

UNIVERSIDAD CATÓLICA SANTO TORIBIO DE MOGROVEJO
FACULTAD DE INGENIERÍA
ESCUELA DE INGENIERÍA DE SISTEMAS Y COMPUTACIÓN



**SISTEMA DE INFORMACIÓN BASADO EN REDES NEURONALES
PARA LA PREDICCIÓN DE RIESGO EN EL OTORGAMIENTO DE
CRÉDITOS PERSONALES EN UNA COOPERATIVA DE AHORRO Y
CRÉDITO EN EL DEPARTAMENTO DE LAMBAYEQUE**

**TESIS PARA OPTAR EL TÍTULO DE
INGENIERO DE SISTEMAS Y COMPUTACIÓN**

**AUTOR
ADRIAN MITHSUO GARCIA NAZARIO**

**ASESOR
LUIS AUGUSTO ZUÑE BISPO**
<https://orcid.org/0000-0001-7838-8656>

Chiclayo, 2020

**SISTEMA DE INFORMACIÓN BASADO EN REDES
NEURONALES PARA LA PREDICCIÓN DE RIESGO EN EL
OTORGAMIENTO DE CRÉDITOS PERSONALES EN UNA
COOPERATIVA DE AHORRO Y CRÉDITO EN EL
DEPARTAMENTO DE LAMBAYEQUE**

PRESENTADA POR:
ADRIAN MITHSUO GARCIA NAZARIO

A la Facultad de Ingeniería de la
Universidad Católica Santo Toribio de Mogrovejo
para optar el título de

INGENIERO DE SISTEMAS Y COMPUTACIÓN

APROBADA POR:

Ricardo David Iman Espinoza
PRESIDENTE

Ernesto Ludwin Nicho Cordova
SECRETARIO

Luis Augusto Zuñe Bispo
VOCAL

DEDICATORIA

A Dios por ser el inspirador y la fuerza para lograr obtener uno de mis anhelos más deseados, con salud, sabiduría e inteligencia y que cada decisión sea la mejor posible.

A mis padres, por su amor, trabajo y esfuerzo en todos estos años, gracias a ellos he logrado llegar hasta aquí y convertirme en lo que soy.

A mi hermana por ser ejemplo y motor de inspiración para seguir adelante en los estudios.

AGRADECIMIENTOS

A mi asesor de tesis por su dedicación y paciencia en cada asesoría, buscando la manera de obtener mejores resultados, en donde cada corrección y sugerencia aportó a lograr los objetivos trazados.

A la empresa por su confianza, brindándome todas las facilidades en el desarrollo del presente trabajo; a cada miembro que aportó con sus conocimientos y experiencias, en especial al jefe de sistemas por su tiempo y apoyo incondicional.

ÍNDICE

RESUMEN.....	8
ABSTRACT	9
I. INTRODUCCIÓN	10
II. MARCO TEÓRICO	12
2.1. ANTECEDENTES.....	12
2.1.1. ANTECEDENTES INTERNACIONALES	12
2.1.2. ANTECEDENTES NACIONALES.....	13
2.1.3. ANTECEDENTES LOCALES	14
2.2. BASES TEÓRICO-CIENTÍFICAS	15
2.2.1. CRISP-DM	15
2.2.1.1. Fases del Crisp-DM.....	16
2.2.2. CREDIT SCORING	17
2.2.2.1. Modelos credit scoring	17
III. METODOLOGÍA	18
3.1. TIPO Y NIVEL DE INVESTIGACIÓN	19
3.1.1. TIPO DE INVESTIGACIÓN	19
3.1.2. NIVEL DE INVESTIGACIÓN.....	19
3.2. DISEÑO DE INVESTIGACIÓN	19
3.3. POBLACIÓN, MUESTRA Y MUESTREO.....	20
3.3.1. POBLACIÓN	20
3.4. CRITERIOS DE SELECCIÓN	21
3.5. OPERACIONALIZACIÓN DE VARIABLES	21
3.5.1. VARIABLES	21
3.5.1.1. Variable independiente	21
3.5.1.2. Variable dependiente	21
3.5.2. INDICADORES (OPERACIONALIZACIÓN DE VARIABLES)	22
3.6. TÉCNICAS E INSTRUMENTOS DE RECOLECCIÓN DE DATOS.....	24
3.7. PROCEDIMIENTOS	25
3.7.1. METODOLOGÍA DE DESARROLLO	25

3.7.2. ANÁLISIS DE RIESGOS	42
3.7.3. PRODUCTO ACREDITABLE.....	42
3.7.4. MANUAL DE USUARIO	43
3.8. MATRIZ DE CONSISTENCIA	44
3.9. CONSIDERACIONES ÉTICAS.....	45
IV. RESULTADOS	46
4.1. EN BASE A LOS OBJETIVOS DE LA INVESTIGACIÓN.....	46
4.1.1. IDENTIFICAR LAS VARIABLES PARA DETERMINAR EL INGRESO DEL CLIENTE	46
4.1.2. IMPLEMENTAR UN SCORING PARA EVALUACIÓN DE CRÉDITOS	46
4.1.3. REDUCCIÓN DE LA TASA DE MOROSIDAD EN LOS CRÉDITOS EN UN PERIODO DETERMINADO.....	46
V. DISCUSIÓN	47
VI. CONCLUSIONES.....	48
VII. RECOMENDACIONES.....	49
VIII. LISTA DE REFERENCIAS	50
IX. ANEXOS.....	51
ANEXO N° 01.....	51
ANEXO N° 03.....	53
ANEXO N° 04.....	56
ANEXO N° 05.....	57

ÍNDICE DE TABLAS

TABLA I – CUADRO COMPARATIVO DE MODELOS CREDIT SCORING	18
TABLA II – DISEÑO DE CONTRASTACIÓN DE HIPÓTESIS	20
TABLA III – MOROSIDAD EN CRÉDITOS PERSONALES	21
TABLA IV - INDICADORES	22
TABLA V - TÉCNICAS E INSTRUMENTOS DE RECOLECCIÓN DE DATOS	24
TABLA VI – SELECCIÓN DE LAS COLUMNAS DE LA BASE DE DATOS	33
TABLA VII – VARIABLES INTRODUCIDAS	38
TABLA VIII – PESO DE VARIABLES INTRODUCIDAS	39
TABLA IX – MATRIZ DE CONFUSIÓN	40
TABLA X – PRECISIÓN DEL MODELO	40
TABLA XI – CARTERA DE CRÉDITOS APLICANDO EL MODELO	41
TABLA XII - MATRIZ DE CONSISTENCIA	44

ÍNDICE DE FIGURAS

FIGURA 1. FASES DEL CRISP-DM [11]	15
FIGURA 2. NÚMERO DE CRÉDITOS MOROSOS SEGÚN TIPO DE CRÉDITO	20
FIGURA 3. PROCESO DE OTORGAMIENTO DE CRÉDITOS	26
FIGURA 4. CARTERA DE PRÉSTAMOS EN RIESGO	30
FIGURA 5. VARIABLES ATÍPICAS – HISTORIAL DE INCIDENCIAS	34
FIGURA 6. VARIABLES ATÍPICAS – INGRESO NETO	34
FIGURA 7. VARIABLES ATÍPICAS – ANTIGÜEDAD DEL PRESTATARIO.....	35
FIGURA 8. VARIABLES ATÍPICAS – DEPÓSITOS.....	35
FIGURA 9. VARIABLES ATÍPICAS – MÁXIMO MONTO DE CUOTA	36
FIGURA 10. VARIABLES ATÍPICAS – CRÉDITOS SUPERIORES	36
FIGURA 11. VARIABLES ATÍPICAS – HISTORIAL CREDITICIO	36
FIGURA 12. VARIABLES ATÍPICAS – APORTES.....	37
FIGURA 13. CURVA ROC.....	41
FIGURA 14. ARQUITECTURA DEL PRODUCTO ACREDITABLE.....	42

RESUMEN

La presente investigación titulada “Sistema de información basado en redes neuronales para la predicción de riesgo en el otorgamiento de créditos personales en una cooperativa de ahorro y crédito en el departamento de Lambayeque” surgió de la necesidad de predecir el riesgo en el otorgamiento de créditos personales en la Cooperativa de Ahorro y Crédito del Departamento de Lambayeque, dedicada a servir a sus socios con productos, servicios financieros y beneficios de calidad para el sector agro productivo y micro financiero con transparencia, solidaridad y responsabilidad, “la morosidad crediticia se ha presentado en estos últimos años debido a que para dar solución a este problema se planteó refinanciar y reprogramar los créditos de sus clientes, sin obtener resultados positivos”.

El objetivo general predecir el riesgo en el otorgamiento de créditos personales en la cooperativa de ahorro y crédito; teniendo como propósito identificar las variables para calcular el ingreso del cliente, implementar un scoring para evaluación de créditos, variables que hacen a un cliente potencial moroso y proyectar la reducción de la tasa de morosidad en los créditos.

En conclusión, determinar el ingreso del cliente en un sector en su mayoría agrícola, requiere de la experiencia y conocimientos de cada parte involucrada, es por eso por lo que toda información aporta y es de relevancia para considerar cada variable que influye para determinar la capacidad de pago de un cliente.

PALABRAS CLAVE: Morosidad, redes neuronales, Scoring, riesgo crediticio, cooperativa de ahorro y crédito.

ABSTRACT

The present research entitled "Information system based on neural networks for the prediction of risk in the granting of personal loans in a savings and credit cooperative in the department of Lambayeque" arose from the need to predict the risk in the granting of personal loans in the Savings and Credit Cooperative of the Department of Lambayeque, dedicated to serving its members with products, financial services and quality benefits for the agro-productive and micro-finance sector with transparency, solidarity and responsibility, "the credit default has been presented in in recent years because to solve this problem was raised to refinance and reschedule the loans of their customers, without obtaining positive results.

The general objective is to predict the risk in the granting of personal loans in the savings and credit cooperative; having as purpose to identify the variables to calculate the income of the client, implement a scoring for evaluation of credits, variables that make a potential client delinquent and project the reduction of the delinquency rate in the credits.

In conclusion, determining the client's income in a sector that is mostly agricultural, requires the experience and knowledge of each party involved, that is why all information contributes and is relevant to consider each variable that influences to determine the capacity of a customer's payment.

KEYWORDS: Delinquency, neural networks, Scoring, credit risk, savings and credit cooperative.

I. INTRODUCCIÓN

El crédito en América Latina viene mostrando en el año 2018 un mejor comportamiento frente a lo registrado en 2017, gracias al mejor desempeño de la economía en los diferentes países. De acuerdo con los pronósticos de BBVA RESEARCH [1], la región crecerá 1,3% en 2018 y 2,1% en 2019, pero dicho crecimiento trae consigo la llamada “morosidad crediticia”, que según [2] morosidad hace referencia a demora, falta de puntualidad en los pagos o en el incumplimiento de las obligaciones.

En América Latina, el Perú es un caso particular; según [3] un estudio realizado por IPSOS, en la actualidad alrededor del 41 % de adultos de 18 a 70 años del Perú urbano es cliente de algún banco, caja o financiera. De acuerdo con un estudio realizado por el Banco Mundial en el 2014, solo dos de cada 10 peruanos contaban con una cuenta bancaria. Según un informe de la Asociación de bancos del Perú (ASBANC), el tercer trimestre del 2017 el 35.93 % de los peruanos formaba parte del sistema financiero. Es decir, ascendimos a casi 4 de cada 10 peruanos haciendo uso un producto bancario actualmente.

