Modelo de scoring para una entidad financiera especializada en el otorgamiento de crédito de vehículos

Por

Valeria Astudillo Girón

Trabajo presentado como requisito parcial para optar al título de Magíster en Administración Financiera

Asesor

Brayan Rojas Ormaza, FMR, Msc.

Universidad EAFIT

Cali, junio, 2022

Resumen

Este trabajo presenta el diseño de un modelo de credit scoring en la etapa de otorgamiento de crédito para una entidad especializada en la financiación de vehículos, con el fin de identificar la probabilidad de incumplimiento de los acreedores vigentes hasta abril de 2020, incluyendo las variables más relevantes acorde a la metodología desarrollada. Lo anterior nace de la necesidad de la entidad por implementar prácticas que le permitan controlar posibles pérdidas en sus activos, teniendo en cuenta que los credit scoring son modelos estadísticos que apoyan la gestión del riesgo de crédito en su medición y monitoreo.

Palabras claves: Riesgo de crédito, Credit scoring, Crédito de vehículo.

Abstract

This paper presents the design of a credit scoring model at the stage of granting credit to an entity specialized in vehicle financing, to identify the probability of default of creditors effective until April 2020, including the most relevant variables according to the methodology developed. This stems from the need of the entity to implement practices that allow it to control possible losses in its assets, taking into account that credit scoring are statistical models that support the management of credit risk in its measurement and monitoring.

Key words: Credit risk, Credit scoring, Vehicle credit.

Tabla de contenido

Índice de tablas

Índice de gráficos

1. Introducción

Para la realización de este proyecto de grado, se tomó como base la información de una empresa del sector financiero especializada en la originación y administración de créditos de vehículos. Esta entidad tiene una trayectoria de más de 20 años en el mercado, su sede principal se encuentra ubicada en la ciudad de Santiago de Cali, Valle del Cauca, y tiene operación de créditos de vehículos a nivel nacional.

En esta entidad hay ausencia de buenas prácticas para la medición y control del riesgo crediticio acorde a la normatividad vigente; por lo tanto, se están generando debilidades al momento de implementar los controles en la etapa de otorgamiento de créditos de vehículo, este hallazgo no garantiza la eficiencia en los activos crediticios de la compañía y podría afectar a largo plazo la rentabilidad financiera de la compañía.

Se debe tener en cuenta que la ausencia de las buenas prácticas para la gestión de riesgo crediticio en dicha entidad se da también por la no obligatoriedad de adoptar un sistema de administración de riesgo crediticio, ya que esta entidad no capta recursos del público y sus ingresos anuales no la acreditan como entidad bancaria, por lo tanto, la Superintendencia Financiera de Colombia (SFC) no la vigila, pero sí la puede requerir ante una petición de un consumidor; sin embargo, para la administración de esta compañía es fundamental acoger las buenas prácticas acorde a la normatividad vigente, dado que esta entidad se fondea a través de la venta de cartera a otras entidades financieras, como los bancos, los cuales sí exigen la implementación de un SARC (Sistema de Administración de Riesgo Crediticio); por ende, las características de los créditos de vehículo que se otorguen deben ajustarse a la eficiencia y calidad que exigen estos aliados.

Este documento presentará el desarrollo de un modelo de credit scoring (sistema automatizado de calificación de crédito), que le permitirá a la entidad identificar la

probabilidad de incumplimiento de un cliente en la etapa de originación del crédito de vehículo y se desarrollará con la información de los acreedores que tienen créditos de vehículo al corte de abril de 2020, a través de una metodología que identificará las principales variables que se deben tener en cuenta al momento de medir la probabilidad de ocurrencia en pago de estos clientes y así poder clasificarlos en diferentes perfiles de riesgo en la etapa de originación de cada uno de los créditos.

Con los resultados obtenidos, se espera emitir recomendaciones hacia la entidad para que evalúe la implementación del modelo dentro del proceso de análisis de los créditos de vehículo y, de esta manera, la compañía pueda transformar sus operaciones, controlando adecuadamente el riesgo crediticio.

Para lograr el desarrollo del modelo, se investigará la literatura relacionada con los modelos de credit scoring, los trabajos similares realizados donde se hayan implementado este tipo de modelos en entidades financieras; y, posteriormente, se estudiará la metodología aplicada para los modelos de riesgo de crédito, incluyendo la selección de variables para estimar el modelo con las respectivas pruebas estadísticas que validan el modelo de manera adecuada.

2. Objetivos

2.1 Objetivo general

Diseñar un modelo de scoring que permita identificar la probabilidad de incumplimiento en la etapa de otorgamiento de un crédito de vehículo, para una entidad del sector financiero.

2.2 Objetivos específicos

- Definir la metodología que suministrará la aplicación del modelo de scoring, diseñado para la etapa de originación de los créditos de vehículos.
- Identificar las variables que se deben incluir en el modelo diseñado.
- Realizar las pruebas correspondientes a los modelos estimados, con el fin de seleccionar el modelo que estadísticamente refleje la realidad.

3. Marco teórico

Para la presente investigación se tienen en cuenta las siguientes referencias conceptuales y métodos de estudio para el diseño de un modelo de scoring:

3.1 Riesgo crediticio

El riesgo crediticio es la posibilidad de que una entidad incurra en pérdidas y se disminuya el valor de sus activos, como consecuencia de que un deudor o contraparte incumpla sus obligaciones (Superfinanciera, 2016, circular externa 025, p. 3).

El riesgo de crédito nace de la necesidad de los acreditados por apalancarse, siendo este un factor natural que está inmerso en la naturaleza del negocio de las entidades financieras.

La gestión adecuada del riesgo crediticio mitiga la materialización de este riesgo (default), previniendo pérdidas económicas para el sector financiero, que podría afectar su estabilidad y posición en el mercado. En ese orden de ideas, la materialización del riesgo de crédito se ha considerado en las últimas décadas como una de las causas de las crisis financieras (Urbina Poveda, 2017).

Una de las variables más importantes para medir el riesgo crediticio es la probabilidad de incumplimiento, la cual indica a las empresas la probabilidad de que la contrapartida no haga frente a sus obligaciones contractuales. Es decisión de cada compañía indicar cuándo se materializa este factor, y puede ser en los siguientes casos: retraso en el pago de más de un mes, retraso en el pago de más de tres meses, entre otros (Samaniego Medina, 2008).

3.2 Sistema de Administración de Riesgo Crediticio – SARC

De acuerdo con lo que establece la Superintendencia Financiera (2016), el SARC otorga a las entidades los mecanismos especiales para la adecuada administración del riesgo crediticio, no sólo desde la perspectiva de su cubrimiento a través de un sistema de provisiones, sino también por medio de la administración del proceso de otorgamiento de créditos y permanente seguimiento de éstos.

Los principales elementos que debe contener un SARC son:

- Políticas y procesos de administración
- Modelos internos o de referencia para la estimación o cuantificación de pérdidas esperadas
- Sistema de provisiones
- Procesos de control interno

3.3 Modelos scoring

Los credit scoring son modelos estadísticos que permiten medir y controlar el riesgo crediticio. Para las entidades financieras es una herramienta de bajo costo, que contribuye a generar eficiencia en la etapa de otorgamiento de crédito hacia los solicitantes.

Los modelos de credit scoring emplean principalmente la información del evaluado, contenida en las solicitudes de crédito y/o en fuentes internas y/o externas de información. Dentro de las metodologías disponibles para estimar este tipo de modelo, se encuentra los modelos probit y logit, los cuales permiten para cada acreditado una probabilidad de default, clasificándolos en grupos de riesgo (Gutiérrez Girault, 2007).

También, los credit scoring son métodos estadísticos utilizados para clasificar a los solicitantes de crédito, o incluso a quienes ya son clientes de la entidad evaluadora, con el fin de tipificarlos en 'buenos' y 'malos' (Hand & Henley, 1997).

