UNIVERSIDAD NACIONAL DE CHIMBORAZO

FACULTAD DE INGENIERÍA

CARRERA DE INGENIERÍA EN SISTEMAS Y COMPUTACIÓN

Proyecto de Investigación previo a la obtención del Título de Ingeniero en Sistemas y Computación.

TRABAJO DE TITULACIÓN

PREDICCIÓN DE CLIENTES POTENCIALES UTILIZANDO EL ALGORITMO K-VECINO MÁS CERCANO EN EL ÁREA DE NEGOCIOS DE LA COAC "RIOBAMBA" LTDA.

AUTORES:

Geovanny Augusto Izurieta Guamán
Raquel Johanna Moyano Arias

TUTOR:

PhD. Miryan Estela Narváez Vilema

Riobamba - Ecuador

2019

VEREDICTO DE LA INVESTIGACIÓN

Los miembros del Tribunal de Graduación del proyecto de investigación de título: **"PREDICCIÓN DE CLIENTES POTENCIALES UTILIZANDO EL ALGORITMO K-VECINO MÁS CERCANO EN EL ÁREA DE NEGOCIOS DE LA COAC "RIOBAMBA" LTDA.",** presentado por: Geovanny Augusto Izurieta Guamán y Raquel Johanna Moyano Arias, dirigida por: PhD. Miryan Estela Narváez Vilema.

Una vez escuchada la defensa oral y revisado el informe final del proyecto de investigación con fines de graduación escrito en la cual se ha constatado el cumplimiento de las observaciones realizadas, remite la presente para uso de custodia en la biblioteca de la facultad de Ingeniería de la UNACH.

Para constancia lo expuesto firman:

PhD. Miryan Estela Narváez Vilema
Director de Proyecto

Firma

MsC. Ana Elizabeth Congacha Aushay
Miembro de Tribunal

Firma

MsC. Lady Marieliza Espinoza Tinoco
Miembro de Tribunal

Firma

AUTORÍA DE LA INVESTIGACIÓN

"La responsabilidad del contenido de este Proyecto de Graduación corresponde exclusivamente a: Geovanny Augusto Izurieta Guamán y Raquel Johanna Moyano Arias con la dirección de la PhD. Miryan Estela Narváez Vilema y el patrimonio intelectual de la misma a la Universidad Nacional de Chimborazo"

PhD. Miryan Estela Narváez Vilema
060357677-8
Tutor de Proyecto de Investigación

Geovanny Augusto Izurieta Guamán
060419117-1
Autor del Proyecto de Investigación

Raquel Johanna Moyano Arias
060505492-3
Autor del Proyecto de Investigación

DEDICATORIA

Dedico este trabajo de investigación a Dios por bendecirme, ser mi guía y fortaleza, a mi madre que con mucho cariño, voluntad y paciencia me ha permitido llegar a cumplir un sueño más, enseñándome a creer que no hay obstáculo en la vida que no se pueda superar, a mi familia y amigos que me han apoyado y brindado su ayuda en todo momento.

GEOVANNY AUGUSTO IZURIETA GUAMÁN

A mis padres, por brindarme su apoyo y consejos para hacer de mí una mejor persona.
A mis hermanos y sobrinos, por su amor y compañía en todo momento.
A mi tutora de tesis, por el arduo trabajo de transmitirme sus diversos conocimientos.

RAQUEL JOHANNA MOYANO ARIAS

AGRADECIMIENTO

Agradezco en primero lugar a Dios por estar conmigo en cada paso que doy, cuidándome y dándome fortaleza para continuar, en segundo lugar, a mi madre que fue parte vital para este logro, a mi familia y amigos por haberme brindado su apoyo incondicional y a mi tutora de tesis, quien me guio en todo momento.

GEOVANNY AUGUSTO IZURIETA GUAMÁN

Agradezco a Dios y mi familia por permitirme cumplir con excelencia mis estudios y el desarrollo de mi tesis de grado, gracias por creer en mí y apoyarme en todo momento.

Gracias a la Universidad Nacional de Chimborazo, noble institución de educación superior, la cual me abrió sus puertas para formarme profesionalmente.

Gracias a mis profesores y tutora de tesis, personas de gran sabiduría, quienes me trasmitieron sus conocimientos para ayudarme a llegar al punto en el que hoy me encuentro.

RAQUEL JOHANNA MOYANO ARIAS

ÍNDICE

ÍNDICE DE TABLAS

ÍNDICE DE ILUSTRACIONES

RESUMEN

La Minería de Datos es el proceso de exploración y análisis de grandes cantidades de datos cuyo objetivo es encontrar patrones, tendencias y relaciones significativas (conocimiento), a través de la aplicación de algoritmos y técnicas que ayuden a obtener información relevante. Uno de estos, es el algoritmo de clasificación k vecino más cercano (KNN, K Nearest Neighbor), el cual clasifica nuevas instancias como la clase mayoritaria de entre los k vecinos más cercanos.

Para el desarrollo de la investigación se empleó la metodología del Descubrimiento de Conocimiento en Bases de Datos (KDD, Knowledge Discovery in Databases), por ser un proceso automático en el que se combinan descubrimiento y análisis. La aplicación del algortimo KNN se hizo a través de *Weka*, herramienta de código abierto basada en Java, excelente en tareas de clasificación.

Con la investigación se evidenció que en la base de datos de la COAC "Riobamba" Ltda., se encuentra almacenada información relevante sobre clientes y el manejo de las cuentas de créditos y ahorros, para así clasificar los datos de mayor importancia que aporten a la investigación de forma que no se afecte a los clientes. Como resultado de la investigación se predijo clientes potenciales, clasificándolos de acuerdo con la información demográfica, económica y aspectos internos de la COAC "Riobamba" Ltda., de manera que se apoye a la institución en la toma de decisiones con respecto a futuras ofertas de crédito.

Se concluye la importancia de aprovechar la información que en cada institución se maneja y más si se trata del sector financiero debido a que se benefician las dos partes, los clientes porque tendrán la facilidad de acceder a más opciones de créditos y las instituciones financieras porque aumentarán su cartera de clientes y brindarán un mejor servicio.

Palabras Clave: KDD, Minería de Datos, KNN, clientes potenciales, créditos.

Abstract

Data Mining is the process of exploring and analyzing large amounts of data to find patterns, trends and significant relationships (knowledge), through the application of algorithms and techniques that help to obtain relevant information. One of these is the K Nearest Neighbor (KNN) classification algorithm, which classifies new instances as the majority class among the nearest k neighbors. For the development of the research, the methodology of Knowledge Discovery in Databases (KDD) was used, as it is an automatic process in which discovery and analysis are combined. The application of the KNN algorithm was made through *Weka*, an open-source tool based on Java, excellent in classification tasks. Through the investigation it was evidenced that in the database of the COAC "Riobamba" Ltda., there is stored relevant information about clients and the management of the credit and savings accounts, in order to classify the most important data that contribute to the investigation in a way that does not affect the clients. As a result of the research, potential clients were predicted and classified according to the demographic, economic, and internal aspects of the COAC "Riobamba" Ltda. so as to support the institution in making decisions regarding future credit offers. It concludes that it is important to take advantage of the information handled in each institution and more if it is the financial sector because both parties benefit, the clients because they will have easy access to more credit options and the financial institutions because they will increase their portfolio of clients and provide a better service.

Keywords: KDD, Data Mining, KNN, potential clients, credits.

Translation reviewed by: Trujillo, Myriam
Linguistic Competences Professor

INTRODUCCIÓN

Hoy en día en todas las instituciones sean financieras, comerciales, industriales se manejan grandes volúmenes de información de clientes como de las propias empresas, almacenadas en bases datos, de allí surge la necesidad de encontrar la manera de obtener ventaja a partir de estos datos. Si bien es cierto mediante consultas sobre estos se pueden obtener algunos resultados, a medida que aumentan los registros estos resultados son cada vez más difíciles de interpretar para las empresas. De ahí nace la necesidad de extraer información relevante que está oculta en las bases de datos, más tarde a este proceso se lo catalogó como el descubrimiento de conocimiento en bases de datos (KDD, Knowledge Discovery in Databases).

El proceso KDD tiene varias fases que se realizan de forma secuencial e iterativa, por lo que es posible repetir una y otra vez las fases hasta obtener el conocimiento que se busca. Es importante diferenciar el proceso KDD de la minería de datos. La minería de datos es una de las fases de KDD y se define como la ciencia que estudia patrones en grandes bases de datos, para lo cual emplea distintos algoritmos entre los cuales se encuentra el algortimo de clasificación k vecino más cercano, al cual va orientada la presente investigación.

La idea básica sobre la que se fundamenta este algoritmo es que un nuevo caso se va a clasificar en la clase más frecuente a la que pertenecen sus *k* vecinos más cercanos. El algoritmo ha sido empleado en estudios de carácter financiero, en estos los resultados alcanzados han sido alentadores debido a la buena precisión y eficiencia del algoritmo.

Dentro de esta investigación los resultados de la aplicación del algoritmo servirán de soporte en la toma de decisiones para futuras ofertas de crédito a clientes potenciales en la COAC “Riobamba” Ltda.

El documento está organizado de la siguiente manera:

El capítulo I, detalla el problema, la justificación, el objetivo general y los objetivos específicos de la investigación.

El capítulo II, presenta una descripción general del marco teórico relacionado con la investigación.

El capítulo III, describe la metodología aplicada durante el desarrollo de la investigación.

El capítulo IV, da a conocer los resultados y la discusión de la investigación.

El capítulo V, establece las conclusiones y recomendaciones del estudio.

CAPITULO I

1. PLANTEAMIENTO DE PROBLEMA

1.1. Problema

El principal y más importante activo de una organización financiera es su cartera de clientes, una empresa carece de razón de ser si no es por ellos. Por otro lado, al ser los clientes el eje central de la entidad es importante dirigirse a los mismos con ofertas de créditos que facilite, en un momento dado tener liquidez para poder comprar, hacer pagos, algún tipo de inversión, etc.

El crédito no sólo ayuda en muchos aspectos empresariales, sino que, además, aporta confianza en el sistema financiero de un país, evita que el tejido industrial del mismo se rompa y, sobre todo, invita a la inversión de todo tipo de empresas extranjeras en el país. Actualmente el principal problema que existe en la mayoría de los países, debido a la crisis, es que las entidades financieras no otorgan créditos con facilidad a sus clientes. Por otra parte, la banca tiene herramientas de gestión para obtener información que ayude a la toma de decisiones, pero no utilizan esa información para gestionar créditos.

Parece que esta tendencia está cambiando ya que las propias entidades se dan cuenta de que, si no facilita el crédito, esto al final repercute negativamente en sus propios balances. Es por ello por lo que entidades como la COAC “Riobamba” Ltda., busca llegar a clientes potenciales ofertando créditos que contribuyan al desarrollo de los clientes y de la Cooperativa.

1.2. Justificación

Con la finalidad de predecir clientes potenciales en la COAC “Riobamba” Ltda, existe la necesidad apremiante de identificar a clientes con características comunes y agruparlos. Por otro lado, es importante determinar un modelo que permita clasificar a los clientes en grupos de clientes potenciales y poder realizar predicciones para la otorgación de futuros créditos.

A nivel nacional no existen registros de investigaciones referentes al tema en cuestión, sin embargo, a nivel internacional sí y estos serán utilizados como base bibliográfica para el cumplimiento del objetivo.

En ese contexto, el presente trabajo utiliza el enfoque de aprendizaje perezoso con el algoritmo de k-vecino más cercano (KNN), que se basa en la búsqueda de un conjunto de prototipos de k más cercanos al modelo a clasificar, por ello las predicciones se realizan en base a ejemplos similares al que se tiene que predecir (clientes potenciales). Por tanto, mediante la observación de los vecinos más próximos de un elemento del que se desconoce su pertenencia a alguna de las poblaciones conocidas, es posible deducir la información necesaria para categorizarlo de manera automática.

Debido a la eficacia, este método se ha convertido en una de las técnicas de clasificación más usadas, contando por cientos las variantes y mejoras que la comunidad científica ha ido desarrollando hasta el día de hoy.

1.3. OBJETIVOS

1.3.1. Objetivo General

Predecir clientes potenciales utilizando el algoritmo K-Vecino más cercano en el área de negocios de la COAC “Riobamba” Ltda.

1.3.2. Objetivos Específicos

- Analizar el proceso de gestión de créditos en la COAC “Riobamba” Ltda.

- Estudiar el algoritmo de predicción k-vecino más cercano para aplicarlo en la base de datos del área de negocios de la COAC “Riobamba” Ltda.

- Interpretar los resultados obtenidos que ayudará a la toma de decisiones para próximas ofertas de crédito a la COAC “Riobamba” Ltda.

CAPÍTULO II

2. MARCO TEÓRICO

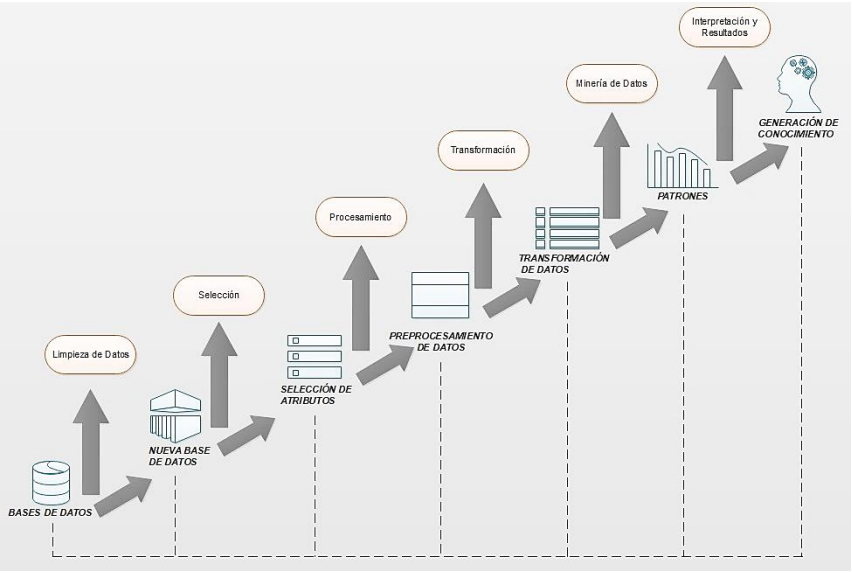
2.1. Proceso de Generación de Conocimiento o KDD

El Descubrimiento de Conocimiento en Bases de Datos (KDD, Knowledge Discovery in Databases) es básicamente un proceso automático en el que se combinan descubrimiento y análisis. El proceso consiste en extraer patrones en forma de reglas o funciones a partir de los datos, para que el usuario los analice. Se puede aplicar en distintos dominios, por ejemplo, para determinar perfiles de clientes fraudulentos, para descubrir relaciones implícitas entre síntomas y enfermedades, para determinar perfiles de estudiantes académicamente exitosos, para determinar patrones de compra de los clientes, etc. (Timarán-Pereira, Hernández-Arteaga, Caicedo-Zambrano, Hidalgo-Troya, & Alvarado-Pérez, 2016)

2.2. Proceso de KDD

Las fases del proceso de Generación de Conocimiento pueden ser recursivas, es decir, se retorna a ellas una y otra vez a medida que se obtienen resultados que requieren replantear las variables iniciales. A continuación, se observan y detallan cada una de las fases.

Ilustración 1: Fases del Proceso KDD



Fuente: (UIAF), Unidad de Información y Análisis Financiero. (2014). Recuperado de https://www.urosario.edu.co/observatorio-de-lavado-de-activos/Archivos_Lavados/Tecnicas-de-mineria-de-datos-para-la-prevencion-de.pdf

Fase Selección de los datos

Esta etapa consiste en la recolección y preparación de los datos, requiere cerca del 90% del tiempo. En la selección de los datos se comprende la problemática asociada a la base de datos y se establecen objetivos, así como también se identifican las variables que serán consideradas para la construcción del modelo de minería de datos.

Fase Pre procesamiento de Datos

La etapa de pre procesamiento de datos consiste en analizar si la base de datos requiere incluir o integrar información o variables que reposan en otras bases de datos, y que será relevante para el modelo de minería de datos. Si es necesario, se realiza un modelo de entidad-relación entre tablas, el cual permite representar las entidades relevantes (representaciones gráficas y lingüísticas) de un sistema, así como sus propiedades e interrelaciones. (Ávila, 2005)

Además, se realiza el reconocimiento y la limpieza de los datos, eliminando inconsistencias y datos erróneos.

Fase Selección de Características

Consiste en encontrar las características más significativas para representar los datos, dependiendo del objetivo del estudio. En este paso se pueden utilizar métodos de transformación para reducir el número efectivo de variables a ser consideradas o para encontrar otras representaciones de los datos.

En esta fase se estandariza o normaliza la base de datos y se disminuye el tamaño de los datos mediante la eliminación de características redundantes.

Fase Minería de Datos

La minería de datos se puede definir como un proceso no trivial de identificación válida, novedosa, potencialmente útil y entendible de patrones comprensibles que se encuentran ocultos en los datos, que, a su vez, facilita la toma de decisiones y emplea técnicas de aprendizaje supervisado y no-supervisado. (Esteban, 2008)

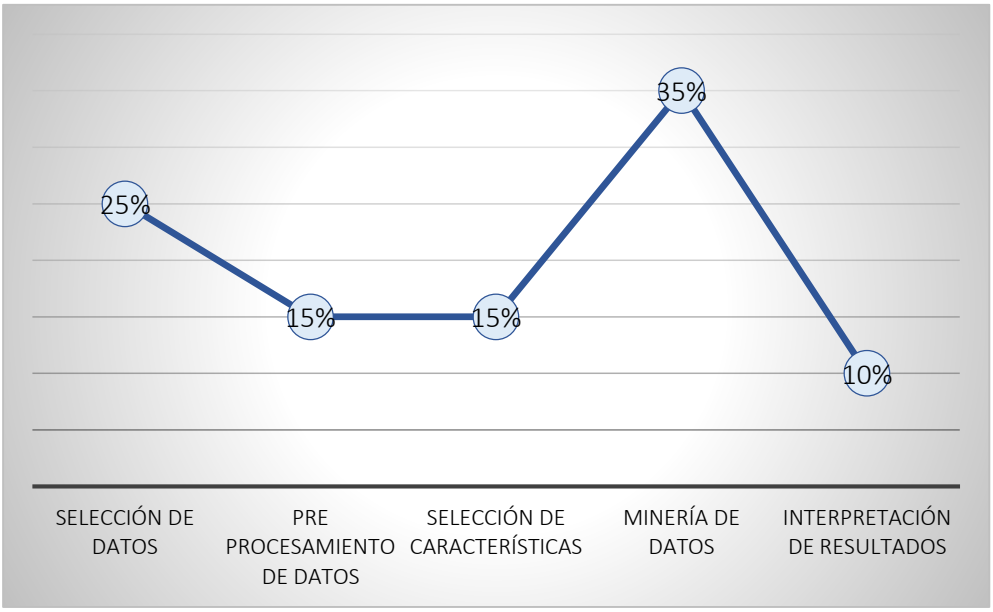
En esta fase se decide si el objetivo del proceso de KDD es: Regresión, Clasificación, Agrupamiento, etc. Además, se elige el algoritmo de Minería de Datos a utilizarse.

Fase Interpretación y Resultados

Se analizan los resultados de los patrones obtenidos en la fase de Minería de Datos, mediante técnicas de visualización y de representación, con el fin de generar conocimiento que aporte mayor valor a los datos. En esta fase se evalúan los resultados con los expertos y, si es necesario, se retorna a las fases anteriores para una nueva iteración. ((UIAF), 2014)

A continuación, se muestra el esfuerzo requerido por cada una de las fases del Proceso de KDD.

