



Construcción de un modelo Scoring de Probabilidad: el caso de la empresa SEGUMAR S.A.

CARRASCO PRECIADO, ANDY

Universidad Agraria del Ecuador (Ecuador)

Correo electrónico: anndy.carrasco.preciado@uagraria.edu.ec

GARCÍA REGALADO, JORGE

Universidad Agraria del Ecuador (Ecuador)

Correo electrónico: jgarcia@uagraria.edu.ec

CORNEJO MARCOS, GINO

Universidad Tecnológica ECOTEC (Ecuador)

Correo electrónico: gcornejo@ecotec.edu.ec

RESUMEN

El objetivo del presente trabajo es la construcción de un modelo credit scoring de probabilidad con la finalidad de minimizar el riesgo de incumplimiento de pago de la cartera de clientes, para lo que se utilizó variables dependientes (cliente "bueno o malo") y como independientes (características de los clientes) para proporcionar un análisis correcto para determinar si la empresa concede o no un crédito. Se aplicó la metodología descriptiva y enfoques cuantitativos y cualitativos tomando como fuentes primarias los datos de la cartera de clientes de la empresa SEGUMAR S.A. La base de datos consiste de la información de 100 personas solicitantes de un crédito y se incluye en la medición de 7 variables para cada persona. Cada solicitante se clasifica en una de dos categorías posibles, "buen cliente" (70 casos) o "mal cliente" (30 casos). Se desarrolló una regla de credit scoring para determinar si un nuevo solicitante es "Bueno" o "Malo" cliente, basándose en los valores de una o más variables explicativas resultantes del modelo final. Este estudio evaluó las características que tienen los clientes al momento de pedir un crédito y según las características de cada cliente se puede realizar predicciones, clasificarlos como un buen o un mal cliente. En los resultados obtenidos del modelo Logit se puede concluir que las variables seleccionadas que se aplicaron en el modelo arrojaron un 76% de éxito que nos permite clasificar a cada uno de nuestros clientes como un buen cliente o mal cliente en nuestro modelo.

Palabras clave: Logit; credit scoring; riesgo crediticio; variable dicotómica; incumplimiento.

Clasificación JEL: C33, G01, G21, P43.

MSC2010: 91G40, 91G50.

Construction of a Probability Scoring model for the company SEGUMAR S.A.

ABSTRACT

The objective of this work is the construction of a probability credit scoring model in order to minimize the risk of default on payment of the customer portfolio, for which dependent (customer good and bad) and independent (characteristic of customer) variables were used to provide a correct analysis to determine whether or not the company grants a loan. The descriptive methodology and quantitative and qualitative approaches were applied taking as primary sources the data of the customer portfolio of the company SEGUMAR S.A. The database consists of the information of 100 people applying for a loan and is included in the measurement of 7 variables for each person. Each applicant is classified into one of two possible categories, "good customer" (70 cases) or "bad customer" (30 cases). A credit scoring rule was developed to determine whether a new applicant is a "Good" or "Bad" customer, based on the values of one or more explanatory variables resulting from the final model. This study evaluated the characteristics that customers have at the time of requesting a loan and that according to the characteristics of each customer it is possible to make predictions, classify them as a good customer or a bad customer. In the results obtained from the Logit model it can be concluded that the selected variables that were applied in the model gave us a 76% success rate that allows us to classify each of our customers as a good customer or bad customer in our model.

Keywords: Logit; credit scoring; credit risk; dichotomous variable; default.

JEL classification: C33, G01, G21, P43.

MSC2010: 91G40, 91G50.



1. Introducción

SEGUMAR S.A es una cooperativa de ahorro y crédito localizada en la ciudad de Guayaquil de la provincia del Guayas (Ecuador), que tiene como principal actividad económica otorgar microcréditos a personas naturales, con el fin de lograr réditos de tales acciones.

En la actualidad, SEGUMAR S.A presenta una deficiente gestión de créditos, lo que obliga a realizar diversas actividades de recuperación de cartera, provocando altos costos administrativos. Por tal motivo, es de gran importancia identificar, medir, mitigar y controlar el riesgo crediticio que presenta cada individuo al momento de solicitar una nueva operación.

Con el fin de medir e identificar el riesgo crediticio, este artículo de investigación tiene como objetivo construir un modelo scoring de probabilidad, puesto que es un sistema de evaluación crediticia que valora automáticamente el riesgo asociado de una solicitud de crédito en condiciones de certidumbre (Puertas & Martí, 2013).

En este estudio, entendemos al riesgo de crédito como la probabilidad de que los clientes no puedan cumplir con sus compromisos de forma parcial o total por falta de liquidez al vencimiento.

Para reducir el riesgo crediticio, es decir, la recuperación de cuentas por cobrar, se recomienda analizar la industria en la que participan los clientes, el comportamiento histórico y las capacidades potenciales de pago de acorde a sus capacidades económicas (Urbina, 2019).

Según Largo (2015), una cartera de créditos es un conjunto de documentos que protegen a los activos financieros de la persona que solicita un crédito, el mismo que está bajo el poder del ofertante. Tomando en consideración la definición del autor podemos deducir que una cartera de créditos es el activo más importante debido a que es el mayor generador de rentabilidad para una empresa, por lo que una incobrabilidad o el incumplimiento de los individuos ya sean empresas o personas, considerado como riesgo crediticio, pondrían en peligro el patrimonio de la empresa.

La puntuación de crédito es cualquier sistema de evaluación de crédito que permite a los prestamistas evaluar que tan bien maneja las obligaciones financieras las personas (Bank of America, 2022). El riesgo dependerá de la solvencia del deudor, el tipo de crédito, los plazos del cliente y otras características y operaciones, que definirán cada observación, es decir, cada solicitud de crédito. Solo cuando la entidad ejecutora de la entidad de crédito actúe como mediador o intermediario, o cuando el crédito cuente con una garantía estatal, la operación crediticia no será riesgosa.

