ha determinado darles un mayor impulso a las líneas de crédito de consumo dentro de su cartera de préstamos.

Por los mencionados motivos, se planteó llevar a cabo la presente investigación cuyo objetivo es diseñar un modelo de scoring que permita medir la exposición al riesgo de crédito a partir del otorgamiento de los créditos de consumo, en la compañía de financiamiento objeto del estudio, por medio del análisis de las técnicas estadísticas sobre las que se desarrollan los modelos de scoring para establecer el que mejor se ajusta a las características del perfil de riesgo de los prestamistas de la entidad financiera sobre la que se apoyará esta investigación e identificar las variables cualitativas y cuantitativas, disponibles en la población en estudio, que pueden incidir en el incumplimiento de las personas naturales que han contratado créditos de consumo con dicha entidad financiera.

Se inició la investigación con una recolección de referencias bibliográficas que incluyeran los temas por abordar, se continuó con la consecución de la base de datos sobre la que se diseñó el modelo de scoring, se depuró la misma y se elaboró la modelación correspondiente y por último se examinaron los resultados obtenidos.

El presente documento comienza con un marco teórico en el que se definen los conceptos básicos, seguido de un análisis de las técnicas estadísticas sobre las que se desarrollan los modelos de scoring para la evaluación de clientes de créditos de consumo en el sector financiero. Luego continúa con la descripción

detallada de la metodología de investigación que se siguió para alcanzar los objetivos planteados. Para finalizar presenta los resultados obtenidos y las conclusiones propuestas.

Marco teórico

1.1. Riesgo de crédito

La evolución económica y social ha favorecido la consolidación del sistema financiero fundamentada en una disminución en la resistencia a apalancarse por parte de las personas naturales y de las empresas. Por tal razón, el riesgo se ha constituido en un “factor estratégico, no solo para las entidades financieras, sino para cualquier organización con independencia del tamaño que posea y el sector en el que realice su actividad” (Samaniego Medina, 2008, pp. 2-3).

Según Echemendía Tocabens (2011), el riesgo se define como la posibilidad de ocurrencia de un evento negativo o la probabilidad de tener un resultado no deseado. El riesgo de crédito, por su parte, se constituye en un tipo de riesgo financiero que consiste en “la posibilidad de que una entidad incurra en pérdidas y se disminuya el valor de sus activos, como consecuencia de que un deudor o contraparte incumpla sus obligaciones” (Superintendencia Financiera de Colombia, 1995, p. 3). Según Samaniego Medina (2008, p. 16), el riesgo de crédito consiste en “riesgo de cambios en la calidad crediticia del emisor”. Si se consideran ambas definiciones, el riesgo de crédito se determina como la probabilidad de que un deudor incurra en impagos derivados de cambios en su calidad crediticia, lo que trae como consecuencia que el prestamista presente pérdidas por dicho incumplimiento.

1.1.1. Factores determinantes del riesgo de crédito

Samaniego Medina (2008) precisa que existen tres factores que permiten medir el riesgo de crédito: probabilidad de incumplimiento, exposición y severidad o tasa de recuperación.

- La probabilidad de incumplimiento se interpreta como la de que el prestamista eluda sus responsabilidades contractuales.
- La exposición es el valor de pérdida en el que se incurre en el momento del incumplimiento de la contraparte.
- Por último, se conoce como severidad al porcentaje de pérdida resultante luego del incumplimiento y la recuperación. De acuerdo con Saavedra García y Saavedra García (2009, p. 299), la recuperación depende de la garantía del crédito puesto que una garantía disminuye el riesgo de crédito si la cobertura de la deuda que proporciona es adecuada y es de fácil realización, es decir, puede convertirse en dinero con facilidad.

1.1.2. Tipos de riesgo de crédito

El riesgo de crédito se puede clasificar en:

- a. **Riesgo de impago:** es el riesgo relacionado con el no pago por parte del prestamista, tanto de capital como de intereses (Samaniego Medina, 2008).

- b. Riesgo de calificación:** incertidumbre relacionada con la alteración de la calidad crediticia del acreditado o emisor (Rayo, 2013).

1.2. Comité de Basilea

Fue creado en diciembre de 1974, luego de la caída del sistema de tasas fijas y la insolvencia de los bancos Bankhaus Herstatt y Franklin National Bank, ubicados en Alemania y en Estados Unidos, en su orden, lo que ocasionó un desequilibrio en el sistema financiero internacional.

Sus integrantes fueron los países miembros del G-10 (Alemania, Bélgica, Canadá, Estados Unidos, Francia, Italia, los Países Bajos, Japón, el Reino Unido y Suecia).

Las funciones del Comité son las siguientes:

- a.** Establecer un foro adecuado para el diálogo de las problemáticas propias de la supervisión.
- b.** Coordinar las responsabilidades de la supervisión entre las autoridades delegadas para el efecto, con el fin de garantizar una supervisión efectiva.
- c.** Instaurar los estándares de supervisión que hacen referencia a la solvencia de las entidades financieras.

La naturaleza del Comité no se basa en una autoridad supranacional de supervisión, sino en la institución de esquemas y recomendaciones. Su

incorporación en la legislación de los países del G-10 ha generado una influencia importante en el resto de la comunidad internacional que ha logrado que hoy en día la mayoría de los países adopten sus premisas (Ustáriz González, 2003). En términos generales, el Comité de Basilea se erige como una institución dedicada a formular las mejores prácticas para la administración de los riesgos financieros en los intermediarios bancarios.

1.2.1. Basilea I

Acuerdo publicado en el año 1988. Su iniciativa radicaba en la necesidad de establecer normas cautelares con respecto al capital mínimo requerido por las entidades bancarias para asumir los riesgos crediticios.

Como punto fundamental del acuerdo se estableció un nivel mínimo de capital del 8% con respecto a los activos ponderados por riesgo (razón de solvencia), con el objetivo de fortalecer el sistema bancario internacional y alinear la regulación bancaria para disminuir la desigualdad competitiva. Se entiende por activos ponderados por riesgo la sumatoria de los activos ponderados según el nivel de riesgo de crédito de cada uno, que se obtiene al multiplicar el valor nominal de cada activo por el coeficiente de ponderación según el nivel de riesgo de crédito que cada activo represente para la entidad.

El cálculo de la razón de solvencia es el siguiente:

Ecuación 1. Razón de solvencia (Basilea I):

$$\text{Razón de solvencia} = \frac{\text{Capital del banco}}{\text{Activos ponderados por riesgo}} \geq 8\%$$

Con posterioridad se llevó a cabo una revisión completa al documento, lo que trajo consigo algunas mejoras, que fueron incluidas en el acuerdo de Basilea II (Herrera García, 2009).

1.2.2. Basilea II

Acuerdo publicado en el año 2004, cuyo propósito fue “construir una base sólida para la regulación prudente del capital, la supervisión y la disciplina de mercado, así como perfeccionar la gestión del riesgo y la estabilidad financiera” (Comité de Supervisión Bancaria de Basilea, 2004, p. 5). Se ratificó la importancia de tres pilares: i) requisitos de capital mínimo, ii) un proceso de examen supervisor y iii) utilización eficaz de la disciplina de mercado. Es importante mencionar aquí que con base en el Acuerdo de Basilea II se instituyeron los sistemas de administración de riesgos (SAR), que son los modelos de evaluación y análisis que se aplican hoy en día en los intermediarios financieros vigilados por la Superintendencia Financiera de Colombia.

Pilar 1: requisitos mínimos de capital

Este requisito establece un capital común del 8%, al igual que en el Acuerdo de Basilea I; la diferencia radica en el cálculo del denominador del coeficiente o razón de solvencia, en el que se incluyeron, además de los activos ponderados por riesgo crediticio, los riesgos de mercado y operativo. En este orden de ideas se propuso una nueva razón de solvencia, calculada así (Banco de Pagos Internacionales, 2001):

Ecuación 2. Razón de solvencia (Basilea II):

Razón de solvencia

$$= \frac{\text{Capital del banco}}{\text{Resgo de crédito} + 12,5 * (\text{riesgo de mercado} + \text{riesgo operativo})}$$

$\geq 8\%$

El Comité postuló nuevas alternativas para el cálculo de los riesgos de crédito y operativo, mientras que el cálculo del de mercado se mantuvo igual al propuesto en Basilea I. No se profundizará en dichos métodos al tener en cuenta que no se encuentran en el alcance de la presente investigación.

Pilar 2: proceso de examen supervisor

La finalidad del segundo pilar es “asegurar que cada banco cuenta con procesos internos confiables para evaluar la suficiencia de su capital, a partir de una apreciación meticulosa de sus riesgos” (Banco de Pagos Internacionales, 2001, p. 7), lo que se hará por medio de la evaluación de los supervisores.

Se plantearon cuatro principios fundamentales:

“Principio 1: los bancos deben tener un proceso de evaluación de la suficiencia de capital total en relación a su perfil de riesgo y una estrategia para mantener sus niveles de capital” (Banco de Pagos Internacionales, 2001, p. 37), esto es, los bancos deberán asegurar que su nivel de capital es el adecuado para su perfil de riesgo y su ciclo económico, así como las estrategias para mantenerlo.

“Principio 2: los supervisores deben examinar y evaluar las estrategias y valoraciones internas de la suficiencia del capital de los bancos, así como su habilidad para vigilar y asegurar la aplicación de los coeficientes de capital de supervisión. Los supervisores deben intervenir si no están satisfechos con el resultado de este proceso” (Banco de Pagos Internacionales, 2001, p. 37). Los supervisores deberán garantizar el cumplimiento del principio 1 mediante la vigilancia y el seguimiento periódico a los bancos.

“Principio 3: Los supervisores deben esperar que los bancos operen por encima de los coeficientes mínimos de capital regulador y tener la habilidad de exigir a los bancos que mantengan capital en exceso del mínimo” (Banco de Pagos Internacionales, 2001, p. 37). Los supervisores se encargarán de velar porque los bancos cumplan el mantenimiento del capital requerido por encima de los estándares mínimos legales.

“Principio 4: los supervisores deben procurar intervenir temprano para evitar que el capital descienda por debajo de los niveles mínimos necesarios para cubrir las características de riesgo de un banco particular y exigir una acción correctiva inmediata cuando el capital no está en el nivel requerido o no es restaurado a ese nivel” (Banco de Pagos Internacionales, 2001, p. 37). Es obligación de los supervisores exigir medidas correctivas a aquellos bancos que presenten niveles de capital por debajo de los mínimos establecidos, con el fin de asegurar el cumplimiento de los tres primeros principios.

Pilar 3: disciplina de mercado

La disciplina de mercado consiste en una continua divulgación de información por parte de los bancos que posibilite que los participantes del mercado se encuentren informados acerca de su condición, su perfil de riesgo y, por ende, su solidez (Comité de Supervisión Bancaria de Basilea, 2004).

1.2.3. Basilea III

Acuerdo publicado en el año 2010, como consecuencia de la crisis financiera del año 2007. Tiene como objetivo establecer lineamientos para “reforzar las normas internacionales de capital y liquidez con el fin de promover un sector bancario más resistente” (Comité de Supervisión Bancaria de Basilea, 2010, p. 1).

Las reformas propuestas buscan mejorar la gestión del riesgo, reforzar las buenas prácticas en las entidades y estimular la transparencia y la adecuada divulgación de la información con el fin de constituir un sistema financiero sólido que se traduzca en crecimiento económico. Según Álvarez Franco y Osorio Betancur (2011), dichas reformas incluyen:

- Fortalecimiento del capital, sustentado en mayor calidad, consistencia y transparencia. Con esta reforma se busca reforzar la capacidad de los bancos para asumir pérdidas y mejorar su solvencia.
- Constitución de colchones de capital en períodos de abundancia que permitan hacer frente a posibles crisis.
- Introducción de un índice o razón de apalancamiento que evite el exceso del mismo.

- Introducción de un estándar de liquidez internacional, tanto para el corto como para el largo plazo, que garantice la capacidad de los bancos de hacer frente a situaciones de tensión.
- Reforzamiento de los tres pilares planteados en el Comité de Basilea.

1.3. Superintendencia Financiera de Colombia

La Superintendencia Financiera de Colombia “es un organismo técnico adscrito al Ministerio de Hacienda y Crédito Público, con personería jurídica, autonomía administrativa y financiera y patrimonio propio” (Superintendencia Financiera de Colombia, 2017). Surgió de la fusión de la Superintendencia Bancaria de Colombia y la Superintendencia de Valores (Manjarrés V., 2012).

La entidad tiene como objetivo principal la supervisión de los principales riesgos a los que se ven expuestas las entidades que conforman el sistema financiero, a saber: de crédito, operativo, de mercado y de lavado de activos.