Ilustración 2: Esfuerzo requerido por cada fase KDD



Realizado por: Los autores

2.3. Minería de Datos

La minería de datos puede definirse como un proceso de descubrimiento de nuevas y significativas relaciones, patrones y tendencias al examinar grandes cantidades de datos. La disponibilidad de grandes volúmenes de información y el uso de herramientas informáticas ha transformado el análisis de datos hacia técnicas englobadas bajo el nombre de minería de datos o Data Mining. (López, 2007)

De esta manera la minería de datos hace hincapié en:

- La escalabilidad del número de atributos y de instancias.
- Algoritmos y arquitecturas.
- La automatización para manejar grandes volúmenes de datos mezclados.

2.3.1. Técnicas de Minería de Datos

Clasificación

La clasificación es la técnica de minería de datos más comúnmente aplicada, que emplea un conjunto de ejemplos preclasificados para desarrollar un modelo que puede clasificar la población de registros en general. Este enfoque emplea con frecuencia el árbol de decisión o los algoritmos de clasificación basados en redes neuronales. (Ramageri, 2010)
La técnica de clasificación se usa en la segmentación de clientes, modelado de negocios, análisis de crédito, entre otras.

Regresión

La regresión es un modelo predictivo de minería de datos también conocido como técnica de aprendizaje supervisado. Esta técnica analiza la dependencia de algunos valores de atributo, que dependen principalmente de valores de otros atributos, presentes en el mismo elemento. En las técnicas de regresión se conocen valores objetivo, por ejemplo, puede predecir el comportamiento del niño basándose en antecedentes familiares. (Deepashri & Ashwini, 2017)

Análisis de datos de series de tiempo

La base de datos de series de tiempo usa secuencias de valores o eventos obtenidos en mediciones repetidas del tiempo. Los valores se miden normalmente en intervalos de tiempo iguales, como por hora, diariamente, semanalmente. Una base de datos de secuencias es cualquier base de datos que consiste en una secuencia de eventos ordenados, muchas veces con nociones concretas de tiempo. (Wolfgang, 2015)

Predicción

Esta técnica descubre la relación entre variables independientes y la relación entre variables dependientes e independientes. La predicción es predecir un estado futuro, en lugar de uno actual. (Wolfgang, 2015)
Sus aplicaciones incluyen la prevención de desastres naturales, epidemias, entre otros. Un ejemplo de aplicación, el volumen de ventas de accesorios para computadoras se puede predecir en función del número de computadoras vendidas en los últimos meses. (Deepashri & Ashwini, 2017)

Agrupamiento o Clustering

El clustering es un modelo no supervisado que se utiliza para identificar grupos (clúster) o patrones de observaciones similares en un conjunto de datos. Una de las técnicas más utilizadas es el método de k-means, que consiste en definir un punto central de cada clúster (centroide) y asignar a cada individuo al clúster del centroide más próximo en función de las distancias existentes entre los atributos de entrada. El algoritmo parte de la fijación de k centroides aleatoriamente y, mediante un proceso iterativo, se asigna cada punto al clúster con el centroide más próximo, procediendo a actualizar el valor de los centroides. Este proceso termina cuando se alcanza un determinado criterio de convergencia. (Management Solutions, 2018)

2.3.2. Aplicaciones de la minería de datos

Existen numerosas áreas donde la minería de datos se aplica, prácticamente en todas las actividades humanas que generen datos:

- Comercio y banca: segmentación de clientes, previsión de ventas, análisis de riegos.
- Medicina y farmacia: diagnóstico de enfermedades y la efectividad de los tratamientos.
- Seguridad y detección de fraude: reconocimiento facial, identificaciones biométricas, accesos a redes no permitidas.

- Recuperación de información no numérica: minería de texto, minería web, búsqueda e identificación de imagen, video, voz y texto de bases de datos multimedia.
- Geología, minería, agricultura y pesca: identificación de áreas de uso para distintos cultivos o de pesca o de explotación minera, en bases de datos de imágenes de satélites.
- Ciencias ambientales: identificación de modelos de funcionamiento de exosistemas naturales y/o artificiales para mejorar su observación, gestión y control.
- Ciencias sociales: estudio de los flujos de la opinión pública, planificación de ciudades (identificación de barrios con conflicto en función de valores sociodemográficos) (Riquelme , Ruiz, & Gilbert, 2006)

2.4. Algoritmo k vecino más cercano (KNN)

Es un paradigma clasificatorio conocido como K-NN (K-Nearest Neighbor). La idea básica sobre la que se fundamenta este algoritmo es que un nuevo caso se va a clasificar en la clase más frecuente a la que pertenecen sus *K* vecinos más cercanos. (Moujahid, Inza , & Larrañaga)

El método del vecino más cercano se puede extender utilizando no uno, sino un conjunto de datos más cercanos para predecir el valor de los nuevos datos, en lo que se conoce como los k-vecinos más cercanos (k-NN o k-Nearest Neighbor). Al considerar más de un vecino, se brinda inmunidad ante ruido y se suaviza la curva de estimación. (Moreno , 2004)

El proceso seguido por el algoritmo KNN es el siguiente:

- Se almacena en una tabla los prototipos clasificados manualmente. A esto lo llamaremos conjunto de referencia (de tamaño N).
- Ante un nuevo patrón a clasificar se calcula su distancia euclídea a los N prototipos del conjunto de referencia y se consideran los K más cercanos.
- Se contabilizan las clases a las que pertenecen esos K prototipos y el nuevo patrón se clasifica con la clase mayoritaria. (Cazorla, Olmo, & Alados-Arboledas, 2005)

2.4.1. Características Generales

- Las reglas de clasificación por vecindad están basadas en la búsqueda en un conjunto de prototipos de los *k* prototipos más cercanos al patrón a clasificar.
- No existe un modelo global asociado a los conceptos a aprender.
- Las predicciones se realizan basándose en los ejemplos más parecidos al que hay que predecir.
- Se conoce como mecanismo de aprendizaje perezoso (*lazy learning*).
- En KNN se debe especificar una métrica para poder medir la proximidad, suele utilizarse por razones computacionales la distancia Euclídea (δ), para este fin. (García Cambronero & Gómez Moreno)

2.4.2. Medidas de distancia de proximidad

Las medidas de distancia de proximidad en KNN son fundamentales para descubrir similitudes entre medidas de dos objetos similares y diferencias entre puntos de datos. Además, tiene como objetivo principal encontrar la distancia apropiada o similar.

Distancia Euclidiana

K-Nearest Neighbor (KNN) se calcula a través de la distancia euclidiana brindando eficiencia y productividad. Es una distancia entre dos puntos en el espacio euclidiano en el que se calcula la raíz de las diferencias cuadradas entre las coordenadas de un par de puntos de datos. La fórmula es:

$$Dist_{xy} = \sqrt{\Sigma_{k=1}^{m}(x_{ik} - x_{jk})^2}$$

Distancia de Manhattan

$$Dist_{xy} = |X_{ik} - X_{jk}|$$

Calcula la diferencia entre las coordenadas de un par de puntos de datos. (Mulak & Talhar, 2013)

2.4.3. Aplicaciones

KNN como técnica de minería de datos tiene una amplia variedad de aplicaciones. Algunas de ellas son:

Minería de Texto

El algoritmo K-Nearest Neighbor (KNN) es uno de los más populares para la categorización de texto o la minería de texto.

Agricultura

En general KNN, se aplica menos que otras técnicas de minería de datos en campos relacionados con la agricultura. Se ha aplicado, por ejemplo, en simulaciones de precipitaciones diarias y otras variaciones climáticas, así como la evaluación de invernaderos forestales y su estimulación de variables. En estas aplicaciones, se utilizan imágenes satelitales, con el objetivo de mapear la cobertura del suelo y el uso de este. Otras incluyen estimación de medidas de agua en el suelo y el pronóstico del clima.

Finanzas

La minería de datos como un proceso común para descubrir patrones y correlaciones útiles tiene su propio nicho en el modelado financiero. Similares a otros métodos computacionales, casi todos los métodos y técnicas de minería de datos se han utilizado en el modelado financiero. Una lista completa incluye una variedad de modelos lineales y no lineales de redes neuronales multicapa, k-medias y agrupamiento jerárquico, k-vecinos más cercanos, análisis de árboles de decisión, regresión (logística y múltiple general), análisis de componentes y aprendizaje bayesiano. La previsión del mercado de valores es una de las tareas financieras más importantes de KNN. Algunas otras aplicaciones del algoritmo en finanzas se mencionan a continuación.

- Previsión del mercado de valores: predecir el precio de un stock, en función del rendimiento de la empresa, medidas y daos económicos.
- Tipo de cambio de moneda.
- Bancarrotas bancarias.
- Comprender y gestionar el riesgo financiero.
- Negociación futura.
- Calificación crediticia.
- Gestión de préstamos.
- Perfil del cliente del banco.
- Análisis de lavado de dinero.

Medicina

- Predecir un segundo ataque si un paciente hospitalizado ha sufrido ya un primer ataque cardiaco. La predicción se apoyará en las mediciones clínicas, demográficas y dietéticas.
- Estimar en una persona diabética la cantidad de glucosa en la sangre, a partir del espectro de absorción infrarroja de la sangra de la misma.
- Identificar los factores de riesgo para el cáncer con base en variables clínicas y demográficas.
- Otras aplicaciones incluyen la detección de intrusiones en sistemas informáticos y el manejo de bases de datos de objetos en movimiento, como computadoras con conexiones inalámbricas.

2.4.4. Ventajas y Desventajas de KNN

Ventajas

KNN tiene varias ventajas principales: efectividad, intuición y rendimiento de clasificación competitiva en muchos dominios. Es robusto para los datos de entrenamiento ruidosos y es efectivo si los datos de entrenamiento son grandes.

Desventajas

A pesar de las ventajas, KNN es muy sensible a las características irrelevantes o redundantes debido a que todas las características contribuyen a la similitud y, por lo tanto, a la clasificación, por lo que, el costo de cálculo es bastante alto debido a que se necesita calcular la distancia de cada instancia de consulta a todas las muestras de entrenamiento. (Department of Economics & Payame Noor University, 2013)

2.4.5. Cross-Validation (Validación Cruzada)

La validación cruzada es un método estadístico para evaluar y comparar algoritmos de aprendizaje dividiendo en dos segmentos: uno usado para aprender o entrenar el modelo y el otro utilizado para validar el modelo. En la validación cruzada típica, los conjuntos de capacitación y validación deben cruzarse en rondas sucesivas de modo que cada punto de datos tenga la posibilidad de ser validado. La forma básica de validación cruzada es la validación k-fold.

En la validación cruzada k-fold los datos se dividen en k segmentos o pliegues de igual tamaño (o casi igual). Posteriormente, se realizan las k iteraciones de entrenamiento y validación de tal manera que dentro de cada iteración se retiene un pliegue diferente de datos para la validación, mientras que los k pliegues restantes se usan para el aprendizaje. En minería de datos y aprendizaje automático, la validación cruzada 10 veces (k=10) es la más común.

La validación cruzada se usa para evaluar o comparar algoritmos de aprendizaje de la siguiente manera: en cada iteración, uno o más algoritmos de aprendizaje usan k1 pliegues de datos para aprender uno o más modelos, y posteriormente se les pide a los modelos aprendidos que hagan predicciones sobre los datos en el pliegue de validación. El rendimiento de cada algoritmo de aprendizaje en cada pliegue se puede rastrear utilizando una métrica de rendimiento predeterminada como es la precisión. (Payan, Lei, & Huan, 2008)

2.5. Software

2.5.1. Microsoft SQL Server

Derivado de Microsoft, Microsoft SQL Server es un sistema gestor de bases de datos relacional (RDBMS), tiene como principal leguaje de consulta a Transact-SQL.

Características

- Soporte para procedimientos almacenados.
- Permite gestionar información de servidores de bases de datos de terceros.
- Incluye potente entorno grafico para administración, permitiendo el uso de comandos MML y DDL de manera gráfica.
- Escalabilidad, seguridad y estabilidad.
- Soporte para transacciones.
- Cliente-servidor, donde los datos e información se albergan en el servidor y los clientes o terminales de la red acceden solo a la información.

Incluye una versión reducida denominada MSDE orientado a proyectos más pequeños como el mismo gestor de base de datos, además Microsoft SQL Server se establece como una alternativa de Microsoft a otros poderosos sistemas gestores de bases de datos como Sybase ASE, Oracle, PostgreSQL o MySQL. (Santamaría & Hernández)

Brinda a las organizaciones herramientas efectivas para proteger, desbloquear y escalar el poder de sus datos, desde computadoras de escritorio, teléfonos y tabletas, hasta centros de datos y nubes privadas y públicas. (Mistry & Misner, 2012)

2.5.2. Software para la Calidad de Datos

2.5.2.1. Definición de Calidad de Datos

En la actualidad la existencia de grandes volúmenes de datos en abrumadoras bases de datos excede la capacidad de poder analizarlos para generar información útil, situación que viven día numerosas organizaciones en el mundo. En este contexto diversas decisiones son tomadas muchas veces en base a la intuición y experiencias

vividas en ambientes reales, por consiguiente, los datos no son limpios, contienen inconsistencias, errores y la falta de valores clave.

La Calidad de Datos conocida también como Data Quality o Customer Data Integration, Depuración y Limpieza de Datos consiste en validar todos los datos que se posean para proceder a corregirlos, normalizarlos y de duplicarlos (eliminar duplicados) a través de las etapas que posee el KDD: pre procesamiento, minería de datos y post procesamiento.

Etapa del Pre procesamiento: prepara los datos para poder ser utilizados en la siguiente etapa, aquí se mencionan los siguientes procesos: selección de variables relevantes para el estudio, limpieza de datos con la finalidad de eliminar datos erróneos o inconsistentes y transformación de datos cuyo fin es reducir el almacén de datos.

Etapa de Minería de Datos: se decide la tarea a realizar teniendo como ejemplo la agrupación o clasificación.

Etapa de Post procesamiento: consiste en la interpretación y evaluación de resultados que son analizados por los expertos. De ser necesario se vuelve a fases preliminares para una nueva iteración. (Camana & Torres, 2017)

2.5.2.2. RapidMiner

El software proporciona procedimientos de minería de datos y aprendizaje automático que incluye la extracción, transformación y la carga de datos, conocido también como proceso ETL, pre procesamiento y visualización de datos, modelado, evaluación e implementación. RapidMiner está escrito en el lenguaje de programación Java. Utiliza esquemas y evaluadores de atributos del entorno de aprendizaje automático de Weka y esquemas de modelado estadístico de R-Project. Se lo utiliza para investigación, educación y capacitación. (Hirudkar & Mrs. Sherekar, 2013)

Especificación Técnica

- Lanzado en 2016
- Licenciado por AGPL Propietario.
- Multiplataforma, es decir se lo puede instalar en cualquier sistema operativo.

Características generales

- RapidMiner tiene un entorno para el aprendizaje automático y los procesos de minería de datos
- Cuenta con operadores flexibles para los formatos de archivos de entrada y salida.
- RapidMiner admite alrededor de veinte y dos formatos de archivos
- RapidMiner incluye muchos algoritmos de aprendizaje de Weka.

Especialización

- RapidMiner brinda soporte a la mayoría de las bases de datos, lo que significa que los usuarios pueden importar información desde una diversidad de bases de datos para ser examinadas y analizadas dentro de la aplicación.
- Provee soluciones especializadas para empresas que incluyen análisis predictivo y computación estadística.

Ventajas

- Tiene facilidad completa para la evaluación del modelo usado validación cruzada y conjuntos de evaluación independientes.
- Incluye más de mil quinientos métodos para la integración de datos, transformación de datos, análisis y modelado, así como la visualización.
- Ofrece numerosos procedimientos, especialmente en el área de selección de atributos y para la detección de valores atípicos que ninguna otra solución ofrece.

Limitaciones

RapidMiner es un paquete software de minería de datos más adecuado para las personas que están acostumbradas a trabajar con archivos de bases de datos, como en entornos académicos o empresariales. La razón de esto es porque el software requiere la capacidad de manipular sentencias y archivos SQL. (Rangra & Dr. Bansal, 2014)

2.5.2.3. EmEditor

Es un editor de texto rápido, liviano pero extensible y de fácil uso para Windows con versiones nativas disponibles de 64 y 32 bits. Poderosa herramienta para edición de texto que ha ganado innumerables premios en todo el mundo. (Emurasoft, 2019)
La mayoría de las operaciones típicas como abrir y guardar archivos, así como buscar y reemplazar cadenas se realizan en una fracción de segundo, y no tratándose solo de archivos normales que pesan kilobytes (kb). EmEditor funciona perfectamente con archivos grandes y extra grandes de hasta 248 gigabytes (Gb) que contienen millones de líneas convirtiéndose en una herramienta para todos sin importar el idioma con el que se trabaje. (Emura, 2011)

2.5.3. Weka

Es una colección de algoritmos de aprendizaje para tareas de minería de datos que se pueden aplicar directamente a un conjunto de datos. Weka contiene un conjunto de interfaces graficas con una colección de algoritmos para análisis de datos y modelado predictivo además de herramientas para la visualización de los datos.

Especificación técnica

- Lanzado por primera vez en 1997.
- Tiene licencia pública general de GNU.
- Multiplataforma

Características Generales

Es una herramienta de código abierto basada en Java que es una colección de muchos algoritmos de minería de datos y aprendizaje automático, incluido el procesamiento previo de datos, clasificación, agrupación y extracción.

Especialización

- Weka es el más adecuado para las reglas de asociación en minería de datos.
- Adecuado con fuertes técnicas para el aprendizaje automático.

Ventajas

- Adecuado para desarrollar nuevos esquemas de aprendizaje automático.
- Weka carga el archivo de datos en formatos ARFF, CSV, C4.5 y binario.
- Aunque es de código abierto, gratuito extensible, puede ser integrado en otros paquetes Java.

Limitaciones

- El lector CSV no es tan robusto como en RapidMiner.
- Conectividad ineficiente con hojas de cálculo Excel y bases de datos no basadas en Java.
- No tiene facilidad de guardar parámetros para que la escala se aplique en futuros conjuntos de datos. (Rangra & Dr. Bansal, 2014)

2.5.4. IBM SPSS Statistics 25

El programa estadístico SPSS (Statical Package for the Social Sciences), es uno de los programas mayormente usados a nivel mundial, los procesos estadísticos que incluye ayudan a entidades que necesitan analizar datos para aplicaciones prácticas o diversas necesidades de investigación. Además, ofrece crear vínculos con herramientas office para finalmente permitir efectuar análisis estadísticos muy complejos.

Usos Potenciales

Permite crear un archivo en forma estructurada, además de organizar una base de datos que puede ser examinada con diferentes métodos estadísticos, permite realizar y capturar análisis de datos sin depender de otros programas y por último hace posible la transformación de un banco de datos creado en Excel a una de SPSS.

Ventajas y Desventajas

El software hace uso de distintos cuadros de diálogo que permiten determinar las acciones a tomar seleccionando análisis útiles, sin embargo, el usuario al no contar con una experiencia previa haciendo uso de SPSS hace complicado escoger las opciones de análisis que este provee. (Castañeda, Cabrera, Navarro, & Wietse de Vries, 2010)

CAPÍTULO III

3. METODOLOGÍA

3.1. Tipo y diseño de investigación

3.1.1. Investigación cuantitativa

El tipo de investigación, según el nivel de medición y análisis de la información es cuantitativa, debido a que se llevó a cabo el estudio de la base de datos del área de negocios de la COAC "Riobamba" Ltda., es decir, se trató de una investigación objetiva en la que se establecieron mediciones reales con lo que se obtuvo una mayor cantidad de datos fiables de manera que se lograron explicaciones contrastadas, estadísticas y clasificables. El resultado de este tipo de investigación por lo general se basa en la estadística.