Según [3] ASBANC hasta el mes de octubre se concluyó que el índice de morosidad promedio en el 2018 alcanzó el 3.14 %. Esta cifra supera el ratio de morosidad bancaria que se cerró en el 2017 con 3.04 %, la cual ya registraba su pico más alto desde el 2005.

La presente tesis denominada “SISTEMA DE INFORMACIÓN BASADO EN REDES NEURONALES PARA LA PREDICCIÓN DE RIESGO EN EL OTORGAMIENTO DE CRÉDITOS PERSONALES EN UNA COOPERATIVA DE AHORRO Y CRÉDITO EN EL DEPARTAMENTO DE LAMBAYEQUE”, se inicia con el propósito de predecir el riesgo en el otorgamiento de créditos personales en la Cooperativa de Ahorro y Crédito del Departamento de Lambayeque, dedicada a servir a sus socios con productos, servicios financieros y beneficios de calidad para el sector agro productivo y micro financiero con transparencia, solidaridad y responsabilidad, “la morosidad crediticia se ha presentado en estos últimos años debido a que para dar solución a este problema se planteó refinanciar y reprogramar los créditos de sus clientes, sin obtener resultados positivos”.

Dado que la mayoría de sus clientes son trabajadores independientes que tienen ingresos variables o dependientes que no están formalizados, causando dificultad para consultar sus fuentes de ingreso (muchas veces superan su capacidad de pago). Otro problema que se presenta en la cooperativa es de no contar con una herramienta que apoye a los analistas a consultar las fuentes de ingreso del solicitante, a detectar atrasos en los pagos, trayendo todo esto como consecuencia cuotas superan la capacidad de pago, clientes no informados sobre el vencimiento

de sus cuotas; cabe resaltar que el sistema de evaluación de solicitudes de crédito que utilizan es obsoleto, ya que no cuenta con los parámetros o variables competentes para poder evaluar al cliente y por consiguiente otorgar o no el crédito, para evitar el aumento de “la morosidad crediticia”.

Ante esta realidad, es importante formular la siguiente pregunta: ¿Cómo se puede predecir el riesgo en el otorgamiento de créditos personales en una cooperativa de ahorro y crédito en el departamento de Lambayeque?

Debido a todo esto se consideró la siguiente hipótesis: “Un sistema de información en el otorgamiento de créditos personales apoyará la predicción de riesgo y beneficiará en una futura reducción de morosidad”.

Se plantea como objetivo general apoyar en la predicción de riesgo en el otorgamiento de créditos personales en la cooperativa de ahorro y crédito; teniendo como propósito identificar las variables para calcular el ingreso del cliente, implementar un scoring para evaluación de créditos, variables que hacen a un cliente potencial moroso y proyectar la reducción de la tasa de morosidad en los créditos.

Desde el punto de vista financiero – económico se justificó puesto que pretende mejorar las colocaciones de sus créditos y así poder reducir la tasa de morosidad a porcentajes tolerables por la entidad.

Desde el punto de vista personal, se justificó la investigación como una oportunidad para adquirir experiencia en nuevos campos de la ingeniería de sistemas que me otorga una ventaja competitiva y me resulta fundamentalmente interesante por la amplia oferta laboral del sistema financiero.

La presente tesis está dividida en nueve (9) capítulos: I) Introducción, II) Marco teórico, III) Metodología, IV) Resultados V) Discusión, VI) Conclusiones, VII) Recomendaciones, VIII) Lista de referencias y IX) Anexos.

En el Capítulo I muestra el contexto con la problemática planteada, en el Capítulo II revisaremos antecedentes y bases teóricas, en el Capítulo III detalla la metodología a utilizar, en el Capítulo IV se presentan los resultados.

En el Capítulo V la discusión, en el Capítulo VI detalla las conclusiones, en el Capítulo VII se muestran las recomendaciones, en el Capítulo VIII se presentan las referencias bibliográficas y en el Capítulo IX se presentan los anexos.

II. MARCO TEÓRICO

2.1. Antecedentes

Se han considerado para esta investigación los siguientes antecedentes:

2.1.1. Antecedentes internacionales

Sánchez [4] desarrolla una simulación del modelo credit scoring aplicadas en el Banco de Santander y de Liberbank, que da respuesta a un problema generalizado en las entidades financieras como es la correcta estimación y valoración del riesgo del cliente o del solicitante del crédito; metodológicamente es econométrica por que se sigue en la implementación práctica de la metodología mencionada. Esta investigación llega a la conclusión que pese a ser un sistema con un elevado coste, la aplicación de este conlleva una significativa reducción de los tiempos de actuación y decisión a la hora de estimar y evaluar el riesgo del crédito.

Lagua [5] desarrolla un modelo credit scoring aplicada a las cooperativas de ahorro y crédito de la ciudad de Ambato – Ecuador, que da respuesta a los problemas más importantes que atraviesan estas organizaciones, que son los niveles altos de morosidad; para ello se hizo uso de una base de datos histórica de los préstamos concedidos de una cooperativa de ahorro y crédito, y mediante el desarrollo del proceso del método Knowledge Discovery in Data bases (KDD), con el uso de una función logarítmica binaria se obtiene los parámetros del modelo. Esta investigación llega a la conclusión de que el modelo de credit scoring, sirve de apoyo al personal del área de negocios y muestra una proyección del comportamiento de pago, ya que el scoring funciona con probabilidades y no con certeza por lo que se debe considerar que el modelo pondera las características de un grupo de prestatarios.

2.1.2. Antecedentes nacionales

Pantoja [6] desarrolla un modelo llamado logit para evaluar el riesgo crediticio de las cajas municipales de ahorro y crédito (CMAC) ya que estas presentan altas tasas de morosidad por la poca innovación en sus metodologías de evaluación crediticia o tecnología crediticia para el otorgamiento de créditos, teniendo como objetivo estimar las probabilidades de incumplimiento del cliente y contribuir a reducir los niveles de riesgo de crédito. Esta investigación llega a la conclusión de que los resultados específicos según la estrategia y metodología, así como etapas necesarias para diseñar un modelo logit y la implementación de esta metodología crediticia avanzada en la CMAC de Huancayo.

Carpio [7] desarrolla un modelo de predicción de la morosidad en el otorgamiento de crédito aplicando la metodología CRIPS-DM como técnica de minería de datos, evaluando cuatro algoritmos scoring: árboles de decisión, redes neuronales, bosques aleatorios y máquinas de soporte vectorial (SVM) en un total de 846 registros de la Caja Rural Los Andes en el año 2015. Todos presentan un grado de certeza por encima del 78%, destacando el de bosques aleatorios con una certeza global de 82%.

Hernández y Quispe [8] se enfocó en la implementación de un Credit Scoring en una Cooperativa de Ahorro y Crédito León XIII de la ciudad de Trujillo, ya que el fundamental problema que se presenta es el índice de morosidad en su cartera de clientes conformado por créditos consumo, pymes e hipotecarios, ha ido incrementándose en los últimos años; aplicándolo a un sistema que permitió disminuir los niveles de morosidad permitiendo mejorar los tiempos de evaluación, además de mejorar la evaluación y las sustentación de los créditos, llegando a la conclusión de que el sistema llegó a ser una herramienta importante para los analistas de crédito de dicha cooperativa.

Rayo [9] propone un modelo de credit scoring para las instituciones de microfinanzas dado que el problema principal se centra en historiales de crédito irregulares, incompleto e información imprecisa; además de contar con una base de datos con poca información económica y financiera; utilizando la fórmula de Basilea II en la normativa SBS. Esta investigación llega a la conclusión con la afirmación del autor, que la integración de BASILEA II en la gestión del riesgo de crédito en IMFS permite discriminación de precio entre clientes derivada de su diferente probabilidad de impago o distinto rating.

2.1.3. Antecedentes locales

Medina y Ulfe [10] desarrollan un modelo scoring que permita predecir el otorgamiento de crédito personal en una Cooperativa de Ahorro y Crédito, con la finalidad de clasificar a los clientes a partir de la probabilidad de default; siendo una investigación aplicada, con propósito predictivo y explicativo, basada en el proceso CRISP-DM para el desarrollo del proyecto. Esta investigación llega a la conclusión de que el credit scoring que se propuso fue una herramienta útil en la evaluación del sujeto a manera de sugerencia de aceptación o rechazo de una futura solicitud de crédito previniendo el sobreendeudamiento e incumplimiento.

2.2. Bases teórico-científicas

2.2.1. Crisp-DM

CRISP-DM (Cross Industry Standard Process for Data Mining) o Modelo de Proceso Estándar para Minería de Datos, es una metodología que normaliza y proporciona una visión general del ciclo de vida de un proyecto de análisis de datos.

El modelo cubre las fases de un proyecto, sus respectivas tareas, y las relaciones entre estas tareas. Todas las relaciones no son posibles de identificar al principio, debido a que las relaciones pueden existir entre cualquier tarea en función de los objetivos, los antecedentes, el interés del usuario, y lo más importante, en los datos. [11].

Las fases no son rígidas, el resultado de cada fase o tarea particular de una fase determinará que debe realizarse a continuación. Siempre se requiere avanzar y retroceder entre las diferentes fases.

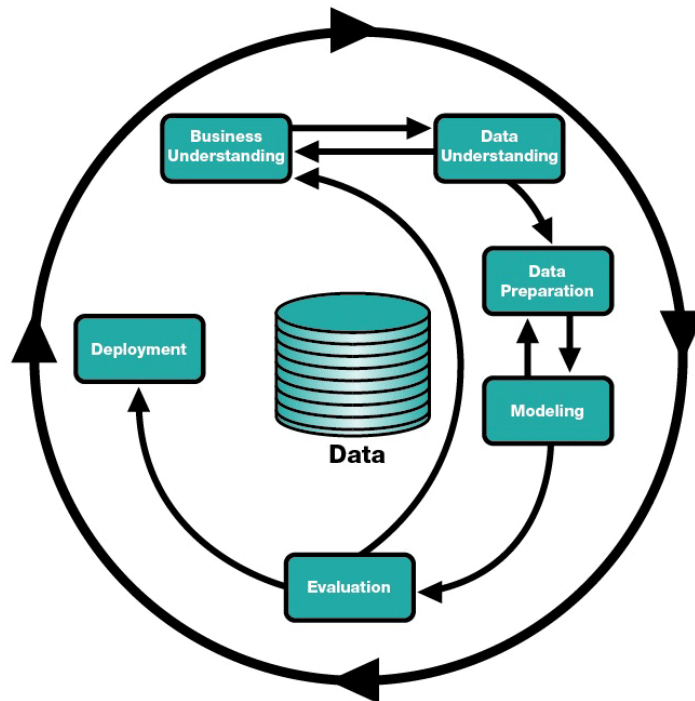


Figura 1. Fases del CRISP-DM [11]

Las flechas indican las dependencias más importantes y frecuentes entre las fases y el círculo externo simboliza la naturaleza cíclica de los proyectos de análisis de datos.

La extracción de datos no termina una vez que se implementa una solución. Las lecciones aprendidas durante el proceso y de la solución implementada puede desencadenar preguntas comerciales nuevas, a menudo más centradas.

En términos generales, el estándar CRIPS-DM es un modelo y una guía, estructurados en seis fases.

