



República Del Ecuador

Universidad Tecnológica Empresarial de Guayaquil - UTEG

Facultad de Posgrado e Investigación

Tesis en opción al título de Magister en:

Sistemas de Información Gerencial

Tema de Tesis:

Modelo para la Evaluación del Riesgo en la Concesión de Operaciones Crediticias del Sector Cooperativo de la Ciudad de Loja, utilizando técnicas y herramientas de Minería de Datos.

Autor:

Ing. Liliana Juliza González Lapo

Director de Tesis:

Econ. Olmedo Farfán González

Septiembre 2019

Guayaquil - Ecuador

Declaración expresa

“La responsabilidad del contenido de esta Tesis de Graduación, me corresponden exclusivamente; y el patrimonio intelectual de la misma a la **“UNIVERSIDAD TECNOLÓGICA EMPRESARIAL DE GUAYAQUIL”**”.

(Reglamento de Graduación de la UTEG)

González Lapo Liliana Juliza
C.I. 1900478973

Dedicatoria

Este trabajo está dedicado primeramente a Dios y a su infinito amor, por haberme concedido la sabiduría necesaria en la ejecución de este trabajo de investigación; salud y bienestar durante mis largos viajes; fortaleza en esas largas noches y por haber puesto en mi camino ángeles que fueron mi sostén durante este trabajo, a la vez que me permite alcanzar una meta más en mi vida.

Cristalizar este sueño implicó dejar gratos momentos en familia, pero todo lo vale cuando el amor de familia siempre está ahí para darnos la motivación que necesitamos en cada nueva meta propuesta, por ello este trabajo también se lo dedico a mi hermosa familia, por ser el soporte y la inspiración para alcanzar mis sueños.

Agradecimiento

Quiero expresar un caluroso agradecimiento primeramente a Dios por la sabiduría y fortaleza brindada; seguidamente a mis padres Santos y Lili por sus sabios consejos, su amor y constante apoyo; a mi novio Santy por creer en mí y motivarme todos los días a alcanzar esta meta y no permitirme desmayar; a mi hermana Alejandra y a mi hermano Ever por la motivación brindada en este proceso; gracias a todos y cada uno de Ustedes por celebrar conmigo cada meta conseguida, por demostrarme que la vida es bella tal y como es, que con sus altos y con sus bajos sigue siendo indescriptiblemente fascinante.

Quiero también extender un agradecimiento especial a mi Director de Tesis el Econ. Olmedo Farfán González, MSc, por su guía brindada para el desarrollo del presente trabajo investigativo, quién además me alentó a culminar con éxito esta etapa.

Fueron muchas las personas que fueron partícipes de este proceso y que estuvieron junto a mí, gracias a todos por alentarme e incentivarme.

Resumen

El nivel de riesgo en las entidades financieras se puede ver afectado por la inadecuada evaluación del perfil del cliente como sujeto de crédito por parte de responsables en la institución. Cooperativa Ltda. denominada así por confidencialidad con la institución, es una Cooperativa de Ahorro y Crédito del Segmento 2 con presencia en la ciudad de Loja; que presta sus servicios en su matriz y en diez agencias dentro la ciudad y fuera de ella, su propósito; impulsar el desarrollo de sectores productivos de la región sur del país.

El presente trabajo estudia la utilidad de técnicas y herramientas de minería de datos como apoyo en el proceso de evaluación del nivel de riesgo en la otorgación de créditos, a través del análisis de la cartera de consumo de Cooperativa Ltda. con la construcción de un modelo de evaluación de riesgo de crédito utilizando la técnica estadística y de minería de datos de regresión logística (logit) y un análisis univariante y bivalente a través del análisis descriptivo y exploratorio de las variables recogidas y seleccionadas, con la herramienta de ciencia de datos Stata.

Con la metodología antes dicha con un estudio descriptivo y exploratorio de las variables en análisis y aplicación del modelo econométrico Logit de regresión logística y utilizando la herramienta Stata, se procedió a estimar un modelo, donde las variables independientes, traten de explicar el comportamiento de la variable dependiente dicotómica o dummy 0 y 1, de Riesgo Crediticio o Default donde 0 es “Mal cliente” y 1 “Buen cliente”.

Como resultado se presenta un modelo con su respectiva prueba de bondad para la calificación de un sujeto de crédito a fin de reducir la incertidumbre de incumplimiento del sujeto de crédito y mitigar pérdidas esperadas, dejando por demostrado la utilidad de la minería de datos en la evaluación del riesgo de crédito.

Palabras clave: Minería de datos, técnicas de minería de datos, riesgo de crédito, regresión logística, logit.

Abstract

The level of risk in financial institutions can be affected by the inadequate evaluation of the client's profile as a subject of credit by people who are in charge at the until. Cooperativa Ltda. It was named like that in order to protect the identify of the institution. It is a Credit Union od segment two in Loja city. It works in its main office and in ten more agencies which work in the city and out ofi t. The purpose is to promove the development of productive sectors in the southern región of the country.

This research studies the helpfulness of data mining techniques and tools to support the evaluation process of the risk level in granting loans, through the analysis of Cooperativa Ltda. Consumer portfolio with the construction of a model of credit risk assessment using the statistical and logistic regression data mining technique (logit) and a univariate and bivariate analysis through the descriptive and exploratory analysis of the collected and selected variables, with the Stata data science tool.

With the methodology mentioned above with a descriptive and exploratory study of the variables in analysis and application of the Logit regression logistic econometric model and using the Stata tool, we proceeded to estimate a model, where the independent variables, try to explain the behavior of the dichotomous or dummy dependent variable 0 and 1, Credit Risk or Default where 0 is "Bad customer" and 1 "Good customer".

As a result, a model with its respective proof of goodness is presented for the qualification of a credit subject in order to reduce the uncertainty of default of the credit subject and mitigate expected losses, leaving the utility of data mining in the evaluation demonstrated of credit risk.

Keywords: Data mining, data mining techniques, credit risk, logistic regression, logit.

ÍNDICE GENERAL DE LA TESIS

Declaración expresa	I
Dedicatoria.....	II
Agradecimiento.....	III
Resumen	IV
Abstract	V
ÍNDICE GENERAL DE LA TESIS	VI
Glosario de términos.....	XI
INTRODUCCIÓN.....	1
1.1. Antecedentes de la investigación	2
1.2. Planteamiento del problema.....	5
1.2.1. Formulación del problema de investigación	7
1.2.2. Sistematización del problema de investigación.....	7
1.3. Objetivos de investigación.....	9
1.3.1. Objetivo General.....	9
1.3.2. Objetivos específicos.....	9
1.4. Justificación de la investigación	10
1.4.1. Justificación teórica	10
1.4.2. Justificación práctica.....	10
1.5. Marco de referencia de la investigación	11
1.5.1. Riesgo de Crédito	13
1.5.2. Economía Popular y Solidaria.....	22
1.5.3. Principales Modelos de Evaluación del Riesgo Crediticio	27
1.5.4. Comparación de modelos aplicados en la evaluación de riesgo crediticio	36
1.5.5. Business Intelligence y Minería de Datos	38
1.5.6. Selección del modelo de evaluación.....	50
1.5.6.1. Definición de la problemática de riesgo crediticio a partir de la utilización de un instrumento.....	50
1.5.6.2. Definición de las variables dependiente e independientes basado en un modelo de atributos para la evaluación de riesgo crediticio	51
1.5.6.3. Diagrama del modelo inicial de evaluación propuesto para operacionalización	52
2.1. Tipo de diseño, alcance y enfoque de la investigación	57
2.1.1. Tipo de estudio	57
2.1.2. Metodología de la investigación.....	58
2.1.2.1. Enfoque de la investigación.....	58

2.2.	Métodos de investigación.....	58
2.3.	Unidad de Análisis, población y muestra.....	60
2.4.	Variables de investigación y operacionalización.....	62
2.5.	Fuentes, técnicas e instrumentos para la recolección de información	63
2.5.1.	Fuentes de información	64
2.5.2.	Técnicas para la recolección de información	65
2.6.	Tratamiento de la información.....	65
3.1.	Análisis de la situación actual.....	67
3.2.	Análisis comparativo, evolución, tendencias y perspectivas.....	74
3.3.	Presentación de resultados y discusión	84
	CONCLUSIONES	90
	RECOMENDACIONES	93
	REFERENCIAS BIBLIOGRÁFICAS.....	94
	ANEXOS.....	107

ÍNDICE DE CUADROS

CUADRO 1 PILARES FUNDAMENTALES BASILEA.....	18
CUADRO 2 SEGMENTACIÓN SECTOR FINANCIERO POPULAR Y SOLIDARIO.....	24
CUADRO 3 TIPOS DE CRÉDITO COOPERATIVAS	25
CUADRO 4 MODELOS DE EVALUACIÓN DE RIESGO DE CRÉDITO.....	28
CUADRO 5 VARIABLES RELEVANTES DE LOS PRINCIPALES MODELOS APLICADOS EN LA EVALUACIÓN DE RIESGO CREDITICIO EN EL SECTOR FINANCIERO.....	36
CUADRO 6 RELACIÓN DE LA VARIABLE DEPENDIENTE E INDEPENDIENTES	53
CUADRO 7 ANALOGÍA ENTRE KDD Y CRISP-DM	59
CUADRO 8 DE LA CALIFICACIÓN CARTERA DE CRÉDITO.....	63
CUADRO 9 ÍNDICES DE MOROSIDAD POR TIPO CARTERA COOPERATIVA LTDA.	71
CUADRO 10 TABLA DE VARIABLES PARA SU ANÁLISIS.....	74
CUADRO 11 DICCIONARIO DE VARIABLES Y CATEGORIZACIÓN.....	75

ÍNDICE DE FIGURAS

FIGURA 1 METODOLOGÍA RESUMIDA KDD (Knowledge Discovery from Databases)	40
FIGURA 2 ARQUITECTURA DE MINERÍA DE DATOS	42
FIGURA 3 CLASIFICACIÓN DE LAS	45
FIGURA 4 RELACIÓN FUNCIONAL DE LA VARIABLE DEPENDIENTE Y VARIABLE INDEPENDIENTE	51
FIGURA 5 CARTERA POR VENCER SECTOR COOPERATIVO	67
FIGURA 6 CARTERA IMPRODUCTIVA SECTOR COOPERATIVO	68
FIGURA 7 CARTERA IMPRODUCTIVA SECTOR COOPERATIVO LOJA.....	68
FIGURA 8 CARTERA POR VENCER SECTOR COOPERATIVO LOJA.....	69
FIGURA 9 VOLUMEN DE OPERACIONES POR PROVINCIA	69
FIGURA 10 MOROSIDAD SEGMENTO 2.....	70
FIGURA 11 CONCENTRACIÓN DE LA CARTERA COOPERATIVA LTDA.	71
FIGURA 12 SUJETOS DE CRÉDITO COOPERATIVA LTDA.	72
FIGURA 13 CARTERA POR VENCER COOPERATIVA LTDA.	72
FIGURA 14 CARTERA POR VENCER COOPERATIVA LTDA. LOJA	73
FIGURA 15 VARIACIÓN DEL INDICADOR DE MOROSIDAD COOPERATIVA LTDA.	73

ÍNDICE DE TABLAS

TABLA 1 DEFINICIÓN DE LA VARIABLE EN ANÁLISIS “DEFAULT”	77
TABLA 2 FRECUENCIAS DE LA VARIABLE ANALIZADA “GENE”	78
TABLA 3 FRECUENCIAS DE LA VARIABLE ANALIZADA “RANG_ED”	78
TABLA 4 FRECUENCIAS DE LA VARIABLE ANALIZADA “EST_CIVIL”	79
TABLA 5 FRECUENCIAS DE LA VARIABLE ANALIZADA “NIV_EST”	79
TABLA 6 FRECUENCIAS DE LA VARIABLE ANALIZADA “OCUP”	79
TABLA 7 FRECUENCIAS DE LA VARIABLE ANALIZADA “TIP_VIV”	80
TABLA 8 FRECUENCIAS DE LA VARIABLE ANALIZADA “TIP_GAR”	80
TABLA 9 TABLA DE CONTINGENCIA VARIABLE “GENE” * “DEFAULT”	81
TABLA 10 TABLA DE CONTINGENCIA VARIABLE “RANG_ED” * “DEFAULT”	81
TABLA 11 TABLA DE CONTINGENCIA VARIABLE “EST_CIVIL” * “DEFAULT”	82
TABLA 12 TABLA DE CONTINGENCIA VARIABLE “NIV_EST” * “DEFAULT”	82
TABLA 13 TABLA DE CONTINGENCIA VARIABLE “OCUP” * “DEFAULT”	83
TABLA 14 TABLA DE CONTINGENCIA VARIABLE “TIP_VIV” * “DEFAULT”	83
TABLA 15 TABLA DE CONTINGENCIA VARIABLE “TIP_GAR” * “DEFAULT”	84
TABLA 16 MODELO CON VARIABLES ANALIZADAS	85
TABLA 17 MODELO PREDICTIVO LOGIT	86
TABLA 18 LOGISTIC MODEL FOR DEFAULT	88
TABLA 19 TEST DE HOSMER Y LEMESHOW	89

Glosario de términos

Riego de crédito.- Se define como la probabilidad de pérdida por el no pago o incumplimiento total o parcial de la obligación crediticia del sujeto de crédito.

Default.- Término usualmente utilizado para evaluación de riesgo crediticio o en finanzas cuando el prestatario o sujeto de crédito no hace frente a su obligación de pago debido a la falta de liquidez o efectivo donde se suspende los pagos.

Minería de datos.- Pertenece al descubrimiento de patrones o comportamientos en grandes volúmenes de datos estructurados o no estructurados que facilite la toma de decisiones.

Técnicas de minería de datos.- Las técnicas son las herramientas aplicadas a la minería de datos procedentes de la inteligencia artificial y la estadística aplicadas a grandes volúmenes de datos para la obtención de patrones, entre ellas están modelos estadísticos, árboles de decisión, regresión lineal, clustering, redes neuronales, entre otros.

Modelos de evaluación de riesgo de crédito.- Son esquemas que facilitan el análisis de ciertos factores de un todo, y en este caso identificar patrones del riesgo de crédito en una cartera, con el ánimo de evitar pérdidas a futuro.

Stata.- Es un Software para estadística and ciencia de datos que permite realizar analítica predictiva con la manipulación de datos para obtener patrones, comportamientos y obtener información útil, y que puede servirle a una institución para iniciar un proyecto de minería de datos.

Qlik Sense.- Es una analítica de datos basada en BI y minería de datos, que permite descubrir conocimiento a través de la exploración y relación de datos.

INTRODUCCIÓN

Las Cooperativas de Ahorro y Crédito pertenecen al Sector Popular y Solidario, son reguladas por la Superintendencias de Economía Popular y Solidaria SEPS, a nivel de toda Latinoamérica, Ecuador es el segundo país con más Cooperativas de Ahorro y Crédito (SEPS, sf), y para nadie es desconocido que uno de los principales riesgos a los que se enfrenta como institución financiera es a que un cliente incumpla con sus pagos (riesgo de crédito).

Esta investigación pretende contribuir a mejorar la calidad de la cartera de crédito en Cooperativas de Ahorro y Crédito y demostrar que el apoyo de técnicas y herramientas de minería de datos pueden ser muy útiles en la construcción de un modelo de evaluación de riesgo de crédito para determinar aquellos patrones, comportamientos de clientes que inciden en la medición de la variable “default” que básicamente es la probabilidad de que un sujeto de crédito incumpla total o parcialmente con sus obligaciones crediticias.

El desarrollo de esta investigación recopila información teórica respecto a modelos de evaluación de riesgo de crédito y minería de datos, y propone un conjunto de variables a ser analizadas para la construcción del modelo con la aplicación de la técnica de minería de datos de regresión logística (logit) y el uso de la herramienta Stata.

El documento se encuentra organizado en tres capítulos: Un primer capítulo que comprende el marco teórico conceptual donde se aborda la importancia y propósito del tema, recopilación bibliográfica relacionada con el objeto de estudio, una comparación de modelos aplicados en la evaluación del riesgo de crédito, así como la selección del modelo para evaluación; un segundo capítulo correspondiente al marco metodológico; y finalmente, el tercer capítulo con los resultados obtenidos, así como las conclusiones y recomendaciones al trabajo desarrollado.

CAPÍTULO I. MARCO TEÓRICO CONCEPTUAL

1. Marco teórico

1.1. Antecedentes de la investigación

La innovación tecnológica tiene un constante avance en la sociedad, su utilidad mejora la vida de muchas personas en distintos roles; pero esta innovación, no fuese posible sin los “datos”, pues ellos se han convertido en la oportunidad que aprovechan personas creativas para el desarrollo de nuevas tecnologías, que beneficien a distintos sectores.

Estas nuevas tecnologías, incluyen herramientas y soluciones prácticas que facilitan la vida de las personas y de las empresas. Los datos se encuentran en todas partes, son omnipresentes y su valor es incalculable; y conocedores de la importancia y el valor que tienen los “datos”, invierten en nuevas tecnologías para el almacenamiento de estos; que posterior a su transformación, puede tener múltiples aplicaciones, brindando nuevos conocimientos y soluciones ilimitadas, inclusive beneficios lucrativos si las empresas saben aprovecharlos y utilizarlos para optimizar sus estrategias.

En este sentido, el ser humano como un requerimiento económico para maximizar sus ganancias dentro las empresas, tomar decisiones con mayor certidumbre, reducir el impacto en sus principales activos de riesgo, etc., ha propiciado la necesidad de estudiar novedosas técnicas para gestionar el flujo de su información (datos transformados) propia de las transacciones de su negocio. La minería de datos es una de ellas, y se la define como el conjunto de metodologías, aplicaciones y tecnologías que permiten reunir, depurar y transformar datos de los sistemas transaccionales e información no estructurada (interna y externa a la compañía) en información estructurada, para su explotación directa o para su análisis y conversión en conocimiento y así dar soporte a la toma de decisiones sobre el negocio (Marcano, Y. & Talavera, R. 2007). El big data se ha convertido en un tema de importancia a nivel mundial, cada día su crecimiento e importancia es innegable (Hernández

& Duque & Moreno, 2017). Esta permite el descubrimiento de nuevo conocimiento, a través de la preparación de datos y obtención de resultados mediante la aplicación de sus diversas técnicas y herramientas. Molina y García (2006) explican que los datos tal cual se almacenan en las bases de datos no suelen proporcionar beneficios directos; su valor real reside en la información que podamos extraer de ellos; información que nos ayude a tomar decisiones o a mejorar la comprensión de los fenómenos que nos rodean. El Descubrimiento de Conocimiento en bases de datos consiste en un análisis automático exploratorio además de un modelado de grandes repositorios de datos que involucra áreas de conocimiento como la inteligencia artificial, aprendizaje automático, estadística, sistemas de gestión de base de datos, técnicas de visualización de datos y otros medios existentes que apoyan las decisiones de una empresa (Cagnina et al 2019).

Sus beneficios han sido aprovechados en diversas áreas, debido a la complejidad del mercado y a las actividades de cada institución. El negocio bancario es una de ellas, debido a sus constantes variaciones y estándares económicos en su principal activo de riesgo, "El crédito". Las entidades financieras operan bajo entornos macroeconómicos altamente volátiles, en donde las tasas de interés, tasas de cambio y algunas otras variables (exógenas y endógenas) presentan fluctuaciones considerables y de difícil predicción (Pérez, Fredy & Fernández, H. 2007). La administración del riesgo de crédito implica un análisis de incumpliendo de una obligación por parte del cliente (Ayala, L. & Sarmiento, L., 2015). es por ello que una de las variables más analizadas asociadas a la calidad de la cartera es "Default", que es la probabilidad de que el deudor incumpla con su obligación, para su análisis se considera variables de los sujetos de crédito de cada portafolio, se ordena a los sujetos según sus categorías de riesgo (Tamariz, 2019).

Un sistema financiero saludable es crucial para el desarrollo de la economía de cualquier país, y es que este sector usa ahorros que pertenecen a sus clientes para la colocación de créditos, de esta forma fortalece a diferentes sectores económicos, proveen del capital necesario a los inversionistas que llevan a cabo actividades productivas que contribuyen al crecimiento de la economía;

de allí la importancia que las instituciones que pertenecen al sistema financiero, establezcan modelos internos o adopten estándares para una adecuada valoración del riesgo crediticio, que permita determinar la probabilidad de impago de un sujeto de crédito, en aras de tomar acciones preventivas y controlar su índice de morosidad.

La Superintendencia de Economía Popular y Solidaria mide el índice de morosidad como el porcentaje de la cartera total improductiva frente a la cartera total bruta. Los ratios de morosidad se calculan para el total de la cartera bruta y por línea de negocio (SEPS, 2015).

El índice de morosidad de los prestatarios es una variable que impacta de manera negativa y significativa a la rentabilidad obtenida por los bancos y es la consecuencia de las decisiones ligadas a la concesión del crédito, (Chavarín, 2015). Según (Guillén Uyen, 2001), la morosidad crediticia constituye una de las más importantes señales que advierten las crisis financieras. Además, se plantea que los problemas en la rentabilidad se agudizan si se toma en cuenta que la entidad financiera incrementa su cartera de créditos impagos, lo que resulta en la caída de las utilidades.

Entonces, un aumento significativo en la morosidad de las instituciones del sistema financiero puede provocar un problema de rentabilidad y liquidez, incluso se traduciría en la generación de pérdidas y déficit de provisiones, hasta pérdidas económicas por ser un riesgo de tipo sistémico.

En este sentido, la presente investigación muestra la aplicación y bondades de técnicas y herramientas de minería de datos en el sistema financiero (Sector Financiero Popular y Solidario), con el ánimo de obtener un modelo predictivo para la evaluación de riesgo crediticio para determinar la probabilidad y variables que explican la variable de "Default" que se define en un buen o mal cliente.

Múltiples estudios se han realizado debido a la importancia del tema y a continuación se pone en manifiesto algunos de ellos:

- Modelo de Evaluación de Créditos Financieros basados en Redes Neuronales orientado a Edpymes, desarrollado por Huayna, A.; Calvo V. & Huiman, J. Universidad Nacional Mayor San Marcos.
- Redes Neuronales y la Evaluación del Riesgo de Crédito de Pérez, F. & Fernández, H. Universidad de Medellín, Colombia.
- Desarrollo de un modelo predictivo para detección de riesgo crediticio utilizando redes neuronales para una entidad financiera. Gaona, R. Universidad Técnica Particular de Loja.
- Métodos para estimar riesgo crediticio en base a Minería de Datos y Teoría de Juegos. Bravo, C. Universidad de Chile.
- El análisis inteligente de datos y su incidencia en el otorgamiento de créditos financieros en la Cooperativa de Ahorro y Crédito Maquita Cushun Ltda, de Hidalgo, A. Universidad Técnica de Ambato.
- Evaluación del Riesgo Crediticio, a través de Credit Scoring mediante Regresión logística: un caso de estudio.
- Un Modelo de Credit Scoring para instituciones de microfinanzas en el marco de Basilea II.
- Aplicación de la Técnica Regresión Logística de la Minería de Datos en proceso de descubrimiento de conocimiento (KDD) en Bases de Datos Operativas Transaccionales de Juan Camilo Giraldo Mejía y Fabio Alberto Vargas Agudelo.

1.2. Planteamiento del problema

La Administración del Riesgo de Crédito es un aspecto fundamental para cualquier institución financiera, dado que su correcta gestión entre otros, supone elevar la calidad de la cartera crediticia, lo que implica además disminuir los índices de cartera improductiva; para la (Superintendencia de Bancos y Seguros del Ecuador, 2014) el riesgo de crédito “es la posibilidad de incurrir en pérdidas, como consecuencia del incumplimiento de las obligaciones por parte del deudor en operaciones de intermediación financiera”.

La evaluación del riesgo de crédito se basa en la probabilidad de que el prestatario o emisor del bono incumpla con sus obligaciones (ocurra un default) (Saavedra, M. & Saavedra, M, 2010). La probabilidad de incumplimiento (default) es una de las variables más analizadas por las instituciones financieras al momento de calificar y aprobar a un sujeto de crédito por su alta vulnerabilidad. Para estimar el nivel de riesgo del sujeto de crédito, se asigna una probabilidad de default, este indicador se usa para medir cuando el prestatario no podrá cumplir con sus obligaciones (Solarte, M. & Cerezo, E., 2018).

La calidad de la cartera puede verse afectada por múltiples razones, algunas investigaciones previas sugieren que el comportamiento de pago de los individuos (y la tasa de default asociada) depende tanto de variables de tipo financiero (en general vinculadas al ingreso) como de variables de carácter personal (edad, estado civil y género, entre otras). (Superintendencia de Bancos e Instituciones Financieras, 2014).

Un estudio desarrollado por la Superintendencia de Bancos e Instituciones Financieras Chile, concluye que, para todos los segmentos crediticios, la tasa de incumplimiento del segmento de mujeres es inferior a la exhibida por el segmento de clientes hombres, por lo que al mismo tiempo se efectuó un análisis, respecto al efecto sobre la tasa de incumplimiento de variables como estado civil y tamaño e ingreso.

Estos antecedentes ponen en manifiesto la necesidad de medir la probabilidad del riesgo default, con el ánimo de determinar cuáles son las variables de carácter cuantitativo y/o cualitativo que inciden y cual no, en un modelo de evaluación del riesgo crediticio.

En el presente trabajo se analiza la concentración y exposición del riesgo, de un sector que en los últimos años ha tenido un importante crecimiento en la economía ecuatoriana. El Sector Cooperativo ha jugado un papel clave en la economía del país, ha llegado a ocupar un poco más del 16% en relación al PIB para diciembre del 2014 (SEPS, 2015), a enero de 2016, estuvo integrada

por un total de 880 Cooperativas de Ahorro y Crédito (SEPS, 2016), conformada por cinco segmentos de conformidad a lo que establece la Junta de Política y Regulación Monetario y Financiera en la Resolución No. 038-2015-F del 13 de febrero del 2015.

El sector financiero popular y solidario muestra una evolución positiva de sus activos, pasivos y patrimonio. Los activos de las cooperativas de ahorro y crédito pasaron del 2013 al 2016, del 7% al 11,2%, llegando a 9.897 millones de dólares, con una cartera de crédito de 6.420 millones de dólares, a diciembre de 2016 (SEPS, 2017), de acuerdo con los datos constantes en los boletines mensuales de la Superintendencia de Economía Popular y Solidaria, el riesgo crediticio, el índice de morosidad en tan solo tres meses tuvo variaciones positivas en los segmentos tales como microcrédito, y consumo prioritario (SEPS, 2016); ello se evidencia que la calidad de la cartera puede verse afectada en cortos períodos de tiempo, debido al giro del negocio siempre este tipo de instituciones siempre se verán expuestas a problemas de liquidez y de rentabilidad sino gestionan adecuadamente la calidad de su cartera, por la probabilidad de incumplimiento de sus clientes crediticios.

1.2.1. Formulación del problema de investigación

¿La calidad de la cartera de crédito puede verse afectada debido a la falta de tecnologías de crédito adecuadas para la evaluación de riesgo default?
¿Cuáles son los determinantes que inciden en la probabilidad del riesgo “default”?

1.2.2. Sistematización del problema de investigación

Una de las mayores preocupaciones de las Instituciones Financieras es el riesgo de crédito, mantener el control del riesgo implica la implementación de modelos, políticas y procedimientos en el proceso de crédito, que contribuyan eficaz y eficientemente a la toma de decisiones en la calidad de la cartera de crédito. La cartera de crédito es el activo de mayor riesgo en una entidad

bancaria, y el índice de morosidad puede sufrir grandes deterioros en lapsos de tiempo muy cortos, lo que puede afectar significativamente la liquidez y la rentabilidad de la institución, y la situación económica de un país.

Cuando los ciclos económicos son muy favorables y el ritmo de colocaciones se acelera, el riesgo del portafolio crediticio tiende a aumentar ya que los créditos otorgados a deudores con dudosa capacidad de pago se incrementan bajo estos escenarios (Aparicio, C & Moreno, H., 2011). Los bancos a fin de que la salud financiera de la institución no se vea afectada deben controlar el crecimiento del indicador de morosidad (Chavarín, 2015). En este sentido, se hace necesario investigar y analizar los factores que inciden en la probabilidad de riesgo “default” que hacen que incrementen la vulnerabilidad de la cartera crediticia (cartera improductiva); dado que diferentes factores pueden afectar la calidad de la misma, factores macroeconómicos como el desempleo, tasas de interés, inflación, etc. o microeconómicas como a) *falta de experticia del oficial que concede el crédito*, en muchos de los casos en su afán de incrementar colocaciones atadas a conseguir un mayor margen de comisiones omiten realizar una adecuada valoración del sujeto crediticio, es por ello que para que el nivel de calidad de la inversión (préstamos y créditos) sea el adecuado es preciso contar con un equipo humano capacitado (UNED, 2015), o muchas veces lo que ocurre es el exceso de confianza de las instituciones hace la experticia del oficial al momento de calificar un crédito; b) *escasas técnicas y metodologías de evaluación por parte de la institución para permanecer en un nivel aceptado de riesgo* (García, J. & Salazar, P., 2005), que suplanten efectivamente el criterio del oficial o del analista de crédito, c) *falta de garantías y procesos suficientes que respalde el crédito* (Vela, L. & Uriol, C. 2012), etc., establecer políticas adecuadas en la gestión de Cartera de crédito favorecerá la determinación del perfil del cliente que está en condiciones económicas adecuadas para cumplir el pago correspondiente (Cortez, D. & Burgos, J. 2016). A esto se suma que d) *la banca privada, y demás instituciones del sistema financiero no usan fuentes de información crediticia integral y eficiente* (ASBA, 2008), además de sistemas inadecuados para la evaluación del riesgo de crédito que proporcione información estratégica para la colocación de créditos y selección de créditos, la incertidumbre en el proceso

de selección, aprobación del crédito conlleva molestias, demoras y problemas no únicamente para el solicitante sino para la institución financiera (Gil, 2017).

En este sentido se hace necesario formular las siguientes interrogantes, ¿Existen técnicas y herramientas de minería de datos para la construcción de modelos de evaluación de riesgo de crédito que permitan identificar comportamientos que expliquen la probabilidad de “default”? ¿Cuáles son las variables y comportamientos que inciden en la medición de la probabilidad de riesgo default? ¿Las Cooperativa de Ahorro y Crédito pueden mejorar sus procesos de evaluación de riesgo de crédito con la tecnología de Minería de datos, que minimice la incertidumbre de pérdidas esperadas?

1.3. Objetivos de investigación

1.3.1. Objetivo General

Efectuar un modelo de evaluación de riesgo crediticio para el sector cooperativo con presencia en la ciudad de Loja, con el uso de técnicas y herramientas de minería de datos, con la intención de determinar patrones y/o comportamientos en los socios que minimice la probabilidad de una pérdida.

1.3.2. Objetivos específicos

- ✓ Analizar modelos previamente estudiados para la evaluación del riesgo crediticio y técnicas de minería de datos.
- ✓ Realizar un análisis de la concentración de riesgo crediticio y la morosidad para el objeto de estudio.
- ✓ Determinar las variables que inciden en la probabilidad de riesgo “default” con el uso de la base de datos de la Cooperativa Ltda, para la construcción del modelo de evaluación de riesgo de crédito para el sector cooperativo, con la aplicación de minería de datos.

1.4. Justificación de la investigación

1.4.1. Justificación teórica

El sector financiero tiene la increíble capacidad de generar grandes volúmenes de datos, lo que a su vez puede representar un gran obstáculo sino se sabe cómo aprovecharlos, es cuando se hace necesario la adopción de ciertas técnicas y/o herramientas que le brinde información útil a los procesos que son los encargados de agregar valor dentro de la empresa, y el Big Data sin duda aporta novedosas capacidades analíticas, que a la vez permiten generar nuevos conocimientos y aplicarlos a la solución de los riesgos que las entidades puedan afrontar, para obtener así ventajas competitivas (EY, 2014) y al ser esta una disciplina que está en constante desarrollo, resulta un tema interesante de estudiar dentro de los sistemas de información.

El tema de estudio se justifica además en la importancia que tiene el control del riesgo crediticio para las Instituciones Financieras, monitorear el riesgo de crédito es fundamental para preservar la estabilidad del sistema financiero (SBS, 2013), y dentro de este aspecto los beneficios que le puede brindar la minería de datos a partir de los datos históricos de créditos ya colocados a instituciones financieras del sector cooperativo, para tomar mejores decisiones en el monitoreo de la cartera y evitar que el riesgo crediticio se dispare por la cultura de impago que puedan tener ciertos clientes.

En este sentido, la presente investigación pretender capturar conocimientos elementales del Big Data y a través de sus métodos y técnicas de minería de datos aportar información útil a estas instituciones con el aporte de un modelo de evaluación de riesgo de crédito.

1.4.2. Justificación práctica

El crédito es uno de los elementos clave en el crecimiento de la economía, pero uno de los principales problemas que afrontan las instituciones crediticias a

nivel nacional es el riesgo de crédito (Default), y esto no es una excepción en el sector cooperativo (sector bastante desarrollado en el país), llegando en ciertas ocasiones a producir hasta el cierre o quiebra de las entidades. Una de las causas principales son los débiles mecanismos de evaluación que existen para la otorgación de créditos, siendo el personal de crédito quien toma la decisión final al momento de seleccionar un cliente. A esto se suma que las entidades financieras cuentan con información (estadísticos) que no se evalúan adecuadamente y en función de los cuales se podría establecer o ajustar nuevas variables que permitan correlacionar y al final controlar los índices de morosidad en las entidades financieras.

En este sentido la presente investigación busca partir de una evaluación de uno de los segmentos de crédito más sensibles asociar conceptos elementales de Minería de datos que es un campo en constante evolución y bastante interesante para la extracción de información relevante de un gran volumen de datos, lo que se pretende es mostrar la aplicación de estas técnicas en la evaluación del riesgo crediticio, a fin de determinar las principales variables que influyen en la calidad de la cartera, mantener el indicador de morosidad controlado le garantiza la sostenibilidad a las instituciones financieras, y a la vez generar la fidelidad de sus cuenta ahorristas que son quienes inyectan el capital necesario para la colocación de créditos con el respaldo de una institución sólida. La minería de datos puede otorgar a la institución ciertos elementos y variables de análisis y reducir la incertidumbre de si el cliente es factible o no de un crédito, incluso predecir situaciones futuras en la calidad de la cartera de crédito, que no resultan tan evidentes con datos masivos y poco ordenados.