El beneficio principal del scoring estadístico es que permite reducir el tiempo en gestiones de cobro, debido a que la calificación de las solicitudes de crédito reduce el número, monto y plazo de los préstamos desembolsados a los solicitantes de alto riesgo, y la calificación de la solicitud puede ayudar a priorizar los esfuerzos que se deben llevar a cabo en la cobranza. Adicionalmente, el modelo de *Scoring* utiliza la misma lógica que el analista de crédito, pues se basa en experiencias y seguimientos de créditos otorgados en el pasado, mediante un análisis de las características de los nuevos solicitantes, con el fin de calificar o descalificar los perfiles (Schreiner, 2002).

Los modelos estadísticos más usados para calcular la probabilidad de incumplimiento son:

3.3.1 Modelo Logit

El modelo logístico arroja un resultado binario, en donde la variable dependiente toma el valor de 0 o de 1, siendo un modelo de predicción de probabilidad de ocurrencia de una variable dicotómica categórica.

La regresión binaria es un tipo de análisis de regresión, donde la variable dependiente es una variable dummy, ejemplo: Buen cliente (0) o Mal cliente (1) (Fernández Castaño & Pérez Ramírez, 2005).

La función logística relaciona la variable dependiente con las variables independientes $\boldsymbol{X_1, X_2 \ldots . X_i}$ a través de la siguiente ecuación:

$$\boldsymbol{Y_i = \frac{1}{1 + \exp(-z)} + u_i}$$

$$\boldsymbol{Y_i = Variable\ dependiente\ (toma\ el\ valor\ de\ 0\ y\ 1)}$$

$$z = Scoring\ logistico\ \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots \beta_K X_K$$

$$u = Variable\ aleatoria\ que\ se\ distribuye\ normalemnte\ N\ (0, \sigma^2)$$

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

3.3.2 Modelo Probit

De la misma forma que el modelo logístico, este modelo arroja un resultado binario y procesa variables aleatorias, para asociar una probabilidad con cada uno de los valores que pueden tomar estas variables.

En el modelo probabilístico la variable discreta *Y** proviene de una variable continua *Y**, de la siguiente manera: (Enchautegui, 2003)

$$Y^* = \beta_0 X_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots \beta_K X_K + \mu$$

$$Y^* = \sum_{K=0}^{k} \beta_K X_K + \mu$$

$$X = Variables\ explicativas, X_0 = 1\ \ Es\ la\ constante$$

$$\mu = \text{El error}$$

$Y^* =$

$Variable\ no\ observada, toma\ el\ valor\ de\ 0\ o\ de\ 1, si\ cruza\ un\ límite\ o\ no$

3.3.3 Comparación de modelos Probit y Logit

- Similitudes: la estimación de los parámetros del modelo Logit y Probit, se efectúan por el método de estimación de máxima verosimilitud, permitiendo que se maximice la función asociada a una muestra de tamaño N.
- Diferencia: la diferencia entre los modelos de predicción (Logit y Probit) radica en su función, ya que la función logística estadísticamente tiene colas más anchas que la probabilística y, la interpretación con la función logística es mucho más práctica, ya que es exponencial y cuenta con una transformación inversa que facilita la interpretación de resultados.

Por lo tanto, se escoge la aplicación de un modelo logístico, ya que si la entidad decide hacer un proceso de automatización para implementar el modelo de credit scoring, con una herramienta de desarrollo interno, será mucho más práctica la aplicación con la metodología del modelo logístico, gracias a sus cálculos y parámetros.

Tabla 1. Trabajos de investigación relacionados con modelos de credit scoring

Modelos de Scoring			
Título	**Metodología**	**Variables**	
		Cualitativas	**Cuantitativas**
Propuesta de un modelo de score de originación para la cartera de consumo de una cooperativa de ahorro y crédito del segmento 3 en el Ecuador. (Aguilar, 2021)	Modelo Logit (0: mal pagador, 1: buen pagador); mediante el cociente de verosimilitud.	Sociodemográficas (sexo, edad, ciudad, estado civil, nivel de escolaridad).	- Score de las centrales de riesgo - Ingreso mensual promedio

Un Modelo de Credit Scoring para instituciones de microfinanzas en el marco de Basilea II. (Rayo Canton, Lara Rubio, & Camino Blasco, 2010)	Modelo de regresión logística binaria (cliente paga 0; cliente no paga 1)	- Actividad económica del cliente - Sexo, edad, estado civil. - Tipo de garantía - Situación laboral del cliente. - Lugar geográfico. - Sector de actividad económica	- Comportamiento en centrales de riesgo - Número de cuotas pagadas - Días de mora - Tasa de interés mensual - Variables macroeconómicas (tasa de desempleo durante las vigencias del crédito, PIB, IPC).
Scoring de crédito: herramienta para la evaluación de riesgo de crédito en entidades financieras. (Carbonell, 2018)	Modelo lineal generalizado (mínimos cuadrados generalizados) de una entidad financiera en Colombia	- Edad, sexo - Situación laboral - Domicilio - Profesión - Actividad que desempeña -Antigüedad en el empleo. - Estrato.	- Ingresos del cliente - Tasa de desempleo - Tasa de inflación - Tasa de interés (BanRep) -PIB - Tasa de cambio - Puntaje en las centrales de riesgo.
Perfil de riesgo de crédito para una cooperativa en Villavicencio a partir de	Modelo Logit: toma el valor de 1 si el cliente es incumplido y 0 si es cumplido	- Edad - Ocupación - Género - Antigüedad laboral	Ingresos del cliente

un modelo logit. (Pardo Carrillo, 2020)		- Estado civil - Personas a cargo	
Diseño de un modelo de scoring para el otorgamiento de crédito de consumo en una compañía de financiamiento colombiana. (Duque & Baena, 2017)	Modelo de regresión logística	- Garantía - Reestructurado - Edad - Ocupación - Nivel Educativo - Estrato socioeconómico - Antigüedad laboral - Estado civil - Sexo - Tipo de vivienda	- Categoría Riesgo - Monto desembolsado - Saldo actual - Plazo (meses de crédito) - Tasa pactada crédito - Días de mora - Ingreso total
Modelo Scoring para el otorgamiento de crédito de las pymes. (Echeverri, 2017)	Modelo Logit dicotómico, la variable dependiente es binaria y toma sólo dos valores.		-Plazo - Cupo - Acierta plus (puntaje data) - Prueba acida - Razón endeudamiento

Construcción de un modelo de scoring para el otorgamiento de crédito en una entidad financiera. (Ochoa, Galeano, & Agudelo, 2010)	Análisis discriminante: se asigna puntajes a cada perfil de cliente	-Ubicación de la oficina - Garantía. - Reestructurado - Edad - Ocupación - Nivel Educativo - Antigüedad Laboral - Estado civil - Género - Personas a cargo - Tipo de vivienda	- Número de días en que el cliente incumple el pago de la obligación - Categoría de calificación de riesgo. - Ingreso total - Plazo del crédito - Capacidad de pago
Modelo de scoring para aprobación de créditos para la cartera de consumo, en una cooperativa de aporte y crédito colombiana. (Castro & Noriega, 2019)	Modelo Logit	- Garantía - Ocupación - Sexo - Edad - Estrato - Tipo vivienda - Estado Civil - Nivel Educativo - Personas a Cargo - Sector económico	- Monto - Plazo - Ingresos - Endeudamiento (Deudas vs salario) - Puntaje acierta - Calificación crediticia (alto, medio, bajo)
Modelo financiero para riesgo de crédito de vehículo del banco DAVIVIENDA. (Rueda Pimiento & Vergel Esteban, 2006)	Modelo Logit y Probit- pruebas con ambas	- Edad - Sexo - Estado civil - Actividad Económica - Experiencia laboral - Sucursal del crédito	- Plazo - Monto - Valor Comercial - Préstamo - Tasa - Ingresos

Un modelo de credit Scoring basado en el conocimiento de la aplicación de Basilea II y su papel innovador en el sector bancario. (Esteve, 2007)	Análisis discriminante: Clasificación de clientes en segmentos y algoritmo de Kohonen	- Año de nacimiento - Niños a cargo -Personas dependientes	- Renta del cónyuge - Renta cliente - Gastos en hipotecas - Gastos en tarjeta de crédito

Fuente: Elaboración propia, 2022.