3.2. Unidad de análisis

La unidad de análisis está representada por la base de datos del área de negocios de la COAC "Riobamba" Ltda., entendiéndose que dentro de ella se maneja información de ahorros y créditos de los clientes.

3.3. Población de estudio y tamaño de la muestra

En la investigación se establece como población de estudio a la base de datos correspondiente al área de negocios de la COAC "Riobamba" Ltda., que abarca información de ahorros y créditos de los clientes.
La muestra corresponde a los clientes de la matriz de la COAC "Riobamba" Ltda., equivalente a 73 529 clientes.

3.4. Técnicas de recolección de datos

3.4.1. Entrevista

Se utilizó la técnica de la entrevista para establecer una relación directa con la COAC "Riobamba" Ltda., con el objetivo de conocer el proceso de gestión de créditos, de

manera que se emplearon preguntas abiertas como instrumento para definir un modelo con diversos parámetros o características que marcaron el inicio del análisis de la investigación.

3.4.2. Estudio de la Base de Datos

Se realizó el estudio de la base de datos de la COAC "Riobamba" Ltda., obteniéndose distintos criterios en base a un exhaustivo análisis de esta, delimitando así las tablas a utilizarse de acuerdo con la actividad económica, calificación, los tipos de créditos a los que acceden, el historial y pago de créditos.

3.5. Técnicas de análisis e interpretación de la información

3.5.1. Herramientas utilizadas

- Microsoft SQL Server
- Software RapidMiner
- Software EmEditor
- Herramienta de Minera de Datos Weka
- IBM SPSS Statistics 25

3.5.2. Metodología aplicada

Para llevar a cabo la investigación y cumplir con los objetivos, se aplicó la metodología de Descubrimiento de Conocimiento en Bases de Datos, KDD (del inglés Knowledge Discovery in Databases) dentro de la cual se cumplieron cinco fases que se indican:

Fase Selección de los datos

En esta fase se comprendió la problemática asociada a la base de datos, recolectando y preparando la información requerida y con la ayuda de la herramienta de base de datos Microsoft SQL Server se establecieron las siguientes tablas a utilizarse.

- Persona
- Clientes

- Saldos
- Historial 2017/2018/2019
- Solicitud Crédito
- Historia Plazo

Las tablas Personas y Clientes se tomaron en cuenta por contener información demográfica de los clientes de la Cooperativa.
Las tablas Saldos e Historial reflejan los tipos de cuentas de ahorro que se manejan y el comportamiento de los ahorros de los clientes.
Por último, las tablas Solicitud Crédito e Historia Plazo reflejan en sus registros información crediticia y comportamiento de pago de los clientes por lo que se consideraron vitales dentro de la investigación.

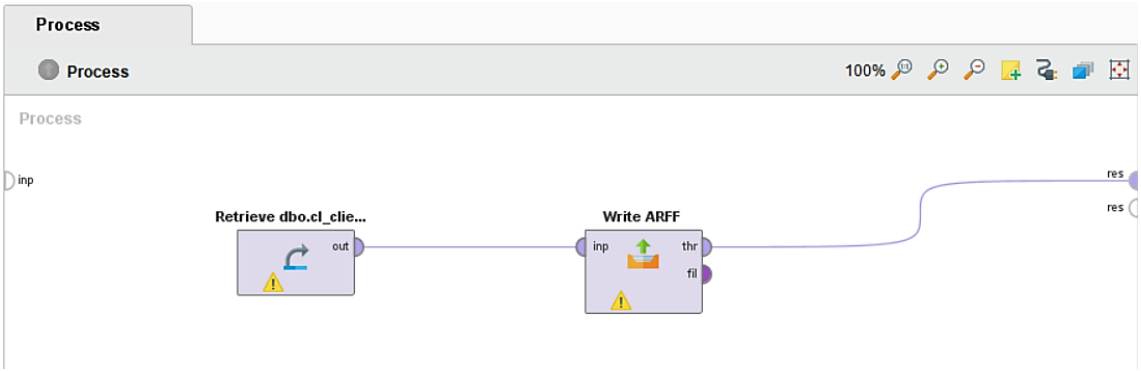
Fase Pre procesamiento de Datos

La fase de pre procesamiento de datos permitió hacer un reconocimiento y limpieza de los atributos y registros contenidos en las tablas seleccionadas, eliminando ruidos e inconsistencias con la ayuda del software de limpieza EmEditor.
Con el software se pudo filtrar la información de las tablas de manera que los registros reflejados sean únicamente de la matriz de la Cooperativa, además, con la ayuda de EmEditor se dividió a las tablas en otras con menos registros para su posterior análisis debido a la cantidad de información.

Fase Selección de Características

Con la ayuda del software RapidMiner se realizó la normalización de los atributos de cada una de las tablas seleccionadas dentro de la investigación (Anexo II).
Con RapidMiner se pudo transformar el tipo de archivo original de las tablas (.csv) al formato .arff, este tipo de formato fue considerado debido a que es uno de los que mejor acepta la herramienta *Weka.*
A continuación, se muestra un ejemplo de lo que se realiza en RapidMiner para la transformación de formatos de archivos.

Ilustración 3: Diagrama de transformación de formato de archivos en RapidMiner

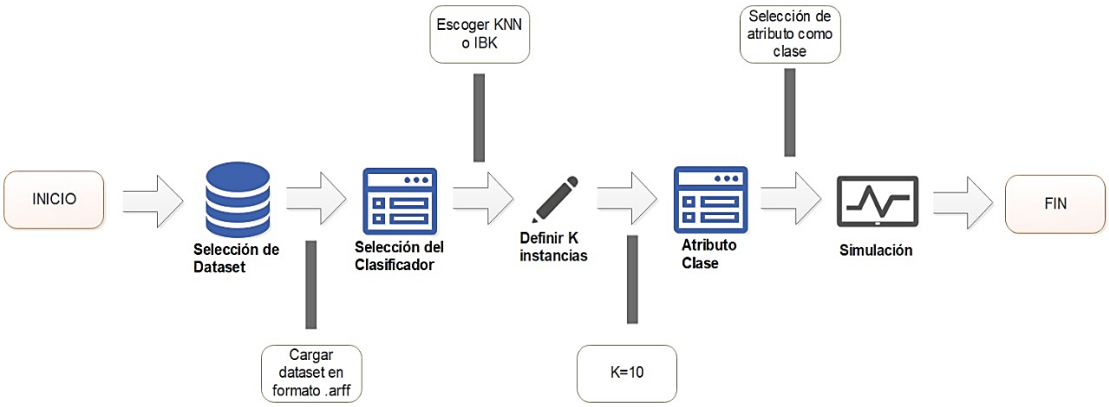


Realizado por: Los autores

Fase Minería de Datos

En esta fase se llevó a cabo la ejecución de las simulaciones de las tablas consideradas en la investigación, para ello se utilizó el software de Minería de Datos Weka, en el cual se siguieron los pasos que se indican en el diagrama:

Ilustración 4: Diagrama de Minería de Datos en Weka



Realizado por: Los autores

Cabe recalcar que se utilizó k=10 para todas las tablas simuladas debido a la cantidad de registros de cada una.

Finalizadas las simulaciones, *Weka* arroja los resultados para cada tabla que se pueden evidenciar más adelante en el apartado de resultados (Capítulo IV).

Fase Interpretación y Resultados

Con el software IBM SPSS Statistics 25, especialista en manejar estadísticas se graficó los resultados para una mejor apreciación de estos.

CAPÍTULO IV

4. RESULTADOS Y DISCUSIÓN

De acuerdo con la metodología KDD utilizada en el presente estudio, se presentan los resultados obtenidos en cada una de las fases.

4.1. Selección de los Datos

La tabla 1, muestra las tablas utilizadas en la investigación con su respectivo número de registros y descripción de cada una de ellas, las tablas Personas y Clientes describen datos demográficos e información de las cuentas de los clientes, las tablas Saldos e Historial hacen referencia al comportamiento de las cuentas de ahorros y las tablas Solicitud Crédito e Historia Plazo indican información crediticia de los clientes.

Tabla 1: Base de Datos COAC "Riobamba" Ltda.

Tabla	N.º Registros		Sucursal	Descripción
Personas	225 672		Todas las sucursales del país	Datos demográficos de las personas que forman parte de la Cooperativa.
Clientes	73 529		Matriz	Información de la cuenta que maneja cada uno de los clientes.
Saldos	779 884		Matriz	Información del tipo de cuenta de ahorros que manejan los clientes.
Historial 2017/2018/2019	2017	5 678 281	Matriz	Comportamiento de las cuentas de ahorros de los años 2017, 2018 y de enero hasta abril 2019.
	2018	5 396 158		
	2019	3 916 502		
Solicitud Crédito	62 317		Matriz	Información crediticia de los clientes.
Historia Plazo	3 175 308		Matriz	Comportamiento de pago de crédito de los clientes.

Realizado por: Los autores

4.2. Pre procesamiento de Datos

Las seis tablas depuradas y limpias pertenecientes a la Base de Datos de la COAC "Riobamba" Ltda., se pueden ver en el Anexo II.

4.3. Selección de Características

Los atributos normalizados de las tablas que se utilizaron para la ejecución de las simulaciones se pueden observar en el Anexo III.

4.4. Minería de Datos

4.4.1. Condición de los equipos con los cuales se realizaron las simulaciones

Para obtener los resultados de las simulaciones realizadas en el software *WEKA* de cada una de las tablas se utilizaron dos equipos portátiles con las siguientes características:

Tabla 2: Características Equipo 1

Característica	Descripción
SO	Windows 10 Home de 64 Bits
CPU	Intel® Core™ i7 Séptima Generación
RAM	16 Gb

Realizado por: Los autores

Tabla 3: Características Equipo 2

Característica	Descripción
SO	Windows 10 Pro de 64 Bits
CPU	Intel® Core™ i5 Quinta Generación
RAM	8 Gb

Realizado por: Los autores

4.4.2. Resultados arrojados por las simulaciones

WEKA arrojó los resultados para cada una de las tablas, como se indica:

Tabla 4: Resultados Tabla Personas

Instancias Clasificadas	225 670	
Instancias Clasificadas Correctas	204 962	90.82%
Instancias Clasificadas Incorrectas	20 708	9.17%

Realizado por: Los autores

La tabla 4, muestra de un total de 225 670 instancias clasificadas, el 90.82% se clasifican como correctas.

Tabla 5: Resultados Tabla Clientes

Instancias Clasificadas	73 527	
Instancias Clasificadas Correctas	44 525	60.55%
Instancias Clasificadas Incorrectas	29 002	39.44%

Realizado por: Los autores

La tabla 5, indica de un total de 73 537 instancias clasificadas, el 60.55% se clasifican como correctas.

Tabla 6: Resultados Tabla Saldos

Instancias Clasificadas	779 882	
Instancias Clasificadas Correctas	771 166	98.88%
Instancias Clasificadas Incorrectas	8 716	1.11%

Realizado por: Los autores

La tabla 6, muestra de un total de 779 882 instancias clasificadas, el 98.88% se clasifican como correctas.

Tabla 7: Resultados Tablas Historial

	Historial 2017		**Historial 2018**		**Historial 2019**	
Instancias Clasificadas	5 678 279		5 396 156		3 916 500	
Instancias Clasificadas Correctas	4 326 787	68.76%	4 162 407	70.36%	3 037 430	71.06%
Instancias Clasificadas Incorrectas	1 351 492	31.24%	1 233 749	29.64%	879 070	28.94%

Realizado por: Los autores

La tabla 7, indica el 68.76% de instancias clasificadas como correctas para el año 2017, el 70.36% para el año 2018 y el 71.06% para los meses de enero a abril de 2019.

Tabla 8: Resultados Tabla Solicitud Crédito

Instancias Clasificadas	62 315	
Instancias Clasificadas Correctas	45 566	73.12%
Instancias Clasificadas Incorrectas	16 749	26.87%

Realizado por: Los autores

La tabla 8, muestra de un total de 62 315 instancias clasificadas, el 73. 12% se clasifican como correctas.

Tabla 9: Resultados Tabla Historia Plazo

Instancias Clasificadas	3 175 306	
Instancias Clasificadas Correctas	3 152 660	99.28%
Instancias Clasificadas Incorrectas	22 646	0.72%

Realizado por: Los autores

La tabla 9, indica de un total de 3 175 306 instancias clasificadas, el 99.28% se clasifican como correctas.

4.5. Tiempo de respuesta de las simulaciones

Tabla 10: Tiempo de respuesta de cada equipo

Tabla	N.º Registros	Equipo 1	Equipo 2	Diferencia
Personas	225 672	2 h 11 min	2 h 45 min	34 min
Clientes	73 529	13 min	20 min	7 min
Saldos	779 884	1 d 2 h	1 d 14 h	12 h
Historial 2017	5 678 281	6 d 17 h	7 d 5 h	11 h
Historial 2018	5 396 158	7 d	7 d 12 h	12 h
Historial 2019	3 916 502	4 d 19 h	5 d 3 h	8 h
Solicitud Crédito	62 317	6 min	10 min	4 min
Historia Plazo	3 175 308	3 d 18 h	5 d 13 h	1 d 19 h

Realizado por: Los autores

La tabla 10, muestra la diferencia entre los tiempos de respuesta de cada equipo para las distintas tablas simuladas, el equipo 1 tarda un tiempo estimado de 6 a 13 minutos

en las tablas con menos de 100 000 registros, de 2 horas a 1 día 2 horas en las que superan los 100 000 registros y las tablas que superan el 1 000 000 de 3 a 7 días.
El equipo 2 completa las simulaciones en un tiempo estimado de 10 a 20 minutos en las tablas con menos de 100 000 registros, en las tablas que superan los 100 000 de 2 horas a 1 día 14 horas y de 5 a 7 1/2 días en las tablas que superan el 1 000 000 de registros.

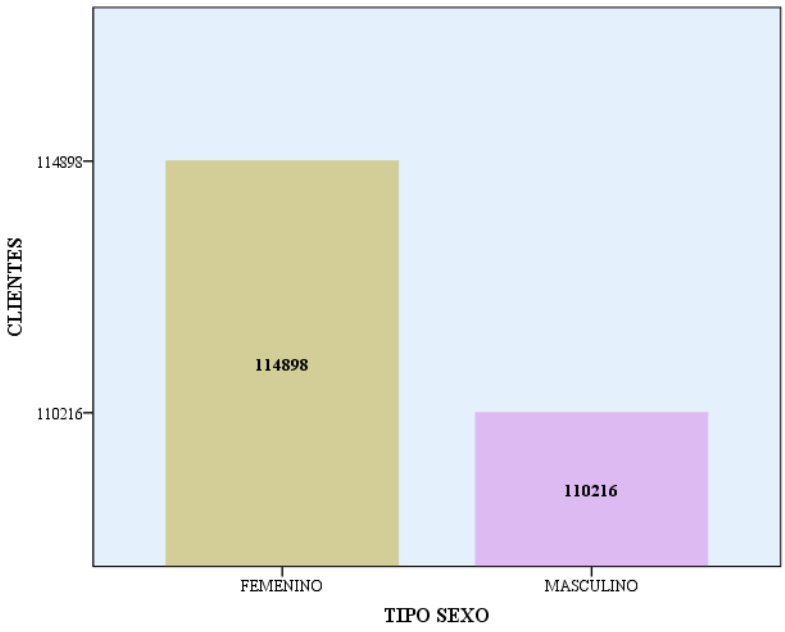
4.6. Interpretación y Resultados

Los resultados que se reflejan en la tabla personas corresponden a los clientes que pertenecen a las 13 sucursales de la COAC "Riobamba" Ltda.

TABLA PERSONAS

Tipo Sexo

Ilustración 5: Grafica Tipo Sexo

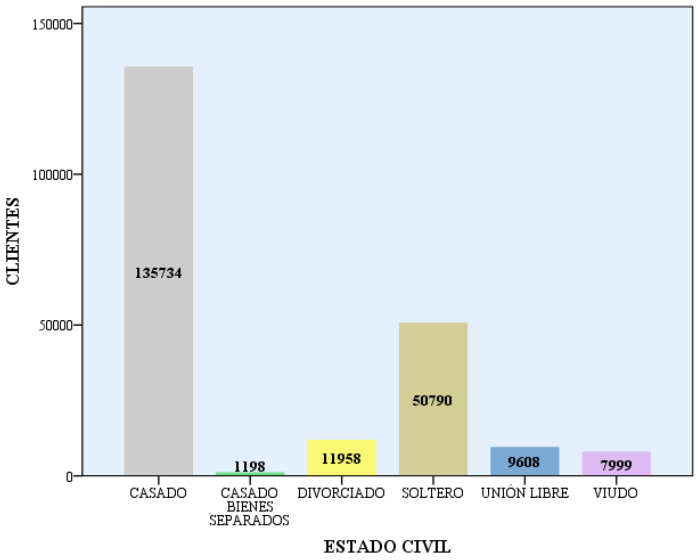


Realizado por: Los autores

La ilustración 5, muestra el número de clientes de género femenino y masculino, se observa que el 51% son mujeres y el 49% son hombres.

Estado Civil

Ilustración 6: Grafica Estado Civil

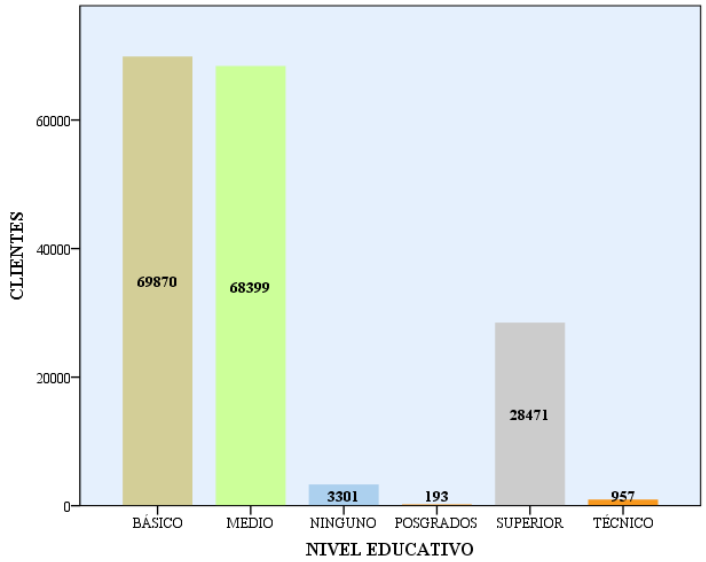


Realizado por: Los autores

La ilustración 6, indica cada categoría de estado civil, se observa que el 62% de clientes son casados, seguido del 23% son solteros.