En general, la gestión crediticia requiere de información que asegure la viabilidad económica y financiera de la organización, y para ello debe contar con procesos adecuados que permitan generar mecanismos de orientación y corrección. En este caso se señaló que los indicadores financieros representan un conjunto de interrelaciones que pueden evaluar la liquidez, solvencia, eficiencia operativa, deuda, desempeño y rentabilidad de una organización.

Al comienzo de su establecimiento, el sistema de calificación crediticia se construyó sobre la base de técnicas estadísticas, especialmente el uso de métodos de análisis discriminatorios. Posteriormente, estos modelos se enriquecieron con las matemáticas, la econometría y la tecnología de inteligencia artificial. En cualquier caso, la construcción de un modelo de calificación crediticia requiere la inclusión de información interna y externa de los clientes en su solicitud de crédito.

El artículo está estructurado de forma secuencial, en donde se redactará con bases científicas las teorías relacionados al Modelo Credit Scoring, modelo de probabilidad lineal y modelo Logit. Posteriormente se desarrollará la metodología que se utilizará y, por último, los resultados y conclusiones del modelo.

En la situación actual, los avances tecnológicos han logrado que los países de la periferia soporten sus decisiones de riesgo crediticio, tomando como base modelos matemáticos que estimen la probabilidad de no pago a partir de comportamientos y características del individuo (Cantón, Rubio, & Blasco, 2010).

Por tal razón, resulta importante apoyar las decisiones con un modelo explicativo que contribuya a los sectores pequeños y medianos, que no sean exclusivos para sectores bancarios, tales como es el segmento de las cooperativas, mutualista, cajas de ahorros y otros entes no formales; siendo este estudio un aporte y evidencia de la extensión de este tipo de modelos y su alcance en un escenario de limitaciones de acceso a la información crediticia y asertividad en la respectiva colocación de crédito bajo una mitigación de riesgo.

Modelo Credit Scoring

El modelo credit scoring es un sistema de calificación crediticia cuyo resultado es la previsión de comportamiento de pago. El modelo de calificación crediticia también se conceptualiza como un método cuantitativo para predecir la probabilidad de que un solicitante de préstamo posponga o deja de cumplir con las obligaciones.

El termino scoring se refiere al uso del conocimiento sobre el desempeño y las características de préstamos antes otorgados que sirven como método para predecir el desempeño de otorgar préstamos a futuro. Scoring estadístico es el uso de conocimientos cuantitativos sobre el desempeño y las características de préstamos pasados registrados en una base de datos para predecir el desempeño futuro de los préstamos.

El sistema de calificación crediticia puede tomar decisiones automáticamente para otorgar o denegar alguna operación de crédito, utiliza información cualitativa y cuantitativa sobre préstamos de crédito.

Existen varias formas de medir la probabilidad de incumplimiento, uno de los principales problemas que surgen al elegir un determinado modelo es la disponibilidad y calidad de la información (características crediticias, información del cliente, garantías, experiencia de pago, economía ambiental) (Espín & Rodríguez, 2013).

Los primeros modelos de calificación crediticia se desarrollaron en la década de 1970 y se utilizan ampliamente aprovechando el avance tecnológico y las herramientas estadísticas, en la década de los noventa. Lara (2010) realizó un estudio sobre las técnicas de calificación crediticia y las dividió en modelos paramétricos y modelos no paramétricos. En el modelo paramétrico, el autor citó modelos lineales (análisis discriminante y modelos de probabilidad lineal). Y técnicas no lineales (modelo logit y modelo probit). La programación lineal, los árboles de decisión y las redes neuronales se construyen con métodos no paramétricos. A continuación, se presentan breves conceptos sobre cada una de las tecnologías anteriores.

La Técnica del Credit Scoring

Credit Scoring permite la cuantificación de la probabilidad de incumplimiento (o cumplimiento) de un individuo analizado, esta técnica requiere el uso de métodos cuantitativos, entre los más conocidos se encuentran los métodos estadísticos paramétricos o no paramétricos (Torrico, 2014).

Una medida de cumplimiento o incumplimiento debe ser diseñada por las personas analizadas y la información debe ser recolectada, para fines de modelado, se representan como variables binomiales discretas con los valores 1 y 0.

Una vez recopilada la información, debe modelarse utilizando una técnica cuantitativa que incluya:

- Paramétrica, que requiere supuestos sobre la información

- Logit (que es el más utilizado)
- Probit
- Log- Log
- Probabilidad Lineal

- No paramétrica, que no requiere supuestos sobre la información

- Redes Neuronales
- Algoritmos genéticos
- Análisis de vecindarios cercanos

Una vez aplicada la técnica, es necesario que la aplicación tenga un conocimiento profundo del contexto en el que se creó el modelo, que suele estar definido en las premisas del modelo, ya que utilizar el modelo desarrollado en un contexto inadecuado puede resultar perjudicial.

Modelos de Probabilidad Lineal

El modelo de probabilidad lineal utiliza el método de regresión de mínimos cuadrados, donde la variable dependiente (variable dummy) toma el valor de uno (1) si un cliente no cumple con sus pagos, o el valor cero (0) si el cliente cumple con su obligación de pago. La ecuación de regresión es una función lineal de las variables explicativas (Lara, 2010).

El modelo de probabilidad lineal considera el modelo de regresión lineal clásico porque la relación que se establece entre las variables es lineal:

$$Y_I = \beta_0 + \beta_1 X_{1I} + \beta_2 X_{2I} + \dots + \beta_K X_{KI} + \varepsilon_I$$

Modelo Logit o Logístico

Logit es un modelo de elección discreto más simple y más utilizado. Es popular porque la fórmula de probabilidad de elección tiene una expresión cerrada y se explica fácilmente.

El modelo de regresión logística permite calcular la probabilidad de que un cliente pertenezca a uno de los grupos a priori (no pagador o pagador). Al clasificar se debe tener en cuenta el comportamiento de una serie de variables independientes para cada valor de observación. La principal ventaja del modelo de regresión logística es que no es necesario realizar supuestos iniciales, como la normalidad de la distribución de las variables y mejorar el tratamiento de variables cualitativas o categóricas. Además, la ventaja de este modelo es que la posibilidad de incumplimiento se puede medir manteniendo las variables explicadas en el rango de variación entre cero y uno (Pantoja, 2016).

