

# Predicción de clientes potenciales utilizando K vecino más cercano en el área de negocios de la Cooperativa Riobamba

## Prediction of potential customers using K nearest neighbor in the Riobamba Cooperative business area

Miryan Estela Narváez Vilema\*, Raquel Johanna Moyano Arias\*, Diego Bernardo Palacios Campana‡, Geovanny Augusto Izurieta Guamán\*

\*Universidad Nacional de Chimborazo, 060108, Riobamba, Ecuador

‡Escuela Superior Politécnica de Chimborazo (ESPOCH), 060155, Riobamba, Ecuador

Email: \*miryan.narvaez@unach.edu.ec, \*jmoyano.fis@unach.edu.ec, ‡dpalacios@epoch.edu.ec, \*gizurieta.fis@unach.edu.ec

**Resumen—** K vecino más cercano (KNN) es uno de los algoritmos que permite el diagnóstico en tiempo real y apoya la toma de decisiones. Para esta investigación, se consideró la base de datos del área de negocio de la Cooperativa de Ahorro y Crédito Riobamba Ltda., un banco de datos que almacena una gran cantidad de información de los clientes. Estos datos se utilizaron para seleccionar información relevante manteniendo y respetando la confidencialidad de los clientes. El objetivo principal del proyecto es predecir clientes potenciales aplicando el algoritmo KNN. Los resultados demuestran que el algoritmo KNN es adecuado para predecir clientes potenciales de acuerdo a sus antecedentes demográficos, económicos y factores internos de la Cooperativa Riobamba Ltda., Resultando esto, un recurso útil para la institución en la toma de decisiones sobre futuras ofertas de crédito. Se destaca la importancia de aprovechar la información que se maneja en cada institución y más si es dentro del sector financiero porque tanto los clientes como la institución se benefician. Lo primero porque tendrían más opciones de crédito y las instituciones financieras porque podrían incrementar su cartera de clientes y mejorar su servicio.

**Palabras Clave—** Clientes Potenciales, Créditos, KDD, K-NN, Minería de Datos.

**Abstract—** K Nearest Neighbor (KNN) is one of the algorithms that enables diagnosis in real time, and supports decisions making. For this research, the database of the business area of the Cooperativa de Ahorro y Crédito Riobamba Ltda. was considered, a databank that stores a large amount of information from customers. This data was used to select relevant information maintaining and respecting the clients' confidentiality. The main objective of the project is to predict potential customers by applying the KNN algorithm. The results demonstrate that k nearest-neighbor is suitable for predicting potential clients were predicted according to their demographic and economic background and internal factors of Cooperativa Riobamba Ltda., resulting this, a useful resource for the institution in making decisions regarding future credit offers. It is highlighted the importance of taking advantage of the information that is managed in each institution and even more if it is within the financial sector because both, clients and the institution benefit. The former since they would have more credit options and the financial institutions because they might increase their portfolio of clients and improve their service.

**Keywords—** Credits, Data Mining, KDD, K-NN, Potential Clients.

### I. INTRODUCCIÓN

Actualmente, las nuevas tecnologías ayudan a la creación, distribución y manipulación de información, facilitando las actividades que se realiza día a día. Todas las instituciones como financieras, comerciales e industriales producen grandes volúmenes de datos, datos que contienen información valiosa que pueden ser muy útil para las mismas, a la hora de tomar decisiones y de resolver problemas.

La minería de datos (Data Mining) permite descubrir conocimiento oculto en grandes cantidades de datos, información valiosa que, al analizarlos, ofrece ventajas competitivas a las organizaciones. A medida que aumentan los datos, casi de forma exponencial son cada vez más difíciles de interpretar, pasando a ser más compleja en naturaleza y forma. La minería de datos busca patrones, comportamientos, secuencias, tendencias, generando modelos que permitan comprender mejor la información, a través de una combinación de tareas como: Extracción de datos, limpieza de datos, selección de características, análisis de resultados. [1]

La minería de datos es una etapa del proceso KDD (Knowledge Discovery in Databases o en español Descubrimiento de Conocimiento en Base de Datos). KDD 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. Este proceso incluye pre-procesamiento, minería de datos y post-procesamiento de datos. KDD se puede aplicar en diferentes campos, por ejemplo, para determinar perfiles de clientes fraudulentos, para descubrir relaciones implícitas existentes entre síntomas y enfermedades, para determinar perfiles de estudiantes académicamente exitosos y para determinar patrones de compra de clientes. [12]