Nivel Educativo

Ilustración 7: Grafica Nivel Educativo

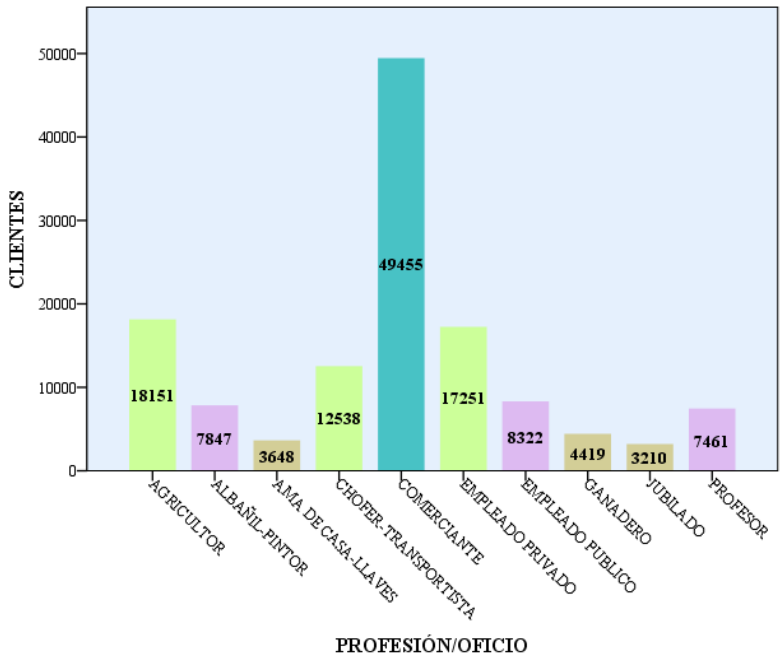


Realizado por: Los autores

La ilustración 7, muestra los niveles educativos de los clientes, el 41% tienen estudios básicos, el 40% estudios medios y el 17% tienen estudios superiores.

Profesión

Ilustración 8: Grafica Profesión

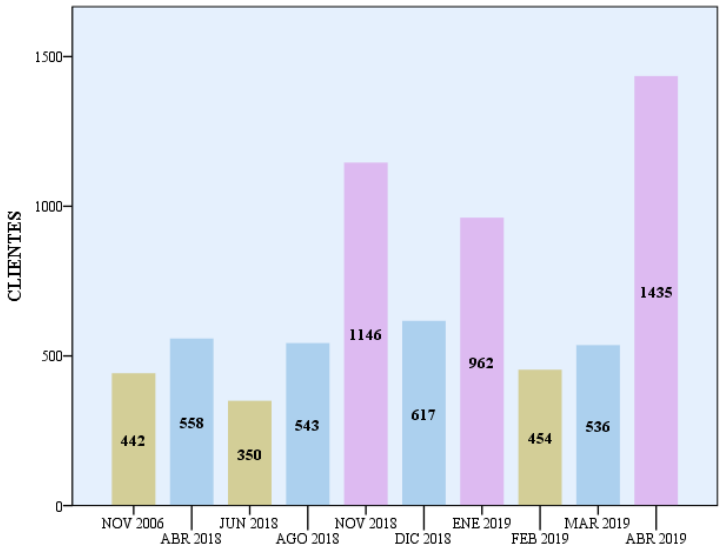


Realizado por: Los autores

La ilustración 8, refleja las distintas profesiones u oficios, se observa que el 37% de clientes se dedican al comercio, el 14% a la agricultura, el 13% tienen sus empleos en el sector privado y el 10% se dedica al transporte.

Fecha de Ingreso

Ilustración 9: Grafica Fecha Ingreso

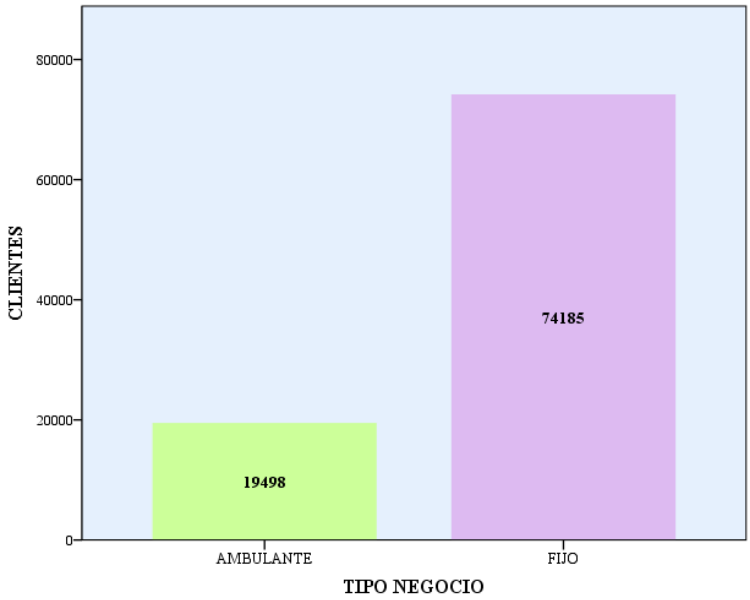


Realizado por: Los autores

La ilustración 9, indica los meses de mayor ingreso de clientes a la Cooperativa, se observa que hubo un 20% de apertura de cuentas en el mes de abril de 2019, seguido del 16% en noviembre de 2018 y un 14% en enero 2019.

Tipo de Negocio

Ilustración 10: Grafica Tipo de Negocio



Realizado por: Los autores

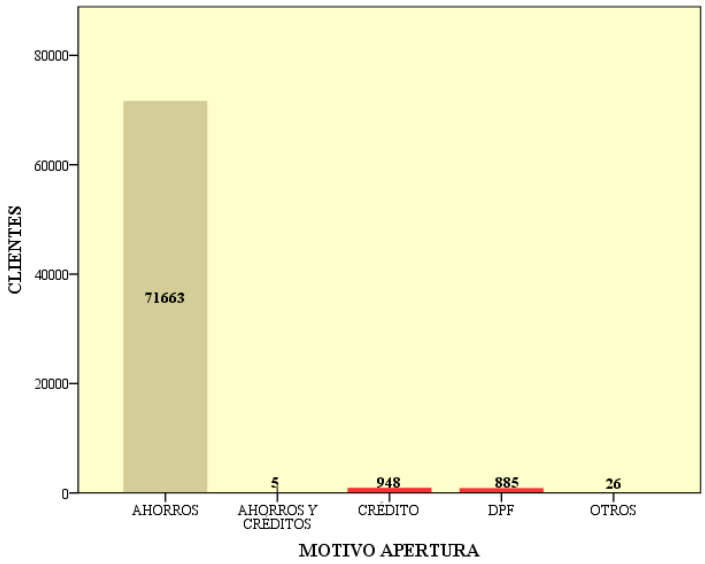
La ilustración 10, muestra el tipo de negocio, se observa que el 79% de los clientes que se dedican al comercio tienen un negocio fijo.

Los resultados que a continuación y hasta finalizar se muestran, corresponden a los clientes de la matriz de la COAC "Riobamba" Ltda.

TABLA CLIENTES

Motivo Apertura vs. Motiva Baja

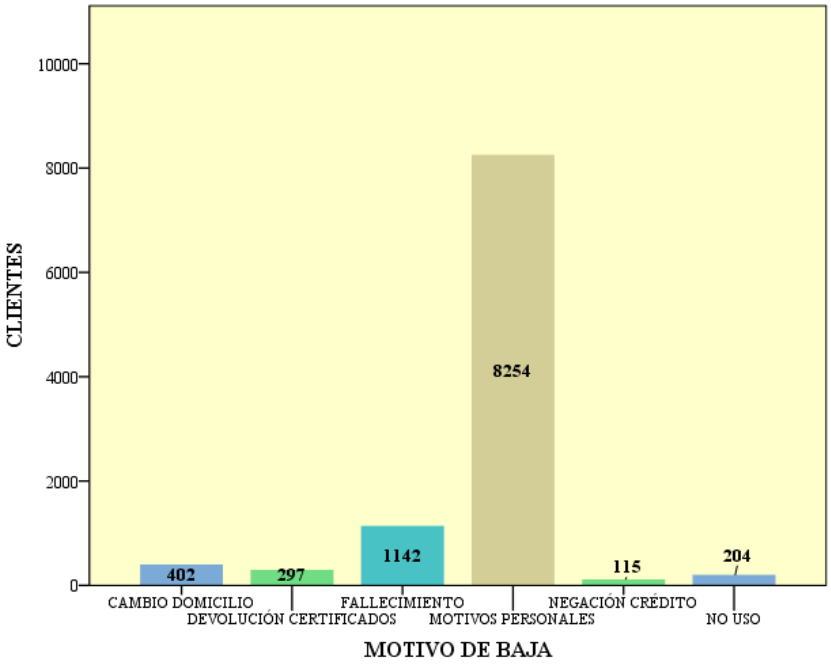
Ilustración 11: Grafica Motivo Apertura



Realizado por: Los autores

La ilustración 11, indica el motivo de apertura de cuentas de los clientes, se observa que para el 97% de clientes el motivo principal de apertura es el de ahorros

Ilustración 12: Grafica Motivo Baja

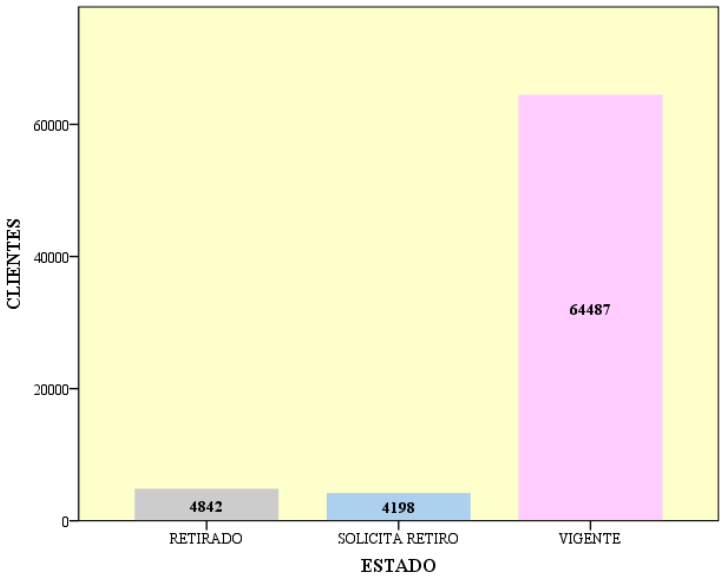


Realizado por: Los autores

La ilustración 12, muestra el motivo de baja de cuentas de los clientes, se observa que el 79% lo hace por motivos personales y el 11% por fallecimiento.

Estado

Ilustración 13: Grafica Estado

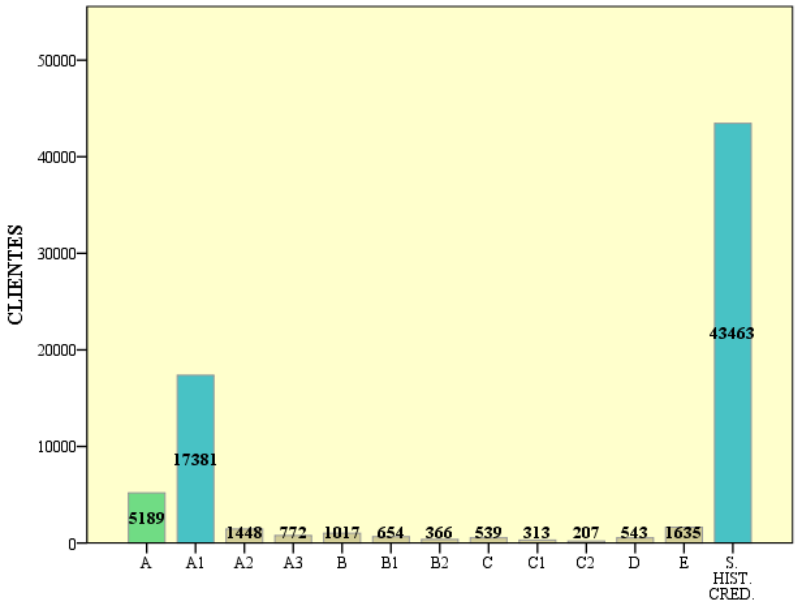


Realizado por: Los autores

La ilustración 13, refleja el estado de la cuenta de los clientes, se observa que el 87% están en estado vigente, el 7% son cuentas retiradas y el 6% en solicitud de retiro.

Calificación Interna

Ilustración 14: Grafica Calificación Interna

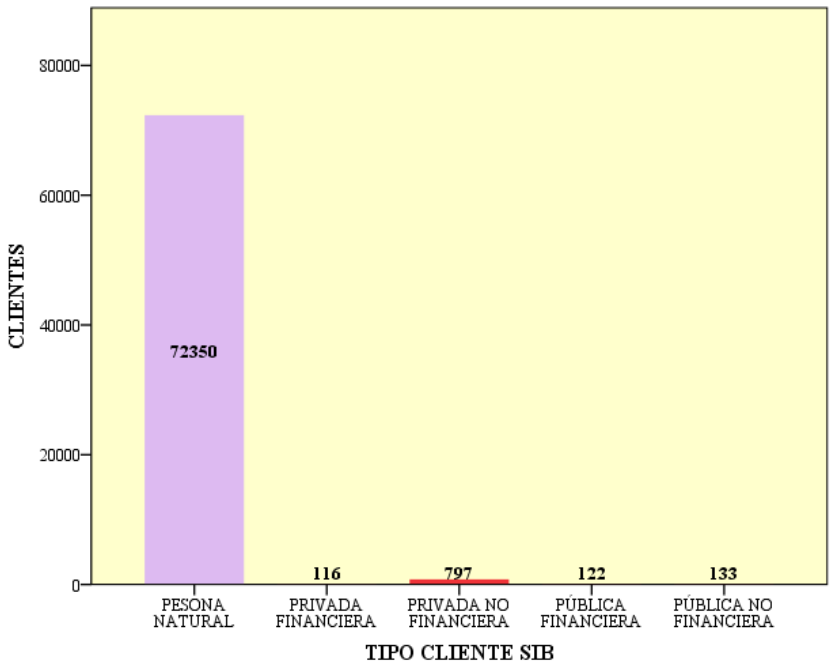


Realizado por: Los autores

La ilustración 14, muestra la calificación interna que se les da a los clientes, se observa que el 59% de clientes no cuenta con historial crediticio, el 24% tienen una calificación de "A1" y un 7% la calificación "A", lo que indica que pertenecen a la clasificación de riesgo normal de morosidad. (Ver Tabla 19)

Calificación Super Intendencia de Bancos (SIB)

Ilustración 15: Grafica Calificación SIB



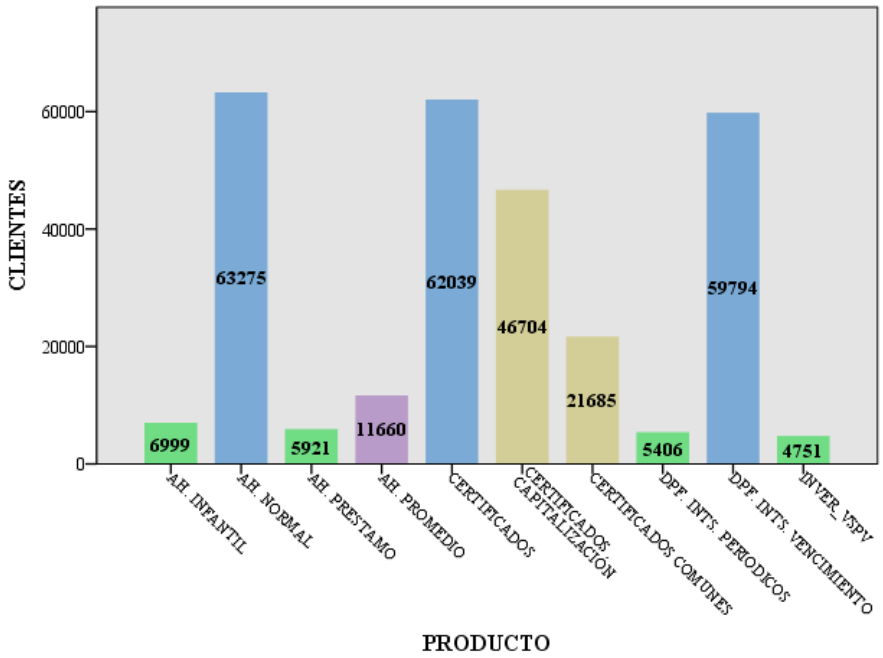
Realizado por: Los autores

La ilustración 15, indica la calificación de los clientes de acuerdo con la SIB, se observa que el 98% son personas naturales, se entiende por persona natural como la persona humana que ejerce derechos y cumple obligaciones a título personal.

TABLA SALDOS

Producto

Ilustración 16: Grafica Producto Ahorros

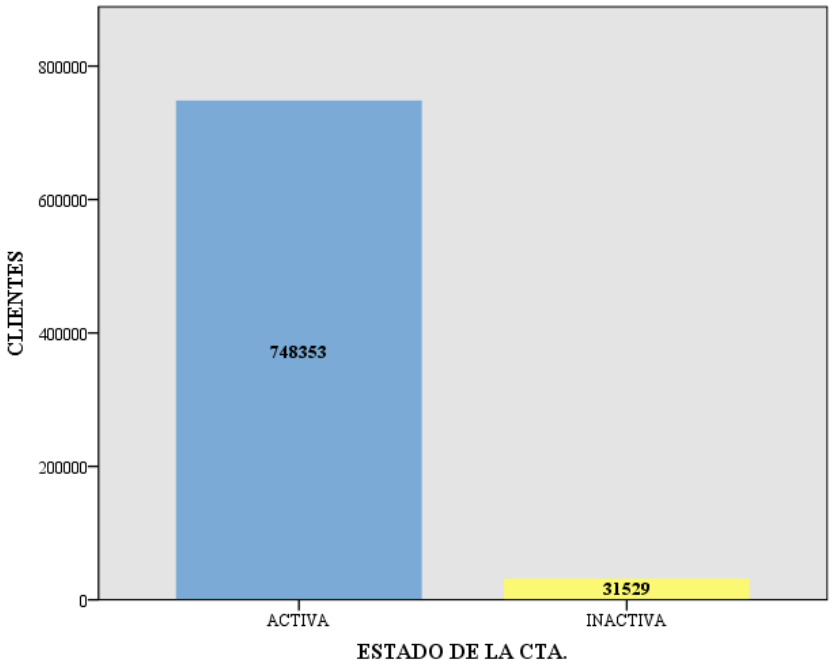


Realizado por: Los autores

La ilustración 16, muestra las distintas categorías de ahorros, el 22% de clientes acceden al ahorro normal, el 22% invierten en certificados y el 21% acceden a DPF. INTS. VENCIMIENTO que son depósitos a plazo fijo con intereses a cobrar cuando vence la póliza.