El modelo Logit o Logístico proporciona la probabilidad de que una determinada observación pertenezca a un determinado grupo y conozca el valor de la variable independiente de la observación. La ecuación del modelo es la siguiente:

$$P_j = \frac{e^{z_j}}{1 + e^{z_j}}, \text{ sabiendo que } Z_j = B_1 X_1 + B_2 X_2 + \dots + B_P X_P$$

O, lo que es lo mismo:

$$\text{Log} = \left(\frac{p_j}{1 - p_j} \right) = B_1X_1 + B_2X_2 + \dots + B_pX_p$$

donde:

X_p = Variables independientes, $p=1, \dots, n$

B_p = Coeficientes estimados o pesos de cada variable, $p=1, \dots, k$

Z_j = Puntuación del individuo o de la observación $j=1, \dots, k$

P_j = Probabilidad del individuo j , de pertenecer a un grupo.

Riesgo Crediticio

Es la posibilidad que sufran pérdidas por incumplimientos del prestatario o contraparte en negocios directos, indirectos o derivados, lo que significa impago, pago parcial o no posibilidad de pago de la deuda pactada.

El riesgo se puede definir como: el impacto o impacto económico actual o potencial (aún no realizado) causado por posibles eventos adversos que se originan por causas internas o externas y pueden ser accidentales o deliberados. Estos eventos van deteriorando los resultados esperados (Álveo, 2011).

Esta connotación se centra en el riesgo porque pueden ocurrir eventos inciertos (amenazas), accidentales y económicamente desfavorables; ésta es una combinación de probabilidad e impacto (grado de daño). Todas las actividades conllevan riesgos, deben aceptarse, mitigarse, transferirse o evitarse, pero no pueden eliminarse (Lagua, 2015).

Factores de Riesgo

Los factores de riesgo son situaciones o circunstancias que aumentan la posibilidad de un determinado riesgo y se pueden dividir en factores internos y factores externos. Los factores internos están directamente relacionados con la gestión de la organización, por lo que son más predecibles y controlables de acuerdo con las capacidades de gestiones de alta dirección de la organización. Uno de los factores más importantes es la decisión de suavizar la política crediticia. La agresividad es el mayor riesgo de la política crediticia es que la cartera de inversiones se concentra en unos pocos negocios crediticios, falta de control o control mínimo en el proceso crediticio, y escaso personal, capacidad de gestión, compromiso y comprensión del área de negocio, acciones intencionadas, etc. (Lagua, 2015).

Marco Regulatorio ecuatoriano sobre Gestión de Riesgos

El Marco regulatorio ecuatoriano para la gestión de riesgos financieros está dictaminado por la Superintendencia de Bancos y Seguros. La base de la norma es el libro I: Normas Generales para la aplicación de la Ley General de Instituciones del Sistema Financiero, específicamente el Título X: De la Gestión y Administración de Riesgos (Superintendencia de banco y seguros, 2010).

Revisión de literatura

En el presente artículo de investigación se procedió a la revisión y recopilación literaria de estudios realizados de la temática que se expone, estudios que se detallan a continuación.

En el trabajo titulado Modelo credit scoring en una entidad micro financiera, que permite a los solicitantes calificados obtener microcréditos, el modelo se basa en un modelo de regresión logística. La evaluación del modelo incluye la aptitud medida por las estadísticas de Hosmer Lemeshow, el poder predictivo medido por R^2 y la discriminación medida por la curva COR y el área bajo la curva COR. El modelo se desarrolló utilizando la información del cliente desde agosto de 2008 hasta julio de 2009, desde agosto hasta octubre de 2009 para probar la validez del pronóstico. Los resultados de la

verificación reflejan la buena correspondencia del modelo y la tasa de éxito general es del 83% (Colonia, 2012).

El trabajo titulado Análisis de riesgo crediticio, propuesta del modelo credit scoring, corresponde a una empresa dedicada a la producción, venta y distribución de productos asfálticos en el sur de Chile. La mencionada empresa no desea revelar su nombre y la denominamos Fantasía S.A. por este motivo. En los últimos años, las ventas de Fantasía han logrado un crecimiento significativo, seguido de una disminución en los niveles de liquidez y una disminución en la calidad de las cuentas por cobrar. Sin embargo, dada la política de liberalización de las cuentas por cobrar, el aumento de las cuentas por cobrar está relacionado con el supuesto mayor riesgo de cobranza.

Además, Fantasía S.A. no cuenta con un sistema de gestión crediticia objetivo y no puede evaluar adecuadamente la calidad y las capacidades crediticias de sus clientes actuales y potenciales. Por ello, este artículo propone a Fantasía un modelo de evaluación crediticia para sus clientes actuales y potenciales, que ha sido ajustado y ponderado de acuerdo a las condiciones reales, reduciendo así el riesgo de crédito o insolvencia. Este trabajo consideró la descripción del modelo de evaluación crediticia, especialmente la descripción del modelo de calificación crediticia. Mediante entrevistas a expertos se definen las variables cuantitativas y cualitativas clave a considerar en el proceso de gestión crediticia.

En cuanto a la calidad del modelo de evaluación crediticia propuesto, mostró que el 81,82% de los préstamos otorgados a los clientes superaron el nivel mínimo de evaluación o aprobación de la empresa (Leal, Aranguiz & Gallegos, 2017).

Según Laguna (2015), una gestión adecuada del riesgo de crédito puede permitir que las cooperativas de crédito logren rentabilidad y sostenibilidad. Durante la fase de diagnóstico, se ha determinado que uno de los temas más importantes que enfrentan estas organizaciones son las altas tasas de morosidad, que pueden generar desconfianza en los clientes y, en algunos casos, el eventual cierre de las agencias de control.

Otra forma de mejorar el desempeño de la cartera de préstamos es aplicar modelos estadísticos, que pueden superar la subjetividad en la toma de decisiones crediticias. Esta investigación tiene como objetivo utilizar herramientas estadísticas de regresión logística binaria (que pueden extraer conocimiento de bases de datos) para diseñar un modelo de calificación crediticia para la gestión del riesgo crediticio en carteras de microempresas. La eficiencia del modelo se verifica utilizando la bondad de las herramientas de ajuste odds, matriz de clasificación, análisis de sensibilidad y aplicación del modelo a casos reales) e información estadística, que ayuda a comprender el pronóstico alcanzado por el modelo.