1.5. Marco de referencia de la investigación

La evaluación y administración del riesgo crediticio ha sido un factor clave desde décadas atrás, basta con mencionar modelos como Rosenberg & Gleit (1994); Merton (1974), Hand & Jacka (1998); Thomas Crook & Edelman (1992); Lewis (1992); Mays (1998), Falkenstein, (1997); Galicia, (2003), Morgan (1997). Sin embargo, dada la importancia de una adecuada administración del

riesgo de crédito y el control de la calidad de la cartera, su vigencia es innegable y cada día con nuevas y mejores metodologías establecer controles más adecuados está al alcance de las instituciones. Según Sudhamathy (2016) la evaluación de riesgo de crédito es crucial, en virtud; que los bancos presentan muchos riesgos que influyen en el impago de un solicitante, y esta evaluación permite al banco determinar si el deudor puede ser moroso o no, y la minería de datos al permitir la extracción de conocimiento útil puede ser una herramienta que permita predecir a los solicitantes. Hussain, (2014) destaca que: “Una solicitud es aceptada o rechazada en base a la expectativa de que el solicitante es capaz o no capaz de pagar su obligación financiera”. Un estudio realizado por Aboobyda & Mohammed (2016) hace hincapié en señalar que: “El proceso de revisión continua y la clasificación de los préstamos permite controlar la calidad de las carteras de crédito y tomar medidas para contrarrestar la caída de la calidad crediticia de las carteras”. Pese a que varios autores señalan la importancia de la evaluación de riesgo de crédito, Angelini et al. (2008) concluye en que el riesgo de crédito genera una amenaza pese a las avanzadas técnicas de evaluación de riesgo crediticio y diversificación. El riesgo financiero es una pérdida monetaria por lo general asociado cuando una persona o empresa no cancela el crédito otorgado por el banco (Díaz & Del Valle Guerra, 2017). El factor interno de mayor relevancia es la deficiente evaluación crediticia realizada por los analistas (Ticona, 2018). Con el ánimo de identificar los patrones que influyen en el deterioro de la calidad de la cartera y la probabilidad del riesgo “default” y establecer modelos de evaluación de riesgo más certeros, varios autores han apostado al uso de técnicas y herramientas de minería de datos para la construcción de modelos de evaluación y la predicción del riesgo crediticio que sirva de apoyo al juicio humano, tal es el caso del estudio realizado por Tello, Eslava y Tobías, (2012) ante la comparación de sus resultados, concluye que algoritmos de minería de datos como ID3 – Árbol de decisión provee una gran precisión en la clasificación de clientes de crédito, por su lado Hussain (2014) muestra las redes neuronales como una de las técnicas de minería de datos y estadísticas más prometedoras para la clasificación de crédito y apoyar en la toma de decisiones, metodologías como logit que utiliza técnicas de minería de tipo predictivo permiten mejorar la administración del riesgo de crédito (Pantoja,

2016) y así como estos podemos mencionar un sin número de investigaciones donde se ha demostrado la utilidad de las técnicas de minería de datos en la banca; la minería de datos hoy por hoy es utilizada como una tecnología de apoyo para explorar, analizar, obtener, comprender, y aplicar conocimiento en datos de grandes volúmenes, (Rodríguez, Y. & Díaz, A. 2011) y la aplicación de estos métodos en la predicción de riesgo crediticio ha arrojado resultados satisfactorios (Lanzarini et al 2018).

La importancia de la gestión y/o administración del riesgo crediticio es asegurar que las instituciones financieras no se vean involucradas en cuantiosas pérdidas económicas por una mala o deficiente gestión de su principal activo de riesgo. Es así que en la presente investigación se pretende analizar los factores que inciden en el deterioro de la calidad de la cartera de crédito de las Cooperativas de Ahorro y Crédito, con el uso de técnicas y herramientas de minería de datos.

En esta sección se presenta revisiones literarias en la administración y evaluación del riesgo crediticio empezando desde Basilea que ha sido un pilar fundamental en la supervisión bancaria y la gestión de riesgos; aborda así mismo la conceptualización del sector cooperativo, se presentan modelos propuestos en estudios realizados para la gestión del riesgo crediticio, hasta abordar la eficacia y utilidad de la minería de datos en la banca en el campo de la evaluación sobre la concesión de un crédito o no.

1.5.1. Riesgo de Crédito

Uno de los principales riesgos en los que incurre una institución financiera es la posibilidad de que el deudor o sujeto de crédito no asuma su obligación de pago, sea que este incumplimiento sea proporcional o total, se generarán pérdidas. En este sentido la (Superintendencia de Bancos y Seguros del Ecuador, 2015) define al riesgo de crédito como “la posibilidad de incurrir en pérdidas, como consecuencia del incumplimiento de las obligaciones por parte del deudor, en operaciones de intermediación financiera. Este incumplimiento

se manifiesta en el no pago, el pago parcial o la falta de puntualidad en el pago de las obligaciones pactadas.” (SBS, 2015).

Saavedra & Saavedra, (2010) explica al Riesgo Crediticio como la incertidumbre o probabilidad de que el prestatario no pueda cumplir con sus obligaciones en los periodos establecidos, se de forma parcial o en su totalidad. Chorafas, (2000) citado por Saavedra & Saavedra, (2010) enfatiza que el riesgo de crédito es la probabilidad de que, a su vencimiento de cuota, una entidad no haga frente, en parte o en su totalidad de su obligación pendiente o rendimiento previamente legalizado con un instrumento financiero, debido a quiebra, iliquidez, falta de voluntad o alguna otra razón. El riesgo de crédito se define como la probable pérdida que puede sufrir el prestamista si las cláusulas contractuales del crédito se incumplen por el sujeto de crédito, incluso si sus contrapartes o terceros obligados (garantes) también incumplen (Vela & Caro, 2015).

Una Institución Financiera puede sufrir pérdidas no controladas por la falta de pago del deudor. Un ejemplo de riesgo crediticio constituye cuando se otorga un crédito de consumo y existe la probabilidad de que el socio en cuestión incumpla y entre en morosidad, y es precisamente esto lo que busca medir el riesgo de crédito bajo diferentes técnicas o modelos. (Córdova, 2009). (Galicía, 2003), citado por Saavedra & Saavedra (2010) manifiesta que el riesgo de crédito puede ser evaluado en tres aspectos:

- **Riesgo de incumplimiento:** se define como la probabilidad de que el prestatario no cumpla su obligación crediticia según los plazos acordados en el contrato, para lo cual las entidades conceden periodo de gracia antes de ser declarado como incumplimiento definitivo.
- **Exposición:** El autor define a este aspecto como la incertidumbre que existe respecto a los montos futuros de riesgo, usualmente una operación crediticia se amortiza de acuerdo con las fechas establecidas para pago y así conocer el remanente de riesgo, caso que no pasa con los créditos que se conceden u otorgan a tarjetahabientes y líneas de

crédito revolvente, donde el crédito fluctúa dependiendo de las necesidades del cliente y se desconoce su fecha exacta de liquidación.

- **Recuperación:** este factor se da posterior a la declaración de incumplimiento y según su autor depende del tipo de garantía que se haya recibido por parte del cliente y la situación a la fecha del incumplimiento lo que hará que su riesgo sea menor.

Las normas Generales para las Instituciones del Sistema Financiero señalan que el riesgo de crédito debe ser evaluado de forma permanente a través de metodologías por cada entidad y por cada tipo de crédito (SBS, 2018). Como antes se describe, varios autores concuerdan que para la banca el conceder recursos financieros en calidad de préstamos, siempre será un riesgo, ocasionado incertidumbre de lo que pueda ocurrir en un futuro, estos valores siendo este su negocio y pilar de supervivencia, su éxito dependerá de la adecuada evaluación del riesgo y que tan efectiva es su administración, el riesgo “default” es inevitable, y para contrarrestar el efecto de una pérdida esperada es necesario la evaluación y el comportamiento de cada cliente para determinar su capacidad de pago para el cumplimiento de la obligación (Cardona, 2006). Santillán y Gonzales (2013) manifestaron que una de las principales causas de la morosidad es una mala evaluación financiera esto sumado a factores exógenos del entorno socioeconómico.

El deterioro de la cartera también se ve amenazado por un conjunto de factores internos y externos (Ticona, 2018). En los últimos años la banca ha avanzado y ha trabajado en la regulación y supervisión del riesgo de crédito, como para Díaz Quevedo (2009) el riesgo crediticio es el elemento más importante siendo vital que la entidad financiera gestione y administre con mejores métodos de evaluación y análisis del riesgo para evitar graves problemas de solvencia. Basilea (1999) recomienda que la administración y modelos para evaluar el riesgo de crédito dependerán del tipo y características de cada institución. Dada la variedad de modelos para la evaluación del riesgo de crédito, su éxito está en la calidad del algoritmo y de un sistema eficiente de análisis de datos (Ludovic & Aranguiz & Gallegos, 2018).

Comité de Basilea y la supervisión del riesgo

Con el pasar de los años el sector financiero ha enfrentado grandes retos en cuanto a gestión de riesgos se refiere, el Banco Internacional de Pagos a través del Comité de Basilea, organismo que surgió en el año de 1975 por los Gobernadores de los Bancos Centrales del Grupo de los 10¹, consciente de este escenario con frecuencia y a través de varias reformas emite directrices y recomendaciones para la gestión de riesgos financieros, donde incluye requerimientos mínimos para la gestión del riesgo crediticio. El otorgamiento de operaciones de crédito para muchas entidades financieras es el instrumento de penetración y profundización de mercado, y tal sentido, la fuente de mayor riesgo de pérdidas y afectación patrimonial, allí que surge la incertidumbre cuando no se cuenta con el personal capacitado, o asesoramiento adecuado para mitigar los riesgos de irrecuperabilidad que puede generar el impago del deudor (Olano, 2018). El Acuerdo de Basilea parte de unos principios establecidos en cuanto al riesgo de crédito, con el ánimo de establecer en el ámbito internacional los requisitos mínimos para que los bancos cuantifiquen y monitoreen aquel riesgo (Pérez, F. & Fernández, H. 2007).

Para el año 2000, Ecuador debido a los conflictos que ocasionaron gobiernos muy permisivos en la regulación financiera, sumado a la crisis financiera internacional, el país se motivó en la implementación y adaptación de las regulaciones internacionales emitidos por el Comité de Regulación Bancaria de Basilea, que para el año de 1996 ya había emitido los 25 principios fundamentales para la supervisión y regulación bancaria, principios que sirven a los países como referencia para evaluar la calidad de sus sistemas supervisores, además de ofrecer un estándar integral sobre el cual sentar bases de regulación, supervisión, buen gobierno y gestión de riesgos para el sector bancario (BIS, 2011). El Comité, con el principio No. 8. Hablaba sobre el Riesgo de crédito haciendo énfasis en que los bancos cuenten con un proceso de control y mitigación del riesgo crediticio, acorde al perfil de riesgo de la institución donde incluya además el riesgo de contraparte.

¹ G10 lo integra: Bélgica, Canadá, Francia, Italia, Japón, Países Bajos, Reino Unido, Estados Unidos y los bancos centrales de Alemania y Suecia

Los acuerdos de Basilea evolucionaron a medida que los retos del sistema financiero se incrementaron a fin de garantizar la capacidad de respuesta de las instituciones financieras ante el riesgo operacional, de crédito y de mercado, en marzo del 2011 el Comité a través de Core Principles Group actualizó los principios básicos y se incrementaron a 29 principios.

En muchos países, el riesgo de crédito es evaluado y modelado en base a las primicias de los Acuerdos de Basilea I y II. Los Métodos IRB Básico y Avanzado, buscando que las entidades del sector financiero estimen las pérdidas esperadas y no esperadas, con el objeto de implementar una adecuada administración del riesgo.

Basilea I documento publicado en el año de 1988 establecía una ponderación para medición del riesgo por categorías para todos los activos que mantenía el banco, según el tipo de riesgo del deudor, esta metodología fijaba un ratio de capital para los bancos en un mínimo del 8% sobre los activos ponderados de riesgo, con un solo fin y era el de regular el nivel de solvencia y disminuir el riesgo de crédito (Pérez, F. & Fernández, H. 2007). Las recomendaciones que contemplaba este acuerdo era fijar un capital y estándares mínimos con la finalidad de que las autoridades nacionales de supervisión de cada institución financiera pueda implementarlos en sus respectivos países (Olano, 2018).

$$\frac{\textit{Capital mínimo regulador}}{RC + RM} \geq 8\%$$

Donde:

RC: Activos ponderados por riesgo crediticio

RM: Activos ponderados por riesgo de mercado

En el Ecuador, según lo establece las Normas Generales para las Instituciones del Sistema Financiero², sobre el patrimonio técnico, las instituciones financieras públicas y privadas, las compañías de arrendamiento mercantil, las

² Normas Generales para las Instituciones del Sistema Financiero, Capítulo I.- Relación entre el patrimonio técnico total y los activos y contingentes ponderados por riesgo para las instituciones del sistema financiero (sustituido con resolución No JB-2001-382 de 8 de octubre del 2001)

compañías emisoras o administradoras de tarjetas de crédito y las subsidiarias o afiliadas del exterior de las instituciones financieras del Ecuador, deben establecer obligatoriamente sobre la base de los estados financieros consolidados e individuales, un nivel mínimo de patrimonio técnico total equivalente al nueve por ciento (9%) de la suma total de los activos y contingentes ponderados por riesgo.

La categorización de los activos según Basilea I, se estableció según su grado de riesgo en (0%, 10%, 20%, 50% y 100%), con el ánimo de estandarizar el cálculo del riesgo crediticio a las ponderaciones; sin embargo, este acuerdo no consideraba la repercusión de la reducción del riesgo en cuanto a la concentración y diversificación de la cartera; así como el requerimiento de capital no podía ser siempre el mismo considerando los perfiles de riesgo de cada institución, siendo así que para junio del 2004 el acuerdo fue sustituido por el Nuevo Acuerdo de Capital conocido como Basilea II, que establece una ponderación del riesgo más minuciosa.

Basilea II se centra en 3 pilares fundamentales: 1) Requerimientos de capital mínimo; 2) Proceso supervisión bancaria y 3) Disciplina de mercado.

CUADRO 1 PILARES FUNDAMENTALES BASILEA

PILAR I	Requerimientos mínimos de capital	Son calculados tomando los activos ponderados de riesgo, con nuevos criterios que reflejen los cambios en el perfil de riesgo de las entidades.	
	Los riesgos que se consideran en el pilar I son tres:		
	Riesgo de Crédito	Método Estandarizado (EE)	Muy similar al Acuerdo de Basilea I, pero se introduce algunas categorías de riesgo y la posibilidad de evaluaciones de riesgo efectuadas por agencias o calificadoras externas
		Método Basado en Calificaciones Internas (IRB)	Propone dos cambios: 1) Básico (FIRB): los bancos estiman únicamente la probabilidad de incumplimiento (o default) para cada activo, mientras los indicadores y ecuaciones son provistos por el Comité de Basilea. 2) Avanzado (AIRB): los bancos estiman todos los indicadores cuantitativos que se requieren en las ecuaciones.

	Riesgo de Mercado	Se mantiene de Basilea I
	Riesgo Operacional	Se considera este riesgo que estaba implícito en los otros riesgos del Acuerdo vigente. Se permiten tres métodos: 1) Indicador básico, 2) Estándar y 3) Avanzado (AMA)
PILAR II	Supervisión Bancaria	Se le otorga un rol fundamental y los principios básicos son: 1) Los bancos deberán contar con un proceso para evaluar la suficiencia de capital total en función de su perfil de riesgo y con una estrategia de mantenimiento de sus niveles de capital. 2) Los supervisores deberán examinar las estrategias y evaluaciones internas de la suficiencia de capital de los bancos así como la capacidad de estos para vigilar y garantizar su cumplimiento y deberán intervenir cuando no queden satisfechos con el resultado. 3) Los supervisores deberán esperar que los bancos operen por encima de los coeficientes mínimos de capital y deberán tener la capacidad de exigirles que mantengan capital por encima del mínimo. 4) Los supervisores deberán intervenir con prontitud para evitar que el capital descienda por debajo de los mínimos y deberán exigir la inmediata adopción de medidas correctivas.
PILAR III	Disciplina de Mercado	Se establecen requerimientos de divulgación de la información con el objetivo de permitir a los participantes del mercado evaluar el perfil de riesgo del banco. Esto por cuanto los nuevos métodos de estimación de riesgo que se introducen dependen en mayor medida de las estimaciones de las propias entidades.

Fuente: Federación Latinoamericana de Bancos – FELABAN, 2003

Cuadro extraído boletín, Basilea II: Hacia un nuevo esquema de medición de riesgos

Elaboración: Autor

Como se observa en la tabla que antecede, la gestión del riesgo de crédito en Basilea II plantea una preocupación mayor en la mitigación del riesgo de crédito estableciéndose el método estandarizado (deudores sufren categorizaciones y se permite calificaciones externas) y las calificaciones internas para que los bancos puedan usar sus propios modelos para el cálculo del requerimiento de capital regulatorio para riesgo de crédito.

Según la normativa de Basilea II, para la aplicación de modelos interno IRB es recomendable que las entidades financieras crediticias dispongan de un modelo de Credit Scoring que les facilite medir la Probabilidad de Incumplimiento del crédito que se pretende conceder (Trejo, J., Ríos, H. & Martínez, M., 2016).

Propone una metodología o un enfoque aún más complejo para garantizar la solvencia y estabilidad del sistema financiero, haciendo énfasis en la necesidad de realizar una valoración, control y mitigación integral de los riesgos a lo que se somete como institución financiera, incluyéndose ya la medición del riesgo operativo que no consideraba Basilea I, el ratio de capital mínimo del 8% se mantiene pero se modifica el cálculo del riesgo de crédito, la nueva metodología es muy cercana a la que establece el modelo Vasicek:

$$\frac{\text{Capital mínimo regulador}}{NCR + RM + RO} \geq 8\%$$

Donde:

NCR: APRC (Activos ponderados por riesgo de crédito considerando las ponderaciones del riesgo en Basilea II) calculado bajo alguna de las nuevas metodologías

RO: Activos Ponderados por Riesgo Operacional

RM: Activos ponderados por riesgo de mercado

Capital Regulatorio: Definición de Basilea I

El Comité de Supervisión Bancaria de Basilea, denomina Basilea III, al conjunto de reformas sobre incremento de capital y liquidez aprobadas por el Comité. Los aspectos que más se destacan en Basilea III, es el requerimiento de capital donde se norma un capital³ mínimo requerido en los bancos, el cual podrá ser superior en función del riesgo, el entorno económico, demás factores internos y/o externos que puedan incidir.

Para complementar las reformas iniciales de Basilea III en el año 2010, el Comité emite en diciembre de 2017, la “Finalización de las reformas poscrisis”, cuyo propósito es restituir la credibilidad en el cálculo de los activos

³ Los bancos financian sus inversiones mediante capital y deuda, por ejemplo, depósitos y captaciones de clientes. El capital puede absorber las pérdidas, reduciendo la probabilidad de que un banco deje de ser viable. El capital regulador comprende: Capital ordinario de Nivel 1 – acciones ordinarias, beneficios no distribuidos y otras reservas; Capital adicional de Nivel 1 – instrumentos de capital sin una fecha fija de vencimiento; Capital de Nivel 2 – deuda subordinada y reservas generales para préstamos incobrables. Los bancos que disponen de más capital regulador están en mejores condiciones para financiar el crecimiento del crédito. BANCO DE PAGOS INTERNACIONALES, 2017. Comité de Supervisión Bancaria de Basilea. Finalización de Basilea III.

ponderados por riesgo (RWA)⁴ y a tal punto perfeccionar la comparabilidad de los coeficientes del capital bancario, para lo cual se adopta medidas tales como (BIS, 2017):

- Mejorar la solidez y sensibilidad al riesgo de los métodos estandarizados para el riesgo de crédito, el riesgo de ajuste de valoración del crédito (CVA) y el riesgo operacional;
- Se empieza a restringir la utilización de los modelos internos de Basilea II, dado que se detectó que en su uso se minimiza las ponderaciones por riesgo no permitiendo una evaluación de los riesgos que se confiable, por lo cual se introduce límites sobre algunos parámetros que son utilizados por los bancos para calcular los requerimientos de capital en calificaciones internas (IRB) para la gestión del riesgo de crédito, y elimina finalmente el uso de métodos basados en modelos internos para el riesgo CVA y el riesgo operacional.
- Introduce un coeficiente de apalancamiento para los bancos más grandes a nivel mundial, a fin de limitar el apalancamiento excesivo y poco necesario.
- Se incorpora un Output floor más fuerte, reprime la reducción de capital regulatorio; el cálculo de los RWA por el uso de modelos internos no podrá ser inferiores al 72.5% de los activos ponderados por riesgo que se obtiene de la utilización del método estándar proporcionado por los organismos supervisores.

Lo antes expuesto concluye en que Basilea I, II y III supone varios desafíos para la banca en el manejo integral de riesgos, y fundamentalmente en uno de los riesgos de mayor preocupación como es el riesgo de crédito. Basilea no pretende estandarizar una metodología de evaluación crediticia, sino por el contrario que estos dependan y sean de responsabilidad de cada institución, sin embargo; Basilea II si sugiere a las instituciones financieras implementar

⁴ RWA son un estimado del riesgo que determina el nivel mínimo de capital regulador que un banco debe mantener para poder hacer frente a pérdidas no esperadas. BANCO DE PAGOS INTERNACIONALES, 2017.

modelos de evaluación del riesgo crediticio considerando la mayor cantidad de variables implícitas a los clientes y como mínimas requeridas de liquidez, rendimientos, deuda, activos e IRB (Internal rating based approach), además de criterios de calificación, estructura y composición propios de la institución. Los esquemas de medición que se pueden extraer del Nuevo Acuerdo de Basilea se agrupan en metodologías tradicionales o estáticos y los internos o dinámicos IRB (Olano, 2018).

1.5.2. Economía Popular y Solidaria

El sistema financiero del país se integra principalmente por instituciones tanto públicas como privadas y el sector de la economía popular y solidaria donde se encuentra el sector cooperativo. Según Arrien (2018) las Cooperativas de Ahorro y Crédito manejan alrededor del 66% de las operaciones de microcrédito.

La historia de la Economía Popular y Solidaria y del cooperativismo particularmente en el Ecuador remonta desde décadas atrás, para ser más exactos de la época precolombina como impulso a los pueblos que necesitaban encontrar un estilo de vida mediante la cooperación, desarrollo y beneficio comunitario de sus integrantes. Aunque con el presente documento no se pretende profundizar en los orígenes y la historia de la Economía Popular y Solidaria y/o del Cooperativismo en el país, si se destaca que este sector ha tenido un gran impulso que va en incremento con el pasar de los años, desde que el estado (Gobierno) tuvo su primera intervención y dictó la primera Ley de Cooperativas en el año 1937, que tuvo como principal objetivo apoyar y organizar los movimientos campesinos de aquel tiempo, además de legitimar este sector.

El Artículo 283 de la Constitución de la República del Ecuador señala que el sistema económico país mantiene dos pilares fundamentales, el social y solidario, que tienen por objetivo garantizar la creación y reproducción de las condiciones que posibiliten el buen vivir de los ciudadanos. Este mismo articulado señala también que la organización del sistema económico podrá ser

de carácter público, privado, mixto, popular y solidaria, y otros que la constitución de la República del Ecuador determine, con ello es claro la importancia de la economía popular, y solidaria, teniendo como precepto la búsqueda del “Buen Vivir”; este sector contribuye mayormente al crecimiento económico, constituyéndose en un factor clave para la economía ecuatoriana. El artículo 309 de la antes señalada Constitución de la República del Ecuador ratifica en señalar que el sistema financiero nacional se integra además de los sectores público, privado, del popular y solidario. El Artículo 1 de la Ley de la Economía Popular y Solidaria define a la Economía Popular y Solidaria como una forma de organización económica, donde sus miembros organizan y desarrollan procesos para satisfacer necesidades y generar nuevos ingresos (SEPS, 2011).

Por su lado el artículo 311 de la Constitución menciona que a este sector lo conformarán las 1) Cooperativas de ahorro y crédito, 2) Las entidades asociativas y/o solidarias, 3) Cajas y bancos comunales, y las 4) cajas de ahorro. El sector procura satisfacer la necesidad de inclusión de los grupos sociales más vulnerables, que a menudo son privados de acceder a servicios financieros como un crédito. Como afirma De la Torre et al. (2012) en un estudio desarrollado por The World Bank, la falta de acceso puede manifestar las barreras monetarias para el uso de los servicios financieros. Después de todo el cooperativismo se constituye en la forma de organización más conocida y reconocida de la economía popular y solidaria, según Coraggio (2011) pone en manifiesto que esta economía en primer lugar es social porque produce sociedad además de utilidades económicas, no está orientada en la ganancia y la acumulación de capital sin límites. Es importante entender que las diversas formas de organización social comunitarias, en las que ha predominado la cooperación, parten de la necesidad del hombre por satisfacer sus necesidades, que hoy en día forman parte de un sistema económico (Torres & Fierro & Alonso 2017).

El Artículo 21 constante en la Ley de la Economía Popular y Solidaria y del Sector Financiero Popular y Solidario por su lado define al sector cooperativo como un conjunto de cooperativas de ahorro y crédito unidad de forma

voluntaria para satisfacer sus necesidades económicas, sociales y culturales en común, con personalidad jurídica de derecho privado e interés social. Las cooperativas, en su actividad y relaciones, además se sujetarán a los principios establecidos y contemplados en el Código Orgánico Monetario y Financiero, la Ley Orgánica de Economía Popular y Solidaria y del Sector Financiero Popular y Solidario, a los valores y principios universales del cooperativismo y a las prácticas de Buen Gobierno Corporativo.

En el Ecuador la Superintendencia encargada de la vigilancia y control de la economía popular y solidaria es la SEPS – Superintendencia de Economía Popular y Solidaria conjuntamente con la Junta de Política y Regulación Monetaria y Financiera que se creó con el Código Orgánico Monetario y Financiero.

El Código Orgánico Monetario y Financiero, en su artículo 447 que se refiere al Capital y segmentación establece que las Cooperativas se situaran en el segmento que así lo determine la Junta. La Junta libró la resolución No. 038-2015-F de 13 de febrero de 2015, “NORMA PARA LA SEGMENTACIÓN DE LAS ENTIDADES DEL SECTOR FINANCIERO POPULAR Y SOLIDARIO”, que en su artículo 1 clasifica a las entidades del sector financiero popular y solidario en cinco segmentos de acuerdo al monto de sus activos (Véase Cuadro No. 2):

CUADRO 2 SEGMENTACIÓN SECTOR FINANCIERO POPULAR Y SOLIDARIO

Segmento	Activos (USD)
1	Mayor a 80'000.000,00
2	Mayor a 20'000.000,00 hasta 80'000.000,00
3	Mayor a 5'000.000,00 hasta 20'000.000,00
4	Mayor a 1'000.000,00 hasta 5'000.000,00
5	Hasta 1'000.000,00
	Cajas de Ahorro, bancos comunales y cajas comunales

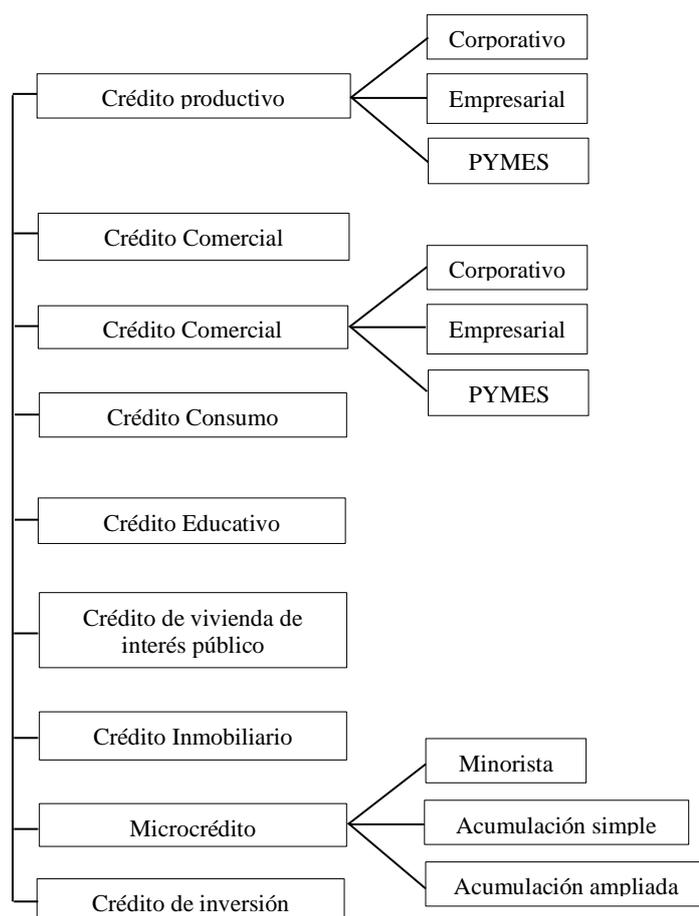
Fuente: Corporación del Seguro de Depósitos (Cosede, 2015)
Elaborado por: Autor

La cartera de crédito de una Cooperativa puede integrarse por distintos tipos de créditos, que pueden verse afectados por un riesgo diferente dependiente del producto, por lo tanto; la definición de portafolios que permita la agrupación o

segmentación de los créditos con similares características, resulta crucial, previo a la administración del portafolio.

Los tipos de crédito (Véase Cuadro No. 3) para las instituciones del sistema financiero de acuerdo con lo que establece la Resolución No. 043-2015-F de 05 de marzo de 2015 de la Junta de Política y Regulación Monetaria y Financiera homologado en la Codificación de Resoluciones Monetarias, Financieras, De Valores y Seguros, Libro I: Sistema Monetario y Financiero, se fija de la siguiente manera:

CUADRO 3 TIPOS DE CRÉDITO COOPERATIVAS



Fuente: JRMF, Codificación de Resoluciones Monetarias, Financieras, de Valores y Seguros, Libro I., 2019.

Elaborado por: Autor

De acuerdo con lo comprendido en el Libro I de la Codificación de Resoluciones Monetarias, Financieras, de Valores y Seguros, 2019, se tiene las siguientes definiciones de los tipos de crédito: **1) Crédito Productivo,**

aquellos que se conceden a personas naturales o jurídicas que se encuentran obligadas a llevar contabilidad a un plazo mayor a un año, propicio para el financiamiento de actividades relacionadas con la producción y comercialización de bienes y/o servicios en sus diferentes fases de fabricación. Los créditos productivos de acuerdo con lo fijado por la Junta de Política y Regulación Monetaria y Financiera se dividen en tres segmentos, *Productivo Corporativo; Productivo Empresarial y Productivo PYMES* (BCE, 2015). Está el **2) Crédito Comercial Ordinario**, para las personas naturales o jurídicas obligados a llevar contabilidad con ventas anuales que superen los USD \$ 100.000.00, cuyos recursos sean requeridos para fines productivos y/o comerciales. Otro tipo de crédito es el **3) Crédito Comercial Prioritario**, básicamente dirigido a personas naturales que estén en la obligación de llevar contabilidad o personas jurídicas con ventas anuales superiores a USD 100,000.00 cuyos recursos deberán ser destinados en la adquisición de bienes y/o servicios de actividades comerciales y productivas que no se contemplen como segmento comercial ordinario (BCE, 2015). **4) Crédito de Consumo Ordinario**, otorgado a personas naturales con garantía prendaria y/o fiduciaria básicamente destinada a la adquisición o comercialización de vehículos livianos de combustible fósil. Otro de los créditos más solicitados es el **5) Crédito de Consumo Prioritario**, otorgado a personas naturales y precisamente para la compra de bienes, servicios o gastos que no se encuentren relacionados o comprendidos dentro de una actividad productiva, comercial y otras compras y gastos no incluidos en el segmento de consumo ordinario. Se continúa con el **6) Crédito Educativo**, el tipo de crédito que fue creado y destinado al financiamiento de la formación y capacitación profesional o técnica. Está el **7) Crédito de Vivienda** un tipo de crédito de Interés Público, destinados a personas naturales para la adquisición o construcción vivienda única y de primer uso, siempre y cuando tengan el amparo en una garantía hipotecaria. **8) Crédito Inmobiliario**, concede con garantía hipotecaria a personas naturales para adquirir bienes inmuebles para la construcción, reparación, remodelación y mejora de inmuebles propios, adquisición de terrenos para construcción de vivienda propia y/o vivienda y que no pertenecen al crédito de vivienda. El **9) Microcrédito**, dirigido a personas naturales y/o jurídicas que mantengan un nivel de ventas o ingresos anuales inferior o igual a

USD 100,000.00, creado para el financiamiento de actividades que fortalezcan la producción y/o comercialización en pequeña escala de las personas naturales que trabajen en el sector comercial, este se subdivide en *Microcrédito Minorista*; *Microcrédito de Acumulación Simple*; y *Microcrédito de Acumulación Ampliada* (BCE, 2015). Finalmente está el **10**) Crédito de Inversión Pública, pensado para el financiamiento de programas, proyectos, obras y servicios enfocados a la provisión de servicios públicos gubernamentales.

1.5.3. Principales Modelos de Evaluación del Riesgo Crediticio

Las cooperativas de ahorro y crédito que requieran fortalecer su presencia en el mercado deben partir de un esquema de mejora continua de sus procesos de colocación, evaluación y aprobación de crédito (Jeres, 2018). La evaluación y gestión del riesgo crediticio o probabilidad de “default” de los clientes de una institución financiera se hace imprescindible para su sobrevivencia.