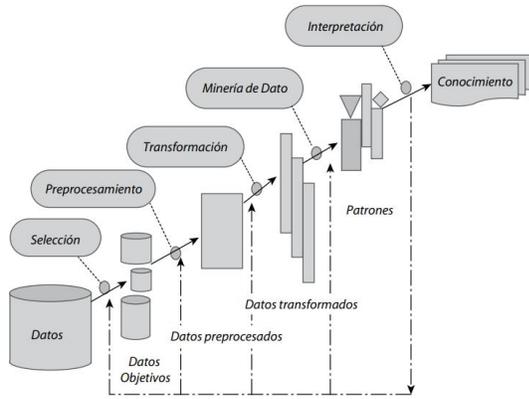


Figura 1. Etapas del proceso KDD. [12]

KDD está compuesto por diferentes etapas que se realizan de forma secuencial e iterativamente, el proceso se muestra en la Fig. 1. Se resume en cinco etapas:

- Selección: En esta etapa se determinan los datos relevantes y el tipo de información a utilizar.
- Preprocesamiento: Esta etapa consiste en la preparación y limpieza de los datos.
- Transformación: Consiste en el tratamiento preliminar de los datos, transformación y generación de nuevas variables a partir de las ya existentes.
- Minería de datos: El objetivo de la etapa es descubrir patrones desconocidos, válidos, potencialmente útiles que están ocultos en los datos.
- Interpretación: Se identifican los patrones obtenidos, la remoción de los patrones redundantes o irrelevantes y la traducción de los patrones útiles en términos que sean entendibles para el usuario.

Es importante diferenciar el proceso KDD de la minería de datos. La minería de datos se define como la ciencia que estudia patrones en grandes bases de datos, con base en esto, se han aplicado técnicas de minería de datos para detección de fraudes, calificaciones crediticias, comportamiento de clientes, etc. [6]

Las técnicas de la minería de datos permiten diseñar estrategias de manejo para explorar grandes bases de datos, de manera automática, con el fin de encontrar patrones repetitivos, tendencias que expliquen el comportamiento de los datos y convertirlos en información productiva [4]. Las técnicas de minería de datos se clasifican en dos categorías: supervisadas y no supervisadas. [5]

Los algoritmos supervisados o predictivos, obtienen a partir de los datos un modelo que relaciona el valor de un atributo llamado etiqueta y los valores de otros atributos (descriptivos), se desarrollan en dos fases: 1) Entrenamiento (construcción de un modelo usando un subconjunto de datos con etiqueta conocida) y 2) Prueba (prueba del modelo sobre el resto de los datos). La misión de los algoritmos de aprendizaje supervisados es encontrar el conjunto de atributos que permite predecir con mayor precisión la clase de cada objeto del conjunto de entrenamiento. [10]

Las técnicas de minería de datos no supervisadas, también

Tabla I  
TÉCNICAS DE MINERÍA DE DATOS

Supervisados	No supervisados
Árboles de decisión	Detección de desviaciones
Redes neuronales	Segmentación
Regresión	Clustering
Serie temporales	Asociación
	Patrones secuenciales

conocidas con el nombre de técnicas de descubrimiento del conocimiento, se utilizan para la detección de patrones ocultos en bases de datos de gran tamaño. Dichos patrones representan por sí mismos información útil que puede ser utilizada directamente en la toma de decisiones [9]. En la Tabla I se muestran las principales técnicas de minería de datos de las dos categorías. [11]

Dentro de las técnicas de minería de datos, se proporcionan modelos que aportan información útil [2], desde un sencillo análisis gráfico, pasando por métodos estadísticos, complementados con métodos y algoritmos de aprendizaje automático que resuelven problemas de clasificación, predicción de valores, detección de patrones, asociación de atributos [3], entre los cuales se encuentra el algoritmo de clasificación K Nearest Neighbors. KNN es una de las técnicas ampliamente utilizada para la clasificación de patrones, al cual va orientada la presente investigación.