Para el caso de estudio se seleccionó una institución del sistema financiero popular y unido de la zona central del país, que brindó una base de datos de 5.078 créditos otorgados durante el período 2010-2014. Los resultados obtenidos permiten la calificación automática y segura de las solicitudes de crédito. Hay que aclarar que estos modelos no reemplazan el trabajo de los oficiales de crédito, solo constituyen herramientas de apoyo para la emisión de préstamos.

En el trabajo titulado Diseño de un scoring de crédito para la Cooperativa de Ahorro y Crédito "Crediamigo" LTDA. Cantón Loja - Provincia de Loja, la empresa brinda servicios financieros a sus empleados y clientes, incluyendo préstamos al consumo y préstamos a microempresas que utilizan garantías personales y garantías hipotecarias. Si bien se administran de manera efectiva a través de procesos y devoluciones de fondos, se diferencian del pago por el sujeto del crédito. La calificación no es alta. Mantiene más del 8% de la cartera de inversiones problemática. Por definición, las pérdidas por impago no pueden transferirse y, por lo tanto, son absorbidas por el patrimonio de la institución. Por tanto, el propósito de este trabajo es asegurar la Cooperativa de Ahorro y Crédito Crediamigo Ltda. Mediante la implementación de un modelo de puntuación de calificación de cartera para sus afiliados, logre en primer lugar, a través del análisis estadístico detallado de la base de datos institucional, conocer

las variables y factores de riesgo crediticio que afectan el normal funcionamiento de la empresa y tomar las medidas necesarias para identificar; medir, mitigar y controlar el riesgo crediticio (González, 2015).

En el trabajo Modelos para otorgamiento y seguimiento en la gestión de riesgo de crédito realizado por Millán y Caicedo (2018), se buscó implementar 3 modelos para la clasificación de solicitantes de créditos; adicional a esto ellos usaron 14 variables que informaban sobre las características socioeconómicas del prestatario y las características propias de la operación crediticia. Con esta investigación previa se pudo corroborar que las variables usadas en el trabajo presente son de características homogéneas, dando un resultado óptimo y esperado.

El siguiente trabajo, titulado Diseño de un modelo de credit scoring aplicado a créditos recurrentes para una Caja Municipal, cubre y nos permite comprender un tema importante en el negocio de las entidades financieras, que es la evaluación y gestión estratégica de los riesgos crediticios en el proceso de concesión de créditos. Hay dos razones principales para este problema. En primer lugar, los clientes desean obtener préstamos rápida y fácilmente. En segundo lugar, porque las entidades financieras pueden tomar medidas para satisfacer este deseo. Para poder gestionar el riesgo de estar inmerso en este proceso, es necesario adoptar una solución que pueda generar un proceso rápido y eficaz, lo cual sea beneficioso tanto para el cliente como para la organización, por lo que se recomienda como objetivo de esta investigación: diseñar un modelo matemático de puntaje crediticio. Es adecuado para otorgar crédito a clientes habituales de bancos municipales. Esto se logra seleccionando las características socioeconómicas y financieras del cliente. Utilizo el método de simulación del sistema para obtener los datos necesarios para obtener el modelo a partir de los datos generados aleatoriamente. Estos datos se procesan mediante un software estadístico llamado Eviews. Además del conocimiento de estadística y econometría, el software también puede generar, analizar y seleccionar variables exógenas que finalmente conducen al modelo logit. La tasa de predicción correcta es del 55,30%. El ajuste es pobre, pero tiene una variable final confiable porque su efecto marginal está en línea con las expectativas. Dado que no se obtiene a través del análisis de datos reales, el modelo no puede ser considerado como un modelo confiable del verdadero proceso de otorgamiento de crédito (Navarro, 2015).

En el presente trabajo, titulado Modelo scoring para recuperar cartera de microcrédito, se menciona la construcción de un modelo Scoring para establecer la probabilidad de pago del cliente cuando se encuentra en mora, dirigido al sector micro crediticio, buscando una herramienta con la que se pueda formalizar la cartera de una manera eficiente, reduciendo los costos e incrementando el nivel de recuperación. A fin de cumplir este objetivo se cuenta con una base de 6 meses (abril- septiembre del 2018), con la que se generan 4 modelos diferentes: regresión logística, regresión probit, árboles de decisión y redes neuronales. Se mide el porcentaje de error en cada uno de ellos, dando como resultado que el porcentaje menor el generado por medio de árboles de decisión con un 11,3 %, dicho modelo refleja un ajuste adecuado, mostrando un nivel de predictibilidad acertado. En cuanto a los modelos restantes ninguno cuenta con un porcentaje mayor al 12,5 %. Por ello y por medio de la regresión logística se determina que la garantía, el género y la mora al momento de evaluación son las variables con mayor incidencia en el no pago de los clientes (Rodríguez, 2018).

De la misma manera Rodríguez, Rendón, Trespalacios y Jiménez (2022), mencionan en su trabajo que los modelos de tipo credit score permiten a los analistas de crédito la cuantificación de los riesgos que implican las operaciones de crédito; estos modelos buscan inferir en la posibilidad de impago de un afiliado.

2. Metodología

En el presente trabajo de titulación, se adoptó el tipo de investigación bibliográfica porque se realizó un análisis exhaustivo y se revisaron los documentos de la empresa SEGUMAR S.A. Reflejados en libros,

estos libros aportaron con información analítica. También se empleó la investigación de campo que permite obtener información directamente de las variables de estudio de la investigación.

El diseño de investigación que se utilizó fue no experimental porque no se manipularon las variables de investigación durante el proceso de investigación.

El enfoque cuantitativo parte de una idea y tiene como objeto seleccionar información numérica sobre diversas variables que han sido previamente analizadas, para poder sistematizarlas con el fin de obtener una realidad de acuerdo al problema que se está investigando.