3.4 Pruebas estadísticas para la selección del modelo

3.4.1 Test de Hosmer-Lemeshow

La prueba de Hosmer y Lemeshow es una prueba de bondad de ajuste, que se realiza a los modelos estimados, la cual comprueba si el modelo puede explicar lo que se observa. La hipótesis nula de esta prueba consiste en que no se presenten diferencias entre los valores esperados y observados, por lo tanto, un p-valor superior a 0,05 indica que lo observado se ajusta suficientemente a lo esperado y al rechazarse esta prueba, se indicaría que el modelo no se ajusta a la realidad (Pérez, 2013).

3.4.2 Curva ROC

La curva ROC- Receiver Operating Characteristic (Característica operativa del receptor) es una prueba que clasifica a los individuos de una población en dos grupos: uno que presenta un evento de interés y otro que no, para esta investigación, sería incumplimiento y no cumplimiento. La curva es el gráfico resultante que representa para cada valor las medidas de sensibilidad (eje Y) y 1-especificidad (eje X) de la prueba. La sensibilidad cuantifica a los individuos que

representan el evento de interés y la especificidad cuantifica la proporción de individuos que no representan el evento.

La curva presenta una gran capacidad de discriminación si la sensibilidad y especificidad se aproximan al 100%. Para saber si la exactitud de la prueba es alta el valor del área de la curva debe estar por encima de 0,9 (Benavides, 2017).

El área bajo la curva es el estadístico que mide la capacidad discriminante de la prueba. Su rango de valores va de 0,5 hasta 1, por lo tanto, entre mayor sea este estadístico mejor exactitud presentará la prueba.

3.4.3 Matriz de confusión

La matriz de confusión es una herramienta que permite evaluar qué tan efectivo es el modelo. Se representa a través de la tabla 2, la cual tiene dos dimensiones, indicando en las columnas las predicciones y en las filas los casos reales.

Tabla 2. Matriz de confusión

		Predicción	
		Positivo	**Negativo**
Real	**Positivo**	Verdaderos positivos	Falsos Negativos
	Negativo	Falsos Positivos	Verdaderos Negativos

Fuente: Elaboración propia, 2022.

Los modelos no son 100% precisos en la mayoría de los casos, es decir, que los falsos positivos y negativos, sean equivalentes a cero. Por lo tanto, los verdaderos positivos y negativos son los datos estimados de manera correcta (González, 2019).

4. Metodología

Se tomó como fuente de información para la estimación del modelo, la base de datos de la entidad, la cual por temas de confidencialidad no se expondrá su nombre, sino que para fines académicos de este proyecto de grado se le nombró entidad financiera. Esta base de datos contiene una población total de 12,153 créditos de vehículos otorgados a personas naturales y jurídicas hasta abril de 2020 con 51 variables (cualitativas y cuantitativas), las cuales se sometieron a un análisis estadístico descriptivo para determinar la relación de cada variable con los créditos vigentes. Este análisis descriptivo se realizó para la base de datos completa (Personas jurídicas y naturales) y para la base de datos de personas naturales únicamente (11,587 registros de créditos de vehículos).

La metodología seleccionada para el diseño del modelo de scoring fue la estimación mediante regresión logística binaria en el software Stata 16; la cual, como se mencionó en el apartado anterior, arroja un resultado binario, y la variable dependiente toma el valor de 0 y 1, para este caso, la variable dependiente se denomina DEFAULT donde (0) es buen cliente y (1) es un mal cliente.

La selección de las variables explicativas se realizó en el software Stata mediante la función Stepwise, la cual utiliza la estrategia Backward (selección de variables), con el fin de encontrar el modelo más reducido que explique los datos (principio de parsimonia). Al obtener las variables adecuadas, se realizó la estimación de dos modelos propuestos: modelo con personas naturales y jurídicas y modelo únicamente con personas naturales. Se escogió el modelo que se ajustaba a la realidad de la operación diaria en el otorgamiento de créditos de vehículos en la entidad financiera y, finalmente, se realizó el análisis de los resultados para emitir las recomendaciones pertinentes.

5. Resultados

5.1 Modelación

Para realizar el modelo final, inicialmente se revisaron y limpiaron los datos con base en el análisis descriptivo de las variables, se depuraron datos vacíos o atípicos y las variables se organizaron acorde a su distribución. La variable dependiente DEFAULT: (0) buen cliente y (1) mal cliente, se definió con los días de mora que presentaba cada crédito, donde un cliente es bueno si al corte de abril de 2020 presentó mora inferior a 31 días y, si presentó mora superior a 31 días al mismo corte, el cliente se establece como malo; lo anterior, teniendo en cuenta que el cobro jurídico en esta entidad inicia a partir de los 31 días en mora. Al momento de analizar esta variable dependiente, se evidenció que era necesario realizar un proceso de balanceo. Este proceso se llevó a cabo mediante el comando iweight en Stata, creando una nueva variable llamada PONDERACIÓN_DÍAS_MORA con el fin de incluir en esta, un factor de expansión que permitiera balancear los datos de la variable dependiente. Este factor de expansión se extrae de la siguiente manera:

$$\frac{\boldsymbol{Clientes\ buenos}}{\boldsymbol{Clientes\ malos}} = \boldsymbol{Factor\ de\ expansi}\text{ó}\boldsymbol{n}.$$

Factor de expansión: número que indica las veces que debo multiplicar los registros de los clientes malos para que sean equivalentes a los clientes buenos.

Esta nueva variable se incluyó con el comando iweight al momento de realizar el proceso de selección de variables (stepwise), para que el software ejecutara el balanceo de la variable dependiente con el modelo estimado.

Posteriormente, en la revisión de las variables explicativas (cualitativas y cuantitativas), se reagruparon variables con nuevas categorías y, para la estimación del modelo, se dicotomizaron con el fin de incluirlas en el análisis de los resultados. Finalmente, la base de datos quedó conformada con las variables explicativas que se detallan a continuación:

5.1.1 Variables explicativas no categóricas

- R_ACIERTA: score de data crédito del cliente al momento de la aprobación del crédito. Se encuentra en un rango de 0 a 950.
- CUOTA_PACTADA: plazo del crédito (tiempo). Se encuentra entre 12 y 72 meses.
- PORC_FINANCIACIÓN: porcentaje de financiación del vehículo. El rango se encuentra entre el 20% y el 100%.
- EDAD: años de vida del cliente (19 a 79 años).