2.2.1.1. Fases del Crisp-DM

- 1) Comprensión del negocio, está enfocado en entender los objetivos y requerimientos del proyecto desde una perspectiva de negocio.
- 2) Comprensión de datos, comienza con la recolección inicial de datos para familiarizarse con ellos, así como identificar problemas de calidad de datos y descubrir las primeras relaciones que permitan formar la hipótesis con respecto a la información oculta.
- 3) Preparación de datos, envuelve todas las actividades necesarias para construir el conjunto de datos final (datos que se introducirán en las herramientas de modelado) a partir de los datos brutos iniciales. Incluye la selección de tablas, registros y atributos, así como la transformación y limpieza de datos para las herramientas de modelado.
- 4) Selección y aplicación de varias técnicas de modelado, utilizando las adecuadas según la problemática planteada y el análisis que se desea hacer. Las técnicas seleccionadas permitirán generar modelos de minería de datos, y con ello descubrir patrones de información implícitos en los datos.
- 5) Evaluación detallada del modelo construido, revisando los pasos realizados para su creación y asegurando de que el modelo logre adecuadamente los objetivos de negocio para poder tomar una decisión sobre el uso de los resultados de la minería de datos.
- 6) Despliegue, el conocimiento adquirido deberá ser organizado, documentado y presentado de forma que el cliente pueda utilizarlo. Es importante que el cliente comprenda qué acciones deben llevarse a cabo para poder utilizar realmente el modelo creado.

2.2.2. Credit scoring

El credit scoring son todos los modelos y técnicas que sirven como herramienta para decidir quién, cuanto y en qué condiciones una entidad está sujeto a crédito, determinando el riesgo para cada uno de los clientes principalmente en los créditos de consumo.

Actualmente es utilizada en instituciones financieras y en la mayoría de los bancos para el otorgamiento de crédito [12]. Sin embargo, los modelos no reemplazan el criterio de los analistas, sino es más bien, es una herramienta de apoyo para optimizar su trabajo.

2.2.2.1. Modelos credit scoring

- 1) Modelo Logit: Los modelos de regresión logística permiten calcular la probabilidad que tiene un cliente para pertenecer a uno de los grupos establecidos a priori (no pagador o pagador). La clasificación se realiza de acuerdo con el comportamiento de una serie de variables independientes de cada observación o individuo.
- 2) Árboles de decisión: un árbol de decisión es un conjunto de condiciones o reglas organizadas en una estructura jerárquica, de tal manera que la decisión final se puede determinar siguiendo las condiciones que se cumplen desde la raíz hasta alguna de sus hojas. En el diseño de aplicaciones informáticas, un árbol de decisión indica las acciones a realizar en función del valor de una o varias variables.
- 3) Bosques Aleatorios también conocidos como Random Forest es una combinación de árboles predictores tal que cada árbol depende de los valores de un vector aleatorio probado independientemente y con la misma distribución para cada uno de estos. El algoritmo para inducir un Random Forest fue desarrollado por Leo Breiman y Adele Cutler.
- 4) Redes Neuronales artificiales: Son un conjunto de algoritmos matemáticos inspirados en el razonamiento del cerebro humano. Están compuestos por un número de procesadores simples (neuronas) que trabajan en paralelo sin ningún tipo de control centralizado.

III. METODOLOGÍA

La implementación de un sistema, apoyado en minería de datos e implementado por patrones de comportamiento, apoyará la predicción de riesgo en el otorgamiento de créditos personales de una cooperativa de ahorro y crédito en el departamento de Lambayeque.

TABLA I – Cuadro comparativo de modelos credit scoring

			Ventajas	Inconvenientes
Técnicas Paramétricas	Lineales	Análisis Discriminante	<ul style="list-style-type: none"> ▪Excelente rendimiento ante muestras grandes. ▪Conveniente en la estimación y mantenimiento. 	<ul style="list-style-type: none"> ▪Problemas estadísticos y estimadores ineficientes ▪No arroja probabilidades de impago
		Modelos de Probabilidad Lineal	<ul style="list-style-type: none"> ▪Excelente rendimiento ante grandes muestras. ▪Se sugieren probabilidades de impago. ▪Parámetros interpretables. 	<ul style="list-style-type: none"> ▪Estimadores ineficientes ▪Las probabilidades estimadas podrían quedar fuera del intervalo (0,1)
	No Lineales	Modelos Logit	<ul style="list-style-type: none"> ▪Excelentes propiedades estadísticas y no son estrictos con las hipótesis sobre los datos. ▪Muestran las probabilidades de impago y gran rendimiento 	<ul style="list-style-type: none"> ▪Dificultad de interpretación de los parámetros
		Modelos Probit	<ul style="list-style-type: none"> ▪Excelentes propiedades estadísticas y no son juiciosos con las hipótesis sobre los datos. ▪Muestran las probabilidades de impago 	<ul style="list-style-type: none"> ▪Dificultad de interpretación de los parámetros ▪Proceso de estimación relativamente complicado

Técnicas no paramétricas	Programación Lineal	<ul style="list-style-type: none"> ▪Excelente para gran cantidad de variables ▪Modelo de gran flexibilidad (no requiere una especificación previa del modelo) 	<ul style="list-style-type: none"> ▪No estima parámetros ni probabilidades de impago ▪Difícil comprensión ▪Inexactitud en la predicción
	Redes Neuronales	<ul style="list-style-type: none"> ▪Excelente rendimiento ante muestras pequeñas ▪Modelo de gran flexibilidad (no requiere una especificación previa del modelo) 	<ul style="list-style-type: none"> ▪No estima parámetros ni probabilidades de impago ▪Difícil comprensión
	Árboles de decisiones	<ul style="list-style-type: none"> ▪El mejor rendimiento para muchos autores ▪Modelo de gran flexibilidad (no requiere una especificación previa del modelo) 	<ul style="list-style-type: none"> ▪No estima parámetros ni probabilidades de impago ▪Difícil comprensión

3.1. Tipo y nivel de investigación

3.1.1. Tipo de investigación

Tipo de investigación aplicada porque hace uso de conocimientos previos y los aplica a una realidad concreta.

3.1.2. Nivel de investigación

Nivel de investigación explicativo.

3.2. Diseño de investigación

Variable independiente	Aplicación	Variable dependiente
<u>O1:</u> Nivel de riesgo en el otorgamiento de créditos.	<u>X:</u> Minería de datos	<u>O2:</u> Sistema de información

TABLA II – Diseño de contrastación de hipótesis

3.3. Población, muestra y muestreo

3.3.1. Población

Se recopiló información de 26745 créditos otorgados en el periodo 2013 – 2018 y clasificados según su tipo obteniendo el siguiente gráfico de resultado:

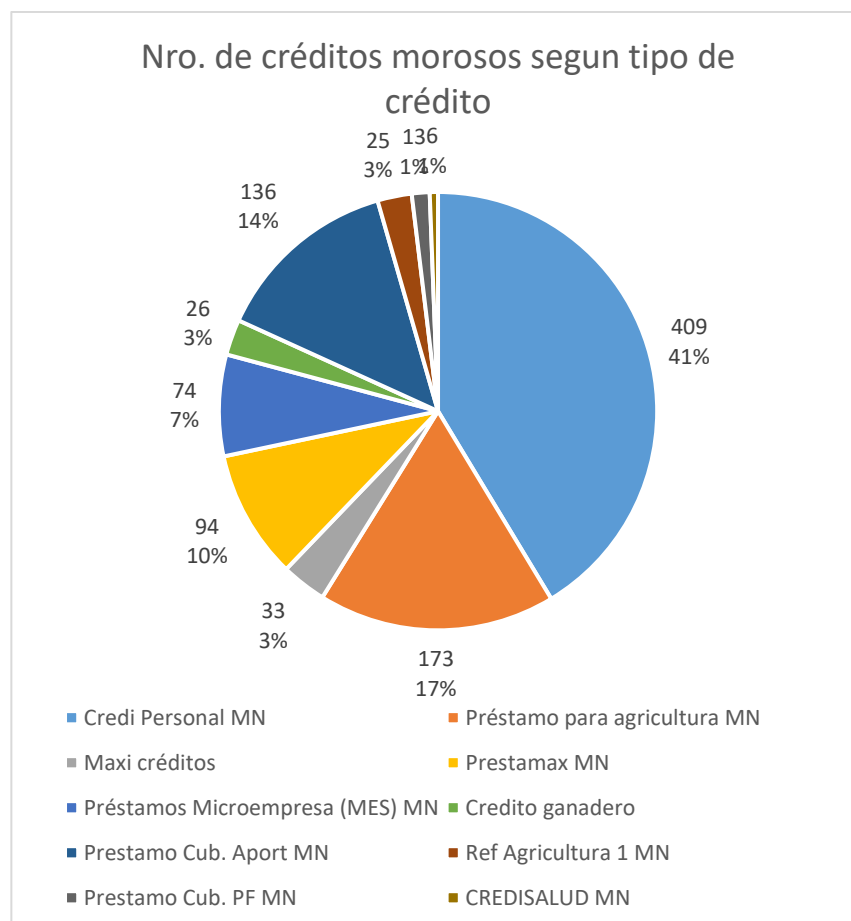


Figura 2. Número de créditos morosos según tipo de crédito

La población para la presente investigación todas aquellas personas que le han sido otorgado un crédito de tipo personal debido a que es donde se encuentra la mayor población morosa y representa el 41% del total de créditos en la cooperativa de ahorro y crédito.

Muestra

Siguiendo con el análisis de 7563 créditos personales, se observa un incremento de la morosidad en los últimos años representados en la siguiente tabla:

2013	2014	2015	2016	2017	2018
6.35%	2.74%	3.17%	5.13%	9.48%	14.12%

TABLA III – Morosidad en créditos personales

La muestra se ha obtenido del periodo 2017 – 2018, con 2639 créditos personales registrados en la base de datos de la cooperativa de ahorro y crédito.

3.4. Criterios de selección

Personas que les han otorgado un crédito de tipo personal en la cooperativa de ahorro y crédito en el periodo 2017 – 2018.

3.5. Operacionalización de variables

Las variables que se han utilizado como elementos básicos en el desarrollo de la hipótesis están identificadas de la siguiente manera:

3.5.1. Variables

3.5.1.1. Variable independiente

Sistema de información.

3.5.1.2. Variable dependiente

Nivel de riesgo en el otorgamiento de créditos personales.

3.5.2. Indicadores (Operacionalización de variables)

TABLA IV - INDICADORES

Objetivo específico	Indicador(es)	Definición conceptual	Unidad de medida	Instrumento	Definición operacional
Identificar las variables para identificar el ingreso del cliente.	Características del cliente	Créditos que presentan incumplimiento en el pago puntual en relación con el total de créditos otorgados	Porcentaje	Reporte de morosidad	Falta de puntualidad o incumplimiento en el pago de una cantidad concedida por la entidad financiera.
Reducción de la tasa de morosidad en los créditos	Precisión del sistema desarrollado	Precisión del sistema para identificar clientes potencialmente morosos.	Porcentaje	Reporte del sistema	Resultado para conseguir la mayor cantidad de aciertos en la predicción de riesgo en el otorgamiento de créditos.
Implementar un scoring para evaluación de créditos	Masculino Femenino	Perfil del cliente que clasifica y define si es un potencial moroso, teniendo en cuenta la ley de protección de datos.		Base de datos	Categoría de género que se utiliza para diferenciar a las personas según su sexo.
	Soltero Casado Conviviente Viudo Divorciado				Condición de la persona según el registro civil en relación de que si cuenta o no con una pareja.
	Menor a 1500 soles. Entre 1500 y 3000 soles. Mayor a 3000 soles.				Remuneración que recibe la persona periódicamente por la realización de su trabajo y es expresado en dinero.
	Vivienda propia Vivienda alquilada				Tipo de lugar en el que habita la persona,

					ocupándola total o parcialmente.
	Tiempo de servicio				Años que viene laborando la persona en su empleo o servicio prestado.
	Desemplead o Empleado Agricultor Profesional Estudiante				Actividad o trabajo que se dedica la persona.
	Garante				Seguridad para el cumplimiento de una obligación de pago.
	Familiares				Personas dentro de la familia que trabajan y aportan con la canasta familiar.