El tipo de investigación es descriptiva porque describe la situación de riesgo crediticio.

Variables dependientes

Clase, que es una variable dicotómica que se utiliza para referirse a clientes "buenos" con un valor de 0 y clientes "malos" con un valor de 1.

Variables independientes

Las Características de los clientes: Edad, Historia Crediticia, Estado civil, Tipo de Vivienda, Teléfono Registrado, Género, Duración de Trabajo.

Población

Los datos corresponden a fuentes secundarias, como la cartera de clientes de la empresa SEGUMAR S.A en el periodo de tiempo comprendido de abril del 2020 a enero del 2021.

Muestra

Para la determinación de la muestra, se estableció un criterio de inclusión necesario que permitió resolver los objetivos planteados con anterioridad, que son:

- Crédito otorgado entre abril del 2020 y enero del 2021.
- Poseer la totalidad de información relacionado a: historia crediticia, duración de trabajo, género, estado civil, edad, tipo de vivienda y teléfono.
- Número de casos de créditos impago-aproximados a los créditos pagados, con el objetivo de evitar problemas de inexactitud por problemas por el tamaño de los grupos (Cantón, Rubio, & Blasco, 2010).

Los casos seleccionados son solo los que tuvieron toda la información necesaria para el estudio, eliminando el resto de los casos que no aporta información seleccionada para el desarrollo del modelo.

Después de discriminar la información de la base de datos, se obtuvo un número total de 134 créditos otorgados que proporciona la totalidad de información necesaria para el estudio entre el periodo estipulado.

La posterior elaboración de número de muestra se dio tomando un criterio del 95% de nivel de confianza con un margen de error de 5%, dando como resultado a un número de muestra de 100 datos.

Tipo de muestra

Para la selección de los casos, se utilizó una muestra aleatoria simple, donde se le asignó un número en orden a los 134 casos y luego con la generación de números aleatorios, se seleccionaron los casos correspondientes a la muestra con un criterio del 95% de nivel de confianza.

Estudios como el de Araújo, Carmona, & Amorim (2007), utilizan una muestra aleatoria simple, realizó un modelo credit Scoring con 200 datos divididos en dos grupos iguales identificados como mora y morosos las cuales son características de los individuos, para posteriormente establecer un modelo de riesgo crediticio.

Técnica de Recolección de Datos

La técnica que se empleó en este trabajo de investigación es a través de fuente secundaria; la fuente de estos datos se encuentra en la cartera de clientes de la empresa SEGUMAR S.A, de las que se seleccionó información numérica, para la resolución de las variables y objetivos planteados.

Para un efectivo procedimiento e interpretación de datos se procedió a utilizar técnicas estadísticas, para el desarrollo de cuadros y gráficos, se utilizó el programa Gretl con un nivel de confianza del 95%. El análisis y la interpretación de los cuadros y gráficos estadísticos permiten formular las respectivas recomendaciones de acuerdo a los resultados reflejados.

Estadística Descriptiva e Inferencial

Se realizó un modelo Scoring de tipo paramétrico donde el modelo Logit permite calcular la probabilidad de que un individuo pertenezca o no a uno de los grupos establecidos a priori. La clasificación se basa en el comportamiento de una serie de variables independientes, que son las características de cada persona. Éste es un modelo de elección binaria en el que la variable dependiente tomará el valor 1 o 0. En nuestro problema, el valor depende de si el individuo ha cumplido con su obligación crediticia.

El modelo Logit permite obtener estimaciones de la probabilidad de un suceso e identificar los factores de riesgo que determinan dichas probabilidades. La modelización Logit es similar a la regresión tradicional, salvo que utiliza como función de estimación la función logística, lo que permite por un lado obtener la estimación de probabilidad, pero al tratarse de un análisis de regresión, permite identificar las variables más relevantes.

En este caso se va a elegir un modelo Logit, es decir, cuando el número de alternativas son dos y son excluyentes entre sí. El modelo presenta las siguientes características: la variable dependiente es binaria y toma sólo dos valores y las variables independientes, que pueden ser cualitativas o cuantitativas y al final arroja un vector de parámetros con valores numéricos, que son los coeficientes para cada una de las variables explicativas. De aquí se puede obtener que a cada valor del vector de parámetros le corresponde una variable explicativa, al tenerse en cuenta todas en conjunto y dar valores a cada una de las variables independientes contenidas en el modelo definitivo, se obtiene el valor de la probabilidad de que un individuo posea la característica de interés estudiada en el modelo.

La ecuación del modelo es la siguiente:

$$P_j = \frac{e^{z_j}}{1 + e^{z_j}}, \text{ sabiendo que } Z_j = B_1X_1 + B_2X_2 + \dots + B_pX_p$$

O, lo que es lo mismo:

$$\text{Log} = \left(\frac{p_j}{1 - p_j} \right) = B_1X_1 + B_2X_2 + \dots + B_pX_p$$

donde:

X_p = Variables independientes, $p=1, \dots, n$

B_p = Coeficientes estimados o pesos de cada variable, $p=1, \dots, k$

Z_j = Puntuación del individuo o de la observación $j=1, \dots, k$

P_j = Probabilidad del individuo j , de pertenecer a un grupo

3. Resultados

3.1. Descripción de la Base de Datos

La base de datos que se utilizó fue la cartera de clientes de la empresa SEGUMAR S.A., la cual consiste en la información de 100 personas solicitantes de un crédito y se incluye en la medición de 7 variables para cada persona. Cada solicitante se clasifica en una de dos categorías posibles, "buen cliente" (70 casos) o "mal cliente" (30 casos).

Se desarrolla una regla de calificación crediticia que determina si un nuevo solicitante es un cliente "bueno" o "malo" según el valor de una o más variables explicativas generadas por el modelo final. Las variables a considerar se describen a continuación:

Tabla 1. Descripción de la Base de Datos.

No.	Nombre de la variable	Descripción	Tipo de variable	Descripción
1	Clase	Clasificación de los solicitantes	Catagórica (Binaria)	1 = Malo 0 = Bueno
2	Historia crediticia	Historial crediticio por cada cliente	Catagórica	0 =No tiene créditos 1 =Todos los créditos pagados debidamente 2 = Créditos existentes debidamente pagados hasta ahora 3 = Retraso en el pago 4 =Cuenta crítica
3	Duración en trabajo	Años en el trabajo actual	Catagórica	1 =Desempleado 2 =Menos de 1 año 3 =Entre 1 y 4 años 4 =Entre 4 y 7 años 5 =Mayor o igual a 7 años
4	Género y Estado civil	Categorías en las que se clasificaron a los clientes de acuerdo a su género y estado civil	Catagórica	1 =Hombre Divorciado\Separado 2 =Mujer Divorciada\Separada\Casada 3 =Hombre soltero 4 =Hombre Casado\Viudo 5 =Mujer Soltera
5	Edad	Edad a la que el cliente solicitó el crédito	Numérica	
6	Tipo Vivienda	Tipo de vivienda en la que reside el cliente	Catagórica	1 =Alquilada 2 =Propia
7	Teléfono	¿Tiene el cliente número de teléfono registrado bajo su nombre?	Binaria	1 =No 2 =Sí

Fuente: Elaboración propia.

En la Tabla 2 se puede observar cómo se detalla el estado civil combinado con el género (variable ya configurada en la base de datos de esa manera) de acuerdo a clientes "Buenos" y "Malos", los datos

se concentran en las categorías Mujer: Divorciada/Separada/Casada, siendo el doble en clientes “Buenos” y también en Hombre: Soltero, siendo en mayor cantidad en los clientes “Buenos”.

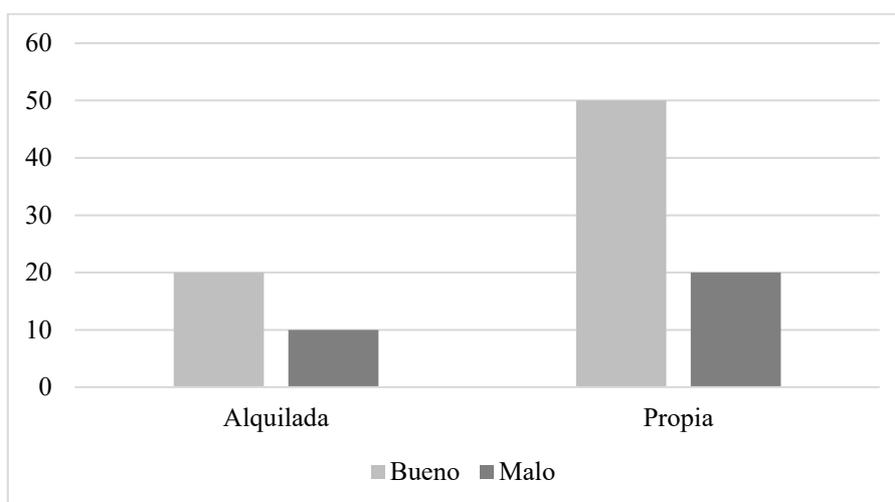
Tabla 2. Género-Estado Civil.

	Clase		Total
	Malo	Bueno	
Hombre: Divorciado/soltero	3	3	6
Mujer: Divorciada/Separada/Casada	12	20	32
Hombre: Soltero	14	40	54
Hombre: Casado/Viudo	2	6	8
Total	31	69	100

Fuente: Elaboración propia.

La Figura 1 muestra la variable tipo de vivienda, que cuenta con dos categorías (Alquilada y Propia), teniendo la mayor concentración de los datos en la categoría de Propia para ambas clases, pero con mayor proporción dentro de los clientes “Buenos”.

Figura 1. Tipo de Vivienda.



Fuente: Elaboración propia.

En la Tabla 3 se observa que la variable Duración en el trabajo actual, que consta de 5 categorías, en los clientes “Malos” se tiene mayor concentración en la categoría de Entre 1 y 4 años, seguido por Menos de un año; dentro de los clientes “Buenos” la mayor concentración de datos está en la categoría Entre 1 y 4 años, pero a esta categoría le sigue la de Mayor o igual a 7 años.

Tabla 3. Duración en el trabajo.

	Clase		Total
	Malo	Bueno	
Desempleado	2	3	5
Menos de 1 año	7	13	20
Entre 1 y 4 años	12	23	35
Entre 4 y 7 años	3	13	16
Mayor o igual a 7 años	6	18	24
Total	30	70	100

Fuente: Elaboración propia.

En la Tabla 4 se puede observar la variable Historia Crediticia, que tiene 5 categorías, y tanto los clientes “Malos” como los “Buenos” se encuentra más de la mitad de ellos dentro de los Créditos existentes debidamente Pagados hasta ahora.

Tabla 4. Historia Crediticia.

	Clase		
	Malo	Bueno	Total
No tiene créditos tomados	2	1	3
Todos los créditos pagados debidamente	2	2	4
Créditos existentes debidamente pagados hasta hora	16	40	56
Retraso en el pago	2	6	8
Cuenta crítica	5	24	29
Total	5	73	100

Fuente: Elaboración propia.

3.2. Selección de Variables Aplicadas al Modelo

Variable dependiente

La variable dependiente del modelo se llama Clase, que es una variable dicotómica que se utiliza para referirse a clientes "buenos" con un valor de 0 y clientes "malos" con un valor de 1.

Variable independiente

Las variables independientes seleccionadas de acuerdo con las características de la base de la empresa SEGUMAR S.A, son:

- **Historia Crediticia.** Está conformada por 4 categorías en donde se muestra si el cliente ha tenido otros créditos:

- 0 = No tiene créditos tomados debidamente.
- 1 = Todos los créditos pagados debidamente.
- 2 = Créditos existentes debidamente pagados hasta ahora.
- 3 = Retraso en el pago.
- 4 = Cuenta crítica

- **Duración en el trabajo.** Está conformada por 5 categorías:

1 = Desempleado.
 2 = Menos de 1 año.
 3 = Entre 1 y 4 años.
 4 = Entre 4 y 7 años.
 5 = Mayor o igual a 7 años

- **Género y Estado Civil.** Está conformada por 5 categorías:

1 = Hombre: Divorciado/Separado.
 2 = Mujer: Divorciada/Separada/Casada.
 3 = Hombre: Soltero.
 4 = Hombre: Casado/ Viudo.
 5 = Mujer: Soltera.