5.1.2 Variables explicativas categóricas

- MODELO_VEH: año de antigüedad del vehículo de cada crédito. Esta variable contiene 6 categorías: 2015 (esta categoría se agrupó con los modelos de vehículos de años anteriores), 2016, 2017, 2018, 2019 y 2020. Cada categoría se renombró para incluirla en la estimación del modelo, en su respectivo orden: MODELOVEH_1, MODELOVEH_2, MODELOVEH_3, MODELOVEH_4, MODELOVEH_5 Y MODELOVEH_6.
- CLASE_GAR: tipo de vehículo para cada crédito. Esta variable contiene 4 categorías: AUTOMÓVIL, CAMIÓN, CAMIONETA, REMOLCADOR y están renombradas para incluirlas en la estimación del modelo, en su respectivo orden: CLASE_GAR_01_1, CLASE_ GAR_01_2, C CLASE_GAR_01_3 Y CLASE_GAR_01_4.
- CAG_ACTIVIDAD_01234: actividad económica del cliente. Esta variable contiene 5 categorías: EMPLEADO, INDEPENDIENTE, PENSIONADO, RENTISTA DE CAPITAL, TRANSPORTADOR. Para la estimación del modelo se renombraron de la siguiente manera, respectivamente: CAG_ACTIVIDAD_01234_1; CAG_ACTIVIDAD_01234_2; CAG_ACTIVIDAD_01234_3; CAG_ACTIVIDAD_01234_4 Y CAG_ACTIVIDAD_ 01234_5.
- ESTADO_CIVIL: estado civil de cada cliente al momento del desembolso del crédito. Esta variable contiene 5 categorías: CASADO, DIVORCIADO, SEPARADO, SOLTERO Y UNIÓN LIBRE. Cada categoría se renombró para incluirla en la

estimación del modelo, en su respectivo orden: ESTADO_CIVIL_01_1, ESTADO_CIVIL_01_2, ESTADO_CIVIL_01_3, ESTADO_CIVIL_01_4 Y ESTADO_CIVIL_01_5.

- DEPARTAMENTO_01: departamento de residencia del cliente al momento del desembolso del crédito. Esta variable contiene 9 categorías: CALDAS, CAUCA, CUNDINAMARCA, NARIÑO, PUTUMAYO, QUINDÍO, RISARALDA, VALLE DEL CAUCA y OTROS, ésta última contiene los departamentos de Antioquia, Atlántico, Bolívar y Santander, se agrupó de esa manera acorde a la distribución de los datos. Cada una se encuentra renombrada para la estimación del modelo, en su respectivo orden: DEPARTAMENTO_01_2, DEPARTAMENTO_01_3, DEPARTAMENTO_01_4, DEPARTAMENTO_01_5, DEPARTAMENTO_01_7, DEPARTAMENTO_01_8, DEPARTAMENTO_01_9, DEPARTAMENTO_01_1 Y DEPARTAMENTO_01_6.
- RANGO_VLRO_012: valor desembolsado de cada crédito. Esta variable fue categorizada para mayor practicidad en el análisis por salarios mínimos (SMMLV 2021: $ 908.526) en 3 categorías: hasta 31 SMMLV, de 32 a 45 SMMLV y mayor a 46 SMMLV, cada una se encuentra renombrada para la estimación del modelo, en su respectivo orden: RANGO_VLRO_012_1, RANGO_VLRO_012_2 Y RANGO_VLRO_012_3. Se agrupó de esa manera para que la entidad tenga mayor facilidad en un largo plazo, de comparar el valor del desembolso del crédito vs los ingresos del cliente.
- TIPIFICACIÓN_DECISOR_01234: esta variable es una herramienta adquirida por la financiera, que trabaja en conjunto con una central de riesgo para indicarle al analista de manera rápida un estudio del crédito. La herramienta arroja distintos resultados que fueron tipificados en varias categorías: aprobado - TIPIFICACIÓN_DECISOR_01234_2, negado TIPIFICACIÓN_DECISOR_01234_3, entra a estudio TIPIFICACIÓN_DECISOR_01234_4 y fuera de servicio TIPIFICACIÓN_DECISOR_01234_5. Esta variable hace parte del estudio del crédito, mas no es determinante para la aprobación o negación del mismo.

5.2 Regresión logística binaria

Como primer paso se realizó una regresión logística binaria, con la base de datos completa, mencionada en el apartado anterior (registros de créditos otorgados a personas naturales y jurídicas), la cual contiene 31 variables y 12,153 registros. La tabla 3 muestra los registros de personas jurídicas denominados empresas, en la variable que describe la actividad económica del cliente (previo a su transformación) correspondiente a 566 créditos de vehículos. Las empresas se discriminaron en esta variable, porque la financiera no segrega los créditos otorgados a empresas en su sistema, éstas se identifican únicamente por medio de su ID, variable que no es significativa y no hace parte del modelo.

Tabla 3. Descripción variable actividad económica sin transformaciones con la base de datos completa

ACTIVIDAD ECONOMICA	**Cantidad Observaciones**	**%**
Empleado	5,992	49.30
Empresa	566	4.66
Independiente	3,867	31.82
Pensionado	355	2.92
Rentista de capital	261	2.15
Transportador	1,112	9.15
Total	**12,153**	**100.00**

Fuente: Cálculos propios a través del software Stata 16.

Posterior a dicotomizar esta variable categórica y realizar el proceso de selección de variables (stepwise), se evidencia que los registros de los créditos de personas jurídicas (empresa), se agrupan en una categoría nombrada “0”, tal como se observa en la tabla 3, esto se da porque hay una correlación entre las variables explicativas con los registros de personas jurídicas.

Tabla 4. Descripción variable actividad económica con transformaciones en la base de datos completa

ACTIVIDAD ECONOMICA	**Cantidad Observaciones**	**%**
Empleado	5,992	49.30
0	566	4.66
Independiente	3,867	31.82
Pensionado	355	2.92
Rentista de capital	261	2.15
Transportador	1,112	9.15
Total	**12,153**	**100.00**

Fuente: Cálculos propios a través del software Stata 16.

En la tabla 5 se muestran los coeficientes de la regresión logística binaria inicial que se estimó, con la base de datos completa, mencionada en el apartado 5.2.

Tabla 5. Modelo logístico inicial

Logistic regression	Number of obs	= 11,587
	LR chi2(31)	= 931.50
	Prob > chi2	= 0.0000
Log likelihood = -692.64392	Pseudo R2	= 0.4021

DEFAULT	**Coef.**	**Std. Err.**	**z**	**P>z**
R_ACIERTA	-.001	.0004794	3.31	0.001
EDAD	.018	.0076331	2.46	0.014
MODELOVEH_2 (2016)	-.687	.2491835	-2.76	0.006
MODELOVEH_3 (2017)	-2.056	.3809443	-5.40	0.000
MODELOVEH_4 (2018)	-2.617	.5181935	-5.05	0.000
MODELOVEH_5 (2019)	-3.338	.592244	-5.64	0.000

MODELOVEH_6 (2020)	-489.231	1.084999 -4.51	0.000
PORC_FINANCIACION (%)	.023	.0060664 3.90	0.000
RANGO_VLRO_012_2 (Desembolsos entre DE 32 A 45 SMMLV)	.773	.2156108 3.59	0.000
RANGO_VLRO_012_3 (Desembolsos MAYORES A 46 SMMLV)	.833	.2351282 3.54	0.000
TIPIFICACION_DECISOR_01234_2 (Aprobado)	-5.475	.6738558 -8.13	0.000
TIPIFICACION_DECISOR_01234_3 (Negado)	-5.689	.6615739 -8.60	0.000
TIPIFICACION_DECISOR_01234_4 (Estudio)	-5.569	.6861936 -8.12	0.000
TIPIFICACION_DECISOR_01234_5 (Fuera de Servicio)	-5.206	1.400947 -3.72	0.000
CAG_ACTIVIDAD_01234_2 (Independiente)	.335	.3015085 1.11	0.266
CAG_ACTIVIDAD_01234_3 (Pensionado)	.523	.2795345 1.87	0.061
CAG_ACTIVIDAD_01234_4 (Rentista de Capital)	-1.364	.8659335 -1.58	0.115
CAG_ACTIVIDAD_01234_5 (Transportador)	-.922	.9459698 -0.98	0.329
ESTADO_CIVIL_01_2 (Divorciado)	.878	.4596311 1.91	0.056
ESTADO_CIVIL_01_3 (Separado)	-.429	.5545553 -0.77	0.439
ESTADO_CIVIL_01_4 (Soltero)	-.136	.1976262 -0.69	0.490
ESTADO_CIVIL_01_5 (Unión Libre)	-.156	.2415215 -0.65	0.517
CLASE_GAR_01_2 (Camión)	-.171	.2744151 -0.62	0.533
CLASE_GAR_01_4 (Remolcador)	-1.186	.5666877 -2.09	0.036

DEPARTAMENTO_01_3 (Cauca)	.456	.404213 1.13	0.259
DEPARTAMENTO_01_4 (Cundinamarca)	-.918	.3347225 -2.74	0.006
DEPARTAMENTO_01_6 (Otros)	-.672	.3157887 -2.13	0.033
DEPARTAMENTO_01_7 (Putumayo)	1.142	.4624854 2.47	0.013
DEPARTAMENTO_01_8 (Quindío)	.074	.4265768 0.17	0.862
DEPARTAMENTO_01_9 (Risaralda)	.349	.2496368 1.40	0.162
CUOTA_PACTADA	-.055	.0076963 -7.14	0.000
_cons	3.578	.9043067 3.96	0.000

Fuente: Cálculos propios a través del software Stata 16.