Según la ley 1328 de 2009, artículo 67, la función de supervisión ejercida por la Superintendencia se basará en los siguientes principios:

- a. Principio de materialidad:** La Superintendencia Financiera de Colombia ejercerá sus funciones y facultades de inspección y vigilancia, así como de control, dando especial relevancia y atención a aquellos hechos que por su naturaleza,

cuantía o circunstancias coyunturales, afecten de manera grave la confianza pública en el sistema financiero, asegurador y en el mercado de valores, pongan en peligro la continuidad del servicio o comporten un riesgo sistémico.”

b. Principio de acceso a la información para la protección de la estabilidad y confianza en el sistema financiero: Cuando a juicio del Superintendente Financiero existan elementos que previsiblemente puedan llegar a vulnerar la estabilidad, seguridad y confianza en el sistema financiero, asegurador o en el mercado de valores, la Superintendencia Financiera de Colombia podrá requerir, en cualquier tiempo, cualquier información de entidades que ordinariamente no se encuentren bajo su inspección y vigilancia o control (Congreso de Colombia, 2009, artículo 67).

Según Manjarrés V. (2012), entre las facultades de la entidad se encuentran:

- Concesión para la creación de establecimientos financieros.
- Autorización de funcionamiento a dichos establecimientos.
- Regulación de las entidades que hacen parte del sistema financiero.
- Asesoramiento a los participantes del sistema financiero.
- De control.
- De crear o proveer información.
- De certificación.
- De interpretación de las normas.

- Jurisdiccionales y sancionatorias: la Superintendencia Financiera tiene la capacidad de tomar decisiones sobre los establecimientos e imponer las respectivas sanciones.

Así mismo, la supervisión ejercida debe ser independiente, universal, integral, autónoma, preventiva, suficiente y especializada (Manjarrés V., 2012).

Objetivos estratégicos

Los objetivos estratégicos establecidos para la Superintendencia Financiera de Colombia son los siguientes (Superintendencia Financiera de Colombia, 2017):

- Fortalecer la gestión funcional de la Superintendencia Financiera de Colombia.
- Desarrollar la supervisión basada en riesgos bajo la metodología MIS (Marco Integral de Supervisión).
- Fortalecer la supervisión consolidada.
- Robustecer los requerimientos prudenciales para las entidades vigiladas.
- Contribuir con mecanismos de inclusión y educación financiera.
- Velar por la protección al Consumidor Financiero.
- Apoyar el desarrollo del Mercado de Capitales.
- Fortalecer la gestión administrativa, financiera y defensa judicial de la Entidad.

1.4. Sistema de Administración de Riesgo Crediticio (SARC)

En el marco del Comité de Basilea II se estableció como segundo pilar el proceso del examen supervisor, que se basa en velar por que los bancos posean los

procesos correctos que les permitan asegurar el mínimo de capital requerido para cubrir los riesgos de sus actividades como prestamistas. Así mismo, tiene como objetivo estimular tanto el uso de técnicas para la gestión y el seguimiento del riesgo como su debida documentación (Comité de Supervisión Bancaria de Basilea, 2004).

Por lo anterior, la Superintendencia Financiera de Colombia, a través de la circular externa 011 (1995) y carta circular 035 (2002), estableció como obligatoria la implementación del Sistema de Administración de Riesgo Crediticio (SARC) para todas las entidades bajo su supervisión. El SARC tiene como propósito “mantener adecuadamente evaluado el riesgo de crédito implícito en los activos” (Superintendencia Financiera de Colombia, 1995, p. 3) y que cuenta con una infraestructura tecnológica óptima y los sistemas necesarios para ello. Dicho sistema fue adoptado por las entidades vigiladas según el marco legal de la circular externa 100 (básica, contable y financiera).

Según Torres Avendaño (2005), el SARC se define como

el conjunto de políticas, procedimientos, metodologías, herramientas informáticas y capital físico y humano de las entidades financieras, y que son dirigidas hacia la adquisición de conocimientos, medición y control de riesgos crediticios, en los que incurre una entidad dentro del giro normal de su negocio. Es un mecanismo integral de administración del riesgo crediticio (p. 125).

La definición del SARC proporcionada por Torres Avendaño es más global y abarca el objetivo correspondiente.

1.4.1. Contenido del SARC

El SARC incluye todas las etapas que componen el ciclo de crédito (otorgamiento, seguimiento y control y recuperación) y cuenta con los siguientes elementos básicos (Superintendencia Financiera de Colombia, 1995):

- Políticas de administración del riesgo crediticio.
- Proceso de administración del riesgo crediticio.
- Modelos para la estimación de las pérdidas esperadas.
- Sistema de provisiones para cubrir el riesgo crediticio.
- Procesos de control interno.

1.4.2. Etapa de otorgamiento

Al considerar el tema que compete a la presente investigación, a continuación se hace referencia a lo establecido en el capítulo II de la circular básica, contable y financiera (Superintendencia Financiera de Colombia, 1995) acerca del otorgamiento de crédito.

Se entiende como otorgamiento de crédito la etapa en la que se decide la aprobación de los créditos por parte de la entidad financiera, con base en el conocimiento del sujeto de crédito o contraparte, de su capacidad de pago y de las características del contrato por celebrar entre las partes, que incluyen, entre otros aspectos, las condiciones financieras del préstamo, las garantías, las fuentes de pago y las condiciones macroeconómicas a las que pueda estar expuesto (Superintendencia Financiera de Colombia, 1995).

Según la Superintendencia Financiera de Colombia (1995), los parámetros por considerar en esta etapa son:

a. Información previa al otorgamiento de crédito: consta de la información mínima requerida que se debe comunicar al potencial sujeto de crédito en cuanto a tasas de interés, base de capital sobre la que se aplicará la tasa de interés, comisiones y recargos, plazo del préstamo, condiciones de prepago, derechos de la entidad acreedora en caso de incumplimiento por parte del deudor, derechos del deudor y toda aquella información relacionada con las condiciones del contrato y su eficaz perfeccionamiento.

b. Selección de variables y segmentación de portafolios: el establecimiento de crédito debe diseñar un modelo de otorgamiento que cumpla los lineamientos exigidos por el ente regulador, con enfoque hacia su nicho de mercado objetivo, la consideración de las variables cualitativas y cuantitativas del sujeto de crédito y la vigilancia para cumplir las políticas de la entidad. Estos modelos deben ser revisados como mínimo dos veces al año, según establece la norma, en los meses de mayo y noviembre.

“Los procesos de segmentación y discriminación de los portafolios de crédito y de sus posibles sujetos de crédito, deben servir de base para su calificación” (Superintendencia Financiera de Colombia, 1995, p. 8).

c. Capacidad de pago del deudor: con el fin de calcular la probabilidad de incumplimiento del posible deudor y así mantener una adecuada administración del riesgo de crédito, la entidad financiera deberá evaluar la capacidad de pago del deudor y de cualquier persona natural o

jurídica que pueda resultar obligada al pago de los créditos (codeudor, deudor solidario, aval, etc.). Para el efecto se debe contar con al menos la siguiente información, que se toma de manera textual del capítulo II de la circular básica contable y financiera (Superintendencia Financiera de Colombia, 1995, pp. 8-9):

- Los flujos de ingresos y egresos, así como el flujo de caja del deudor y/o del proyecto financiado o a financiar.
- La solvencia del deudor, a través de variables como el nivel de endeudamiento y la calidad y composición de los activos, pasivos, patrimonio y contingencias del deudor y/o del proyecto.
- Información sobre el cumplimiento actual y pasado de las obligaciones del deudor. La atención oportuna de todas las cuotas o instalamentos, entendiéndose como tales cualquier pago derivado de una operación activa de crédito, que deba efectuar el deudor en una fecha determinada, independientemente de los conceptos que comprenda (capital, intereses, o cualquier otro). Adicionalmente, la historia financiera y crediticia, proveniente de centrales de riesgo, calificadoras de riesgo, del deudor o de cualquier otra fuente que resulte relevante.
- El número de veces que el crédito ha sido reestructurado y las características de la(s) respectiva(s) reestructuración(es). Se entenderá que entre más operaciones reestructuradas tenga un mismo deudor, mayor será el riesgo de no pago de la obligación.
- En la evaluación de la capacidad de pago de entidades públicas territoriales, las entidades vigiladas deberán verificar el cumplimiento de las

condiciones establecidas en las leyes 358 de 1997, 550 de 1999 y 617 de 2000, y de las demás normas que las reglamenten o modifiquen.

- Los posibles efectos de los riesgos financieros a los que está expuesto el flujo de caja del deudor y/o del proyecto a financiar, considerando distintos escenarios en función de variables económicas (tasas de interés, tasas de cambio, crecimiento de los mercados, etc.) que puedan afectar el negocio o la capacidad de pago del deudor, según el caso. Igualmente, se debe examinar la calidad de los flujos de caja teniendo en cuenta la volatilidad de los mismos. Dentro de estos riesgos se deben analizar, según resulten relevantes:

- Posibles “descalces” de monedas, plazos y tasas de interés en la estructura de balance y en operaciones fuera de balance, como por ejemplo la incidencia de los derivados financieros.

- Para aquellos créditos con tasa de interés variable o indexado a la UVR u otro índice, proyecciones y escenarios posibles de evolución de las cuotas según el comportamiento esperado de las tasas de interés, de la tasa de cambio, la inflación y otras variables que puedan afectar directamente el servicio de la deuda.

- Para créditos denominados en monedas extranjeras, el riesgo de mercado derivado de la volatilidad de las tasas de cambio respectivas y su posible impacto sobre la capacidad de pago del deudor. En el caso de operaciones de crédito con el exterior, análisis propios y del mercado sobre el riesgo del país en el cual está domiciliado el deudor, con el objeto de identificar los riesgos de transferencia y convertibilidad de las divisas requeridas para atender el crédito.

- Los riesgos de contagio, legales, operacionales y estratégicos a los que puede estar expuesta la capacidad de pago del deudor o el proyecto a financiar. En este contexto es necesario evaluar, entre otros, la información relacionada con el conglomerado económico al que pertenece el deudor.

d. Garantías que respaldan la operación: se entiende como garantía aquel respaldo que satisface el pago o cumplimiento de la obligación adquirida. La Superintendencia Financiera de Colombia (1995) estableció que es necesario conocer el valor de las garantías que respaldan la operación con el fin de calcular las pérdidas esperadas y las provisiones. Para ello se hace fundamental conocer los siguientes factores relacionados con la garantía: naturaleza, valor, cobertura y liquidez de la misma.

1.5. Crédito de consumo en Colombia

Los créditos de consumo son aquellos que se ofrecen a personas naturales con el fin de financiar la adquisición de bienes o servicios. Existen varios tipos de crédito de consumo en Colombia:

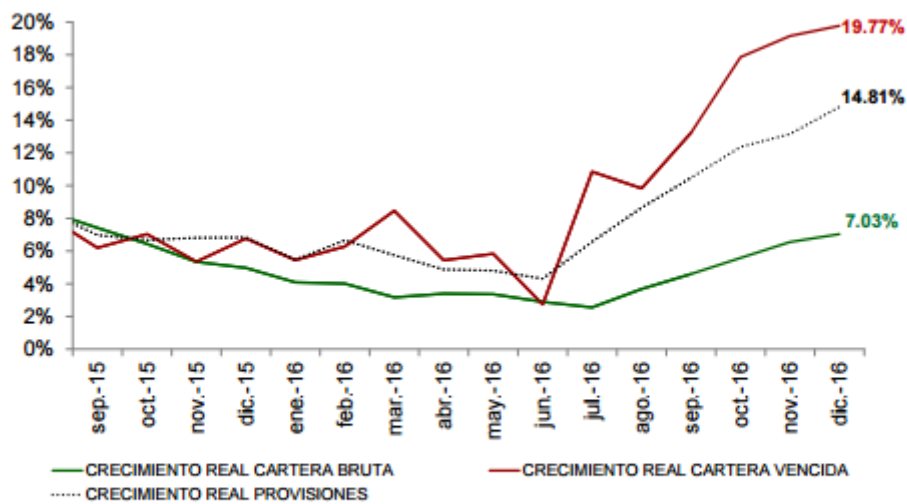
- **Crédito de consumo de libre inversión:** es un producto mediante el cual la entidad financiera otorga un crédito de libre destinación a un cliente persona natural.
- **Crédito de consumo rotativo:** es una línea de crédito de libre inversión que permite que, a medida que se vayan cancelando las cuotas, el cupo se libere para ser utilizado de nuevo.