El algoritmo KNN o en español K vecinos más cercanos, es un clasificador no paramétrico, está considerado como uno de los de mejor desempeño en el ámbito del aprendizaje automático [7]. KNN construye un modelo sencillo para resolver problemas de predicción, es ampliamente utilizado para tareas de clasificación supervisada. [8]

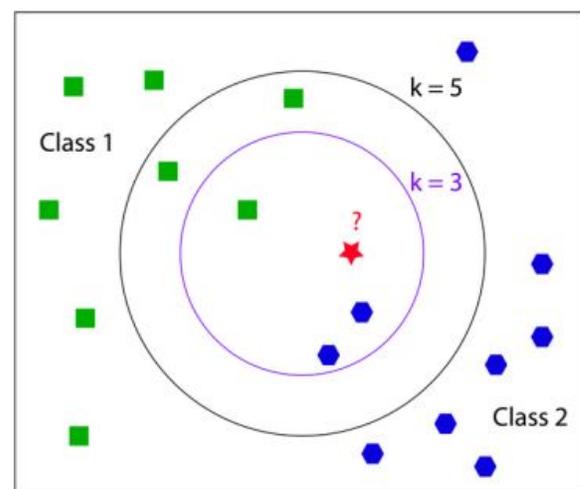


Figura 2. Ejemplo clasificación KNN.

La Fig. 2 muestra un ejemplo de clasificación KNN, que recibe dos características clase 1 y clase 2 como entrada y devuelve una de las siguientes clases como salida: verde, azul. Dado un conjunto entrenado  $T$  y una nueva entrada  $x$

= (clase 1, clase 2) (representada por una estrella roja en la figura), el clasificador KNN busca los  $k$  vecinos más cercanos de  $x$  en  $T$  y entre ellos selecciona la clase más frecuente. La clase más frecuente está definida por una función de similitud representada por la distancia. El valor de  $k$  se puede configurar según las necesidades del usuario. Como resultado del ejemplo, el clasificador KNN clasifica  $x$  como verde. [13]

La idea básica en la que se fundamenta el algoritmo es que, kNN decide que  $x$  pertenece a la categoría de voto mayoritario de los  $k$  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. [14], [15]

La Cooperativa de Ahorro y Crédito (COAC) Riobamba Ltda., es una entidad dedicada a las finanzas populares y solidarias, creada para impulsar el desarrollo socio-económico de los socios a través de cuentas de ahorro y depósitos a plazo fijo; así como al otorgamiento de créditos para diversas actividades, fomentando de esta manera el progreso y desarrollo económico de la ciudad.

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 de la COAC Riobamba.

## II. METODOLOGÍA

Para llevar a cabo la investigación y cumplir con los objetivos, se aplicó la metodología de Descubrimiento de Conocimiento en Bases de Datos (KDD), dentro de la cual se cumplieron cinco fases:

### A. Fase Selección de Datos

En esta fase se determinó las tablas a utilizar y con la ayuda de la herramienta Microsoft SQL Server se establecieron las siguientes.

- Persona
- Clientes
- Saldos
- Historial 2017-2019
- Solicitud Crédito
- Historia Plazo

### B. Fase Preprocesamiento de Datos

La fase de preprocesamiento permitió reconocer y limpiar los atributos y registros de las tablas seleccionadas, eliminando ruidos e inconsistencias con la ayuda del software de limpieza EmEditor.

EmEditor permite filtrar información de las tablas, de manera que los registros reflejados sean únicamente de la matriz de la Cooperativa, además, se dividieron las tablas para el análisis debido a la cantidad de información.

### C. Fase de Transformación

Con la ayuda del software RapidMiner se realizó la normalización de los atributos de cada una de las tablas seleccionadas, y se transformó el 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, se muestra en la Fig. 3.

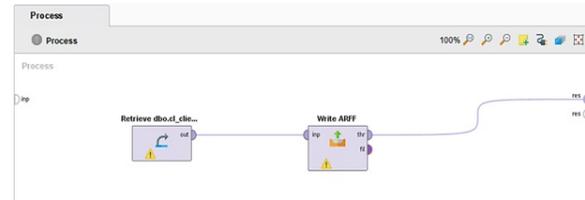


Figura 3. Transformación de formato.

### D. Fase Minería de Datos

En esta fase se llevó a cabo las simulaciones de todas las tablas consideradas en la investigación, para ello se utilizó el software de Minería de Datos Weka. Los pasos a seguir se presenta en el diagrama, Fig. 4.