Las 11,587 observaciones descritas en la tabla 5 son equivalentes a la cantidad de observaciones iniciales: 12,153, menos la cantidad de registros de empresas: 566 (12,153- 566 = 11,587), es decir, la base de datos de créditos otorgados a personas naturales únicamente. Así las cosas, se procede a realizar un segundo modelo con los registros de personas naturales, para validar que no se presente correlación con ninguna de las variables explicativas seleccionadas.

El primer paso para la estimación de un segundo modelo, como ya se mencionó, es realizar el proceso de selección de variables (stepwise), el cual propone el siguiente modelo logístico descrito en la tabla 6:

Tabla 6. Modelo logístico final

Logistic regression	Number of obs	=	11,587
	LR chi2(29)	=	894.94
	Prob > chi2	=	0.0000
Log likelihood = -670.00986	Pseudo R2	=	0.4004

DEFAULT	**Coef.**	**Std. Err.**	**Z**	**P>z**
R_ACIERTA	-.001	.0004997	-2.98	0.003
CUOTA_PACTADA	-.053	.0078304	-6.89	0.000
MODELOVEH_5 (2019)	-3.379	.5928813	-5.70	0.000
MODELOVEH_6 (2020)	-4.927	1.086.464	-4.54	0.000
MODELOVEH_4 (2018)	-2.647	.518579	-5.11	0.000
MODELOVEH_3 (2017)	-2.224	.4036236	-5.51	0.000
PORC_FINANCIACION	.022	.0063412	3.40	0.001
MODELOVEH_2 (2016)	-.702	.2500263	-2.81	0.005
DEPARTAMENTO_01_4 (Cundinamarca)	-.909	.3354704	-2.71	0.007
RANGO_VLRO_012_3 (Desembolsos MAYORES A 46 SMMLV)	.880	.2390122	3.69	0.000
RANGO_VLRO_012_2 (Desembolsos entre DE 32 A 45 SMMLV)	.817	.216279	3.78	0.000
CLASE_GAR_01_4 (Remolcador)	-135.146	.6460767	-2.09	0.036
EDAD	.020	.0074522	2.76	0.006
CAG_ACTIVIDAD_01234_3 (Pensionado)	-171.309	.8413541	-2.04	0.042
DEPARTAMENTO_01_6 (Otros)	-.635	.3158131	-2.01	0.044

CAG_ACTIVIDAD_01234_4 (Rentista de Capital)	-127.919	.9101117	-1.41	0.160
DEPARTAMENTO_01_7 (Putumayo)	1.172	.4648536	2.52	0.012
TIPIFICACION_DECISOR_01234_3 (Negado)	-5.715	.6702924	-8.53	0.000
TIPIFICACION_DECISOR_01234_4 (Estudio)	-5.593	.6934658	-8.07	0.000
TIPIFICACION_DECISOR_01234_2 (Aprobado)	-5.494	.683185	-8.04	0.000
TIPIFICACION_DECISOR_01234_5 (Fuera de Servicio)	-514.326	1.401.937	-3.67	0.000
DEPARTAMENTO_01_9 (Risaralda)	.365	.2512085	1.46	0.145
CAG_ACTIVIDAD_01234_2 (Independiente)	.164	.1890141	0.87	0.385
ESTADO_CIVIL_01_2 (Divorciado)	1.000	.4566516	2.19	0.028
ESTADO_CIVIL_01_3 (Separado)	-.287	.5547209	-0.52	0.605
DEPARTAMENTO_01_3 (Cauca)	.481	.4058417	1.19	0.236
CAG_ACTIVIDAD_01234_5 (Transportador)	-.292	.3252838	-0.90	0.369
DEPARTAMENTO_01_8 (Quindío)	.093	.4286622	0.22	0.827
CLASE_GAR_01_2 (Camión)	-.180	.2878223	-0.63	0.530
_cons	3.811	.8891877	4.29	0.000

Fuente: Cálculos propios a través del software Stata 16.

A diferencia del modelo inicial, este modelo trabaja sobre la población total con 29 variables explicativas, sin omitir observaciones de las variables categóricas.

5.2.1 Coeficientes de las variables explicativas

El modelo logístico estimado predice la probabilidad de incumplimiento de los solicitantes de crédito mediante el signo resultante en los coeficientes. Los cuales se pueden observar en la salida que arroja el software Stata (Anexo 1).

El resultado de las variables explicativas del modelo se clasifica entre las que aumentan y disminuyen la probabilidad de incumplimiento del solicitante del crédito:

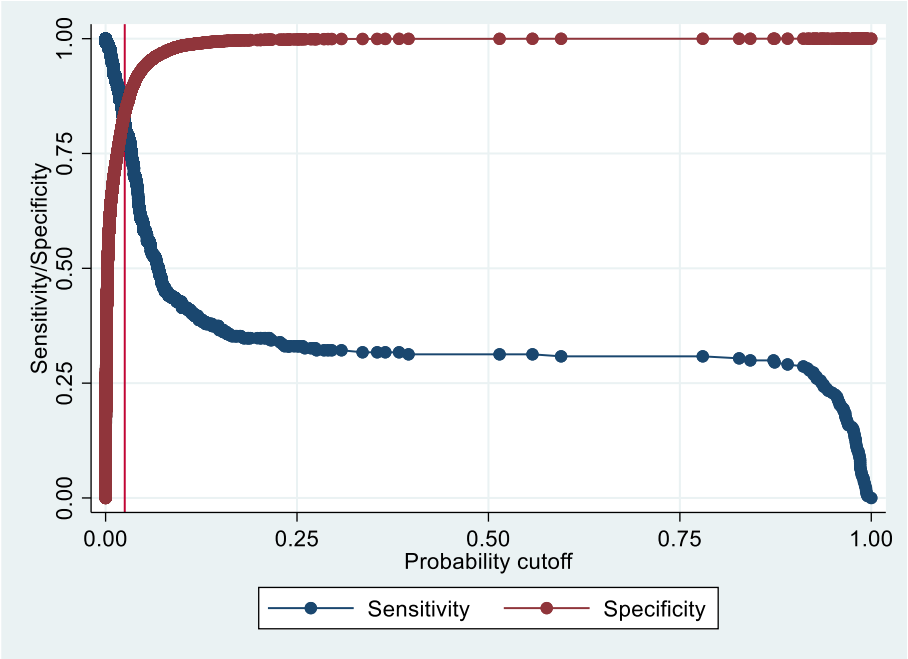
- Variables que aumentan la probabilidad de incumplimiento: PORC_FINANCIACIÓN, RANGO_VLRO_012_3, RANGO_VLRO_012_2, EDAD, DEPARTAMENTO_01_7, DEPARTAMENTO_01_3, DEPARTAMENTO_01_9, CAG_ACTIVIDAD_01234_2, ESTADO_CIVIL_01_2, DEPARTAMENTO_01_8. Estas variables indican que la probabilidad de incumplimiento del solicitante del crédito aumenta si el monto del crédito es superior a 32 SMMLV; si el porcentaje de financiación del vehículo es alto; si el cliente es divorciado; si el crédito se desembolsa en los departamentos de Putumayo, Cauca, Risaralda y Quindío; si el cliente genera ingresos como independiente y entre más años de vida tenga hay mayor probabilidad de incumplimiento.
- Variables que disminuyen la probabilidad de incumplimiento: R_ACIERTA, CUOTA_PACTADA, MODELOVEH_5, MODELOVEH_6, MODELOVEH_4, MODELOVEH_3, MODELOVEH_2, DEPARTAMENTO_01_4, CLASE_GAR_01_4, CAG_ACTIVIDAD_01234_3, DEPARTAMENTO_01_6, CAG_ACTIVIDAD_01234_4, TIPIFICACIÓN_DECISOR_01234_3, TIPIFICACIÓN_DECISOR_01234_4, TIPIFICACIÓN_DECISOR_01234_2, TIPIFICACIÓN_DECISOR_01234_5, ESTADO_CIVIL_01_3, DEPARTAMENTO_01_3, CAG_ACTIVIDAD_01234_5. Estas variables indican que un mayor puntaje de acierta disminuye la probabilidad de incumplimiento, al igual que si el cliente es pensionado, rentista de capital, transportador y está separado; los créditos que se desembolsan en los departamentos de Cundinamarca, Antioquia, Atlántico, Bolívar, Santander y Tolima, tienen menor probabilidad de incumplimiento. Las variables relacionadas con el vehículo indican que si el vehículo es un remolcador o