- **Tipo de Vivienda.** Está conformada por 2 categorías:

1= Alquilada
 2= Propia

- **Teléfono.** Está conformada por 2 categorías:

1= Sí
 2= No

3.3. Estimación del Modelo en Gretl

Para construir el modelo se utiliza la base de datos de la empresa SEGUMAR S.A, teniendo como variable dependiente a la variable Clase, para los clientes incumplidos está la etiqueta de Malos con un valor de 1, y para los clientes cumplidos como Buenos con el valor de 0; incluyendo la lista de variables independientes.

$$P_j = \frac{e^{z_j}}{1 + e^{z_j}}$$

Clase(0: Buencliente; 1: Mal Cliente)

$$= -0,442219 \text{ Historia Crediticia} - 0,239800 \text{ Duración de Trabajo} \\ + 0,264952 \text{ Género y Estado civil} - 0,339963 \text{ Vivienda} + 0,00217914 \text{ Edad}$$

En el siguiente cuadro, se muestran los resultados del Modelo Logit:

Media de la vble. dep.	0,300000	D.T. de la vble. dep.	0,460566
R-cuadrado de McFadden	0,361218	R-cuadrado corregido	0,353374
Log-verosimilitud	-57,34683	Criterio de Akaike	128,6937
Criterio de Schwarz	146,9299	Crit. de Hannan-Quinn	136,0742

Número de casos 'correctamente predichos' = 76 (76,0%)
 f(beta'x) en la media de las variables independientes = 0,461

Contraste de razón de verosimilitudes: Chi-cuadrado(6) = 7,4792 [0,2788]

En la Tabla 5 se puede observar que el modelo logit tiene un tamaño muestral de 100 observaciones y además tiene una constante con un valor p asintótico de 0,4694 con un nivel de significancia del 5%, y sus variables independientes o explicativas como la historia crediticia tiene un valor p asintótico de 0,0749 con un nivel de significancia del 10%, Duración de Trabajo tiene un valor p asintótico de 0,2740 con una significancia del 1%, Género y Estado civil tiene un valor p de 0,5124 con una significancia del 1%, vivienda tiene un valor 9 asintótico de 0,1355 con un nivel de significancia

del 5%. Y la variable Clase queda explicada en un rango entre 27,88 % y el 36,12 % por las variables explicativas del modelo.

Tabla 5. Modelo Logit.

Modelo 1: Logit, usando las observaciones 1-100				
Variable dependiente: Clase				
Desviaciones típicas basadas en el Hessiano				
	<i>Coficiente</i>	<i>Desv. Típica</i>	<i>z</i>	<i>valor p</i>
const	1,39600	1,92964	0,7234	0,4694 **
Historia Crediticia	-0,442219	0,248279	-1,781	0,0749 *
Duración en el trabajo actual	-0,239800	0,219197	-1,094	0,2740 ***
Género Estado Civil	0,264952	0,404482	0,6550	0,5124 ***
Vivienda	-0,339963	0,227765	-1,493	0,1355 **
Edad	0,00217914	0,0217280	0,1003	0,9201
Teléfono	-0,121791	0,518215	-0,2350	0,8142

Fuente: Elaboración propia.

Tabla 6. Predicciones.

		Predicho	
		0	1
Observado	0	68	2
	1	22	8

Fuente: Elaboración propia.

El número de casos correctamente predichos fue de 76 aciertos lo que representa un 76%.

4. Conclusiones

En primer lugar, el desarrollo de un modelo de calificación crediticia depende de la empresa y de los resultados esperados, pues según la investigación realizada, existen múltiples modelos disponibles, así como sus ventajas y desventajas, pues si bien esto es cierto, permiten que el riesgo pueda ser cuantificado como probabilidad; además debe ser consistente y claro; también requiere una serie de datos de alta calidad y amplia para poder ser más certeros.

Dado que no se logró obtener mayor información del que utiliza la empresa SEGUMAR S.A, para la toma de decisión en lo que se refiere a otorgarle o no un crédito a una persona bajo los criterios que utilizaron, únicamente se desarrolló el modelo planteado en el trabajo de titulación y utilizarlo en la comparación de los clientes de la base de datos que fue facilitada. Esto arrojó una predicción del modelo de un 76%, es decir, que el 76% de la categoría en la que se encuentran los clientes ya sean en incumplimiento o no, es la misma que muestra el modelo utilizado en esta investigación, específicamente el Logit.

Las variables que se utilizaron para hacer el análisis de crédito están referenciadas directamente al crédito tales como historia crediticia y se consideraron características propias del individuo como estado civil, edad y género, duración de trabajo, características que se consideraron importantes y relevantes para un Credit Scoring

Por otro lado, se evidencia que las variables usadas en este estudio son relevantes, ya que en otros estudios similares se usan variables socioeconómicas de los clientes y en base los hallazgos obtenidos se pueden considerar como cliente (bueno o malo) a tal caso de diligenciarle un crédito.

A nivel general, se concluye que el uso del Credit Scoring genera beneficios a la empresa, siempre y cuando se cuente con la información necesaria y confiable para el desarrollo del mismo.

Se concluye un aporte significativo para los sectores no formales de la actividad financiera en Guayaquil obteniéndose 76 clientes “buenos” de cada 100; lo que hace que este tipo de modelo sean confiables y válidos para la actividad; considerando aspectos restrictivos de la información asociada a cada individuo. No obstante, el nivel de adopción de riesgo es discrecional a la entidad.