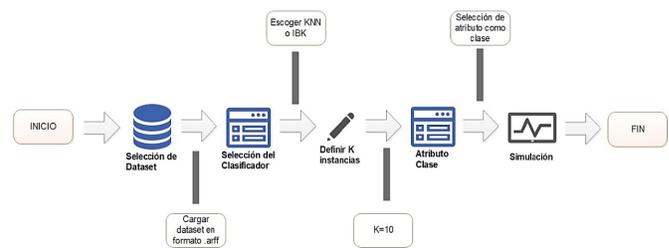


Figura 4. Proceso Minería de Datos.

Se define  $k=10$  para las simulaciones de todas las tablas, debido a la cantidad de registros de cada una.

### E. Fase Interpretación

Con el software IBM SPSS Statistics 25, especialista en análisis estadísticos de datos, se graficó los resultados para una mejor apreciación.

### F. Herramientas utilizadas

- Microsoft SQL Server, es un sistema de gestión de bases de datos relacionales, su principal lenguaje de consulta es Transact-SQL. [17]
- Software RapidMiner, plataforma de código abierto para la preparación de datos, aprendizaje automático y la implementación de modelos. [18]
- Software EmEditor, es un editor de texto rápido, ligero, extensible y fácil de usar para Windows, ayuda a editar o resaltar errores automáticamente. [19]
- Herramienta de Minería de Datos Weka, es una plataforma de software destinada al aprendizaje automático y la minería de datos. [20]
- IBM SPSS Statistics 25, analiza bases de datos de gran magnitud, permite consultar datos y formular hipótesis, ejecuta procedimientos para aclarar las relaciones entre variables, identificar tendencias y realizar predicciones. [16]

### III. RESULTADOS

En este apartado se presenta el proceso de extracción de conocimiento el cual está enmarcado en la fase minería de datos.

Para el proceso de extracción y selección de los datos se utilizó la herramienta integrada al gestor de bases de datos Microsoft SQL Server. La Tabla II, muestra las tablas utilizadas con su respectivo número de registros:

- Las tablas Personas y Clientes contienen información demográfica de los clientes.
- Las tablas Saldos e Historial reflejan los tipos de cuentas que se manejan y el comportamiento de ahorros de los clientes.
- Las tablas Solicitud Crédito e Historia Plazo evidencian en los registros, información crediticia y comportamiento de pago de los clientes, por lo que se consideran vitales dentro de la investigación.

Tabla II  
BASE DE DATOS COAC RIOBAMBA

Tablas	Nº Registros
Personas	225 672
Clientes	73 529
Saldos	779 884
Historial 2017	5 678 281
Historial 2018	5 396 158
Historial 2019	3 916 502
Solicitud Crédito	62 317
Historia Plazo	3 175 308

#### A. Algoritmo KNN

KNN es el algoritmo más simple en aprendizaje automático, este método se puede utilizar para extraer patrones desconocidos, útiles y comprensibles que están ocultos en los datos.

Se empleó el método de predicción del vecino más cercano basado en la distancia euclidiana. La distancia euclidiana está entre dos puntos P (p1, p2... pn) y Q (q1, q2... qn) en el espacio euclidiano, la ecuación es, Ec. 1.

$$d(P, Q) = \sqrt{\sum_{i=0}^n (P_i - Q_i)^2} \quad (1)$$

El esquema general de búsqueda se puede resumir como:

- Paso 1: Cada una de las nuevas instancias se verifica con los casos ya disponibles, en función de la asignación de distancia y se clasifica mediante el valor *k*.
- Paso 2: La distancia será menor, si las instancias son más similares y viceversa.
- Paso 3: Según la distancia, el valor *k* y la instancia, se asignan a una clase específica.
- Paso 4: La predicción se basa en el valor *k* (*k* representa el número de vecinos más cercanos).

- Paso 5: Determinar el valor de *k* para el conjunto de datos del área de negocios de la cooperativa para la precisión de la clasificación. [14]

Para obtener los resultados de las simulaciones realizadas en el software WEKA se utilizaron dos equipos portátiles con las siguientes características, se muestran en la Tabla III.