Sin duda la administración o gestión del riesgo crediticio durante las últimas décadas ha sido una de las áreas con más crecimiento (Trejo, J., Martínez, M., & Venegas, F., 2017). Para estimar este riesgo, se han diseñado múltiples modelos estadísticos, en base a un conjunto de atributos, patrones, que permiten calificar objetivamente la calidad de un cliente (Hand & Henley, 1997; Schreiner, 2002). Establecer las variables para una adecuada gestión del crédito es su principal propósito. (Winffel Basso, 2013) citado por (Bonilla, Crespo & León, 2018) indica que la necesidad de establecer métodos más eficientes de evaluación de riesgo crediticio, con el paso de los años, se han generado diversos modelos que pueden ser analizados desde el enfoque tradicional y moderno. Para la gestión del riesgo de crédito, se han implementado muchos modelos y algoritmos que buscan brindar apoyo en la calificación crediticia, incluyendo técnicas estadísticas, algoritmo genético y redes neuronales (Millán & Cerezo, 2018).

La estimación del riesgo para un crédito de cualquier tipo se realiza mediante el cálculo de la probabilidad de incumplimiento, sin embargo; identificar aquellos clientes que no generan un negocio para la entidad requiere adoptar mejores

lineamientos en su tratamiento (Trejo & Ríos & Martínez, 2016). Los modelos de evaluación y valoración del riesgo de crédito tienen como finalidad identificar los determinantes, factores o variables del riesgo de crédito en las carteras de cada institución (Saavedra, M. & Saavedra, M, 2010). En la determinación de modelos de evaluación del riesgo crediticio además de las variables que estos identifican, como hemos visto en Basilea es importante considerar criterios de calificación, estructura y composición propios de la institución; dado que, la medición del riesgo y su probabilidad de incumplimiento puede variar dependiendo de cada institución y la calidad de su cartera para la adopción de modelos internos.

En el cuadro No. 4. Que se muestra a continuación se da a conocer los principales modelos de evaluación, resultado de la investigación de (Saavedra, M. & Saavedra, M, 2010), que sigue la clasificación de (Galicia, 2003), entre modelos tradicionales y modelos modernos:

CUADRO 4 MODELOS DE EVALUACIÓN DE RIESGO DE CRÉDITO

Modelos tradicionales	Modelos modernos
	* Modelo KMV
	* Modelo de valuación de Merton
* Sistemas expertos (Galicia, 2003)	* Modelo Credimetrics de J.P. Morgan (1997b)
* Sistemas de calificación	* Modelo Credit Risk + (Morgan, 1997a)
	* Modelo de retorno sobre capital ajustado al riesgo (Falkenstein, 1997)
	* Modelo CyRCE

Fuente: Adoptado de Saavedra, M. & Saavedra, M, 2010 en base a (Galicia,2003)
Elaborado por: Autor

Empezando por los modelos tradicionales, los sistemas expertos y los sistemas de calificación scoring, rating, son los modelos de evaluación de riesgo crediticio que principalmente se caracterizan por su uso frecuente, sin embargo; mantiene algunas limitaciones, como el que considera apreciaciones de tipo subjetivo como el juicio, criterio o experiencia del analista u oficial de crédito en la fase de calificación y/o concesión de créditos. Para Berlanga

(2016) Un sistema experto necesita del conocimiento de especialistas en evaluación de riesgo de crédito y de un sistema que automáticamente pueda modificar y ampliar este conocimiento frente a la práctica, además de la dificultad para evaluar el desempeño de estos tipos de sistemas. Los modelos expertos utilizan la inteligencia artificial, técnicas estadísticas avanzadas para estimar la probabilidad de default (incumplimiento).

La solvencia y capacidad de una institución financiera frente al riesgo crediticio puede ser medida a través de modelos internos de las instituciones, estos modelos utilizan sistemas de calificación rating o scoring, efectuados por parte de compañías calificadores de riesgos y que además cuentan con la calificación de los organismos de control de cada país y se emplean únicamente para facilitar a los analistas la decisión de conceder o no un crédito, y siempre se acompañan de políticas de crédito propias de la institución. En cualquier caso, la construcción de un modelo de scoring requiere de información interna y externa del cliente, contenida en sus solicitudes del crédito (Rayo et al., 2010). Los modelos de credit scoring los cuales proponen automatizar el proceso de gestión de créditos en cuanto a conceder o no una determinada operación crediticia sujeto a un conjunto de variables relevantes de decisión (Ludovic & Aranguiz & Gallegos 2018).

Entre los modelos de enfoque moderno o contemporáneo, denominados así porque utilizan más variables para su desarrollo incluso son capaces de proporcionar estimaciones de Pérdidas No Esperadas. Según la clasificación de (Galicia, 2003) tomada por (Saavedra, M. & Saavedra, M, 2010), está el Modelo KMV, el modelo estructural de KMV Merton (Siglas de los socios fundadores de la empresa KMV - Kecholfer, McQuown, y Vasicek) se basa originalmente en el modelo de Robert Merton (1974), considera variables económicas – financieras presentes y pasados buscando modelar los activos y los pasivos de una empresa para identificar la probabilidad de incumplimiento (Martínez y Ballón, 2010). Los criterios para medición de KMV, se basan en el análisis de datos de las fluctuaciones del precio de mercado de las acciones emitidas por la compañía (Author, A., 2017). El modelo pretende inferir el valor de los activos de la empresa estimando la volatilidad del patrimonio y su

propio valor para obtener la probabilidad de incumplimiento default, tomando en cuenta el comportamiento crediticio de los deudores, toma doctrinas del modelo de Frecuencias de Incumplimiento Esperado (EDF, por su sigla en inglés), además de considerar la diversificación requerida en los portafolios de deuda (Saavedra, M. & Saavedra, M, 2010). El Modelo de Merton de pasivo contingente surge como una alternativa para medir la probabilidad de incumplimiento de las empresas. Supone que las empresas tienen dos formas de financiación a través de acciones y de deuda (Munafo, F., 2018). Según el principal supuesto del modelo, la empresa se ve en default cuando sus activos están por debajo de sus pasivos y es que si los pasivos superan críticamente los activos su patrimonio neto será nulo y la empresa de declarará en “*default*”. La estimación de riesgo default permite establecer una correlación de variables observadas y no observadas.

El modelo moderno de CrediMetrics de J.P. Morgan y Modelo Credit Risk, son los dos modelos internos más conocidos al pretender la aplicación de la teoría de portafolios en la gestión del riesgo crediticio, estos modelos intentan aplicar la diversificación como se lo hace en el riesgo de mercado, pero al ser dos escenarios distintos su adaptación no resulta ser tan sencilla. CreditMetrics 1997 desarrolla el modelo para evaluación de riesgo crediticio en base a las siguientes etapas: 1) Especifica un sistema de calificaciones y una matriz de transición utilizando la información de las agencias calificadoras de (Moody's o Standard&Poor's); 2) Establece un horizonte de tiempo que por lo general es de un año; 3) Desarrolla un modelo de valoración; 4) Analiza los cambios en el valor de la cartera de créditos; y 5) define el incumplimiento como el momento en el cual el valor de los activos se encuentra por debajo del valor nominal de los créditos⁵ (Caicedo, Mercè & Casanovas, 2011)., este modelo agrupa los créditos asignando una calificación a cada cliente, sin embargo; presenta una gran debilidad y es que asume que en base a sus datos históricos todas las instituciones financieras y todos los créditos presentan la misma probabilidad de incumplimiento con la misma calificación, enfocándose principalmente en los cambios que puedan surgir de la calificación real y la probabilidad en la

⁵ JP Morgan and Company (1997).

migración de la categoría del riesgo de la empresa (Author, A., 2017). En el caso de CreditRisk otro de los modelos modernos definido por su autor lo que hace es agrupar los créditos por la misma calificación que dependen de los mismos factores de riesgo y tienen la misma probabilidad de incumplimiento, asumiendo que la variable o factor de incumplimiento de un deudor es la misma para los otros deudores que forman parte del mismo grupo, su principal objetivo es calcular la distribución de las pérdidas de una cartera crediticia usando las matemáticas actuariales (Haaf, H. y Tasche, D., 2002).

El Modelo moderno o contemporáneo de retorno sobre capital ajustado al riesgo, es usado frecuentemente por calificadoras y bancos para evaluar la calidad y rentabilidad de una cartera diversificada midiendo la exposición al riesgo, se basa principalmente en los factores que determinan la distribución de valores futuros de la cartera del modelo CreditMetrics (Cardona, Franco, y Ossa, 2015). Otro de los modelos contemporáneos es el denominado CyRCE - Capitalización y Riesgo de Crédito, que es un modelo de Riesgo de Crédito para Mercados Emergentes resultado de una investigación realizada por el Banco de México, la información que se utiliza para este modelo es escasa y en ocasiones de mala calidad, un modelo de incumplimiento que se puede medir a través de su media y su varianza; el resultado de la segmentación de la cartera asume que las diferentes probabilidades de “default” o no pago y su correlación dentro de un mismo segmento son iguales para todos los créditos, hecho que no sucede realmente en las instituciones financieras. La metodología que usa CYRCE es similar a la de KMV. Es un modelo indicativo de probabilidad de incumplimientos que muchas veces dificulta predecir la quiebra de un banco; el modelo instaura la relación entre Riesgo de Crédito y Capital, además evalúa si el capital que asigna el banco o institución financiero a la cartera crediticia es suficiente para cubrir las posibles pérdidas (Rincon y Torres, 2015).

Debido a la necesidad de determinar modelos y perfiles de riesgos más exactos, de los modelos más conocidos y de mayor credibilidad incluso de mayor uso por las instituciones financieras del país están los modelos Credit Score, logit, probit, estos estiman la probabilidad de riesgo “*default*” o

probabilidad de incumplimiento, es decir que los deudores y/o solicitantes incumplan con sus obligaciones crediticias en un momento determinado de su obligación (buen o mal cliente), son modelos que usualmente son relacionados con el Data mining o minería de datos por pretender la búsqueda de patrones o comportamientos que se asocien a la probabilidad de default. También se definen como “métodos estadísticos utilizados para clasificar a los solicitantes de crédito, o incluso a quienes ya son clientes de la entidad evaluadora, que distinga los buenos deudores de los malos deudores (Millán & Cerezo, 2018).

García, M. & Sánchez, C. (2005) muestran una clasificación similar a la de Saavedra, M. & Saavedra, M, (2010) que toma la clasificación de Galicia (2003), enfocándose principalmente en los modelos que pueden ser usados en credit score analizado desde modelos tradicionales y modelos modernos: Entre los modelos tradicionales y más conocidos destaca el modelo de las Cinco C para la evaluación del riesgo de crédito, donde los factores para la medición del riesgo crediticio, son: Carácter, Capital, Capacidad, Colateral, y Condiciones; su aplicación puede llegar a ser compleja por las características de cada cliente, y porque desde décadas atrás el criterio y la intuición del analista de crédito puede prevalecer en este tipo de modelos, algo que ha sido complicado eliminar pese a los modelos de evaluación de riesgo crediticio existentes. Los modelos modernos contrario a los tradicionales se basan en modelos estadísticos, matemáticos, econométricos, y actualmente de inteligencia artificial, entre ellos tenemos: Z-Score, Creditmetrics y Var, modelos Z, también usan modelos de respuesta binaria. La investigación “Un Modelo de Credit Scoring para instituciones de microfinanzas en el marco de Basilea II”, 2010; muestra una clasificación de los modelos Credit Scoring en la banca de conformidad con la metodología estadística aplicada: a) Análisis Discriminante; b) Modelos de Probabilidad Lineal; c) Modelos Logit; d) Modelos de programación lineal; e) Redes neuronales; f) Árboles de decisión.

Los resultados de la investigación de Rodríguez-Guevara, D.E., Becerra-Arévalo, J.A. y Cardona-Valencia, D. (2017) señala que los modelos credit score más importantes hasta el año 2015 se clasifican en: modelos paramétricos, no paramétricos y los modelos semi paramétricos, donde se

busca dar respuesta a la operación matemática $f(Y,X)$, en este sentido, el riesgo como variable dependiente está en función de una serie de variables independientes cualitativas (género, estrato social, etc). y/o cuantitativas (ingresos, gastos, cargas familiares, etc) en el caso de personas naturales, y en el caso de una persona jurídica razones e indicadores financieras.

Gutiérrez, M. (2007) en su investigación denominada "Credit scoring models: what, how, when, and for what purpose", señala que entre las técnicas o metodologías más usadas están: análisis discriminante, regresión lineal, regresión logística, modelos probit, modelos logit, métodos no paramétricos de suavizado, métodos de programación matemática, modelos basados en cadenas de Markov, algoritmos como los árboles de decisión, sistemas expertos, algoritmos genéticos, redes neuronales, modelos probit, no excluyendo además, el juicio humano, como un método complementario de evaluación (Gutiérrez, M. 2007, p. 4). Logit, un modelo de probabilidad variable binaria que utiliza la técnica de regresión logística corresponde a un tipo de análisis de regresión donde la variable dependiente es una variable dummy dicho de otra forma: código 0 (Buen Cliente) o 1 (Mal Cliente) (Fernández Castaño, Horacio, Pérez Ramírez, Fredy Ocaris, 2005). La metodología Logit es un modelo estadístico que permite identificar las variables más importantes, con el análisis de las variables independientes con la dependiente (Valencia, 2017). Este modelo reduce las restricciones de un modelo de probabilidad lineal. El modelo logit dicotómico es un modelo de elección binaria, el cual tiene características específicas como son la elección de las variables explicativas, los supuestos relacionados con la forma funcional de las variables (Moreno, 2013). La predicción de la variable dependiente Default de probabilidad binaria (buenos y malos clientes) es un proceso que usualmente se utiliza para la construcción de un scoring, ocupa un papel muy importante en el sistema de administración del riesgo de crédito, debido a que permite anticipar el posible comportamiento del cliente desde el inicio de la relación financiera (Moreno, 2013). La construcción de modelos scoring y aplicación de modelos de regresión han ido en aumento en los últimos años dada su eficiencia predictiva, tal es el caso que el Banco Central del Ecuador en el año 2016 dado un ejercicio de alerta temprana en el sistema bancario ecuatoriano propone el uso

de análisis logístico o los modelos probit y logit para calcular la probabilidad que el sistema financiero entre en una crisis financiera, así como este existen múltiples análisis económicos del Banco Central del Ecuador donde a través de modelos Logit y Probit realiza análisis de determinantes.

Los modelos logit remedian la limitante de heterocedasticidad de los modelos lineales, por lo que se asume que el modelo es homocedástico y que los efectos marginales se pueden interpretar, para el caso de las variables categóricas de más de dos opciones de respuesta (BCE, 2018). (Bonilla, Crespo & León, 2018) tomando como sustento lo citado por (Carey 2001) y (Florez, 2007), hace hincapié que ante la necesidad minuciosos modelos de evaluación de riesgo crediticio como los antes recopilados, el entorno económico demanda modelos de medición y/o evaluación crediticia más flexibles y desarrollados, ante ello; y tomando las recomendaciones de Basilea II, la presencia y desarrollo de Modelos internos han sido apoyados en distintos países, y en el caso específico de Ecuador por la Superintendencia de Bancos – SB, la Superintendencia de Economía Popular y Solidaria – SEPS, y Banco Central del Ecuador – BCE.

En cuanto a la administración del riesgo crediticio, el sistema financiero ecuatoriano con la resolución JB-2011-1897 de 15 de marzo de 2011, sustituida con la JB-2011-2034 de 25 de marzo de 2011, proponía un modelo de las 5 C's con la aplicación de metodologías y/o modelos internos. La Superintendencia de Bancos dentro del Título IX De la Gestión y Administración de Riesgos, de las Normas de Control para las Entidades de los Sectores Financiero Público y Privado, vigente que incluye la resolución No JB-2003-602 de 9 de diciembre del 2003; establece algunos esquemas mínimos para las entidades controladas para la Administración del Riesgo de Crédito (riesgo de contraparte y pérdidas esperadas) que no son una camisa de fuerza, ni un modelo único, pues reconoce que cada entidad mantiene su propio perfil de riesgo y podrán adaptar sus metodologías internas.

El acuerdo internacional denominado “Nuevo Acuerdo de Capital” recomienda a las entidades financieras de los países miembros una revisión de sus

capitales mínimos para cubrir los riesgos, además de disponer de herramientas que les permitan y facilite establecer modelos de medición scoring y ratings con el propósito de discriminar entre clientes según su perfil de riesgo (Trejo et al., 2017), por lo que dichos modelos deberán adaptarse a las necesidades, objetivos de cada institución. La norma ecuatoriana de la SBS señala que la metodología implantada por la entidad deberá considerar criterios cuantitativos y cualitativos, que permitan controlar el riesgo de los diferentes portafolios de crédito, recomienda el análisis de estados financieros, flujos de efectivo, calidad de la gerencia, sistemas de apoyo para evaluación crediticia como “credit scoring”, sin perjuicio de los criterios establecidos en la “Calificación de activos de riesgo y constitución de provisiones por parte de las entidades de los sectores público y privado bajo el control de la Superintendencia de Bancos”. La Resolución No. 129-2015-F de 23 de septiembre de 2015 expide las “Normas para la Gestión del Riesgo de Crédito en las Cooperativas de Ahorro y Crédito”, que establece parámetros mínimos para la gestión de riesgo crediticio, la norma señala que para una adecuada gestión los segmentos 1 y 2 deberán establecer los perfiles de riesgo, con características cuantitativas y cualitativas de los socios tales como: edad, actividad económica, género, etc.; límites de exposición de riesgo para el segmento 1 y 2; tasas de interés según la segmentación de la cartera; mercado objetivo (zonas geográficas, sector socioeconómico) de los segmentos 1 y 2; la importancia de que todas las operaciones crediticias estén garantizadas; fija además la escala con los criterios de calificación de la cartera de crédito y contingentes según la morosidad y segmento de crédito que desde el año 2019 se contiene en el Libro I de la Codificación de Resoluciones Monetarias, Financieras, De Valores y Seguros.

En base a la normativa antes señalada se destaca que la normativa ecuatoriana no fija un modelo único como metodología de evaluación de riesgo crediticio, propone modelos internos de evaluación de gestión de riesgos basados en el enfoque regulatorio de Basilea (riesgo de crédito, mercado y operacional), dichos principios empezando con la Resolución JB-2001-382 de 08 de octubre de 2001 se han ido incorporando con el paso de los años, fijando para las instituciones financieras ecuatorianas variables mínimas para una

adecuada gestión del riesgo de crédito, pero que estas pueden ser adaptadas a la naturaleza y condiciones de cada institución, preceptos ahora regulados y homologados en el Libro I De la Codificación de Resoluciones Monetarias, Financieras, de Valores y Seguros.

1.5.4. Comparación de modelos aplicados en la evaluación de riesgo crediticio

En la evaluación de riesgo crediticio y en la viabilidad de aprobar o rechazar una solicitud de crédito a personas naturales o jurídicas para la predicción del riesgo "Default", como hemos visto en el repaso de teorías investigadas en el presente trabajo investigativo, existen una variedad de metodologías.

En el Cuadro No. 5 que se muestra a continuación se presente un resumen de los diferentes modelos y/o metodologías investigadas para la evaluación del riesgo crediticio aplicados al sector financiero, con las variables cuantitativas o cualitativas consideradas por cada autor para su análisis.

CUADRO 5 VARIABLES RELEVANTES DE LOS PRINCIPALES MODELOS APLICADOS EN LA EVALUACIÓN DE RIESGO CREDITICIO EN EL SECTOR FINANCIERO

Año	Autor	Modelo	Variables empleadas
2007	Gutierrez Girault Matias Alfredo	Credit Scoring	<p>* Información constante en solicitudes de crédito y/o fuentes internas y/o externas de la entidad</p> <p>* Pymes: Activos/IPC, pasivos/activos, crecimiento de los ingresos netos, inventarios/costo de mercaderías vendidas, indicador de prueba ácida, ganancias retenidas/activos, efectivo/activos, crecimiento y/o incremento de las ventas, indicador ratio de la cobertura de la deuda, ganancias obtenidas, apalancamiento, tamaño de la empresa y liquidez de la misma. (Falkenstein, 2000)</p> <p>* Personas naturales: edad, estado civil, cantidad de personas a cargo, empleo actual, tiempo que ha permanecido en el domicilio declarado, género, nivel educativo, promedio de gastos e ingresos mensuales, arriendo o propietario de vivienda, ocupación, cuentas, calificación buró crediticio. Boyes, Hoffman y Low (1989) y Greene (1992)</p>

			<ul style="list-style-type: none"> * Comportamiento actual y pasado del cliente: Historial de pagos o scored, capital adeudado, historial crediticio, nuevo crédito, tipo de crédito. (FICO credit risk score)
2010	Rayo, S., Lara, J., Camino, D. Ochoa, J.C., Galeano, W., & Agudelo, L.G.		<ul style="list-style-type: none"> Toma variables a partir del proceso de la promoción, evaluación y concesión de la operación crediticia: * Variables del cliente * Variables de la operación de crédito (solicitud de crédito) * Variables del ciclo económico
2005	Fernández Castaño, Horacio, Pérez Ramírez, Fredy Ocaris	Logit (Los modelos de regresión y de probabilidad predictiva son muy usados actualmente por el BCE para análisis económicos)	Considera variables cuantitativas y cualitativas de la operación de un sujeto de crédito, la mayor cantidad de variables según la correlación existente entre ellas, para la determinación y estimación del riesgo "Default".
2010	Rayo, S., Lara, J., Camino, D.		Considera el comportamiento de una serie de variables independientes de cada observación. Al medir la probabilidad de incumplimiento mantiene la variable dependiente dentro de un rango entre cero y uno.
2013	Moreno Valencia Sandra		Variable de respuesta "Default" y atributos independientes, como características demográficas y financieras del cliente.
2010	María Luisa Saavedra García & Máximo Jorge Saavedra García	Sistemas Expertos	* Capacidad
			* Capital
			* Colateral
			* Carácter
		Sistemas de Calificación	* Condiciones
			* Evaluaciones de riesgo conceptualmente sólidos
			* Técnicas de mitigación de riesgo
		KMV	* Sistemas de Gestión
* Relación existente entre el valor del mercado de capital y el valor de mercado de sus activos			
CyRCE	* Relación existente entre la volatilidad de los activos del capital de la compañía		
	* Relación existente entre el riesgo de crédito y el capital requerido para afrontar riesgos		
	* Relación entre el riesgo de crédito y los límites individuales de cada segmento de la cartera crediticia.		
			* Probabilidad de impago, incumplimiento
2015	Superintendencia de Economía Popular y Solidaria (SEPS)	Modelos internos	* Límites de exposición al riesgo de crédito
			* Criterios tales como: Monto, plazo, garantías, tipo de producto, destino del financiamiento, etc.
			* Criterios tales como: Zonas geográficas, sectores socioeconómicos, etc.
			* Perfil del riesgo (características del socio) tales como: edad, actividad económica, género, etc. (Resolución No. 129-2015-F)
2018	Superintendencia de Bancos Ecuador		Variables a partir de las fases de evaluación; aprobación; instrumentación y desembolso; seguimiento; recuperación y control:

			* Criterios o características cuantitativas y cualitativas básicas del sujeto de crédito
			* Criterios para aceptación de garantías
			* Constitución de provisiones
			* Criterios de calificación, recuperación, castigos y/o reestructuraciones
			* Burós, niveles de créditos internos y/o externos (Credit Scoring)
			* Análisis de estados financieros
			* Flujos de caja del proyecto
			* Comportamiento histórico
			* Calidad de la gerencia (Codificación de las Normas de la Superintendencia de Bancos, Libro I)

Fuente: Datos recopilados de la investigación

Elaborado por: Autor

1.5.5. Business Intelligence y Minería de Datos

La inteligencia de negocios nace de la necesidad de gestionar el conocimiento empresarial a partir de datos estructurados y no estructurados (Tello & Perusquia, 2015), también se considera como el conjunto de estrategias empresariales que mediante la creación del conocimiento ayuda y facilita la toma de decisiones, utilizando el análisis de los datos; las organizaciones hoy en día han desarrollado múltiples aplicaciones tecnológicas que facilitan este proceso y es que los datos se han constituido en parte de los activos no tangibles más importantes de las organizaciones que tiene interés en multiplicar sus competencias y oportunidades en los negocios.

Según el Data Warehouse Institute, la inteligencia de negocios o BI por sus siglas en inglés Business Intelligence (BI) es un término que abarca los procesos, las herramientas, y tecnologías que convierte a los datos en información, y esta información en conocimiento valioso para conducir y guiar de forma eficaz las actividades de los negocios. BI abarca las tecnologías de datawarehousing entre ellos procesos en el 'back end¹⁰', informes, consultas, análisis y demás herramientas que sirven para mostrar información y los procesos en el 'front end'." (EBI, 2005). Son componentes el Business Intelligence el Data Warehouse, Data mart y Data Mining, es una herramienta que facilita la toma de decisiones (Pilamunga, 2019).

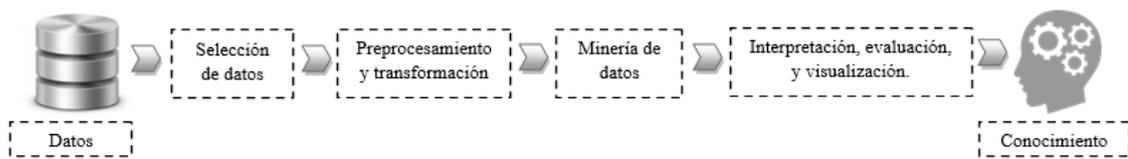
Según Laudon (2012), las principales herramientas para la inteligencia de negocios incluyen software de consultas e informes de bases de datos, herramientas para el análisis de datos multidimensional y herramientas para la minería de datos. Se denomina BI (Business Intelligence o Inteligencia de Negocios) al conjunto de estrategias, acciones y herramientas enfocada la administración y creación de conocimiento mediante el análisis de datos existentes dentro de una organización o empresa (Apale & Olivares & Rodríguez & Alor & Muñoz, 2016).

Para nadie es desconocido que cada día en el mundo se generan miles de datos que son almacenados en grandes bases de datos, pero consigo también la dificultad de analizarlos, si no se sabe cómo usarlos pueden convertirse en un verdadero dolor de cabeza para las empresas. Debido al crecimiento de datos estructurados como no estructurados, se ha despertado un gran interés en la gestión del conocimiento dentro de las instituciones, respecto a su procesamiento, uso, análisis e interpretación de datos, como un mecanismo que facilite la toma de decisiones. De acuerdo a lo antes expuesto, el proceso de KDD (Knowledge Discovery in Databases) se refiere al proceso de identificación o extracción de conocimiento de las bases de datos empresariales, comprende una serie de pasos que son necesarios para llegar al conocimiento esperado que sirva para la toma de decisiones.

Según Fayyad et al. (2002), el KDD es el proceso no insignificante de identificar patrones, comportamientos válidos, novedosos, útiles y comprensibles a partir de los datos que por sí solos son irrelevantes. (KDD) es el encargado de la preparación de los datos y la interpretación de los resultados obtenidos (Giraldo & Vargas, 2017). El paso indispensable del proceso de KDD está en la preparación de los datos, tarea que no es tan sencilla. Una vez que se ha obtenido y preparados los datos adecuados se procede a la minería de datos, en este paso se seleccionarán las herramientas y técnicas adecuadas para lograr los objetivos que se pretenden alcanzar (García & Molina, 2006).

A continuación, se muestra de forma resumida y para una mejor comprensión la metodología KDD:

FIGURA 1 METODOLOGÍA RESUMIDA KDD (Knowledge Discovery from Databases)



Fuente: Adaptado de (Apale & Olivares & Rodríguez & Alor & Muñoz, 2016)

Elaborado por: Autor

El proceso conlleva varias fases o etapas: El proceso de KDD consta de una secuencia iterativa de etapas: integración y recopilación de datos; selección, limpieza y transformación de datos; minería de datos; evaluación; difusión, uso y monitorización de modelos (De Battista et al, 2016).

La **recopilación de datos** posiblemente sea la base que más tiempo requiere en el proceso KDD, aquí se determina la fuente de información que será el insumo para recoger los datos y extraer el conocimiento requerido, estas pueden ser: bases de datos, fuentes internas o externas con datos estructurados o no estructurados. En esta etapa se establecen las fuentes de datos y el tipo de información a utilizar, los datos son extraídos para el análisis desde las fuentes de datos (Torres & Farroñay, 2015). Para (Hernández et al. 2004) la recopilación de datos consiste en identificar las fuentes de información útiles en el estudio, descubrir dónde encontrarlas y la mejor manera de acceder a ellas; la integración de datos busca transformar los datos de diferentes fuentes en un formato común (Bernstein & Haas, 2008) previo a su análisis; este último usualmente mediante una bodega de datos conocida como OLAP, OLTP, Datawarehouse, etc. con el fin de unificar toda la información recogida.

La **Selección, limpieza y transformación de los datos** es la más importante pues es aquí donde se tratan, corrigen los datos incorrectos seleccionando aquellas variables o atributos relevantes que facilite la tarea de minería. (Betancur, sf.). Esta etapa trata de eliminar inconsistencias existentes en los datos, discretizar atributos poco necesarios e incluir los datos faltantes y dejar las variables listas y preparadas que serán utilizadas para la construcción del modelo previamente investigado, en esta fase también será necesario

identificar la distribución y simetría de los datos, a través de la correlación de variables con reglas de asociación, clustering, técnicas estadísticas, etc.

La **Minería de Datos**, en este punto se da la selección de las tareas y/o técnicas de minería (regresión, clasificación, agrupamiento, árboles de decisión, etc) que serán utilizadas para la extracción del conocimiento requerido, la herramienta que se selecciones dependerá de los patrones que se desea obtener o el tipo de conocimiento que se desea descubrir.

En la fase de la **evaluación y validación** se seleccionarán el modelo de los resultados que se obtengan de la búsqueda de comportamientos y/o patrones con aplicación del paso anterior.

En la **Difusión, interpretación y utilización** se reconoce el modelo con la obtención del conocimiento que será usado para la toma de decisiones, en esta fase se verificará la utilidad o se rechazará el modelo.

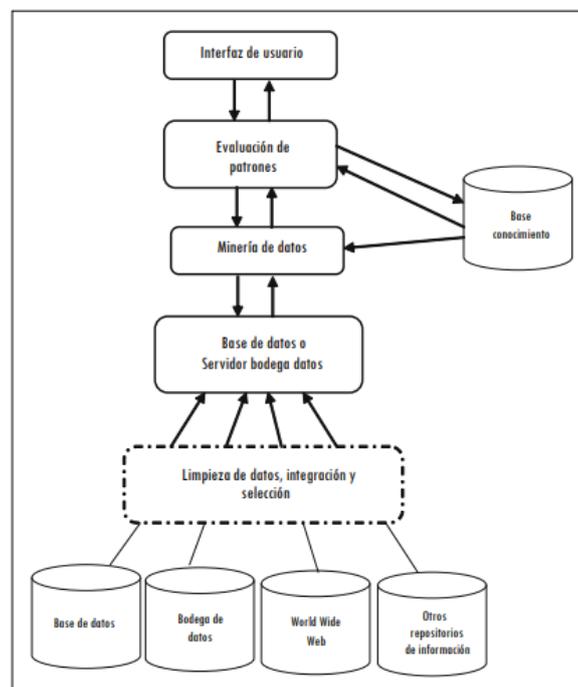
La minería de datos cuando bien aplicada produce magia al descubrir información de un conglomerado de datos que a simple vista pareciera ser inmanejable, constituye la tercera etapa de análisis del gran proceso KDD. La identificación de patrones, comportamientos en las bases de datos pueden ser los factores que estas requieren para mejorar procesos de riesgo en las empresas a través del conjunto de herramientas que facilitan la obtención de conocimiento. El objetivo final es generar modelos que permitan comprender y predecir el comportamiento de los procesos a partir de los registros almacenados en las BD (estructuradas y no estructuradas), a los cuales se aplican métodos matemáticos y computacionales expertos en el descubrimiento de patrones inicialmente ocultos (Kotu & Deshpande, 2015). La minería de datos nace de la necesidad de conocer información útil a partir de los bases de datos o Datawarehouse (Hernández & Duque & Moreno, 2017), con el crecimiento de los datos disponibles. La minería de datos es capaz de navegar a través de los datos y generar información precisa y acertada a partir del apoyo de tres tecnologías: Recolección de datos, multiprocesador y algoritmos de minería de datos (Bramer, 2007). La Minería de Datos es posible

aplicarse en cualquier tipo de repositorio o almacén de datos tales como BD Relacionales, Data Warehouses, BD Transaccionales, en la Web y en Sistemas de BD avanzados como BD Objeto-relacionales, BDEspaciales, BD Temporales, BD de Textos, BD Multimedia (De Battista et al, 2016).

La arquitectura de la minería de datos según (Roddick & Lees, 2001) comprende:

- Bases de conocimiento. - Que constituye la gestión y obtención de los datos en almacenes de datos donde sea posible su recuperación con el ánimo de evaluarlos y establecer relaciones o patrones que generen conocimiento que apoye la toma de decisiones.
- Algoritmos o técnicas de minería de datos. - Son aquellos que permiten el análisis de los datos a través de comportamientos, patrones, relaciones, etc.
- Evaluación de patrones. - En esta etapa se evalúa e interpreta los patrones obtenidos.

FIGURA 2 ARQUITECTURA DE MINERÍA DE DATOS



Fuente: (Han y Kamber, 2006, p. 9).

El proceso de minería de datos incluye una variedad de técnicas, algoritmos que aplicados a un conjunto de datos obtenidos de un almacén de datos dan como resultado modelos contruidos desde patrones obtenidos en el procesamiento de los datos, generando nuevo conocimiento, dicho de otra forma la aplicación de algoritmos de minería de datos obtiene patrones de los datos y, en ese sentido; crea modelos que apoyen la toma de decisiones y que contribuyan a mejorar los índices de competitividad o de un problema en particular (Roddick & Lees, 2001). Las tecnologías y algoritmos sofisticados y novedosos son indispensables para procesar eficientemente lo que se conoce como Big Data (García & Ramírez & Luego & Herrera, 2016).