Referencias

- Álveo, E. (2011). *Gestión integral de riesgo en bancos e instituciones financieras*. [Sitio web] <https://www.gestiopolis.com/gestion-integral-de-riesgo-en-bancos-e-instituciones-financieras/>
- Araújo, E., Carmona, C., & Amorim, A. (2007). Aplicação de modelos credit scoring na análise da inadimplência de uma instituição de microcrédito. *Revista Ciências Administrativas*, 13(1), 110-121. <https://www.redalyc.org/articulo.oa?id=475647703009>
- Bank of America (2022). *Mejores Hábitos Financieros*. <https://bettermoneyhabits.bankofamerica.com/es/credit/how-credit-score-is-calculated#:~:text=lograr%20el%20éxito,-¿Qué%20es%20un%20puntaje%20de%20crédito%3F,ha%20manejado%20sus%20obligaciones%20financieras.>
- Cantón, S., Rubio, J., & Blasco, D. (2010). Un Modelo de Credit Scoring para instituciones de microfinanzas en el marco de Basilea II. *Journal of Economics, Finance and Administrative*, 15(28), 84-124. doi:http://www.scielo.org.pe/scielo.php?script=sci_arttext&pid=S2077-18862010000100005&lng=es&tlng=en.
- Colonia, D.V. (2012). *Válidez de pronóstico del modelo credit scoring en una entidad microfinanciera (Tesis de grado)*. Universidad Nacional de Trujillo (Lima, Perú). <http://dspace.unitru.edu.pe/bitstream/handle/UNITRU/10073/COLONIA%20IPARRAGUIRRE%20C%20Denis%20Vanesa.pdf?sequence=1&isAllowed=y>
- Espin, O., & Rodríguez, V. (2013). Metodología para un scoring de clientes sin referencias crediticias. *32(137-162). Cuadernos de Economía*, 32(59), 137-162. https://www.researchgate.net/publication/262434088_METODOLOGIA_PARA_UN_SCORING_DE_CLIENTES_SIN_REFERENCIAS_CREDITICIAS
- González, X.V. (2015). *Diseño de un scoring de crédito para la cooperativa de ahorro y crédito "Crediamigo" LTDA (Tesis de maestría)*. Universidad San Francisco de Quito, Colegio de Postgrados (Quito, Ecuador). <http://repositorio.usfq.edu.ec/bitstream/23000/5053/1/121568.pdf>
- Lagua, V.M. (2015). *Diseño de un modelo de credit scoring en la gestión de riesgo crediticio en la cartera de microempresa de las cooperativas de ahorro y crédito*. Pontificia Universidad Católica

- del Ecuador (Ambato, Ecuador).
<https://repositorio.pucesa.edu.ec/bitstream/123456789/1524/1/76061.pdf>
- Lara, J. (2010). *La gestión de riesgo de crédito en las instituciones de microfinanzas (Tesis Doctoral)* Univesidad de Granada (España).
<https://digibug.ugr.es/bitstream/handle/10481/5648/18892656.pdf?sequence=1&isAllowed=y>
- Largo, M. (20 de Junio de 2015). *Contabilidad Bancaria y de Seguros*. [Entrada de blog] http://contabancariaseguros.blogspot.com/2015/06/3_47.html
- Leal, A.L., Aranguiz, M.A., & Gallegos, J. (2017). Análisis de Riesgo Crediticio. Propuesta del Modelo Credit Scoring. *Revista Facultad de Ciencias Económicas*, 26(1), 181-207.
<https://doi.org/10.18359/rfce.2666>
- Millán, J.C., & Caicedo, E. (2018). Modelos para otorgamiento y seguimiento en la gestión de riesgo de crédito. *Revista de Métodos cuantitativos para la economía y la empresa*, 25, 23-41.
<https://dialnet.unirioja.es/servlet/articulo?codigo=6499643>
- Navarro, H. E. (2015). *Diseño de un modelo scoring aplicado a creditos recurrentes para Caja Municipal (Tesis de grado)*. Universidad Nacional de Piura (Perú).
<http://repositorio.unp.edu.pe/bitstream/handle/UNP/674/IND-NAV-GUE-15.pdf?sequence=1&isAllowed=y>
- Pantoja, P.M. (2016). *Propuesta de un modelo logit para evaluar el riesgo crediticio en las Cajas Municipales de ahorro y crédito: Caso de la Caja Municipal de Huancayo, periodo 2011-2015 (Tesis de grado)*. Universidad San Ignacio de Loyola (Lima, Perú).
<https://repositorio.usil.edu.pe/server/api/core/bitstreams/9de9df94-ecc4-4dac-8e91-8ace940d16dd/content>
- Puertas, R., & Martí, M. (2013). Análisis del Credit Scoring. *Revista de Administração de Empresas*, 53(3), 33-315. <https://www.redalyc.org/articulo.oa?id=155127485011>
- Rodriguez, A.L. (2018). *Modelo scoring para recuperar cartera de microcrédito. (Tesis de maestría)*. Fundación Universitaria Los Libertadores (Bógota, Colombia).
https://repository.libertadores.edu.co/bitstream/handle/11371/2082/Rodriguez_Angelica_2018.pdf?sequence=1&isAllowed=y
- Rodríguez, D.E., Rendón, J.F., Trespalacios, A., & Jiménez, E.A. (2022). Modelación de riesgo de crédito de personas naturales. Un caso aplicado a una caja de compensación familiar colombiana. *Revista de Metodos cuantitativos para la economía y la empresa*, 33, 29-48.
<https://www.upo.es/revistas/index.php/RevMetCuant/article/view/5146/5390>
- Superintendencia de banco y seguros (2010). *Libro I: Normas generales para las instituciones del sistema financiero*. https://www.superbancos.gob.ec/bancos/wp-content/uploads/downloads/2017/06/L1_X_cap_I.pdf
- Torrico, S.E. (2014). Macro credit scoring como propuesta para cuantificar el riesgo de credito. *Revista Investigación y Desarrollo* 2(14), 42-63
http://www.scielo.org.bo/scielo.php?script=sci_arttext&pid=S2518-44312014000200004&lng=es&tlng=es.
- Train, K. (2009). *Métodos de elección discreta con simulación* (2ª ed.).
<https://eml.berkeley.edu/books/choice2nd/Combined.pdf>

Urbina, M. (2019). Riesgo de Crédito: Evidencia en el sistema bancario ecuatoriano. *Boletín de coyuntura*(23), 4-9. <http://dx.doi.org/10.31164/bcoyu.23.2019.842>