Tabla III  
CARACTERÍSTICAS DE EQUIPOS

Equipos	Características	Descripción
Equipo 1	SO	Windows 10 Home de 64 Bits
	CPU	Intel <sup>®</sup> Core™ i7 Séptima Generación
	RAM	16 Gb
Equipo 2	SO	Windows 10 Pro de 64 Bits
	CPU	Intel <sup>®</sup> Core™ i5 Quinta Generación
	RAM	8 Gb

#### B. Resultados Información de clientes

Tabla IV  
RESULTADOS PERSONAS

Personas	Total	%
Instancias Clasificadas Instancias	225 670	100.00%
Clasificadas Correctas Instancias	204 962	90.82%
Clasificadas Incorrectas	20 708	9.17%

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

Tabla V  
RESULTADOS CLIENTES

Clientes	Total	%
Instancias Clasificadas Instancias	73 527	100.00%
Clasificadas Correctas	44 525	60.55%
Instancias Clasificadas Incorrectas	29 002	39.44%

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

#### C. Resultados Información Ahorros

Tabla VI RESULTADOS SALDOS

Clientes	Total	%
Instancias Clasificadas Instancias	779 882	100.00%
Clasificadas Correctas Instancias	771 166	98.88%
Clasificadas Incorrectas	8 716	1.11%

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

Tabla VII  
RESULTADOS SOLICITUD CRÉDITO

Cientes	Total	%
Instancias Clasificadas	62 315	100.00%
Clasificadas Correctas	45 566	73.12%
Instancias Clasificadas Incorrectas	16 749	26.87%

Tabla VIII  
RESULTADOS HISTORIA PLAZO

Cientes	Total	%
Instancias Clasificadas	3 175 306	100.00%
Clasificadas Correctas	3 152 660	99.28%
Clasificadas Incorrectas	22 646	0.72%

Tabla IX  
TIEMPO DE RESPUESTA DE CADA EQUIPO

Tablas	Nº Registros	Equipo 1	Equipo 2	Diferencia
Personas	225 672	2 h 11 min	2 h 45 min	34 min
Cientes	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

D. Resultados Información Créditos

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

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

La Tabla IX, muestra la diferencia de los tiempos de respuesta de cada equipo, la diferencia del equipo 2 es de 1 día 19 horas si las tablas superan los 1 000 000 registros.

E. Interpretación de Resultados

Para el análisis de los datos se utilizó el software estadístico IBM SPSS Statistics 25, para crear gráficas y presentar los resultados obtenidos. Por la gran cantidad de información que almacena la cooperativa, se consideró analizar los datos que corresponden a los clientes de la matriz, equivalente a 73 529, se tomó como línea base tres periodos 2017, 2018 y 2019.

La Fig. 5, 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.

La Fig. 6, muestra las distintas categorías de créditos, se observa que el 30% acceden a Micro Quiro que son créditos para microempresarios con garantía quirografaria, el 26% acceden a Prioritario Quiro que son créditos para empleados con garantía quirografaria.

La Fig. 7, muestra los tipos de crédito, el 48% de clientes acceden al Microcrédito, que va dirigido a pequeños y medianos negocios sean estos formales e informales, el 47% acceden al de Consumo, crédito destinado para todas aquellas

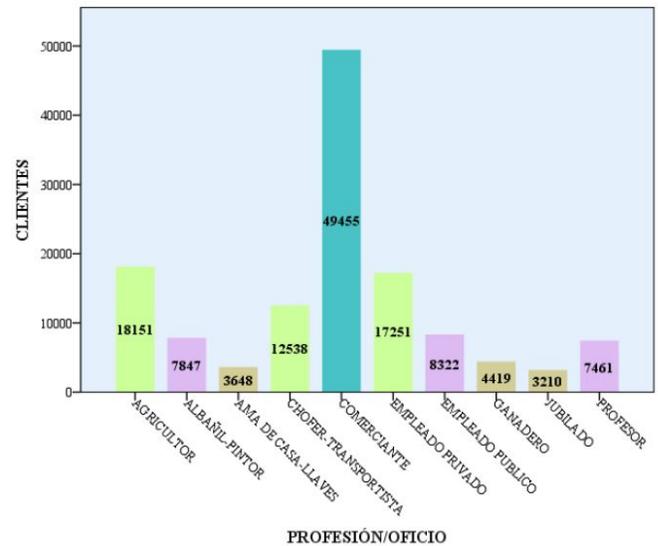


Figura 5. Profesion.