3.6. Técnicas e instrumentos de recolección de datos

A continuación en la siguiente tabla se muestra las técnicas e instrumentos que fueron útiles para la recolección de datos.

TABLA V - Técnicas e instrumentos de recolección de datos

Técnicas	Instrumentos	Elementos de la población	Propósito
Entrevistas	Entrevista Estructurada	Gerencia	Proceso de recolección de datos para comprender el negocio.
Entrevistas	Entrevista semi - Estructurada	Analistas de créditos	Comprensión del proceso de otorgamiento de créditos del área de créditos.
Entrevistas	Entrevista semi - Estructurada	Área de sistemas	Proceso de recolección de la base de datos para la comprensión de los datos.

3.7. Procedimientos

3.7.1. Metodología de desarrollo

A continuación, se mencionan las actividades que se realizaron en cada una de las iteraciones de la metodología a seguir, en este caso CRISP-DM:

1. Fase #1: Compresión del negocio

Como toda organización que compite hoy en día en el mercado financiero, las cooperativas de ahorro y crédito deben poseer una estrategia sustentada en procesos de planeamiento estratégico a fin de generar su sostenibilidad en el tiempo. La “Cooperativa de Ahorro y Crédito” tiene como objetivo lograr un nuevo posicionamiento con la finalidad de incrementar el crecimiento de las principales ratios de la cooperativa y por ende obtener mayor rentabilidad, la misma que se traducirá en mayores beneficios para los socios. Sin embargo los niveles de morosidad reportados a Diciembre del 2017 fueron del 6.44%, este índice debe ser controlado, en el año 2018. La Mora meta proyectada al cierre 2018 es 6.57%. Este punto requiere de especial atención en tres aspectos:

- a) Es un activo improductivo que le resta ingresos a la COOPAC.
- b) Le exige a la COOPAC mayores provisiones que afectan resultados y por lo tanto al patrimonio, por lo cual se va debilitando.
- c) Hay una mora latente en las cuentas refinanciadas y reprogramados, sobre las que se debe poner toda la atención de vigilancia permanente y recuperación ya que de caer en mora nuevamente elevaría el índice de manera inmediata.

Una solución propuesta por el área de sistemas es la creación de Scoring para evaluación de créditos con el objeto de controlar el

incremento de la Mora actual. Provisionalmente, el estudio se considerará un éxito si:

- La solución implementada es de utilidad y dentro del presupuesto.
- La proyección de morosidad total al 2018 es inferior al 6.57%.
- Se mejora la calidad de sus colocaciones de créditos.

Determinar la estructura de la organización

El proceso de otorgamiento de créditos de la Cooperativa de Ahorro y Crédito es el siguiente:

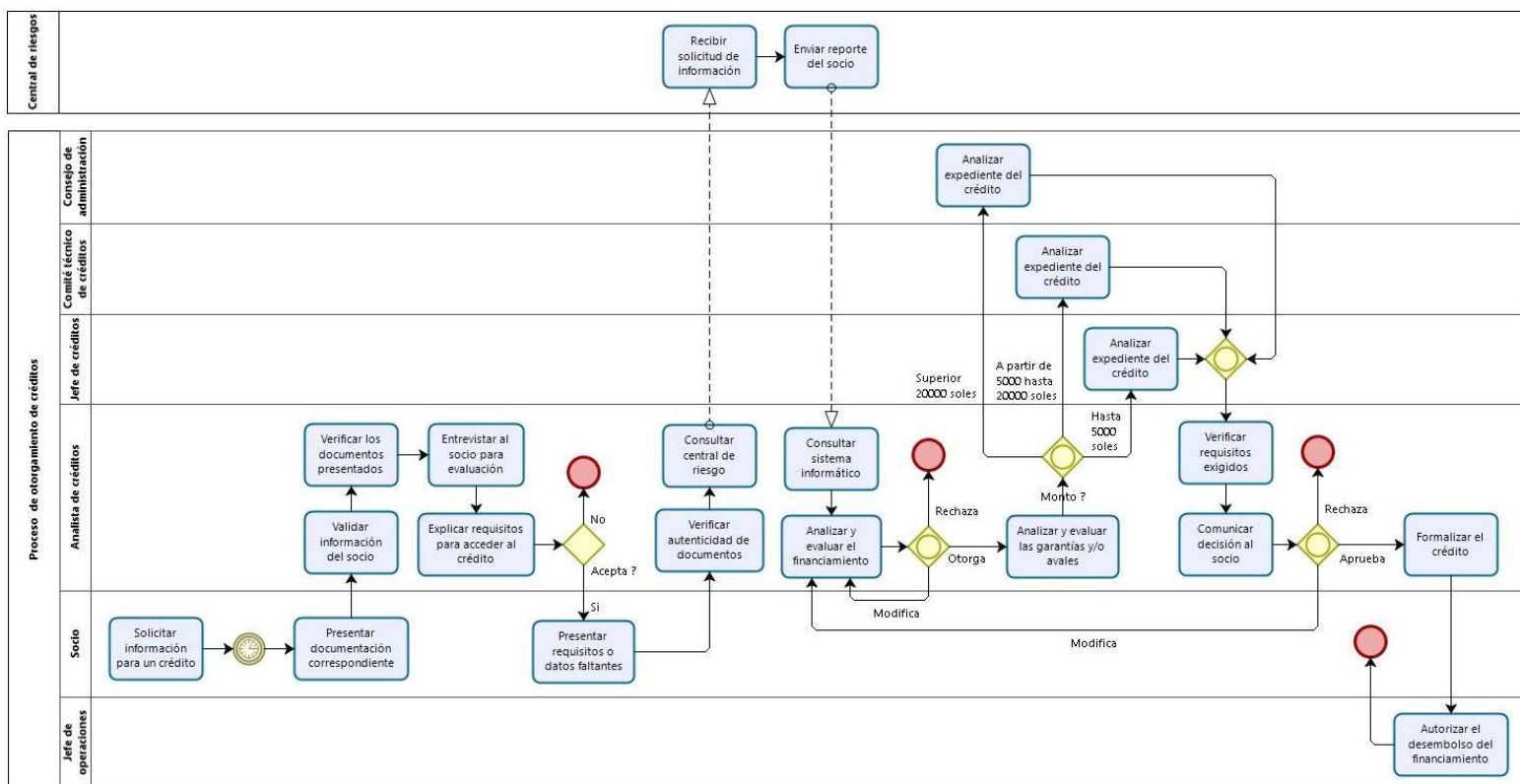


Figura 3. Proceso de otorgamiento de créditos

1. **Consejo de Administración:** Es el órgano de gobierno responsable del funcionamiento administrativo de la cooperativa. Está compuesto por siete miembros a cargo de la presidenta: Fanny Soledad Millones Chancafe. Entre sus funciones generales destaca:

- Aprobar los planes y presupuestos anuales de ingresos y egresos de la Cooperativa, sus modificaciones y transferencias.

2. **Gerencia General:** Es el órgano de máxima dirección ejecutiva; a ella compete la responsabilidad del desarrollo de las actividades y operaciones de la Cooperativa. Cuenta con el servicio del Gerente General Prof. Segundo Atilano Castillo Espinoza. Tiene como responsabilidad, la más importante:

- Supervisar el cumplimiento de los procesos y procedimientos relacionados con las operaciones de crédito.

3. **Unidad de Riesgos:** Es la unidad orgánica de prevención, encargada de la identificación, tratamiento y evaluación de todos los riesgos que enfrenta la Cooperativa. Cuenta con el servicio de un Jefe de Unidad de Riesgos C.P.C Carmen R. Galloso Coronado. Según la Res. SBS-3780-2011, sobre Reglamento de la Gestión de Riesgo de Créditos, una de las principales funciones de la Unidad de Riesgo es:

- Monitorear el riesgo de crédito y el mantenimiento de éste dentro del nivel de tolerancia de la empresa.

4. **Asesoría Legal:** Es la unidad orgánica asesora, encargada de velar por el asesoramiento especializado en derecho en materia cooperativa, civil, comercial, penal, laboral, administrativa y

tributaria. Cuenta con el servicio de un Jefe del Área de Asesoría Legal Abg. Francisco Aurelio Baca Castillo.

- Interpretar leyes, decretos, normas, disposiciones que sean aplicables en cualquier aspecto de las operaciones de rutina de la cooperativa.

5. Área de Tecnologías de Información: Es la unidad orgánica de apoyo, que se encarga del diseño, desarrollo de sistemas, programación implementación y mantenimiento de los elementos que constituyen la infraestructura informática de la Cooperativa. Cuenta con el servicio de un Jefe de Tecnologías de Información – Sistemas Ing. Fernando Seclen Eneque, y un Asistente de Ayuda & Control de Calidad a Sistemas Ing. Renato Pérez Fiestas. Tienen por función:

- Evaluar e informar a la Gerencia General sobre las nuevas tecnologías existentes y cómo estos contribuirían al cumplimiento de los objetivos de la cooperativa.

6. Área de Créditos: Se encarga de administrar y controlar los procesos de admisión, evaluación, aprobación y ejecución eficiente de los créditos de acuerdo al Reglamento General de Créditos; así como, de ejecutar y cumplir con las políticas de créditos de la Cooperativa. Cuenta con el servicio de un Jefe de Área de Créditos Luis Eduardo Gil Núñez, Analistas de Créditos quienes ofertan los productos financieros para Agricultura, Ganadería, Créditos de Consumo (Personal, Convenios, Cubiertos), Microempresa, Pyme u otros productos nuevos. El objetivo del puesto es:

- Planificar, dirigir, coordinar y controlar las actividades del Área de Crédito de la Cooperativa en sus diferentes modalidades, gestionar las recuperaciones de los créditos, utilizando criterios definidos en las políticas, reglamentos

y procedimientos, evaluar los productos y servicios crediticios procurando obtener la máxima rentabilidad al menor riesgo posible en las colocaciones.

7. Analista de Créditos: Ejecutan los procesos relacionados con la Gestión de Créditos, según el Reglamento General de Créditos. Es el encargado de:

- Completar adecuadamente los formatos y formularios de crédito para cada socio o potencial socio.

8. Comité de Riesgo: Es un órgano constituido en el marco de la gestión integral de Riesgo de la cooperativa, cuyo funcionamiento responde a la necesidad de evaluar y proponer decisiones que atañen a los riesgos significativos a los que está expuesta la cooperativa, y que tienen que ver con su operatividad. Define el nivel de tolerancia y el grado de exposición al riesgo que la cooperativa está dispuesta a asumir en el desarrollo de las actividades.

9. Comité de Crédito: Es un órgano eminentemente técnico y está a cargo de resolver las solicitudes de crédito de los socios y proponer al Consejo de Administración todas las normas y Reglamentos crediticios. Se encarga de entrevistar, si es necesario, a los Socios que soliciten créditos a fin de contar con los mejores elementos de juicio para la aprobación del Crédito.

10. Comité de Morosidad: Tiene responsabilidad de resolver la situación de créditos de los socios que han incumplido sus pagos puntualmente, por lo que este deberá realizar un seguimiento consciente, veraz y efectivo para mantener una cartera sana cumpliendo en todo momento los Reglamentos y Normas vigentes.

Describir el área problemática

El crédito se define como la capacidad que tienen los socios para obtener dinero prestado, respaldado por una adecuada garantía y con el compromiso de pagar el monto recibido así como los intereses y gastos generados en un plazo contractual previo derivado de un acuerdo establecido entre el socio y la Cooperativa.