- **Crédito de libranza:** es un mecanismo de financiación para clientes empleados o pensionados, que autorizan a su empleador o a su fondo de pensiones para que se deduzca de su nómina o mesada pensional las cuotas correspondientes al préstamo otorgado y se les transfieran los recursos a la entidad financiera.
- **Crédito de vehículo:** es un producto mediante el que la entidad financia la adquisición de vehículos nuevos o usados no generadores de renta a sus clientes personas naturales.
- **Crédito de consumo para compra de cartera:** financiación a personas naturales para la cancelación de sus deudas con otras entidades financieras.

Según el informe de actualidad del sistema financiero colombiano (Superintendencia Financiera de Colombia, 2016), la cartera de consumo presentó un incremento real del 7,03% durante el año 2016 en comparación con el año inmediatamente anterior, lo que sitúa la cartera con un saldo de \$113,4 billones. Por su parte, el crecimiento de la cartera de vivienda fue de 6,70%, con un saldo de \$55,5 billones de pesos. A su vez, la cartera de consumo representó un 27,5% del total de la cartera de créditos.

En cuanto a la tasa de interés promedio, fue del orden de 24,9% efectiva anual.

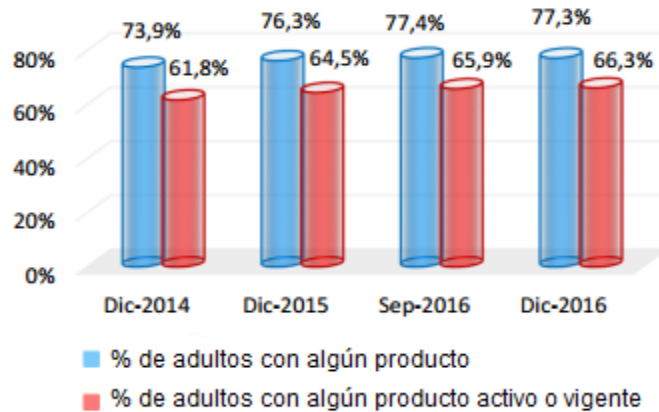
Gráfica 1. Cartera de consumo en Colombia



Fuente: Superintendencia Financiera de Colombia (2016, p. 9)

En cuanto al índice de inclusión financiera del año 2016, se presentó un incremento del 3% con respecto al año anterior y fue de 77,3% de adultos con algún producto financiero, es decir, 25,7 millones de adultos, y 66,3% adultos (aumento del 4,4%) con algún producto financiero activo o vigente, lo que se traduce en 22 millones de adultos, según el reporte trimestral de inclusión financiera (Banca de las Oportunidades, 2016). Lo anterior significa que cerca de 7,6 millones de adultos en Colombia no cuentan con productos financieros formales.

Gráfica 2. Porcentaje de adultos colombianos con algún producto con inclusión de todas las entidades que otorgan créditos



Fuente: Banca de las Oportunidades (2016, p. 3)

Por rango de edad, los usuarios activos de productos financieros son en su mayoría adultos entre 26 y 40 años (37,6%) y entre 41 y 65 años (41,1%). Los jóvenes de 18 a 25 años representan el 13,3%, mientras que los adultos mayores de 65 años solo el 7,9% (Banca de las Oportunidades, 2016).

1.6. Modelos de scoring

Según Gutiérrez Girault (2007), los modelos de scoring de crédito se comenzaron a implementar en los años setenta; sin embargo, esta práctica no se generalizó sino hasta los años noventa debido a los avances tecnológicos y estadísticos.

El objetivo de estos modelos se centra en asociar una probabilidad de incumplimiento al solicitante, con base en una ponderación de sus características cualitativas y cuantitativas, así como en el comportamiento de pagos que ha

presentado en obligaciones anteriores (canceladas) o actuales y vigentes al tener en cuenta las condiciones del crédito solicitado. Dicha evaluación permite disminuir el riesgo de incumplimiento de la deuda mediante un otorgamiento acertado porque de esta manera se le evita a la entidad financiera incurrir en pérdidas provenientes de la mayor provisión de los créditos colocados.

Tal y como lo describe Gutiérrez Girault (2007),

Los métodos o modelos de credit scoring, a veces denominados score-cards o classifiers, son algoritmos que de manera automática evalúan el riesgo de crédito de un solicitante de financiamiento o de alguien que ya es cliente de la entidad. Tienen una dimensión individual, ya que se enfocan en el riesgo de incumplimiento del individuo o empresa, independientemente de lo que ocurra con el resto de la cartera de préstamos (p. 5).

Por su parte, Puertas Medina y Martí Selva (2013), describen los credit scoring como:

Todo sistema de evaluación crediticia que permite valorar de forma automática el riesgo asociado a cada solicitud de crédito. Riesgo que estaría en función de la solvencia del deudor, del tipo de crédito, de los plazos, y de otras características propias del cliente y de la operación, que van a definir cada observación, es decir, cada solicitud de crédito (p. 304).

Se toman como definiciones de scoring las mencionadas en los párrafos anteriores toda vez que se complementan.

Para la construcción de dichos modelos se utilizan diversas técnicas estadísticas como las que se detallan a continuación.

1.6.1. Análisis discriminante

El análisis discriminante permite determinar las características diferenciadoras de los miembros de un grupo u otro.

La variable dependiente representará la pertenencia a los grupos y tendrá tantos valores discretos como grupos. Las variables en las que se diferencian los grupos se conocen como variables independientes, de clasificación o discriminantes. Estas variables deben ser cuantitativas continuas o admitir un tratamiento numérico con significado (Universidad Complutense de Madrid, Departamento de Sociología IV, 2017).

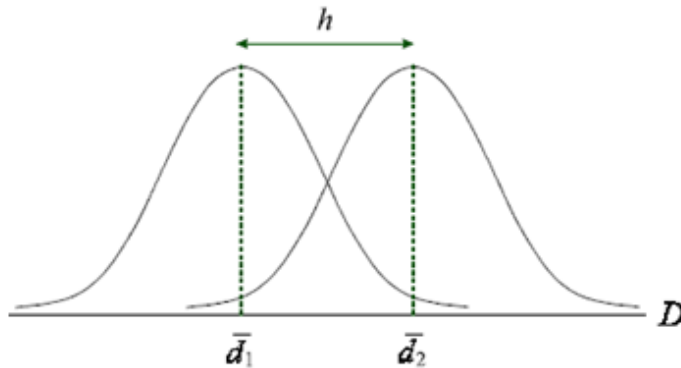
El análisis discriminante tiene dos objetivos: predictivo y descriptivo. El primero es aquel que permite llevar a cabo la diferenciación o clasificación de los casos en los distintos grupos. El segundo corresponde al que permite establecer las características que diversifican a cada grupo (Pérez Ramírez y Támara Ayús, 2012). Así mismo, este análisis proporciona una función lineal con base en la información contenida en las variables independientes que permite alcanzar los objetivos mencionados. Para el caso de dos grupos se tiene una combinación lineal de la siguiente forma (Universidad Complutense de Madrid, Departamento de Sociología IV, 2017):

Ecuación 3. Función discriminante

$$D = b_1X_1 + b_2X_2$$

Donde X_1 y X_2 representan los grupos y b_1 y b_2 son las ponderaciones de las variables independientes que consiguen hacer que los sujetos de uno de los grupos obtengan puntuaciones máximas en D y los sujetos del otro grupo puntuaciones mínimas. A continuación se muestra una gráfica de la función discriminante; los grupos corresponden a los histogramas y las líneas punteadas a las proyecciones de los centroides (Universidad Complutense de Madrid, Departamento de Sociología IV, 2017).

Gráfica 3. Histogramas de cada grupo y centroides representados en la función discriminante



Fuente: Universidad Complutense de Madrid, Departamento de Sociología IV (2017, p. 4)

Los centroides de cada grupo (X_1 y X_2) están representados por \bar{d}_1 y \bar{d}_2 , cuyas ecuaciones se obtienen al sustituir el valor de las medias de cada grupo en las variables X_1 y X_2 . Así (Universidad Complutense de Madrid, Departamento de Sociología IV, 2017):

Ecuación 4. Centroide del grupo 1:

$$\bar{d}_1 = b_1 \bar{x}_1^{(1)} + b_2 \bar{x}_2^{(1)}$$

Ecuación 5. Centroide del grupo 2:

$$\bar{d}_2 = b_1 \bar{x}_1^{(2)} + b_2 \bar{x}_2^{(2)}$$

La distancia entre los dos centroides (h) debe ser máxima para que de esta manera los grupos estén lo más distantes que sea posible (Universidad Complutense de Madrid, Departamento de Sociología IV, 2017).

Ecuación 6. Distancia entre centroides:

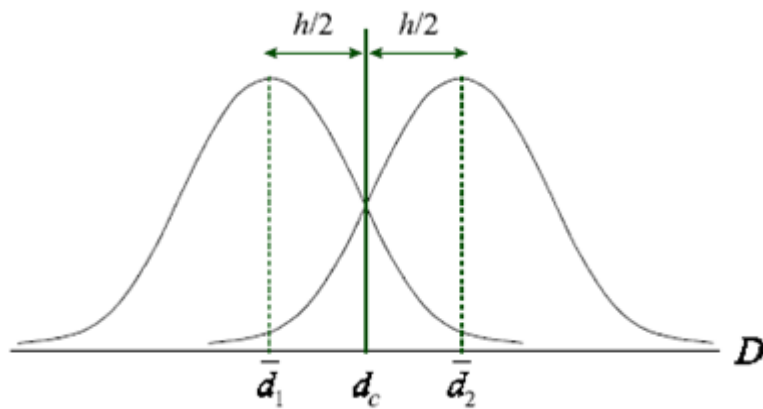
$$h = \bar{d}_1 - \bar{d}_2$$

Ahora bien, una vez obtenida la función discriminante es recomendable clasificar los casos en los que se estableció dicha función para así comprobar el grado de asertividad.

Según la Universidad Complutense de Madrid, Departamento de Sociología IV (2017), existen varias reglas de clasificación:

- a. Utilización de un punto de corte equidistante de ambos centroides. Esta regla consiste en establecer un punto equidistante entre los centroides (d_c). De esta manera, los casos cuyas puntuaciones sean mayores al punto de corte se asignarán al grupo superior pero si sus puntuaciones son inferiores se adjudicarán al inferior.

Gráfica 4. Utilización de un punto de corte equidistante de ambos centroides



Fuente: Universidad Complutense de Madrid, Departamento de Sociología IV (2017, p. 40).

La principal limitación de esta regla consiste en que solo se puede utilizar cuando se tienen dos grupos. De igual manera, en caso de que los grupos sean de distinto tamaño, el punto de corte (d_c) se debe calcular con la distancia ponderada.

b. La segunda regla está basada en la teoría bayesiana. Según Alonso Cánovas y Tubau Sala (2002), la probabilidad se define mediante cuatro axiomas:

(1) la probabilidad de un suceso es un número mayor o igual que cero, (2) la probabilidad del suceso seguro es igual a 1, (3) la probabilidad de la unión de dos sucesos incompatibles es igual a la suma de las probabilidades de cada uno de ellos (p. 27).

Y, por último, “si A y B son dos sucesos, entonces la probabilidad del suceso A , supuesto que se cumpla B (probabilidad de A condicionado a B), se representa así $p(A|B)$ ” (p. 28).

Ecuación 7. Probabilidad de A condicionado a B

$$p(A|B) = p(A \text{ y } B) / p(B)$$

Según Alonso Cánovas y Tubau Sala (2002), el teorema de Bayes consiste en una fórmula a partir de la cual es posible conocer una probabilidad condicionada a otras probabilidades (probabilidad a posteriori):

Ecuación 8. Teorema de Bayes (probabilidad de la hipótesis H_i dados los datos D)

$$p(H_i|D) = \frac{p(H_i)p(D|H_i)}{p(D)}$$

Esto es, la probabilidad de ocurrencia de una hipótesis H_i dada la ocurrencia de los datos D .

En caso de existir n hipótesis (o grupos) globalmente exhaustivas y mutuamente excluyentes la fórmula es:

Ecuación 9. Probabilidad de ocurrencia de los datos, dadas n hipótesis

$$p(D) = p(H_1)p(D|H_1) + p(H_2)p(D|H_2) + \dots + p(H_n)p(D|H_n)$$

Al reemplazar en la fórmula de Bayes se obtiene:

Ecuación 10. Teorema de Bayes (probabilidad de la hipótesis H_i dada la ocurrencia de D , con n hipótesis)

$$p(H_i|D) = \frac{p(H_i)p(D|H_i)}{p(H_1)p(D|H_1) + p(H_2)p(D|H_2) + \dots + p(H_n)p(D|H_n)}$$

1.6.2. Modelo probabilístico

Modelo de decisión que permite procesar variables aleatorias y asociar una probabilidad con cada uno de los valores que pueden tomar dichas variables. La respuesta de este modelo constituye una respuesta o resultado binario, es decir, la variable dependiente solo toma dos valores 0 o 1 para indicar éxito o fracaso, lo que se conoce como variable dicotómica (Greene, 2003). Este modelo utiliza una función de distribución normal estándar.