La minería de datos comprende todo un conjunto de técnicas con el ánimo de crear mecanismos adecuados de orientación, entre ellas está la estadística, el reconocimiento de patrones, la clasificación y la predicción (Rodríguez, Y. & Díaz, A. 2009). Adopta técnicas estadísticas, de inteligencia artificial, bases de datos, entre otros, para la obtención de modelos, patrones o comportamientos en bases de datos (Siebes, 2000). Existen diversas clasificaciones de técnicas de minería de datos. La clasificación de las técnicas y/o algoritmos de minería de datos dependerá de su autor. (Weiss & Indurkha, 1998) señala que los algoritmos de minería de datos se clasifican en dos grandes categorías: supervisados o predictivos y no supervisados o de descubrimiento del conocimiento.

Los algoritmos supervisados o predictivos son utilizados para predecir el valor de un atributo (etiqueta) de un conjunto de datos, de otros atributos conocidos (descriptivos) para inducir la relación entre el atributo (etiqueta desconocido) y los otros atributos conocidos. El aprendizaje supervisado se lleva a cabo en dos fases, la primera llamada entrenamiento donde se construye el modelo usando un subconjunto de datos co etiqueta conocida (atributo sobre el cual se predecirá un valor) y una segunda fase de prueba del modelo sobre el resto de los datos. (Moreno et al. 2004). Si la etiqueta es discreta el modelo será de clasificación mientras que si es continua tendremos un modelo de regresión. El modelo obtenido se utilizará posteriormente para realizar predicciones con datos no etiquetados (Moreno & López, s.f.).

Algunas de las técnicas de minería de datos de tipo supervisadas son: árboles de decisión, inducción neuronal, regresión, etc. Los métodos no supervisados o descubrimiento del conocimiento son utilizados para descubrir patrones o comportamiento de grandes bases de datos, utilizan datos actuales no utiliza datos históricos (Moreno et al. 2004). Entre los más conocidos están: Agrupamiento (clustering), reglas de asociación, patrones secuenciales, detección de desviaciones, etc.

(Pérez & Santín, 2008), distingue o clasifica a las técnicas de minería de datos entre predictivas, descriptivas y auxiliares, donde las dos primeras son empleadas para el descubrimiento y las últimas se usan para verificación.

Técnicas predictivas.- Las técnicas o métodos predictivos o lo que para (Weiss & Indurkha, 1998) se conoce como métodos de aprendizaje supervisado, se basan en entrenar un conjunto de datos o un modelo, utilizando algún algoritmo, con el objeto de predecir una variable de interés o estimar valores de salida. Supone la existencia de variables dependientes e independientes para su análisis. Un modelo para que se aceptado como válido debe cumplir con las fases de identificación objetiva (identificar el mejor modelo que se ajuste a los datos), estimación (cálculo de los parámetros del modelo seleccionado), diagnosis (contrastar la validez del modelo estimado) y predicción (utilización del modelo identificado, estimado y validado para predecir un valor o valores futuros sobre la base de las variables dependientes (Pérez & Santín, 2008).

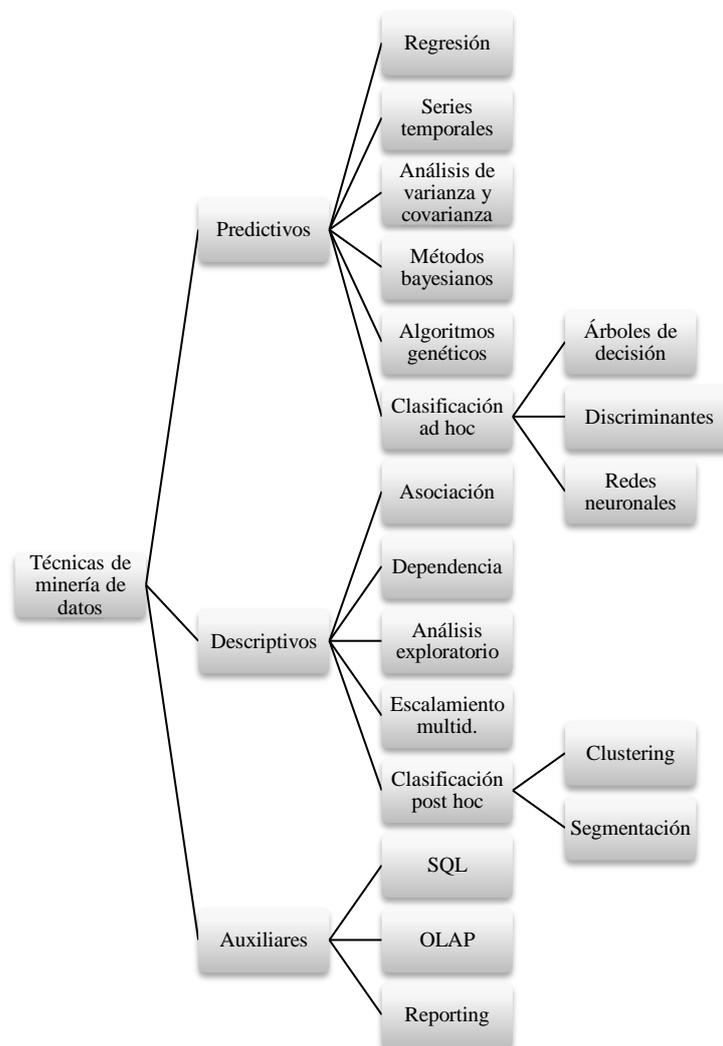
Técnicas descriptivas.- Los métodos descriptivos o de aprendizaje no supervisado, no supone la existencia de variables dependientes o independientes y tampoco supone la existencia de un modelo previo a los datos (Pérez & Santín, 2008). Su objetivo es buscar patrones para describir datos futuros, por ejemplo un restaurante busca identificar sus clientes en base a sus preferencias a fin de mejorar y personalizar su servicio.

Tanto las técnicas descriptivas como las técnicas predictivas buscan el descubrimiento del conocimiento de un conjunto de datos (Novoa, 2016).

Técnicas auxiliares.- Las técnicas auxiliares son técnicas de apoyo usadas o enfocadas en la verificación de los modelos.

En la figura No. 3 se muestra la clasificación de las técnicas de minería de datos según (Pérez & Santín, 2008):

FIGURA 3 CLASIFICACIÓN DE LAS PRINCIPALES TÉCNICAS DE MINERÍA DE DATOS



Fuente: Adaptado de Pérez & Santín, 2008
Elaborado por: Autor

A continuación, se presenta una breve descripción de las principales técnicas de minería de datos:

Regresión: El método de regresión es un método estadístico – predictivo, que es usado para establecer la relación entre una variable dependiente y un número finito de variables independientes x_1, x_2, \dots, x_k ; este modelo de regresión tiene un solo nodo primario o explicativo sobre otras variables independientes y es insuficiente en datos multidimensionales donde se relacionan más de dos variables.

El procedimiento para estimar la fórmula de la regresión lineal es el procedimiento de mínimos cuadrados (García & Molina, 2006). Con la recta de la regresión lineal que se instaura de la fórmula $y = a + bx$, donde y es la variable dependiente, x la variable independiente y en el caso específico de a y b asumen el valor de las constantes o valores de los parámetros, es así que se determina la relación existencias entre las dos variables que están en análisis.

La regresión logística es otra de las técnicas de análisis predictivo en minería de datos de tipo no lineal, esta es una técnica estadística que permite la predicción de los valores de una variable dicotómica o categórica con la aplicación de logit o probit en función de otra serie de variables explicativas igualmente continua o categórica (Villarino, 2015). La regresión logística es una de las herramientas estadísticas con mejor capacidad para el análisis de datos en investigación de diferente contexto, de ahí su amplia utilización (Giraldo & Vargas 2017).

Series temporales: Este tipo de algoritmos son utilizados para observar el cambio de una variable en estudio a través del tiempo. Se estudian principalmente con el objetivo de extraer información de algún fenómeno del pasado e intentar predecir el futuro (Valero, C & González, M., sf.). Su aplicación es extensa en muchas áreas del conocimiento, donde se requiere predecir el comportamiento de variables en un determinado momento (González & Soto, 2013).

Análisis de varianza y covarianza: También conocido como ANOVA (Varianza) y ANCOVA (Covarianza), función es comparar las medias de un grupo de datos y establecer igualdades o diferencias en su comportamiento.

Para la aplicación de ANOVA se debe cumplir algunas condiciones como que el conjunto de datos en análisis debe mantener una distribución normal, ser separado de otro conjunto, y las varianzas obtenidas de cada conjunto de datos deben mantener similitud, pues su objetivo principal es determinar diferencias significativas (Boqué & Maroto, sf.).

Métodos bayesianos: Entre los diferentes algoritmos de minería de datos están las redes bayesianas, este análisis se trata de un procedimiento basado en la estadística y la probabilidad con el ánimo de predecir el valor de variables no observadas y definir y/o explicar las observadas. Una red bayesiana permite inferir relaciones de dependencia en un conjunto de variables.

Las metodologías bayesianas obtienen de una distribución de datos la incertidumbre en las estimaciones para inferir sobre variables desconocidas (Pereira da Silva, 2016).

Algoritmos genéticos: Son de las principales técnicas de minería de datos con aplicación de inteligencia artificial, utilizan información histórica para el análisis de la evolución de poblaciones simulando en una máquina el proceso genético natural del ser humano, para la construcción de algoritmos (Sosa, 2007). Son técnicas de minería de datos que tiene su origen en la genética o evolución natural y en la teoría Darwinista que considera como principio que no todos los individuos de una población son iguales pero que cada uno de ellos representa una posibilidad, este tipo de técnicas busca descubrir individuos aptos e idóneos dentro de una población en estudio y los no aptos en base a la selección de un criterio.

Clasificación ad hoc: Estas son de las técnicas de minería más usadas en la evaluación de riesgo crediticio, incluyen técnicas como árboles de decisión, análisis discriminante, y redes neuronales, este tipo de técnicas predictivas tienen como propósito la construcción de un modelo a través de la clasificación de datos y determinación de comportamientos.

Árboles de decisión: Esta técnica es considerada como una de las más relevantes e integrales en el manejo de datos y una de las más utilizadas por el conjunto de valores que puede tomar una variable en análisis, además; de ser una de las técnicas de clasificación supervisada más eficaz. (Bouza, C. & Santiago, A., 2012) señala que los árboles de decisión permiten la segmentación de los grupos más importantes; la clasificación al permitir la asignación de ítems a uno de los grupos; predicción al establecer reglas para ciertos comportamientos; reducción y simplificación de datos a fin de obtener los más importantes para la creación de un modelo; permitir la identificación – interrelación al definir las variables más importantes y cuál es su relación existente; finalmente su recodificación para discretizar variables o determinar criterios cualitativos disminuyendo la posibilidad de perder datos relevantes.

Visualmente el árbol de decisión comienza por un nodo inicial que contiene toda la información, luego se ramifica en nodos de decisión tantos como resultados posibles, los que a su vez crean nodos adicionales como el nodo de probabilidad que indica en que punto del proceso ocurre un evento aleatorio, el nodo terminal que no requiere una división adicional, y las ramas que muestran los caminos para la toma de una decisión o algún evento aleatorio, y muestran además de las relaciones entre las alternativas de decisión y los eventos (Berlanga et al. 2013). Los árboles de decisión pueden dividirse en CHAID (Chi-square automatic interaction detector), CHAID exhaustivo, Árboles de clasificación y regresión (CRT-Classification and regression trees) y QUEST (Quick, unbiased, efficient, statistical tree); hay muchos algoritmos específicos de árboles de decisiones entre los cuales se puede destacar ID3 desarrollado por J. Ross Quinlan; C4.5, C5 y J48 para casos de evaluación de riesgo crediticio que realiza su predicción en función de la variable objetivo “default” o incumplimiento, cualquiera de los árboles de decisión expuestos obtendrá un resultado de n nodos terminales y sus probabilidades con un n número de valores posibles.

Los árboles de decisión son una herramienta bastante eficaz y efectiva para la predicción de probabilidad de incumplimiento, cuyos resultados predictivos permiten la planeación de estrategias comerciales de venta de servicios,

estrategias de cobranza, entre otras (Cardona, P., 2004). Los árboles de decisión debido a su habilidad de generar modelos predictivos son de las técnicas y/o herramientas estadísticas y de minería de datos más usadas en distintas áreas. Este método se encuentra clasificado dentro de las técnicas no paramétricas de cálculo de scoring de crédito, su propósito es predecir una variable objetivo a través de las independientes (Millán & Cerezo, 2018).

Redes neuronales: Esta técnica de inteligencia artificial se emplea para predecir el comportamiento de datos, emulando el sistema nervioso siendo capaz de aprender con la información que le es proporcionada.

Son un conjunto de algoritmos matemáticos que encuentran las relaciones no lineales entre conjuntos de datos; suelen ser utilizadas como herramientas para la predicción de tendencias y como clasificadoras de conjuntos de datos (Pérez, F. & Fernández H. 2007).

Las redes neuronales pueden ser diseñadas para varios entornos ayudando a la gestión y toma de decisiones, es capaz de descubrir patrones y descubrir comportamiento de las variables ingresadas (Peña & Orellana, 2018).

Análisis exploratorio: Este tipo de técnicas estadísticas son bastante útiles dentro de la minería de datos, su propósito principal es determinar las relaciones entre las variables cuando no hay o estas no están totalmente específicas en la naturaleza de estas relaciones, se basan en métodos estadísticos computacionales o no y pueden llegar hasta una exploración multivariante para la identificación de patrones. Entre las técnicas estadísticas se encuentran distribuciones de variables, correlaciones con matrices de contingencias, análisis de factores, etc. (Rodríguez Suárez, Yuniét, Díaz Amador, 2009).

El análisis exploratorio revela información sobre el comportamiento de los atributos, el análisis exploratorio no constituye un análisis lo suficiente fuerte como para soportar de manera definitiva la toma de decisiones, sin embargo, es relevante en cualquier estudio de datos, dado que es este análisis quien

entrega el primer acercamiento al comportamiento de los atributos y variables objetivos (Sánchez et al., 2018).

Reglas de Asociación: Las técnicas de asociación, técnicas descriptivas que pretenden la búsqueda de patrones en un conjunto de datos de gran tamaño, estableciendo relaciones o combinaciones que deben mantener una probabilidad bastante confiable para la toma de decisiones.

Una regla de asociación puede relacionar dos o más variables de interés y se compone de dos partes atributo y valor; cuyo propósito es encontrar los comportamientos que más se repiten en una base de datos (Malberti & Beguerí, 2015).

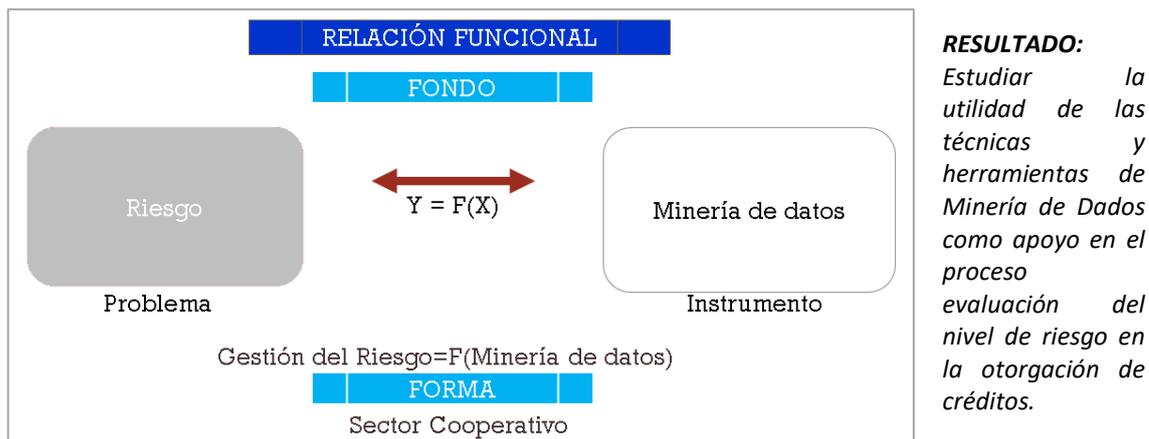
Clustering “Segmentación”: Son conocidas también como de agrupamiento y son técnicas estadísticas que permiten ver la similitud existente de elementos en un conjunto de datos (García & Molina, 2006), donde miembros de un grupo esté lo más cercano posible a otro y los grupos excluyentes o diferentes se encuentren alejados uno del otro, son usados para la exploración de datos e identificación de anomalías a la vez que se crea predicciones (Gallego & Navarro & Castillo, 2015).

1.5.6. Selección del modelo de evaluación

1.5.6.1. Definición de la problemática de riesgo crediticio a partir de la utilización de un instrumento

Con el ánimo de determinar comportamientos que incidan en la evaluación de riesgo crediticio en una Cooperativa de Ahorro y Crédito, se hace necesaria la aplicación de un instrumento de estudio para la realización de la presente investigación.

FIGURA 4 RELACIÓN FUNCIONAL DE LA VARIABLE DEPENDIENTE Y VARIABLE INDEPENDIENTE



*Fuente: Problemática de trabajo de investigación
Elaborado por: Autor*

El instrumento de estudio se sustenta en teorías, modelos, metodologías, investigaciones similares aplicadas y desarrolladas a partir del objetivo general, específico y formulación del problema. Dada la necesidad, previo a la aplicación del instrumento fue necesario la investigación de metodologías relacionadas a la variable dependiente “Default” en la concesión y evaluación de una operación crediticia, que expliquen dicho fenómeno y que además contribuya a generar un modelo para medir el riesgo crediticio a través de la aplicación de técnicas y herramientas de minería de datos, que puedan ser implementadas en una institución financiera como apoyo en la administración del riesgo crediticio.

1.5.6.2. Definición de las variables dependiente e independientes basado en un modelo de atributos para la evaluación de riesgo crediticio

Para la determinación de las variables independientes o explicativas de la probabilidad de incumplimiento de un cliente “Default”, para el efecto se tomó en consideración estudios realizados como el de Gutiérrez, M. (2007) que contiene análisis de varios autores para la construcción de Modelos Credit Scoring en la evaluación del riesgo crediticio y para estimar la probabilidad de “default” en individuos y PYMES, y para el caso específico de la presente

investigación que se realiza un análisis de personas naturales, el autor obtiene variables socioeconómicas como edad, empleo, nivel de estudios, estado civil, promedio de gastos e ingresos, etc, para análisis, así mismo para este tipo de modelos Saavedra, M. & Saavedra M. (2010) sugiere el análisis de las cinco “C”: capacidad, capital, colateral, carácter, y condiciones, que fueron analizadas en el caso de estudio de la presente investigación para la determinación de variables.

Se toma en consideración además las recomendaciones de modelos internos constantes en la Codificación de Resoluciones Monetarias, Financiera, De Valores y Seguros vigente que toma como preceptos a Basilea en la evaluación del riesgo crediticio, donde se recomienda variables asociadas al perfil del socio tales como: edad, género, actividad económica, etc. Siendo así que se consideró y analizó las variables constantes en la solicitud de crédito de la institución tomada como caso para este estudio.

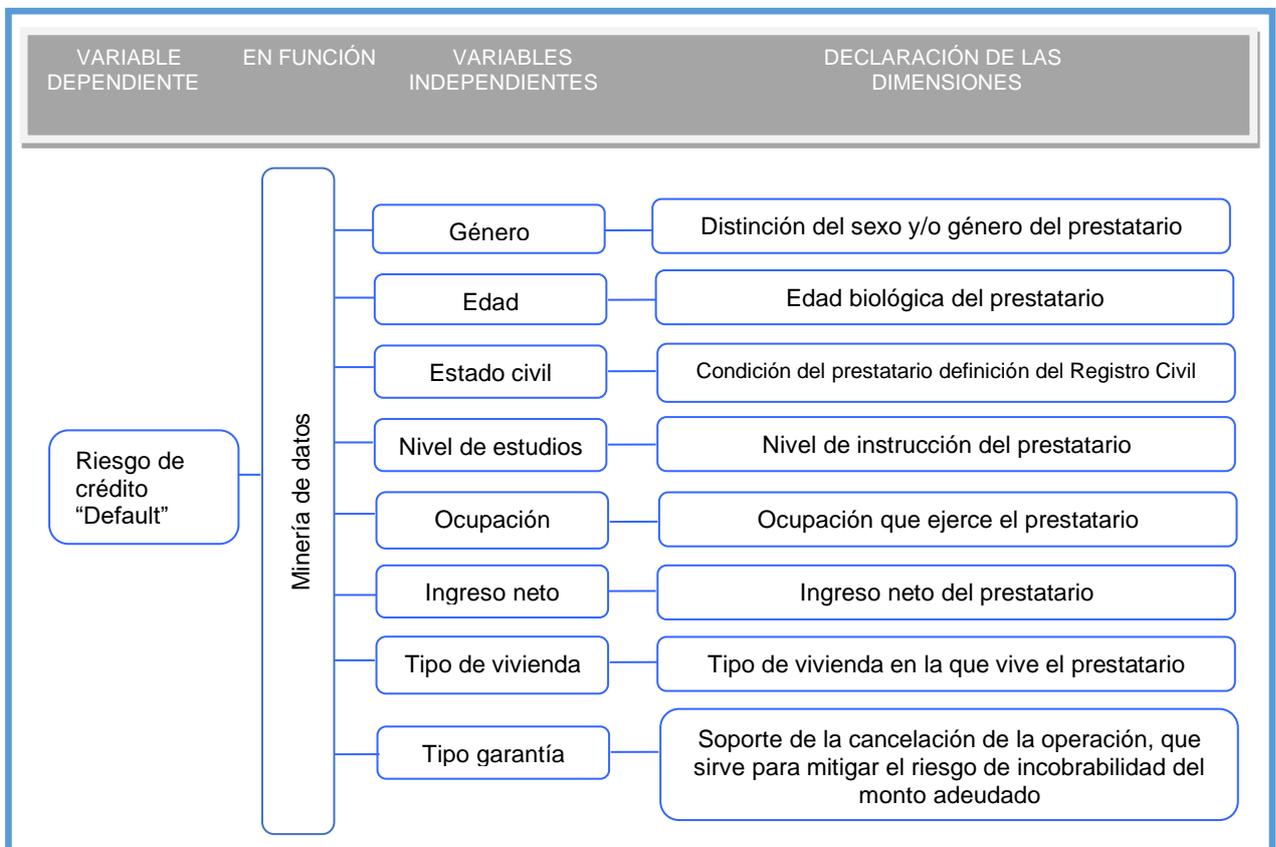
Se consideró el estudio de Rayo, S., Lara, J., Camino, D., (2010) que considera varios autores para la definición de variables a partir de los datos constantes en la solicitud de crédito de cada institución. Se consideró además el estudio de Fernández Castaño, Horacio, Pérez Ramírez, Fredy Ocaris (2005), Moreno Valencia Sandra (2013) en el análisis de la definición de la variable dependiente dummy y aplicación del modelo logit con aplicación de técnicas de tipo predictivo de minería de datos, este tipo de modelos siguen siendo muy utilizados por el Banco Central del Ecuador en sus análisis económicos e identificación de determinantes o comportamientos sujetos a estudio.

1.5.6.3. Diagrama del modelo inicial de evaluación propuesto para operacionalización

En el cuadro No. 6 se presenta el modelo de evaluación del nivel de riesgo para operacionalización en la otorgación de créditos para el Sector Cooperativo, basado en diferentes teorías, modelos, y/o metodologías recopiladas en el presente trabajo investigativo, considerando principalmente la metodología logit que usa variables similares a modelo scoring para personas

naturales modelo también recomendado por el Órgano Rector para Cooperativas de Ahorro y Crédito en el país que toma principalmente los principios de Basilea en su normativa; se consideró además estudios como el de Gutiérrez, M. (2007); Rayo, S., Lara, J., Camino, D., (2010) y Ochoa, J.C., Galeano, W., & Agudelo, L.G. (2010). que contienen análisis de varios autores para la construcción de modelos de riesgos en la determinación de variables explicativas para la evaluación del riesgo crediticio para estimar la probabilidad de riesgo “default” en individuos (personas naturales) que es la base del presente estudio con aplicación de la metodología de regresión logística bajo el modelo logit, y se consideró además el estudio de Fernández Castaño, Horacio, Pérez Ramírez, Fredy Ocaris (2005) que sigue la metodología logit, en el análisis de la definición de la variable dependiente dummy “default”.

CUADRO 6 RELACIÓN DE LA VARIABLE DEPENDIENTE E INDEPENDIENTES



Fuente: Datos recopilados en la investigación

Elaborado por: Autor

Aunque en un modelo de evaluación de riesgo de crédito pueden ser múltiples las variables objeto de estudio y análisis dependiendo del segmento de crédito

del que se trate, en la presente investigación se ha optado por hacer un análisis de las variables y comportamientos - datos recopilados en la investigación de personas naturales (individuos) que inciden en la evaluación del riesgo de crédito en la concesión de operaciones crediticias y definición de clientes del segmento de consumo de Cooperativa Ltda. cuyo perfil del cliente cumple con características de ser personas naturales que preferencialmente son asalariados, dedican a trabajar en el sector público o privado, rentistas, jubilados, estudiantes o amas de casa, entre otros; uno de los segmentos más solicitados en las instituciones financieras y con variaciones en sus indicadores de riesgo.

En el proceso de esta evaluación se consideró aspectos cualitativos y cuantitativos en la tasa “default” constantes en la solicitud de crédito de la institución (bases de datos) como caso de estudio sustentándose en las metodologías antes estudiadas; considerando como “Default” el comportamiento y calificación del cliente según lo establece los órganos rectores y el actual libro 1 de la Codificación de Resoluciones Monetarias, Financieras, De Valores y Seguros.

Se ha seleccionado 8 variables independientes con las respectivas dimensiones que involucra cada una de estas a fin de cuantificar la importancia de estas variables en la evaluación del riesgo de crédito y cuantificación del riesgo “Default” y dejar por demostrado la utilidad de técnicas y herramientas de minería de datos en este tipo de estimaciones y modelos econométricos con la aplicación de conceptos estudiados anteriormente, no se pretende demostrar un modelo de evaluación de riesgo profundo y experimentado con la descripción de las razones del porqué de las variaciones y probabilidad en las variables explicativas.

El presente trabajo es un caso de estudio para demostrar la utilidad de técnicas y herramientas de minería de datos:

Default: Define la variable dependiente con la clasificación de “Buen Cliente” o “Mal cliente”.

Género (gene): Se refiere a la definición del sexo y/o género de masculino o femenino que se le asigne a una persona natural.

Edad (rang_ed): Corresponde a la edad biológica de la persona, sin embargo; es importante mencionar que parte de las condiciones que conceden las instituciones financieras en la concesión de un crédito es fijar un límite de edad máximo menos el plazo de la operación crediticia.

Estado civil (est_civil): Se refiere a la definición jurídica de cada persona que los diferencia en la sociedad civil y que le confiere derechos y obligaciones, atributos de una persona que tienen carácter intransferibles e intransmisibles, se analizan atributos como soltero, casado, unión libre, etc.

Nivel de estudios (niv_est): Se refiere al nivel de instrucción que mantiene el prestatario al momento de la concesión de la operación crediticia. Un buen perfil puede asumirse en un sujeto que mantiene un buen nivel de estudios y que además mantiene un buen puesto, por lo que se decide analizar esta variable, que contempla atributos como: sin estudio, primaria, secundaria, universitaria, etc.

Ocupación (ocup): La ocupación u oficio podría asumirse vinculante en la concesión de una operación crediticia considerando que de esta dependerá su nivel de ingresos. Entre estas pueden determinarse atributos como asalariados (empleados públicos y privados), comerciantes, transportistas, personas que se dediquen a producción agrícola, amas de casa, estudiantes, etc.

Ingreso neto (ing_net): Diferencia entre los ingresos y gastos declarados por el sujeto de crédito al momento de la otorgación de la operación.

Tipo de vivienda (tip_viv): Atiende a la clasificación del tipo de vivienda en el que habita el prestatario al momento del inicio de la obligación de pago, se analiza atributos como: arrendada, vive con familiares, propia, etc.

Garantía (tip_gar): Parte de las condiciones en el otorgamiento de crédito son las garantías, todas las operaciones de crédito independientemente del segmento que se trate deberán estar garantizadas, son las instituciones quienes definen en sus políticas y manuales internos y tecnología de crédito los criterios para dichas garantías de conformidad a lo expuesto en el Art. 214 del Código Orgánico Monetario Financiero y las normas para la gestión del riesgo de crédito en las Cooperativas de Ahorro y Crédito. Sin embargo; pese a estar garantizado toda operación tiene su riesgo, es medida a través de sus atributos de garantía personal y real.

CAPÍTULO II. MARCO METODOLÓGICO

2. Marco metodológico

2.1. Tipo de diseño, alcance y enfoque de la investigación

En este capítulo se explica brevemente la metodología usada en el desarrollo de la presente investigación para la presentación de un modelo de evaluación de riesgo crediticio en el sector cooperativo, aplicando técnicas y herramientas de minería de datos.

2.1.1. Tipo de estudio

La información de la presente investigación es de tipo cualitativa y cuantitativa, por lo que su análisis se basa en un estudio descriptivo, exploratorio – correlacional.

Descriptivo por cuanto se presenta una revisión teórica sobre el riesgo crediticio y modelos aplicados para su evaluación, así mismo se hace un análisis descriptivo de la situación cooperativa; se analiza y describe las variables explicativas a la variable criterio “default”, representando las características y/o variables que influyen en su capacidad de pago; exploratorio - correlacional por medir la semejanza o asociación positiva o negativa de las variables de la investigación sobre otra, con el ánimo de identificar patrones de los socios prestamistas mediante el uso de la técnica de minería de datos de estadística exploratoria y de regresión logística aplicable a un modelo logit de riesgo crediticio para personas naturales en la definición del perfil del riesgo.

Para el modelo de evaluación del riesgo de crédito y predicción del riesgo de incumplimiento se utilizó la técnica de minería de datos predictiva de regresión logística (Logit) como técnica estadística y de minería, además de la técnica de minería de datos de tipo descriptivo de análisis exploratorio (tablas de contingencia). Logit al ser un modelo de probabilidad binaria, es uno de los modelos que mejor se ajusta para este tipo de análisis permitiendo predecir los

valores o probabilidades de variables dicotómicas y/o categóricas en relación a la variable dependiente “Default”, esto en la presente investigación permitió una adecuada organización de la información para la determinación del modelo predictivo final. Finalmente se realizaron pruebas a fin de determinar la bondad del modelo obtenido con la base de datos de la cartera de consumo seleccionada de la Cooperativa Ltda.

2.1.2. Metodología de la investigación

2.1.2.1. Enfoque de la investigación

El enfoque de la investigación para la definición de un modelo de evaluación de riesgo crediticio considera la evaluación de variables de tipo cualitativo y cuantitativo, considerando las teorías estudiadas para el modelamiento del riesgo de crédito bajo la metodología logit y las técnicas de minería de datos de tipo predictivo (regresión logística) y la técnica de tipo descriptivo (análisis exploratorio).

Se procedió con la recolección y análisis de datos de la Cooperativa Ltda. y bajo una comprensión de la metodología KDD, con el uso de técnicas de minería de datos seleccionadas para dar solución a nuestra problemática se efectuó la operacionalización de variables en el sistema Stata para la presentación de los resultados obtenidos con la interpretación del conocimiento arrojado (modelo predictivo).

2.2. Métodos de investigación

La metodología aplicada al presente trabajo de investigación se basó en la concepción y comprensión de Business Intelligence y del conocido KDD⁶

⁶ KDD: Es el proceso de descubrir conocimiento en grandes bases de datos con la identificación de patrones y que incluye una serie de etapas siendo una de ellas la minería de datos.

(Knowledge Discovery in Databases) y del proceso de minería de datos⁷, considerando implícitamente el método CRISP-DM⁸ para la solución de la problemática planteada. Se utilizaron además los métodos teóricos inductivo – deductivo, analíticos, métodos estadísticos que contribuyan al análisis de los datos. A continuación, se muestra la relación entre KDD y CRISP-DM:

CUADRO 7 ANALOGÍA ENTRE KDD Y CRISP-DM

CRISP-DM	KDD
Comprensión del negocio	Definición del objetivo KDD
Comprensión de datos	Selección e integración de los datos de la fuente
	Limpieza y preprocesamiento de datos (eliminar inconsistencias)
Preparar datos	Transformación
Modelamiento e interpretación	Minería de datos (selección y aplicación de algoritmos)
	Interpretación de comportamientos
Evaluación	Utilización del conocimiento

Fuente: Adaptado de Azevedo, A. y Santos, M. (2008).

Elaborado por: Autor

La presente investigación no pretende probar y realizar un back-testing del modelo obtenido en la Cooperativa Ltda. sujeta a estudio respecto al último paso del KDD que concierne a la utilización del conocimiento y aplicación.

Método deductivo: Este método trata de la obtención de conclusiones particulares a partir de un todo, en este caso específico se inició con la revisión de la literatura existente de Riesgo Crediticio; Modelos de evaluación de riesgo y Minería de datos; se revisó modelos previamente estudiados para la evaluación de riesgo crediticio y la determinación de variables de estudio con el análisis de teorías, metodologías, postulados, principios, aplicados a la problemática en estudio. A continuación se buscó las fuentes de datos necesarias, recolectados los datos y una vez definidos los objetivos de la investigación se procedió con la limpieza y preprocesamiento de los datos con el ánimo de normalizar la data (eliminar datos innecesarios y rellenar los faltantes, así como pruebas estadísticas para establecer las variables candidatas y eliminar variables innecesarias); luego se procedió con la

⁷ Minería de datos: Constituye una parte del proceso KDD que corresponde a la selección y uso de algoritmos como parte de este gran proceso para generar conocimiento.

⁸ CRISP-DM: Modelo analítico de la minería de datos que involucra un proceso integral de análisis de datos para la construcción de un modelo hasta su etapa final.

transformación de datos; hecho posterior se seleccionó la herramienta tecnológica que sería usada para hacer minería de datos con la selección de algoritmos o técnicas, estadísticos de minería y finalmente para la aplicación de las técnicas seleccionadas e interpretación de los resultados arrojados con los patrones que genere este análisis.

Método histórico: Esta investigación recopiló datos históricos de operaciones crediticias concedidas en Cooperativa Ltda. con corte a mayor de 2019 con un histórico de un año de conformidad a las recomendaciones que otorga del Comité de Basilea para la estimación de variables en modelos de evaluación de riesgo crediticio para instituciones financieras.