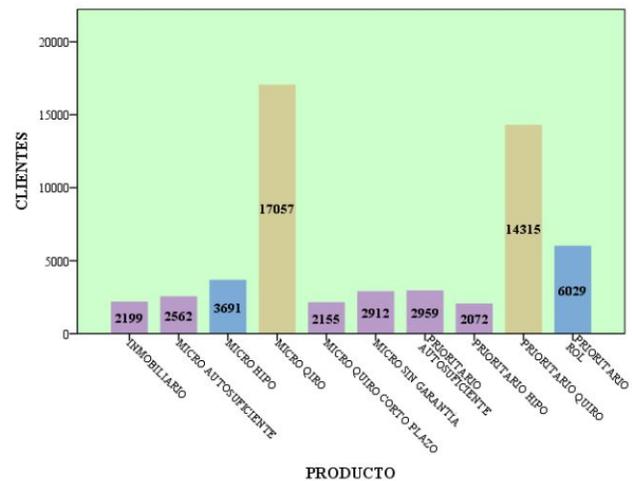


Figura 6. Solicitud de Crédito.

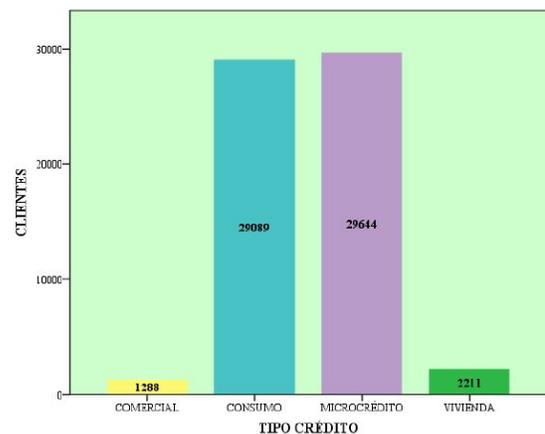


Figura 7. Tipo de Crédito.

personas naturales que trabajan en relación de dependencia o

perciben un sueldo.

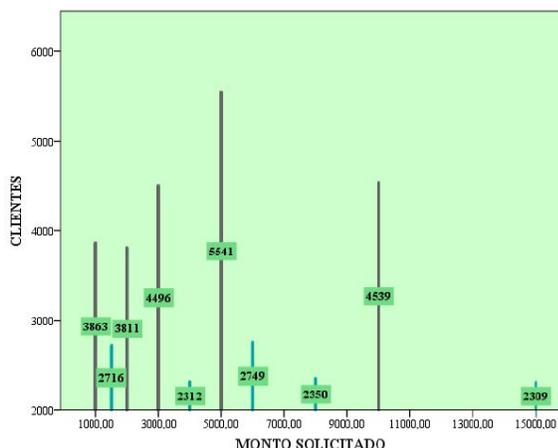


Figura 8. Monto Solicitado de Crédito.

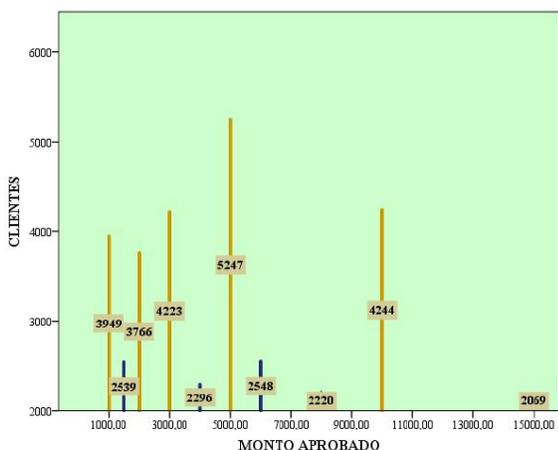


Figura 9. Monto Aprobado de Crédito.

Las Fig. 8 y 9, 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 presenta en la imagen al mismo porcentaje se aprueban los créditos.

#### IV. 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.

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.

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.

#### REFERENCIAS

- [1] Mauricio Miranda, and Jheser Guzmán. *Análisis de la deserción de estudiantes universitarios usando técnicas de minería de datos*. Formación universitaria, 10(3), p. 61-68, 2017.
- [2] Cesar Higinio Menacho Chiok. *Predicción del rendimiento académico aplicando técnicas de minería de datos*. Anales Científicos. Universidad Nacional Agraria La Molina, 78(1), p. 26-33, 2017.
- [3] Sergio Valero, Alejandro Salvador, and Marcela García. *Minería de datos: predicción de la deserción escolar mediante el algoritmo de árboles de decisión y el algoritmo de los k vecinos más cercanos*. