

ESCUELA SUPERIOR POLITÉCNICA DEL LITORAL

Facultad de Ciencias Sociales y Humanísticas

Propuesta metodológica para evaluar el riesgo de impago en clientes sin referencias crediticias de una empresa distribuidora y comercializadora de productos de uso agrícola, salud humana y salud animal

PROYECTO INTEGRADOR

Previo la obtención del Título de:

Economía

Presentado por:

Michael Arístides Benavides Ramírez

Nicole Madelaine Garcés Ronquillo

GUAYAQUIL - ECUADOR

Año: 2019

DEDICATORIA

El presente proyecto lo dedico a Dios por guiarme y protegerme para llegar a este momento tan importante, a mi mami Katty Ronquillo por ser mi razón de vivir y superación, a mis hermanos Katty y Marco por ser parte de mi vida, a mis doce perritos por acompañarme siempre y ser mi felicidad, a Jorge Coronel y a Eduardo Moncayo por haber hecho el esfuerzo de llevarme a la universidad, a mi novio por su apoyo y enseñanzas constantes.

Nicole Garcés Ronquillo

El presente proyecto lo dedico principalmente a mi madre Rocío, abuelos Teresa y Arístides, Elsa y Manuel; a mis hermanos John, Anthony, y Mathías, a mi papá John y a toda mi familia que estuvo presente en esta etapa de mi vida; a mis amigos por apoyarme en temas que no necesariamente eran académicos, a mi novia por su apoyo, constancia en nuestra relación, motivación y ayuda en el proceso de estudio.

Michael Benavides Ramírez

DECLARACIÓN EXPRESA

"Los derechos de titularidad y explotación, nos corresponde conforme al reglamento de propiedad intelectual de la institución; *Michael Benavides, Nicole Garcés* y damos nuestro consentimiento para que la ESPOL realice la comunicación pública de la obra por cualquier medio con el fin de promover la consulta, difusión y uso público de la producción intelectual"



Michael Benavides Ramirez



Nicole Garcés Ronquillo

EVALUADORES



MSc. Juan Carlos Campuzano
PROFESOR DE LA MATERIA



MSc. Juan Carlos Campuzano
PROFESOR TUTOR

RESUMEN

El presente trabajo de titulación tiene como objetivo principal proponer un modelo de score de crédito para evaluar el riesgo de impago en clientes sin referencias crediticias de una empresa distribuidora y comercializadora de productos de uso agrícola, salud humana y salud animal mediante el análisis cualitativo y cuantitativo. Esto permite incrementar la cantidad de clientes que cumplan con sus obligaciones de pago dentro del plazo establecido. Existen varios métodos como los econométricos, estadísticos y actualmente los de inteligencia artificial para desarrollar el modelo que predice el comportamiento de pago. Como resultado al aplicar el modelo, se clasificó correctamente el 77% de solicitantes con un error estimado de 23%. Además, si se hubiera empleado esta metodología en el periodo de un año se hubiera reducido en promedio la cartera vencida en \$7,183,769 y se hubiera clasificado correctamente a 1309 clientes en promedio al negar el crédito. Con todo esto, se logró crear una base de datos para organizar la información proveniente de las solicitudes de crédito y establecer perfiles de los clientes.

Palabras Clave: Modelos de score de crédito, Riesgo de impago, Solicitantes, Cartera vencida.

Código JEL: C10, C13, C19, C25.

ABSTRACT

The main objective of this investigation is to propose a credit score model to evaluate the risk of non-payment in customers without credit references from a distributor and marketer of products for agricultural use, human health and animal health through qualitative and quantitative analysis. This allows to increase the number of clients that fulfill their payment obligations within the established term. There are several methods such as econometric, statistical and currently artificial intelligence to develop the model that predicts payment behavior. As a result, when applying the model, 77% of applicants were correctly classified with an estimated error of 23%. In addition, if this methodology had been used in the one-year period, the overdue portfolio would have been reduced by an average of \$ 7,183,769 and 1309 clients would have been correctly classified by denying the credit. With all this, it was possible to create a database to organize information from credit applications and establish customer profiles.

Keywords: *Credit score models, Risk of default, Applicants, Past due portfolio.*

JEL Code: C10, C13, C19, C25.

ÍNDICE GENERAL

EVALUADORES.....	4
RESUMEN.....	I
ABSTRACT	II
ÍNDICE GENERAL	III
ÍNDICE DE GRÁFICOS.....	V
ÍNDICE DE TABLAS.....	VI
CAPÍTULO 1.....	1
1. Introducción.....	1
1.1 Descripción del problema	1
1.2 Justificación del problema	2
1.3 Objetivos	2
1.3.1 Objetivo General	2
1.3.2 Objetivos Específicos.....	3
1.4 Marco teórico.....	3
1.4.1 Definición de cartera	3
1.4.1.1 Cartera normal	3
1.4.1.2 Cartera vencida.....	3
1.4.1.3 Cartera dudosa.....	3
1.4.1.4 Cartera eliminada	3
1.4.2 Divisiones de la empresa	3
1.4.3 Puntuación o score de crédito.....	4
1.4.4 Modelos de score crediticio	5
1.4.5 Score de crédito y su uso en la empres.....	7
CAPÍTULO 2.....	9
2. Metodología.....	9
2.1 Descripción de variables del modelo Logit	10
2.1.1 Tipo de contribuyentes.....	10
2.1.1.1 Personas naturales	10
2.1.1.2 Personas jurídicas.....	10
2.1.2 Divisiones de la empresa XYZ.....	11
2.1.2.1 Agro.....	11
2.1.2.2 Farma.....	11
2.1.2.3 Salud animal.....	11

2.1.3	Estado del Registro Único del Contribuyente (RUC)	11
2.1.3.1	Activo.....	11
2.1.3.2	Pasivo.....	11
2.1.3.3	Suspendido	11
2.1.4	Letra de cambio	11
2.1.5	Pagaré	11
2.1.6	Estado civil.....	12
2.1.6.1	Soltero.....	12
2.1.6.2	Casado.....	12
2.1.6.3	Divorciado	12
2.1.6.4	Viudo.....	12
2.1.7	Obligación con el Servicio de Rentas Internas del Ecuador (SRI)	12
2.1.8	Solicitud de crédito.....	12
2.1.9	Días de crédito.....	12
2.1.10	Garantías (Prenda comercial).....	12
2.1.11	Cupo otorgado.....	12
2.2	Descripción del modelo	13
2.3	Análisis univariado.....	13
2.4	Análisis multivariado.....	18
2.5	Validación.....	18
2.6	Limitaciones del modelo	21
CAPÍTULO 3.....		22
3.	Resultados Y ANÁLISIS	22
CAPÍTULO 4.....		29
4.	Conclusiones Y Recomendaciones.....	29
4.1	Conclusiones.....	29
4.2	Recomendaciones.....	29
BIBLIOGRAFÍA.....		30
APÉNDICES.....		31

ÍNDICE DE GRÁFICOS

Gráfico 1 Porcentaje de clientes con cartera vencida por división, estado del RUC y tipo de contribuyente	17
Gráfico 2 Porcentaje de clientes con cartera vencida por división.....	17
Gráfico 3 Monto total de cartera vencida por división.....	17
Gráfico 4 Área bajo la curva de ROC de los datos de prueba por validación cruzada de tipo k-folds con 5 iteraciones vs área bajo la curva del modelo usando todos los datos	19
Gráfico 5 Área bajo la curva de ROC del modelo planteado	20
Gráfico 6 Comparación de perfiles a través de la probabilidad de pago adicional en clientes	28

ÍNDICE DE TABLAS

Tabla 1 Estadísticas descriptivas de clientes con cartera vencida en la división Agro, con corte al 31 julio del 2019.....	14
Tabla 2 Estadísticas descriptivas de clientes con cartera vencida en la división Salud animal con corte al 31 julio del 2019	15
Tabla 3 Estadísticas descriptivas de clientes con cartera vencida en la división Farma, con corte al 31 julio del 2019	16
Tabla 4 Promedio de cartera vencida por división y estado del RUC	18
Tabla 5 Matriz de correlaciones de variables continuas	18
Tabla 6 Prueba de Hosmer-Lemeshow Goodness of Fit.....	21
Tabla 7 Errores estándar robustos en las variables del modelo	22
Tabla 8 Resultados de la estimación del modelo en odds-ratio.....	24
Tabla 9 Efectos marginales de las interacciones del modelo	25
Tabla 10 Efectos marginales de las interacciones del modelo	26

CAPÍTULO 1

1. INTRODUCCIÓN

Las empresas e instituciones que otorgan crédito a sus clientes enfrentan la posibilidad de que estos no cumplan con la obligación adquirida en el plazo establecido o con el pago total, peor aun cuando no poseen información crediticia. Dado ese riesgo, es importante contar con una herramienta que ayude a decidir de una forma eficaz a quien se le aprueba o no el crédito. Por este motivo, se han usado metodologías cuantitativas y cualitativas como los modelos de score crediticio para pronosticar e identificar la probabilidad de un cliente no pague en el plazo establecido o su riesgo de impago.

En los últimos años, la empresa XYZ ha evidenciado un incremento en su cartera vencida debido a las fallas que existen en el proceso de otorgamiento de crédito y a problemas macroeconómicos que enfrenta el país. Frente a ello, surge la necesidad de implementar un modelo de score de crédito para clientes sin referencias crediticias para que junto el juicio de las personas encargadas de aprobar los créditos se aumente la cantidad de clientes que cumplen con sus pagos a tiempo.

A través de la información obtenida de las solicitudes de crédito, en el presente trabajo se aplicará diferentes métodos estadísticos con las variables que inciden en la probabilidad de no pago para seleccionar el modelo de score crediticio con menor error de predicción y así predecir el comportamiento de impago de los nuevos clientes.