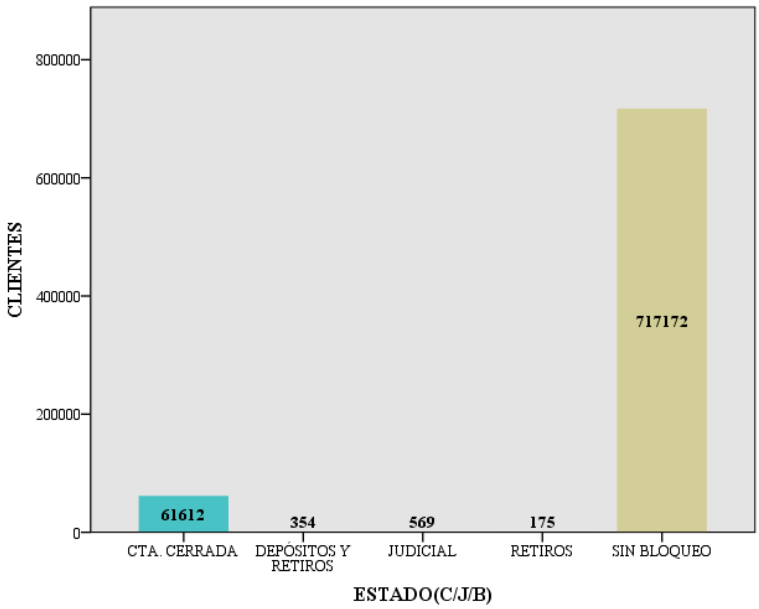
Tipo Cuenta vs. Estado

Ilustración 17: Grafica Estado de la Cuenta



Realizado por: Los autores

Ilustración 18: Grafica Estado (Cerrada/Judicial/Bloqueo)

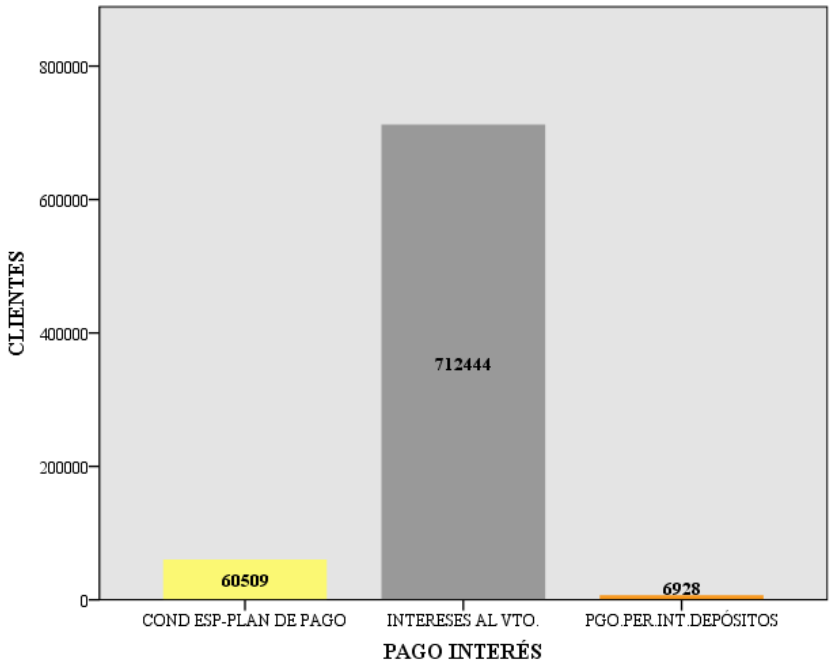


Realizado por: Los autores

Las ilustraciones 17 y 18, indican las categorías del estado de una cuenta, del 96% de las cuentas que se encuentran activas, el 92% están sin bloqueo.

Tipo Pago de Intereses

Ilustración 19: Grafica Tipo Pago Intereses



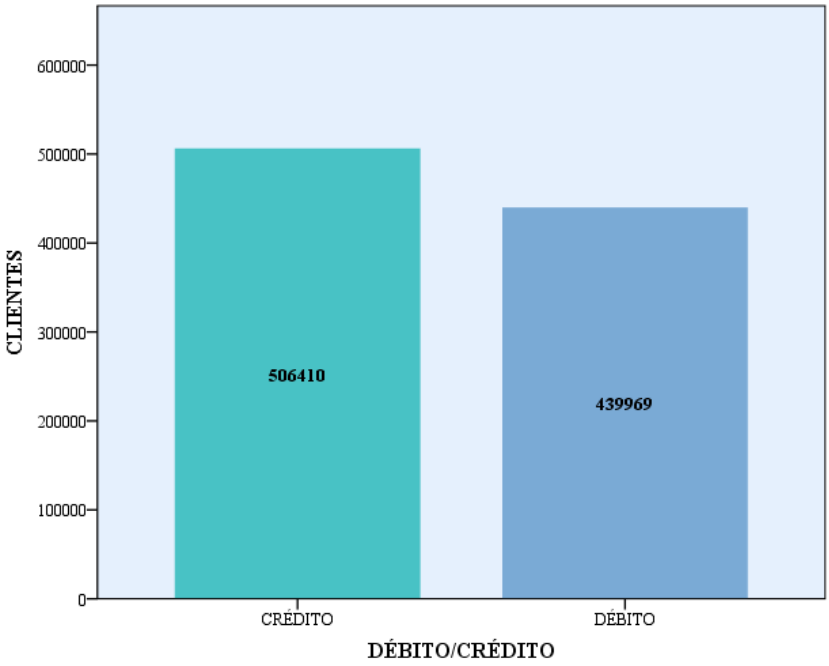
Realizado por: Los autores

La ilustración 19, muestra los tipos de pago de interés a los clientes, se observa que al 91% se les paga “Intereses al Vencimiento” lo que significa que se cancela al cliente el pago cuando se cumpla la fecha de vencimiento del producto, dependiendo de la línea de ahorro al que accedan. (Ver Ilustración 16)

TABLA HISTORIAL 2017

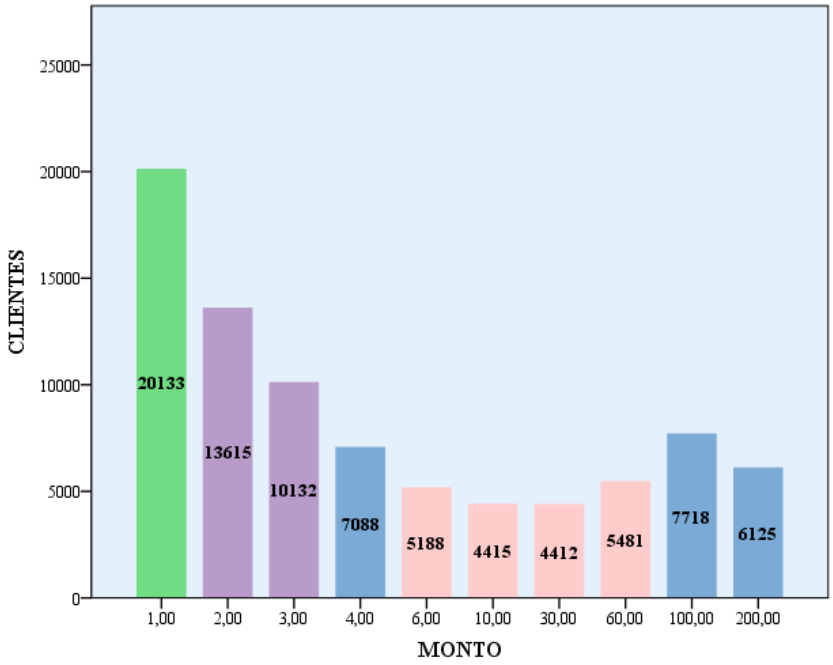
Débito/Crédito vs. Monto

Ilustración 20: Grafica Débito/Crédito 2017



Realizado por: Los autores

Ilustración 21: Grafica Montos 2017



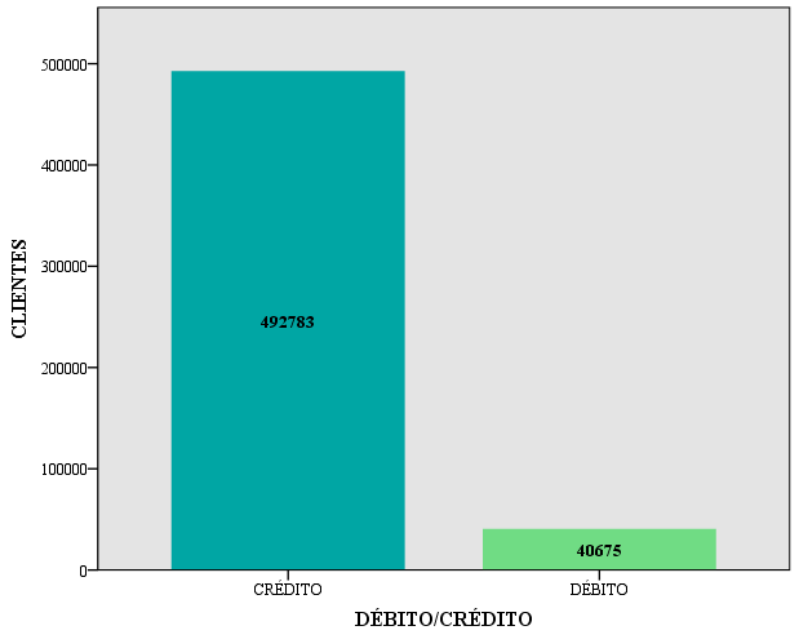
Realizado por: Los autores

Las ilustraciones 20 y 21, indican una comparación entre los débitos (ingresos) y créditos (egresos) que a diario se realizan y los montos que se debitan o acreditan a los clientes, se observa que el 54% de los clientes tienen más egresos, de estos la COAC “Riobamba” Ltda., debita al 24 % de clientes un dólar, al 16% dos y al 12% tres dólares.

TABLA HISTORIAL 2018

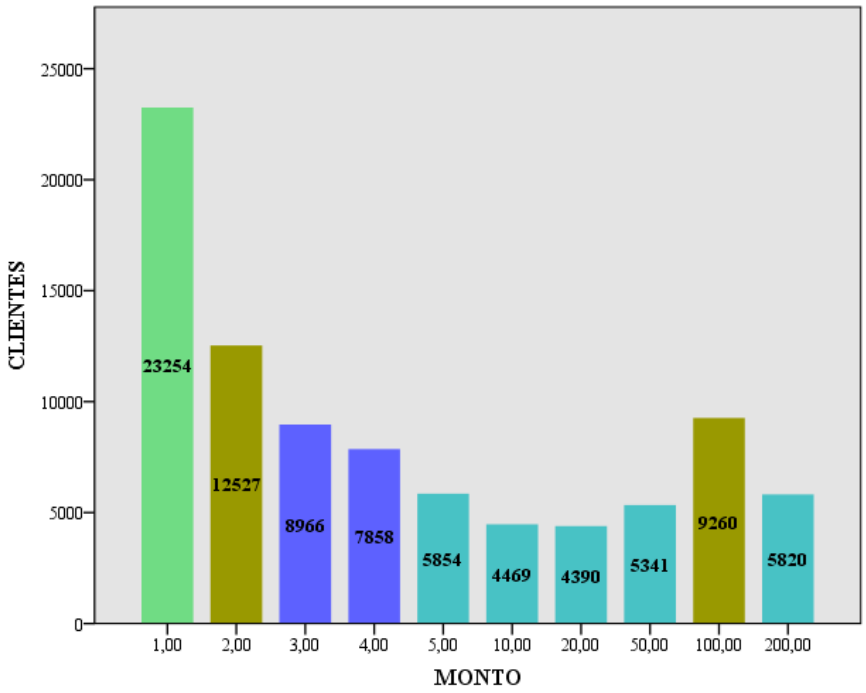
Débito/Crédito vs. Monto

Ilustración 22: Grafica Débito o Crédito 2018



Realizado por: Los autores

Ilustración 23: Grafica Montos 2018



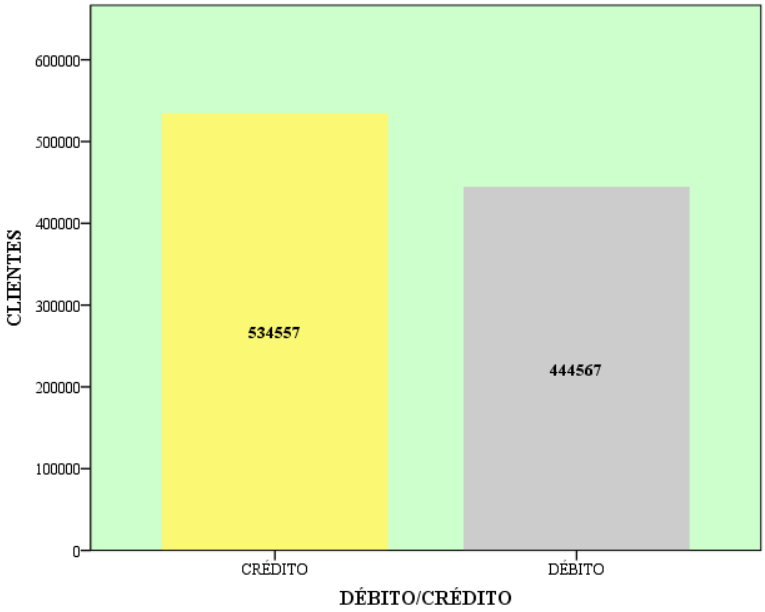
Realizado por: Los autores

Las ilustraciones 22 y 23, indican una comparación entre los débitos (ingresos) y créditos (egresos) que a diario se realizan y los montos que se debitan o acreditan a los clientes, se observa que el 92% de los clientes tienen más egresos, de estos la COAC "Riobamba" Ltda., debita al 27 % un dólar, al 14% dos y al 11% 100 dólares.

TABLA HISTORIAL 2019

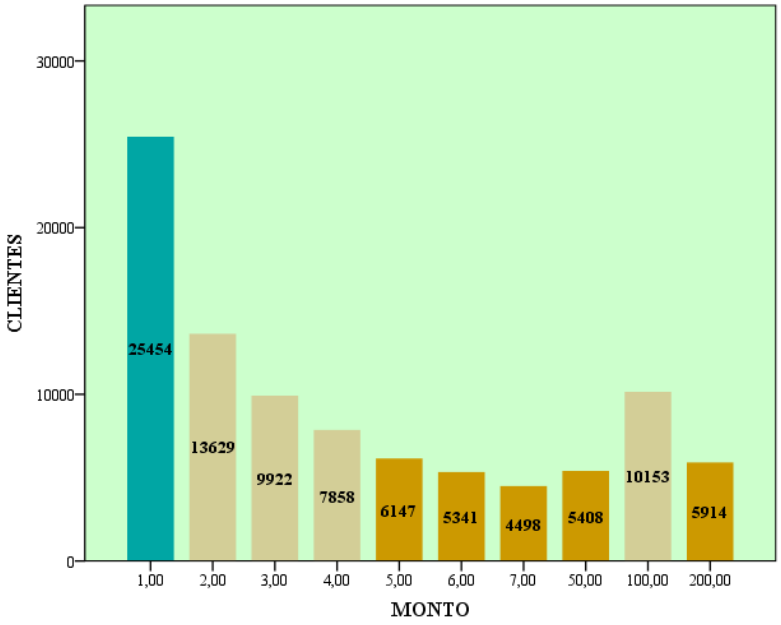
Débito/Crédito vs. Monto

Ilustración 24: Grafica Débito o Crédito 2019



Realizado por: Los autores

Ilustración 25: Grafica Montos 2019



Realizado por: Los autores

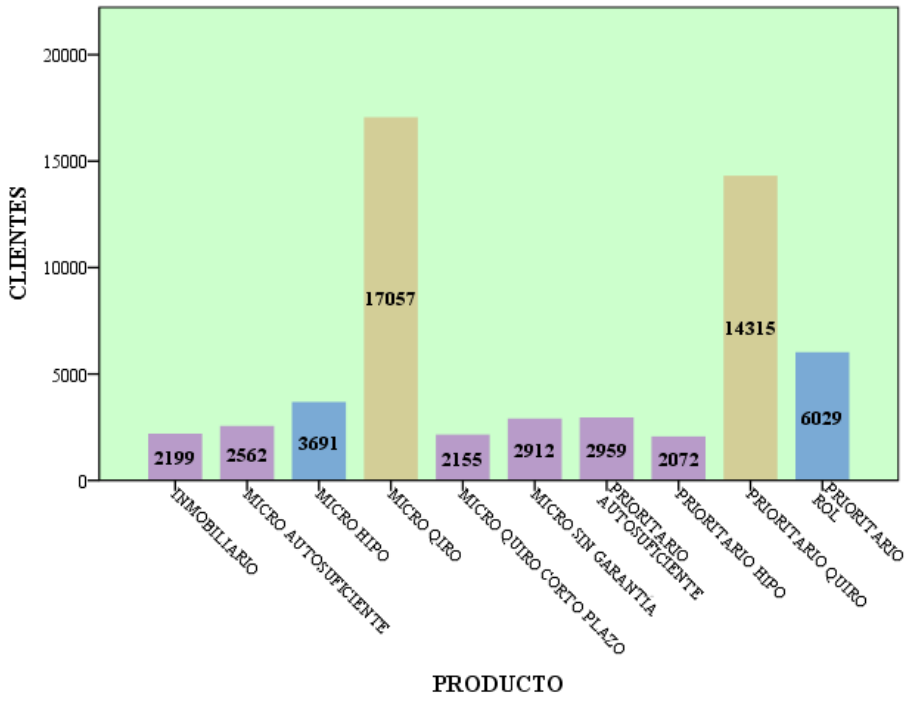
Las ilustraciones 24 y 25, indican una comparación entre los débitos (ingresos) y créditos (egresos) que a diario se realizan y los montos que se debitan o acreditan a los clientes, se observa que el 55% de los clientes tienen más egresos, de estos

la COAC “Riobamba” Ltda., debita al 27 % un dólar, al 14% dos y al 11% 100 dólares.

TABLA SOLICITUD DE CRÉDITO

Producto

Ilustración 26: Grafica Producto Créditos

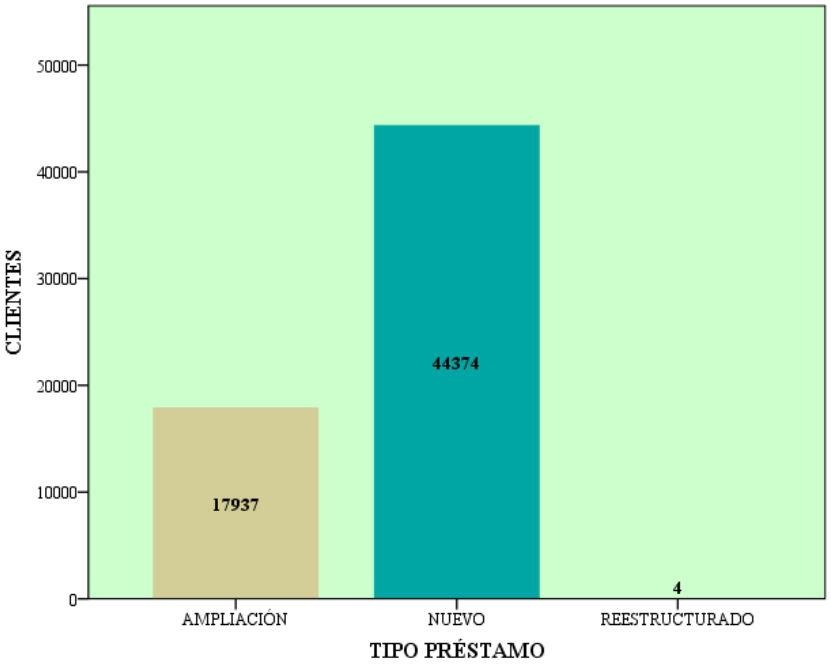


Realizado por: Los autores

La ilustración 26, muestra las distintas categorías de créditos, se observa que el 30% acceden a Micro Quiro que son créditos para microempresarios con garantía quirografaria, además el 26% acceden a Prioritario Quiro que son créditos para empleados con garantía quirografaria (solo basta la firma de quien presta el producto).