El área de Créditos cumple con las funciones asignadas y dispone de su propia toma de decisiones enmarcadas dentro de su ámbito de ejecución establecido en el Reglamento General de Créditos, cabe recalcar que el otorgamiento de créditos a los socios se efectuará teniendo en cuenta la disponibilidad de recursos financieros que cuenta la Cooperativa.

Cartera de Prestamos en Riesgo	0		0	
Total Moroso-antes de depurar	840,962.32	6.44%	1,084,648.47	6.57%
Saldo-Estimaciones(antes de depurar)	1,109,352.00		1,553,092.80	
Total Bruto	13,063,183.71	100.00%	16,500,000.00	100.00%

Figura 4. Cartera de préstamos en riesgo

El cuadro anterior nos muestra las proyecciones plasmadas en el Plan Estratégico Empresarial y Presupuesto 2018, donde la morosidad proyectada al cierre del 2018 es de 6.57% equivalente a S/ 1,084,648.47. La creación de un Scoring para evaluación de créditos apoyaría la reducción de morosidad y otorgaría utilidades así como más beneficios para los socios. Actualmente, la cooperativa no está utilizando ninguna técnica de minería de datos.

2. Fase #2: Compresión de los datos

Para la evaluación de créditos implementando la solución de un Scoring, se utiliza varios orígenes de datos, incluyendo:

Expedientes de créditos. Los expedientes contienen toda información de los créditos otorgados, que serán tomadas de la base de datos de créditos, sin embargo, la evaluación de los socios así como su capacidad de pago se tomarán de estos expedientes como parte del proceso de preparación de los datos.

Adquisición de los datos. Si se le otorga un crédito a un socio, se guarda toda la información relativa al crédito.

Base de datos de cuentas. Las cuentas de ahorros a plazo y de aportaciones son de gran utilidad cuando se solicita un crédito. Sirven como encaje al momento de realizar un préstamo. Es necesario tener en cuenta que las aportaciones es la contribución obligatoria del socio como cuota que se depositan en su cuenta individual.

Base de datos de socios. Esta base de datos contiene información de los socios asociados a la cooperativa. Estos registros no son fiables, ya que esta información solo es tomada cuando se presenta una solicitud de admisión. Es necesario actualizar la información tomada de los expedientes de créditos.

La empresa otorga todas las facilidades y recursos para implementar la solución y sea de utilidad.

2.1. Descripción de los datos

Vigentes al 25 de Enero del 2019, se cuenta con 7968 socios hábiles en la cooperativa. De los cuales a 4359 se le han otorgado 27279 créditos en los últimos 5 años.

Las columnas a considerar se obtienen de las tablas socios y cuentas, y se convertirán en variables con los que se evaluará a un cliente para el desarrollo del modelo y son los siguientes:

Documentos del socio:

- Datos personales, familiares, económicos, de producción, legal y social del socio o trabajador y fiador solidario.
- DNI, legales, garantías, registrales, y otros.

- Domiciliaria, tanto del titular así como de su garante de ser el caso y del inmueble o bien mueble que va a ser sujeto de garantía.

Evaluación del Socio:

- Obtener su calificación en el sistema interno (SIAF) y financiero (reporte central de riesgo).
- Obtener su nivel de endeudamiento directo e indirecto en el sistema financiero (reporte central de riesgo), en monto y por número de instituciones.
- Levantar información sobre los activos de su negocio personal.
- Indagar con el Socio sobre las deudas pendientes con proveedores, a fin de cruzarlas con la posición de los activos.
- Determinar el costo de ventas y/o producción contrastándolo con el promedio del sector.
- Determinar la carga familiar (número de dependientes y sus edades) y su correspondiente nivel de gasto (alimentación, vivienda, educación, salud, transporte, servicios, otros).
- Elaborar los estados financieros básicos en función a la información levantada (balance, estados de resultados y flujo de caja mayor a 12 meses).
- Calcular los principales indicadores financieros y comentarios.

2.2. Selección de las columnas de la base de datos

Se tomaron en consideración las siguientes columnas y fueron clasificadas según su tipo de variable: Numérica, dicotómica y categórica.

TABLA VI – Selección de las columnas de la base de datos

Columna	Descripción	Tipo Variable
Sexo	Género del prestatario.	Dicotómica. (0) Hombre, (1) Mujer
Edad	En el momento de la solicitud del crédito.	Numérica.
Estado Civil	Estado civil.	Dicotómica. (0) Soltero, (1) Unidad Familiar
Carga Familiar	Nro. de miembros que dependen del prestatario.	Numérica.
Fondo Mortuario	Fondo mortuario.	Dicotómica. (0) No, (1) Si
Grupo	Nro. de miembros que pertenecen al grupo del prestatario.	Numérica.
Antiguo	Tiempo del prestatario como cliente en la entidad.	Numérica.
Condición	Destino del crédito	Catagórica. (0) Consumo, (1) Agricultura, (2) Comercio, (3) Hipotecario
Monto	Importe del crédito	Numérica.
Cuotas	Nro. de cuotas mensuales del crédito solicitado.	Numérica.
Historial	Nro. de créditos concedidos con anterioridad.	Numérica.
Tipo Operación	Tipo de operación del préstamo.	Dicotómica. (0) Con incidencias (Refinanciado, ampliado), (1) Normal
Créditos Simultáneos	Nro. de créditos simultáneos del prestatario.	Numérica.
CT	Nro. total de cuotas pagadas en historial crediticio.	Numérica.
CA	Nro. de cuotas con atraso.	Numérica.
CM	Nro. de cuotas incurridas en morosidad.	Numérica.
MY	Nro. de días de la mayor mora.	Numérica.
Aportes	Importe de aportes del prestatario.	Numérica.
Depósitos	Importe de ahorros del prestatario.	Numérica.
Puntualidad	Nro. total de aportes pagados con puntualidad.	Numérica.
Ingreso Neto	Ingreso neto del prestatario.	Numérica.

3. Fase #3: Preparación de los datos

3.1 Limpieza de datos

En esta fase se excluyen todos los créditos que presentan valores atípicos, representados en las siguientes imágenes. Donde: los puntos verdes son los créditos morosos y los azules representan los créditos normales, es decir, no presentan ningún tipo de inconvenientes.

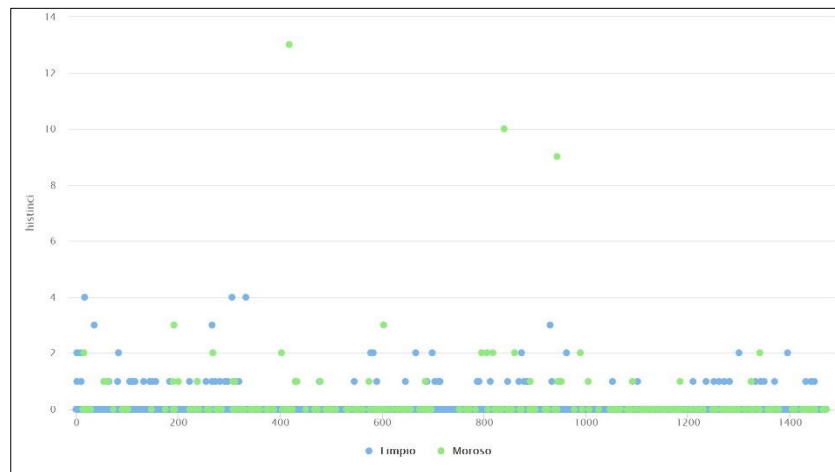


Figura 5. Variables atípicas – Historial de incidencias

El historial de incidencias es el número de créditos que están o han estado con complicaciones en el cumplimiento puntal de su crédito ya sea por ser aplazado, refinanciado, reprogramado o moroso. Se establece un límite de 4, donde si este indicador es superado, el cliente va a ser considerado moroso.

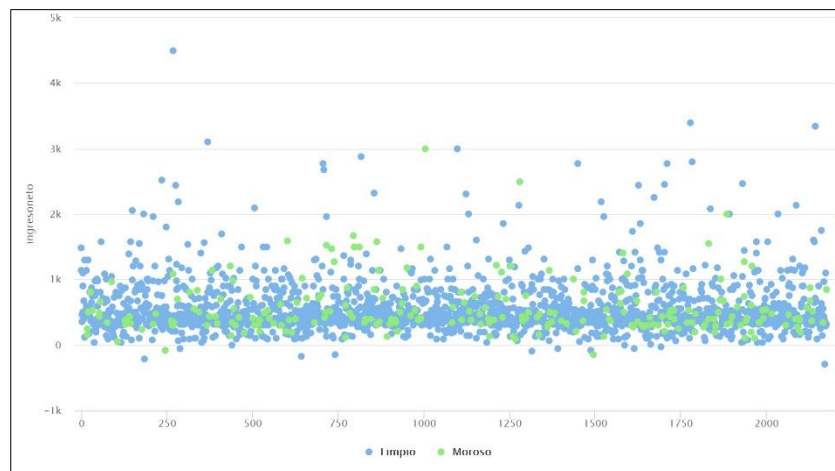


Figura 6. Variables atípicas – Ingreso neto

El ingreso neto del cliente, es el resultado de sus ingresos menos sus egresos, considerando la carga familiar por la canasta básica familiar. Se establece un límite de 2000 soles.

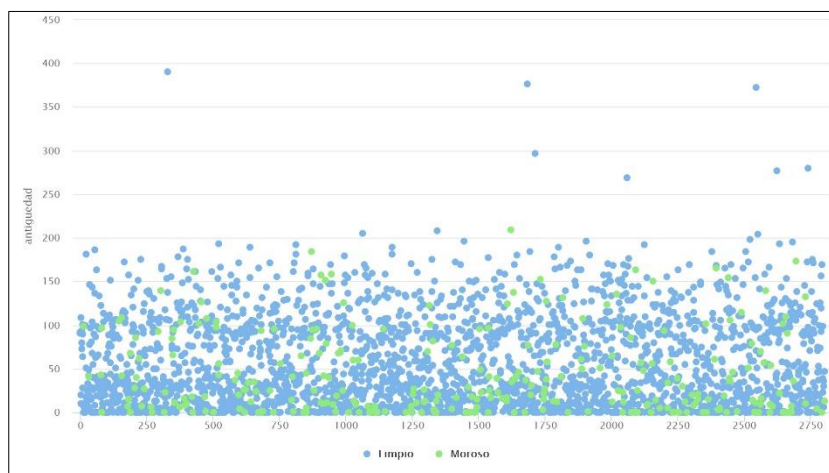


Figura 7. Variables atípicas – Antigüedad del prestatario

La variable antigüedad es el tiempo del prestatario como cliente de la entidad y se limitará a 10 años.

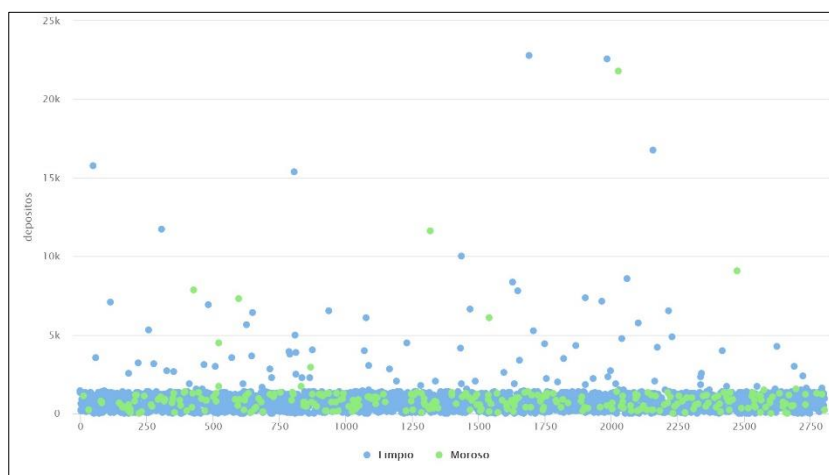


Figura 8. Variables atípicas – Depósitos

Los depósitos ya sea por cuentas de ahorros o a plazo fijo no será considerado como variable en el modelo, sin embargo, se observa que los clientes que cuentan con depósitos son menos morosos y es una variable por considerar para que el analista pueda tomar una decisión.