camión disminuye la probabilidad de incumplimiento, al igual que los modelos de los vehículos superiores al año 2016. Por su parte, las categorías de la variable Decisor, indican que independiente del resultado que arroje la herramienta, sea negativo, positivo, entra a estudio o fuera de servicio, la probabilidad de incumplimiento del cliente disminuye.

5.3 Pruebas estadísticas

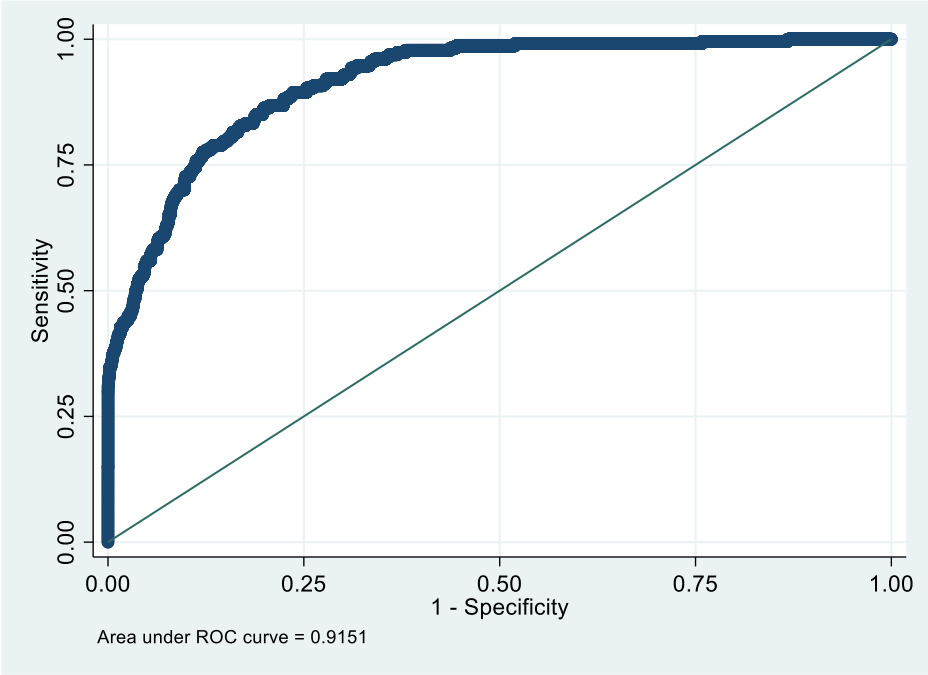
Una vez obtenido los coeficientes, se valida el punto de corte del modelo final, probando el valor donde se cruzan las curvas de sensibilidad y especificidad, donde se obtuvo un valor igual a 0.025 (Ver gráfico 1); posterior a esto, se realizan las pruebas estadísticas mencionadas en el marco teórico y se comparan los resultados para los modelos propuestos:

Gráfico 1. Punto de corte modelo logístico final



Fuente: Cálculos propios a través del software Stata 16.

Gráfico 2. Curva ROC modelo logístico final



Fuente: Cálculos propios a través del software Stata 16.

En el gráfico 2 se observa un área bajo la curva del modelo final, equivalente a 0,9151.

El área bajo la curva del modelo inicial es equivalente a 0,9141 (Ver gráfico en Anexo 2).

El área bajo la curva de ambos modelos arroja resultados cercanos a 1, indicando que hay capacidad de discriminación entre los clientes buenos y malos.

También, para medir la capacidad predictiva de los modelos, se analizó la tabla 7 de clasificación:

Tabla 7. Clasificación del modelo final

True D está definido como DEFAULT

	True		
Classified	D	-D	Total
+	185	1846	2031
-	42	9514	9556
Total	227	11360	11587

Sensitivity		81.50%
Specificity		83.75%
Positive predictive value		9.11%
Negative predictive value		99.56%
False + rate for true ~D		16.25%
False - rate for true D		18.50%
False + rate for classified	+	90.89%
False - rate for classified	-	0.44%
Correctly classified		83.71%

Fuente: Cálculos propios a través del software Stata 16.

El porcentaje de sensibilidad (sensitivity) hace referencia a los casos positivos (clientes malos), el cual indica que del total de clientes malos (227), el 81.50% (185) fueron correctamente clasificados por el modelo. Por su parte, el porcentaje de especificidad (specificity) hace alusión a los casos negativos (clientes buenos), indicando que del total de clientes buenos (11,360), el 83.75% (9,556) fueron correctamente clasificados por el modelo.

A nivel general, el porcentaje correcto de clasificación predice los datos de los clientes buenos y malos, es decir, que los datos correctamente clasificados son los 9514 más los 185, que equivalen a un 83.71% del total de los datos (11,587). Este resultado indica que el modelo es más específico que sensible y tiene más

capacidad de predecir exactamente a los clientes buenos, aunque la diferencia entre la sensibilidad y la especificidad sea únicamente del 2.25%.

5.3.1 Test de Hosmer-Lemeshow modelo final

La prueba de Hosmer y Lemeshow arrojó un nivel de significancia equivalente a $\boldsymbol{P} = \mathbf{0.5752} > \mathbf{0.05}$, es decir, que no se rechaza la hipótesis nula, indicando que el modelo se ajusta suficientemente a los datos (Ver anexo 4).

5.4 Entrenamiento y validación del modelo logístico

Con el fin de evaluar las predicciones del modelo logístico estimado, se escogió una muestra del total de la base correspondiente al 70% (8,037 observaciones), para realizar el entrenamiento del modelo, estimando nuevamente los coeficientes de la regresión logística y aplicándole la prueba de bondad de ajuste Hosmer-Lemeshow; con el 30% restante (3,550 observaciones), se realizó la validación del modelo verificando la correcta clasificación de los datos. La elección de la muestra para el 70% y 30% se realizó directamente en el software Stata. En la tabla 8 se observa el resultado del Test de Hosmer- Lemeshow para el modelo de entrenamiento:

Tabla 8. Test de Hosmer-Lemeshow modelo entrenamiento

number of observations	=	8037
number of groups	=	10
Hosmer-Lemeshow chi2(8)	=	5.89
Prob > chi2	=	0.6590

Fuente: Cálculos propios a través del software Stata 16.

Al igual que el modelo completo, la prueba de Hosmer-Lemeshow indica que el modelo de entrenamiento se ajusta suficientemente a los datos, teniendo en cuenta su valor $\boldsymbol{P} = \mathbf{0.6590} > \mathbf{0.05}$.