1.1 Descripción del problema

La empresa XYZ cuenta con varios años en el mercado y se dedica a la comercialización y distribución de diversos productos de salud humana, salud animal y agro, se caracteriza por ofrecer crédito directo a los clientes. Esta facilidad de cancelación ha sido beneficiosa para las dos partes, sin embargo, la empresa ha tenido que afrontar pérdidas por la falta de pago, siendo la consecuencia de no realizar un análisis previo para otorgar el crédito y tampoco contar con la gestión documental de las solicitudes. Con lo mencionado anteriormente, deja sin

información oportuna a los evaluadores sobre la situación actual y expectativas de pago de los clientes, por consiguiente, la empresa no reporta la situación legal de las actividades comerciales al buró de crédito. En la actualidad, una empresa privada realiza un análisis de score crediticio sobre los clientes o posibles clientes, no obstante, el análisis de los clientes frecuentes no refleja la relación comercial con la empresa porque no poseen el historial crediticio.

1.2 Justificación del problema

La falta de información, de un método cualitativo y cuantitativo para tomar decisiones, son unas de las principales razones por la que la empresa mantiene cartera eliminada y dudosa; debido a que existen irregularidades y limitaciones en el análisis de otorgamiento de crédito para los clientes. Resultado de esto la empresa con corte al 31 de julio del 2019 tiene \$4,742,277 en cartera eliminada, siendo la división “Agro” la que tiene mayor concentración de esta (\$2,855,236).

Además, 10% de los clientes están en mora y en fase previa a proceso judicial por tener una cartera vencida mayor a cinco años, con una deuda de \$8,017,287, mientras que 0.8% de los clientes están en mora y en proceso judicial por tener una cartera vencida mayor a 5 años, el monto de la deuda total de este tipo de clientes es de \$2,921,224. La cartera en estado pre legal se la ha provisionado con \$9.3 millones y \$10 millones la cartera en estado legal.

Otra categoría para considerar en la gestión de cartera de la empresa es la cartera dudosa, el monto de esta asciende a \$1,294,296, y 80% de este monto pertenece a la división “Agro”.

La propuesta de un modelo de score crediticio dotará de una herramienta que permita agilizar la toma de decisiones, obtener la probabilidad de no pago en los plazos establecidos o de tener una cartera vencida, dudosa y/o castigada para otorgar o rechazar la solicitud de crédito de forma lógica, objetiva y eficiente.

1.3 Objetivos

1.3.1 Objetivo General

Proponer un modelo de score crediticio de acuerdo con las necesidades de la empresa mediante el análisis cualitativo y cuantitativo para incrementar la cantidad de clientes que cumplan con sus obligaciones de pago dentro del plazo establecido.

1.3.2 Objetivos Específicos

1. Organizar la información personal y financiera de los clientes para estructurar una base de datos.
2. Establecer perfiles de los clientes según las variables que inciden en la probabilidad de no pago de acuerdo con las divisiones de la empresa para anticipar el nivel de cartera eliminada en los clientes con cartera normal y dudosa.
3. Comparar los principales modelos de acuerdo con la literatura a través de criterios de información para seleccionar el modelo de score crediticio con menor error de predicción.

1.4 Marco teórico

1.4.1 Definición de cartera

1.4.1.1 Cartera normal

Estado inicial de cartera que no presenta retraso en los pagos o amortizaciones.

1.4.1.2 Cartera vencida

Deuda de la porción de clientes que reportan atraso en el cumplimiento de sus obligaciones de pago.

1.4.1.3 Cartera dudosa

Deuda con alta probabilidad de no ser cobrada, una vez clasificada como eliminada, debe realizarse la respectiva provisión y llevarlas al gasto con el objetivo de reconocer financieramente su pérdida.

1.4.1.4 Cartera eliminada

Cartera mayor a cinco años con pedido formal por parte del departamento de contabilidad para eliminarse, se denomina cartera incobrable, es el valor que la empresa considera que no puede recuperar, y por tanto se debe provisionar.

1.4.2 Divisiones de la empresa

Los productos se dividen en tres categorías: agro, salud humana y salud animal. En la división “agro” se encuentra la distribución y comercialización de las semillas, agroquímicos y equipos para la fumigación, el plazo de su crédito es de 60 días promedio. En “farma” se ofrece una gran variedad de productos en líneas de insumos hospitalarios, OTC (venta libre), farmacéutico y fórmula infantil, con

plazo de pago promedio de 45 días. En la categoría “salud animal” están los productos de cuidado animal y cuenta con crédito de 40 días promedio de plazo.

1.4.3 Puntuación o score de crédito

La puntuación o score de crédito se introdujo por primera vez en la década de 1940 y a través de los años ha evolucionado y desarrollado significativamente. En la década de 1960, con la creación de tarjetas de crédito, bancos y otros emisores de tarjetas de crédito se tomó en cuenta de las ventajas de la calificación crediticia en el proceso de otorgamiento de créditos (Tan, Goh, & Koh, 2006).

Tan, Goh & Koh (2006) sostienen que en la década de 1980 la calificación crediticia se usó para otros fines, como ayuda en la decisión de aprobar solicitudes de préstamos personales y en recientes años se ha utilizado para préstamos de vivienda, pequeñas empresas solicitudes y renovaciones de préstamos y seguros. Un modelo de puntuación de crédito proporciona una estimación del riesgo de crédito del prestatario, es decir, “la probabilidad de que el prestatario reembolsará el préstamo según lo prometido, basado en una serie de características cuantificables del prestatario” (Kleimeier & Dinh, 2007).

La puntuación de crédito es el conjunto de modelos de decisión y su subyacente que son de gran ayuda para los prestamistas en la concesión de créditos al consumo, deciden quién obtendrá crédito, cuánto crédito deben obtener, y qué estrategias operativas mejorarán la rentabilidad del prestatario para el prestamista (Thomas, Edelman, & Crook, 2002).

Un acreedor es capaz de generar ingresos cuando predice con éxito la solvencia crediticia y el riesgo de incumplimiento de los solicitantes dependiendo de los factores predictores por defecto. La calificación crediticia es una técnica adecuada que conecta estos factores con la probabilidad de incumplimiento (Lieli & White, 2010)

Un buen modelo de score de crédito es aquel que identifica como buenos a los clientes que cumplen con regularidad el pago del crédito, e identifica como malos a aquellos que no fueron capaces de pagar. El modelo cuenta con un margen de error, debido a esto, es posible que califique como buenos a clientes que han entrado en default o insolvencia y después de un tiempo determinado la deuda se convierte en incobrable, por ello es importante realizar análisis adecuados que permitan minimizar dichos errores.

1.4.4 Modelos de score crediticio

En la década de los 70's se comenzó a evaluar el riesgo de crédito mediante la utilización de modelos Scoring para conocer las probabilidades de default y clasificar correctamente a los deudores y solicitantes de crédito, pero esta técnica se popularizó en los años 90's con los avances y mejoras de recursos estadísticos, computacionales y debido a la necesidad que presentaba la industria bancaria en el estudio de crédito de ser más eficaz y eficientes. A pesar de contar con diversos modelos scoring, el juicio humano sigue siendo utilizado mediante reglas que la entidad aplica de manera sistemática para filtrar solicitudes o deudores (Girault, 2007)

En los últimos años los modelos de "credit scoring" también denominados "score-cards" o "clasifiers" han sido utilizados para la medición y control del riesgo dado que son "métodos estadísticos utilizados para clasificar a los solicitantes de crédito, o incluso a quienes ya son clientes de la entidad evaluadora, entre las clases de riesgo bueno o malo" (Hand & Henley, 1997).

Históricamente, el análisis discriminante y la regresión lineal han sido las técnicas más utilizadas para construir estos modelos. Otras técnicas incluyen regresión logística, análisis probit, métodos no paramétricos de alisado especialmente k-vecinos más cercanos, programación matemática, modelos de cadena de Markov, partición recursiva, sistemas expertos, algoritmos genéticos y redes neuronales (Hand & Henley, 1997).

Según B.W. Yap et al (2011) las redes neuronales artificiales (ANN) han sido criticadas por su enfoque de "caja negra" y dificultades interpretativas. Algunas de las técnicas desarrolladas recientemente para la construcción modelo de puntuación de crédito son: regresión adaptativa multivariada por splines (MARS), árbol de clasificación y regresión (CART), razonamiento basado en casos (CBR) y máquina de vectores de soporte (SVM).

Para identificar el proceso de referenciación de los modelos usados para el score de crédito es importante tener en cuenta la explicación de algunos autores sobre los métodos usados discriminando los usos y los enfoques en que se desenvuelven.

García & Sánchez (2005) sostienen que según la literatura los modelos de score de crédito se pueden clasificar en dos tipos: modelos tradicionales y modelos de

enfoque moderno. Entre los modelos tradicionales se destaca el modelo denominado experto que analiza el carácter, capital, capacidad, colateral y ciclo, con esa información el experto según su juicio determina si se otorga o no el crédito, pero es difícil de aplicar por las diferencias que existen entre clientes y es costoso para las empresas.

Por el contrario, los modelos de enfoque moderno son de aplicación matemática, como los modelos Z Score, modelos Z, modelos de respuesta binaria, Creditmetrics y VarR (Value at Risk) (García & Sánchez, 2005).

En cambio, Mpofu & Mukosera (2012) manifiestan que los modelos se pueden dividir en dos métodos diferentes: basados en estadísticas y métodos basados en aprendizaje automático e Inteligencia Artificial. Algunos ejemplos de modelos de score de crédito basados en estadísticas son: análisis discriminante lineal, árboles de decisión, análisis de cadena de Markov, análisis probit y regresión logística. Por otra parte, los métodos basados en Inteligencia Artificial que los investigadores sugieren y exploran son redes neurológicas artificiales, algoritmos genéticos y sistema inmune artificial. El enfoque de la Inteligencia Artificial se utiliza para generar la mejor combinación de pesos que puede producir el mayor porcentaje de correcta predicción, esto se logra al identificar el conjunto de factores que se utilizará y el peso para cada factor se establecerá de forma aleatoria. Las redes neuronales artificiales son técnicas computacionales que presentan un modelo matemático basado en la estructura neural de organismos inteligentes, una red neuronal artificial procesa ciertas características y produce respuestas como las del cerebro humano.