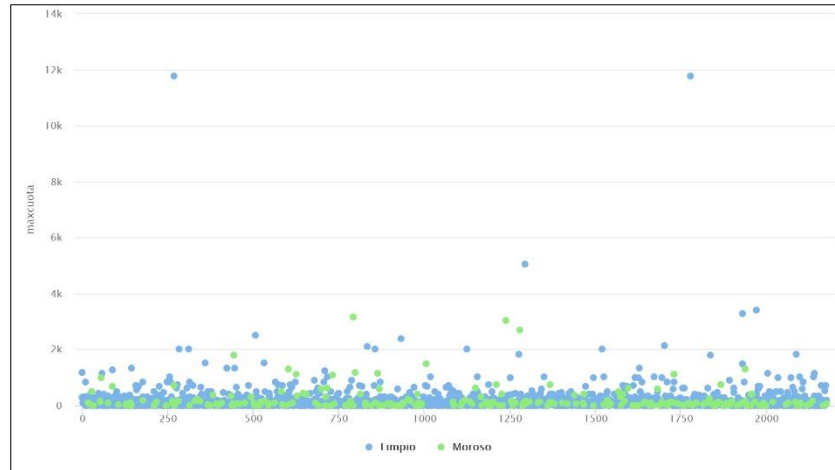


Figura 9. Variables atípicas – Máximo monto de cuota

Se excluye 3 créditos que superan los 5000 soles como monto de cuota.

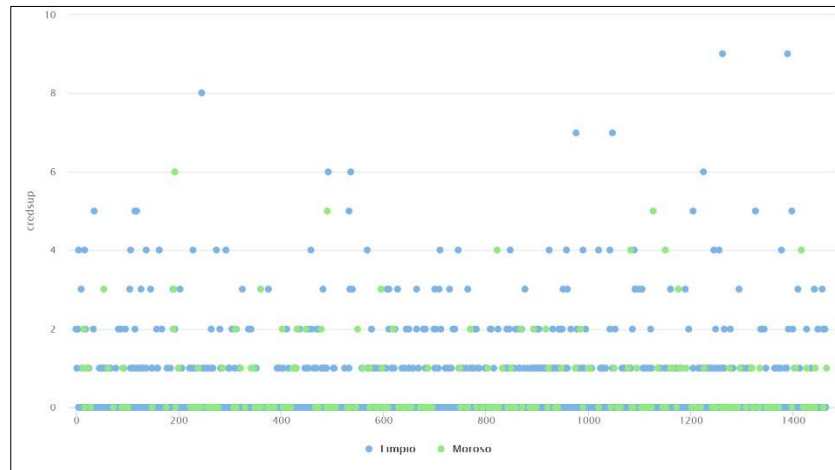


Figura 10. Variables atípicas – Créditos superiores

Se limita a 5 el número de créditos superiores a la cantidad solicitada.

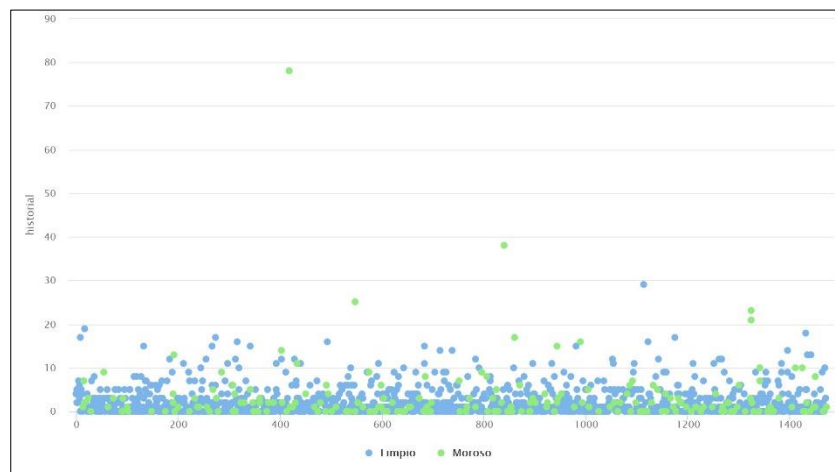


Figura 11. Variables atípicas – Historial crediticio

El cliente que supera los 20 créditos en los últimos 5 años es considerado moroso.

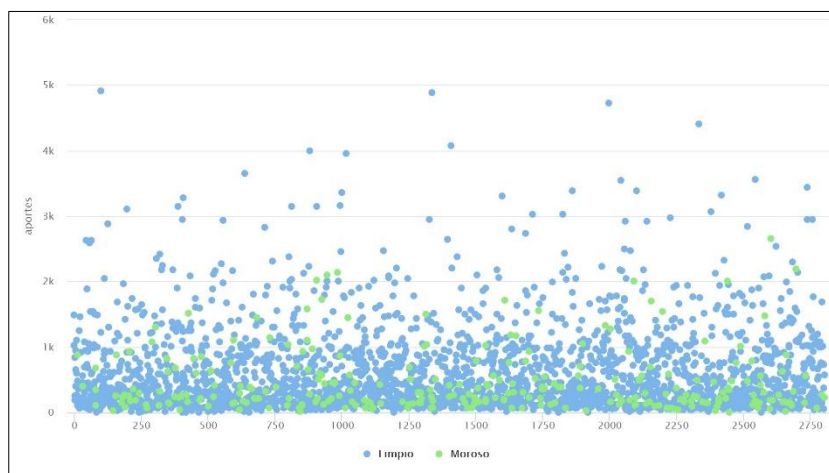


Figura 12. Variables atípicas – Aportes

Los aportes del cliente, es una contribución obligatoria mensual equivalente a 10 soles que se depositan en una cuenta individual por cada cliente. Se establece un límite de 1200 soles, y cuando se supere este monto se convertirá en una variable para que el analista pueda tomar una decisión.

4. Fase #4: Selección de la técnica de modelado

Deep Learning

O Aprendizaje profundo es una actualización innovadora del enfoque de redes neuronales artificiales. Descubre una estructura compleja en grandes conjuntos de datos utilizando el algoritmo de propagación hacia atrás para indicar cómo una máquina debe cambiar sus parámetros internos que se utilizan para calcular la representación en cada capa a partir de la representación en la capa anterior.

Muchos problemas supervisados y no supervisados se han resuelto con éxito mediante un aprendizaje profundo. La calificación crediticia es uno de los desafíos importantes para el proceso de toma de decisiones de las instituciones crediticias.

Se seleccionó este modelo debido a su flexibilidad y exactitud al manejar muestras pequeñas, las mismas se detallan en la *Tabla N.º 01*.

Construcción del modelo

En base a las columnas seleccionadas (detalladas en la *Tabla N.º 06*) y posterior a la limpieza de datos, se han introducido al modelo las siguientes variables.

TABLA VII – Variables introducidas

	Descripción	Tipo de variable
puntualidad	Nro. de aportes pagados puntualmente en los últimos 6 meses.	Numérica
histinci	Nro. de incidencias en su historial crediticio	Numérica
ctotales	Total de cuotas pagadas.	Numérica
migrupo	Nro. de miembros avales.	Numérica
ccanceladas	Nro. de cuotas canceladas a la fecha.	Numérica
aportes	Monto total de aportaciones.	Numérica
antiguedad	Tiempo del prestatario en la entidad.	Numérica
cuotas30	Nro. de cuotas en 30 días	Numérica
credsimul	Nro. de créditos simultáneos.	Numérica
estadocivil	Estado civil.	Catagórica
histulti	Nro. de créditos en el último año.	Numérica
edad	Edad del cliente	Numérica
maxmora	Máximo nro. de días incurrido en morosidad	Numérica
catrasadas	Nro. de cuotas atrasadas en su historial.	Numérica
mediamora	Media de los días de mora	Numérica
histultiinc	Nro. de créditos con incidencias en el último año.	Numérica
cmorosas	Nro. de cuotas incurridas en mora.	Numérica
maxcredito	Máximo crédito otorgado al cliente.	Numérica
credsupinc	Nro. de créditos superiores con incidencias.	Numérica
fultincidencia	Nro. de meses desde la última incidencia.	Numérica
monto	Monto solicitado.	Numérica
ingresoneto	Ingreso neto del cliente	Numérica
credsup	Nro. de créditos superiores.	Numérica

Implementación del modelo

Se utilizó la librería H2O con los algoritmos MLP para resolver patrones de comportamiento a causa de que los créditos introducidos en el modelo son a base de su historial crediticio.

Asimismo, se determina el peso que se presenta en cada variable y representado en la siguiente tabla.

TABLA VIII – Peso de variables introducidas

Variable	Peso
puntualidad	0.1680
histinci	0.1521
ctotales	0.1418
migrupo	0.1121
ccanceladas	0.1013
aportes	0.0981
antiguedad	0.0650
cuotas30	0.0621
credsimul	0.0592
estadocivil	0.0546
histulti	0.0436
edad	0.0413
maxmora	0.0410
catrasadas	0.0383
mediamora	0.0381
histultiinc	0.0368
cmorosas	0.0366
maxcredito	0.0324
credsupinc	0.0250
fultincidencia	0.0190
monto	0.0186
ingresoneto	0.0176
credsup	0.0145

5. Fase #5: Evaluación del modelo

La matriz de confusión generada por el aprendizaje profundo es el siguiente:

TABLA IX – Matriz de confusión

		Real	
		Limpio	Moroso
Predicción	Limpio	714	53
	Moroso	55	69

Donde:

TABLA X – Precisión del modelo

Precisión Global	0.8788
Error Global	0.1212
Precisión Positiva	0.9285
Precisión Negativa	0.5656
Falsos Positivos	0.4344
Falsos Negativos	0.0715

Los resultados del aprendizaje profundo son buenos, se puede interpretar de la matriz de confusión que la precisión global es de un 87.88% y un error 12.12%. En cuanto a la precisión de los clientes limpios es de 92.85% y de los morosos un 56.56%. El modelo predice un 43.44% como cliente limpio a un cliente moroso, sin embargo solo un 7.15% a un cliente moroso siendo en este caso limpio.

Realizamos el análisis del modelo mediante el uso de la Curva ROC, en la que podemos apreciar que el modelo tiene un grado de certeza del 80.4% lo que es muy bueno y aceptable porque tiene un valor superior al 70%.

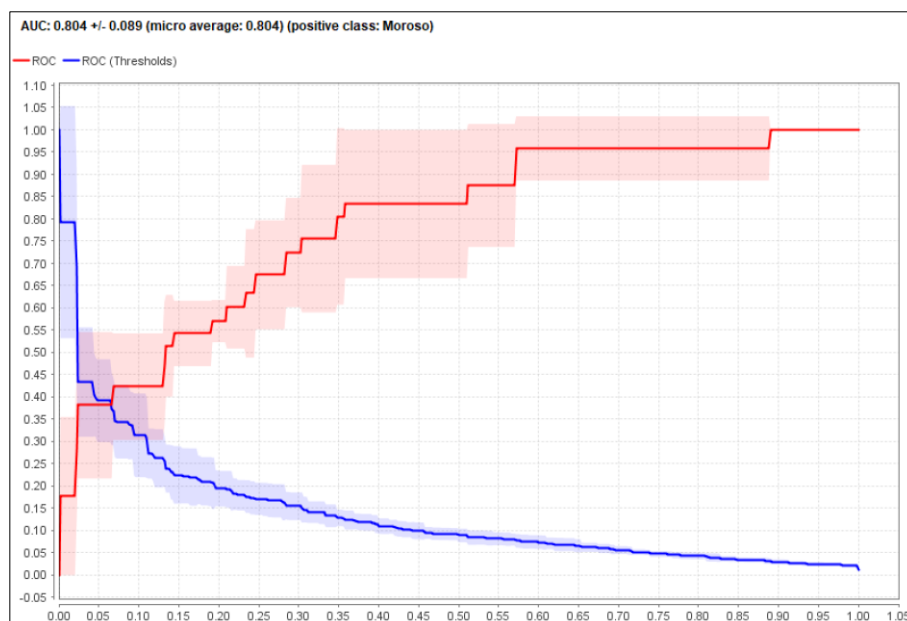


Figura 13. Curva ROC

Sin embargo, es necesario realizar una validación para ver la reducción de la morosidad en la cartera de créditos. Tratamos de ver dos aspectos muy importantes:

- a) **Total bruto:** Total de préstamos otorgados en los 2 últimos años.
- b) **Total moroso:** Suma del monto total de los préstamos otorgados por el modelo que entran en morosidad.

TABLA XI – Cartera de créditos aplicando el modelo

	Bruto	Moroso	
Real	S/2,326,727.49	S/369,082.47	15.86%
Predicción	S/1,843,121.23	S/88,834.19	4.82%

Aquí los resultados son excelentes, muestran una reducción del 329.12% en base a créditos reales frente a la predicción generada en el modelo, da como resultado un porcentaje del 4.82%, muy por debajo de lo proyectado por la entidad.

3.7.2. Análisis de riesgos

El análisis de riesgos en el desarrollo de la presente tesis se efectuó con la finalidad de identificar las fases, entregables y objetivos afectados durante desarrollo de la presente tesis, las mismas de detallan en el *Anexo N.º 02*.

3.7.3. Producto acreditable

1. Arquitectura

De diseñó una arquitectura idónea para el funcionamiento del sistema basado en tecnologías Microsoft y en software abierto como Python y H2O, el cual se detalla en el siguiente gráfico.

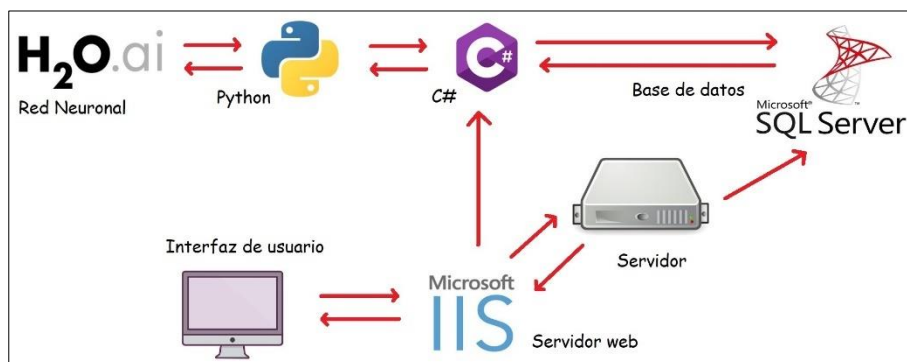


Figura 14. Arquitectura del producto acreditable

2. Infraestructura tecnológica

Considerando la arquitectura anteriormente descrita, se definen las características de cada uno de sus componentes:

H2O: Plataforma de código abierto especializado en machine learning y en inteligencia artificial. Se puede ejecutar desde el paquete estadístico R, Python y otros.

Python: Es un lenguaje de programación en consola que nos permitirá la comunicación entre el modelo y la aplicación.

C#: Lenguaje de programación orientada a objetos en donde se establece la conexión con la base de datos y el servidor web.

Servidor: Es el encargado de atender las peticiones de los usuarios.

SQL Server: Sistema gestor de base de datos relaciones que alberga toda la información de la entidad.

IIS: Servidor Web.

3.7.4. Manual de usuario

Se elaboró un manual de usuario con la finalidad de ayudar a los usuarios en el uso del sistema para poder ingresar el monto solicitado y el número de cuotas que se implementó, la cual se muestra en el *Anexo N.º 03*.

3.8. Matriz de consistencia

TABLA XII - MATRIZ DE CONSISTENCIA

PROBLEMA	OBJETIVOS	HIPÓTESIS	VARIABLES
FORMULACIÓN DEL PROBLEMA	OBJETIVO GENERAL	HIPÓTESIS	VARIABLES DE ESTUDIO
¿Cómo se puede apoyar en la predicción de riesgo en el otorgamiento de créditos personales en una cooperativa de ahorro y crédito en el distrito de Lagunas?	Apoyar en la predicción de riesgo en el otorgamiento de créditos personales en una cooperativa de ahorro y crédito.	Un sistema de información en el otorgamiento de créditos personales apoyará la predicción de riesgo y beneficiará en una futura toma de decisión.	VARIABLE INDEPENDIENTE
			Nivel de riesgo en el otorgamiento de créditos personales
			VARIABLE DEPENDIENTE
			Sistema de información
<u>OBJETIVOS ESPECÍFICOS</u>	<u>DESCRIPCIÓN DEL LOGRO DE LOS OBJ. ESPECÍFICOS</u>		<u>INDICADORES</u>
Identificar las variables para calcular el ingreso del cliente			Características del cliente
Implementar un scoring para evaluación de créditos			Aciertos del modelo desarrollado
Reducción de la tasa de morosidad en un periodo determinado.			Aciertos del modelo desarrollado

3.9. Consideraciones éticas

A continuación, se listan los aspectos que se han considerado para la protección y bienestar de los participantes de esta investigación, en este caso la política de protección de datos personales enmarcado en la vigente Ley 29733 tiene como objetivo proteger los datos de las personas naturales. Los nombres y apellidos se obviarán, la edad o ingreso económico serán clasificados; así como cualquier información personal que las identifica o las hace identificables, garantizando su seguridad.

- ✓ Aplicación de técnicas de recolección de datos: Encuestas, entrevistas, etc.
- ✓ Seguridad de la información.
- ✓ Protección de contraseñas y datos registrados.
- ✓ Resguardo de los datos y secreto de la información.

IV.RESULTADOS

4.1. En base a los objetivos de la investigación

4.1.1. Identificar las variables para determinar el ingreso del cliente

En base a los antecedentes se determinó las variables que influyen en el otorgamiento de créditos personales con historial crediticio mostrado en el *Tabla N.º 06*.

Así mismo, se identificó con el apoyo de los analistas y aceptado por el jefe de créditos, las variables que serán consideradas en el mantenimiento y actualización del sistema, puesto que actualmente no se registran los suficientes datos. Estas variables están adecuadas a la realidad problemática y al entorno en el cual trabajan, siendo de prioridad la determinación los ingresos y egresos del cliente, contemplados en el *Anexo N.º 01*.

4.1.2. Implementar un scoring para evaluación de créditos

El sistema desarrollado mejorará las colocaciones del crédito de acuerdo con los resultados mostrados en la *FASE#5 Evaluación del Modelo*. Ha sido implementado de acuerdo con normativas de la entidad y bajo supervisión del Área de TI, donde se emitió una constancia mostrado en el *Anexo N.º 05*.

4.1.3. Reducción de la tasa de morosidad en los créditos en un periodo determinado

Se establece como periodo los primeros meses de implementación del sistema, sin embargo, a la fecha no hay resultados confiables, no obstante, la entidad afirma que el sistema ha mejorado las colocaciones y se espera que en próximo mes se obtenga los primeros resultados.

V. DISCUSIÓN

En el siguiente capítulo se realizó un análisis de los resultados del Sistema de información basado en redes neuronales para la predicción de riesgo en el otorgamiento de créditos personales en una cooperativa de ahorro y crédito en el departamento de Lambayeque. Se realizó una evaluación, en el que se midieron los siguientes indicadores:

Indicador 1: Características del cliente.

Para cumplir con este indicador se procedió a una recolección de datos: variables consideradas por los antecedentes, variables propias del mercado, bases de datos de la cooperativa y con el apoyo de los analistas de créditos y jefe de créditos el cual dio su aprobación adjunta en el Anexo *N.º 01*.

Indicador 2: Aciertos del modelo desarrollado.

Para cumplir con este indicador se probó el modelo con la matriz de confusión detallada en la *TABLA IX*, el cual dio más del 50% en aciertos para potenciales moros. Sin embargo en la cartera de créditos detallado en la *TABLA XI* se presenta una reducción de más del 250% lo que supone un porcentaje de morosidad aceptable para la entidad ya que se encontraría en un 6% frente a lo actual.

VI. CONCLUSIONES

- 1.** Las redes neuronales son funcionales para hallar patrones de datos y de difícil comprensión como es el caso del riesgo crediticio. Es de relevancia comprender como actúan estos modelos y aplicarlos a los diferentes campos de la computación. En nuestro caso en particular la reducción de la morosidad supone un gran avance para la entidad, puesto que las mejores colocaciones de créditos permiten crecer acorde al mercado nacional y obtener menos riesgo al momento de otorgar o no un crédito personal.

- 2.** Determinar el ingreso del cliente en un sector en su mayoría agrícola, requiere de la experiencia y conocimientos de cada parte involucrada, es por eso por lo que toda información aporta y es de relevancia para considerar cada variable que influye para determinar la capacidad de pago de un cliente.

- 3.** El modelo de credit scoring es una herramienta complementaria al proceso de otorgamiento de crédito, apoya al personal del área de negocios y da a conocer un panorama del futuro comportamiento de pago, supone la mejora y la capacidad de otorgar crédito en toda entidad financiera, es por ello por lo que las principales entidades dedican tiempo e inversión.

VII. RECOMENDACIONES

1. Se recomienda el ingreso de las nuevas variables, ya que identificadas no hace dependencia a su historial crediticio y se concentra en conocer mucho mas al cliente, es por eso por lo que cada parte comprometida sea consciente de los valores ingresados y no mostrar alteraciones en la información.
2. Se recomienda implementar controles para determinar las fuentes de ingreso, en especial a los clientes que son independientes porque están susceptibles a dejar de percibir ingresos.
3. Se recomienda implementar un sistema de estado de cuenta, apoyaría a los clientes a estar comprometidos con la entidad y constantemente alertados del vencimiento de sus cuotas o del pago puntal de sus aportaciones que les permitirá recaudar activos para la entidad.