Método analítico: En la presente investigación se hizo necesario desagregar el objeto de estudio para cuantificar el comportamiento y las probabilidades de las variables explicativas que afectan a la variable dependiente “default”, para demostrar la utilidad de algoritmos de minería de datos y estadísticos en el sistema seleccionado.

2.3. Unidad de Análisis, población y muestra

Para la estimación del modelo predictivo para evaluación de riesgo de crédito basado en una metodología logit crediticio se optó como fuente principal una malla de datos concedida por la Cooperativa de Ahorro y Crédito del Segmento 2 con presencia en la ciudad de Loja, que por confidencialidad no se revela su razón social y/o nombre comercial, y que para este trabajo se ha denominado Cooperativa Ltda.

Con fecha 27 de mayo de 2019 mediante solicitud se requirió al gerente de la Cooperativa una malla de datos de los créditos concedidos y que la institución mantiene con corte al 30 de mayo de 2019, se solicitó que dichos sujetos contengan mínimo un año de comportamiento, con sus variables cuantitativas y cualitativas ligadas al perfil del acreditado según consta en la “Solicitud de Crédito” y tecnología crediticia de la institución (Capacidad, Capital, Colateral, Carácter, Condiciones, otros), es importante mencionar que el Art. 353 del

Código Orgánico Monetario Financiero y Financiero señala que los depósitos y captaciones del sistema financiero nacional están sujetos a sigilo bancario, más no hace referencia al caso específico de colocaciones, en su parte pertinente de este artículo señala además que *“Las entidades del sistema financiero nacional podrán dar a conocer las operaciones anteriores en términos globales, no personalizados ni parcializados, solo para fines estadísticos o de información, cuando exista un interés público”* (Código Orgánico Monetario Financiero, 2014), lo que busca la norma antes citada es precautelar los derechos financieros de cada individuo y específicamente a sus datos personales y su situación en el sistema financiero, sin embargo; el presente estudio requiere información general del comportamiento del sujeto de crédito para fines de evaluación lo que es permitido por la ley, más no se transgrede ni se afecta a ningún individuo, en tal virtud; y por ser únicamente para fines educativos no se solicitó datos personales del individuo como cédulas, nombres y/o apellidos; una vez evaluado dicho requerimiento y por no contravenir norma alguna, la Cooperativa autoriza el acceso a los datos solicitados y con fecha 18 de junio de 2019 con correo institucional se nos remite la data solicitada (no incluye datos personales de los sujetos de créditos) únicamente variables de tipo cualitativo y cuantitativo, sin embargo; pese a lo antes manifestado el presente trabajo no concede el nombre de la Cooperativa por fines de confidencialidad y reserva absoluta, dado que lo que se pretende con este trabajo es demostrar la utilidad de técnicas y herramientas de minería de datos en la construcción de modelos de evaluación de riesgo que sirvan de apoyo para instituciones del sistema financiero, más no profundizar en el comportamiento crediticio de la institución que nos facilitó la información.

Esta base datos inicialmente contaba con 28194 registros y 33 variables incluyendo variables de tipo informativo que no aportaban al modelo, con un histórico de un año de las operaciones crediticias existentes con corte a mayo de 2019 en atención a las recomendaciones de organismos expertos para este tipo de análisis.

Para efectos de este trabajo se tomó la información actual de las operaciones crediticias del segmento de consumo (personas naturales) existentes a mayo

de 2019 con su respectiva calificación de crédito del sujeto de crédito a la fecha.

La muestra obtenida luego de la limpieza, preprocesamiento y transformación de datos fue de 2145 registros de socios prestamistas en su calidad de personas naturales que han sido sujetos de crédito de consumo en Cooperativa Ltda, a fin de analizar su comportamiento de pago desde la concesión de la operación de crédito y determinar el modelo predictivo.

Para la determinación de variables candidatas para el modelo inicial se consideró las teorías y metodologías estudiadas en la presente investigación y la información de la solicitud de crédito de Cooperativa Ltda., siendo así que en la fase de normalización, limpieza, preprocesamiento y transformación de los datos se eliminó variables informativas, incompletas, inconsistentes y que no aportaban a la construcción del modelo de acuerdo a corridas estadísticas previas, a fin de normalizar la data a la cual se le aplicaría minería de datos con técnicas estadísticas con la herramienta seleccionada, quedando como se lo expuso antes una muestra total de 2145 registros y 8 variables cualitativas y cuantitativas de socios prestamistas que conforman la base de datos de la cartera y que se les ha concedido crédito de consumo en Cooperativa Ltda.

2.4. Variables de investigación y operacionalización

La presente investigación para la construcción de un modelo de evaluación de riesgo crediticio comprende un análisis a personas naturales (créditos individuales de consumo) que suscribieron operaciones crediticias con la institución Cooperativa Ltda. (caso de estudio), el corte se lo realiza a mayo de 2019 con un histórico de un año, analizándose las variables constantes en la base de datos de la institución en un proceso de normalización con aplicación a teorías previamente revisadas en el marco teórico de esta investigación.

Con el ánimo de realizar la estimación la probabilidad de incumpliendo “Default” y sus variables explicativas se usó la herramienta estadística Stata, así como

Qlik Sense para análisis financieros de la concentración de crédito en el sector cooperativo y Cooperativa Ltda.

Para el análisis de la variable “Default” y definición de clientes buenos y malos se ha considerado los criterios de calificación de créditos de la institución en análisis en función de la morosidad que contempla el Libro 1 de la Codificación de Resoluciones Monetarias, Financieras, de Valores y Seguros.

CUADRO 8 DE LA CALIFICACIÓN CARTERA DE CRÉDITO

NIVEL DE RIESGO		DÍAS DE MOROSIDAD
Riesgo normal	A	de 0 hasta 30
Riesgo potencial	B	de 31 hasta 60
Riesgo deficiente	C	de 61 hasta 120
Dudoso recaudo	D	de 121 hasta 180
Pérdida	E	mayor a 180

Fuente: JPRMF, Codificación de Resoluciones Monetarias, Financieras, de Valores y Seguros, 2019

Elaborado por: Autor

En base las recomendaciones del Comité de Supervisión Bancaria Basilea para la presente estimación se ha considerado un horizonte en el tiempo de un año, con corte a mayo de 2019.

Para la operacionalización de variables en base a la revisión literaria previamente estudiada se ha definido la relación constante en el “Cuadro 6 Relación de la variable dependiente e independiente” antes expresada.

2.5. Fuentes, técnicas e instrumentos para la recolección de información

Para la evaluación del riesgo crediticio se ha considerado un enfoque basado en la propia problemática y realidad de la institución.

Los datos que se utilizaron para probar el presente modelo fueron tomados de la Cooperativa Ltda. con presencia en la ciudad de Loja del Segmento 2, cartera de Consumo (Personas Naturales), que por asuntos de confidencialidad solicitados por la misma cooperativa su nombre no será proporcionado.

Las fuentes, técnicas e instrumentos utilizados fueron:

2.5.1. Fuentes de información

La revisión literaria de la presente investigación se clasifica en primaria y secundaria:

Fuentes primarias

Como fuentes primarias se identificó:

- Matriz de créditos de una Cooperativa de Ahorro y Crédito con presencia en la ciudad de Loja
- Revisión de investigaciones y artículos científicos para evaluación de modelos de riesgo crediticio
- Información de la Superintendencia de Economía Popular y Solidaria
- Información del Banco Central del Ecuador
- Organismos que regulan a las Cooperativas de Ahorro y Crédito

Fuentes secundarias

Como fuente secundaria se identificó:

- Información documental de otras fuentes de información
- Información de artículos científicos para identificación de modelos y variables de evaluación de riesgo crediticio.
- Libros de técnicas y herramientas de minería de datos.
- Tesis relacionadas al tema de investigación y problema planteado.

2.5.2. Técnicas para la recolección de información

2.5.2.1. Técnica de investigación estadística

Esta técnica fue usada en el análisis de los datos recolectados de la Cooperativa Ltda., para efectuar análisis descriptivos, exploratorios - correlacionales como técnica estadística de minería de datos para medir el comportamiento de las variables y atributos considerados para el presente estudio.

2.5.2.2. Técnica de investigación documental

La técnica de investigación documental fue una parte fundamental en este proceso investigativo, siendo de utilidad para la recopilación de información, metodologías, teorías, etc. que sustenten, den solución y explicación a la problemática planteada y objeto de estudio, el conocimiento fue recopilado de tesis, artículos científicos, entes gubernamentales, informes financieros, organismos de control y regulación, entre otros.

2.6. Tratamiento de la información

La base de datos para análisis de las variables que influyen en la estimación de la variable "Default" corresponde a una cartera de consumo y fue proporcionada con una Cooperativa de Ahorro y Crédito del Segmento 2 con presencia en la ciudad de Loja desde hace varios años, por razones de confidencialidad no se revela la razón social y/o nombre comercial.

La forma de tratar los datos de la presente investigación fue mediante una comprensión de la metodología KDD, mediante el análisis, limpieza, pre procesamiento y transformación de los mismos a fin de normalizar los datos, eliminar datos innecesarios y rellenar los datos faltantes previo a selección y análisis de variables en estudio sustentadas en la revisión literaria y análisis estadísticos para la aplicación de técnicas de minería de datos de tipo

predictivo y descriptivo con el software estadístico seleccionado “Stata”, una gran herramienta de ciencia de datos con múltiples técnicas y modelos para realizar análisis predictivo de grandes bases de datos, este permitió obtener un modelo predictivo con análisis de las variables independientes en relación a la variable dependiente “default” (buen o mal cliente), este es un paquete idóneo para que una institución que no requiere grandes inversiones efectúe minería de datos con técnicas de tipo predictivo.

Además, se utilizó Qlik Sense para análisis financieros en su versión gratuita y personal, que es un analítico de datos con aplicación de BI con visualizaciones inteligentes, bastante interactiva para responder a interrogantes en la data analizada.

CAPÍTULO III RESULTADO Y DISCUSIÓN

3. Resultado y discusión

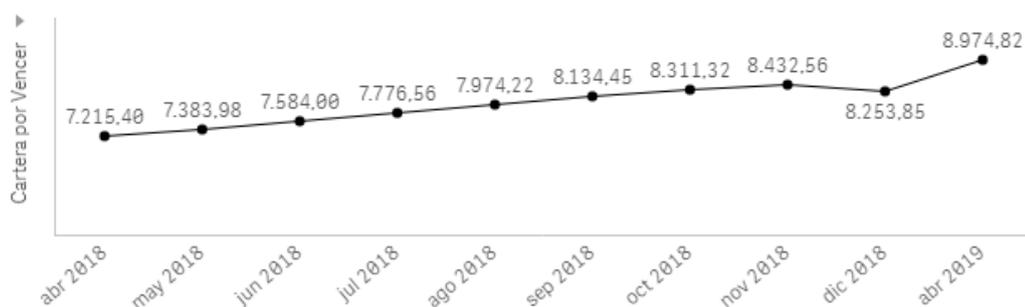
3.1. Análisis de la situación actual

3.1.1. Entorno Cooperativo

El Sistema Financiero Nacional desde décadas pasadas cumple un papel crucial en el desarrollo de la economía nacional y aún más con la vigencia del Código Orgánico Monetario y Financiero en el mes de septiembre del año 2014 donde el sistema empezó a tener nuevos desafíos.

Si se realiza un análisis de la cartera por vencer del Sector Cooperativo en el país del último año con corte abril 2019 esta cartera según datos obtenidos de las SEPS con los datos históricos se observa un crecimiento de la cartera en el transcurso del tiempo, con una variación anual positiva del 24,40% y una variación mensual positiva del 1.10%, como se observa en la Figura No. 5.

FIGURA 5 CARTERA POR VENCER SECTOR COOPERATIVO

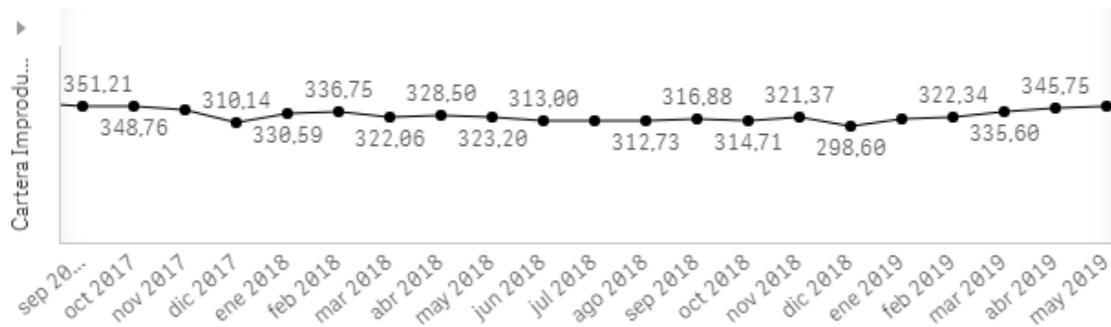


Fuente: SEPS, 2019 (Productos estadísticos)
Elaborado por: Análisis autor con aplicación QlikSense

Respecto a la cartera improductiva que es aquella que no genera una renta financiera y que se integra de la cartera vencida y la cartera que no devenga intereses, en un escenario nacional del sector cooperativo se presentan variaciones positivas y negativas importantes.

FIGURA 6 CARTERA IMPRODUCTIVA SECTOR COOPERATIVO

Cartera Histórica

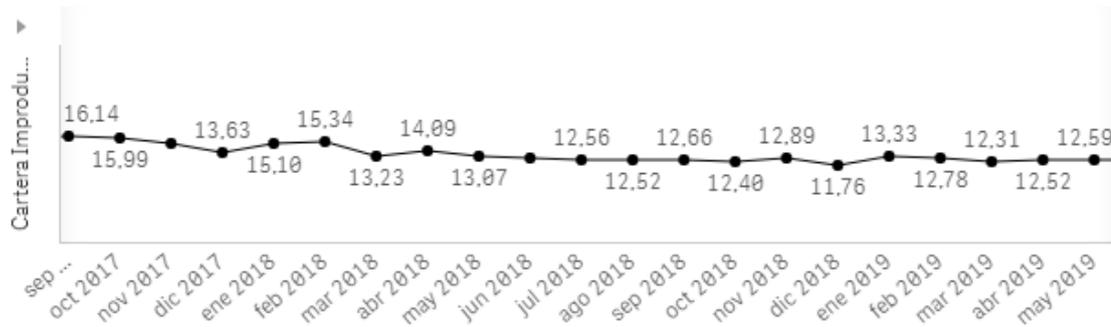


Fuente: SEPS, 2019 (Productos estadísticos)
Elaborado por: Autor

En Loja la cartera improductiva presenta una gráfica con tendencias similares, está al alza y a la baja por lo que las cooperativas tienen que hacer esfuerzos para enfrentar este tipo de situaciones:

FIGURA 7 CARTERA IMPRODUCTIVA SECTOR COOPERATIVO LOJA

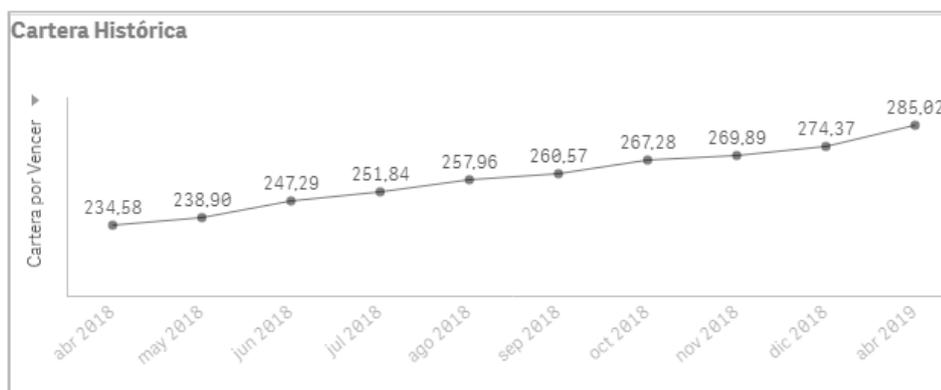
Cartera Histórica



Fuente: SEPS, 2019 (Productos estadísticos)
Elaborado por: Autor

En el caso de la cartera por vencer con una variación positiva del 21.50% y mensual del 0.90%. A continuación, en la Figura No. 8 véase el deterioro de la cartera por vencer en el sector Cooperativo con presencia en la ciudad de Loja:

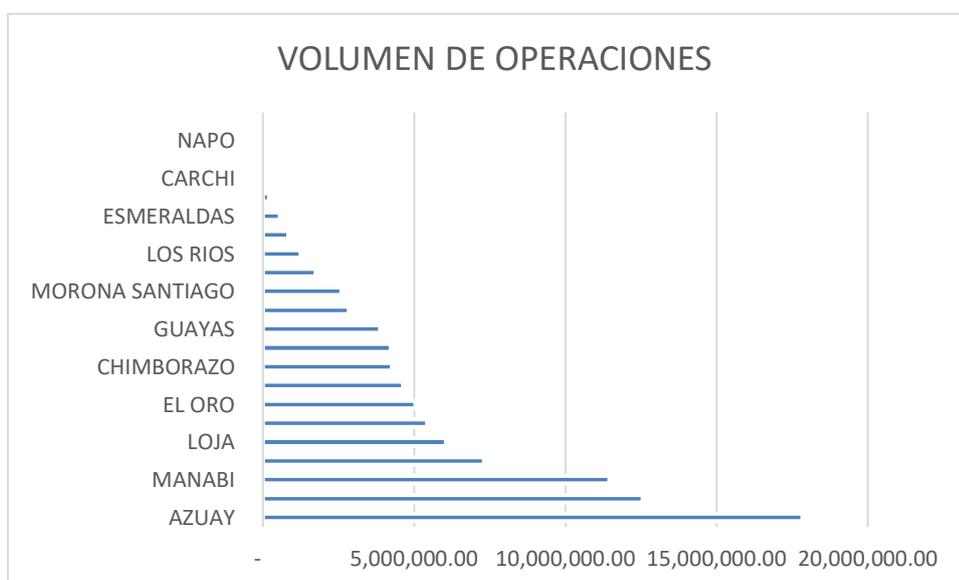
FIGURA 8 CARTERA POR VENCER SECTOR COOPERATIVO LOJA



Fuente: SEPS, 2019 (Productos estadísticos)
Elaborado por: Análisis autor con aplicación QlikSense

El Sector de Cooperativas de Ahorro y Crédito del primer piso incluye los segmentos 1 y 2. Las Cooperativas de Ahorro y Créditos de los segmentos 1 y 2 son los más representativos a nivel de activos y colocación de operaciones crediticias. El volumen de crédito para el segmento 2 del 01 al 30 de abril de 2019 mantiene un total general de USD \$ 92,965,826.40 millones de dólares. Donde Loja mantiene una participación del 6.50%. Véase en la Figura No. 9 la tendencia en el volumen de colocaciones por provincia.

FIGURA 9 VOLUMEN DE OPERACIONES POR PROVINCIA

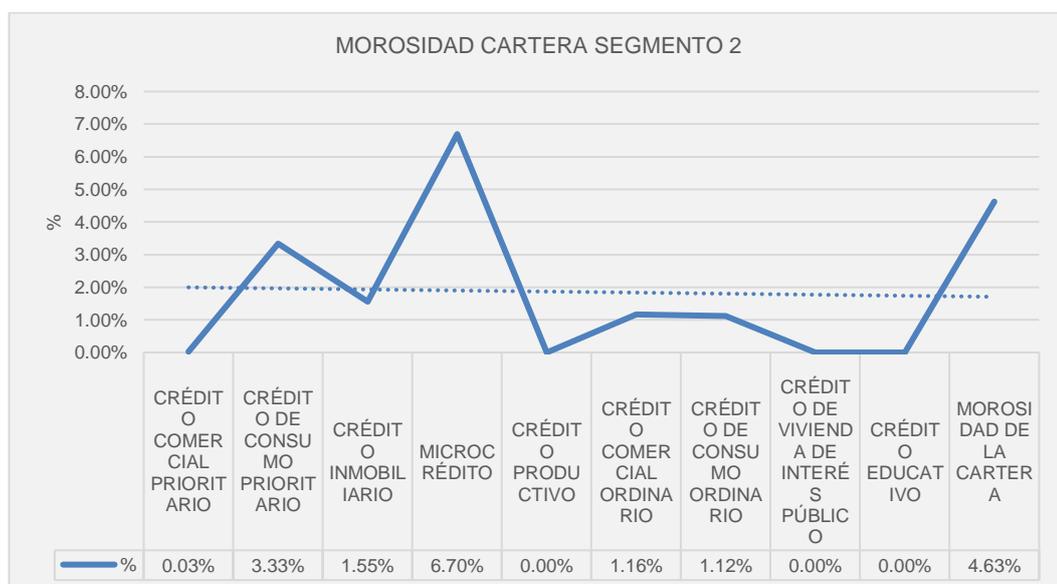


Fuente: SEPS, 2019 (Productos estadísticos)
Elaborado por: Autor

De acuerdo con la (Superintendencia de Bancos y Seguros del Ecuador, 2014), el riesgo de crédito puede medirse mediante 2 indicadores, el primero de ellos es el indicador de morosidad, el cual se mide como la relación que existe entre la cartera bruta y la improductiva.

De acuerdo con la información que proporciona la Superintendencia de Economía Popular y Solidaria – SEPS el indicador de morosidad a nivel Cooperativo para el Segmento 2 para el mes de marzo de 2019 cerró en 4.63%. (Véase Figura No. 10 Morosidad Segmento 2).

FIGURA 10 MOROSIDAD SEGMENTO 2



Fuente: SEPS, 2019 (Productos estadísticos)

Elaborado por: Autor

El mayor indicador de morosidad global de las Cooperativas de Ahorro y Crédito del Segmento 2 se registra en el segmento de Microcrédito con un 6.70% seguido por el segmento de consumo prioritario con un 3.33%.

Los índices de morosidad de la Cooperativa Ltda. A un mes determinado del presente año, según datos publicados por la Superintendencia de Economía Popular y Solidaria se registran en 3.68% de acuerdo con el siguiente detalle:

CUADRO 9 ÍNDICES DE MOROSIDAD POR TIPO CARTERA COOPERATIVA LTDA.

INDICES DE MOROSIDAD	%
MOROSIDAD DE CREDITO DE CONSUMO PRIORITARIO	2.09%
MOROSIDAD DE LA CARTERA DE CREDITO INMOBILIARIO	2.84%
MOROSIDAD DE LA CARTERA DE MICROCREDITO	8.88%
MOROSIDAD DE LA CARTERA TOTAL	3.68%

Fuente: SEPS, 2019 (Productos estadísticos)

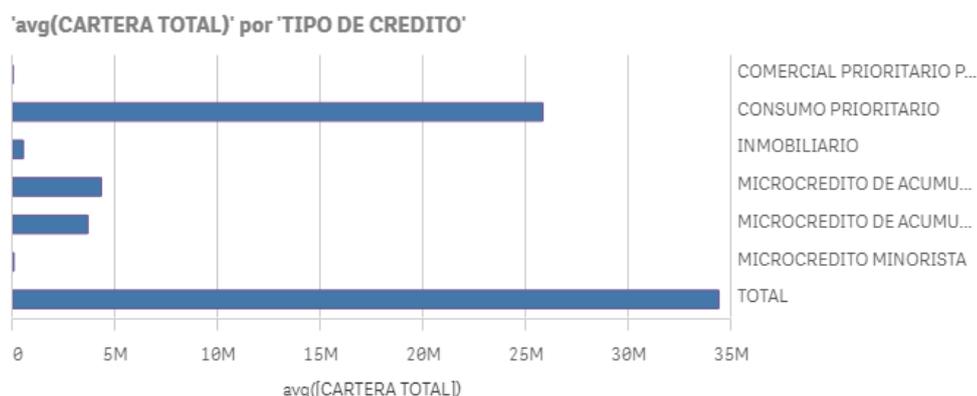
Elaborado por: Autor

En relación a un mes anterior del antes analizado presenta una variación negativa cerrando para este mes en 3.80% en la cartera total. En un escenario comparativo con el mismo mes al 2018 es del del 4.43%, y para el año 2017 con un cierre del indicador de mora del 6.75%, presenta una disminución lo que es positivo para la institución, lo que demuestra que se han tomado estrategias para este rubro que es bastante volátil e incierto en sus deterioros.

3.1.2. Situación Cooperativa Ltda.

En cuanto a la concentración de la cartera de la Cooperativa Ltda. Al mes de abril de 2019 en mayor proporción se encuentra en el segmento de consumo prioritario seguido de microcrédito (Véase Figura No. 11), por lo que es importante que en estos segmentos la cooperativa aplique metodologías apropiadas para la evaluación de los sujetos de crédito, considerando que en el apartado anterior se denotó que en estos dos segmentos se registra el mayor indicador de morosidad.

FIGURA 11 CONCENTRACIÓN DE LA CARTERA COOPERATIVA LTDA.



Fuente: SEPS, 2019 (Productos estadísticos)

Elaborado por: Análisis autor con aplicación QlikSense

Precisamente es el segmento de consumo prioritario (Véase figura No. 12) dentro de la Cooperativa Ltda. El que mayor recuento de operaciones y sujetos de crédito mantiene para todos los meses del año 2019, veamos un análisis de un mes del presente año:

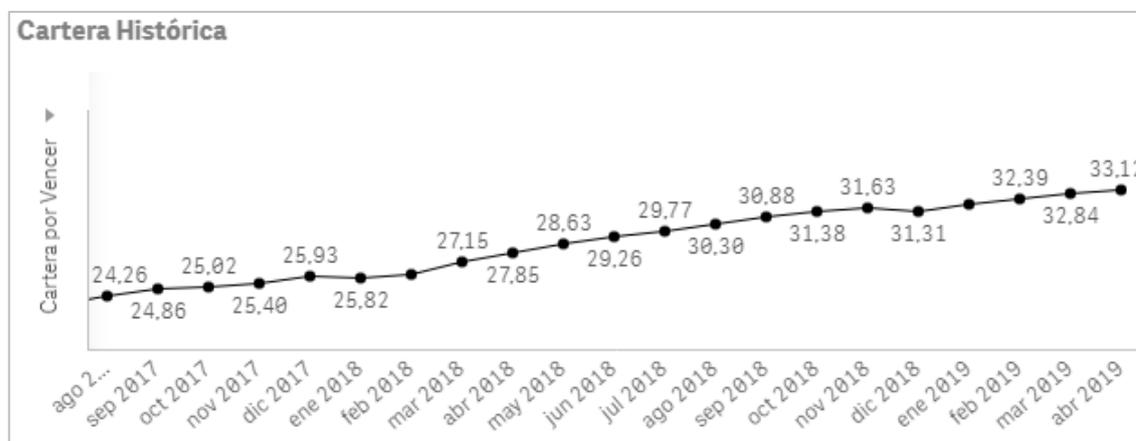
FIGURA 12 SUJETOS DE CRÉDITO COOPERATIVA LTDA.

TIPO DE CREDITO <input type="text"/>	Sum([NUMERO SUJETOS CREDITO])
COMERCIAL PRIORITARIO PYMES	4
CONSUMO PRIORITARIO	3007
INMOBILIARIO	20
MICROCREDITO DE ACUMULACION AMPLIADA	258
MICROCREDITO DE ACUMULACION SIMPLE	1202
MICROCREDITO MINORISTA	65
TOTAL	4556

Fuente: SEPS, 2019 (Productos estadísticos)
Elaborado por: Análisis autor con aplicación QlikSense

En un escenario más amplio en Cooperativa Ltda. (Véase Figura No. 13) se denota un incremento de la cartera por vencer en el transcurso del tiempo con una variación positiva del 18.90% y mensual del 0.8%.

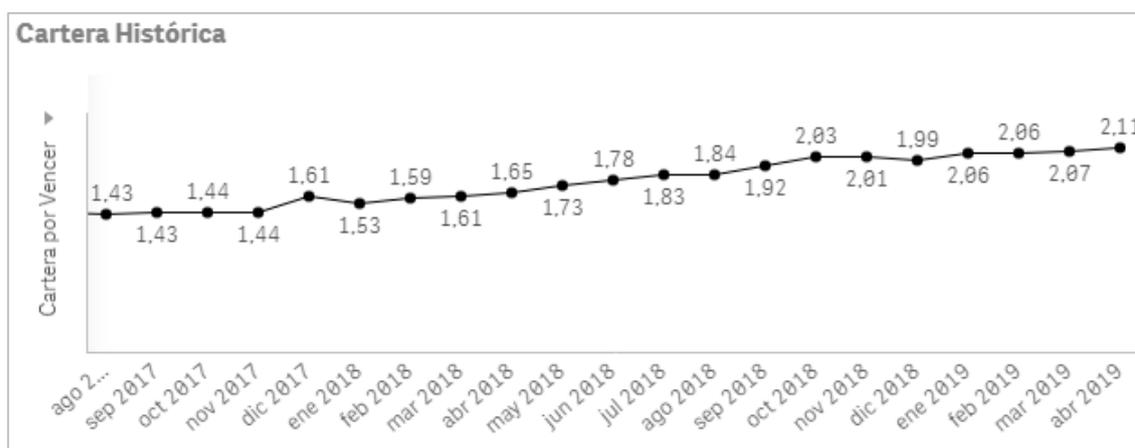
FIGURA 13 CARTERA POR VENCER COOPERATIVA LTDA.



Fuente: SEPS, 2019 (Productos estadísticos)
Elaborado por: Análisis autor con aplicación QlikSense

A diferencia de todos los cantones donde tiene presencia Cooperativa Ltda. en la ciudad de Loja mantiene un escenario similar con una variación positiva anual del 27.50% y mensual del 1.60% (Véase Figura No. 14).

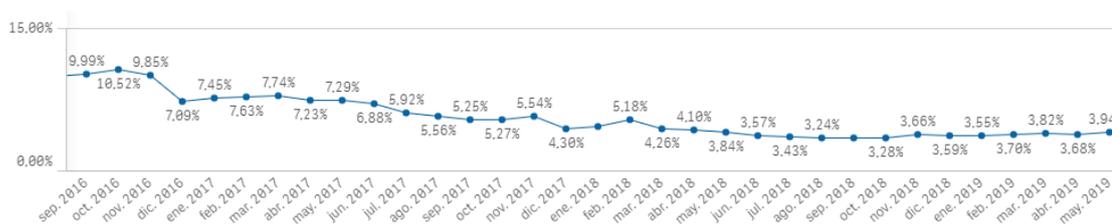
FIGURA 14 CARTERA POR VENCER COOPERATIVA LTDA. LOJA



Fuente: SEPS, 2019 (Productos estadísticos)
Elaborado por: Análisis autor con aplicación QlikSense

En cuanto a la calidad de la cartera y cartera improductiva que resulta de la cartera vencida más la cartera que no devenga intereses, a continuación; se muestra gráficamente (Véase Figura No. 15) la variación del índice de morosidad en los últimos años de la Cooperativa Ltda.

FIGURA 15 VARIACIÓN DEL INDICADOR DE MOROSIDAD COOPERATIVA LTDA.



Fuente: SEPS, 2019 (Productos estadísticos)
Elaborado por: Análisis autor con aplicación QlikSense

Acorde a lo antes expuesto, se puede concluir que el índice de morosidad en la Cooperativa Ltda. presenta variaciones significativas, pero con una tendencia positiva, sin embargo; la calidad de la cartera y la cartera riesgosa es bastante vulnerable como se denota en la gráfica anterior más aún lo será sino se posee métodos eficientes de evaluación de riesgo de crédito, por lo que es crucial que

se adopten estrategias que consigan resultados aún mayores de disminución de estos indicadores, con atención en los segmentos de mayor concentración que usualmente son concedidos a personas naturales (créditos individuales).

3.2. Análisis comparativo, evolución, tendencias y perspectivas

Para la definición del modelo de score de crédito se ha considerado la base de datos de cartera de créditos de consumo de una cooperativa de ahorro y crédito con corte a mayo de 2019. Una vez depuradas y normalizada la data de la Cooperativa con el descarte de variables informativas, inconsistentes en una primera corrida estadística y análisis inicial de la data para ajuste del modelo y relación entre variables, se ha definido las siguientes entre cualitativas y cuantitativas para su análisis y mediante las cuales se pretende definir los comportamientos e interacciones existentes. El modelo inicial se define en ocho variables explicativas y una variable dicotómica o binaria dependiente de la siguiente manera:

CUADRO 10 TABLA DE VARIABLES PARA SU ANÁLISIS

Variable	Nomenclatura	Descripción	Tipo	Categorías
Default	default	Binaria	Dependiente	Buen cliente / Mal cliente
Género	gene	Binaria	Independiente	Masculino / Femenino
Edad	rang_ed	Intervalo	Independiente	De: 18 hasta 32 / De: 33 hasta 47 / De: 48 hasta 62 / De: 63 O más
Estado civil	est_civil	Nominal	Independiente	Soltero / Unión Libre y Casado / Casado con disolución conyugal / Divorciado / Viudo
Nivel de estudios	niv_est	Ordinal	Independiente	Sin estudio / Primaria / Secundaria / Formación Intermedia Técnica, Superior y Cuarto Nivel
Ocupación	ocup	Nominal	Independiente	Estudiante, Ama de Casa, Jubilado / Asalariados / Comercio de bienes y servicios y actividades independientes / Transporte pasajeros y de carga / Construcción, Explotación, Industriales / Producción agrícola, ganadera, silvicultura y pesca
Ingreso neto	ing_net	Continua	Independiente	

Tipo de vivienda	tip_viv	Nominal	Independiente	Prestada / vive con familiares / arrendada / propia hipotecada o no hipotecada
Tipo de garantía	tip_gar	Binaria	Independiente	Personal / Real

Fuente: Datos recopilados en la investigación

Elaborado por: Autor

El propósito principal del presente análisis fue evaluar las características y composición de las variables cualitativas y cuantitativas de la información disponible mediante estadística descriptiva y un análisis exploratorio de la calidad de información de las variables que potencialmente pueden formar parte del modelo final de riesgo de crédito.