Mueres et al. (2005) realizaron un estudio empírico con la finalidad de determinar el modelo de scoring que mejor discrimine entre clientes morosos y no morosos, utilizando dos técnicas estadísticas: regresión logística y análisis discriminante. Para ello utilizaron una muestra de clientes de entidades financieras de Castilla y León que incumplieron sus obligaciones de pago y clientes que pagaron sus deudas a su vencimiento. Las variables para ambos modelos fueron seleccionadas con el método stepwise o pasos sucesivos. Obteniendo como resultado que los dos modelos tienen una capacidad predictiva del 97,1%. Diferenciándose en la predicción de clientes buenos de malos, puesto que el análisis discriminante predice el 100% los buenos clientes y 88,9% los malos. Por otro lado, la regresión logística predice el 98,1% y 94,4% respectivamente.

Comparando otros modelos “los modelos de variable dependiente limitada constituyen un mejor mecanismo que aquellos de regresión lineal, debido a que usualmente la variable dependiente es discreta (clientes buenos y malos)” (Espin-García & Rodríguez-Caballero, 2013). Sin embargo, Henley (1995) encontró que la regresión logística superó a la regresión lineal pero no por mucho como se esperaba. Esto se debe a que una gran cantidad de créditos de la muestra se encontraban entre los cuantiles 0,2 y 0,8 en los que ambas distribuciones son prácticamente idénticas. A pesar de que la preocupación en el riesgo de créditos es el comportamiento de las colas, es común seguir realizando el modelaje vía regresiones logísticas.

Baesens (2003) realizó un estudio a partir de ocho bases de datos usando ocho métodos diferentes y 17 modelos para evaluar la precisión del scoring. Los métodos y modelos fueron: dos modelos lineales, regresión logística, programación lineal, cuatro diferentes variantes de máquinas de soporte vectorial, cuatro diferentes árboles de decisión, dos variantes de la técnica de vecinos más cercanos (nearest neighbours), redes neuronales y dos técnicas bayesianas de segmentación. El autor confirmó que el scoring menos eficiente se consiguió con el método de Naïve Bayes y los mejores resultados se obtuvieron con regresiones logísticas, redes neuronales o árboles de clasificación.

Según Bolton (2010), el método por regresión logística en la elaboración de un score de crédito es mejor en comparación con otros métodos, es fácil de elaborar, de ejecutarlo y esta técnica ha sido ampliamente reconocida como método de elección por parte de la industria bancaria.

1.4.5 Score de crédito y su uso en la empres

Las entidades que desean evaluar y gestionar el riesgo de crédito en sus clientes siempre deberán considerar variables de tipo cualitativo y cuantitativo, pero dependiendo de las necesidades y del uso que se quiera dar. Dichas variables son independientes del tipo de empresa a la que se le hará el estudio (Thomas, Edelman, & Crook, 2002).

“Los estudios de personas naturales toman variables enfocadas a las características de la persona, la descripción del ser resultando en una infinidad de variables cualitativas (por ejemplo: genero, estrato, municipalidad, raza),

o cuantitativas (por ejemplo: ingresos, egresos, patrimonio, salario)” (Rodríguez, Becerra, & Cardona, 2017).

Gonçalves & Braga (2008) sostienen que un estudio de score de crédito para una persona jurídica, se usa variables de indicadores financieros: la razón corriente, la liquidez, activos, encaje bancario, cobertura en pasivos, provisión, rentabilidad de activos y patrimonio entre otros.

En el caso de desarrollar el modelo de score de crédito para clientes sin referencias crediticias, la información proveniente de las solicitudes de crédito servirá para crear bases de datos y así poder enmarcar el perfil sociodemográfico del cliente. El primer filtro que se debe aplicar sobre el total de clientes debe ayudar a eliminar a los que cuenten con un estatus de inactividad dentro del portafolio de la institución financiera (Espin-García & Rodríguez-Caballero, 2013). Espin & Rodríguez (2013) mencionan que en la información sociodemográfica proveniente de las solicitudes de crédito es muy probable encontrar falsedad en la información, como ingresos demasiados altos, número de dependientes económicos elevado, entre otros, dado esto se recomienda considerar ciertos casos como anormales para evitar complicaciones en los análisis posteriores.

CAPÍTULO 2

2. METODOLOGÍA

La empresa XYZ ofrece crédito directo a sus clientes y su aprobación de dicho valor se da sin previo análisis y sin la correcta gestión documental de la solicitud, por lo que los analistas de esta área no cuentan con la información suficiente y oportuna sobre la situación actual y expectativas de pago de los clientes en consecuencia enfrenta grandes pérdidas, a ello se suma que contrata los servicios de una empresa privada quien realiza un análisis de score crediticio pero la calificación otorgada a los clientes no refleja su situación comercial.

Se realizó un análisis cualitativo y cuantitativo para incrementar la cantidad de clientes que cumplan con sus obligaciones de pago dentro del plazo establecido, que permitió seleccionar el modelo de score crediticio de acuerdo con las necesidades de la empresa.

Para este estudio se tomó en consideración los clientes sin referencias crediticias, se utilizó una base de datos con 3663 registros proporcionada por la empresa XYZ sobre los campos presentes en la solicitud de crédito que el cliente presenta antes de otorgarle o negarle el crédito, con fecha de corte al 31 de julio del 2019, se seleccionaron aquellos registros que tenían información en todos los campos.

Se depuró la base de estudio a 3658 registros, debido a que no contaban con suficiente información, de estos 1700 eran clientes que tienen cartera vencida, correspondiente al 46.47% de la muestra, mientras que 1958 eran clientes que no tenían cartera vencida representa el 53.53% del total de registros.

La base se diseñó con el objetivo de permitir que las observaciones estén balanceadas, la muestra cuenta con 1097 personas jurídicas (28%) y 2561 personas naturales (72%), porcentajes similares a los de la población de estudio, además toda la información está en la misma línea temporal de esta forma todos los registros o clientes son comparables y se define si el cliente tiene cartera vencida de acuerdo con el histórico de pagos de la empresa.

Para la elaboración de la variable dependiente binaria (Cartera vencida), se generó una nueva variable que tomó el valor de 1 si el cliente tiene una cartera vencida mayor a cero, y 0 si en caso contrario.

Los datos se obtuvieron de fuentes secundarias internas, específicamente de las solicitudes de créditos, este es un requisito previo que debe llenar el cliente. Con esa información, se procedió a digitalizarla y conformar la base de datos con una muestra aleatoria de 3658 clientes. Se encontró problemas con algunas variables numéricas ya que contenían respuestas dicotómicas y algunas variables categóricas con respuestas numéricas, por lo que se procedió a realizar el cambio correspondiente.

Se organizó la información personal y financiera de los clientes obtenida en las solicitudes de crédito para estructurar una base de datos mediante el Excel versión 2016 y se ingresaron los datos al programa Stata 15, de acuerdo con los resultados obtenidos en el modelo seleccionado se procedió a establecer perfiles de los clientes según las variables que incidieron en la probabilidad de no pago de acuerdo con las divisiones de la empresa.

Luego de ingresado los datos al Stata 15 se comparó varios modelos logit, probit y análisis discriminante lineal con distintas variables y se seleccionó el modelo de score crediticio con menor error de predicción, según el criterio de información menor y la mayor área bajo la curva de ROC. El seleccionado fue un modelo logit con las variables descritas anteriormente, este fue comparado y analizado con el modelo actual de la empresa que se pretendía usar como complemento en la toma de decisiones para seleccionar a los clientes que solicitan un crédito.

2.1 Descripción de variables del modelo Logit

2.1.1 Tipo de contribuyentes

2.1.1.1 Personas naturales

Según el Servicio de Rentas Internas del Ecuador (SRI) “las personas naturales son todas las personas nacionales y extranjeras que realizan actividades económicas en nuestro país obligadas o no a llevar contabilidad” (Ecuador, 2015).

2.1.1.2 Personas jurídicas

Según el Servicio de Rentas Internas del Ecuador (SRI) “las personas jurídicas son las sociedades, tales como las instituciones del sector público, las personas jurídicas bajo control de la Superintendencias de Compañías y de Bancos, las organizaciones sin fines de lucro, entre otras” (Ecuador, 2015).

2.1.2 Divisiones de la empresa XYZ

2.1.2.1 Agro

Se encuentra la distribución y comercialización de las semillas, agroquímicos y equipos para la fumigación con plazo de crédito promedio de 60 días.

2.1.2.2 Farma

Se ofrece una gran variedad de productos en líneas de insumos hospitalarios, OTC (venta libre), farmacéutico y fórmula infantil, con plazo de pago de 45 días promedio.

2.1.2.3 Salud animal

Diversos productos de cuidado animal y cuenta con crédito de 40 días promedio de plazo.

2.1.3 Estado del Registro Único del Contribuyente (RUC)

2.1.3.1 Activo

“Se denomina activo en el momento en que se inscribe en el RUC y desde ese instante deberá cumplir con las obligaciones tributarias” (Ecuador, 2015).

2.1.3.2 Pasivo

“Si el contribuyente fallece, el RUC se cancela y se ubicará en estado pasivo” (Ecuador, 2015).

2.1.3.3 Suspendido

“Si el contribuyente suspende temporalmente su actividad deberá comunicar al SRI para que su registro pase a un estado de suspensión” (Ecuador, 2015).

2.1.4 Letra de cambio

Es un documento mercantil en el cual la persona denominada librador da la orden de pago de una determinada cantidad en una fecha establecida al librado.

2.1.5 Pagaré

Es un documento en el que una persona se compromete a pagar una determinada cantidad antes de la fecha de vencimiento.

2.1.6 Estado civil

2.1.6.1 Soltero

Se denomina soltero a la persona que no ha contraído matrimonio.

2.1.6.2 Casado

Estado civil de la persona que ha contraído matrimonio.

2.1.6.3 Divorciado

Persona que legalmente ha dado por terminado el vínculo matrimonial.

2.1.6.4 Viudo

Persona que no ha vuelto a casarse después del fallecimiento del cónyuge.

2.1.7 Obligación con el Servicio de Rentas Internas del Ecuador (SRI)

“La obligación tributaria se entiende como el vínculo jurídico, personal, existente entre el Estado y los contribuyentes, por medio de la cual se debe aportar a la sociedad a través del pago de los impuestos y el cumplimiento voluntario de las obligaciones tributarias que existen en nuestro país” (Ecuador, 2015).

2.1.8 Solicitud de crédito

Solicitud realizada por la empresa que debe llenar la persona que desee que le otorguen un crédito.

2.1.9 Días de crédito

Plazo máximo otorgado para pagar la deuda. La división agro, farma y salud animal cuentan con 60, 45 y 40 días de plazo promedio respectivamente.

2.1.10 Garantías (Prenda comercial)

Contrato en el que un bien mueble se da como garantía de una obligación adquirida que, al momento de no cumplir con la obligación, la persona tendrá el derecho de aquel bien.

2.1.11 Cupo otorgado

Cantidad máxima de dinero que se le otorga a la persona para que compre productos de la empresa.

2.2 Descripción del modelo

El modelo propuesto es el siguiente:

$$Prob(Carteravencida=1) = \frac{1}{1 + e^{B_1x_1 + B_2x_2 + \dots + B_ix_k}} + E_i \quad (1)$$

Donde, $i=3,4, 5 \dots 9$; $k=4, 5, 6 \dots 11$

x_1 : Variable binaria de "Persona", 1 si es persona natural y 0 si es jurídica

x_2 : Variable categórica de "División", 1 si es agro, 2 si es farma y 3 si es salud animal

x_3 : Variable categórica del "estado de Ruc", 1 si es activo, 2 si es pasivo, 3 si está suspendido, 4 si no posee.

x_4 : Variable binaria de "Letra de cambio", 1 si tiene y 0 si no tiene

x_5 : Variable binaria de "Pagaré", 1 si tiene y 0 si no tiene

x_6 : Variable categórica de "Estado civil", 1 soltero, 2 casado 3 divorciado y 4 viudo

x_7 : Variable binaria de "Obligación con el SRI", 1 si tiene y 0 si no tiene.

x_8 : Variable binaria de "Solicitud llena", 1 si la llenó y 0 si no la llenó

x_9 : Variable continua de "Dias de crédito"

x_{10} : Variable continua de "Garantías (Prenda commercial)"

x_{11} : Variable continua del "Cupo otorgado"

2.3 Análisis univariado

Se realizó el análisis univariado de las principales variables que aparecen en el modelo propuesto.

A continuación, se presentan las estadísticas descriptivas de las variables de interés por divisiones de la empresa, enfocándonos principalmente en los clientes que tienen cartera vencida.

Tabla 1 Estadísticas descriptivas de clientes con cartera vencida en la división Agro, con corte al 31 julio del 2019

Variables	Media	Desviación estándar	Máximo	Mínimo	Observaciones
Cartera vencida	35355.61	139000	249000	0.010	988
Tipo contribuyente (Binaria)	0.70	0.45	1	0	988
Letra de cambio (Binaria)	0.11	0.31	1	0	988
Pagaré (Binaria)	0.53	0.49	1	0	987
Obligaciones con el SRI (Binaria)	0.15	0.36	1	0	988
Solicitud de crédito (Binaria)	0.91	0.27	1	0	987
Días de crédito	57.52	45.61	210	0	988
Garantía	7889.51	39909.55	600,000	0	988
Cupo	28921.40	75011.09	1,000,000	0	988

Lo más relevante de estos datos, es la media de cartera vencida en la división Agro, y el contraste que tiene con la media del cupo otorgado, además los días promedio de crédito son 58, los mismos que varían por líneas de negocio (tipo de negocio) y por canales. Otra observación relevante es que el 70.4% de las observaciones de esta división son personas naturales, mientras que el restante son personas jurídicas.

Además, el 52.8% de los clientes presentaron un pagaré, el 15.5% tiene obligaciones con el SRI y el 91.6% presentaron solicitud de crédito.

Tabla 2 Estadísticas descriptivas de clientes con cartera vencida en la división Salud animal con corte al 31 julio del 2019

Variables	Media	Desviación estándar	Máximo	Mínimo	Observaciones
Cartera vencida	6375.69	39208.19	740592.8	0.01	482
Tipo contribuyente (Binaria)	0.76	0.42	1	0	482
Letra de cambio (Binaria)	0.03	0.18	1	0	482
Pagaré (Binaria)	0.21	0.40	1	0	482
Obligaciones con el SRI (Binaria)	0.15	0.36	1	0	482
Solicitud de crédito (Binaria)	0.92	0.25	1	0	482
Días de crédito	27.42	27.52	360	0	482
Garantía	1804.68	10959.06	200000	0	482
Cupo	25087.88	194000	4,000,000	0	482

Con respecto a esta división, la media de la cartera vencida es de \$1695 USD, y el contraste que tiene con la media de cupo otorgado es diferente al de la división agro, debido a que en este caso la media del cupo otorgado es mayor que la media de cartera vencida, los días promedio de crédito son 55, estos varían de acuerdo con la línea de negocio (tipo de negocio) y por canales.

El 75% de las observaciones de esta división son personas naturales, el restante son jurídicas.

Además, el 51% de los clientes presentaron un pagaré, el 10% tiene obligaciones con el SRI y 93% presentaron solicitud de crédito.

Tabla 3 Estadísticas descriptivas de clientes con cartera vencida en la división Farma, con corte al 31 julio del 2019

VARIABLES	Media	Desviación estándar	Máximo	Mínimo	Observaciones
Cartera vencida	1695	3317.31	23938.20	0.010	230
Tipo contribuyente (Binaria)	0.75	0.43	1	0	230
Letra de cambio (Binaria)	0.06	0.24	1	0	230
Pagaré (Binaria)	0.51	0.50	1	0	230
Obligaciones con el SRI (Binaria)	0.10	0.30	1	0	230
Solicitud de crédito (Binaria)	0.93	0.24	1	0	230
Días de crédito	55.10	25.59	120	0	230
Garantía	653.95	1454.06	10000	0	230
Cupo	5673.08	10640.43	12000	0	230

La muestra de clientes de la división Farma tiene una cartera vencida promedio de \$6376 USD, al igual que la división de salud animal la media de cupo otorgado es mayor que la media de cartera vencida, los días promedio de crédito son 27, al igual que todas las divisiones de la empresa, varían de acuerdo con la línea de negocio (tipo de negocio) y por canales.

El 76% de las observaciones de esta división son personas naturales, el restante son jurídicas.

Además, en este caso, solo el 21% de los clientes presentaron un pagaré; el 15% tiene obligaciones con el SRI y el 93% presentaron solicitud de crédito.

Un análisis del estado del RUC en los clientes mostró la falta de control y seguimiento de los clientes a los que se le otorga crédito, esto en mención a que existen clientes con cartera vencida que actualmente tienen el RUC suspendido e incluso no cuentan con información en Servicio de Rentas Internas, esta información se puede observar en el gráfico 1.

Con respecto al porcentaje de cartera vencida y el monto por división, en agro el 52% de sus clientes tienen cartera vencida con un monto total de \$5,800,000, en salud animal, el 49% de sus clientes tienen cartera vencida con un monto total de \$2,798,000, mientras que en la división farma, el 37% de los clientes tienen cartera

vencida con un monto total de \$746,366 tal como lo muestra el gráfico 2 y 3 respectivamente.

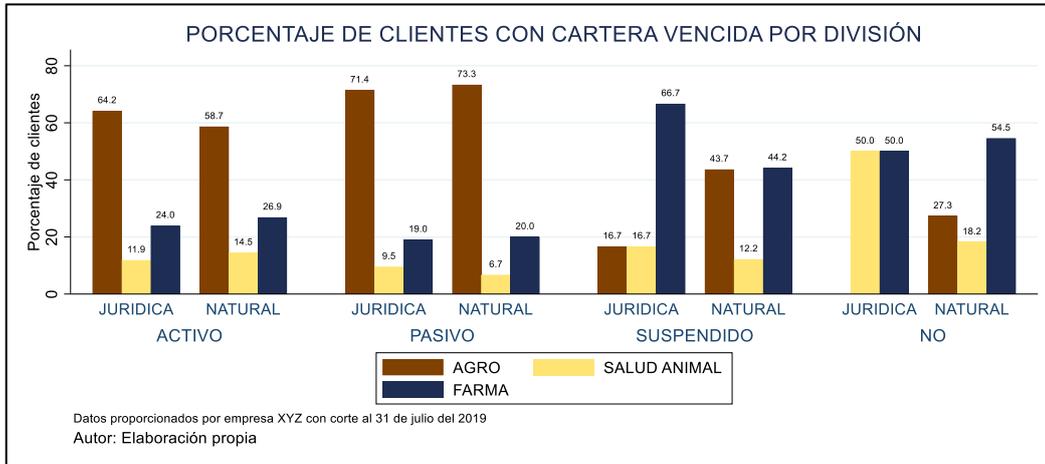


Gráfico 1 Porcentaje de clientes con cartera vencida por división, estado del RUC y tipo de contribuyente

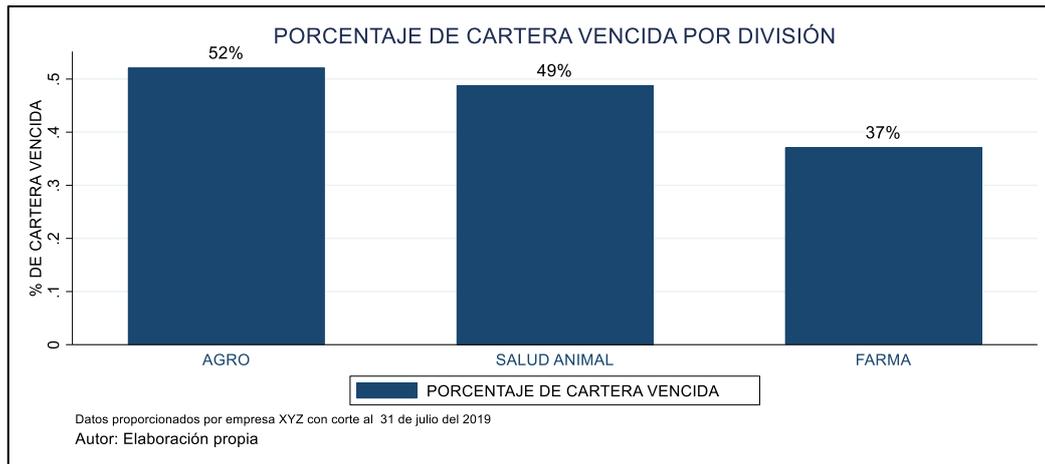


Gráfico 2 Porcentaje de clientes con cartera vencida por división

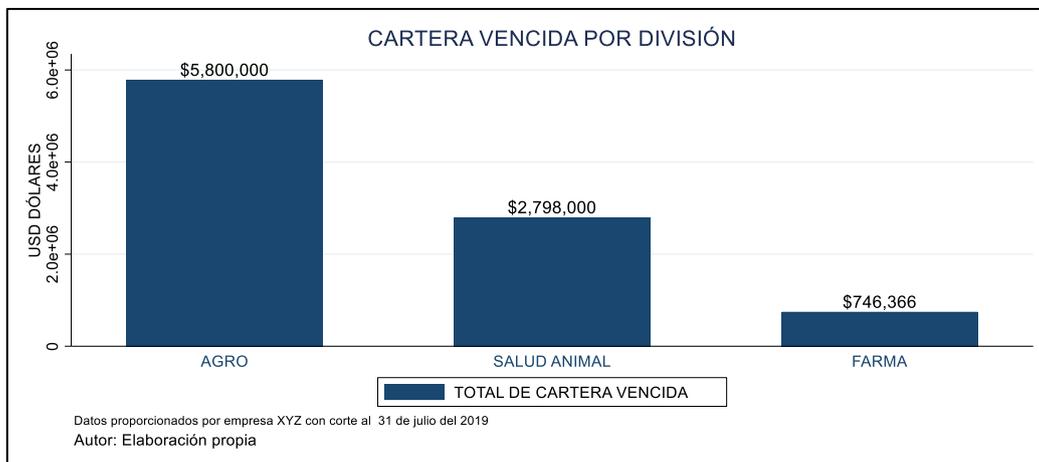


Gráfico 3 Monto total de cartera vencida por división

Con respecto al monto de cartera vencida según el estado del RUC, el promedio de cartera vencida en clientes de la división agro con RUC suspendido es de \$13886.20, en clientes de salud animal es de \$1130.31, mientras que en la división farma es de \$1811.33, más detalles en la tabla 4.

Tabla 4 Promedio de cartera vencida por división y estado del RUC

Estado de RUC	División		
	Agro	Salud animal	Farma
Activo	38133.66	1669.77	7638.73
Pasivo	18024.80	6778.32	975.31
Suspendido	13886.20	1130.31	1811.33
No	683.57	2660.58	3658.43

2.4 Análisis multivariado

En esta sección se presenta la matriz de las correlaciones de las variables continuas.

Tabla 5 Matriz de correlaciones de variables continuas

	Cartera vencida	Días de crédito	Garantía	Cupo
Cartera vencida	1			
Días de crédito	0.1481	1		
Garantía	0.2372	0.0971	1	
Cupo	0.3826	0.1911	0.2711	1

La correlación mayor, es la que existe entre la cartera vencida y el cupo otorgado, seguida por la existente entre garantía y cupo; sin embargo, no representa un problema de correlación alta entre variables independientes o multicolinealidad dado que no es una correlación muy alta, mayor a 0.50.

2.5 Validación

Para validar el poder predictivo se usó validación cruzada k-fold de 5 iteraciones, este método de validación consiste en dividir o segmentar el conjunto de datos en 5 grupos, en este caso (k grupos, en general) de igual o similar tamaño.

Se utilizaron 4 subconjuntos para el desarrollo del modelo y el subconjunto restante se utiliza como grupo de validación, se repitió este método 5 veces hasta

que cada uno de los subconjuntos haya sido utilizado una vez como grupo de validación, al final del proceso se obtiene una medida del error o una medida de rendimiento, el cual es determinado como el promedio de las 5 medidas calculadas.

Los resultados de la validación cruzada con k-folds demostró que el modelo tiene un buen poder predictivo, y se lo puede catalogar como un buen modelo debido a que el área bajo la curva de ROC del modelo probado en los datos de prueba de las 5 iteraciones supera el 0.70, en el gráfico 4 se puede confirmar lo mencionado.

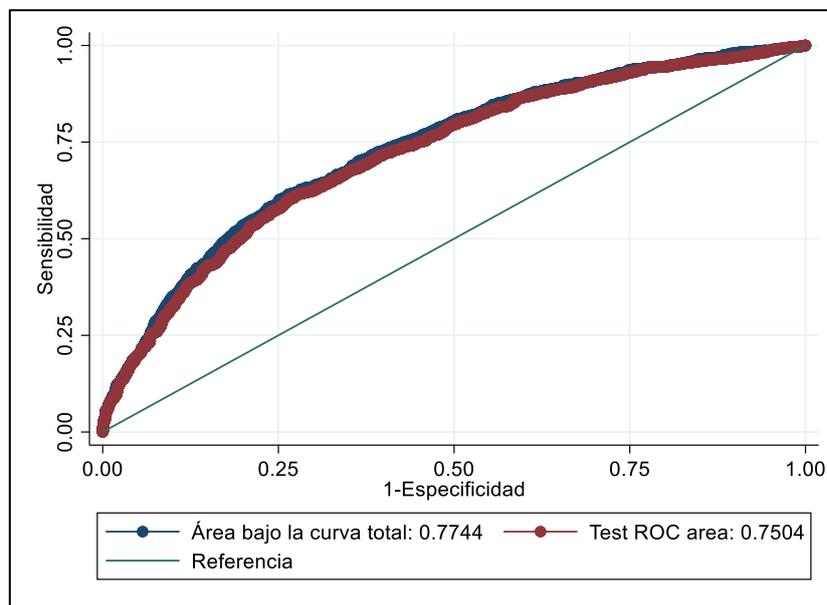


Gráfico 4 Área bajo la curva de ROC de los datos de prueba por validación cruzada de tipo k-folds con 5 iteraciones vs área bajo la curva del modelo usando todos los datos

La validación permite evaluar los dos aspectos principales de la eficiencia de un modelo predictivo: la discriminación y calibración.

La discriminación se refiere a la capacidad del modelo para clasificar correctamente a los individuos que tienen cartera vencida y a los que no tienen, en este caso la discriminación del modelo presentado se realizó mediante el área bajo la curva de ROC (Receiver Operating Characteristic), esta refleja la probabilidad de que observaciones seleccionadas al azar sean asignadas correctamente al grupo que pertenecen (clientes con cartera vencida o no), o en otras palabras asigne la probabilidad más alta al individuo que realmente tiene cartera vencida.

En el gráfico 5, se presenta el área bajo la curva de ROC del modelo planteado.

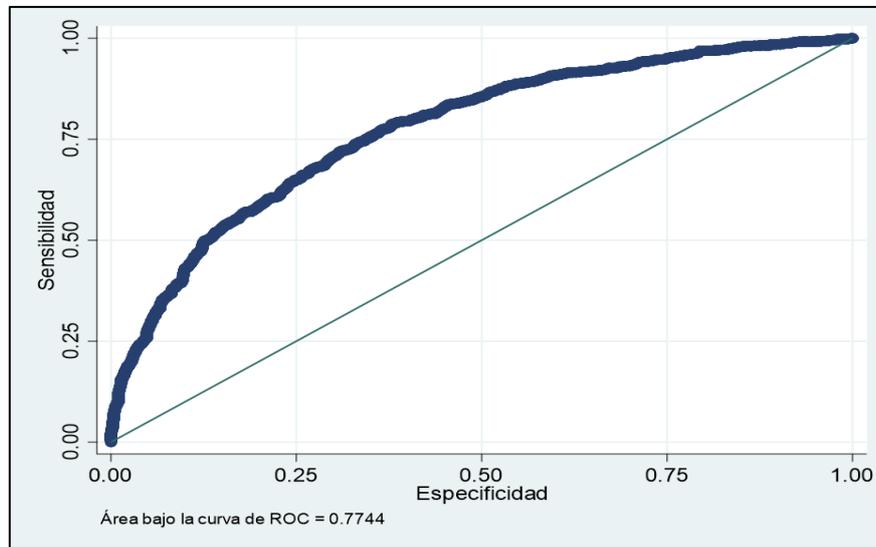


Gráfico 5 Área bajo la curva de ROC del modelo planteado

En este caso, el área bajo la curva de ROC es de 0.7744, se considera aceptable de acuerdo con el criterio general, en donde se considera que un modelo es aceptable si su área bajo la curva de ROC es mayor a 0.7, es considerado bueno si su área bajo la curva es mayor a 0.8 y muy bueno si su área bajo la curva es mayor a 0.85.

Sin embargo, hay que tener cuidado, porque un área bajo la curva de ROC en el modelo de desarrollo o con datos de entrenamiento, no necesariamente puede ser bueno, sino que puede ser señal de que el modelo está sobre ajustado.

Con respecto a la calibración, este da un diagnóstico o relación entre las predicciones estimadas por el modelo y los resultados, la prueba estadística que se usó para evaluar la calibración del modelo es la prueba de Hosmer-Lemeshow Goodness of Fit (GOF), si el ajuste del modelo es bueno (Hipótesis nula) un valor alto de la probabilidad predicha se asociará con los clientes clasificados correctamente con cartera vencida.

En el presente modelo el p-value de la prueba fue de 0.6895, tal como lo indica la tabla 6.

Tabla 6 Prueba de Hosmer-Lemeshow Goodness of Fit

Tabla contraída en cuantiles de probabilidades estimadas	
Número de observaciones =	3658
Número de grupos =	5
Hosmer-Lemeshow chi2(5) =	3.07
Prob > chi2 =	0.6895

2.6 Limitaciones del modelo

Hay que considerar que para tener un modelo robusto o más real se debe considerar dentro la muestra a personas o clientes que llenaron su solicitud de crédito pero que fueron negadas, para incluir el comportamiento de este tipo de clientes y ver sus características, además debemos tomar en cuenta que este modelo predecirá el riesgo de impago en clientes con características similares a los de la muestra, por ejemplo, si se desea agregar otra división a la empresa, el poder predictivo de este modelo será muy restringido a las divisiones ya existentes y a clientes cuyas características difieran a la de los clientes incluidos en la muestra, además este análisis solo es permitido a clientes que presentan completa la solicitud de crédito.

El riesgo considerado en este modelo está relacionado con características observables, a pesar de que el modelo sea eficiente hay que considerar otras variables latentes momento de emitir un juicio, además se supone que las relaciones entre las variables son lineales, porque establecer relaciones no lineales entre las variables puede ser un proceso costoso.

Otra debilidad importante es el horizonte de tiempo en el que se desarrolló el modelo o en el que está la muestra, porque un modelo validado, con buena calibración y discriminación en momentos donde la economía es estable es muy probable que no sea eficiente en evaluar el riesgo de impago de clientes en momentos donde se esté viviendo una crisis, es decir es muy probable que las condiciones o variables de los clientes hayan cambiado respecto a las condiciones iniciales, esto pasa debido a que no se consideran las variables coyunturales del país.

Otra limitación significativa es que el modelo responderá a un riesgo relativamente alto de impago debido al tipo de clientes que por lo general no tienen referencias

crediticias tales como: agricultores, microempresarios, emprendedores, entre otros.

Por último, es imposible considerar variables polémicas o políticamente no correctas para usarlas en el modelo, tales como la raza, la religión o preferencias sexuales, esto de una u otra manera podrían aportar explicación al modelo al usarlas como variables de control, suponiendo que son iguales en el tiempo, sin embargo, es muy difícil obtenerlas.

CAPÍTULO 3

3. RESULTADOS Y ANÁLISIS

Los resultados del modelo fueron contundentes, individualmente las variables de interés fueron significativas, de igual manera las variables con interacción, lo que proporcionó un análisis marginal interesante para la toma de decisiones, que actualmente no ha sido considerado como criterio o juicio complementario al momento de otorgar crédito a clientes sin referencia crediticia.

Los errores estándar robustos se presentan en la tabla 6 y los resultados del modelo a través de odds ratios se presentan en la tabla 7.

Tabla 7 Errores estándar robustos en las variables del modelo

VARIABLES	Cartera Vencida
Persona jurídica#salud animal	0.09 (0.204)
Persona natural#farma	0.08 (0.162)
Persona jurídica#agro	0.62*** (0.114)
Persona natural#salud animal	0.56*** (0.148)
Persona jurídica#farma	0.12 (0.121)
Estado de ruc = 2, pasivo	0.40* (0.242)
Estado de ruc = 3, suspendido	0.71*** (0.121)
Estado de ruc = 4, no	-0.60* (0.345)

Letra de cambio = 1, si	0.34**
	(0.146)
Pagaré = 1, si	0.18**
	(0.082)
Soltero#si_obligacion	0.54***
	(0.188)
Casado#no_obligacion	-0.08
	(0.086)
Casado #si_obligacion	0.66***
	(0.168)
Divorciado#no_obligacion	-0.05
	(0.176)
Divorciado#si_obligacion	0.63
	(0.390)
Viudo#no_obligacion	-0.07
	(0.326)
Viudo#si_obligacion	1.38
	(1.132)
Sin info#no_obligacion	0.13
	(0.141)
Sin info#si_obligacion	0.11
	(0.357)
Solicitud de crédito = 1, si	0.49***
	(0.141)
Dias crédito	0.02***
	(0.001)
Garantía	0.00***
	(0.000)
Cupo	0.00***
	(0.000)
Constant	-1.82***
	(0.176)
Observations	3,658
Log-Likelihood Full Model	-2234
Chi-square test	355.5
AIC	1.238
BIC	-25298
Pseudo R2	0.115
PctCorr	76.85%

Errores estándar robustos en paréntesis

*** p<0.01, ** p<0.05, * p<0.1

Tabla 8 Resultados de la estimación del modelo en odds-ratio

Logistic regression		Number of obs	=	3,658	
		Wald chi2(23)	=	355.46	
		Prob > chi2	=	0.0000	
Log pseudolikelihood =	- 2234.4776	Pseudo R2	=	0.1154	
	Robust				
Cartera Vencida	Odds Ratio	Std. Err.	z P>z	[95% Conf.	Interval]
PERSONA#DIVISION					
Juridica#salud animal	1.089969	.2220578	0.42 0.672	.7311395	1.624906
Juridica#farma	1.081162	.1749851	0.48 0.630	.7872713	1.484762
Natural#agro	1.863308	.2116082	5.48 0.000	1.491479	2.327834
Natural#salud animal	1.745366	.2591651	3.75 0.000	1.304649	2.33496
Natural#farma	1.128302	.1367766	1.00 0.319	.8896921	1.430905
RUC					
Pasivo	1.494579	.3614067	1.66 0.097	.93044	2.400763
Suspendido	2.039724	.2463047	5.90 0.000	1.60985	2.584386
No	.5496999	.1894825	-1.74 0.083	.2797145	1.08028
LETRA DE CAMBIO					
Si	1.403	.2054514	2.31 0.021	1.052956	1.869413
PAGARÉ					
Si	1.192113	.0973941	2.15 0.031	1.015723	1.399134
ESTADOCIVIL#OBLIGACIÓN					
Soltero#si	1.717382	.3236028	2.87 0.004	1.187071	2.484603
Casado#no	.9213221	.0795618	-0.95 0.343	.7778663	1.091234
Casado#si	1.937854	.3246445	3.95 0.000	1.395472	2.691045
Divorciado#no	.9544583	.1677607	-0.27 0.791	.6763091	1.347004
Divorciado#si	1.873162	.7312461	1.61 0.108	.871531	4.025946
Viudo#no	.9301504	.3029895	-0.22 0.824	.4912253	1.761269

Viudo#si	3.973812	4.499219	1.22 0.223	.4319787	36.55547
No#no	1.141903	.1607403	0.94 0.346	.8665822	1.504697
No#si	1.114675	.3977802	0.30 0.761	.5538483	2.243396
SOLICITUD					
Si	1.631557	.2296868	3.48 0.001	1.238146	2.14997
Dias de credito	1.016494	.0012894	12.90 0.000	1.01397	1.019024
Garantia	1.000019	7.15e-06	2.62 0.009	1.000005	1.000033
Cupo	1.000008	2.89e-06	2.93 0.003	1.000003	1.000014
Constante	.1615108	.0284259	-10.36 0.000	.1143906	.228041

Se puede observar que las variables de interés son significativas, sin embargo, los coeficientes, en este caso los odds-ratio no dicen nada sobre los efectos marginales, por esa razón se encontraron los efectos marginales con la intención de comparar las variables con interacción de manera que se establecen perfiles de clientes y su probabilidad de tener cartera vencida, de acuerdo con las variables utilizadas.

En la tabla 8 se puede observar la significancia de los efectos marginales de las interacciones y posteriormente se hace la comparación de efectos marginales en la probabilidad de la variable dependiente.

Tabla 9 Efectos marginales de las interacciones del modelo

	Delta- method					
Margin	Std. Err.	z	P>z	[95% Conf.	Interval]	
PERSONA#DIVISION						
Juridica#agro	.4002978	.0188694	21.21	0.000	.3633144	.4372811
Juridica#salud animal	.4182842	.0381055	10.98	0.000	.3435987	.4929697
Juridica#farma	.4165789	.0281027	14.82	0.000	.3614986	.4716591
Natural#agro	.5341118	.0137708	38.79	0.000	.5071215	.561102
Natural#salud animal	.5198416	.0243713	21.33	0.000	.4720746	.5676085
Natural#farma	.4255737	.0153119	27.79	0.000	.395563	.4555844
ESTADOCVIL#OBLIGACION						
Soltero#no	.4543946	.0142236	31.95	0.000	.4265169	.4822723
Soltero#si	.5705455	.0376381	15.16	0.000	.4967762	.6443149

Casado#no	.4370739	.0116216	37.61	0.000	.4142961	.4598518
Casado#si	.5961772	.0322859	18.47	0.000	.5328979	.6594565
Divorciado#no	.4445217	.0344086	12.92	0.000	.3770822	.5119612
Divorciado#si	.5890065	.0813318	7.24	0.000	.4295991	.7484138
Viudo#no	.4390808	.0672279	6.53	0.000	.3073165	.570845
Viudo#si	.7364815	.198087	3.72	0.000	.348238	1.124725
No#no	.4827414	.0258323	18.69	0.000	.4321111	.5333717
No#si	.4775644	.0747307	6.39	0.000	.3310949	.624034
ESTADOCVIL#PERSONA						
Soltero#juridica	.4129336	.0190397	21.69	0.000	.3756165	.4502508
Soltero#natural	.4963541	.0140831	35.24	0.000	.4687518	.5239564
Casado#juridica	.401216	.0164082	24.45	0.000	.3690566	.4333754
Casado#natural	.4840561	.0123233	39.28	0.000	.4599029	.5082093
Divorciado#juridica	.4067195	.0341977	11.89	0.000	.3396933	.4737457
Divorciado#natural	.4898099	.0322112	15.21	0.000	.4266771	.5529426
Viudo#juridica	.4211266	.0651734	6.46	0.000	.2933891	.5488641
Viudo#natural	.5022492	.0639265	7.86	0.000	.3769555	.6275428
No#juridica	.4262528	.027474	15.51	0.000	.3724047	.480101
No#natural	.510186	.0261399	19.52	0.000	.4589527	.5614193
RUC						
Activo	.4475474	.0081131	55.16	0.000	.431646	.4634489
Pasivo	.5334248	.0512789	10.40	0.000	.43292	.6339296
Suspendido	.5996653	.0236203	25.39	0.000	.5533704	.6459601
No	.3281763	.0634862	5.17	0.000	.2037458	.4526069

Tabla 10 Efectos marginales de las interacciones del modelo

	Unadjusted		
Contrast	Std. Err.	[95% Conf.	Interval]
CARTERA VENCIDA			
PERSONA			
NATURAL vs JURIDICA	.3786153	.0962699	.1899297 .5673009
DIVISION			
FARMA vs SALUD ANIMAL	-.2221821	.1320575	-.4810099 .0366458
FARMA vs AGRO	-.2118017	.0959388	-.3998383 .0237651
SALUD ANIMAL vs AGRO	.0103804	.1220446	-.2288227 .2495834
PERSONA#DIVISION			
(NATURAL#FARMA) vs (NATURAL#AGRO)	-.5016395	.0962769	-.6903389 .3129402

(NATURAL#FARMA) vs (NATURAL#SALUD ANIMAL)	-.4362508	.1320079	-.6949814	- .1775201
(NATURAL#SALUD ANIMAL) vs (NATURAL#AGRO)	-.0653888	.1280905	-.3164415	.185664
(JURIDICA#FARMA) vs (JURIDICA#SALUD ANIMAL)	-.0081134	.2250253	-.4491549	.4329281
(NATURAL#FARMA) vs (JURIDICA#SALUD ANIMAL)	.0345642	.1949173	-.3474666	.4165951
(NATURAL#FARMA) vs (JURIDICA#FARMA)	.0426776	.1502222	-.2517526	.3371078
(JURIDICA#FARMA) vs (JURIDICA#AGRO)	.0780361	.1618491	-.2391823	.3952546
(JURIDICA#SALUD ANIMAL) vs (JURIDICA#AGRO)	.0861495	.2037285	-.313151	.48545
(NATURAL#FARMA) vs (JURIDICA#AGRO)	.1207137	.1212234	-.1168798	.3583073
(NATURAL#SALUD ANIMAL) vs (JURIDICA#SALUD ANIMAL)	.470815	.2111435	.0569814	.8846486
(NATURAL#SALUD ANIMAL) vs (JURIDICA#FARMA)	.4789283	.175641	.1346784	.8231783
(NATURAL#AGRO) vs (JURIDICA#SALUD ANIMAL)	.5362038	.1926067	.1587016	.9137059
(NATURAL#AGRO) vs (JURIDICA#FARMA)	.5443171	.1487328	.2528062	.835828
(NATURAL#SALUD ANIMAL) vs (JURIDICA#AGRO)	.5569645	.1484875	.2659344	.8479946
(NATURAL#AGRO) vs (JURIDICA#AGRO)	.6223532	.1135659	.3997682	.8449383
RUC				
NO vs ACTIVO	-.5983828	.3447017	-1.273986	.0772201
SUSPENDIDO vs PASIVO	.3109701	.2640072	-.2064745	.8284148
PASIVO vs ACTIVO	.4018445	.2418117	-.0720977	.8757867
SUSPENDIDO vs ACTIVO	.7128146	.1207539	.4761412	.949488

Como se puede observar en la tabla 9, hay diferencias significativas entre los clientes por división y por estado de RUC, estas variables no han sido tomadas en cuenta de manera crítica en decisiones de crédito, sin los clientes se diferencian entre sí de

acuerdo con estas variables; de esta manera podemos armar perfiles de clientes, tales como los indica el gráfico 6.



Gráfico 6 Comparación de perfiles a través de la probabilidad de pago adicional en clientes

De acuerdo a los resultados, un cliente que es persona natural tiene una probabilidad de tener cartera vencida adicional de 37.86%, respecto a un cliente que es persona jurídica, además si el cliente es persona natural y solicita un crédito en la división agro tiene una probabilidad de tener cartera vencida adicional de 54.43% respecto a una persona jurídica que solicita crédito en la división farma; un cliente con RUC suspendido tiene una probabilidad de tener cartera vencida adicional de 71.28% respecto a un cliente con RUC activo; un cliente que es viudo tiene una probabilidad de tener cartera vencida adicional de 36.38% respecto a un cliente casado y finalmente; un cliente con obligaciones con el SRI tiene una probabilidad de tener cartera vencida adicional de 67.73% respecto a una persona que no tiene obligaciones con el SRI.

La propuesta metodológica del score de crédito no significó desembolso significativo de dinero ni de recursos tecnológicos, salvo una organización interna del proceso de ingreso información de las solicitudes de crédito a través de la gestión documental y posterior digitalización de la información para la creación y actualización de la base de datos que se creó al desarrollarse el modelo expuesto, por lo que es viable económicamente y tecnológicamente debido al poco uso de recursos y porque la aplicación del modelo significaría una eficacia del 77% en la clasificación correcta de clientes con cartera vencida y con ello un incremento del presupuesto de cobros por el incremento de clientes que pagan sus deudas dentro del plazo establecido.

CAPÍTULO 4

4. CONCLUSIONES Y RECOMENDACIONES

4.1 Conclusiones

A través de la aplicación de este proyecto se logró crear una base de datos para organizar la información de solicitudes de crédito aprobadas, necesarias para evaluar clientes futuros; se demostró estadísticamente que los riesgos crediticios varían de acuerdo a la división en la que se solicita, tipo de contribuyente, estado civil y estado de RUC, finalmente se respaldó a la literatura que existe sobre la temática, debido a que, de acuerdo a la comparación entre el modelo Logit, Probit y Análisis Discriminante Lineal, el modelo Logit con las variables mencionadas en el segundo capítulo es más eficiente de acuerdo a criterios de información (AIC y BIC).

4.2 Recomendaciones

Se recomienda utilizar la metodología propuesta en la empresa XYZ como un complemento en el juicio utilizado para el otorgamiento de créditos, además en el futuro cercano se debe proponer una metodología para analizar clientes con historial crediticio, a través del análisis de variables financieras, comerciales, socioeconómicas e incluso macroeconómicas que capturen el efecto que tienen en la probabilidad de que el deudor tenga cartera vencida.

Finalmente, en el largo plazo, considerando que la empresa tiene que actualizar procesos de información, operativos y en el departamento de créditos y cobranzas se debe implementar metodologías de aprendizaje automático o alternativas a métodos tradicionales, tales como redes neuronales, máquinas de soporte vectorial (Support Vector Machines, SVMs), random forest (Bosques aleatorios) MARS (Multivariate Adaptive Regression Splines), entre otros, debido a que la literatura sugiere utilizarlos por ser más eficientes que los métodos tradicionales, es decir tienen un error de predicción menor.

BIBLIOGRAFÍA

- Ecuador, S. d. (Quito de Octubre de 2015). *SRI*. Obtenido de <https://www.sri.gob.ec/web/guest/home>
- Espin-García, O., & Rodríguez-Caballero, C. (2013). Metodología para un scoring de clientes sin referencias crediticias. *Cuadernos de Economía*, 139-165.
- García, M., & Sánchez, C. (3 de Diciembre de 2005). *Riesgo de crédito en México: aplicación del modelo CreditMetrics*. Obtenido de Universidad de las Américas Puebla: http://catarina.udlap.mx/u_dl_a/tales/documentos/laex/garcia_s_m/
- Girault, M. G. (Octubre de 2007). *Modelos de credit scoring: qué, cómo, cuándo y para qué*. Obtenido de Munich Personal RePEc Archive: <https://mpra.ub.uni-muenchen.de/16377/>
- Hand, D. J., & Henley, W. E. (Julio de 1997). *Statistical Classification Methods in Consumer Credit Scoring: a Review*. Obtenido de Royal Statistical Society: <https://rss.onlinelibrary.wiley.com/doi/10.1111/j.1467-985X.1997.00078.x>
- Kleimeier, S., & Dinh, T. H. (2007). A credit scoring model for Vietnam's retail banking market. *International Review of Financial Analysis*, 471-495.
- Lieli, R., & White, H. (2010). The construction of empirical credit scoring rules based on maximization principles. *Journal of Econometrics*, 110-119.
- Rodríguez, D., Becerra, J., & Cardona, D. (2017). MODELOS Y METODOLOGÍAS DE CREDIT SCORE PARA PERSONAS NATURALES: UNA REVISIÓN LITERARIA. *Revista CEA*, 13-28.
- Tan, W. C., Goh, C. P., & Koh, H. C. (2006). A Two-step Method to Construct Credit. *International Journal of Business and Information* , 96-118.
- Thomas, L. C., Edelman, D. B., & Crook, J. N. (2002). *Credit Scoring and Its Applications*. SIAM.

APÉNDICES

APÉNDICE A: Porcentaje de clientes por división



APÉNDICE B: Comparación de modelos Logit y Probit (Sin interacciones)

Variables	Logit	Logit	Probit	Probit
	Cartera Vencida	Cartera Vencida	Cartera Vencida	Cartera Vencida
Persona = 1, natural	0.42*** (0.086)	0.42*** (0.087)	0.24*** (0.052)	0.24*** (0.052)
Division = 2, salud animal	-0.01 (0.109)	-0.01 (0.109)	-0.01 (0.067)	-0.01 (0.067)
Division = 3, farma	- 0.35*** (0.085)	- 0.35*** (0.085)	- 0.21*** (0.051)	- 0.21*** (0.051)
Estado de ruc = 2, pasivo	0.37 (0.243)	0.37 (0.243)	0.22 (0.148)	0.22 (0.149)
Estado de ruc = 3, suspendido	0.70*** (0.119)	0.70*** (0.120)	0.42*** (0.073)	0.42*** (0.073)
Estado de ruc = 4, no	-0.55 (0.345)	-0.55 (0.344)	-0.32* (0.194)	-0.33* (0.193)
Letra de	0.35**	0.36**	0.23**	0.23**

cambio = 1, si				
	(0.146)	(0.146)	(0.090)	(0.090)
Pagare = 1, si	0.18**	0.19**	0.12**	0.12**
	(0.081)	(0.082)	(0.050)	(0.050)
Estado civil = 2, casado		-0.06		-0.03
		(0.081)		(0.049)
Estado civil = 3, divorciado		-0.02		-0.01
		(0.162)		(0.099)
Estado civil = 4, viudo		0.02		0.01
		(0.301)		(0.191)
Estado civil = 5, no		0.07		0.05
		(0.136)		(0.083)
Solicitud de credito = 1, si	0.46***	0.51***	0.28***	0.32***
	(0.129)	(0.141)	(0.077)	(0.086)
Obligaciones con el sri = 1, si	0.59***	0.59***	0.35***	0.36***
	(0.111)	(0.112)	(0.068)	(0.068)
Dias de credito	0.02***	0.02***	0.01***	0.01***
	(0.001)	(0.001)	(0.001)	(0.001)
Garantia	0.00***	0.00***	0.00***	0.00***
	(0.000)	(0.000)	(0.000)	(0.000)
Cupo	0.00***	0.00***	0.00**	0.00**
	(0.000)	(0.000)	(0.000)	(0.000)
Constant	- 1.68***	- 1.71***	- 1.00***	- 1.02***
	(0.139)	(0.166)	(0.081)	(0.100)
Observations	3,660	3,658	3,660	3,658
Log- likelihood full model	-2244	-2242	-2250	-2248
Chi-square test	349.2	351.4	386.3	388.8
Aic	1.238	1.240	1.241	1.243
Bic	-25370	-25315	-25359	-25304
Pseudo R2	0.112	0.112	0.110	0.110
Pctcorr	67.81	67.63	67.73	67.50

APÉNDICE C: Comparación de modelos Logit (Con interacciones)

VARIABLES	MODELOS LOGIT						
	Cartera Vencida						
0b.PERSONA#1b.DIVISION			0.00	0.00	0.00	0.00	0.00
			(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
0b.PERSONA#2.DIVISION			0.08	0.09	0.09	0.08	0.34
			(0.203)	(0.203)	(0.204)	(0.203)	(0.217)
0b.PERSONA#3.DIVISION			0.08	0.08	0.08	0.08	0.46**
			(0.162)	(0.162)	(0.162)	(0.162)	(0.180)
1.PERSONA#1b.DIVISION			0.62***	0.62***	0.62***	0.62***	0.89***
			(0.114)	(0.114)	(0.114)	(0.113)	(0.145)
1.persona#2.division			0.55***	0.55***	0.56***	0.55***	0.77***
			(0.148)	(0.149)	(0.148)	(0.148)	(0.167)
1.persona#3.division			0.11	0.11	0.12	0.11	0.44***
			(0.121)	(0.121)	(0.121)	(0.121)	(0.145)
Estado de ruc = 2, pasivo	0.37	0.37	0.40	0.40	0.40*	0.40*	0.38
	(0.243)	(0.243)	(0.244)	(0.245)	(0.242)	(0.243)	(0.287)
Estado de ruc = 3, suspendido	0.70***	0.70***	0.72***	0.72***	0.71***	0.71***	0.81***
	(0.120)	(0.120)	(0.120)	(0.121)	(0.121)	(0.121)	(0.140)
Estado de ruc = 4, no	-0.55	-0.55	-0.58*	-0.56	-0.60*	-0.59*	-0.06
	(0.345)	(0.344)	(0.347)	(0.349)	(0.345)	(0.346)	(0.354)
Letra de cambio = 1, si	0.35**	0.36**	0.34**	0.34**	0.34**	0.34**	0.13
	(0.146)	(0.146)	(0.146)	(0.146)	(0.146)	(0.146)	(0.185)
Pagarã% = 1, si	0.19**	0.19**	0.18**	0.17**	0.18**	0.18**	0.25***
	(0.082)	(0.082)	(0.082)	(0.082)	(0.082)	(0.082)	(0.096)
1b.ESTADOCVIL#0b.SOLICITUD				0.00			
				(0.000)			
1b.ESTADOCVIL#1.SOLICITUD				1.21***			
				(0.388)			

2.ESTADOCVIL#0b.SOLICITUD				0.77*			
				(0.441)			
2.estadocvil#1.solicitud				1.12***			
				(0.388)			
3.ESTADOCVIL#0b.SOLICITUD				0.28			
				(0.720)			
3.estadocvil#1.solicitud				1.17***			
				(0.414)			
4.ESTADOCVIL#0b.SOLICITUD				0.51			
				(1.132)			
4.estadocvil#1.solicitud				1.21**			
				(0.494)			
5.ESTADOCVIL#0b.SOLICITUD				0.84**			
				(0.410)			
5.estadocvil#1.solicitud				1.18***			
				(0.408)			
Obligaciones con el sri = 1, si	0.59***	0.59***	0.59***	0.59***			0.57***
	(0.112)	(0.112)	(0.111)	(0.112)			(0.125)
Dias credito	0.02***	0.02***	0.02***	0.02***	0.02***	0.02***	0.02***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Garantia	0.00***	0.00***	0.00**	0.00***	0.00***	0.00**	0.00*
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Cupo	0.00***	0.00***	0.00***	0.00***	0.00***	0.00***	0.00**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Estado civil = 2, casado	-0.06	-0.06	-0.06				-0.10
	(0.081)	(0.081)	(0.081)				(0.094)
Estado civil = 3, divorciado	-0.01	-0.02	-0.03				0.04
	(0.162)	(0.162)	(0.162)				(0.187)
Estado civil = 4, viudo	0.02	0.02	0.01				0.05
	(0.301)	(0.301)	(0.305)				(0.399)
Estado civil = 5, no	0.06	0.07	0.08				-0.18
	(0.136)	(0.136)	(0.136)				(0.156)

Solicitud de crédito = 1, si	0.51*** (0.141)	0.51*** (0.141)	0.50*** (0.141)		0.49*** (0.141)		0.05 (0.161)
Persona = 1, natural	0.41*** (0.087)	0.42*** (0.087)					
División = 2, salud animal	-0.01 (0.109)	-0.01 (0.109)					
División = 3, farma	-0.35*** (0.085)	-0.35*** (0.085)					
Estado final = 1, completo	-0.98 (0.609)						
1b.ESTADOCVIL#0b.OBLIGACION					0.00 (0.000)		
1b.ESTADOCVIL#1.OBLIGACION					0.54*** (0.188)		
2.ESTADOCVIL#0b.OBLIGACION					-0.08 (0.086)		
2.estadocvil#1.obligacion					0.66*** (0.168)		
3.ESTADOCVIL#0b.OBLIGACION					-0.05 (0.176)		
3.estadocvil#1.obligacion					0.63 (0.390)		
4.ESTADOCVIL#0b.OBLIGACION					-0.07 (0.326)		
4.estadocvil#1.obligacion					1.38 (1.132)		
5.ESTADOCVIL#0b.OBLIGACION					0.13 (0.141)		
5.estadocvil#1.obligacion					0.11 (0.357)		
Estado civil						0.02 (0.032)	

0b.OBLIGACION#0b.SOLICITUD						0.00	
						(0.000)	
0b.OBLIGACION#1.SOLICITUD						0.43***	
						(0.145)	
1.OBLIGACION#0b.SOLICITUD						0.30	
						(0.355)	
1.obligacion#1.solicitud						1.06***	
						(0.178)	
Oficial de crã%odito							0.07**
							(0.031)
Constant	-1.71***	-1.71***	-1.84***	-2.52***	-1.82***	-1.82***	-1.98***
	(0.166)	(0.166)	(0.176)	(0.388)	(0.176)	(0.196)	(0.226)
Observations	3,658	3,658	3,658	3,658	3,658	3,658	2,838
Log-Likelihood Full Model	-2241	-2242	-2237	-2234	-2234	-2237	-1738
Chi-square test	351.4	351.4	350.7	361.7	355.5	351.5	306.6
AIC	1.241	1.240	1.238	1.238	1.238	1.236	1.245
BIC	-25301	-25315	-25317	-25298	-25298	-25350	-18865
Pseudo R2	0.113	0.112	0.114	0.115	0.115	0.114	0.105
PctCorr	67.69	67.63	67.39	67.28	76.86	67.50	66.81

APÉNDICE D: Estimación de Análisis Discriminante Lineal

Linear discriminant analysis
Resubstitution classification summary

Key				
Number Percent				
True CARTRAVENCID A	Classified	NO	SI	Total
		NO	1,570 80.06	391 19.94
SI	805 47.44	892 52.56	1,697 100.00	
Total	2,375 64.93	1,283 35.07	3,658 100.00	
Priors	0.5361	0.4639		

Leave-one-out classification table

Key					
Number Percent					
True CARTRAVENCID A	LOO Classified	NO	SI	Unclassified	Total
		NO	1,567 79.91	394 20.09	0 0.00
SI	815 48.03	881 51.92	1 0.06	1,697 100.00	
Total	2,382 65.12	1,275 34.86	1 0.03	3,658 100.00	

APÉNDICE E: Estimación de Análisis Discriminante LOGARÍTMICO

Logistic discriminant analysis
Resubstitution classification summary

Key				
Number Percent				
True CARTRAVENTID A	Classified			Total
	NO	SI		
NO	1,551 79.09	410 20.91		1,961 100.00
SI	765 45.08	932 54.92		1,697 100.00
Total	2,316 63.31	1,342 36.69		3,658 100.00
Priors	0.5361	0.4639		

Error rate estimated by error count

Note: 5 observations were not classified and are not included in the table

	CARTRAVENTIDA		Total
	NO	SI	
Error rate	.209077	.4507955	.3212138
Priors	.5360853	.4639147	

APÉNDICE F: Efectividad del modelo seleccionado (Tabla de confusión)

Clasificados positivos si la probabilidad del cliente con cartera vencida	$\geq .5$	
Sensibilidad	Pr(+ D)	75.92%
Especificidad	Pr(\sim D)	77.61%
Valor predictivo positivo	Pr(D +)	70.37%
Valor predictivo negativo	Pr(\sim D -)	67.05%
Tasa de falsos positivos para verdaderos valores negativos	Pr(+ \sim D)	22.39%
Tasa de falsos negativos para verdaderos valores positivos	Pr(- D)	44.08%
Falsos positivos clasificados como positivos	Pr(\sim D +)	31.63%
Falsos positivos clasificados como negativos	Pr(D -)	32.95%
Correctamente clasificados		76.85%