El modelo fue desarrollado por Robertson y Sparck Jones a finales de los años setenta. Según Blázquez Ochando (2012), su funcionamiento se basa en la representación binaria de los documentos en los que se lleva a cabo una consulta de términos.

El resultado de la consulta puede ser 1 en caso que el término se encuentre en el conjunto de documentos o 0 en caso de que no. Se consideran las siguientes premisas:

- Los documentos se clasifican en relevantes e irrelevantes.

- La respuesta obtenida del conjunto de documentos relevantes se denomina conjunto de respuesta ideal.
- Se denomina consulta ideal a aquella que proporciona el conjunto de respuesta ideal.

Los errores del modelo probabilístico se distribuyen de manera normal, con media cero y desviación estándar de uno.

Según Enchautegui (2003, p. 4), “en el modelo probabilístico la variable discreta Y proviene de una variable continua Y^* , así”:

Ecuación 11. Función de la variable continua Y^* (modelo probit)

$$Y^* = \beta_0 X_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + \mu$$

$$Y^* = \sum_{k=0}^k \beta_k X_k + \mu$$

Donde X son variables explicativas, $X_0 = 1$ representa la constante, y μ el término de error. Y^* no es observada; no hay datos para Y^* . Lo que se observa es una variable dicótoma Y que asume un valor de 0 o de 1 dependiendo de si la variable latente Y^* cruza o no un límite.” (Enchautegui, 2003, p. 4).

En este orden de ideas, Y tomará el valor de uno si:

$$Y^* = \sum_{k=0}^k \beta_k X_k + \mu > 0$$

$$= \mu \geq - \sum_{K=0}^k \beta_k X_k$$

Y tomará el valor de cero sí:

$$Y^* = \sum_{K=0}^k \beta_k X_k + \mu \leq 0$$

$$= \mu \leq - \sum_{K=0}^k \beta_k X_k$$

1.6.3. Modelo logístico

Al igual que el modelo probabilístico, el modelo logístico provee un resultado binario, es decir, la variable dependiente asume el valor de 0 o de 1. Es un modelo de predicción de probabilidad de ocurrencia de una variable dicotómica categórica. La regresión logit utiliza una función de distribución logística.

El modelo se basa en la siguiente función:

Ecuación 12. Función de la variable Y (modelo logit)

$$Y = \frac{1}{1 + e^{-z}} + \mu$$

Donde z : scoring logístico ($z = \beta_0 X_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k$) (Fernández Castaño y Pérez Ramírez, 2005).

La función logística tiene colas más anchas que la probabilística.

1.6.4. Redes neuronales artificiales

El modelo de redes neuronales artificiales (ANN, según las iniciales de la expresión completa en inglés) es un sistema de computación que se fundamenta en el funcionamiento de las redes neuronales biológicas del cerebro humano, de las que imita su comportamiento (Basogain Olabe, 1998).

Según el citado autor, las ANN presentan tres características propias del cerebro humano:

- **Aprender:** las ANN adquieren su conocimiento de la experiencia, que se desarrolla por medio de ejercicios en los que se presentan entradas y salidas deseadas con el fin de lograr que el modelo encuentre la relación entre las mismas.
- **Generalizar:** las ANN permiten ofrecer salidas o respuestas correctas dentro de un margen.
- **Abstraer:** estas redes son capaces de identificar los aspectos o cualidades comunes de un conjunto de variables de entrada.

Funcionamiento y elementos de las redes neuronales artificiales

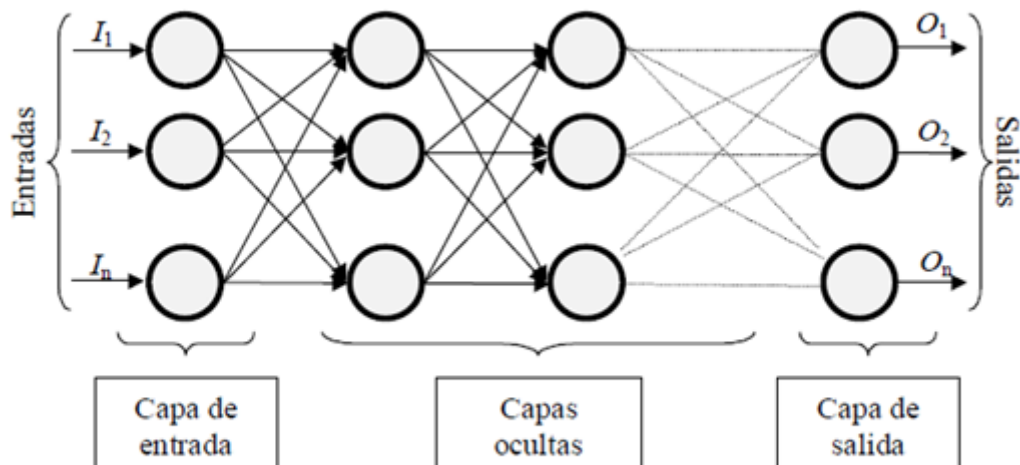
De acuerdo con Matich (2001), el modelo de redes neuronales artificiales se diseña con base en el sistema neuronal humano; sin embargo, su alcance es limitado si se tiene en cuenta la complejidad del sistema nervioso humano. Las neuronas artificiales funcionan de la siguiente manera:

- a.** Reciben un conjunto de variables que ingresan por la “capa de entrada”.
- b.** Las entradas se multiplican por su ponderación.
- c.** Dichas ponderaciones se suman.

Los pasos 2. y 3. se llevan a cabo en las “capas ocultas”, en las que se encuentran las neuronas interconectadas que ejecutan el procesamiento de la información.

- d.** Se genera una respuesta en la “capa de salida”.

Gráfica 5. Esquema de una red neuronal



Fuente: Matich (2001, p. 12)

Entrenamiento de una red neuronal

Al considerar la capacidad de aprender de las redes neuronales a través de la experiencia, deben ser entrenadas con anticipación a través de ejemplos en los que se incluyan entradas y sus respectivas salidas deseadas, con el fin de lograr de esta manera un ajuste en las ponderaciones que permitan alcanzar el resultado requerido (Basogain Olabe, 1998).

Según Matich (2001), existen dos tipos de entrenamiento:

- **Supervisado:** consiste en un proceso de acondicionamiento controlado por un agente externo, que verifica que la salida de la red coincida con la salida que se quiere. En caso de que no corresponda se procede a ajustar los pesos de las conexiones.
- **No supervisado:** este tipo de entrenamiento no involucra un agente externo, por lo que la red no recibe información de su entorno distinta a la

entrada de datos, de modo que la red se encarga de abstraer familiaridades o características similares entre los datos de entrada.

Aspectos metodológicos

Si se considera lo establecido por la Superintendencia Financiera de Colombia en cuanto al Sistema de Administración de Riesgo Crediticio (SARC), se hace necesario para las entidades diseñar un modelo para el otorgamiento y el seguimiento de crédito mediante la utilización de herramientas estadísticas que faciliten el cumplimiento de dicho objetivo. En esta fase del trabajo investigativo se define el proceso de desarrollo de modelo de scoring propuesto para la entidad financiera en estudio, de la que no se revelará la razón social por motivos de seguridad de la información. La implementación de dicho modelo quedará a discrecionalidad de la entidad financiera.

La investigación que se llevó a cabo en el presente trabajo fue de tipo cualitativo, toda vez que para su realización se planteó un problema de estudio concreto seguido de una serie de objetivos que permitieran la solución del mismo. Para alcanzar dichos objetivos se hizo una revisión de la literatura que se constituyó en el marco teórico de la investigación, se tomaron variables cuantitativas y cualitativas de deudores (datos suministrados por la entidad financiera), que, luego de ser sometidas a rigurosas pruebas de correlación estadística, lograron determinar la relación existente entre la variable en análisis y la probabilidad de que el deudor presentara incumplimiento en la atención a sus obligaciones con la entidad financiera en un lapso de 12 meses.

Para la elaboración de este trabajo de grado se inició con la construcción del marco teórico con base en las múltiples referencias bibliográficas consultadas de bases de datos electrónicas de la Universidad EAFIT a través de la herramienta *Descubridor*, las que, además de integrar bases de datos bibliográficas y repositorios institucionales de la universidad, son recursos de acceso abierto para los estudiantes, lo que facilita la investigación. Se utilizó la herramienta de *Google*

Académico, que facilita la búsqueda de tesis, libros, resúmenes y artículos que tienen como fuente organizaciones académicas. Y, por último, se consultó en la base de datos *Scientific Electronic Library Online* o *SciELO*, que permite el acceso a ediciones completas de revistas científicas. De esta manera se estableció el marco teórico sobre el que se apoyó la investigación y que proporcionó las bases para escoger entender el riesgo de crédito y su importancia y para seleccionar el mejor modelo estadístico para el diseño de scoring de crédito.

Para cumplir los objetivos planteados fue necesario contar con la colaboración de una entidad financiera del valle de Aburrá, especializada en la colocación de créditos de consumo en dicha zona. Dicho intermediario bancario facilitó el acceso a la base de datos de deudores, previo acuerdo de confidencialidad entre los autores del trabajo de grado y la cooperativa de crédito, con una fecha de corte específica, y diversa información de los deudores (sin incluir nombres y números de identificación), que, según el criterio de los administradores de la entidad, pudiese tener significancia para explicar la no atención a las obligaciones por parte de los clientes. Se continuó, por tanto, con el análisis de los datos suministrados descritos a continuación:

1.7. Descripción de la base de datos

La base de datos sobre la que se hizo la modelación fue suministrada por la entidad financiera de la que se hace mención en el párrafo anterior. Dicha base de datos cuenta con un total de 24.786 clientes con fecha de corte de diciembre de 2014. Para cada cliente se identificaron 28 variables entre cuantitativas y cualitativas.

La variable dependiente se denomina incumplimiento, que se considera como un indicador de que el cliente sobrepasó un número de días preestablecido por la norma y en el que la entidad financiera argumenta que el deudor entra en mora o en incumplimiento de sus obligaciones con el crédito otorgado. Se entiende como

incumplimiento en los créditos de consumo el referente a aquellas obligaciones que se encuentren en mora mayor de 90 días (Superintendencia Financiera de Colombia, 1995).

Las variables que explican el modelo (independientes) son aquellas que determinan el comportamiento de la variable independiente. A continuación, en la tabla 1 se presentan las variables contenidas en la base de datos y se distinguen aquellas que fueron incluidas en el modelo:

Tabla 1. Variables incluidas en el modelo

Variable	Descripción	Incluida
Cédula	Identificador del cliente. No tiene afectación en el modelo	No
Oficina	Consiste en el lugar en el que se realizó la solicitud del crédito por parte del cliente. Hay un total de nueve oficinas distribuidas en el valle de Aburrá: oficinas 1, 2, 3, 4, 5, 6, 7, 8 y 9. La oficina a la que pertenezca el cliente tomará el valor de 1 en el modelo para cada caso y las demás el valor de cero	Sí
Categoría	Se refiere a la categoría de riesgo de cada cliente según su comportamiento de pago crediticio. Estas categorías se encuentran relacionadas en la circular externa 029 de mayo de 2007 de la Superintendencia Financiera de Colombia. Se clasifican en cinco categorías, a saber (Superintendencia Financiera de Colombia, 1995): <ul style="list-style-type: none"> - Categoría A o “riesgo normal”: créditos 	No

con una atención adecuada, toda vez que el deudor presenta una capacidad de pago apropiada.