Este proceso de entrenamiento y validación se realizó de igual manera en el modelo inicial, donde la prueba de Hosmer-Lemeshow arrojó un resultado equivalente a $\boldsymbol{P} = \mathbf{0.0483} < \mathbf{0.05}$, indicando que se rechaza la hipótesis nula, es decir, que el modelo no se ajusta suficientemente a los datos observados. La tabla del test se encuentra en el Anexo 3.

Tabla 9. Clasificación del modelo-validación

	True		
Classified	D	~D	Total
+	51	514	565
-	15	2970	2985
Total	66	3484	3550

Sensitivity	77.27%
Specificity	85.25%
Positive predictive value	9.03%
Negative predictive value	99.50%
False + rate for true ~D	14.75%
False - rate for true D	22.73%
False + rate for classified +	90.97%
False - rate for classified -	0.50%
Correctly classified	85.10%

Fuente: Cálculos propios a través del software Stata 16.

La tabla 9 indica que para la muestra de validación (30%), el porcentaje de sensibilidad (sensitivity) indica que del total de clientes malos (66), el 77.27% (51) fueron correctamente clasificados por el modelo. Por su parte, el porcentaje de especificidad (specificity) señala que del total de clientes buenos (3,484), el 85.25% (2970) fueron correctamente clasificados por el modelo.

A nivel general, el porcentaje correcto de clasificación equivale a un 85.10% del total de los datos (3,550). Este resultado indica que el modelo es más específico que sensible, al igual que el modelo completo.

Adicional a la estimación del modelo logístico en Stata, se creó una plantilla en Excel para evaluar a los solicitantes de crédito de forma manual, con las variables que arrojó el modelo final y así obtener una probabilidad al momento del estudio del crédito en la etapa de otorgamiento (Ver anexo 5).

6. Conclusiones y recomendaciones

El modelo de scoring diseñado en este proyecto de investigación ayudó a predecir la probabilidad de incumplimiento de 29 variables explicativas que, de acuerdo con la metodología aplicada, son las más relevantes en la etapa de otorgamiento del crédito de vehículo, para personas naturales en la entidad financiera donde se ejecutó el trabajo.

La base de datos con la que se ejecutó el modelo inicial contenía 12,153 créditos de vehículos otorgados a personas jurídicas y naturales con vigencia abril 2020; sin embargo, al realizar las modelaciones se identificó que, para el diseño del modelo, el software no leía adecuadamente los créditos otorgados a personas jurídicas, siendo este hallazgo una de las principales limitaciones para realizar el modelo con la base de datos completa. En ese sentido, se evidenció que la manera en que se captura la información del cliente en el sistema al momento de realizar el estudio de crédito en esta financiera, impide una segregación adecuada para el análisis adecuado de datos, ya que no hay distinción entre las variables que se tienen en cuenta para el estudio de crédito de las personas jurídicas y las personas naturales, así el análisis del crédito se realice de forma distinta para ambos casos. Es decir, los créditos otorgados a empresas en esta financiera, no se analizan de la misma forma que los créditos otorgados a personas naturales, sin embargo, se captura la misma información del acreedor en el sistema, lo cual restringe predicciones más asertivas para modelos de riesgo de crédito.

De acuerdo con las investigaciones que se llevaron a cabo para el desarrollo de este modelo, compañías del mismo sector para la aprobación y captura de información de los créditos otorgados a personas jurídicas, tienen en cuenta variables financieras de la empresa como prueba acida, razón de endeudamiento, márgenes de utilidad, entre otras. Así las cosas, se recomienda crear una metodología distinta en el proceso de captura de datos de personas jurídicas, la

cual permita realizar análisis descriptivos asertivos, con el fin de obtener modelaciones adecuadas y así poder predecir las probabilidades de incumplimiento de las empresas solicitantes de crédito; ya que no fue posible realizar las estimaciones correspondientes con personas jurídicas porque la entidad no cuenta con esta información.

Por su parte, el modelo final arrojó la predicción de la probabilidad de incumplimiento de 11,587 créditos de vehículos otorgados a personas naturales, indicando una especificidad del 83,7%, incluyendo las variables explicativas más representativas para el otorgamiento del crédito. Dentro del análisis de los resultados de los coeficientes de estas variables, se observó que todas las categorías de la variable TIPIFICACIÓN_DECISOR_01234 (herramienta que arroja un estudio del crédito rápido al analista) disminuyen la probabilidad de incumplimiento del acreedor, independientemente del resultado que indique esta herramienta; hay créditos que se aprueban con estado de rechazo en el Decisor, lo cual indica que no es una variable determinante para la decisión final del crédito. Por lo tanto, se recomienda evaluar el uso y costo que se le está dando actualmente a esta herramienta en la financiera, analizando si realmente dentro del flujo de la operación agrega valor en la etapa de otorgamiento, o si el uso de esta herramienta puede agregar valor a otras líneas de crédito que no requieran un estudio de crédito tan minucioso.

Si la organización considera implementar este modelo de scoring para la etapa de otorgamiento de los créditos de vehículos a personas naturales, se recomienda realizar una modelación con la cartera vigente para identificar las brechas entre el modelo diseñado en este trabajo vs el modelo con la cartera vigente de personas naturales, teniendo en cuenta que este es el primer modelo que se realiza de forma experimental en la financiera. Adicionalmente, es importante evaluar la inclusión de variables macroeconómicas (tasa de desempleo, PIB, IPC) dentro de la información capturada en el sistema, y así lograr un resultado íntegro para futuros modelos estadísticos.

Referencias

Aguilar, M. Á. (2021). *Propuesta de un modelo de score de originación para la cartera de consumo de una cooperativa de ahorro y crédito del segmento 3 en el Ecuador.* Quito: Universidad Andina Simón Bolívar.

Benavides, A. V. (2017). *Curvas ROC y sus aplicaciones.* (Trabajo fin de grado inédito). Sevilla: Universidad de Sevilla.

Calderón Romero, L. (2017). *El modelo credit scoring como alternativa de evaluación crediticia en Agrobanco.* Lima: Quipukamayoc.

Carbonell, M. C. (2018). *Scoring de crédito: Herramienta para la evaluación de riesgo de crédito en entidades financieras.* Bogotá: Pontificia Universidad Javeriana.

Castro, V. M., & Noriega, M. J. (2019). *Modelo de Scoring para aprobación de créditos para la cartera de consumo en una cooperativa de aporte y crédito colombiana.* Bogotá: Pontificia Universidad Javeriana.

Duque, L. A., & Baena, D. R. (2017). *Diseño de un modelo de scoring para el otorgamiento de crédito de consumo en una compañia de financiamiento Colombiana.* Medellín: Universidad EAFIT.

Echeverri, A. V. (2017). *Modelo Scoring para el otorgamiento de crédito de las pymes.* Medellín: Universidad EAFIT.

Enchautegui, M. E. (2003). *Módulo de estudio sobre modelos probit y logit.* San Juan de Puerto Rico: Departamento de Economía Universidad de Puerto Rico.

Esteve, E. M. (2007). *Un modelo de credit Scoring basado en el conocimiento de la aplicación de Basilea II y su papel innovador en el sector bancario.* Obtenido de Dialnet: https://dialnet.unirioja.es/servlet/articulo?codigo=2499466

Fernández Castaño, H., & Pérez Ramírez, F. O. (2005). El modelo logístico: una herramienta estadística para evaluar el riesgo de crédito. Medellín: *Revista Ingenieras.*

Goh, R., & Lee, L. (2020). Una nota sobre el bosque aleatorio como el futuro. Malasia: *Journal of Analysis and Applications.*

González, L. (17 de 05 de 2019). *aprendeIA*. Obtenido de aprendeIA: https://aprendeia.com/matriz-de-confusion-machine-learning/

Gutiérrez Girault, M. A. (2007). *Modelos de credit scoring: Qué, cómo, cuándo y para qué.* Munich: Munich personal Repec archive.