Donde el diccionario para las variables y definición de las variables y categorías para la herramienta seleccionada (Stata) para esta investigación se lo ha definido de la siguiente manera:

CUADRO 11 DICcionario DE VARIABLES Y CATEGORIZACIÓN

Variable dependiente "Default":

DESCRIPCIÓN	VALOR	CALIFICACIÓN
Mal cliente	0	B hasta E
Bueno cliente	1	A

Variables independientes:

DESCRIPCIÓN	VALOR
Variable independiente género	
Nomenclatura: gen	
Masculino	0
Femenino	1

DESCRIPCIÓN	VALOR
Variable independiente edad	
Nomenclatura: rang_ed	
De: 18 hasta 32	0
De: 33 hasta 47	1
De: 48 hasta 62	2
De: 63 o más	3

DESCRIPCIÓN	VALOR
Variable independiente estado civil	
Nomenclatura: est_civil	
Soltero	0
Unión Libre y Casado	1
Casado con disolución conyugal	2
Divorciado	3
Viudo	4

DESCRIPCIÓN	VALOR
Variable independiente nivel de estudios	
Nomenclatura: niv_est	
Sin estudio	0
Primaria	1
Secundaria	2
Formación intermedia técnica, superior y cuarto nivel	3

DESCRIPCIÓN	VALOR
Variable independiente ocupación	
Nomenclatura: ocup	
Estudiantes, amas de casa y jubilados	0
Asalariados	1
Comercio bienes y servicios, Actividades independientes	2
Transporte terrestre aéreo marítimo de carga y pasajeros	3
Construcción, explotación, industriales	4
Producción agrícola, ganadera, silvicultura y pesca	5

DESCRIPCIÓN	VALOR
Variable independiente ingreso neto	
Nomenclatura: ing_net	
Diferencia entre ingresos y gastos declarados por el socio prestamista	

DESCRIPCIÓN	VALOR
Variable independiente tipo de viviente	
Nomenclatura: tip_viv	
Prestada	0
Vive con familiares	1
Arrendada	2
Propia hipotecada y no hipotecada	3

DESCRIPCIÓN	VALOR
Variable independiente tipo de garantía	
Nomenclatura: tip_gar	
Personal	0
Real	1

Fuente: Datos recopilados en la investigación
Elaborado por: Autor

Análisis Descriptivo y exploratorio de las variables analizadas

Definido la base de datos muestral y las variables potenciales candidatas se procedió a examinarla con un análisis descriptivo y exploratorio para conocer una poco más a detalle las variables disponibles.

Se procedió con el análisis de la distribución de frecuencias de las variables analizadas en la muestra seleccionada. En el caso de la variable dependiente “Default” que cumple un rol crucial en el presente estudio, considerada una variable cualitativa binaria o dicotómica, definida entre dos valores 0 y 1, donde 0 representa “Mal cliente” y 1 “Buen Cliente” de acuerdo a la calificación de créditos considerando los días de morosidad constantes en el Libro 1 de la Codificación de Resoluciones Monetarias, Financieras, de Valores y Seguros y que se define en que el socio prestamista no cumpla con sus obligaciones pactadas total o parcialmente en los plazos pactados.

Como se observa en la tabla siguiente que incluye el total de la muestra, hay 53 personas (2.47%) considerados como malos clientes es decir que mantienen una morosidad que va desde los 31 días hasta mayor a 180 días (Calificación B hasta E) y 2092 personas (97.53%) considerados como buenos clientes y que su nivel de riesgo es definido con calificación A por ir desde los 0 hasta los 30 días de mora. Esto corrobora generalmente una buena calidad de sus socios prestamistas en el segmento de consumo.

TABLA 1 DEFINICIÓN DE LA VARIABLE EN ANÁLISIS “DEFAULT”

default	Freq.	Percent	Cum.
0	53	2.47	2.47
1	2,092	97.53	100.00
Total	2,145	100.00	

Fuente: Datos recopilados en la investigación
Elaborado por: Autor

En este apartado vamos a describir las variables independientes consideradas en el modelo inicial:

En el análisis de la variable “gene” de la muestra total de créditos de consumo, el género definido como masculino representan el 55.85% es decir 1198 personas; y el 44.5% representan el género femenino es decir 947 personas, lo que denota que para el segmento de consumo el género masculino tiene una mayor participación de la cartera.

TABLA 2 FRECUENCIAS DE LA VARIABLE ANALIZADA “GENE”

gene	Freq.	Percent	Cum.
Masculino	1,198	55.85	55.85
Femenino	947	44.15	100.00
Total	2,145	100.00	

Fuente: Datos recopilados en la investigación

Elaborado por: Autor

En el caso de la variable independiente de intervalo “rang_ed”, muestra que la mayor participación registrada de créditos colocados está entre los 33 a 47 años de edad con un 37.67%, seguido de un 30.91% entre los 18 a 32 años.

TABLA 3 FRECUENCIAS DE LA VARIABLE ANALIZADA “RANG_ED”

rang_ed	Freq.	Percent	Cum.
De18a32años	663	30.91	30.91
De33a47años	808	37.67	68.58
De48a62años	507	23.64	92.21
De63omas	167	7.79	100.00
Total	2,145	100.00	

Fuente: Datos recopilados en la investigación

Elaborado por: Autor

La variable cualitativa con varias categorías de tipo nominal “est_civil” registra la mayor participación para unión libre y casados con un 44.20% seguido de personas en estado civil solteros con un 40.79%.

TABLA 4 FRECUENCIAS DE LA VARIABLE ANALIZADA “EST_CIVIL”

est_civil	Freq.	Percent	Cum.
Soltero	875	40.79	40.79
UnionLCasd	948	44.20	84.99
CasadoDisCony	80	3.73	88.72
Divorciado	186	8.67	97.39
Viudo	56	2.61	100.00
Total	2,145	100.00	

Fuente: Datos recopilados en la investigación

Elaborado por: Autor

En la cartera de consumo en la muestra total seleccionada, las personas con nivel de educación de secundaria representan el 43.82% seguido de un 35.20% de personas con formación técnica, superior y de cuarto nivel.

TABLA 5 FRECUENCIAS DE LA VARIABLE ANALIZADA “NIV_EST”

niv_est	Freq.	Percent	Cum.
sinest	39	1.82	1.82
prim	411	19.16	20.98
sec	940	43.82	64.80
tecEducSCuart	755	35.20	100.00
Total	2,145	100.00	

Fuente: Datos recopilados en la investigación

Elaborado por: Autor

En el caso de la variable de tipo nominal categórica “ocup”, la mayor concentración dentro la muestra seleccionada son los asalariados (empleados públicos y privados) con un 41.40% seguido de un 24.48% de personas que se dedican al comercio de bienes y servicios y otras actividades independientes.

TABLA 6 FRECUENCIAS DE LA VARIABLE ANALIZADA “OCUP”

ocup	Freq.	Percent	Cum.
EstAmacJub	238	11.10	11.10
Asal	888	41.40	52.49
ComActInd	525	24.48	76.97
Transp	138	6.43	83.40
Const_Exp_Ind	148	6.90	90.30
Pagosp	208	9.70	100.00
Total	2,145	100.00	

Fuente: Datos recopilados en la investigación

Elaborado por: Autor

La variable nominal categórica “tip_viv” registra la mayor participación en personas con vivienda propia esté o no hipotecada con un 46.57% seguido de un 41.21% de personas que viven con familiares.

TABLA 7 FRECUENCIAS DE LA VARIABLE ANALIZADA “TIP_VIV”

tip_viv	Freq.	Percent	Cum.
Prestada	8	0.37	0.37
VivFam	884	41.21	41.59
Arrend	254	11.84	53.43
PropHyNoH	999	46.57	100.00
Total	2,145	100.00	

Fuente: Datos recopilados en la investigación
Elaborado por: Autor

En el caso de la variable binaria “tip_gar” muestra que dentro de la cartera de consumo un 59.72% es decir 1281 personas que han sido objeto de crédito en Cooperativa Ltda. mantiene una garantía sobre firmas o de tipo personal, esto mucho depende de las políticas internas de cada institución sobre los montos solicitados.

TABLA 8 FRECUENCIAS DE LA VARIABLE ANALIZADA “TIP_GAR”

tip_gar	Freq.	Percent	Cum.
Personal	1,281	59.72	59.72
Real	864	40.28	100.00
Total	2,145	100.00	

Fuente: Datos recopilados en la investigación
Elaborado por: Autor

Tablas de contingencias de variables analizadas cualitativas

En este apartado se hizo un análisis exploratorio de las variables cualitativas en análisis para el modelo planteado con tablas cruzadas o también denominadas tablas de contingencia. En esta fase aún no se hizo una eliminación de variables candidatas con pruebas de significancia chi-cuadrado dado que esto

se lo hará con la aplicación de la técnica de regresión logística (Logit) al analizar todo el conjunto de variables candidatas con los valores del estadístico z y p-value, donde $z > 2$ y $p\text{-value} < 0,05$.

Al realizar la tabla de contingencia a la variable cualitativa “gene” con la variable de incumplimiento “Default”, muestra que un total de 32 personas del género masculino y 21 personas del género “Femenino” son “Malos clientes” y 1166 personas del género masculino y 926 del género femenino son “Buenos Clientes”.

TABLA 9 TABLA DE CONTINGENCIA VARIABLE “GENE” * “DEFAULT”

gene	default		Total
	0	1	
Masculino	32	1,166	1,198
Femenino	21	926	947
Total	53	2,092	2,145

Fuente: Datos recopilados en la investigación

Elaborado por: Autor

En el caso de la variable “rang_ed” el mayor indicador de “Mal Cliente” lo registra el intervalo de persona que van desde los 18 hasta los 32 años y “Buen Cliente” el intervalo de personas que van desde los 33 a 47 años.

TABLA 10 TABLA DE CONTINGENCIA VARIABLE “RANG_ED” * “DEFAULT”

rang_ed	default		Total
	0	1	
De18a32años	22	641	663
De33a47años	20	788	808
De48a62años	10	497	507
De63omas	1	166	167
Total	53	2,092	2,145

Fuente: Datos recopilados en la investigación

Elaborado por: Autor

Al cruzar la variable cualitativa “est_civil” con la variable dependiente “Default” los resultados muestran que la mayor cantidad de personas con una alta morosidad catalogados como “Mal Cliente” está en los solteros y tienen una mayor probabilidad de incumplimiento, y para la definición de “Buen Cliente” con menor probabilidad de incumplimiento está en los socios prestamistas de Unión Libre y Casados.

TABLA 11 TABLA DE CONTINGENCIA VARIABLE “EST_CIVIL” * “DEFAULT”

est_civil	default		Total
	0	1	
Soltero	34	841	875
UnionLCasd	11	937	948
CasadoDisCony	2	78	80
Divorciado	5	181	186
Viudo	1	55	56
Total	53	2,092	2,145

Fuente: Datos recopilados en la investigación

Elaborado por: Autor

De la misma forma al realizar el cruce de la variable independiente “niv_est” con “default” podemos visualizar que arrojan resultados similares entre el valor de 0 y 1 y tanto un “Buen Cliente” como un “Mal Cliente”, ambos valores se concentran en personas con un nivel de estudios de secundaria seguido de prestamistas con educación técnica, superior y de cuarto nivel y las otras categorías, por lo que a simple vista se deduce que esta variable no discrimina el incumplimiento “Default”, por lo que esta variable bien podría ser desechada del modelo previo a la aplicación del modelo Logit de regresión logística.

TABLA 12 TABLA DE CONTINGENCIA VARIABLE “NIV_EST” * “DEFAULT”

niv_est	default		Total
	0	1	
sinest	1	38	39
prim	13	398	411
sec	20	920	940
tecEducSCuart	19	736	755
Total	53	2,092	2,145

Fuente: Datos recopilados en la investigación

Elaborado por: Autor

De la relación y cruce entre la variable “ocup” y “default” ocurre caso similar que con la variable “niv_est” donde esta variable no discrimina la variable dependiente “default” y la concentración de “Buen cliente” y “Mal Cliente” se registra en los mismos niveles.

TABLA 13 TABLA DE CONTINGENCIA VARIABLE “OCUP” * “DEFAULT”

ocup	default		Total
	0	1	
EstAmacJub	5	233	238
Asal	21	867	888
ComActInd	17	508	525
Transp	2	136	138
Const_Exp_Ind	4	144	148
Pagsp	4	204	208
Total	53	2,092	2,145

Fuente: Datos recopilados en la investigación
Elaborado por: Autor

De la tabla contingencia entre la variable independiente “tip_viv” y la variable dependiente “default” se denota que la definición de “Mal Cliente” se registra en mayor proporción en personas que viven con familiares, y para la definición de “Buen Cliente” y que tienen menor probabilidad de incumplimiento se registra en personas que tienen casa propia con hipoteca o sin hipoteca.

TABLA 14 TABLA DE CONTINGENCIA VARIABLE “TIP_VIV” * “DEFAULT”

tip_viv	default		Total
	0	1	
Prestada	1	7	8
VivFam	23	861	884
Arrend	13	241	254
PropHyNoH	16	983	999
Total	53	2,092	2,145

Fuente: Datos recopilados en la investigación
Elaborado por: Autor

La tabla cruzada de la variable independiente “tip_gar” no discrimina la variable dependiente “default” y sus resultados se presentan en igual proporción entre el valor de 0 y 1.

TABLA 15 TABLA DE CONTINGENCIA VARIABLE “TIP_GAR” * “DEFAULT”

tip_gar	default		Total
	0	1	
Personal	37	1,244	1,281
Real	16	848	864
Total	53	2,092	2,145

Fuente: Datos recopilados en la investigación
Elaborado por: Autor

3.3. Presentación de resultados y discusión

Interpretación del modelo Logit

La técnica de minería de datos usada en la presente investigación para la medición del riesgo de crédito “Default” fue de tipo predictivo (regresión logística) con la aplicación del modelo Logit con el ánimo de medir las probabilidades e identificar los comportamientos que inciden en la problemática planteada.

Una vez descritas las variables que se utilizarán en el presente trabajo de investigación, tomando en cuenta el modelo econométrico Logit mencionado con anterioridad y utilizando la herramienta estadística y de ciencia de datos Stata, se procede a estimar un modelo, donde las variables independientes ya expuestas, traten de explicar el comportamiento de la variable dependiente dicotómica y/o binaria 0 y 1, que en este caso es el Riesgo Crediticio o Default (Buenos y Malos clientes).

La Tabla No. 16 muestra la descripción del modelo con todas las variables consideradas con sus determinados valores: coeficientes, errores estándar, valores estadísticos z, valores de probabilidad p-value y los intervalos de confianza, que en este caso está dado por el 95%. Así también, de manera

general, en la parte superior derecha, se expone el número de observaciones, valores de chi2 y su probabilidad, y Pseudo R2 que es un porcentaje que trata de explicar las variaciones que tienen las variables independientes sobre la variable dependiente.

TABLA 16 MODELO CON VARIABLES ANALIZADAS

Logistic regression	Number of obs	=	2145
	LR chi2(21)	=	35.63
	Prob > chi2	=	0.0240
Log likelihood = -230.65626	Pseudo R2	=	0.0717

default	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
1.gene	.2752502	.3056977	0.90	0.368	-.3239063	.8744067
rang_ed						
1	-.0084143	.3465557	-0.02	0.981	-.6876509	.6708223
2	-.1775308	.4966905	-0.36	0.721	-1.151026	.7959648
3	1.240159	1.095364	1.13	0.258	-.9067143	3.387033
est_civil						
1	.9619406	.401626	2.40	0.017	.1747681	1.749113
2	.3048295	.7742155	0.39	0.694	-1.212605	1.822264
3	.3524983	.5328423	0.66	0.508	-.6918534	1.39685
4	.3730648	1.093959	0.34	0.733	-1.771055	2.517184
niv_est						
1	.131633	1.072156	0.12	0.902	-1.969754	2.23302
2	.7970381	1.063427	0.75	0.454	-1.287241	2.881317
3	.4445961	1.073566	0.41	0.679	-1.659555	2.548748
ocup						
1	.1772976	.5437112	0.33	0.744	-.8883567	1.242952
2	-.1493742	.5430247	-0.28	0.783	-1.213683	.9149346
3	.6538761	.8780585	0.74	0.456	-1.067087	2.374839
4	.2885857	.7184709	0.40	0.688	-1.119591	1.696763
5	.4231897	.6980926	0.61	0.544	-.9450467	1.791426
ing_net	.0007524	.000378	1.99	0.047	.0000115	.0014932
tip_viv						
1	1.744409	1.116926	1.56	0.118	-.4447254	3.933543
2	.9572441	1.131903	0.85	0.398	-1.261245	3.175734
3	1.766168	1.144109	1.54	0.123	-.4762434	4.00858
1.tip_gar	.3708418	.3248788	1.14	0.254	-.2659089	1.007592
_cons	.2676251	1.626727	0.16	0.869	-2.920702	3.455952

Fuente: Datos recopilados en la investigación
Elaborado por: Autor

Para verificar que cada variable independiente sea estadísticamente significativa, es decir, puedan predecir el riesgo crediticio o default, se considera analizar para cada variable, los valores del estadístico z y p-value, donde $z > 2$ y $p\text{-value} < 0,05$.

Evaluando cada variable según los criterios antes mencionados, se procede a eliminar una a una las variables que no son significativas para el modelo econométrico (Anexo No. 4). Del número total de variables independientes consideradas para estimar el modelo econométrico default, las únicas que resultaron estadísticamente significativas fueron: estado civil considerando la categoría “Unión Libre y Casados” y el tipo de vivienda en sus categorías “Vive con familiares” y “Casa propia Hipotecada y no Hipotecada”, expuestas en la Tabla No. 17.

TABLA 17 MODELO PREDICTIVO LOGIT

Logistic regression	Number of obs	=	2145
	LR chi2(3)	=	20.02
	Prob > chi2	=	0.0002
Log likelihood = -238.46092	Pseudo R2	=	0.0403

default	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
1.est_civil	1.031788	.3570559	2.89	0.004	.3319713	1.731605
tip_viv						
1	.8252616	.3482561	2.37	0.018	.1426921	1.507831
3	.9812219	.3817025	2.57	0.010	.2330987	1.729345
_cons	2.612803	.2825802	9.25	0.000	2.058956	3.16665

Fuente: Datos recopilados en la investigación

Elaborado por: Autor

Para evaluar el modelo en conjunto, se toma en consideración la Probabilidad de chi2, donde esta no supere a 0,05, figurando que las variables independientes están asociadas a la variable dependiente y estas son diferentes de cero. Para el presente caso de estudio, esta probabilidad es menor al valor determinado, por lo que los datos se ajustan al modelo.

Por otro lado, basándose en Pseudo R2, el 4% de las variaciones de la variable default puede ser explicada por las variaciones de las variables estado civil y tipo de vivienda. En este caso el porcentaje es muy bajo, esto puede deberse a que se necesitaría variables más significativas para el modelo considerando que el riesgo "Default" asocia múltiples factores socio económicos del cliente e internos que incluso la institución no podría estar considerando dentro de sus bases de datos, es decir; variables que pueda explicar o predecir la variable default con mayor aceptación.

Analizando los coeficientes y las relaciones de cada una de las variables independientes con la variable default, se tiene que sin considerar el estado civil y el tipo de vivienda que tiene el cliente, en promedio, existe una probabilidad de aproximadamente 2,61% de que existan buenos clientes, es decir, clientes que mantengan en su buró crediticio una calificación A. Así también, si el cliente tiene como estado civil, ya sea unión libre o casado, manteniendo constante las demás variables, en promedio, la razón de probabilidad es de aproximadamente 1%, estableciendo una relación positiva entre el estado civil y la variable default, lo que nos da a entender que existe una mayor probabilidad que el cliente mantenga una calificación A en su buró crediticio si mantiene una unión libre o es casado. Finalmente, en cuanto al tipo de vivienda, si los clientes viven con familiares o tienen una casa propia hipotecada o no hipotecada, manteniendo constante las demás variables, en promedio, la razón de probabilidad es de aproximadamente 0,83% y 0,98% respectivamente, indicando una relación positiva entre el tipo de vivienda y la variable default, es decir, existe una probabilidad mayor de que en estas situaciones los clientes mantengan una buena calificación en su buró crediticio.

Evaluación de la bondad del modelo

Otra manera de determinar si los valores ingresados en el modelo están correctamente clasificados, es utilizando la siguiente tabla (Véase Tabla No. 18) en la que se describe un porcentaje que trata de explicar que tan buen especificado está el modelo propuesto.

TABLA 18 LOGISTIC MODEL FOR DEFAULT

Logistic model for default

Classified	True		Total
	D	~D	
+	2092	53	2145
-	0	0	0
Total	2092	53	2145

Classified + if predicted $\Pr(D) \geq .5$

True D defined as default $\neq 0$

Sensitivity	$\Pr(+ D)$	100.00%
Specificity	$\Pr(- \sim D)$	0.00%
Positive predictive value	$\Pr(D +)$	97.53%
Negative predictive value	$\Pr(\sim D -)$.%
False + rate for true ~D	$\Pr(+ \sim D)$	100.00%
False - rate for true D	$\Pr(- D)$	0.00%
False + rate for classified +	$\Pr(\sim D +)$	2.47%
False - rate for classified -	$\Pr(D -)$.%
Correctly classified		97.53%

Fuente: Datos recopilados en la investigación

Elaborado por: Autor

Para este caso de estudio, tal como lo muestra la Tabla No. 18 el modelo econométrico logit de riesgo crediticio prevé el 97,53% de las observaciones correctamente.

Test de Hosmer y Lemeshow

Una prueba para estimar la bondad de ajuste de un modelo logit, es el Test de Hosmer-Lemeshow, el cual evalúa si las frecuencias observadas coinciden con las esperadas en una distribución de datos de un modelo logit, determinando que, si la probabilidad de este test es mayor al 5%, el modelo tiene un buen ajuste, de lo contrario, no puede ser un modelo estadísticamente significativo y no puede ser aceptado.

TABLA 19 TEST DE HOSMER Y LEMESHOW

<u>Logistic model for default, goodness-of-fit test</u>	
(Table collapsed on quantiles of estimated probabilities)	
(There are only 5 distinct quantiles because of ties)	
number of observations =	2145
number of groups =	5
Hosmer-Lemeshow chi2(3) =	1.20
Prob > chi2 =	0.7530

Fuente: Datos recopilados en la investigación

Elaborado por: Autor

El test de Hosmer-Lemeshow realizado al modelo econométrico propuesto indica que existe un buen ajuste de los datos en el modelo, es decir, es un modelo que puede ser aceptado en cuanto cumple con los parámetros establecidos, puesto que la probabilidad es de 75,30%, la cual supera al 5%.

Adicional a las pruebas de bondad realizadas para probar el modelo predictivo obtenido, se solicitó el criterio de dos expertos quienes hicieron una revisión visual del procedimiento aplicado y resultados obtenidos (Véase Anexo 5).

CONCLUSIONES

- El sector cooperativo como quedó demostrado en el presente trabajo se enfrenta a uno de sus principales riesgos, el de crédito, y un modelo logit por su gran poder analítico y predictivo sin desplazar el trabajo del analista de crédito favorece a estas instituciones disminuyendo la probabilidad de pérdidas futuras.
- La cartera de consumo corresponde al segmento de créditos concedidos a personas naturales, al mes de abril de 2019 del total de la cartera de la Cooperativa Ltda. la mayor concentración se registra en consumo prioritario, donde se evidencia calificaciones que van desde la A hasta la E, a mayor riesgo incrementará la probabilidad de pérdidas esperadas. La calidad de la cartera total presenta variaciones significativas, pero tendencia positiva en los últimos años, en un escenario de un año del segmento de consumo al cierre del año 2018 se registra un deterioro del 0.29%.
- La metodología logit se asocia íntimamente a un scoring y permite un análisis particular de las variables asociadas al riesgo default (probabilidad de que el cliente no cumpla con sus obligaciones total o parcialmente), sin embargo; se requiere gran cantidad de información actualizada y confiable del prestatario a fin de realizar una predicción más exacta.
- La construcción de un modelo de evaluación de riesgo crediticio bajo la metodología logit con la aplicación de técnicas de minería de regresión logística requiere de una variable dependiente con la categorización entre buenos y malos clientes, y la calificación interna cumple un papel vital para esta clasificación.
- Los resultados del modelo dependen de la situación de la institución analizada y del entorno socioeconómico de sus clientes, en este modelo se puede incluir la mayor cantidad de variables cualitativas o cuantitativas que se deseen analizar con el ánimo de establecer una mejor relación predictiva en armonía con lo que se pretende conocer con exactitud. Mucho dependerá de lo que se busca con exactitud, puede incluir variables del cliente (perfil), variables de la tecnología crediticia de

cada entidad en sus distintas fases todos dependerá de los criterios con lo que se desee evaluar al sujeto de crédito.

- El proceso de construcción de un modelo logit para la evaluación de riesgo crediticio con ciencia de datos y técnicas predictivas involucra el entendimiento y comprensión general de un proceso de minería de datos dentro del conocido y gran KDD (Knowledge Discovery in Databases), dado que se requiere la limpieza y preprocesamiento de datos y la conversión de atributos continuos a categóricos donde se requiera para un mejor análisis y aplicación de técnicas predictivas y de ciencia de datos.
- La técnica de regresión logística (Logit) seleccionada resulta ser una gran herramienta estadística y predictiva para la evaluación del riesgo de crédito y es que este tipo de técnicas es usado cuando se requiere el análisis y predicción de una variable binaria o dummy como en nuestro caso 0 – Mal cliente y 1 – Buen cliente con la determinación de p de variables explicativas o aleatorias de tipo cualitativo y cuantitativo y si a estos resultados y scoring de incumplimiento obtenido por el modelo logit se le establece puntos de corte para todos los prestatarios de la base de datos dicha información puede servirle a la institución para aprobar o rechazar un cliente.
- Para la aplicación de un modelo Logit así mismo es necesario realizar previamente un análisis del comportamiento individual de las variables analizadas que puede basarse en una técnica estadística y de minería de datos de análisis exploratorio, para determinar las relaciones existentes con la variable dependiente, en este caso pues se ha optado por las matrices de contingencia o tablas cruzadas, esto también se conoce como análisis bivariante.
- Así mismo en la construcción de estos modelos se hace vital la aplicación de pruebas de bondad a fin de medir la aplicabilidad del modelo bajo coeficientes de correlación en los datos muestrales.
- Las relaciones de tablas cruzadas a la variable “gene” demuestra que el género masculino en la institución tiene una mayor probabilidad de incumplimiento “default”, sin embargo; esto también se asocia a que la

institución en el segmento de consumo registra un 55.85% de hombres frente a un 44.50% de mujeres.

- Variables como el nivel de estudio (tip_viv), ocupación (ocup), tipo de garantía (tip_gar), realmente no aportan ni discriminan la variable “default” y en la definición de “Buen Cliente” y “Mal Cliente”, no siendo relevantes dentro del modelo.
- El análisis con la técnica de regresión logística de Logit arroja que las variables estadísticamente significativas son: estado civil considerando la categoría “Unión Libre y Casados” y el tipo de vivienda en sus categorías “Vive con familiares” y “Casa propia Hipotecada y no Hipotecada. Los coeficientes y relaciones muestran que sin considerar el estado civil y el tipo de vivienda que tiene el cliente, en promedio, existe una probabilidad de aproximadamente 2,61% de que existan buenos clientes, si el cliente tiene como estado civil, ya sea unión libre o casado, manteniendo constante las demás variables se establece una relación positiva marcado una mayor probabilidad de ser buenos clientes.
- Si los clientes viven con familiares o tienen una casa propia hipotecada o no hipotecada, manteniendo constante las demás variables, en promedio, la razón de probabilidad es de aproximadamente 0,83% y 0,98% respectivamente, indicando una relación positiva entre el tipo de vivienda y la variable default.
- El criterio de Hosmer-Lemehov arroja un valor de 75.30% que concluye un buen ajuste del modelo obtenido dándose por aceptado. No obstante, se obtuvo un bajo nivel predictivo de Pseudo R2 del 4%, recordemos que un modelo para la evaluación de riesgo se debe básicamente a los datos con los que cuenta la institución analizada y el resultado y el modelo puede ser otro si se realiza el mismo análisis en otra institución, mientras mayor sea la información disponible mejor será el índice de predicción.
- La presente investigación deja el interés para seguir estudiando el beneficio de las técnicas y herramientas de minería de datos en la evaluación de riesgo de crédito con técnicas más sofisticadas de predicción.

RECOMENDACIONES

- Recomendar la aplicación de la minería de datos predictivas para disminuir la probabilidad de pérdidas esperadas respecto a la calidad de la cartera y las variaciones de morosidad dado que estos podrían mejorarse con la aplicación de modelos internos de evaluación de riesgo de crédito acogiendo las recomendaciones de Basilea con ayuda de técnicas y herramientas con capacidad de predicción.
- Que las instituciones financieras entre ellas las Cooperativas de Ahorro y Crédito capture en sus bases de datos la mayor cantidad de variables asociadas al cliente (tecnologías de crédito adecuadas), con el ánimo de obtener patrones y/o comportamientos de medición a la variable “Default”, pues mientras mayor sea la información disponible mejor será la predicción que arroje la construcción del modelo.
- La Cooperativas de Ahorro y Crédito pueden optar por la conformación de bases de datos únicas que sirvan para la construcción de modelos de evaluación de riesgo de crédito, dado que dentro del proceso KDD y construcción de un modelo, la etapa más compleja es precisamente la preparación de los datos que será tratada con la aplicación de técnicas de minería.
- La Cooperativa de Ahorro y Crédito pueden elaborar un modelo de evaluación de riesgo crédito para cada uno de los segmentos a fin de obtener los comportamientos propios de cada cartera, con la ayuda de técnicas de minería de datos predictivas, descriptivas, entre otras.
- Recomendar la aplicación de un modelo de regresión logística logit para Cooperativas de Ahorro y Crédito, dado que aporta en la predicción de riesgo de crédito por ser de tipo no lineal y permite obtener las probabilidades de incumplimiento; y que si a este modelo obtenido de la metodología logit se agrega puntos de corte resultará un scoring de utilidad para la institución para la aprobación o rechazo de clientes.

REFERENCIAS BIBLIOGRÁFICAS

- Aboobyda, J. H. y Tarig, M. A. (2016). Developing Prediction Model of Loan Risk in Banks Using Data Mining. *Machine Learning and Applications: An International Journal (MLAIJ)*. Vol.3, No.1, March 2016. Pag. 1-9. Recuperado de: <http://aircconline.com/mlaij/V3N1/3116mlaij01.pdf>
- Angelini, E., Tollo, G., y Roli, A., 2008. A neural network approach for credit risk evaluation. *Q. Rev. Econ. Finance* 48, 733–755
- Aranda, Yadira Robles, & Sotolongo, Anthony R.. (2013). Integración de los algoritmos de minería de datos 1R, PRISM e ID3 a PostgreSQL. *JISTEM - Journal of Information Systems and Technology Management*, 10(2), 389-406. <https://dx.doi.org/10.4301/S1807-17752013000200012>
- Arrien, I. (2018). Sector Financiero en Ecuador. *ICEX: España exportaciones e inversiones*, 1-9.
- Apale, Y. & Olivares, B. & Rodriguez, L. & Alor, G. & Muñoz, H. (2016). Propuesta de arquitectura para un módulo de inteligencia de negocios basado en minería de datos. Instituto Tecnológico de Orizaba. ISSN 1870-4069. México.
- Aparicio, C. & Moreno, H. (2011). Calidad de la cartera crediticia bancaria y el ciclo económico: una mirada al gasto en provisiones bancarias en el Perú (2001-2011). Superintendencia de Banca, Seguros y Administradoras Privadas de Fondos de Pensiones. Enlace disponible en: http://www.sbs.gob.pe/Portals/0/jer/ddt_ano2011/3_Aparicio_y_Moreno_2011.pdf
- Asociación de Supervisores Bancarios de las Américas, ASBA. (2008). Sanas prácticas de gestión y supervisión del riesgo de crédito en las Américas. Enlace disponible en: <http://www.asbasupervision.com/es/bibl/i-publicaciones-asba/i-1-grupos-de-trabajo/772-gt06-1/file>
- Author, A. Title of the paper. In: *European Financial Systems (2017). Proceedings of the 14th International Scientific Conference*, Brno: Masaryk University, 2017, part 2, pp. 408. . ISBN 978-80-210-8609-8, ISBN 978-80-210-8610-4 (online : pdf).
- Ayala, L. & Sarmiento, L., (2015). Modelo Logístico y Discriminante para el otorgamiento de un CrediAhorro, innovación financiera.
- Azevedo, A. y Santos, M. (2008). KDD, SEMMA and CRISP-DM: A parallel overview. IADIS European Conference Data Mining.
- Banco Central del Ecuador, 2015. Resolución No. 043-2015-F. La Junta de Política y Regulación Monetaria y Financiera.

- Banco Central del Ecuador, 2018. Cuestiones económicas. Vol. 28, Nº 2, 2018. ISSN: 2477-9059. Quito. Ecuador.
- Banco de Pagos Internacionales BIS, 2001. Visión General del Nuevo Acuerdo de Capital de Basilea, Documento Consultivo.
- Banco de Pagos Internacionales BIS, 2006. Metodología de los Principios básicos.
- Banco de Pagos Internacionales BIS, 2006. Principios Básicos para una supervisión bancaria eficaz.
- Banco de Pagos Internacionales BIS, 2011, Principios Básicos para una supervisión bancaria eficaz, Obtenido de: https://www.bis.org/publ/bcbs213_es.pdf
- Banco de Pagos Internacionales, 2017. Comité de Supervisión Bancaria de Basilea. Finalización de Basilea III. Enlace disponible en: https://www.bis.org/bcbs/publ/d424_inbrief_es.pdf
- Basilea. (1999). Principios para La Administración Del Riesgo De Crédito, Pag. 36.
- Berlanga, A. (2016). El camino desde la inteligencia artificial al Big Data.
- Berlanga, V.; Rubio, M.; Vilá, Ruth. (2013). Cómo aplicar árboles de decisión en SPSS. Universitat de Barcelona. Revista d'Innovació i Reserca en Educació.
- Bernstein, P. A.; Haas, L. M. (2008). Bernstein & Haas, 2008. "Information integration in the Enterprise". Communications of the ACM, Volume 51 Issue 9, pp: 72-79. ACM 2008.
- Betancur, D. (sf). Modelo basado en agentes para las etapas de recopilación e integración de datos en el proceso de KDD. Universidad Nacional de Colombia. Medellín.
- Bonilla, J., Crespo, C., & León, Y. (2018). Determinación del riesgo de una incorrecta evaluación del entorno económico en el análisis crediticio de una institución financiera en Ecuador. Primera edición Mayo 2018. ISBN 978-9942-770-96-7.
- Boqué, R. & Maroto, A. sf. El Análisis de la varianza ANOVA. Grupo de Quimiometría y Cualimetría. Universitat Rovira i Virgili. Pl. Imperial Tàrraco, 1. 43005-Tarragona. Enlace disponible en: <http://www.quimica.urv.cat/quimio/general/anovacast.pdf>
- Bouza, C. & Santiago, A., 2012. La Minería de Datos: Árboles de Decisión y su aplicación en estudios médicos. Modelación Matemática de Fenómenos del Medio Ambiente y la Salud. Universidad de la Habana Cuba & Universidad Autónoma de Guerrero México.
- Bramer, M. (2007). Date for data mining. En Principles of data mining. London: Springer, 2007, pp. 11-20.