- Categoría B o “riesgo aceptable, superior al normal”: créditos con una apropiada atención. Sin embargo, se evidencian debilidades que pueden afectar la capacidad de pago del deudor y posiblemente incurrir en futuros impagos

- Categoría C o “riesgo apreciable”: créditos en los que se evidencian insuficiencias en la capacidad de pago del deudor

- Categoría D o “riesgo significativo”: la probabilidad de recaudo del crédito es altamente dudosa al considerar las deficiencias en mayor grado del deudor

- Categoría E o “riesgo de incobrabilidad”: es aquel crédito que se cataloga como incobrable

Esta variable no afecta el modelo

Tiempo de maduración	Es el tiempo, medido en meses, en que el cliente ha tenido su crédito vigente. Este valor puede ser superior al plazo del crédito otorgado cuando el cliente entra en mora	Sí
Monto	Consiste en el valor inicial de dinero otorgado al cliente en calidad de préstamo. Varía entre \$101.000 y \$61.800.000 (en pesos colombianos)	Sí
Saldo	Es el valor adeudado por el cliente a la fecha de corte de la base de datos. Varía entre \$19 y	Sí

\$56.549.783 (en pesos colombianos)

Plazo	Es el tiempo, en meses, concedido a cada cliente para el pago del préstamo. Sus valores oscilan entre 2 y 180	Sí
Tasa de interés	Corresponde a la tasa de interés mensual vencida aplicada a cada crédito. Esta variable toma valores desde 0,16% hasta 3,30%	Sí
Cuota	Es el valor de la cuota del crédito expresada en pesos colombianos	Sí
Garantía	Consiste en el respaldo ofrecido por el deudor a la entidad financiera. Se divide en garantía personal (avalista o codeudor) y garantía real (bien inmueble, por ejemplo). Tomará el valor de uno en el modelo si corresponde a garantía real. De lo contrario, asumirá el valor de cero	Sí
Días de mora	Es el número de días de retraso de los pagos que presenta cada cliente. Para la base de datos en estudio oscila entre 0 y 1.784 días calendario. Esta variable no tiene afectación en el modelo	No
Forma de pago	Consiste en la forma en la que se efectuarán los pagos de las cuotas del crédito. Puede ser por ventanilla, es decir, el deudor le paga en forma directa a la entidad financiera, o por nómina, lo que significa que cada mes se descontará del sueldo del deudor la cuota del crédito que será pagada a la entidad por medio de la empresa en la que labora el deudor. Esta variable no tiene afectación en el modelo	Sí

Reestructurado	<p>Consiste en la modificación de las condiciones iniciales del crédito con el fin de asegurar el pago del mismo por parte del deudor. Esta variable solo puede tomar dos valores: reestructurado y no reestructurado</p> <p>La variable tomará el valor de uno en el modelo si el crédito fue reestructurado. De lo contrario, asumirá el valor de cero</p>	Sí
Edad	<p>Es el número de años que tiene el deudor. Varía entre 19 y 83</p>	Sí
Ocupación	<p>Describe la actividad a la que se dedica el prestatario. Las categorías son:</p> <ul style="list-style-type: none"> - Ama de casa - Desempleado - Empleado - Estudiante - Independiente - Jubilado - Pensionado <p>La ocupación del cliente tomará el valor de uno en el modelo para cada caso y las demás el valor de cero</p>	Sí
Nivel educativo	<p>Es el grado de escolaridad del deudor.</p> <ul style="list-style-type: none"> - Ninguno - Primaria - Bachillerato - Técnico - Tecnológico - Universitario 	Sí

	<ul style="list-style-type: none"> - Posgrado <p>El nivel educativo del cliente tomará el valor de uno en el modelo y los demás asumirán el valor de cero</p>	
Ingreso total	Equivale a los ingresos mensuales recibidos por el deudor. Este valor fluctúa entre 0 y \$15.100.000 (en pesos colombianos)	Sí
Egreso total	Son los gastos mensuales en los que incurre el deudor. Este valor fluctúa entre 0 y \$11.800.000 (en pesos colombianos)	Sí
Estrato socioeconómico	<p>Corresponde al estrato socioeconómico del deudor. Se divide en:</p> <ul style="list-style-type: none"> - Estrato 1: bajo bajo - Estrato 2: bajo - Estrato 3: medio bajo - Estrato 4: medio - Estrato 5: medio alto - Estrato 6: alto <p>El estrato al que pertenezca el cliente tomará el valor de uno en el modelo y los demás permanecerán en cero</p>	Sí
Antigüedad laboral	Consiste en el tiempo que lleva el deudor vinculado como empleado de una empresa. Se mide en años y varía entre 0 y 46	Sí
Estado civil	<p>Se divide en:</p> <ul style="list-style-type: none"> - Casado - Eclesiástico - Separado 	Sí

- Soltero
- Unión libre
- Viudo

La variable que corresponda al estado civil del cliente tomará el valor de uno en el modelo y las demás permanecerán en cero

Sexo	<p>Se refiere al género del deudor, es decir, masculino o femenino</p> <p>Esta variable tomará el valor de uno en el modelo en caso de ser masculino</p>	Sí
Personas a cargo	<p>Es el número de personas que dependen económicamente del deudor. Oscila entre 0 y 25</p>	Sí
Tipo de vivienda	<p>Es el tipo de vivienda en la que vive el deudor. Se tienen las siguientes opciones:</p> <ul style="list-style-type: none"> - Arrendada - Familiar - Propia - Ninguna de las anteriores (<i>null</i>) <p>La variable que corresponda al tipo de vivienda del cliente tomará el valor de uno en el modelo las demás permanecerán en cero</p>	Sí
Tipo de contrato	<p>Contrato laboral del deudor:</p> <ul style="list-style-type: none"> - Término definido o fijo - Término indefinido - Prestación de servicios - Jubilado - <i>Null</i>, para aquellos casos diferente a empleado y jubilado 	Sí

	La variable que corresponda al tipo de contrato del cliente asumirá el valor de uno en el modelo y las demás serán iguales a cero	
Aportes	Hace referencia al promedio de ahorro de los clientes en los productos de captaciones ofrecidos por la entidad No afecta el modelo	No
Número de créditos	Corresponde al número de créditos que tiene o ha tenido el cliente con la entidad Esta variable no tiene incidencia sobre el modelo	No
Antigüedad en la entidad	Es el número de años en el que el deudor ha estado vinculado como cliente a la entidad. Fluctúa entre 0 y 49	Sí
Incumplimiento	Indicador de si el cliente entró en impago o no	Variable de respuesta

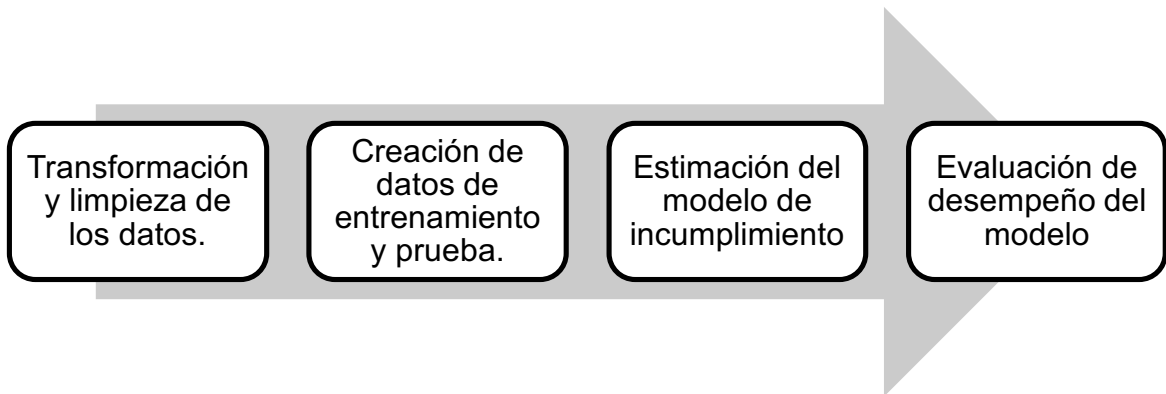
Fuente: elaboración propia

Por último se llevó a cabo la modelación de la base de datos descrita en el siguiente apartado.

1.8. Metodología de modelación

Con la información obtenida se procedió más tarde a realizar las siguientes labores ilustradas en la figura 1.

Figura1. Proceso metodológico



Fuente: elaboración propia

a. Transformación y limpieza de los datos: en esta fase se conocieron de primera mano las variables con las que se realizaría el trabajo de grado, su significado y la forma en la que habían sido capturadas por la entidad financiera, el cálculo de las mismas –para aquellas que aplicara–, y la utilidad que para el intermediario financiera representaban. Se realizó una depuración de la información en la que se presentaban datos atípicos, valores nulos o vacíos en cuanto a valores que correspondan a su variable explicativa. Este paso garantizó que la masa de los datos que se ingresó al modelo estuviera limpia y fuera apta para estimar un modelo.

b. Creación de datos de entrenamiento y prueba: en esta etapa se definió el modelo más adecuado para dar cumplimiento a los objetivos planteados al considerar las necesidades de la entidad financiera sobre la que se apoyó la investigación. Luego se ejecutó una estrategia de validación cruzada, en la que, a partir de toda la base de datos, se escogió un porcentaje del 80% para realizar el entrenamiento del modelo (esto es, estimar los coeficientes de la regresión) y el 20% restante para llevar a

cabo la estimación de incumplimiento con el modelo entrenado y evaluar qué tan acertada fue la predicción de la regresión logística obtenida.

c. Estimación del modelo de incumplimiento: una vez se obtuvo el 80% de la base de datos limpia y lista para modelar, se utilizó el lenguaje estadístico *R* para realizar las operaciones matemáticas y estadísticas necesarias para ejecutar la estimación del mejor modelo de regresión que representara la relación entre las variables explicativas y la variable de respuesta de incumplimiento. De igual manera, se analizaron los resultados, así como la importancia de las variables.

d. Evaluación del desempeño del modelo: una vez estimado el modelo se procedió a evaluarlo a través de una matriz de confusión o cruzada³ en la que se observaron las cantidades de aciertos y errores que tuvo el modelo al pronosticar la probabilidad de incumplimiento del 20% restante de la base de datos. De igual forma, se aplicó una prueba de Kolmogorov-Smirnov o prueba K-S que permite inferir acerca de la similitud entre la distribución real de los datos de incumplimiento reales y la distribución del incumplimiento pronosticado por el modelo.

Presentación y análisis de resultados

En esta sección del documento se presentan los resultados obtenidos de la investigación en cuanto a la elección del modelo que mejor se ajusta al perfil de riesgo de los prestamistas de la entidad financiera sobre la que se apoya el trabajo de grado, los coeficientes estimados que permitieron definir el modelo o la fórmula matemática del scoring, la definición de las variables más significativas y, por

³ Matriz de confusión o matriz cruzada: herramienta que permite evaluar el desempeño del modelo basado en aprendizaje. Cada columna de la matriz representa el número de predicciones y cada fila los casos reales. De esta manera se observa en cuántos incurre el modelo.

último, la evaluación de asertividad del modelo a través de una validación cruzada y de una prueba de bondad de ajuste (la de Kolmogorov-Smirnov).

1.9. Elección del mejor modelo

Como antes se mencionó, es obligación de las entidades financieras establecer un modelo de otorgamiento de crédito que les permita clasificar y calificar a los potenciales sujetos de crédito con el fin último de minimizar la probabilidad de que dichos sujetos entren en incumplimiento.

Ahora bien, con base en lo establecido en el modelo de referencia para la cartera de consumo utilizado por la Superintendencia Financiera de Colombia (1995) en el anexo 5 de la circular 100 (básica, contable y financiera), se presenta un modelo de regresión logística que otorga un puntaje basado en las características de cada deudor, definido según la siguiente fórmula:

Ecuación 13. Puntaje del deudor

$$Puntaje = \frac{1}{1 + e^{-z}}$$

Donde z representa las variables que determinan el grado de riesgo de cada cliente y las ponderaciones de cada una. Para el presente trabajo investigativo se consideró la totalidad de las variables y se descartaron aquellas cuya significancia fue baja para el modelo.

1.9.1. Regresión logística

La regresión logística es un tipo de análisis de regresión utilizado para predecir el resultado de una variable categórica en función de las variables independientes o predictoras. Es útil para modelar la probabilidad de un evento que ocurre como función de otros factores. La regresión logística analiza datos distribuidos en forma binomial, es decir, que pueden tomar valores únicamente de 0 o 1 (bueno o malo, enfermo o sano, sube o baja, entre otras posibilidades).

Los modelos logit modelan las probabilidades binomiales desconocidas como una función lineal de los factores o variables independientes denominados x_i .

La formulación matemática del respectivo modelo es:

Ecuación 14. Probabilidad logística

$$\text{logit}(p_i) = \ln\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 x_{1,i} + \dots + \beta_k x_{k,i}$$

Que es equivalente a:

$$p_i = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_{1,i} + \dots + \beta_k x_{k,i})}}$$

Esta forma funcional por lo común se representa de la siguiente forma:

Ecuación 15. Regresión logística

$$Y = \frac{1}{1 + e^{-f(x)}}$$

donde $f(x)$ es una función analítica en x .