Tipo Préstamo

Ilustración 27: Grafica Tipo Préstamo

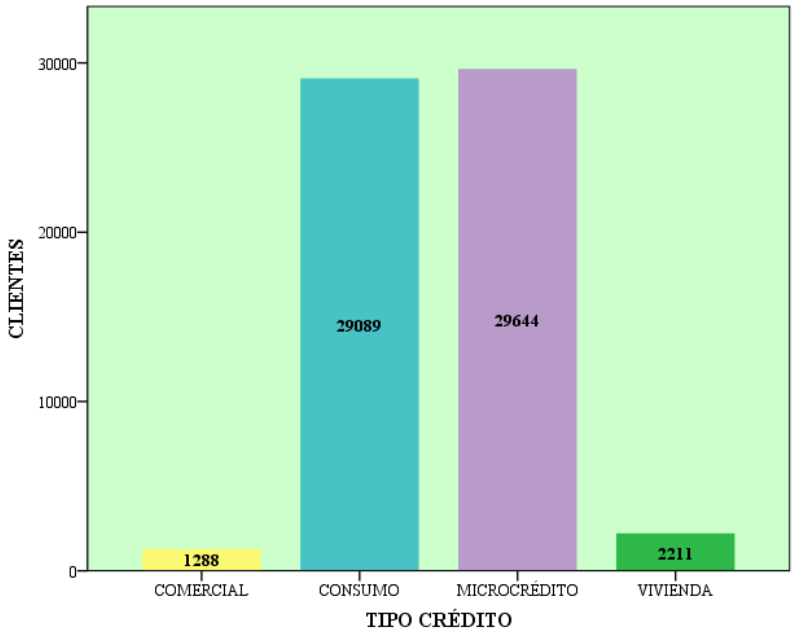


Realizado por: Los autores

La ilustración 27, indica los tipos de préstamo, se observa que el 71% de clientes acceden a préstamos nuevos, el 29% a ampliaciones que consisten en aumentar el importe prestado.

Tipo Crédito

Ilustración 28: Grafica Tipo Crédito

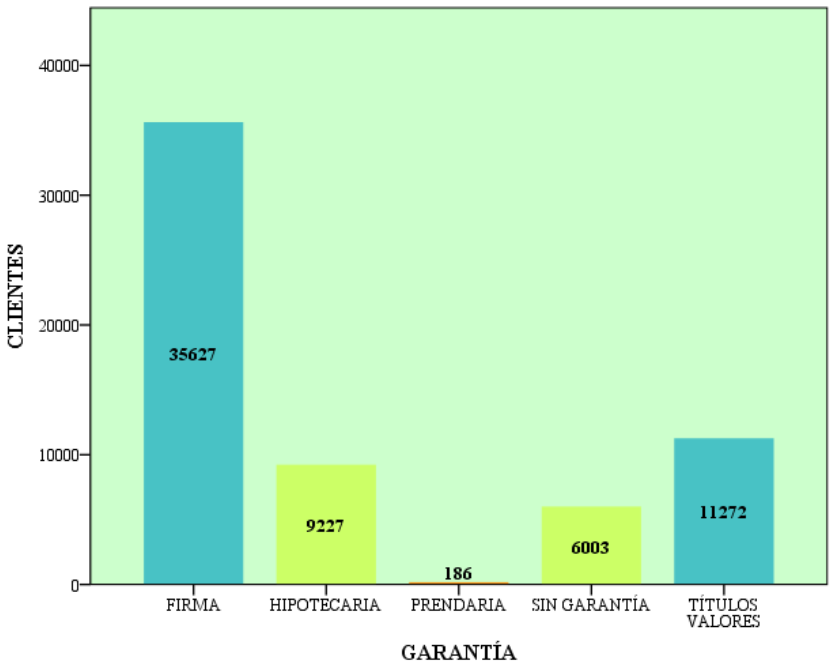


Realizado por: Los autores

La ilustración 28, muestra los tipos de crédito, se observa que el 48% de clientes acceden al Microcrédito que es un crédito que va dirigido a pequeños y medianos negocios sean estos formales e informales, el 47% acceden al de Consumo, crédito destinado para todas aquellas personas naturales que trabajan en relación de dependencia o perciben un sueldo.

Tipo Garantía

Ilustración 29: Grafica Tipo Garantía

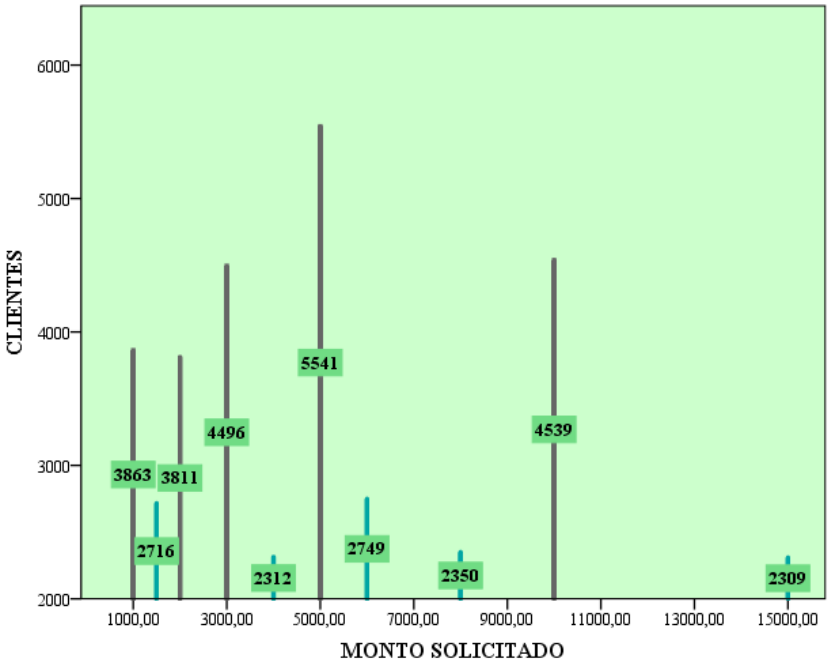


Realizado por: Los autores

La ilustración 29, indica los tipos de garantía, se observa que el 57% de clientes prestan la firma y el 18% los títulos de propiedad.

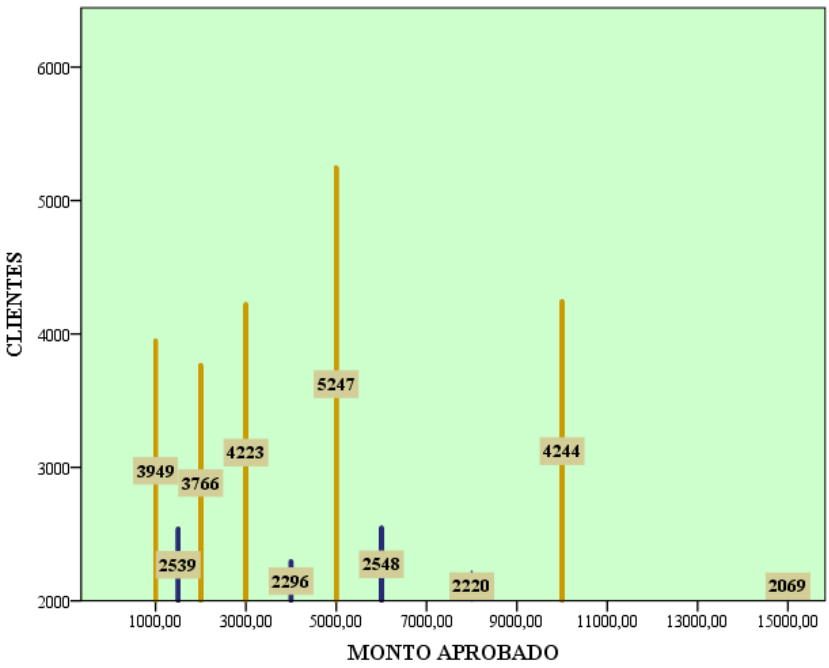
Monto Solicitado vs. Monto Aprobado

Ilustración 30: Grafica Monto Solicitado de Crédito



Realizado por: Los autores

Ilustración 31: Grafica Monto Aprobado de Crédito



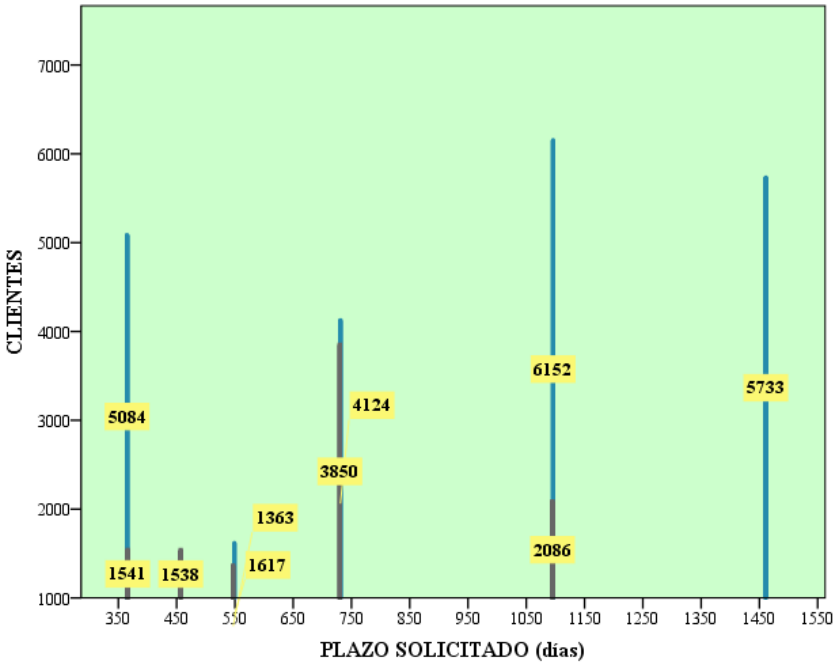
Realizado por: Los autores

Las ilustraciones 30 y 31, muestran una comparación entre los montos solicitados y aprobados de crédito, el 42% de clientes solicitan montos que van desde los 3

000 a 10 000 dólares y como se observa a este mismo porcentaje se les aprueban estos montos.

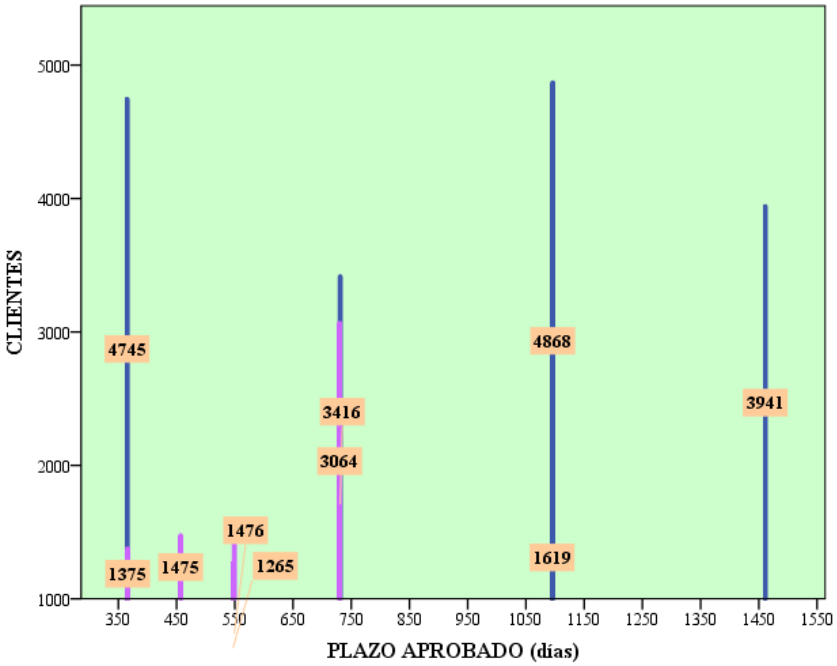
Plazo Solicitado vs. Plazo Aprobado

Ilustración 32: Grafica Plazo Solicitado de Crédito



Realizado por: Los autores

Ilustración 33: Grafica Plazo Aprobado de Crédito



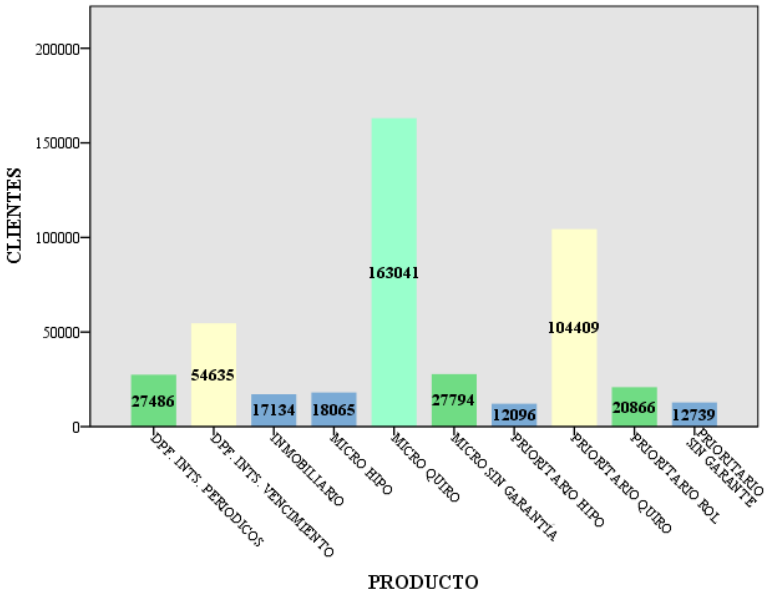
Realizado por: Los autores

Las ilustraciones 32 y 33, muestran una comparación entre los plazos solicitados y aprobados de crédito, el 51% de clientes solicitan plazos que van desde los 3 a 4 años y como se observa al 49% se les aprueban estos plazos.

TABLA HISTORIA PLAZO

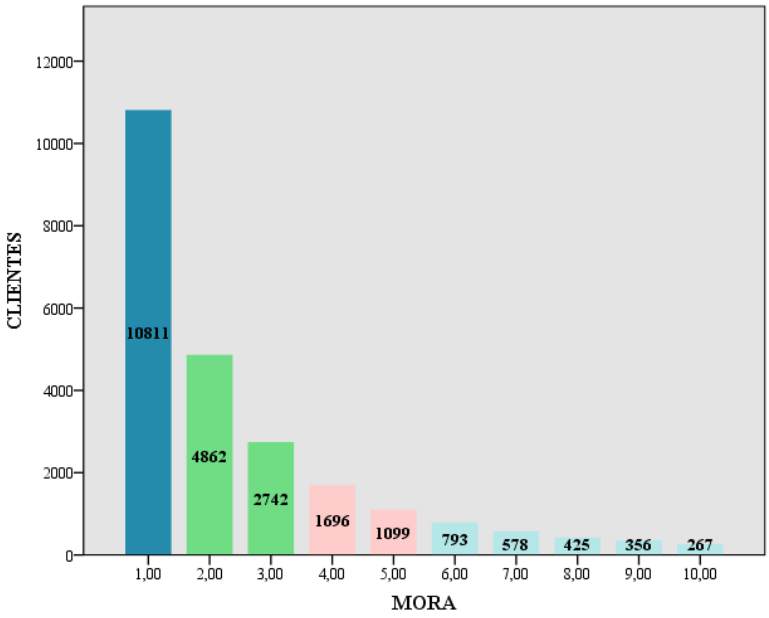
Producto vs. Mora

Ilustración 34: Grafica Producto Crédito



Realizado por: Los autores

Ilustración 35: Grafica Mora Pagada



Realizado por: Los autores

Las ilustraciones 34 y 35, indican una comparación entre las distintas categorías de crédito y la mora que pagan de estos, se observa que los tipos de créditos a los que más acceden son Micro Quiro (36%) y Prioritario Quiro (23%), de acuerdo con ello el 48% pagan mora de un dólar, el 21 % de dos dólares y el 12% tres dólares, lo que indica que pagan al día las cuotas de crédito.

DISCUSIÓN

La presente investigación guarda relación con estudios realizados a nivel internacional, uno de ellos la "Predicción de incumplimiento de pago de clientes de tarjetas de crédito, con aplicación del algoritmo del k-vecino más cercano y Clas- FriedmanAligned-ST.", realizado en la Universidad Nacional de San Agustín, Arequipa, Perú en el año 2017, los resultados experimentales de la investigación fueron alentadores mostrando que el modelo que emplearon fue capaz de alcanzar una buena precisión en la predicción de clientes y las deudas. Por otra parte, en la Universidad Distrital Francisco José De Caldas, Bogotá, Colombia en el año de 2017 se llevó a cabo la "Implementación de la Técnica de lo K-Vecinos en un Algortimo de Recomendación para un Sistemas de Compras utilizando NFC y Android", de la investigación se concluyó que la utilización del algoritmo K-Vecino más cercano proporcionó a los comerciantes una ventaja estratégica frente a sus competidores puesto que les permitió posicionar una mayor cantidad de productos al alcance de sus clientes.

En el presente proyecto de investigación se aplicó el algoritmo k vecino más cercano en el área de negocios de la COAC "Riobamba" Ltda., con el objetivo de predecir clientes potenciales a los cuales llegar con ofertas de crédito. Los resultados arrojados al aplicar el algortimo en *Weka* fueron eficaces debido a que se clasificaron las instancias como correctas en un rango del 70% al 99% de éxito, lo que lleva a concluir que las predicciones sobre clientes potenciales son efectivas. Los resultados obtenidos en esta y en anteriores investigaciones son evidencia de la validez y eficacia del algoritmo de minería de datos k vecino más cercano.

CAPÍTULO V

5. CONCLUSIONES Y RECOMENDACIONES

5.1. CONCLUSIONES

- La aplicación del algoritmo k vecino más cercano en la investigación fue eficaz, debido a que se clasificaron las instancias de la base de datos como correctas, en un rango del 70% al 99%, lo que permitió perfilar con éxito a los clientes potenciales a los cuales llegar con nuevas ofertas de ahorro y crédito.

- Los meses de mayor apertura de cuentas fueron: abril 2019 con un 20%, 16% en noviembre de 2018 y 14% en enero de 2019, de estos clientes el 51% son mujeres y el 49% son hombres, en cuanto al estado civil el 62% son casados y el 23% solteros.

- El 41% de los clientes tienen estudios básicos, el 40% medios y el 17% superiores. De estos, el 37% se dedican al comercio, el 14% a la agricultura, el 13% tienen sus empleos en el sector privado y el 10% se dedica al transporte. De los clientes que se dedican al comercio el 79% cuentan con un negocio propio, por lo que se concluye que son buenos perfiles para acceder a los servicios de la Cooperativa.

- El 97% de clientes apertura sus cuentas por ahorros y el 79% dan de baja por motivos personales y 11% por fallecimiento, en cuanto a la calificación interna que se les otorga a los clientes, el 24% tiene una calificación de “A1” y el 7% “A” lo que los convierte en buenos perfiles al pertenecer a la clasificación de riesgo normal de morosidad (0-30 días/plazo mora).

- Las líneas de ahorro a las que más acceden los clientes son: el 22% acceden al ahorro normal, el 22% invierten en certificados y el 21% acceden a depósitos de plazo fijo.

- En cuanto a líneas de créditos, 30% de clientes acceden a Micro Quiro y el 26% a Prioritario. En cuanto a tipos de crédito, el 48% accede al Microcrédito y el 47% al de Consumo. Además, al 57% de clientes se otorga créditos con la garantía de

la firma y al 18% con los títulos de propiedad. Los montos que se otorgan son de 3 000 a 10 000 dólares para plazos de pago de tres a cuatro años, igualmente se evidencia que el 81% de clientes pagan valores de mora de uno a tres dólares, de lo que se concluye, no se retrasan en el pago de las cuotas.

5.2. RECOMENDACIONES

- Aplicar el algoritmo k vecino más cercano (KNN) en investigaciones dentro de las cuales se manejen grandes volúmenes de datos, debido a la tolerancia al ruido que tiene KNN.