- Caicedo Cerezo, Edinson, Mercè Claramunt Bielsa, M, & Casanovas Ramón, Monserrat. (2011). Medición del riesgo de crédito mediante modelos estructurales: una aplicación al mercado colombiano. *Cuadernos de Administración*, 24(42), 73-100. Retrieved February 20, 2019, from http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S0120-35922011000100004&lng=en&tlng=es.
- Cardona, L., Franco, I. y Ossa, J. (2015). Modelo de valoración ajustado a la evolución del riesgo crediticio para carteras colectivas que invierten en títulos valores no tradicionales en la comisionista global Securities Colombia de la Ciudad de Medellín. Especialización en Riesgos Financieros. Recuperado de: <http://hdl.handle.net/11407/2173>
- Cardona, P., (2004). Aplicación de árboles de decisión en modelos de riesgo crediticio. *Revista Colombiana de Estadística*. Volumen 27 No. 2 Págs. 139-151.
- Cardona, Z. (2006). La diversificación del riesgo en la cartera de créditos del sector financiero con base en la teoría de portafolios. *AD-MINISTER*, N. 9, julio-diciembre, 2006.
- Carlos Aparicio¹ y Hesione Moreno (2011). Calidad de la cartera crediticia bancaria y el ciclo económico: una mirada al gasto en provisiones bancarias en el Perú (2001-2011). Superintendencia de Banca, Seguros y Administradoras Privadas de Fondos de Pensiones. Obtenido de: http://www.sbs.gob.pe/Portals/0/jer/ddt_ano2011/3_Aparicio_y_Moreno_2011.pdf
- Constitución de la República del Ecuador (2008). Asamblea Constituyente.
- Coraggio, J.L. (2011). *Economía Social y Solidaria: El trabajo antes que el capital*. Quito: Ediciones Abya-Yala.
- COSEDE, 2015. Resolución No. 038-2015-F. Junta de Regulación Monetaria y Financiera.
- Código Orgánico Monetario y Financiero (2014). Asamblea Nacional. República del Ecuador.
- Chorafas, D. (2000). *Managing credit risk, analyzing rating and pricing the probability of de-fault*. Londres: Euromoney Institutional Investor PLC.
- Chavarín Rodríguez, Rubén. (2015). "Morosidad en el pago de créditos y rentabilidad de la banca comercial en México". *Revista mexicana de economía y finanzas*, 10(1), 73-85. Recuperado en 30 de mayo de 2019, de http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S1665-53462015000100073&lng=es&tlng=pt.
- Cortez, D. & Burgos, J. (2016). La gestión de cartera de crédito y el riesgo crediticio como determinante de morosidad o liquidez de las empresas comerciales.

- Revista Observatorio de la Economía Latinoamericana, Ecuador, (Septiembre 2016).
- De la Torre, A., & Alain, I., & Schmukler, S (2012). El Desarrollo Financiero en América Latina y El Caribe. El Camino por delante. Estudios del Banco Mundial sobre América Latina y El Caribe. The World Bank.Washington, D.C.
- De Battista, A. & Cristaldo, P. & Ramos, L. & Nuñez, J. & Retamar, S. & Bouzenard, D. (2016). Minería de Datos aplicada a datos masivos. Universidad Tecnológica Nacional. Argentina.
- Díaz y Quevedo, (2009). Determinantes del ratio de morosidad en el sistema financiero Boliviano. Banco Central de Bolivia.
- Díaz, C. & Del Valle Guerra, Y. (2017). Financial Risk in Credits to the Consumption of the Venezuelan Banking System 2008-2015. Revista Científica Electrónica de Ciencias Humanas. Scientific e-journal of Human Sciences. PPX200502ZU1935 / ISSN 1856-1594 / By Fundación Unamuno.
- Enterprise Business Intelligence EBI, (2005): Strategies and Technologies for Deploying BI on an Enterprise Scale, Wayne W. Eckerson y Cindi Howson, TDWI Report Series, Agosto 2005.
- Ernst & Young, EY (2014). Big Data en el Sector Financiero Español. EY Building a better working world. Enlace disponible en: [http://www.ey.com/Publication/vwLUAssets/EY-big-data-en-el-sector-financiero-espanol/\\$FILE/EY-big-data-en-el-sector-financiero-espanol.pdf](http://www.ey.com/Publication/vwLUAssets/EY-big-data-en-el-sector-financiero-espanol/$FILE/EY-big-data-en-el-sector-financiero-espanol.pdf)
- E.J. Hernández-Leal, N.D. Duque-Méndez y J. Moreno-Cadavid (2017). “Big Data: una exploración de investigaciones, tecnologías y casos de aplicación”, Tecnológicas, vol. 20, no. 39, mayo - agosto, 2017.
- Fayyad, Usama, Grinstein, George & Wierse, Andreas. Information visualization in data mining and knowledge discovery. San Francisco: Academic Press, 2002. 373p.
- FELABAN, 2003. Basilea II: Hacia un nuevo esquema de medición de riesgos. Superintendencia de Entidades Financieras y Cambiarias Gerencia de Análisis del Sistema. Boletín. Diciembre 2003. Enlace disponible en: https://www.felaban.net/archivos_boletines_clain/archivo20140722153358PM.pdf.
- Fernández Castaño, Horacio, Pérez Ramírez, Fredy Ocaris (2005), El modelo logístico: una herramienta estadística para evaluar el riesgo de crédito. Revista Ingenierías Universidad de Medellín [en línea] 2005, 4 (enero-junio) Disponible en:<<http://www.redalyc.org/articulo.oa?id=75040605>> ISSN 1692-3324
- García, J. & Molina, J. (2006). *Técnicas de Análisis de Datos*. Universidad Carlos III de Madrid.

- García, M. & Sánchez, C. (2005). Riesgo de Crédito en México: aplicación del Modelo CreditMetrics. Universidad de las Américas Puebla. Págs. 44-74. Recuperado de: http://catarina.udlap.mx/u_dl_a/tales/documentos/laex/garcia_s_m/
- García, S. & Ramírez, S. & Luengo, J. & Herrera, F. (2016). Big data: Preprocesamiento y calidad de datos. Universidad Granada, España.
- García, J. & Salazar, P. (2005). Métodos de Administración y Evaluación de Riesgos. Universidad de Chile. Enlace disponible en: http://repositorio.uchile.cl/tesis/uchile/2005/garcia_j2/sources/garcia_j2.pdf
- Gallego, J. & Navarro, L. & Castillo, L. (2015). Aplicación de técnicas de minería de datos en Atención Primaria en Salud (APS) para el análisis de riesgos en mujeres gestantes de la población manizaleña atendida por Assbasalud. *Revista Biosalud* 2015; 14(2): 71-78. DOI: 10.17151/biosa.2015.14.2.7
- Gil, Y. (2017). Evaluación del riesgo crediticio en entidades bancarias en el área de microfinanzas utilizando árboles de decisión. Tesis de grado. Universidad Santo Tomas. Bucaramanga.
- Giraldo, J. & Vargas, F (2017). Aplicación de la técnica regresión logística de la minería de datos en el proceso de descubrimiento de conocimiento (KDD) en bases de datos operativas o transaccionales. *Perspectiv@s. Revista de Tecnología e Información* 2017 | Año 14 | N° 13 | 51-55
- González, M. & Soto, C. 2013. Minería de datos para series temporales. Universidad Carlos III de Madrid. Universidad Central "Marta Abreu" de las Villas. Pag. 13-32. Doi: 10.13140/RG.2.1.2571.9841.
- Guillén Uyen, J. (2001). Morosidad crediticia y tamaño. Un análisis de la crisis bancaria peruana. Obtenido de Banco Central de Reserva del Perú: <http://www.bcrp.gob.pe/docs/Publicaciones/Documentos-de-Trabajo/2001/DocumentoTrabajo-05-2001.pdf>
- Gutiérrez, M. (2007). Credit scoring models: what, how, when and for what purposes. Banco Central de la República Argentina. Munich Personal RePEc Archive. October 2007. Págs. 1-32. Recuperado de: https://mpra.ub.uni-muenchen.de/16377/1/MPRA_paper_16377.pdf
- Haaf, H. & Tasche, D., (2002). Calculating Value-at-Risk contributions in CreditRisk+*. Pag. 1-11. <http://arxiv.org/abs/cond-mat/0112045>, 2000.
- Han, J. y Kamber, M. (2006). *Data mining: Concepts and techniques*. 7th. ed. Morgan Kaufmann.
- Hand D. y Henley W., (1997), *Statistical Classification Methods in Customer Credit Scoring: A review*, *Journal of the Royal Statistical Association Series A*. Vol. 160

- Hand, D. J. and Henley, W. E (1997). "Statistical Classification Methods in Consumer Credit Scoring: a Review". Royal Statistical Society, 160 (1997), Part 3, Págs. 523-541.
- Hernández et al. (2004). Hernández, J.; Ramírez, M.J.; Ferri, C. (2004). "Introducción a la Minería de Datos". Prentice Hall - Pearson Education (Madrid, España) 2004. ISBN: 84-205-4091-9
- Hernández , E. & Duque, N. & Moreno, J. (2017). Big Data: una exploración de investigaciones, tecnologías y casos de aplicación. TecnoLógicas, vol. 20, no. 39, mayo - agosto, 2017.
- Hussain, B. (2014). Credit risk assessment model for Jordanian commercial banks: Neural scoring approach. El Sevier. Pag. 1-9. Disponible en www.sciencedirect.com.
- Jeres, E. (2018). "Sistema de mejora continua y su incidencia en los procesos crediticios del sector cooperativo perteneciente al segmento 2, del cantón Ambato.". Tesis de pregrado. Universidad Técnica de Ambato. Ambato.
- Junta de Regulación Monetaria Financiera - JRMF (2019). Codificación de Resoluciones Monetarias, Financieras, de Valores y Seguros. Libro I: Sistema Monetario y Financiero.
- Kotu, V. & Deshpande, B. (2015). Predictive analytics and data mining. Concepts and practice with rapidMiner. Análisis predictivo y minería de datos. Conceptos y práctica con rapidMiner. Waltham: Morgan Kaufmann.
- Lanzarini, L., Haspeue, W., Villa, A., Basgall, M., Molina, R., Rojas, L., Corvi, J., Jimbo, P., Fernández, A., Puente, C., Olivas, J. (2018). Minería de Datos y Big Data. Aplicaciones en riesgo crediticio, salud y análisis de mercado. RedUNCI - UNNE - ISBN 978-987-3619-27-4.
- Laudon, K. & Laudon, J. (2012). *Sistemas de Información Gerencial*. Décimosegunda edición. México. Pearson Educación.
- Leal, A., Aranguiz, M. & Gallegos, J. (2018) Análisis de riesgo crediticio, propuesta del modelo Credit Scoring. En: Revista de la Facultad de Ciencias Económica: Investigación y Reflexión. rev.fac.cienc.econ, XXVI (1), DOI: <https://doi.org/10.18359/rfce.2666>
- Ludovic, A. & Aranguiz, M. & Gallegos, M. (2018). Análisis de riesgo crediticio, propuesta del modelo credit scoring. Revista de la Facultad de Ciencias Económicas: Investigación y Reflexión, XXVI(1), undefined-undefined. [fecha de Consulta 01 de mayo de 2019]. ISSN: 0121-6805. Disponible en: <https://www.redalyc.org/articulo.oa?id=909/90953767010>

- Malberti, M. & Beguerí, G. 2015. Reglas de Asociación con los datos de una biblioteca universitaria. Association Rules for university library data. Revista Cubana de Ciencias Informáticas. Vol. 9, No. 4, Octubre-Diciembre, 2015. ISSN: 2227-1899 | RNPS: 2301. Pág. 30-45. Disponible en: <http://scielo.sld.cu/pdf/rcci/v9n4/rcci03415.pdf>
- Marcano Aular, Yelitza Josefina, & Talavera Pereira, Rosalba. (2007). Minería de Datos como soporte a la toma de decisiones empresariales. Opción, 23(52), 104-118. Disponible en: http://www.scielo.org.ve/scielo.php?script=sci_arttext&pid=S1012-15872007000100008&lng=es&tlng=es.
- Martínez C., Ó. A., y Ballón F., R. A. (2010). MODELO KMW - MERTON PARA LA MEDICION DEL RIESGO CREDITICIO DE LAS RESERVAS INTERNACIONALES DEL BANCO CENTRAL DE BOLIVIA. Revista de Análisis Del Banco Central de Bolivia. Recuperado de: http://www.revistasbolivianas.org.bo/scielo.php?pid=S2304-88752010000100006&script=sci_arttext
- Millán, J. & Cerezo, E. (2018): Modelos para otorgamiento y seguimiento en la gestión del riesgo de crédito, Revista de Métodos Cuantitativos para la Economía y la Empresa, ISSN 1886-516X, Universidad Pablo de Olavide, Sevilla, Vol. 25, pp. 23-41
- Molina J. y García, J. (2006). Técnicas de análisis de datos. Disponible en: <http://ocw.uc3m.es/ingenieria-informatica/analisis-de-datos/libroDataMiningv5.pdf> (Libro on-line).
- Moreno, M. y López, V. s.f. Uso de técnicas no Supervisadas en la Construcción de Modelos de Clasificación en Ingeniería del Software. Departamento de Informática y Automática. Universidad de Salamanca.
- Moreno, M.; Miguel, L.; García, F.; Polo, J. (2004). Aplicación de técnicas de minería de datos en la construcción y validación de modelos predictivos y asociativos a partir de especificaciones de requisitos de software. Universidad de Salamanca.
- Moreno, S. (2013). El Modelo Logit Mixto para la construcción de un Scoring de Crédito. Universidad Nacional de Colombia . Medellín, Colombia.
- Munafo, F., (2018). Aplicación del Modelo de Merton utilizando VBA. Revista de Investigación en Modelos Financieros Año 7 Volumen I (2018-I). Universidad de Buenos Aires. ISSN: 250-687X. Enlace disponible en: <http://www.economicas.uba.ar/wp-content/uploads/2017/11/Munafo-F.-Aplicaci%C3%B3n-del-modelo-de-Merton-utilizando-VBA.pdf>
- Novoa, M. (2016). Minería de datos. Aplicaciones de técnicas descriptivas. Escuela Técnica Superior de Ingeniería Universidad de Sevilla. Sevilla.

- Ochoa, J.C., Galeano, W., & Agudelo, L.G. (2010). Construcción de un modelo de scoring para el otorgamiento de crédito en una entidad financiera. *Perfil de Coyuntura Económica*, 16, 191-222. Universidad de Antioquia.
- Olano, E. (2018). Implementación del Acuerdo de Basilea III en la Regulación Bancaria del Perú. Tesis de grado. Universidad Católica Santo Toribio de Mogrovejo. Perú.
- Pantoja, P. (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. Perú.
- Pereira da Silva, H. (2016). Aplicación de modelos bayesianos para estimar la prevalencia de enfermedad y la sensibilidad y especificidad de tests de diagnóstico clínico sin gold standard. Tesis doctoral. Universitat de Barcelona. Barcelona. Enlace disponible en: http://diposit.ub.edu/dspace/bitstream/2445/122049/1/HDPdS_TESIS.pdf
- Pérez, F., & Fernández, H. (2007). Las redes neuronales y la evaluación del riesgo de crédito. *Revista Ingenierías Universidad de Medellín*, 6(10), 77-91. Disponible en: http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S1692-33242007000100007&lng=en&tlng=es.
- Pérez, F. & Fernández, H. (2007). LAS REDES NEURONALES Y LA EVALUACIÓN DEL RIESGO DE CRÉDITO. *Revista Ingenierías Universidad de Medellín*, 6(10), 77-91. Retrieved January 06, 2019, from http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S1692-33242007000100007&lng=en&tlng=es.
- Pérez Ramírez, Fredy Ocaris, & Fernández Castaño, Horacio. (2007). LAS REDES NEURONALES Y LA EVALUACIÓN DEL RIESGO DE CRÉDITO. *Revista Ingenierías Universidad de Medellín*, 6(10), 77-91. Retrieved May 25, 2019. Recuperado de: http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S1692-33242007000100007&lng=en&tlng=es.
- Pérez, C. & Santín, D. (2008). Minería de Datos. Técnicas y Herramientas. Madrid. Thomson Ediciones Paraninfo S.A.
- Peña, M. & Orellana, J. (2018). Red neuronal para clasificación de riesgo en cooperativas de ahorro y crédito. Universidad de Cuenca. DOI: 10.24133/ctespe.v13i1.710
- Pilamunga, R. (2019). Integración de una herramienta Business Intelligence al core financiero para toma de decisiones en el área de cartera crediticia, para la cooperativa de ahorro y crédito indígena SAC Ltda. Trabajo de Graduación. Universidad Técnica de Ambato.
- Díaz Quevedo, O. (2009). "Determinantes del Ratio de Morosidad en el Sistemas Financiero Boliviano". Banco Central de Bolivia. Bolivia.

- Ramos, L., Schab, E., Rivera, R., Richard, C., Cristaldo, P., Núñez, J., Daián, G., Ríos, J., Retamar, S., Casanova, C., De Battista, A., Cagnina, L., Herrera, N. (2019) Descubrimiento de Conocimiento en Bases de Datos. Universidad Nacional de San Luis. Argentina.
- Rayo, S., Lara, J. y Camino, D., (2010), Un Modelo de Credit Scoring para instituciones de micro finanzas en el marco de Basilea II. Journal of Economics, Finance and Administrative Science, vol.15, p.28. Pags. 1-36.
- Rincon, A., y Torres, C. (2015). Clasificación de perfiles de riesgos a través de la Metodología de Árboles de decisión. pp. 2. polux.unipiloto.edu.co:8080/00002514.pdf
- Roddick, J. & Lees, B. (2001). Paradigms for spatial and spatio-temporal data mining. In MILLER, H. y HAN, J. Geographic data mining and knowledge discovery. London: Taylor & Francis.
- Rodríguez, Y. & Díaz, A. (2009). Herramientas de minería de datos. Revista Cubana de Ciencias Informáticas, [S.I.], v. 3, oct. ISSN 2227- 1899. Disponible en: <http://rcci.uci.cu/index.php/rcci/article/view/78>.
- Rodríguez-Guevara, D.E., Becerra-Arévalo, J.A. y Cardona-Valencia, D. (2017). Modelos y metodologías de credit score para personas naturales: una revisión literaria. Revista CEA, 3(5), 13-28.
- Rodríguez Suárez, Yuniét, Díaz Amador (2009). Anolandy, Herramientas de Minería de Datos. Revista Cubana de Ciencias Informáticas [en línea] 2009, 3 (Julio-Diciembre) Disponible en: <<http://google.redalyc.org/articulo.oa?id=378343637009>> ISSN 1994-1536
- Sánchez, C. & Giraldo, L. &, Piedrahita, C. & Bonet, I. & Lochmüller, C. & Tabares, S. & Aejandro, P. (2018). Análisis comparativo entre: “el análisis exploratorio de datos” y los modelos de “árboles de decisión” y “kmeans”. Revista Espacios. Vol. 39 (Nº 28) Año 2018 • Pág. 21.
- Saavedra, M. & Saavedra M, 2010. Modelos para medir el riesgo de crédito de la Banca. Universidad La Salle, Dirección de Posgrado e Investigación, México. Obtenido de: <http://www.scielo.org.co/pdf/cadm/v23n40/v23n40a13.pdf>
- Santillán N. T. y Gonzales Y. G., (2013). Morosidad y su incidencia en la rentabilidad de la caja municipal de ahorro y crédito Santa SA, distrito de Huarmey 2012. Tesis para optar el grado de Contador Público. Universidad Antenor Orrego – Trujillo.
- Schreiner, M., (2002), Ventajas y desventajas del scoring estadístico para las microfinanzas, microfinancerisk Management, Washington University in St. Louis, pp. 1-40.

- Siebes, A. (2000). Data Mining and Statistics Thuraisingham, B. Data Mining. Technologies, Techniques, Tools and Trends CRC Press LLC, 1999.
- Sosa, M. (2007). Inteligencia artificial en la gestión financiera empresarial. Universidad del Norte. Pag. 153-186. Enlace disponible en: <http://rcientificas.uninorte.edu.co/index.php/pensamiento/article/viewFile/3518/2252>
- Sudhamathy, G. (2016). Credit Risk Analysis and Prediction Modelling of Bank Loans Using R. International Journal of Engineering and Technology (IJET). Vol 8 No 5. Pag. 1-13. Recuperado de <http://www.enggjournals.com/ijet/docs/IJET16-08-05-414.pdf>.
- Superintendencia de Bancos e Instituciones Financieras (SBIF) – Chile, 2014. Variables que afectan la Tasa de Incumplimiento de Créditos Chilenos, Dirección de Estudios y Análisis Financiero, Departamento de Estudios. Obtenido de: https://sbif.cl/sbifweb3/internet/archivos/publicacion_10408.pdf
- Superintendencia de Bancos del Ecuador (2014). El Sistema Financiero. En A. Pena, Reporte de Estabilidad Financiera 2013 (págs. 41-80). Quito: Autor. Obtenido de http://www.sbs.gob.ec/medios/PORTALDOCS/downloads/La%20SBS/reporte_estabilidad_2013.pdf
- Superintendencia de Bancos y Seguros del Ecuador. (2014). Matrices de Transición y Análisis de Cosechas en el Contexto de Riesgo de Crédito, 1–22.
- Superintendencia de Bancos del Ecuador (2018). Libro I.- Normas de Control para las entidades de los sectores financieros público y privado. Título IX De la Gestión y Administración de Riesgos. Capítulo II De las Administración del Riesgo de Crédito.
- Superintendencia de Economía Popular y Solidaria (2011). Ley Orgánica de Economía Popular y Solidaria. Enlace disponible en: <http://www.seps.gob.ec/documents/20181/25522/Ley%20Orga%CC%81nica%20de%20Economi%CC%81a%20Popular%20y%20Solidaria.pdf/0836bc47-bf63-4aa0-b945-b94479a84ca1>
- Superintendencia de Economía Popular y Solidaria, (2015). Análisis de riesgo de crédito del Sector Financiero Popular y Solidario. Intendencia de Estadísticas, Estudios y Normas. Enlace disponible en: <http://www.seps.gob.ec/documents/20181/26626/Riesgo%20de%20Cr%C3%A9dito%20Sector%20Financiero%20Popular%20y%20Solidario-%20actualizado%2011%20agosto-2015.pdf/ea0f593f-0d1a-4f25-81f9-0317e9877d30>
- Superintendencia de Economía Popular y Solidaria. (2015). Economía Solidaria experiencias y conceptos. Enlace disponible en: <http://www.seps.gob.ec/documents/20181/26626/Economi%CC%81a%20Solid>

aria-Experiencias%20y%20Conceptos.pdf/2705a873-b256-4f6a-8dfd-59370b09f0d4

Superintendencia de Economía Popular y Solidaria. (2016). Dirección Nacional de Estadística y Estudios. Listado de Cooperativas Financieras. Disponible en: http://www.seps.gob.ec/documents/20181/341872/COOPERATIVAS+FINANCIERAS+CLASIFICADAS+POR+SEGMENTOS_04012016.xlsx/ac3630cb-6c44-4a0e-b198-879d5b008083

Superintendencia de Economía Popular y Solidaria. (2017). Boletín SEPS. Disponible en: <http://www.seps.gob.ec/documents/20181/455927/Boleti%CC%81n+SEPS+N%C2%BA005-02-03-2017.pdf/939e79f3-2907-46a9-b255-64f96197f898?version=1.0>

Superintendencia de Bancos y Seguros SBS (2013). Matrices de Transición y Análisis de Cosechas en el Contexto de Riesgo de Crédito. Superintendencia de Bancos y Seguros del Ecuador. Enlace disponible en: http://www.superbancos.gob.ec/medios/PORTALDOCS/downloads/articulos_financieros/Estudios_Especializados/EE1_2014.pdf

Superintendencia de Economía Popular y Solidaria SEPS (2015). Resolución No. 129-2015-F. Junta de Política y Regulación Monetaria y Financiera.

Superintendencia de Economía Popular y Solidaria SEPS, (2019). Productos estadísticos. Enlace disponible: <https://www.seps.gob.ec/>

Superintendencia de Economía Popular y Solidaria SEPS, sf.. Noticias. Enlace Disponible en: <https://www.seps.gob.ec/noticia?ecuador-tiene-un-total-de-887-cooperativas-de-ahorro-y-credito>

Tamariz, S. (2019). Elaboración de un modelo econométrico de riesgo de crédito para el sector automotriz caso: Indian Motos Inmot S.A. Universidad del Azuay. Cuenca. Ecuador.

Tello, E. & Perusquia, J., (2015). Inteligencia de Negocios :Estrategia para el Desarrollo de Competitividad en Empresas de base de tecnológica. Universidad Autónoma de Baja California, México. Pp. 8. Enlace disponible: <https://www.sciencedirect.com/science/article/pii/S0186104215000807>

Tello, M., Eslava Blanco, H., & Tobias, L. (2013). Análisis y evaluación del nivel de riesgo en el otorgamiento de créditos financieros utilizando técnicas de minería de datos. *Visión Electrónica*, 7(1), 13-26. <https://doi.org/10.14483/22484728.4389>

Ticona, P. (2018). La Calidad de La Cartera de Créditos y su incidencia en la solvencia financiera de las Cooperativas de Ahorro y Crédito de la Región Puno. Universidad Nacional del Altiplano. ISSN: 2077-8686. Enlace disponible en:

<http://www.revistaepgunapuno.org/index.php/SECONOMICO/article/view/621/202>

- Trejo, J. & Martínez, M. & Venegas, F. (2017). Administración del riesgo crediticio al menudeo en México: una mejora econométrica en la selección de variables y cambios en sus características. *Contaduría y administración*, 62(2), 377-398. <https://dx.doi.org/10.1016/j.cya.2017.01.003>. Enlace disponible: http://www.scielo.org.mx/scielo.php?script=sci_abstract&pid=S0186-10422017000200377&lng=es&nrm=iso
- Torres, E. & Farroñay, J. (2015). Implementación de minería de datos para detectar patrones de comportamiento de clientes morosos en empresa de crédito Crediserv Eirl – Chiclayo. Tesis de grado. Universidad Nacional “Pedro Ruiz Gallo”. Perú.
- Torres Peñafiel, Natali, & Fierro López, Pablo Enrique, & Alonso Alemán, Alodia (2017). Balance de la economía popular y solidaria en Ecuador. *Economía y Desarrollo*, 158(1), undefined-undefined. [fecha de Consulta 1 de mayo de 2019]. ISSN: 0252-8584. Disponible en: <https://www.redalyc.org/articulo.oa?id=4255/425553381014>
- Trejo-García, José Carlos, Ríos-Bolívar, Humberto, & Martínez-García, Miguel Ángel. (2016). Análisis de la Administración del Riesgo Crediticio en México para Tarjetas de Crédito. *Revista mexicana de economía y finanzas*, 11(1), 103-121. Recuperado de: http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S1665-53462016000100103&lng=es&tlng=pt.
- Universidad Nacional de Educación a Distancia, UNED (2015). El Riesgo de Crédito. Recuperado el 10 de octubre de 2017. Enlace disponible: http://portal.uned.es/pls/portal/docs/PAGE/UNED_MAIN/LAUNIVERSIDAD/DEPARTAMENTOS/0503/ASIGNATURAS/GESTION%20DE%20EMPRESAS%20FINANCIERAS/CAP%C3%8DTULO%206.2.PDF
- Valero, C & González, M., sf. Paquete para la clasificación de series temporales en Weka. Universidad Central de las Villas. Departamento de Inteligencia Artificial. Cuba.
- Valencia, A. (2017). Modelo Scoring para el otorgamiento de crédito de las pymes. Tesis Maestría. Universidad EAFIT. Medellín.
- Vela, L. & Uriol J. (2012). Los factores que determinan la calidad de la cartera crediticia de las entidades microfinancieras de la Amazonía peruana. Universidad Nacional Pedro Ruiz Gallo. Perú
- Vela, S. & Caro, A. (2015). Herramientas financieras en la evaluación del riesgo de crédito. Universidad Inca Garcilaso de la Vaca. Fondo Editorial.
- Villarino, G. (2015). Metodología de Minería de Datos para el estudio de tablas de siniestralidad vial. Trabajo de fin de Máster. Facultad de Estudios Estadísticos.

Universidad Complutense Madrid. Enlace disponible en:
https://eprints.ucm.es/34870/1/TFM_GuillermoVillarino_Nov_2015.pdf

Weiss, S.M. y Indurkha, N., 1998. *"Predictive Data Mining. A Practical Guide"*. Morgan Kaufmann Publishers. San Francisco. 1998.