Estimar los coeficientes de la función $f(x) = \beta_0 + \beta_1 x_{1,i} + \dots + \beta_k x_{k,i}$ es el objetivo del algoritmo de aprendizaje, con el fin de que puede cuantificar el efecto de cada variable respecto a la variable predictora, que en este caso es la probabilidad de que el cliente entre en incumplimiento.

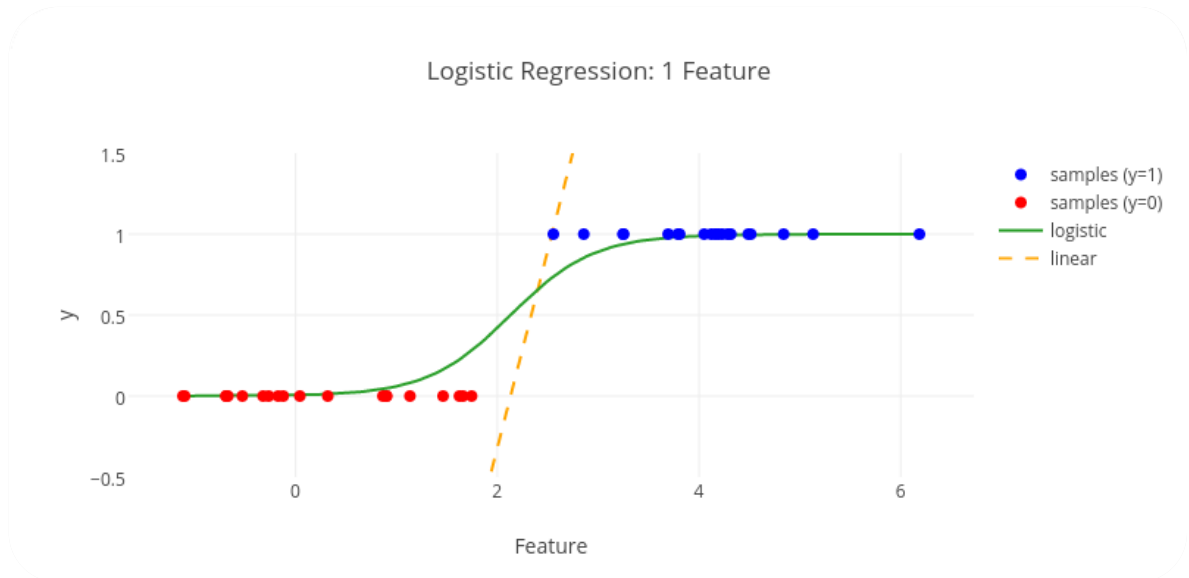
En el lenguaje estadístico *R* se utiliza la función *glm* de la biblioteca *caret*, que estima un modelo lineal generalizado aplicado para variables dicotómicas, es decir, un modelo logit. Para ello utiliza el método de mínimos cuadrados iterativos rebalanceados (IRLS), que se basa en resolver de manera iterativa un problema de la forma:

Ecuación 16. Problema de mínimos cuadrados lineales ponderados

$$\beta^{t+1} = \arg \min_{\beta} \sum_{i=1}^n w_i(\beta^t) |y_i - f_i(\beta)|^2$$

Este proceso se repite hasta que se alcance una convergencia en los β con un nivel de tolerancia dado.

Gráfica 6. Regresión logística



Fuente: elaboración propia en R

1.10. Coeficientes estimados y modelo de scoring

Una vez se realiza la estimación del modelo logístico, la convergencia de los parámetros se logra mediante el ajuste de los siguientes coeficientes con sus respectivos intervalos de confianza ilustrados en los anexos 1 y 2.

En el anexo 1 se observan los coeficientes que el modelo de regresión logística estima. Entre ellos se puede destacar la significancia de algunas de las variables. Es importante resaltar la importancia de las variables faciales del crédito (tiempo de maduración, monto, saldo, plazo, tasa, etc.).

De igual forma, es relevante notar que variables como el nivel educativo en primaria fue significativa al predecir el incumplimiento, así como el estado civil, el tipo de vivienda, el sexo y los egresos del deudor.

Por otro lado, cabe resaltar cómo las variables de la oficina en la que se solicitó el crédito no fueron significantes, así como el estrato al que pertenece el deudor, su ocupación o su tipo de contrato.

Por lo tanto, la fórmula del modelo diseñado para la entidad financiera en cuestión se puede escribir de la siguiente forma:

$$\begin{aligned}
Z = & -32.96639704 - 0.281539333 * Oficina2 - 0.074140452 * Oficina3 - 0.126837205 \\
& * Oficina4 - 0.463050236 * Oficina5 - 0.34119933 * Oficina6 - 0.490683307 \\
& * Oficina7 - 0.030087883 * Oficina8 + 0.508074694 * Oficina9 \\
& + 0.648597039 * TiempoMaduracion - 0.000005378 * Monto + 0.000005452 \\
& * Saldo - 0.162601975 * Plazo + 11.78988735 * Tasa + 0.000023539 * Cuota \\
& - 11.05038201 * GarantiaReal + 0.257058061 * Reestructurado \\
& + 0.002510371 * Edad + 0.920881721 * Desempleado - 0.914240768 \\
& * Empleado - 0.619927638 * Estudiante - 0.234375007 * Independiente \\
& - 1.399037093 * Jubilado - 0.344063076 * Pensionado - 0.260362306 \\
& * EducacionNinguna - 0.111850452 * EducacionPosgrado - 1.014133446 \\
& * EducacionPrimaria - 0.410761399 * EducacionTecnica - 0.348536631 \\
& * EducacionTecnologico - 0.328657962 * EducacionUniversitario \\
& + 0.000000126 * Ingresos - 0.00000151 * Egresos + 0.196025944 * Estrato2 \\
& + 0.138271406 * Estrato3 - 0.153405226 * Estrato4 + 0.072464242 \\
& * Estrato5 - 0.783930522 * Estrato6 - 0.032515884 * Antiguedad \\
& - 0.989877499 * Eclesiatistico + 1.233018769 * Separado + 0.323238392 \\
& * Soltero - 0.118456377 * UnionLibre + 0.823585746 * Viudo \\
& + 0.538918383 * Masculino - 0.012939372 * PersonasCargo - 0.386093463 \\
& * ViviendaFamiliar + 1.528882044 * ViviendaNULL - 0.758151315 \\
& * ViviendaPropia + 0.163683968 * ContratoNULL + 0.103413865 \\
& * ContratoServicios + 0.173584009 * ContratoDefinido + 0.513591197 \\
& * ContratoIndefinido + 0.02286393 * AntiguedadCooperativa
\end{aligned}$$

Donde para calcular la probabilidad de incumplimiento es:

$$P(\text{default}) = \frac{1}{1 + e^{-Z}}$$

Es importante aclarar que el resultado del modelo, tal como es la discriminación entre clientes con alta y baja probabilidad de incumplimiento, se da en el horizonte de un año.

Se resalta que el modelo de regresión logística parte de una base con las variables categóricas cuyo nivel de riesgo se encuentra inmerso en el intercepto, que son: oficina 1, ocupación ama de casa, nivel educativo bachillerato, estrato 1, estado civil casado, sexo femenino, tipo de vivienda arrendada y tipo de contrato jubilado.

El modelo propuesto permite otorgar una probabilidad de incumplimiento a los solicitantes de crédito de consumo con base en los perfiles de riesgo de la compañía de financiamiento sobre la que se desarrolló la presente investigación dado que los datos suministrados corresponden a la base de clientes actuales de la entidad.

1.11. Importancia de las variables

Una de las ventajas del modelo de regresión logística es que permite identificar la importancia de las variables al momento de explicar el incumplimiento o la variable de respuesta. Para los modelos lineales, como la regresión logística con mínimos cuadrados, se utiliza el valor absoluto del estadístico t para cada parámetro del modelo, junto con el valor absoluto de los coeficientes correspondientes al modelo estimado y la suma ponderada de los coeficientes de regresión absoluta, donde los pesos de la ponderación son una función de la reducción de las sumas de cuadrados a través del número de componentes de los mínimos cuadrados; por lo tanto, la contribución de los coeficientes se pondera en forma proporcional a la reducción de las sumas de cuadrados.

Entre las diez variables más importantes se definió que las siguientes fueron las que más influyeron a la hora de disminuir o aumentar la probabilidad:

Variables que aumentaron la probabilidad de incumplimiento

- a. **Tiempo de maduración:** un mayor tiempo de maduración de la deuda generó mayor incertidumbre acerca del pago oportuno de la misma.
- b. **Saldo del crédito:** se debe a que a mayor saldo que se tenga de un crédito fue más probable que el deudor no fuese capaz de cumplir su deuda.
- c. **Cuota del crédito:** de igual forma que el anterior, cuanto mayor sea la cuota fue más difícil cumplir la obligación para un deudor.
- d. **Estado civil (separado, viudo o casado):** hacen referencia a perfiles con mayores obligaciones financieras.
- e. **Egresos totales:** disminuyeron la capacidad de pago de la deuda.
- f. **Nivel educativo (primaria, bachillerato):** niveles educativos bajos tendieron a tener menores ingresos y, por tanto, menos capacidad de pago.
- g. **Género del deudor (masculino):** los hombres tuvieron mayor tendencia a ser peores impagos que las mujeres.
- h. **Antigüedad como empleado:** tiempo total que tenía un trabajador que le prestaba sus servicios a una empresa.

i. **Ocupación (desempleado):** al no tener ingresos, es de esperar que su capacidad de pago se viese comprometida.

Variables que disminuyeron la probabilidad de incumplimiento

a. **Monto del crédito:** a medida que el monto fue de menor cantidad, más fácil sería el pago de la deuda y, por lo tanto, su impago disminuyó.

b. **Plazo del crédito:** de igual forma que el anterior, a menor plazo existió mayor compromiso e incertidumbre en los pagos del crédito.

c. **Indicador de la garantía del crédito:** cuando se vio comprometida una garantía real del deudor para el pago la deuda, el deudor sintió mayor obligación de cumplir.

d. **Nivel educativo (profesional, posgrado):** niveles altos de educación tendieron a tener mejores ingresos y mayor estabilidad en sus pagos mensuales.

e. **Ingresos totales:** mejoraron la capacidad financiera del deudor.

f. **Género del deudor (femenino):** como antes se mencionó, las mujeres tuvieron un perfil de cumplir sus deudas a tiempo mejor que los hombres.

g. **Estado civil (unión libre):** perfil demográfico con menores obligaciones que casado, separado o viudo.

h. Ocupación (empleado, jubilado, pensionado, Independiente, etc.): fueron ocupaciones en las que el deudor devengó un ingreso y, por tanto, le otorgaron capacidad para el pago de su deuda.

1.12. Validación cruzada

Si se recuerda la metodología, para ajustar el modelo se utilizó el 80% de los datos, lo que corresponde a 19.828 registros de deudores. Los restantes 4.957 registros en la base de datos se emplearon para hacer validación cruzada y determinar la capacidad predictora del modelo. Este porcentaje se escogió como regla de oro para el entrenamiento de modelos basado en el principio 80/20 de Pareto⁴.

Es decir que, con el modelo antes estimado se calificaron los 4.957 deudores, se les calculó la probabilidad predicha por el modelo para incumplimiento y se comparó con la realidad si el deudor había entrado en incumplimiento o no. En este proceso se definió que el umbral de probabilidad para que un deudor se considerase en incumplimiento fue del 50%, es decir, que si el modelo predijo que la probabilidad del deudor de entrar en incumplimiento superaba este límite, se consideraría dicho deudor como un posible incumplimiento. Cabe aclarar que este parámetro es en cierta parte subjetivo, puesto que depende en gran medida del apetito de riesgo que la entidad defina en sus políticas.

⁴ El principio de Pareto recibe su nombre del economista Wilfredo Pareto tras su estudio de distribución de los ingresos en el que se estableció que la minoría de la población poseía la mayor parte de la riqueza. Este concepto se aplica a la calidad, por lo que se puede afirmar que si un problema tiene múltiples causas, el 20% de las causas resuelven el 80% del problema y el 80% restante de las causas resuelve el 20% restante del problema.

El proceso anterior arrojó la tabla 2, llamada matriz de confusión o de contingencia, que ayuda a ver qué tan bien se desempeñó el modelo en su fase predictiva.

De estas tablas se puede observar que, de todos los casos de predicción, el modelo pronosticó 4.862 deudores “buenos” cuando en efecto no presentaron incumplimiento. De igual forma, 54 pronósticos de incumplimiento resultaron ser ciertos, lo que indica que $4.862 + 54 = 4.916$ fueron los aciertos del modelo, que corresponde al 99,19% de las veces.