- Utilizar una arquitectura de hardware y software más eficiente a la utilizada en la presente investigación de manera que se ahorre tiempo y esfuerzo en la obtención de resultados al aplicar el algoritmo KNN.

- Para una mejor aplicación del algoritmo, en el caso de tener bases de datos con tablas que superen el millón de registros, dividir estas en tablas con menos registros y analizarlas por separado.

- Innovar y mantener los beneficios de las líneas de ahorros que existen, debido a que la calificación de los clientes que acceden a estas líneas es excelente, por lo que no sería difícil llegar a ellos con nuevas prestaciones.

- Analizar la posibilidad de aumentar el monto aprobado de crédito y el plazo de pago de estos para los clientes que solicitan los créditos Micro Quiro, Prioritario Quiro, Microcrédito y crédito de Consumo debido a que, como se aprecia en los resultados son los créditos a los que más acceden los clientes. Además, promocionar los beneficios de las otras líneas de crédito de manera que tenga mayor afluencia de clientes.

- Se recomienda profundizar la problemática en investigaciones futuras y se aplique en otras instituciones de forma que la investigación sirva como base para futuros estudios.

BIBLIOGRAFÍA

Cazorla, A., Olmo, F. J., & Alados-Arboledas, L. (2005). Estimación de la cubierta nubosa en imágenes de cielo mediante el. 336-337.

(UIAF), U. d. (2014). *Técnicas de minería de datos para la detección y prevención del lavado de activos y la financiación del terrorismo (LA/FT).* Colombia.

Ávila, J. S. (2005). *Sistema de Administración de Red (S.A.R.) Versión 1.0.* México.

Camana, R., & Torres, R. (2017). Descubrimiento del estilo de aprendizaje de estudiantes de la carrera de Tecnología en Análisis de Sistemas. *In Crescendo*, 199-200.

Castañeda, M. B., Cabrera, A. F., Navarro, Y., & Wietse de Vries. (2010). *Procesamiento de datos y análisis estadísticos utilizando SPSS.* Porto Alegre: ediPUCRS.

Deepashri, K. S., & Ashwini, K. (2017). Survet on Techniques of Data Mining and its Applications. *International Journal of Emerging Research in Management & Technology*, 199.

Department of Economics, & Payame Noor University. (2013). Application of K-Nearest Neighbor (KNN) Approach for Predicting Economic Events: Theoretical Background. *S B Imandoust et al. Int. Journal of Engineering Research and Applications*, 609.

Emura, Y. (14 de Noviembre de 2011). *Emurasoft*. Obtenido de Emurasoft: https://www.emeditor.com/reviews/emeditor-simply-the-best-text-editor-for-windows/

Emurasoft. (2019). *EmurasoftEmEditor*. Obtenido de EmurasoftEmEditor: https://www.emeditor.com

Esteban, Á. (2008). *Principios de marketing.* ESIC Edi-torial. Tercera edición.

García Cambronero, C., & Gómez Moreno, I. (s.f.). ALGORITMOS DE APRENDIZAJE: KNN&KMEANS.

Hirudkar, A. M., & Mrs. Sherekar, S. S. (2013). Comprative Analysis of Data Mining Tools and Techniques for Evaluating Performance of Data System. *International Journal od Computer Science And Applications*, 233.

López, C. P. (2007). *Minería de Datos: técnicas y herramientas.* Editorial Paraninfo.

Management Solutions. (2018). *Machine Learning, una pieza clave en la transformación de los modelos de negocio.* España.

Mistry, R., & Misner, S. (2012). Introducing Microsoft® SQL Server 2012. *Microsoft Press*.

Moreno , F. (2004). Clasificadores eficaces basados en algoritmos rápidos de búsqueda del vecino más cercano. 56-89.

Moujahid, A., Inza , I., & Larrañaga, P. (s.f.). Clasificadores K-NN. 1.

Mulak, P., & Talhar, N. (2013). Analysis of Distance Measures Using K-Nearest Neighbor Algorithm on KDD Dataset. *International Journal of Science and Research (IJSR).*

Payan, R., Lei, T., & Huan, L. (2008). Cross-Validation. 1.

Ramageri, B. M. (2010). DATA MINING TECHNIQUES AND APPLICATIONS. *Indian Journal of Computer Science and Engineering*, 302.

Rangra, K., & Dr. Bansal, K. L. (2014). Comparative Study of Data Mining Tools. *Interenational Journal of Advanced Research in Computer Science and Software Engineering*, 220.

Riquelme , J. C., Ruiz, R., & Gilbert, K. (2006). Minería de Datos: Conceptos y Tendencias. *Revista Iberoamericana de Inteligencia Artificial*, 14.

Santamaría, J., & Hernández, J. (s.f.). SQL SERVER VS MySQL. 1.

Timarán-Pereira, S. R., Hernández-Arteaga, I., Caicedo-Zambrano, S. J., Hidalgo-Troya, A., & Alvarado-Pérez, J. C. (2016). El proceso de descubrimiento de conocimiento en bases de datos. *Ediciones Universidad Cooperativa de Colombia*, 64-65.

Wolfgang, K. H. (2015). Time Series Data Mining Methods: A Review.

ANEXOS

ANEXO I:

Tablas utilizadas de la Base de Datos de la COAC "Riobamba" Ltda.

Tabla 11: Tabla Personas BD

Tipo Sexo	Tipo Estado Civil	Tipo Actividad	Patrimonio	Numero Personas a Cargo	Antigüedad	Tipo Nivel Educativo	Ingresos Sueldos	Tipo Vivienda	Profesión u Oficio	Local Terreno Negocio	Fecha Ingreso	Tipo Negocio
2.0	3.0	2.0	36935.0	3.0	12.0	4.0	500.0	6.0	42.0	7.0	1/24/18 12:00 AM	2.0
1.0	3.0	3.0	2500.0	0.0	0.0	2.0	0.0	0.0	0.0	0.0	4/28/04 12:00 AM	0.0
2.0	3.0	2.0	0.0	0.0	0.0	7.0	0.0	6.0	42.0	0.0	6/20/19 12:00 AM	0.0
1.0	3.0	3.0	200.0	0.0	0.0	4.0	0.0	0.0	0.0	0.0	11/08/2006 0:00	0.0
1.0	5.0	2.0	64245.0	0.0	12.0	2.0	350.0	3.0	42.0	3.0	11/14/17 12:00 AM	2.0
1.0	3.0	2.0	4924.0	2.0	0.0	4.0	400.0	3.0	42.0	3.0	6/28/18 12:00 AM	2.0
1.0	5.0	3.0	1000.0	0.0	0.0	2.0	0.0	3.0	0.0	0.0	7/24/08 12:00 AM	0.0
2.0	3.0	1.0	100.0	0.0	0.0	0.0	600.0	6.0	5.0	0.0	6/20/19 12:00 AM	0.0

1.0	3.0	2.0	55586.0	3.0	0.0	4.0	360.0	6.0	3.0	7.0	10/24/14 12:00 AM	2.0
1.0	2.0	1.0	6000.0	0.0	27.0	5.0	980.0	6.0	60.0	0.0	04/01/2016 0:00	0.0
1.0	5.0	3.0	1000.0	0.0	0.0	4.0	0.0	0.0	0.0	0.0	1/30/85 12:00 AM	0.0
2.0	3.0	1.0	135629.0	0.0	4.0	5.0	664.0	6.0	95.0	0.0	10/16/17 12:00 AM	0.0
2.0	6.0	1.0	120575.0	0.0	8.0	5.0	463.0	6.0	95.0	0.0	2/16/18 12:00 AM	0.0
2.0	3.0	1.0	7248.0	0.0	212.0	4.0	246.0	6.0	95.0	1.0	10/03/2012 0:00	2.0
1.0	3.0	1.0	400000.0	0.0	36.0	5.0	947.0	6.0	155.0	0.0	07/04/2017 0:00	0.0
2.0	3.0	2.0	37206.0	0.0	0.0	2.0	335.0	6.0	3.0	3.0	3/30/12 12:00 AM	2.0
1.0	3.0	3.0	0.0	0.0	0.0	7.0	0.0	6.0	0.0	0.0	6/20/19 12:00 AM	0.0
1.0	3.0	2.0	20000.0	1.0	0.0	4.0	0.0	6.0	6.0	4.0	06/07/2011 0:00	1.0
2.0	5.0	3.0	100.0	0.0	0.0	4.0	0.0	0.0	0.0	0.0	7/19/00 12:00 AM	0.0
2.0	5.0	1.0	22000.0	0.0	12.0	4.0	720.0	1.0	128.0	0.0	10/19/15 12:00 AM	0.0
1.0	3.0	2.0	3000.0	0.0	0.0	4.0	0.0	6.0	42.0	0.0	2/19/18 12:00 AM	0.0

1.0	3.0	2.0	200.0	0.0	0.0	2.0	0.0	1.0	59.0	4.0	5/24/18 12:00 AM	2.0
1.0	3.0	3.0	25000.0	0.0	0.0	4.0	0.0	0.0	0.0	0.0	03/02/2005 0:00	0.0
2.0	3.0	2.0	100.0	0.0	0.0	0.0	0.0	6.0	42.0	6.0	10/01/2018 0:00	2.0
…	…	…	…	…	…	…	…	…	…	…	…	…

Tabla 12: Tabla Clientes BD

Fecha Apertura	Fecha Cierre	Fecha UltMov.	Cod. Estado	Tipo Persona	Motivo Baja	Socio	Fecha Re-Apertura	Tipo Cliente SIB	Patrimonio	Sucursal Cliente	Calificacion Cliente	Motivo Apertura
03/01/1988 0:00	10/08/2008 0:00	8/22/07 12:00 AM	3.0	1.0	9.0	1.0	6/20/19 12:00 AM	1.0	1000.0	1.0	0.0	1.0
03/01/1988 0:00	6/20/19 12:00 AM	10/24/06 12:00 AM	5.0	1.0	0.0	1.0	6/20/19 12:00 AM	1.0	2671.0	1.0	2.0	1.0
03/04/1988 0:00	6/20/19 12:00 AM	02/03/1999 0:00	5.0	1.0	0.0	1.0	6/20/19 12:00 AM	1.0	22000.0	1.0	0.0	1.0
03/11/1988 0:00	6/20/19 12:00 AM	9/29/11 12:00 AM	5.0	1.0	0.0	1.0	6/20/19 12:00 AM	1.0	1000.0	1.0	2.0	1.0
03/11/1988 0:00	6/20/19 12:00 AM	08/02/2005 0:00	5.0	1.0	0.0	2.0	6/20/19 12:00 AM	1.0	22030.0	1.0	0.0	1.0
03/12/1988 0:00	6/20/19 12:00 AM	07/05/1995 0:00	5.0	1.0	0.0	1.0	6/20/19 12:00 AM	1.0	1000.0	1.0	0.0	1.0

3/14/88 12:00 AM	6/20/19 12:00 AM	6/23/05 12:00 AM	5.0	1.0	0.0	1.0	6/20/19 12:00 AM	1.0	10000.0	1.0	0.0	1.0
3/15/88 12:00 AM	03/10/2016 0:00	11/29/10 12:00 AM	2.0	1.0	9.0	1.0	6/20/19 12:00 AM	1.0	15990.0	1.0	5.0	1.0
03/02/1988 0:00	12/02/2007 0:00	6/28/07 12:00 AM	2.0	1.0	9.0	1.0	6/20/19 12:00 AM	1.0	122990.0	1.0	1.0	1.0
3/22/88 12:00 AM	6/20/19 12:00 AM	2/17/99 12:00 AM	5.0	1.0	0.0	1.0	6/20/19 12:00 AM	1.0	24000.0	1.0	0.0	1.0
3/22/88 12:00 AM	6/20/19 12:00 AM	08/11/2005 0:00	5.0	1.0	0.0	1.0	6/20/19 12:00 AM	1.0	306000.0	1.0	8.0	1.0
01/07/1994 0:00	05/12/1995 0:00	12/24/08 12:00 AM	5.0	1.0	0.0	1.0	6/20/19 12:00 AM	1.0	19088.0	1.0	0.0	1.0
04/04/1989 0:00	01/04/2013 0:00	08/09/2005 0:00	2.0	1.0	4.0	2.0	6/20/19 12:00 AM	1.0	1000.0	1.0	0.0	1.0
04/04/1989 0:00	03/03/2019 0:00	4/27/99 12:00 AM	2.0	1.0	9.0	2.0	6/20/19 12:00 AM	1.0	1000.0	1.0	0.0	1.0
04/04/1989 0:00	6/20/19 12:00 AM	11/29/10 12:00 AM	5.0	1.0	0.0	1.0	6/20/19 12:00 AM	1.0	6000.0	1.0	0.0	1.0
06/02/2001 0:00	09/07/1994 0:00	07/07/2005 0:00	5.0	1.0	0.0	1.0	6/20/19 12:00 AM	1.0	120000.0	1.0	0.0	1.0
04/12/1989 0:00	6/20/19 12:00 AM	07/07/2001 0:00	5.0	1.0	0.0	1.0	6/20/19 12:00 AM	1.0	1000.0	1.0	0.0	1.0
4/14/89 12:00 AM	6/20/19 12:00 AM	04/12/2011 0:00	5.0	1.0	0.0	1.0	6/20/19 12:00 AM	1.0	864770.0	1.0	2.0	1.0
5/23/01 12:00 AM	12/07/1995 0:00	08/10/2005 0:00	5.0	1.0	0.0	1.0	6/20/19 12:00 AM	1.0	523730.0	1.0	2.0	1.0

4/18/89 12:00 AM	6/20/19 12:00 AM	7/28/05 12:00 AM	5.0	1.0	0.0	1.0	6/20/19 12:00 AM	1.0	1000.0	1.0	1.0	1.0
4/19/89 12:00 AM	01/03/200 8 0:00	2/28/11 12:00 AM	5.0	1.0	9.0	1.0	04/02/201 3 0:00	1.0	20000.0	1.0	0.0	1.0
4/20/89 12:00 AM	6/20/19 12:00 AM	11/06/200 4 0:00	5.0	1.0	0.0	1.0	6/20/19 12:00 AM	1.0	401590.0	1.0	2.0	1.0
4/20/89 12:00 AM	06/05/201 3 0:00	08/05/200 4 0:00	2.0	1.0	9.0	2.0	6/20/19 12:00 AM	1.0	1000.0	1.0	0.0	1.0
4/25/89 12:00 AM	6/20/19 12:00 AM	07/11/200 5 0:00	5.0	1.0	0.0	1.0	6/20/19 12:00 AM	1.0	100000.0	1.0	0.0	1.0
…	…	…	…	…	…	…	…	…	…	…	…	…

Tabla 13: Tabla Saldos BD

Sucursal	Producto	Monto Original	Cuota (Prest.Amort. o DPF)	Fecha Apertura	Nro. Retiros Mes	Tipo Interes	Tipo Pago Intereses	Estado (Cerrado/ Judicial/ Bloqueo)	Cod. Calificacion	Inactiva	Monto Aprobado Prestamo
1.0	0.0	0.0	0.0	4/30/08 12:00 AM	68.0	0.0	0.0	0.0	0.0	1.0	0.0
1.0	0.0	0.0	0.0	8/13/05 12:00 AM	51318.0	0.0	0.0	0.0	0.0	1.0	0.0
1.0	0.0	0.0	0.0	3/20/06 12:00 AM	0.0	0.0	0.0	0.0	0.0	1.0	0.0

1.0	0.0	0.0	0.0	08/12/2005 0:00	322.0	0.0	0.0	0.0	0.0	1.0	0.0
1.0	701.0	0.0	0.0	8/13/05 12:00 AM	15136.0	0.0	0.0	0.0	0.0	1.0	0.0
1.0	701.0	0.0	0.0	03/07/2006 0:00	15045.0	0.0	0.0	0.0	0.0	1.0	0.0
1.0	0.0	0.0	0.0	3/31/08 12:00 AM	1.0	0.0	0.0	0.0	0.0	1.0	0.0
1.0	0.0	0.0	0.0	08/12/2005 0:00	701.0	0.0	0.0	0.0	0.0	1.0	0.0
1.0	0.0	0.0	0.0	04/12/2006 0:00	221.0	0.0	0.0	0.0	0.0	1.0	0.0
1.0	0.0	0.0	0.0	08/12/2005 0:00	122.0	0.0	0.0	0.0	0.0	1.0	0.0
1.0	0.0	0.0	0.0	8/15/05 12:00 AM	3523.0	0.0	0.0	0.0	0.0	1.0	0.0
1.0	0.0	0.0	0.0	08/12/2005 0:00	27987.0	0.0	0.0	0.0	0.0	1.0	0.0
1.0	0.0	0.0	0.0	08/12/2005 0:00	86005.0	0.0	0.0	0.0	0.0	1.0	0.0
1.0	0.0	0.0	0.0	8/17/05 12:00 AM	518.0	0.0	0.0	0.0	0.0	1.0	0.0
1.0	0.0	0.0	0.0	09/01/2005 0:00	1.0	0.0	0.0	0.0	0.0	1.0	0.0
1.0	0.0	0.0	0.0	08/12/2005 0:00	21.0	0.0	0.0	0.0	0.0	1.0	0.0

1.0	701.0	0.0	0.0	8/13/05 12:00 AM	18553.0	0.0	0.0	0.0	0.0	1.0	0.0
1.0	701.0	0.0	0.0	8/13/05 12:00 AM	15407.0	0.0	0.0	0.0	0.0	1.0	0.0
1.0	0.0	0.0	0.0	11/19/08 12:00 AM	0.0	0.0	0.0	0.0	0.0	1.0	0.0
1.0	701.0	0.0	0.0	03/07/2008 0:00	2303.0	0.0	0.0	0.0	0.0	1.0	0.0
1.0	0.0	0.0	0.0	8/15/05 12:00 AM	3350.0	0.0	0.0	0.0	0.0	1.0	0.0
1.0	0.0	0.0	0.0	08/12/2005 0:00	7.0	0.0	0.0	0.0	0.0	1.0	0.0
1.0	0.0	0.0	0.0	08/12/2005 0:00	70289.0	0.0	0.0	0.0	0.0	1.0	0.0
1.0	0.0	0.0	0.0	08/12/2005 0:00	34.0	0.0	0.0	0.0	0.0	1.0	0.0
…	…	…	…	…	…	…	…	…	…	…	…

Tabla 14: Tabla Historial 2017 BD

NUMERO_MOVIMIENTO	FECHA_PROCESADO	SUCURSAL	DEBITO_CREDITO	MONTO
3.0	03/01/2017 0:00	1.0	1.0	70.0
15.0	03/01/2017 0:00	1.0	2.0	70.0
3.0	03/01/2017 0:00	1.0	1.0	50.0
15.0	03/01/2017 0:00	1.0	2.0	50.0
3.0	03/01/2017 0:00	1.0	1.0	200.0
15.0	03/01/2017 0:00	1.0	2.0	200.0
3.0	03/01/2017 0:00	1.0	1.0	150.0
15.0	03/01/2017 0:00	1.0	2.0	150.0
5.0	03/01/2017 0:00	1.0	1.0	175.0
1.0	03/01/2017 0:00	1.0	2.0	92.0
3.0	03/01/2017 0:00	1.0	1.0	88.0
4.0	03/01/2017 0:00	1.0	2.0	167.0
7.0	03/01/2017 0:00	1.0	2.0	3.0
5.0	03/01/2017 0:00	1.0	2.0	1737.0
6.0	03/01/2017 0:00	1.0	1.0	368757.0
7.0	03/01/2017 0:00	1.0	1.0	17.0
3.0	03/01/2017 0:00	1.0	1.0	100.0
15.0	03/01/2017 0:00	1.0	2.0	100.0
15.0	03/01/2017 0:00	1.0	2.0	300.0
3.0	03/01/2017 0:00	1.0	1.0	100.0
15.0	03/01/2017 0:00	1.0	2.0	100.0
16.0	03/01/2017 0:00	1.0	2.0	100.0
17.0	03/01/2017 0:00	1.0	1.0	100.0