VIII. LISTA DE REFERENCIAS

- [1] M. Garzón, «BBVA Continental,» 2018. [En línea]. Available: <https://www.bbva.com/es/el-credito-en-america-latina-crece-a-mejor-ritmo-que-la-economia/>. [Último acceso: 19 04 2019].
- [2] R. A. Española, Española, Real Academia, 2019.
- [3] E. Bobadilla, «Gestión,» Witthaya Prasongsin, 25 03 2019. [En línea]. Available: <https://gestion.pe/publireportaje/2018-aumento-morosidad-credicia-peru-255634>. [Último acceso: 19 04 2019].
- [4] P. Sánchez Bilbao, «CREDIT SCORING,» España, 2015.
- [5] V. M. Lagua Tubón, «DISEÑO DE UN MODELO DE CREDIT SCORING EN LA GESTIÓN DE RIESGO CREDITICIO EN LA CARTERA DE MICROEMPRESA DE LAS COOPERATIVAS DE AHORRO Y CRÉDITO,» Ecuador, 2015.
- [6] P. M. Pantoja Vilchez, «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,» Lima, 2016.
- [7] J. C. Carpio Ticona, «MODELO DE PREDICCIÓN DE LA MOROSIDAD EN EL OTORGAMIENTO DE CRÉDITO FINANCIERO APLICANDO METODOLOGÍA CRISP-DM,» Universidad Andina "Néstor Cáceres Velásquez", Juliaca, 2016.
- [8] C. F. Hernández Calle y L. F. Quispe Becerril, «SISTEMA DE CREDIT SCORING PARA MINIMIZAR EL RIESGO CREDITICIO EN LA CARTERA PYME DE LA COOPERATIVA DE AHORRO Y CREDITO LEON XIII,» TRUJILLO, 2014.
- [9] S. Rayo Canton, «DISEÑO DE UN SISTEMA DE CREDIT SCORING PARA ENTIDADES DE MICROFINANZAS,» AREQUIPA- PERÚ, 2016.
- [10] M. d. P. Medina Rodríguez y H. G. Ulfe Rentería, «“MODELO DE CREDIT SCORING PARA PREDECIR EL,» Lambayeque - Perú, 2015.
- [11] The CRISP-DM consortium, «CRISP-DM,» Smart Vision Europe, 2015. [En línea]. Available: <http://crisp-dm.eu>. [Último acceso: 31 Mayo 2018].
- [12] N. L. Adriazola Román, «CONSTRUCCIÓN DE UN MODELO RATING DE ADMISIÓN PARA LA CLASIFICACIÓN DE RIESGO CRÉDITO,» Universidad de Chile, Santiago de Chile, 2015.
- [13] J. Salkind, de *Metodos de investigación*, México, Prentice Hall, 1999.
- [14] P. M. P. Vilchez, «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,» Lima, 2016.
- [15] Rational Software Corporation, «Rational Unified Process,» Rational, California, 1998.

IX. ANEXOS

ANEXO N° 01
RECOLECCIÓN DE DATOS

Variables		
Central de riesgo	Limpio Con problemas financieros Nro de puntos positivos Nro de puntos negativos	NORMAL POTENCIAL DEFICIENTE. PUDOS O PERDIDA.
Sexo	Se encuentra en la base de datos	
Edad		
Estado Civil		*En la fecha de la solicitud
Nivel de instrucción	Primario e inferiores Secundario Instituto/Técnico Superior Pregrado Superior Posgrado	
Ocupación	Desempleado Independiente Dependiente Obrero/Agricultor	
Contratado	Si No	
Tiempo de empleo	Mes/año de inicio	*Ninguno si es desempleado
Vivienda	Propia Alquilada	
Tiempo de residencia	Mes/año de residencia	
Ingreso mensual		
Otros ingresos		*Ejm: Ingreso neto del conyuge previamente calculado
Gastos	Impuesto a la renta Agua Luz Otros (Gas, autovaluo, alquiler) Fijo, cable, internet Celular	*Cuando ganan a partir de S/ 2,101 mensual (2019)
Carga familiar	Nro de miembros	
Gastos familiares	Nro de miembros * >328	*Canasta básica familiar S/328. Según INEI 2016

Créditos en otras entidades		*Total de créditos en otras entidades
Cuotas en otras entidades		*Suma de cuotas en otras entidades
Garantía	Valor aprox. de la garantía	
Garante	Codigo del grupo	*Se encuentra en la base de datos
	Nro de miembros	*Se encuentra en la base de datos

COOPERATIVA DE AHORRO Y CRÉDITO
SAN FRANCISCO DE MOCIGUAY

LUIS FERNANDO GIL RÚNIZ
JEFE DE CRÉDITOS

ANEXO N° 03

MANUAL DE USUARIO

1. Generalidades

1.1. Descripción

El sistema desarrollado predice el riesgo crediticio del cliente al solicitar un préstamo, es el principal indicador para la toma de decisión de los analistas de crédito.

1.2. Requisitos

- Se necesita una computadora con sistema operativo Windows Sever 2012 a más, 8GB RAM.
- Se tiene que instalar la librería H2O.ai, gestor de base de datos SQL Server y el Administrador de Internet Information Services (IIS) para que el sistema funcione.

IMPORTANTE: No necesita un inicio de sesión, ya que el sistema está incorporado en el de la Cooperativa.

2. Funcionalidad

Botón que inicializa el proceso de predicción del riesgo en el crédito



3. Inicio

La interfaz se divide en 3 secciones las cuales son: Datos del cliente/crédito, resultado de la predicción y variables especiales.

La interfaz de usuario se divide en tres secciones principales:

- Datos del cliente / crédito:** Incluye campos para Código (7022), Monto (4000), Cuotas (24), Estado Civil (0) Sin unidad familiar, Ingreso Fijo (1100), Carga Familiar (0) e Ingreso Neto (962). Un botón azul "Predecir" está ubicado al lado del campo Código.
- Resultados:** Muestra el resultado de la predicción: "MOROSO". Se muestran también los valores de P0 (0.989887034) y P1 (0.010112966).
- Variables especiales:** Muestra los valores de cinco variables: Aportes S/ (180), Depósitos S/ (0), Fondo Mortuario (NO), Nro. de incidencias (1) e Historial crediticio (1).

© 2019 - COOPAC

A. Sección 1: Datos del cliente / crédito

Los datos del cliente/créditos es la información que se ingresará para la evaluación del crédito. Consta de 7 campos y un botón, definidos de la siguiente manera:

1. Código del socio a evaluar.
2. Monto solicitado en el crédito.
3. Numero de cuotas mensuales en la que pretende devolver.
4. Estado Civil, donde: (0) Sin unidad familiar (Soltero(a)) y (1) Con unidad familiar.
5. Ingreso fijo del cliente.
6. Carga familiar. Nro. de miembros del que dependen del prestatario
7. Ingreso neto del cliente previamente calculado.

Datos del cliente / crédito

Nuevo Crédito

1	Código *	<input type="text" value="7022"/>	Predecir
2	Monto *	<input type="text" value="4000"/>	
3	Cuotas *	<input type="text" value="24"/>	
4	Estado Civil *	<input type="text" value="(0) Sin unidad familiar"/>	
5	Ingreso Fijo *	<input type="text" value="1100"/>	
6	Carga Familiar *	<input type="text" value="0"/>	
7	Ingreso Neto *	<input type="text" value="962"/>	

B. Sección 2: Resultado de la predicción

El resultado de la predicción se representa en 3 campos:

1. P0: es la probabilidad del crédito como potencial moroso.
2. P1: es la probabilidad del crédito sea cumplido.
3. Predicción: es el resultado en palabras que determina el otorgamiento del crédito.

Resultados

1	P0	<input type="text" value="0.989887034"/>	
2	P1	<input type="text" value="0.010112966"/>	
3	Predicción	<input type="text" value="MOROSO"/>	Resultado de la predicción

C. Sección 3: Variables especiales

Las variables especiales son valores considerados en el que analista tenga que tomar una decisión en base a sus aportes o en garantía de sus depósitos. Así también los créditos a partir de la quinta incidencia y cuando supere los 20 créditos en 5 años son considerados automáticamente morosos. Estas variables son 5:

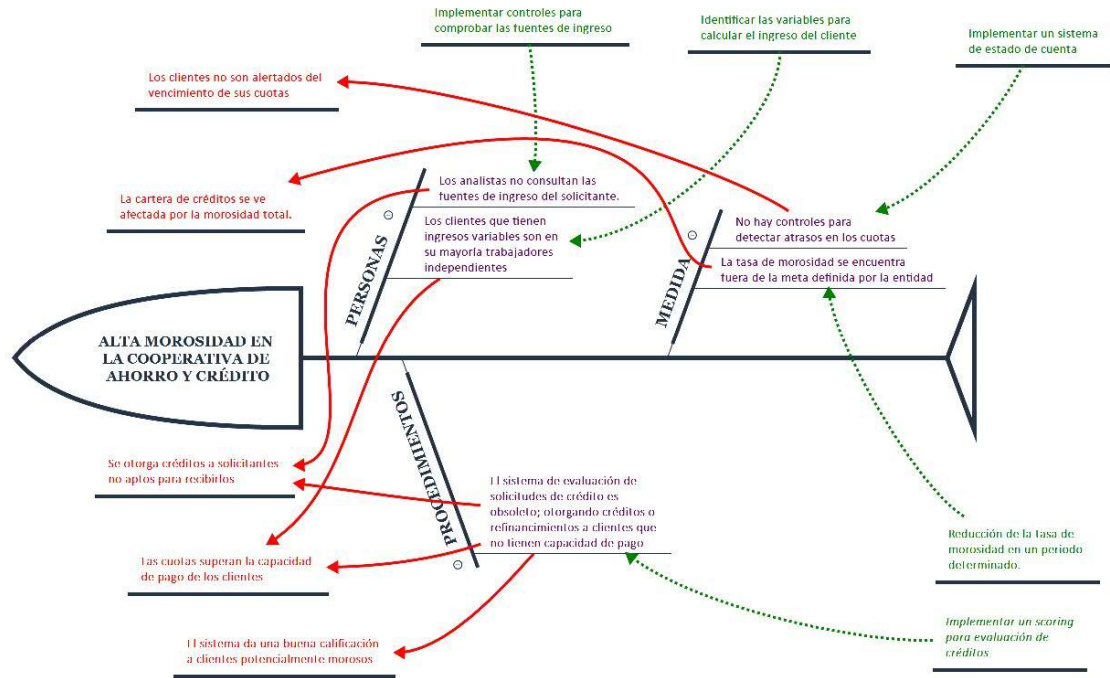
1. Aportes: Importe de aportes del prestatario.
2. Depósitos: Importe de ahorros del prestatario.
3. Fondo Mortuario: Seguro del cliente en caso de fallecimiento.
4. Nro. de incidencias en su historial crediticio
5. Historial crediticio, nro. de créditos otorgados en los últimos 5 años.

Variables	
1 Aportes S/	<input type="text" value="180"/>
2 Depósitos S/	<input type="text" value="0"/>
3 Fondo Mortuario	<input type="text" value="NO"/>
4 Nro. de incidencias	<input type="text" value="1"/>
5 Historial crediticio	<input type="text" value="1"/>

Variables
especiales

ANEXO N° 04

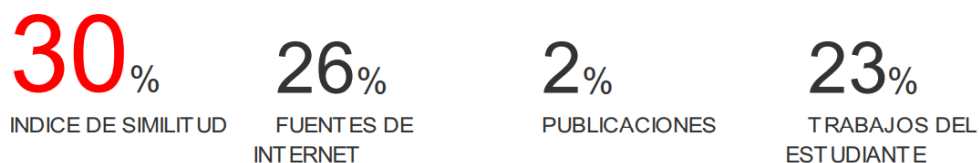
DIAGRAMA ISHIKAWA



ANEXO N° 05
INFORME TURNITIN

Informe de Tesis

INFORME DE ORIGINALIDAD



FUENTES PRIMARIAS

1	repositorio.uancv.edu.pe Fuente de Internet	3%
2	gestion.pe Fuente de Internet	1%
3	www.sbs.gob.pe Fuente de Internet	1%
4	tesis.usat.edu.pe Fuente de Internet	1%
5	repositorio.usil.edu.pe Fuente de Internet	1%
6	repositorio.pucesa.edu.ec Fuente de Internet	1%
7	Submitted to Universidad ESAN -- Escuela de Administración de Negocios para Graduados Trabajo del estudiante	1%
8	Submitted to Universidad Tecnológica del Perú Trabajo del estudiante	1%