Por otro lado, el modelo predijo 36 casos en los que el deudor iba a ser bueno, pero entró en incumplimiento, y en cuatro casos pronosticó lo contrario: que iban a entrar en incumplimiento pero resultaron buenos. Lo anterior dio un total de 40 casos en los que modelo erró en su predicción, que corresponde a un 0,81% de las veces.

De acuerdo con lo que acaba de expresarse, se concluye que el modelo acertó en más del 99% de las veces para pronosticar un incumplimiento y, por lo tanto, se considera apto para predecir.

Tabla 2. Matriz de confusión en cantidades

		Predicción	
		Bueno	Malo
Real	Bueno	4.862	4
	Malo	36	54

Fuente: elaboración propia

Tabla 3. Matriz de confusión en porcentaje

		Predicción	
		Bueno	Malo
Real	Bueno	98,10%	0,08%
	Malo	0,73%	1,09%

Fuente: elaboración propia

1.13. Prueba de Kolmogorov-Smirnov

La prueba de Kolmogorov-Smirnov (prueba de K-S) determina la bondad de ajuste de dos distribuciones de probabilidad entre sí a partir de estadísticos no paramétricos. En esencia, la prueba de K-S realiza inferencias acerca de la distribución de unos datos en asemejarse a otra distribución. Para el caso de

aplicación, la prueba de K-S indica si las distribuciones de las predicciones de incumplimiento se asemejan, desde el punto de vista estadístico, a los datos reales.

Por lo tanto, se supone que la distribución de los datos reales se distribuye en forma binomial con parámetro p , lo que se define como:

$$\theta_{Reales} \sim Binomial(p)$$

La prueba de K-S lleva a cabo inferencias acerca de las siguientes hipótesis nula y alternativa:

$$H_0: \theta_{Logit} \sim \theta_{Reales}$$

$$H_a: \theta_{Logit} \neq \theta_{Reales}$$

Para el efecto, la prueba utiliza un estadístico no paramétrico calculado como:

Ecuación 17. Estadístico

$$D = \sup_{1 \leq i \leq n} |\widehat{F}_n(x_i) - F_0(x_i)|$$

Donde x_i es el i -ésimo valor observado en los datos del pronóstico, $\widehat{F}_n(x_i)$ es un estimador de la probabilidad de observar valores menores o iguales que x_i y $F_0(x_i)$ es la probabilidad de observar valores menores o iguales que x_i cuando H_0 es cierta, es decir, en los datos reales.

Si los valores del pronóstico $\widehat{F}_n(x_i)$ son similares a los datos reales $F_0(x_i)$, el valor de D será pequeño. Cuanto mayor sea la diferencia entre la distribución empírica y la real mayor será el valor de D .

Por lo tanto, si el valor de D es menor que un nivel de D_α , con un α de significancia estadística dada del 95% se acepta la hipótesis nula H_0 y de lo contrario se rechaza la hipótesis nula H_0 . De manera equivalente, si el valor p (la probabilidad de obtener un resultado al menos como el que en realidad se ha obtenido, si se supone que la hipótesis nula es cierta) es mayor que el nivel de significancia α se acepta la hipótesis nula y si el valor p es menor que α se rechaza la misma.

Al efectuar la prueba de K-S se obtuvieron los siguientes resultados:

Gráfica 7. Resultado de la prueba de K-S

```
Two-sample Kolmogorov-Smirnov test
data: Predicciones and Reales
D = 0.0061, p-value = 0.8
alternative hypothesis: two-sided
```

Fuente: elaboración propia en R

Se constata que el valor del D es igual a 0,0061, que es muy pequeño, lo que indica que los valores del pronóstico son similares a los datos reales. Por otro lado, el valor p de la prueba de K-S es igual a 0,8, que es mayor que el nivel de significancia estadística del 0,05 (5%), lo que conduce a que la hipótesis nula es cierta, de lo que se infiere que $H_0: \theta_{Logit} \sim \theta_{Reales}$, es decir, que los datos del modelo logístico replican en comportamiento estadístico a los datos reales de incumplimiento.

Conclusiones y recomendaciones

Los modelos de scoring son herramientas estadísticas que permiten predecir la probabilidad de incumplimiento de un deudor, con base en sus características cualitativas y cuantitativas y en información de las condiciones del crédito solicitado. Su uso fue generalizado en los años noventa y hoy en día forman parte fundamental del proceso de otorgamiento. Razón de lo anterior es el crecimiento constante en la cartera de crédito de consumo que conduce, a su vez, a una supervisión más estricta por parte de la Superintendencia Financiera de Colombia.

El modelo que se presenta en este trabajo buscó predecir, con base en la información histórica disponible de los clientes de crédito de consumo, tanto cuantitativa como cualitativa, si los mismos tenían una elevada probabilidad de caer en incumplimiento o no, lo que arrojó como resultado un modelo de 23 variables consideradas significativas a la hora de evaluar a un posible deudor. No obstante lo anterior, este modelo por el momento solo aplica a la entidad en la que se desarrolló el scoring, toda vez que la población puede sufrir variaciones en su comportamiento al iterar el modelo en otros intermediarios financieros.

Para la realización del presente trabajo fue fundamental la investigación acerca los conceptos básicos de riesgo de crédito, la modificación en la supervisión de las entidades financieras como consecuencia del Comité de Basilea que originó la creación del Sistema de Administración de Riesgo Crediticio (SARC) y, por supuesto, los modelos de scoring más utilizados en el sector financiero. Asimismo, se debe señalar la importancia de la metodología de modelación utilizada en la que se inició con el proceso de limpieza de la base de datos, se continuó con la elección del modelo más adecuado, el entrenamiento del mismo y por último se

obtuvo el scoring con sus respectivas pruebas de asertividad, con lo que se alcanzó el objetivo planteado.

Se analizaron en total cuatro técnicas estadísticas para modelos de scoring: el análisis discriminante, el modelo probabilístico, el modelo logístico y las redes neuronales artificiales. El modelo utilizado se escogió al tener en cuenta lo establecido en el modelo de referencia para la cartera de consumo utilizado por la Superintendencia Financiera de Colombia, en el que se presenta un modelo de regresión logística que otorga un puntaje basado en las características de cada deudor. El modelo de regresión logística permite predecir el comportamiento de una variable en función de otros factores.

El alcance del trabajo de grado es de tipo experimental, por lo que se considera adecuado en caso tal que la empresa en la que se desarrolló el scoring decida implementarlo en sus procesos de otorgamiento de cartera, de modo que su aplicación sea de forma gradual, con el propósito de que sea posible comparar en paralelo los resultados arrojados por el modelo en las diferentes solicitudes de crédito y las recomendaciones emitidas en igual sentido cuando se utiliza el proceso de otorgamiento tradicional. Adicionalmente, es recomendable un ajuste del scoring en el que se puedan observar las brechas entre el modelo tradicional y el aquí desarrollado, de tal forma que el segundo sea calibrado dependiendo de las necesidades de la entidad financiera. Se recomienda como mínimo un lapso de seis meses y hacer ajustes periódicos por lo menos una vez al año.

Al tener en cuenta que desde la perspectiva de la normatividad impuesta por la Superintendencia Financiera de Colombia el proceso de seguimiento y control toma gran cantidad de las variables utilizadas en la etapa de otorgamiento de créditos, en una segunda fase del scoring aquí desarrollado se podría buscar la

manera de crearle algunas modificaciones para que se utilice en ambos procesos de la línea misional de riesgo de crédito.

Si bien no hace parte del alcance de este trabajo, es importante mencionar que la implementación del modelo de scoring de otorgamiento aquí propuesto puede conducir a mejoras significativas en la productividad del área encargada del otorgamiento de crédito y, por ende, a la disminución en tiempos de respuesta para las promesas de servicio, tanto a clientes internos como externos.

El modelo se desarrolló de tal forma que los clientes con baja probabilidad de incumplimiento fueran catalogados en el puntaje de 50% hacia abajo y que valores mayores que el 50% fueran considerados como clientes con alta probabilidad de incumplimiento. No obstante, al tener presente la cercanía entre ambos límites se acordó con la administración de la entidad financiera en la que se elaboró el mismo mantener un margen de holgura prudencial, en el que se tomarán, como ya está definido, clientes con baja probabilidad de caer en incumplimiento aquellos cuyo puntaje sea menor o igual que el 50% y clientes malos desde el 65% hacia arriba. Aquellos deudores que se encuentren en el intervalo entre el 51% y el 64% se considerarán como en zona gris, es decir, en el rango en el que se deben analizar los casos por fuera del scoring.

Referencias

- Alonso Cánovas, D., y Tubau Sala, E. (2002). Inferencias bayesianas: una revisión teórica. *Anuario de Psicología*, 33(1), 25-47. Recuperado el 27 de abril de 2017, de: <http://revistes.ub.edu/index.php/Anuario-psicologia/article/view/8786>
- Álvarez Franco, S. I., y Osorio Betancur, A. (2011). *Medición del riesgo crédito en Colombia - Hacia Basilea III* (trabajo de grado, Ingeniería Administrativa, Escuela de Ingeniería de Antioquia, Envigado). Recuperado el 12 de diciembre de 2016, de: <https://repository.eia.edu.co/handle/11190/1118>
- Banca de las Oportunidades (2016). *Reporte trimestral de inclusión financiera*. Recuperado el 5 de mayo de 2017, de: <http://bancadelasoportunidades.gov.co/es/articulos/sitio/reportes-trimestrales-de-inclusion-financiera-marzo-de-2017>
- Banco de Pagos Internacionales (2001). *Visión general del nuevo acuerdo de capital de Basilea*. Basilea: Banco de Pagos Internacionales. Recuperado el 25 de abril de 2017, de: http://www.bis.org/publ/bcbsca02_s.pdf
- Basogain Olabe, X. (1998). *Redes neuronales artificiales y sus aplicaciones*. Bilbao: Universidad del País Vasco, Escuela Superior de Ingeniería de Bilbao. Recuperado el 11 de diciembre de 2016, de: http://www.ciberesquina.una.edu.ve:8080/2014_2/350_E.pdf
- Blázquez Ochando, M. (2012, 19 de diciembre). *Modelo probabilístico (web log post)*. Recuperado el 26 de marzo de 2017, de: <http://ccdoc-tecnicasrecuperacioninformacion.blogspot.com.co/2012/12/modelo-probabilistico.html>
- Comité de Supervisión Bancaria de Basilea (2004). *Aplicación de Basilea II: aspectos prácticos*. Basilea: Banco de Pagos Internacionales. Recuperado el 8 de mayo de 2017, de: <http://www.bis.org/publ/bcbs109esp.pdf>

- Comité de Supervisión Bancaria de Basilea (2010). *Basilea III: marco regulador global para reforzar los bancos y sistemas bancarios*. Basilea: Banco de Pagos Internacionales. Recuperado el 28 de abril de 2017, de:
http://www.bis.org/publ/bcbs189_es.pdf
- Congreso de Colombia (2009). *Ley 1318, de 15 de julio de 2009, por medio de la cual se dictan normas en materia financiera, de seguros, del mercado de valores y otras disposiciones*. Bogotá: Congreso de Colombia. Recuperado el 26 de agosto de 2016, de:
http://www.secretariassenado.gov.co/senado/basedoc/ley_1328_2009.html
- Enchautegui, M. E. (2003). Módulo de estudio sobre modelos probit y logit. *Departamento de Economía, Universidad de Puerto Rico. Recinto de Río Piedras*. Recuperado el 20 de marzo de 2017, de:
<http://studylib.es/doc/7400986/n%C3%BAm.-7-m%C3%B3dulos-de-estudio-sobre-modelos-probit-y-logit>
- Echemendía Tocabens, B. (2011). Definiciones acerca del riesgo y sus implicaciones. *Revista Cubana de Higiene y Epidemiología*, 49(3), 470-481. Recuperado el 24 de diciembre de 2016, de:
<http://www.redalyc.org/pdf/2232/223221363013.pdf>
- Fernández Castaño, H., y Pérez Ramírez, F. O. (2005). El modelo logístico: una herramienta estadística para evaluar el riesgo de crédito. *Revista Ingenierías Universidad de Medellín*, 4(6), 55-75. Recuperado el 24 de abril de 2017, de: <http://www.redalyc.org/articulo.oa?id=75040605>
- Greene, W. H. (2003). *Econometric analysis*. Upper-Saddle River, NJ: Prentice-Hall.
- Gutiérrez Girault, M. A. (2007). *Modelos de credit Scoring - Qué, cómo, cuándo y para qué*. Munich: Munich Personal RePEc Archive. Recuperado el 11 de diciembre de 2016, de: <https://mpra.ub.uni-muenchen.de/16377/>