3.0	03/01/2017 0:00	1.0	1.0	150.0
…	…	…	…	…

Tabla 15: Tabla Historial 2018 BD

NUMERO_MOVIMIENTO	FECHA_PROCESADO	SUCURSAL	DEBITO_CREDITO	MONTO
3.0	04/01/2018 0:00	1.0	2.0	250.0
12.0	04/01/2018 0:00	1.0	1.0	250.0
6.0	04/01/2018 0:00	1.0	2.0	3.0
16.0	04/01/2018 0:00	1.0	1.0	555.0
1.0	04/01/2018 0:00	1.0	2.0	299.0
2.0	04/01/2018 0:00	1.0	1.0	299.0
3.0	04/01/2018 0:00	1.0	1.0	223.0
5.0	04/01/2018 0:00	1.0	1.0	165.0
5.0	04/01/2018 0:00	1.0	2.0	4.0
15.0	04/01/2018 0:00	1.0	2.0	0.0
22.0	04/01/2018 0:00	1.0	1.0	4.0
17.0	02/01/2018 0:00	1.0	1.0	100.0
18.0	02/01/2018 0:00	1.0	1.0	175.0
19.0	02/01/2018 0:00	1.0	2.0	177.0
23.0	02/01/2018 0:00	1.0	2.0	182.0
3.0	02/01/2018 0:00	1.0	2.0	1.0
12.0	02/01/2018 0:00	1.0	1.0	1.0
16.0	02/01/2018 0:00	1.0	2.0	150.0
17.0	02/01/2018 0:00	1.0	1.0	150.0

2.0	02/01/2018 0:00	1.0	1.0	265.0
3.0	04/01/2018 0:00	1.0	1.0	750.0
15.0	04/01/2018 0:00	1.0	2.0	750.0
3.0	04/01/2018 0:00	1.0	2.0	20.0
12.0	04/01/2018 0:00	1.0	1.0	20.0
…	…	…	…	…

Tabla 16: Tabla Historial 2019 BD

NUMERO_MOVIMIENTO	FECHA_PROCESADO	SUCURSAL	DEBITO_CREDITO	MONTO
32.0	04/01/2019 0:00	1.0	2.0	43.0
31.0	04/01/2019 0:00	1.0	2.0	374.0
30.0	04/01/2019 0:00	1.0	1.0	375.0
29.0	04/01/2019 0:00	1.0	2.0	230.0
28.0	04/01/2019 0:00	1.0	2.0	10.0
27.0	04/01/2019 0:00	1.0	2.0	1201.0
26.0	04/01/2019 0:00	1.0	1.0	1281.0
25.0	04/01/2019 0:00	1.0	2.0	432.0
24.0	04/01/2019 0:00	1.0	2.0	7683.0
23.0	04/01/2019 0:00	1.0	1.0	7914.0
22.0	04/01/2019 0:00	1.0	2.0	2420.0
21.0	04/01/2019 0:00	1.0	2.0	983.0
20.0	04/01/2019 0:00	1.0	1.0	1415.0
19.0	04/01/2019 0:00	1.0	2.0	1406.0
18.0	04/01/2019 0:00	1.0	1.0	1407.0

17.0	04/01/2019 0:00	1.0	2.0	3100.0
16.0	04/01/2019 0:00	1.0	1.0	3111.0
15.0	04/01/2019 0:00	1.0	2.0	2027.0
14.0	04/01/2019 0:00	1.0	1.0	2070.0
13.0	04/01/2019 0:00	1.0	2.0	3461.0
12.0	04/01/2019 0:00	1.0	1.0	3523.0
11.0	04/01/2019 0:00	1.0	2.0	1675.0
10.0	04/01/2019 0:00	1.0	2.0	1257.0
9.0	04/01/2019 0:00	1.0	2.0	1184.0
…	…	…	…	…

Tabla 17: Tabla Solicitud Crédito BD

Ayu.Cod. Sucursal	Tipo Prestamo	Ayu.Producto	Monto Solicitado	Periodo Solicitado	Tipo Credito	Periodo Aprobado	Monto Aprobado	Valuo Hipotecario
1.0	3.0	412.0	1500.0	30.0	4.0	30.0	700.0	0.0
1.0	3.0	412.0	1500.0	30.0	4.0	30.0	1500.0	0.0
1.0	3.0	211.0	1000.0	30.0	2.0	30.0	1000.0	0.0
1.0	1.0	211.0	600.0	30.0	2.0	30.0	600.0	0.0
1.0	3.0	412.0	1600.0	30.0	4.0	30.0	1600.0	0.0
1.0	3.0	411.0	5500.0	30.0	4.0	30.0	5500.0	0.0
1.0	3.0	211.0	600.0	30.0	2.0	30.0	600.0	0.0
1.0	3.0	211.0	400.0	30.0	2.0	30.0	400.0	0.0
1.0	3.0	412.0	1500.0	30.0	4.0	7.0	1500.0	0.0
1.0	3.0	111.0	3000.0	30.0	1.0	30.0	3000.0	0.0

1.0	3.0	211.0	1500.0	30.0	2.0	30.0	1500.0	0.0
1.0	3.0	412.0	1000.0	30.0	4.0	30.0	1000.0	0.0
1.0	3.0	411.0	3000.0	30.0	4.0	30.0	3000.0	0.0
1.0	1.0	211.0	4000.0	30.0	2.0	30.0	4000.0	0.0
1.0	3.0	412.0	500.0	30.0	4.0	30.0	500.0	0.0
1.0	3.0	412.0	500.0	30.0	4.0	30.0	500.0	0.0
1.0	3.0	211.0	6000.0	30.0	2.0	30.0	6000.0	0.0
1.0	3.0	211.0	1000.0	30.0	2.0	30.0	900.0	0.0
1.0	3.0	211.0	1000.0	30.0	2.0	30.0	900.0	0.0
1.0	3.0	430.0	11500.0	30.0	4.0	30.0	11500.0	0.0
1.0	1.0	413.0	10000.0	30.0	2.0	30.0	10000.0	0.0
1.0	3.0	413.0	8000.0	30.0	4.0	30.0	8000.0	0.0
1.0	3.0	211.0	1500.0	30.0	2.0	30.0	1000.0	0.0
1.0	3.0	212.0	10000.0	30.0	2.0	30.0	10000.0	0.0
…	…	…	…	…	…	…	…	…

Tabla 18: Tabla Historia Plazo BD

SUCURSALMOV	PRODUCTO	PERIODO	MONTOORIG	MORAPAGADA	TIPOPRESTAMO
1.0	61.0	30.0	600.0	0.0	2.0
1.0	61.0	30.0	300.0	0.0	2.0
1.0	61.0	30.0	800.0	0.0	2.0
1.0	61.0	30.0	400.0	0.0	2.0
1.0	61.0	30.0	1000.0	0.0	2.0

1.0	61.0	30.0	400.0	0.0	2.0
1.0	61.0	30.0	1000.0	0.0	2.0
1.0	61.0	30.0	1800.0	0.0	2.0
1.0	61.0	30.0	1400.0	0.0	2.0
1.0	61.0	30.0	600.0	0.0	2.0
1.0	61.0	30.0	300.0	0.0	2.0
1.0	61.0	30.0	6000.0	0.0	2.0
1.0	61.0	30.0	25000.0	0.0	2.0
1.0	61.0	30.0	12500.0	0.0	2.0
1.0	61.0	30.0	4000.0	0.0	2.0
1.0	61.0	60.0	30000.0	0.0	2.0
1.0	61.0	180.0	3000.0	0.0	2.0
1.0	61.0	180.0	3000.0	0.0	2.0
1.0	61.0	180.0	3000.0	0.0	2.0
1.0	61.0	180.0	4010.0	0.0	2.0
1.0	61.0	180.0	3000.0	0.0	2.0
1.0	61.0	30.0	60000.0	0.0	2.0
1.0	61.0	90.0	10000.0	0.0	2.0
1.0	61.0	30.0	5000.0	0.0	2.0
…	…	…	…	…	…

ANEXO II:

Fase Pre procesamiento de Datos

Depuración del conjunto de datos de cada una de las tablas consideradas en la investigación respecto a valores atípicos, faltantes y erróneos.

Tabla 19: Atributos de la Tabla Personas

Atributo	Tipo de Dato	Descripción
Tipo Sexo	Integer	Tipo sexo de la persona.
Tipo Estado Civil	Integer	Estado civil de la persona.
Tipo Actividad	Integer	Actividad de la persona
Patrimonio	Integer	Patrimonio que posee la persona.
Numero Personas a Cargo	Integer	Número de personas a cargo.
Antiguedad	Integer	Antigüedad de la persona en el trabajo que realiza.
Tipo Nivel Educativo	Integer	Nivel educativo de la persona.
Ingreso Sueldos	Integer	Ingresos de la persona.
Tipo Vivienda	Integer	Tipo de vivienda de la persona.
Profesion u Oficio	Integer	Profesión u oficio de la persona.
Local Terreno Negocio	Integer	Local, terreno o negocio de la persona.
Fecha Ingreso	Date	Fecha de ingreso a la Cooperativa de la persona.
Tipo Negocio	Integer	Tipo de negocio de la persona.

Tabla 20: Atributos de la Tabla Clientes

Atributo	Tipo de dato	Descripción
Fecha Apertura	Date	Fecha en la que se abre la cuenta.
Fecha Cierre	Date	Fecha en la que se cierra la cuenta.
Fecha UltMovi	Date	Fecha del último movimiento de la cuenta.
Cod Estado	Integer	Estado en que se encuentra la cuenta.
Tipo Persona	Integer	Tipo de persona a la que pertenece la cuenta.
Motivo Baja	Integer	Motivo de baja de la cuenta.
Socio	Integer	Si es o no socio de la entidad.
Fecha Re-Apertura	Integer	Fecha de re-apertura de la cuenta.
Tipo Cliente SIB	Integer	Tipo de cliente de acuerdo con la Superintendencia de Bancos.
Patrimonio	Integer	Patrimonio que posee el socio.

Sucursal Cliente	Integer	Sucursal a la que pertenece el cliente.
Motivo Apertura	Integer	Motivo de apertura de la cuenta.

Tabla 21: Atributos de la Tabla Saldos

Atributo	Tipo de Dato	Descripción
SUCURSAL	Integer	Sucursal a la pertenece el socio.
PRODUCTO	Integer	Tipo de producto.
CUENTA	Integer	Número de cuenta del socio.
Monto Original	Integer	Monto original de la cuenta.
Cuota (Prest Amort o DPF)	Integer	Tipo de cuota.
Fecha Apertura	Date	Fecha en la que se abre la cuenta.
Nro Retiros Mes	Integer	Numero de retiros al mes.
Tipo Interes	Integer	Tipo de interés.
Tipo Pago Interes	Integer	Tipo de Pago para interés.
Estado (Cerrado Judicial Bloqueo)	Integer	Estado de la cuenta.
Cod Calificacion	Integer	Código de calificación del cliente.
INACTIVA	Integer	Estado de la cuenta.
Monto Aprobado Prestamo	Integer	Monto aprobado del préstamo.

Tabla 22: Atributos de las Tablas Historial 2017/2018/2019

Atributo	Tipo de Dato	Descripción
NUMERO_MOVIMIENTO	Integer	Número de movimientos realizados.
FECHA_PROCESADO	Date	Fecha procesada del movimiento.
SUCURSAL	Integer	Sucursal en donde se hizo el movimiento.
DEBITO_CREDITO	Integer	Débito o crédito.
MONTO	Integer	Monto desembolsado.

Tabla 23: Atributos de la Tabla Solicitud Crédito

Atributo	Tipo de Dato	Descripción
Ayu Cod Sucursal	Integer	Código de la sucursal.
Tipo Prestamo	Integer	Tipo préstamo.
Ayu Producto	Integer	Tipo de producto.
Monto Solicitado	Integer	Monto solicitado para crédito.
Tipo Credito	Integer	Tipo crédito.

Plazo Aprobado	Integer	Monto aprobado para crédito.
Monto Aprobado	Integer	Monto aprobado para crédito.
Plazo Solicitado	Integer	Monto solicitado para crédito.
Tipo Garantia	Integer	Tipo de garantía para crédito.

Tabla 24: Atributos de la Tabla Historia Plazo

Atributo	Tipo de Dato	Descripción
SUCURSALMOV	Integer	Sucursal en la que se hizo el movimiento.
PRODUCTO	Integer	Tipo de producto.
PERIODO	Integer	Periodo en días.
MONTOORIG	Integer	Monto original.
MORAPAGADA	Integer	Mora pagada.
TIPOPRESTAMO	Integer	Tipo préstamo.

ANEXO III:

Fase Selección de Características

Normalización de atributos de las tablas consideradas en la investigación.

Tabla 25: Atributos Normalizados Tabla Personas

Atributo	Tipo de Contenido	Valores	Descripción
Tipo Sexo	Discreto	0	Sin Información
		1	Femenino
		2	Masculino
Tipo Estado Civil	Discreto	0	Sin Información
		1	Unión Libre
		2	Casado Separa Bienes
		3	Casado (a)
		4	Divorciado (a)
		5	Soltero (a)
		6	Viudo (a)
Tipo Actividad	Discreto	1	Dependiente
		2	Independiente
		3	Otros
Tipo Nivel Educativo	Discreto	0	Sin Información
		1	Postgrados (Masterados)
		2	Medio (Bachiller)
		3	Ninguno
		4	Básico (Primario)
		5	Superior (Universitario)
		6	Superior (Técnico)
		7	Desconocido
Tipo Vivienda	Discreto	0	Sin Información
		1	Arrienda
		2	En promesa de venta
		3	Familiar
		4	Propia Hipotecada
		5	Prestada
		6	Propia no Hipotecada
		7	Anticresis
Profesion u Oficio	Discreto	42	Comerciante
		3	Agricultor
		59	Empleado Privado
		48	Chofer-Transportista
		60	Empleado Público
		5	Albañil-Pintor
		155	Profesor
		67	Ganadero

		6	Ama de Casa-Llaves
		95	Jubilado
Local Terreno Negocio	Discreto	0	Sin Información
		1	Alquilado
		2	En Promesa de Venta
		3	Familiar
		4	Ninguno
		5	Financiada
		6	Otro
		7	Propio
		8	Anticresis
Tipo Negocio	Discreto	0	Sin Información
		1	Ambulante
		2	Fijo

Tabla 26: Atributos Normalizados Tabla Clientes

Atributo	Tipo de Contenido	Valores	Descripción
Cod Estado	Discreto	1	Inactivo
		2	Solicito Retiro
		3	Retirado
		4	Suspenso
		5	Vigente
Tipo Persona	Discreto	1	Natural
		2	Jurídico
Motivo Baja	Discreto	0	Sin Información
		1	Cambio Domicilio
		2	Devolución de Certificados
		3	Mala Conducta Socio
		4	Fallecimiento
		5	Mala Atención
		6	Negación de Crédito
		7	No usa la Cuenta
		8	Calamidad Domestica
		9	Motivos Personales
Socio	Discreto	1	Socio
		2	Cliente
Tipo Cliente SIB	Discreto	1	Persona Natural
		2	Pública Financiera
		3	Privada Financiera
		4	Publica no Financiera
		5	Privada no Financiera

		6	Fondos de jubilación priv.
		7	Fondos de Inversión
Sucursal Cliente	Discreto	1	Casa Central
Motivo Apertura	Discreto	1	Ahorros
		2	Créditos
		3	Ahorros y Créditos
		4	DPF
		5	Otros

Tabla 27: Normalización Calificación Interna Clientes

Atributo	Tipo Contenido	Valores	Descripción	Días Morosidad
Cod Calificacion	Continuo	0	Sin historial crediticio	
		1	A (Riesgo Normal)	0-30
		2	A1	0
		3	A2	1-8
		4	A3	9-15
		5	B (Riesgo Potencial)	31-60
		6	B1	16-30
		7	B2	31-45
		8	C (Riesgo Deficiente)	61-120
		9	C1	46-70
		10	C2	71-90
		11	D (Dudoso Recaudo)	91-120
		12	E (Perdida)	Mayor a 120

Tabla 28: Atributos Normalizados Tabla Saldos

Atributo	Tipo Contenido	Valores	Descripción
SUCURSAL	Discreto	1	Casa Central
PRODUCTO	Discreto	11	AH. Normal
		14	Certificados
		62	DPF. INST. Vencimiento
		16	Certificados
		15	Capitalización
		25	Certificados Comunes
		21	AH. Promedio
		64	AH. Préstamo
		803	DPF. INST. Periódicos
			Inver_Vspv
Tipo Interes	Discreto	0	No calcula

		1	Descuento Comercial
		2	Descuento Efectivo Anual
		3	Efectivo anual
		4	Nominal o Lineal
		5	Descuento Racional
Tipo Pago Interes	Discreto	0	Intereses al Vto.
		1	Cond Esp-Plan de Pago
		2	Amortizable Francés
		3	Amortizable Hamburgués
		4	Pgo.Per.Int.Depósitos
		5	Pgo.Per.Int.Préstamos
Estado (Cerrado Judicial Bloqueo)	Discreto	0	Sin Bloqueo (Normales)
		1	Cuenta Cerrada
		2	Judicial
		3	Retiros
		4	Depósitos y Retiros
INACTIVA	Discreto	1	Activo
		2	Inactivo

Tabla 29: Atributos Normalizados Tabla Historial 2017/2018/2019

Atributo	Tipo Contenido	Valores	Descripción
SUCURSAL	Discreto	1	Casa Central
DEBITO_CREDITO	Discreto	1	Débito
		2	Crédito

Tabla 30: Atributos Normalizados Tabla Solicitud Crédito

Atributo	Tipo de Contenido	Valores	Descripción
Ayu Cod Sucursal	Discreto	1	Casa Central
Tipo Prestamo	Discreto	1	Ampliación
		2	Reestructurado
		3	Nuevo
Ayu Producto	Discreto	411	Micro Quiro
		211	Prioritario Quiro
		224	Prioritario Rol
		413	Micro Hipo
		220	Prioritario Autosuficiente
		423	Micro Sin Garantía

		420	Micro Autosuficiente
		311	Inmobiliario
		412	Micro Quiro Corto Plazo
		212	Prioritario Hipo
Tipo Credito	Discreto	1	Comercial
		2	Consumo
		3	Vivienda
		4	Microcrédito
		5	Línea de Crédito
		6	Carta Garantía
Tipo Garantia	Discreto	0	Sin Garantía
		1	Firma
		2	Prendaria
		3	Hipotecaria
		4	Títulos Valores

Tabla 31: Atributos Normalizados Tabla Historia Plazo

Atributo	Tipo Contenido	Valores	Descripción
SUCURSALMOV	Discreto	1	Casa Central
PRODUCTO	Discreto	411	Micro Quiro
		211	Prioritario Quiro
		62	DPF. INTS. Vencimiento
		61	DPF. INTS. Periódicos
		423	Micro sin Garantía
		224	Prioritario Rol
		413	Micro Hipo
		311	Inmobiliario
		212	Prioritario Hipo
		223	Prioritario sin Garante
TIPOPRESTAMO	Discreto	1	Ampliación
		2	Reestructurado
		3	Nuevo