ANEXOS

ANEXO No. 1

Matriz auxiliar de operacionalización en el diseño del trabajo de investigación

PROBLEMA	OBJETIVOS	OPERACIONALIZACIÓN DE VARIABLES				
Formulación del problema	Objetivo general	Variable dependiente	Variable independiente	Nomenclatura	Dimensión	Descripción
¿La calidad de la cartera de crédito puede verse afectada debido a la falta de tecnologías de crédito adecuadas para la evaluación de riesgo default?	Efectuar un modelo de evaluación de riesgo crediticio para el sector cooperativo con presencia en la ciudad de Loja, con el uso de técnicas y herramientas de minería de datos, con la intención de determinar patrones y/o comportamientos en los socios que minimice la probabilidad de una pérdida.	Riesgo de crédito "default"	Género	gen	Masculino	Definición del género del prestatario
					Femenino	
			Edad	rang_ed	De: 18 Hasta: 32	Rango de edad en la que se encuentra el prestatario
					De: 33 Hasta: 47	
					De: 48 Hasta: 62	
					De: 63 o más	
			Estado civil	est_civil	Soltero	Estado civil del prestatario en el Registro Civil
					Unión libre y casado	
					Casado con disolución conyugal	
					Divorciado	
			Nivel de estudios	niv_est	Sin estudio	Nivel educativo del prestatario
					Primaria	
Secundaria						
Formación intermedia técnica, Educación Superior y Cuarto Nivel						
Ocupación	ocup	Estudiante, Ama de Casa, y Jubilados	Ocupación principal del prestatario			
		Asalariados				
		Comerciantes bienes y servicios y actividades independiente				
		Transportistas (Terrestre, Áreo, Marítimo, de pasajeros y carga)				
¿Cuáles son las variables y comportamientos que inciden	Realizar un análisis de la concentración				Actividades de construcción; explotación; industriales	

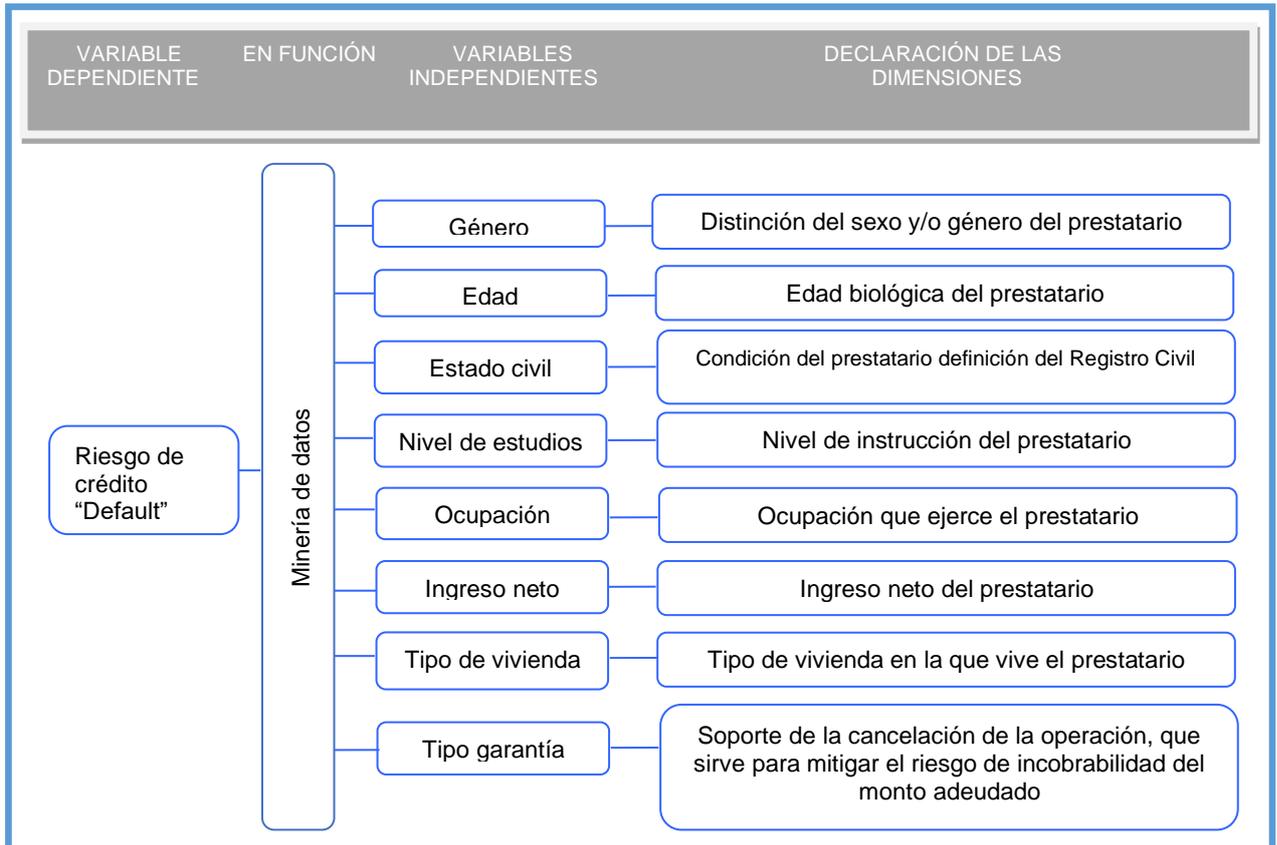
en la medición de la probabilidad de riesgo default?	de riesgo crediticio y la morosidad para el objeto de estudio.				Actividades de producción agrícola, ganadera, silvicultura, y pesca	
			Ingreso neto	mon_gast	Ingreso neto del prestatario	Ingresos-Gastos
			Tipo de vivienda	tip_viv	Prestada	Tipo de vivienda en el que habita el prestatario
					Vive con familiares	
Arrendada						
				Propia hipotecada o no hipotecada		
¿Las Cooperativa de Ahorro y Crédito pueden mejorar sus procesos de evaluación de riesgo de crédito con Minería de datos?	Determinar las variables que inciden en la probabilidad de riesgo "default" con el uso de la base de datos de la Cooperativa Ltda, para la construcción del modelo de evaluación de riesgo de crédito para el sector cooperativo, con la aplicación de minería de datos.		Tipo garantía	tip_gar	Personal	Garantía o respaldo crediticio a través de una firma solidaria
					Real	Garantía o respaldo crediticio a través de un bien prendado o hipotecado

FUENTE: Datos recopilados de la investigación

ELABORADO POR: Autor

ANEXO No. 2

Modelamiento de variables y declaración de dimensiones de la investigación



Fuente: Datos recopilados de la investigación
Elaborado: Autor

ANEXO No. 3

Matriz auxiliar de variables, dimensiones, indicadores y fuentes y técnicas de recolección de datos

Variable dependiente	Variable independiente	Nomenclatura	Tipo	Dimensión	Indicador Stata	Descripción	Técnicas	Instrumentos	Fuente
Riesgo de crédito "default"	Género	gen	Binaria	Masculino	0	Definición del género del prestatario	Estadística	Base de datos	Primaria / Cooperativa
				Femenino	1				
	Edad	rang_ed	Intervalo	De: 18 Hasta: 32	0	Rango de edad en la que se encuentra el prestatario	Estadística	Base de datos	Primaria / Cooperativa
				De: 33 Hasta: 47	1				
				De: 48 Hasta: 62	2				
				De: 63 o más	3				
	Estado civil	est_civil	Nominal	Soltero	0	Estado civil del prestatario en el Registro Civil	Estadística	Base de datos	Primaria / Cooperativa
				Unión libre y casado	1				
				Casado con disolución conyugal	2				
				Divorciado	3				
				Viudo	4				
	Nivel de estudios	niv_est	Ordinal	Sin estudio	0	Nivel educativo del prestatario	Estadística	Base de datos	Primaria / Cooperativa
				Primaria	1				
				Secundaria	2				
				Formación intermedia técnica, Educación Superior y Cuarto Nivel	3				
Ocupación	ocup	Nominal	Estudiante, Ama de Casa, y Jubilados	0	Ocupación principal del prestatario	Estadística	Base de datos	Primaria / Cooperativa	
			Asalariados	1					
			Comerciantes bienes y servicios y actividades independiente	2					

				Transportistas (Terrestre, Área, Marítimo, de pasajeros y carga)	3				
				Actividades de construcción; explotación; industriales	4				
				Actividades de producción agrícola, ganadera, silvicultura, y pesca	5				
	Ingreso neto	ing_net	Continua	Ingreso neto del prestatario	Continua	Ingresos-Gastos	Estadística	Base de datos	Primaria / Cooperativa
	Tipo de vivienda	tip_viv	Nominal	Prestada	0	Tipo de vivienda en el que habita el prestatario	Estadística	Base de datos	Primaria / Cooperativa
Vive con familiares				1					
Arrendada				2					
Propia hipotecada o no hipotecada				3					
	Tipo garantía	tip_gar	Binaria	Personal	0	Garantía o respaldo crediticio a través de una firma solidaria	Estadística	Base de datos	Primaria / Cooperativa
				Real	1	Garantía o respaldo crediticio a través de un bien prendado o hipotecado	Estadística	Base de datos	Primaria / Cooperativa

Fuente: Datos recopilados de la investigación

Elaborado: Autor

ANEXO No. 4

Corridas estadísticas

Logistic regression	Number of obs	=	2145
	LR chi2(21)	=	35.63
	Prob > chi2	=	0.0240
Log likelihood = -230.65626	Pseudo R2	=	0.0717

default	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
1.gene	.2752502	.3056977	0.90	0.368	-.3239063	.8744067
rang_ed						
1	-.0084143	.3465557	-0.02	0.981	-.6876509	.6708223
2	-.1775308	.4966905	-0.36	0.721	-1.151026	.7959648
3	1.240159	1.095364	1.13	0.258	-.9067143	3.387033
est_civil						
1	.9619406	.401626	2.40	0.017	.1747681	1.749113
2	.3048295	.7742155	0.39	0.694	-1.212605	1.822264
3	.3524983	.5328423	0.66	0.508	-.6918534	1.39685
4	.3730648	1.093959	0.34	0.733	-1.771055	2.517184
niv_est						
1	.131633	1.072156	0.12	0.902	-1.969754	2.23302
2	.7970381	1.063427	0.75	0.454	-1.287241	2.881317
3	.4445961	1.073566	0.41	0.679	-1.659555	2.548748
ocup						
1	.1772976	.5437112	0.33	0.744	-.8883567	1.242952
2	-.1493742	.5430247	-0.28	0.783	-1.213683	.9149346
3	.6538761	.8780585	0.74	0.456	-1.067087	2.374839
4	.2885857	.7184709	0.40	0.688	-1.119591	1.696763
5	.4231897	.6980926	0.61	0.544	-.9450467	1.791426
ing_net	.0007524	.000378	1.99	0.047	.0000115	.0014932
tip_viv						
1	1.744409	1.116926	1.56	0.118	-.4447254	3.933543
2	.9572441	1.131903	0.85	0.398	-1.261245	3.175734
3	1.766168	1.144109	1.54	0.123	-.4762434	4.00858
1.tip_gar	.3708418	.3248788	1.14	0.254	-.2659089	1.007592
_cons	.2676251	1.626727	0.16	0.869	-2.920702	3.455952

Logistic regression

Number of obs = 2145

LR chi2(20) = 35.63

Prob > chi2 = 0.0170

Log likelihood = -230.65656

Pseudo R2 = 0.0717

default	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
1.gene	.2752309	.3056866	0.90	0.368	-.3239039	.8743656
rang_ed						
2	-.1714324	.4285012	-0.40	0.689	-1.011279	.6684145
3	1.245895	1.069595	1.16	0.244	-.8504729	3.342263
est_civil						
1	.960502	.3972766	2.42	0.016	.1818542	1.73915
2	.3029486	.7703769	0.39	0.694	-1.206962	1.81286
3	.3499081	.5221436	0.67	0.503	-.6734746	1.373291
4	.3717204	1.092573	0.34	0.734	-1.769683	2.513124
niv_est						
1	.1319976	1.072038	0.12	0.902	-1.969159	2.233154
2	.7982137	1.062317	0.75	0.452	-1.28389	2.880317
3	.4450029	1.073422	0.41	0.678	-1.658865	2.548871
ocup						
1	.1761175	.5415116	0.33	0.745	-.8852257	1.237461
2	-.1500999	.5421874	-0.28	0.782	-1.212768	.9125679
3	.6534326	.877834	0.74	0.457	-1.06709	2.373956
4	.2876043	.7173123	0.40	0.688	-1.118302	1.693511
5	.4223409	.697219	0.61	0.545	-.9441833	1.788865
ing_net	.0007518	.0003772	1.99	0.046	.0000125	.0014911
tip_viv						
1	1.746772	1.112698	1.57	0.116	-.4340763	3.92762
2	.9593585	1.128586	0.85	0.395	-1.25263	3.171347
3	1.766744	1.143904	1.54	0.122	-.4752656	4.008754
1.tip_gar	.3699615	.3228575	1.15	0.252	-.2628276	1.002751
_cons	.2629383	1.615215	0.16	0.871	-2.902826	3.428702

Logistic regression

Number of obs = 2145

LR chi2(19) = 35.62

Prob > chi2 = 0.0118

Log likelihood = -230.66389

Pseudo R2 = 0.0717

default	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
1.gene	.2736196	.3054227	0.90	0.370	-.3249979 .8722372
rang_ed					
2	-.1710876	.4286012	-0.40	0.690	-1.011131 .6689554
3	1.244449	1.069719	1.16	0.245	-.8521611 3.341059
est_civil					
1	.9586514	.3968753	2.42	0.016	.18079 1.736513
2	.3040605	.7703054	0.39	0.693	-1.20571 1.813831
3	.3491896	.522263	0.67	0.504	-.6744271 1.372806
4	.3750691	1.092303	0.34	0.731	-1.765806 2.515944
niv_est					
2	.6753665	.3758003	1.80	0.072	-.0611884 1.411922
3	.3222688	.4085298	0.79	0.430	-.478435 1.122973
ocup					
1	.1760626	.5415948	0.33	0.745	-.8854437 1.237569
2	-.1488247	.542177	-0.27	0.784	-1.211472 .9138228
3	.6556349	.8777108	0.75	0.455	-1.064647 2.375916
4	.289293	.7172678	0.40	0.687	-1.116526 1.695112
5	.4174901	.6960969	0.60	0.549	-.9468347 1.781815
ing_net	.0007507	.0003772	1.99	0.047	.0000113 .00149
tip_viv					
1	1.745656	1.112586	1.57	0.117	-.434973 3.926285
2	.9604031	1.128517	0.85	0.395	-1.251449 3.172256
3	1.76822	1.143775	1.55	0.122	-.4735368 4.009977
1.tip_gar	.3687756	.3227327	1.14	0.253	-.2637689 1.00132
_cons	.3873816	1.26265	0.31	0.759	-2.087368 2.862131

Logistic regression

Number of obs = 2145

LR chi2(18) = 35.54

Prob > chi2 = 0.0081

Log likelihood = -230.70235

Pseudo R2 = 0.0715

default	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
1.gene	.284978	.3028186	0.94	0.347	-.3085355	.8784915
rang_ed						
2	-.1654622	.4283541	-0.39	0.699	-1.005021	.6740964
3	1.293067	1.057153	1.22	0.221	-.7789153	3.36505
est_civil						
1	.9593511	.3970348	2.42	0.016	.1811771	1.737525
2	.296565	.7698613	0.39	0.700	-1.212335	1.805465
3	.347429	.5228078	0.66	0.506	-.6772554	1.372114
4	.3685079	1.092445	0.34	0.736	-1.772645	2.509661
niv_est						
2	.6724607	.3757676	1.79	0.074	-.0640303	1.408952
3	.3060809	.404559	0.76	0.449	-.4868402	1.099002
ocup						
1	.2931092	.3274692	0.90	0.371	-.3487186	.934937
3	.7717251	.7660425	1.01	0.314	-.7296906	2.273141
4	.4034257	.5804494	0.70	0.487	-.7342343	1.541086
5	.5258067	.5688182	0.92	0.355	-.5890564	1.64067
ing_net	.0007443	.0003761	1.98	0.048	7.18e-06	.0014815
tip_viv						
1	1.756352	1.111844	1.58	0.114	-.4228213	3.935526
2	.9687159	1.128035	0.86	0.390	-1.242191	3.179623
3	1.776363	1.143492	1.55	0.120	-.4648411	4.017566
1.tip_gar	.3705083	.3226991	1.15	0.251	-.2619703	1.002987
_cons	.2681134	1.183976	0.23	0.821	-2.052436	2.588663

Logistic regression

Number of obs = 2145

LR chi2(17) = 35.41

Prob > chi2 = 0.0055

Pseudo R2 = 0.0713

Log likelihood = -230.76432

default	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
1.gene	.2919647	.302251	0.97	0.334	-.3004364 .8843658
rang_ed					
2	-.136918	.4205818	-0.33	0.745	-.9612431 .6874072
3	1.358266	1.04295	1.30	0.193	-.6858786 3.402411
est_civil					
1	.9300858	.3876359	2.40	0.016	.1703334 1.689838
2	.2639617	.7639474	0.35	0.730	-1.233348 1.761271
3	.317365	.515144	0.62	0.538	-.6922988 1.327029
niv_est					
2	.668888	.3758897	1.78	0.075	-.0678423 1.405618
3	.3061238	.404792	0.76	0.449	-.4872539 1.099502
ocup					
1	.2833118	.3263503	0.87	0.385	-.3563231 .9229466
3	.7650722	.7658264	1.00	0.318	-.73592 2.266064
4	.4013026	.580159	0.69	0.489	-.7357881 1.538393
5	.5123531	.5675031	0.90	0.367	-.5999325 1.624639
ing_net	.0007456	.0003765	1.98	0.048	7.62e-06 .0014836
tip_viv					
1	1.754787	1.1115	1.58	0.114	-.4237124 3.933287
2	.9689723	1.127776	0.86	0.390	-1.241428 3.179373
3	1.793742	1.142443	1.57	0.116	-.4454057 4.03289
1.tip_gar	.3726969	.3225129	1.16	0.248	-.2594167 1.00481
_cons	.277606	1.183259	0.23	0.815	-2.041539 2.596751

Logistic regression
 Log likelihood = -230.81668

Number of obs = 2145
 LR chi2(16) = 35.31
 Prob > chi2 = 0.0036
 Pseudo R2 = 0.0711

default	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
1.gene	.2884236	.3019827	0.96	0.340	-.3034517	.8802989
3.rang_ed	1.409979	1.030508	1.37	0.171	-.6097796	3.429738
est_civil						
1	.9128161	.3834877	2.38	0.017	.1611941	1.664438
2	.235629	.7588967	0.31	0.756	-1.251781	1.723039
3	.2747227	.4973318	0.55	0.581	-.7000297	1.249475
niv_est						
2	.67606	.3752364	1.80	0.072	-.0593898	1.41151
3	.3177336	.4032999	0.79	0.431	-.4727196	1.108187
ocup						
1	.2772927	.3257644	0.85	0.395	-.3611938	.9157792
3	.7550384	.7650885	0.99	0.324	-.7445075	2.254584
4	.4064541	.579987	0.70	0.483	-.7302996	1.543208
5	.4979751	.5655609	0.88	0.379	-.6105039	1.606454
ing_net	.0007373	.000375	1.97	0.049	2.35e-06	.0014722
tip_viv						
1	1.743525	1.110464	1.57	0.116	-.4329435	3.919994
2	.9545573	1.126407	0.85	0.397	-1.25316	3.162275
3	1.74509	1.131448	1.54	0.123	-.4725066	3.962686
1.tip_gar	.3629127	.3209432	1.13	0.258	-.2661244	.9919498
_cons	.2898569	1.182057	0.25	0.806	-2.026933	2.606647

Logistic regression
 Log likelihood = -230.86781

Number of obs = 2145
 LR chi2(15) = 35.21
 Prob > chi2 = 0.0023
 Pseudo R2 = 0.0708

default	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
1.gene	.2939191	.3015513	0.97	0.330	-.2971105	.8849487
3.rang_ed	1.406632	1.030338	1.37	0.172	-.6127941	3.426059
est_civil						
1	.8904532	.3770702	2.36	0.018	.1514092	1.629497
3	.2554581	.4938612	0.52	0.605	-.7124921	1.223408
niv_est						
2	.6771017	.3751427	1.80	0.071	-.0581644	1.412368
3	.3258262	.4025189	0.81	0.418	-.4630963	1.114749
ocup						
1	.2696017	.324859	0.83	0.407	-.3671103	.9063137
3	.7555285	.7650519	0.99	0.323	-.7439458	2.255003
4	.4009158	.5797625	0.69	0.489	-.7353977	1.537229
5	.4918511	.5651049	0.87	0.384	-.6157341	1.599436
ing_net	.0007462	.0003744	1.99	0.046	.0000124	.0014799
tip_viv						
1	1.751027	1.110208	1.58	0.115	-.4249396	3.926994
2	.9682684	1.125625	0.86	0.390	-1.237917	3.174454
3	1.769525	1.128904	1.57	0.117	-.4430859	3.982136
1.tip_gar	.3680551	.3206433	1.15	0.251	-.2603942	.9965044
_cons	.2835002	1.181689	0.24	0.810	-2.032568	2.599568

Logistic regression

Number of obs = 2145
 LR chi2(14) = 34.92
 Prob > chi2 = 0.0015
 Pseudo R2 = 0.0703

Log likelihood = -231.00922

default	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
1.gene	.29919	.3013584	0.99	0.321	-.2914617	.8898417
3.rang_ed	1.427613	1.029441	1.39	0.166	-.5900536	3.44528
1.est_civil	.8482151	.3692045	2.30	0.022	.1245875	1.571843
niv_est						
2	.6702499	.3749829	1.79	0.074	-.064703	1.405203
3	.3253998	.4023706	0.81	0.419	-.463232	1.114032
ocup						
1	.2712519	.3250042	0.83	0.404	-.3657447	.9082484
3	.7534852	.7649853	0.98	0.325	-.7458584	2.252829
4	.3898424	.5793297	0.67	0.501	-.745623	1.525308
5	.4851912	.5651539	0.86	0.391	-.62249	1.592872
ing_net	.000751	.0003745	2.01	0.045	.000017	.0014849
tip_viv						
1	1.742348	1.109363	1.57	0.116	-.4319647	3.91666
2	.9747376	1.125038	0.87	0.386	-1.230296	3.179771
3	1.781053	1.12788	1.58	0.114	-.4295514	3.991658
1.tip_gar	.3775602	.3203317	1.18	0.239	-.2502784	1.005399
_cons	.3129805	1.179487	0.27	0.791	-1.998772	2.624733

Logistic regression

Number of obs = 2145
 LR chi2(13) = 34.44
 Prob > chi2 = 0.0010
 Pseudo R2 = 0.0693

Log likelihood = -231.25136

default	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
1.gene	.2502666	.2943087	0.85	0.395	-.3265678	.827101
3.rang_ed	1.391014	1.028041	1.35	0.176	-.6239093	3.405938
1.est_civil	.8439475	.3692058	2.29	0.022	.1203173	1.567578
niv_est						
2	.6466238	.3736973	1.73	0.084	-.0858094	1.379057
3	.2855964	.400443	0.71	0.476	-.4992575	1.07045
ocup						
1	.2085543	.3152952	0.66	0.508	-.409413	.8265216
3	.6635371	.7551896	0.88	0.380	-.8166073	2.143681
5	.4058909	.5548883	0.73	0.464	-.6816702	1.493452
ing_net	.0007442	.0003736	1.99	0.046	.0000119	.0014765
tip_viv						
1	1.767068	1.10817	1.59	0.111	-.4049053	3.939042
2	.9991626	1.123788	0.89	0.374	-1.203421	3.201746
3	1.808495	1.126642	1.61	0.108	-.3996836	4.016673
1.tip_gar	.3690707	.3202472	1.15	0.249	-.2586023	.9967438
_cons	.4074376	1.171629	0.35	0.728	-1.888912	2.703787

Logistic regression

Number of obs = 2145
 LR chi2(12) = 34.00
 Prob > chi2 = 0.0007
 Pseudo R2 = 0.0684

Log likelihood = -231.47093

default	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
1.gene	.237494	.2939675	0.81	0.419	-.3386716	.8136597
3.rang_ed	1.358204	1.026723	1.32	0.186	-.6541362	3.370545
1.est_civil	.8495607	.3693841	2.30	0.021	.1255813	1.57354
niv_est						
2	.6820114	.370153	1.84	0.065	-.0434751	1.407498
3	.3500075	.3894677	0.90	0.369	-.4133352	1.11335
ocup						
3	.5744089	.7443695	0.77	0.440	-.8845285	2.033346
5	.3333081	.5459622	0.61	0.542	-.7367581	1.403374
ing_net	.0007373	.0003739	1.97	0.049	4.42e-06	.0014702
tip_viv						
1	1.750332	1.107498	1.58	0.114	-.4203245	3.920989
2	.9793775	1.123341	0.87	0.383	-1.222331	3.181085
3	1.784823	1.126046	1.59	0.113	-.4221867	3.991833
1.tip_gar	.4006627	.3166788	1.27	0.206	-.2200163	1.021342
_cons	.4855171	1.16619	0.42	0.677	-1.800174	2.771209

Logistic regression

Number of obs = 2145
 LR chi2(11) = 33.60
 Prob > chi2 = 0.0004
 Pseudo R2 = 0.0676

Log likelihood = -231.67218

default	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
1.gene	.2456248	.2934114	0.84	0.403	-.3294509	.8207005
3.rang_ed	1.357569	1.02668	1.32	0.186	-.6546872	3.369824
1.est_civil	.8610648	.3690054	2.33	0.020	.1378275	1.584302
niv_est						
2	.6497	.3670233	1.77	0.077	-.0696525	1.369052
3	.2971688	.3814201	0.78	0.436	-.4504009	1.044739
3.ocup	.5398476	.7426285	0.73	0.467	-.9156776	1.995373
ing_net	.0007352	.0003722	1.98	0.048	5.71e-06	.0014646
tip_viv						
1	1.763222	1.10751	1.59	0.111	-.4074585	3.933902
2	.9920193	1.123254	0.88	0.377	-1.209518	3.193556
3	1.804084	1.125902	1.60	0.109	-.4026445	4.010812
1.tip_gar	.3848046	.3157886	1.22	0.223	-.2341297	1.003739
_cons	.5331358	1.16423	0.46	0.647	-1.748712	2.814984

Logistic regression

Number of obs = 2145
 LR chi2(10) = 32.99
 Prob > chi2 = 0.0003
 Pseudo R2 = 0.0664

Log likelihood = -231.97865

default	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
1.gene	.2102358	.2910859	0.72	0.470	-.360282	.7807536
3.rang_ed	1.349475	1.026852	1.31	0.189	-.6631185	3.362068
1.est_civil	.857783	.3692457	2.32	0.020	.1340748	1.581491
niv_est						
2	.6572496	.3663171	1.79	0.073	-.0607187	1.375218
3	.2814421	.3807492	0.74	0.460	-.4648126	1.027697
ing_net	.0007639	.0003729	2.05	0.040	.0000331	.0014948
tip_viv						
1	1.800453	1.10723	1.63	0.104	-.3696777	3.970584
2	1.016411	1.123133	0.90	0.365	-1.18489	3.217712
3	1.83078	1.125662	1.63	0.104	-.3754762	4.037036
1.tip_gar	.3758332	.316086	1.19	0.234	-.2436839	.9953503
_cons	.5312021	1.165184	0.46	0.648	-1.752517	2.814921

Logistic regression

Number of obs = 2145
 LR chi2(9) = 32.46
 Prob > chi2 = 0.0002
 Pseudo R2 = 0.0653

Log likelihood = -232.24245

default	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
3.rang_ed	1.361869	1.027104	1.33	0.185	-.6512181	3.374957
1.est_civil	.8418642	.3683721	2.29	0.022	.1198681	1.56386
niv_est						
2	.6608304	.3660959	1.81	0.071	-.0567043	1.378365
3	.3160008	.3774722	0.84	0.403	-.4238311	1.055833
ing_net	.0007682	.0003732	2.06	0.040	.0000367	.0014997
tip_viv						
1	1.818539	1.10546	1.65	0.100	-.3481233	3.985201
2	1.043803	1.121123	0.93	0.352	-1.153557	3.241164
3	1.867189	1.123	1.66	0.096	-.3338499	4.068229
1.tip_gar	.3698874	.3155099	1.17	0.241	-.2485007	.9882755
_cons	.5825306	1.162537	0.50	0.616	-1.696	2.861061

Logistic regression

Number of obs = 2145
 LR chi2(8) = 31.77
 Prob > chi2 = 0.0001
 Pseudo R2 = 0.0639

Log likelihood = -232.588

default	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
3.rang_ed	1.309065	1.025035	1.28	0.202	-.6999669	3.318097
1.est_civil	.8323517	.3681344	2.26	0.024	.1108216	1.553882
2.niv_est	.4711216	.2961329	1.59	0.112	-.1092882	1.051531
ing_net	.0008196	.0003721	2.20	0.028	.0000903	.0015488
tip_viv						
1	1.777756	1.105246	1.61	0.108	-.3884859	3.943997
2	1.020711	1.121398	0.91	0.363	-1.177188	3.218609
3	1.781649	1.118378	1.59	0.111	-.4103325	3.97363
1.tip_gar	.402643	.3133189	1.29	0.199	-.2114507	1.016737
_cons	.7835505	1.141004	0.69	0.492	-1.452777	3.019878

Logistic regression

Number of obs = 2145
 LR chi2(7) = 31.10
 Prob > chi2 = 0.0001
 Pseudo R2 = 0.0626

Log likelihood = -232.92196

default	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
3.rang_ed	1.312649	1.024885	1.28	0.200	-.696088	3.321386
1.est_civil	.8421107	.3679021	2.29	0.022	.1210358	1.563186
2.niv_est	.4609278	.2955361	1.56	0.119	-.1183123	1.040168
ing_net	.0008075	.0003734	2.16	0.031	.0000755	.0015394
tip_viv						
1	.8049497	.354976	2.27	0.023	.1092096	1.50069
3	.8072281	.387178	2.08	0.037	.0483731	1.566083
1.tip_gar	.3938779	.3125871	1.26	0.208	-.2187816	1.006537
_cons	1.769217	.4012227	4.41	0.000	.982835	2.555599

Logistic regression

Number of obs = 2145
 LR chi2(6) = 29.45
 Prob > chi2 = 0.0001
 Pseudo R2 = 0.0593

Log likelihood = -233.7481

default	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
3.rang_ed	1.332637	1.02393	1.30	0.193	-.674228	3.339502
1.est_civil	.8580577	.3678003	2.33	0.020	.1371824	1.578933
2.niv_est	.431733	.294838	1.46	0.143	-.1461387	1.009605
ing_net	.000776	.0003679	2.11	0.035	.0000548	.0014971
tip_viv						
1	.7511421	.3527388	2.13	0.033	.0597869	1.442497
3	.8333858	.3871427	2.15	0.031	.0746001	1.592172
_cons	1.947213	.3754399	5.19	0.000	1.211365	2.683062

Logistic regression

Number of obs = 2145

LR chi2(5) = 26.85

Prob > chi2 = 0.0001

Pseudo R2 = 0.0540

Log likelihood = -235.0443

default	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
1.est_civil	.8673087	.3662096	2.37	0.018	.1495511	1.585066
2.niv_est	.3996208	.2943875	1.36	0.175	-.1773682	.9766098
ing_net	.0007578	.0003639	2.08	0.037	.0000446	.0014709
tip_viv						
1	.7552575	.3522443	2.14	0.032	.0648713	1.445644
3	.9287084	.3838212	2.42	0.016	.1764327	1.680984
_cons	1.988785	.3734967	5.32	0.000	1.256745	2.720825

Logistic regression

Number of obs = 2145

LR chi2(4) = 24.97

Prob > chi2 = 0.0001

Pseudo R2 = 0.0502

Log likelihood = -235.98745

default	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
1.est_civil	.8611669	.3664254	2.35	0.019	.1429862	1.579348
ing_net	.0007079	.0003613	1.96	0.050	-2.10e-07	.001416
tip_viv						
1	.8210727	.3489575	2.35	0.019	.1371285	1.505017
3	.9311534	.3836347	2.43	0.015	.1792432	1.683064
_cons	2.162078	.3544076	6.10	0.000	1.467451	2.856704

ANEXO No. 5 Certificaciones del modelo obtenido

ROGER AUGUSTO CHAMBA GONZÁLEZ

CERTIFICA:

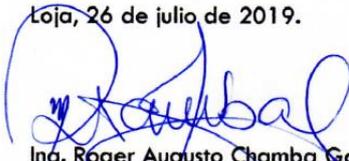
En conocimiento del contenido del documento desarrollado para optar por el título de Magister en Sistemas de Información Gerencial en la Universidad Tecnológica Empresarial de Guayaquil, denominado "Modelo para la evaluación del riesgo en la concesión de operaciones crediticias del Sector Cooperativo de la Ciudad de Loja, utilizando técnicas y herramientas de minería de datos", elaborado por la ingeniera Liliana Juliza González Lapo, me permito manifestar:

1. El planteamiento del problema detallado en el documento de tesis en referencia es relevante porque se plantea la importancia de que las Cooperativas de Ahorro y Crédito del Sector Financiero Popular y Solidario en la ciudad de Loja; busquen medir la probabilidad del incumplimiento (default) mediante técnicas y herramientas de minería de datos en las operaciones de crédito.
2. Los procedimientos realizados por la ingeniera Liliana Juliza González Lapo, al momento del diseño y validación de la estructura de información para el análisis y definición del incumplimiento (default); han sido desarrollados considerando criterios técnicos y estadísticos, los cuales le han permitido determinar el punto de inflexión a partir del cual una cooperativa de ahorro y crédito puede identificar si una operación de crédito se podría encontrar como incumplida. A su vez en el análisis se ha podido determinar las perspectivas de variables candidatas que aportan y aquellas que no inciden en el modelo de evaluación del riesgo crediticio.
3. La técnica de minería de datos que se ha utilizado en el desarrollo del documento de tesis en referencia fue de tipo predictivo (regresión logística) con la aplicación del modelo Logit, el cual es útil para este tipo de análisis y para la construcción de modelos de evaluación de riesgo en la concesión de operaciones crediticias.

En el desarrollo del documento de tesis, el modelo Logit ha sido de mucha utilidad pues permite medir las probabilidades e identificar los comportamientos de las variables consideradas más significativas para el modelo econométrico; y, conforme se detalla en este documento, los resultados obtenidos por la autora expresan que la probabilidad de χ^2 del modelo no supera a 0.05, aunque se observa que el indicador de la variable Pseudo R2 es muy bajo.

4. Un referente adicional a las características del modelo elaborado por la ingeniera Liliana Juliza González Lapo, es que el resultado de las pruebas efectuadas mediante el test de Hosmer-Lemeshow, el cual evalúa si las frecuencias observadas coinciden con las esperadas en una distribución de datos de un modelo logit, fue del 75.30%, por lo que se considera el modelo tiene buen ajuste.
5. El desarrollo de un modelo para la evaluación del riesgo en la concesión de operaciones crediticias del Sector Cooperativo de la Ciudad de Loja, tal como el planteado, es de gran importancia para las entidades financieras en referencia, pero es necesario manifestar que al ser la aversión al riesgo de crédito muy diferente para cada una de las cooperativas, el perfeccionamiento y fortalecimiento del modelo es propio de cada una de ellas por lo que necesariamente se debe desarrollar las variables que le permitan mejorar los indicadores de predictibilidad y determinación de probabilidad del incumplimiento (default) del modelo planteado en la tesis.

Loja, 26 de julio de 2019.



Ing. Roger Augusto Chamba González, Mae.

ASESOR - CONSULTOR - ESPECIALISTA EN ADMINISTRACIÓN INTEGRAL DE RIESGOS.

Telf.: 072102708 - Celular: 0994722443

Email: rchamba@gmail.com

CERTIFICACIÓN Y VALIDACIÓN DEL TRABAJO INVESTIGATIVO DE TESIS

En calidad de asesora y consultora independiente;

CERTIFICO:

1. El presente trabajo de investigación realizado por la Ing. Liliana González Lapo, para obtener el título de Magister en Sistema de Información Gerencial en la Universidad Tecnológica Empresarial de Guayaquil, denominado **"Modelo para la evaluación del riesgo en la concesión de operaciones crediticias del Sector Cooperativo de la Ciudad de Loja, utilizando técnicas y herramientas de minería de datos"**, considera una recopilación de información literaria como numérica importante para cumplir los objetivos que plantea.
2. La técnica de minería de datos que utiliza en el presente trabajo de investigación, está basada en la realización de un modelo econométrico logístico o logit, el cual permite establecer una relación no lineal entre la variable dependiente cualitativa Default (Riesgo Crediticio) y las variables independientes, dando como resultado, que las únicas variables estadísticamente significativas son: estado civil y tipo de vivienda de los clientes, presentando una relación positiva entre ellas.

Así también, al evaluar el modelo en conjunto, establece que las variables independientes se asocian a la variable dependiente y estas son diferentes de cero, determinando que los datos se ajustan al modelo. Por otro lado, para este caso de estudio, el 4% de las variaciones de la variable dependiente Default es explicada por las variaciones de las variables independientes: estado civil y tipo de vivienda; un porcentaje que no demuestra ser representativo para el modelo, puesto que puede existir algún tipo de variable que demuestre un porcentaje más alto de asociación.

Para comprobar que los datos expuestos en un modelo econométrico son utilizados correctamente, se utilizan diferentes tipos de test, los cuales sirven para evaluar y aceptar los resultados que se obtienen. Para este trabajo de investigación, la Ing. ha utilizado el test de Hosmer-Lemeshow, el cual le permitió comprobar que existe un buen ajuste de los datos, es decir, es un modelo que puede ser aceptado porque cumple con los parámetros establecidos.

3. Los resultados que se obtienen en el presente tema de investigación, son aceptados y validados, puesto que cumplen con todas las pruebas de significancia. Es bueno acotar, que los resultados pueden variar según la población escogida.

Finalmente, para la Cooperativa que ha considerado la Ing. y los datos con los cuales ha trabajado, el estado civil de un individuo y el tipo de vivienda que posee, pueden influir sobre el Riesgo Crediticio.

Loja, 26 de julio de 2019

Atentamente



Econ. María Gabriela Encalada Correa
ASESORA – CONSULTORA INDEPENDIENTE
Telf: 072615796 – Cell: 0990947410
Email: mencaladacorrea@gmail.com