- Herrera García, B. (2009). La supervisión de los bancos y el rol del Comité de Basilea para la supervisión bancaria. *Contaduría y Administración*(212), 41-48. Recuperado el 2 de mayo de 2017, de:
<http://www.ejournal.unam.mx/rca/212/RCA21204.pdf>
- Manjarrés V., F. A. (2012). *Estructura y regulación del sistema financiero y bursátil*. Saarbrücken: Editorial Académica Española.
- Matich, D. J. (2001). *Redes neuronales: conceptos básicos y aplicaciones* (trabajo de curso, Informática aplicada a la Ingeniería de Procesos, orientación I, Universidad Tecnológica Nacional, Rosario). Recuperado el 11 de diciembre de 2016, de:
https://www.frro.utn.edu.ar/repositorio/catedras/quimica/5_anio/orientadora1/monograis/matich-redesneuronales.pdf
- Pérez Ramírez, F. O., y Támara Ayús, A. L. (2012). Análisis discriminante como seleccionador de variables influyentes en el cálculo de la probabilidad de incumplimiento. *Revista Ciencias Estratégicas*, 20(27), 307-322. Recuperado el 25 de abril de 2017, de:
<https://revistas.upb.edu.co/index.php/cienciasestrategicas/article/viewFile/1476/1437>
- Presidencia de la República (1993). *Decreto 663, de 2 de abril de 1993, por medio del cual se actualiza el estatuto orgánico del sistema financiero y se modifica su titulación y numeración* (Diario Oficial N° 40.820). Recuperado el 25 de abril de 2017, de:
http://www.secretariassenado.gov.co/senado/basedoc/estatuto_organico_sistema_financiero.html
- Puertas Medina, R., y Martí Selva, M.L. (2013). Análisis del credit scoring. *Revista de Administração de Empresas*, 53(3), 303-315. Recuperado el 5 de noviembre de 2017, de:

http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0034-75902013000300007&lng=en&tlng=en

Rayo, S. (2013). *Gestión avanzada de riesgo de crédito. Seminario para gerentes de riesgos de las entidades de microfinanzas*. Lima: Superintendencia de Bancos y Seguros y Banco Interamericano de Desarrollo. Recuperado el 22 de diciembre de 2016, de: http://www.sbs.gob.pe/Portals/0/jer/pres_doc_basilea/l%20CONCEPTOS.pdf

Saavedra García, M. L. y Saavedra García, M. J. (2009). *Modelos para medir el riesgo de crédito en la banca. Cuadernos de Administración*, 23(40), 295-319. Recuperado el 07 de octubre de 2017, de: <http://www.scielo.org.co/pdf/cadm/v23n40/v23n40a13.pdf>

Samaniego Medina, R. (2008). *El riesgo de crédito en el marco del Acuerdo de Basilea II*. Madrid: Delta Publicaciones Universitarias.

Superintendencia Financiera de Colombia (1995). *Circular básica, contable y financiera (circular externa 100 de 1995)*. Bogotá: Superintendencia Financiera de Colombia. Recuperado el 24 de diciembre de 2016, de: <https://www.superfinanciera.gov.co/jsp/loader.jsf?!Servicio=Publicaciones&Tipo=publicaciones&IFuncion=loadContenidoPublicacion&id=15466>

Superintendencia Financiera de Colombia (2002). *Carta circular externa número 035 de 2002*. Bogotá: Superintendencia Financiera de Colombia. Recuperado el 24 de diciembre de 2016, de: http://www.fasecolda.com/files/7413/8438/0663/ce_100_de_1995_cap_ii.pdf

Superintendencia Financiera de Colombia (2016). *Actualidad del sistema financiero colombiano*. Recuperado el 5 de mayo de 2017, de: <https://www.superfinanciera.gov.co/jsp/loader.jsf?!Servicio=Publicaciones&Tipo=publicaciones&IFuncion=loadContenidoPublicacion&id=10088362>

Superintendencia Financiera de Colombia (2017). *Acerca de la SFC*. Recuperado el 28 de abril de 2017, de:

<https://www.superfinanciera.gov.co/jsp/loader.jsf?!Servicio=Publicaciones&Tipo=publicaciones&IFuncion=loadContenidoPublicacion&id=60607>

Torres Avendaño, G. I. (2005). El Acuerdo de Basilea: estado del arte del SARC en Colombia. *Ad- Minister*, 6, 114-134. Recuperado el 8 de mayo de 2017, de:

<http://publicaciones.eafit.edu.co/index.php/administer/article/view/669/596>

Universidad Complutense de Madrid, Departamento de Sociología IV. *Departamento Sociología IV. Manuales y guías para el análisis de datos en SPSS*. Recuperado el 24 de abril de 2017, de: http://pendientedemigracion.ucm.es/info/socivmyt/paginas/D_departamento/materiales/datos_multivariante.htm

Ustáriz González, L. H. (2003). El Comité de Basilea y la supervisión bancaria. *Vniversitas*, 52(105), 431-462. Recuperado el 2 de mayo de 2017, de:

<http://revistas.javeriana.edu.co/index.php/vnijuri/article/view/14855/11995>

Anexos

Anexo 1. Coeficientes estimados

Variable	Coeficiente	Error estándar	Valor de z	Pr(> z)	Significancia
(Intercepto)	- 32,96639704	9,22776296	-3,57	0,00035	***
Oficina 2	- 0,281539333	0,300640737	-0,94	0,34903	
Oficina 3	- 0,074140452	0,39437077	-0,19	0,85088	
Oficina 4	- 0,126837205	0,372028899	-0,34	0,73315	
Oficina 5	- 0,463050236	0,331618957	-1,4	0,16261	
Oficina 6	-0,34119933	0,271449696	-1,26	0,20877	
Oficina 7	- 0,490683307	0,491053149	-1	0,31768	
Oficina 8	- 0,030087883	0,322174413	-0,09	0,92559	
Oficina 9	0,508074694	0,345314427	1,47	0,1412	
Tiempo de maduración	0,648597039	0,031580645	20,54	< 0,00000000000000002	***
Monto	-	0,000000317	-16,98	<	***

	0,000005378			0,000000000000000002	
Saldo	0,000005452	0,000000312	17,5	< 0,000000000000000002	***
Plazo	- 0,162601975	0,015202379	-10,7	< 0,000000000000000002	***
Tasa	11,78988735	4,089557664	2,88	0,00394	**
Cuota	0,000023539	0,000002822	8,34	< 0,000000000000000002	***
Garantía real	- 11,05038201	1,565052436	-7,06	1,7E-12	***
Reestructurado	0,257058061	2,04989119	0,13	0,90021	
Edad	0,002510371	0,012238162	0,21	0,83747	
Ocupación desempleado	0,920881721	0,918915375	1	0,31628	
Ocupación empleado	- 0,914240768	0,827061025	-1,11	0,26898	
Ocupación estudiante	- 0,619927638	1,821928457	-0,34	0,73366	
Ocupación independiente	- 0,234375007	0,830117845	-0,28	0,77768	
Ocupación jubilado	- 1,399037093	0,895966941	-1,56	0,11841	
Ocupación	-	1,171449963	-0,29	0,76898	

pensionado	0,344063076				
Nivel educativo ninguno	- 0,260362306	0,645973938	-0,4	0,68691	
Nivel educativo posgrado	- 0,111850452	2,26621196	-0,05	0,96064	
Nivel educativo primaria	- 1,014133446	0,397677057	-2,55	0,01077	*
Nivel educativo técnico	- 0,410761399	0,898331746	-0,46	0,64749	
Nivel educativo tecnológico	- 0,348536631	0,245451166	-1,42	0,15561	
Nivel educativo universitario	- 0,328657962	0,296215825	-1,11	0,26721	
Ingresos totales	0,000000126	0,000000202	0,63	0,53079	
Egresos totales	-0,00000151	0,000000524	-2,88	0,00392	**
Estrato 2	0,196025944	0,540983969	0,36	0,71709	
Estrato 3	0,138271406	0,536764408	0,26	0,79671	
Estrato 4	- 0,153405226	0,599990952	-0,26	0,7982	
Estrato 5	0,072464242	0,743944399	0,1	0,9224	
Estrato 6	- 0,783930522	1,689497157	-0,46	0,64265	
Antigüedad del	-	0,018964294	-1,71	0,08642	.

empleado	0,032515884				
Estado civil eclesiástico	- 0,989877499	1,629077354	-0,61	0,54343	
Estado civil separado	1,233018769	0,309683226	3,98	6,84678E-05	***
Estado civil soltero	0,323238392	0,220204439	1,47	0,14213	
Estado civil unión libre	- 0,118456377	0,332489459	-0,36	0,72164	
Estado civil viudo	0,823585746	0,452997802	1,82	0,06905	.
Sexo masculino	0,538918383	0,194260154	2,77	0,00553	**
Número de personas a cargo	- 0,012939372	0,108057826	-0,12	0,90469	
Tipo de vivienda familiar	- 0,386093463	0,419079389	-0,92	0,3569	
Tipo de vivienda NULL	1,528882044	0,413423532	3,7	0,00022	***
Tipo de vivienda propia	- 0,758151315	0,440288724	-1,72	0,08508	.
Tipo de contrato NULL	0,163683968	0,923898646	0,18	0,85938	
Tipo de contrato servicios	0,103413865	1,307270928	0,08	0,93695	
Tipo de contrato término definido	0,173584009	0,985454103	0,18	0,86018	

Tipo de contrato término indefinido	0,513591197	0,919346244	0,56	0,5764	
Antigüedad en la cooperativa	0,02286393	0,057517952	0,4	0,69099	

Fuente: elaboración propia

Anexo 2. Intervalos de confianza

Variable	Límite inferior: 2,5%	Límite superior: 97,5%
(Intercepto)	-51,0524801	-14,88031398
Oficina 2	-0,87078435	0,307705683
Oficina 3	-0,847092957	0,698812054
Oficina 4	-0,856000447	0,602326038
Oficina 5	-1,113011448	0,186910976
Oficina 6	-0,873230958	0,190832298
Oficina 7	-1,453129794	0,47176318
Oficina 8	-0,66153813	0,601362364
Oficina 9	-0,168729145	1,184878534
Tiempo	0,586700113	0,710493965
Monto	-0,000005999	-4,7576E-06
Saldo	0,000004842	6,0631E-06

Plazo	-0,192398091	-0,13280586
Tasa	3,774501619	19,80527309
Cuota	0,000018008	0,00002907
Garantía real	-14,11782842	-7,982935604
Reestructurado	-3,760654843	4,274770966
Edad	-0,021475986	0,026496728
Ocupación desempleado	-0,880159319	2,721922761
Ocupación empleado	-2,53525059	0,706769055
Ocupación estudiante	-4,190841796	2,950986519
Ocupación independiente	-1,861376087	1,392626073
Ocupación jubilado	-3,155100028	0,357025842
Ocupación pensionado	-2,640062813	1,951936662
Nivel educativo ninguno	-1,52644796	1,005723348
Nivel educativo posgrado	-4,553544275	4,329843371
Nivel educativo primaria	-1,793566155	-0,234700738
Nivel educativo técnico	-2,171459267	1,34993647
Nivel educativo tecnológico	-0,829612077	0,132538815
Nivel educativo universitario	-0,90923031	0,251914386
Ingresos totales	-0,000000269	0,000000522

Egresos totales	-0,000002536	-4,842E-07
Estrato 2	-0,864283151	1,256335039
Estrato 3	-0,913767502	1,190310314
Estrato 4	-1,329365884	1,022555432
Estrato 5	-1,385639987	1,530568472
Estrato 6	-4,095284101	2,527423056
Antigüedad del empleado	-0,069685217	0,00465345
Estado civil eclesiástico	-4,182810441	2,203055442
Estado civil separado	0,626050799	1,839986739
Estado civil soltero	-0,108354377	0,754831161
Estado civil unión libre	-0,770123741	0,533210987
Estado civil viudo	-0,064273631	1,711445122
Sexo masculino	0,158175477	0,919661289
Número de personas a cargo	-0,22472882	0,198850075
Tipo de vivienda familiar	-1,207473973	0,435287046
Tipo de vivienda <i>NULL</i>	0,71858681	2,339177278
Tipo de vivienda propia	-1,621101357	0,104798727
Tipo de contrato <i>NULL</i>	-1,647124103	1,974492039
Tipo de contrato servicios	-2,458790072	2,665617802

Tipo de contrato término definido	-1,757870541	2,105038559
Tipo de contrato término indefinido	-1,28829433	2,315476725
Antigüedad en la cooperativa	-0,089869184	0,135597044

Fuente: elaboración propia