Hand , D., & Henley, W. (1997). Statisticall Classification Methods in Consumer CreditScoring: a Review. Reino Unido: *Royal Statistical Society* 160(3).

Leal Fica, A., Aránguiz Casanova, M., & Gallegos Mardones, J. (2018). Análisis de riesgo crediticio propuesta del modelo scoring. Bogotá: *Revista Facultad de ciencias económicas: Investigación y reflexión.*

Ochoa, J., Galeano, W., & Agudelo, L. (2010). Construcción de un modelo de scoring para el otorgamiento de crédito en una entidad financiera. Medellín: *Perfil de Coyuntura Económica,* (16), pp. 191-222.

Pardo Carrillo, O. S. (2020). Perfil de riesgo de crédito para una cooperativa en Villavicencio a partir de un modelo Logit. Villavicencio: *Universidad & Empresa.*

Pérez, J. L. (19 de diciembre de 2013). *LA ESTADISTICA: UNA ORQUESTA HECHA INSTRUMENTO.* Obtenido de https://jllopisperez.com/2013/12/19/test-de-hosmer-y-lemeshow/

Rayo Canton, S., Lara Rubio, J., & Camino Blasco, D. (2010). Un modelo de Credit Scoring Para instituciones de microfinanzas en el marco de Basilea II. Lima: *Journal of Economics, Finance and Administrative Science.*

Rueda Pimiento, L. R., & Vergel Esteban, S. J. (2006). *Modelo financiero para riesgo de crédito de vehículo del Banco Davivienda S.A.* Bucaramanga: Universidad Autónoma de Bucaramanga.

Samaniego Medina, R. (2008). *El riesgo de crédito en el marco del acuerdo Basilea II.* Madrid: Delta publicaciones universitarias.

Schreiner, M. (2002). *Ventajas y desventajas del Scoring Estadístico para las Microfinanzas.* Washington: Microfinance Risk Management.

Superintendencia financiera (2016). *Circular externa 025.* Obtenido de https://www.redjurista.com/Documents/circular_25_de_2016_superfinanciera_-_superintendencia_financiera.aspx#/

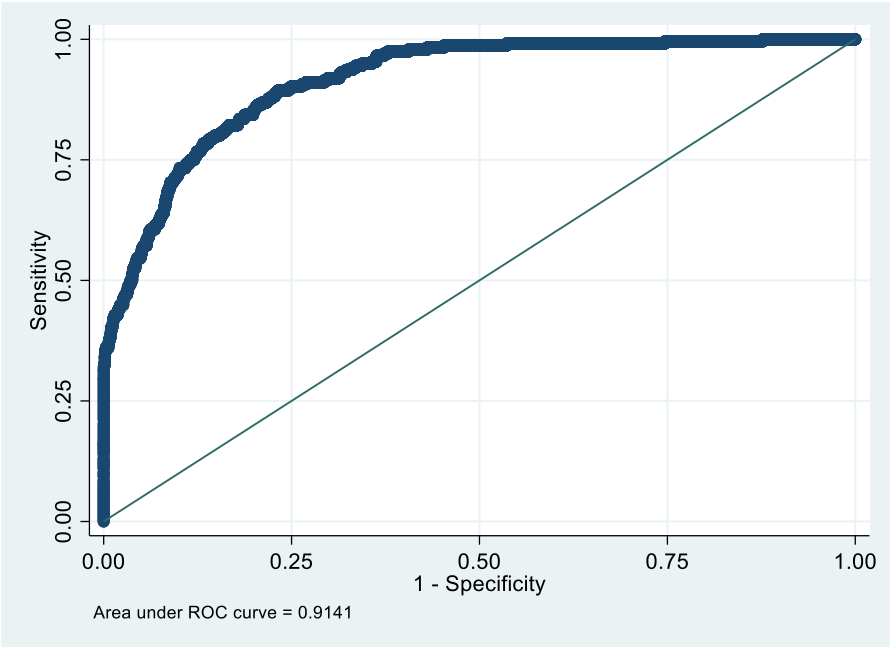
Urbina Poveda, M. A. (2017). *Determinantes del riesgo de crédito bancario: evidencia en latinoamérica.* Santiago de Chile: Universidad de Chile.

Zhang, R., & Qiu, Z. (2020). *Optimización de hiperparámetros de redes neuronales con inteligencia de enjambre: un marco novedoso para la calificación crediticia.* Shanghai: Creative Commons.

Anexos

Anexo 1. Curva ROC modelo inicial

Área bajo la curva = 0.9141



Fuente: Cálculos propios a través del software Stata 16.

Anexo 2. Tabla de clasificación modelo inicial

	True		
Classified	D	~D	Total
+	192	1851	2043
-	44	9500	9544
Total	236	11351	11587

Sensitivity	81.36%
Specificity	83.70%
Positive predictive value	9.19%
Negative predictive value	99.55%
False + rate for true ~D	16.30%
False - rate for true D	18.64%
False + rate for classified +	90.81%
False - rate for classified -	0.45%
Correctly classified	83.65%

Fuente: Cálculos propios a través del software Stata 16.

Anexo 3. Test de Hosmer-Lemeshow modelo inicial de entrenamiento

number of observations	=	8037
number of groups	=	10
Hosmer-Lemeshow chi2(8)	=	15.61
Prob > chi2	=	0.0483

Fuente: Cálculos propios a través del software Stata 16.

Anexo 4. Test de Hosmer-Lemeshow modelo final

number of observations	=	11587
number of groups	=	10
Hosmer-Lemeshow chi2(8)	=	6.65
Prob > chi2	=	0.5752

Fuente: Cálculos propios a través del software Stata 16.

Anexo 5. Plantilla Scoring Excel

NOMBRE EMPRESA

Plantilla Scoring Crédito de Vehiculo

Analisis de crédito

Código: FO-ACV-12
Versión: 01
Fecha de Elaboración:22/05/2022
Fecha de Actualización:

Dependencia: Vehiculos
Area: Crédito

R_ACIERTA	CUOTA_PACTADA	PORC_FINANCIACION	MODELOVEH_2	RANGO_VLRO_012_3	CLASE_GAR_01_4	EDAD	TIPIFICACION_DECISOR_01234_3	CAG_ACTIVIDAD_01234_2	DEPARTAMENTO_01_3	CLASE_GAR_01_2	Corte_Formula	Probabilidad	Observaciones
624	60	90	2016	1	0	52	1	1	1	1	-1,7160	0,152391279	

Tabla Modelo Scoring

Variable	Coeficiente
R_ACIERTA	-0,0015
CUOTA_PACTADA	-0,0539
MODELOVEH_5	-3,3796
MODELOVEH_6	-4,9277
MODELOVEH_4	-2,6474
MODELOVEH_3	-2,2246
PORC_FINANCIACION	0,0215
MODELOVEH_2	-0,7027
DEPARTAMENTO_01_4	-0,9094

NOMBRE EMPRESA

Plantilla Scoring Crédito de Vehiculo

Analisis de crédito

Código: FO-ACV-12
Versión: 01
Fecha de Elaboración:22/05/2022
Fecha de Actualización:

Dependencia: Vehiculos
Area: Crédito

R_ACIERTA	CUOTA_PACTADA	PORC_FINANCIACION	MODELOVEH_2	RANGO_VLRO_012_3	CLASE_GAR_01_4	EDAD	TIPIFICACION_DECISOR_01234_3	CAG_ACTIVIDAD_01234_2	DEPARTAMENTO_01_3	CLASE_GAR_01_2	Corte_Formula	Probabilidad	Observaciones
624	60	90	2016	1	0	52	1	1	1	1	-1,7160	0,152391279	

Fuente: Elaboración propia Excel, 2022.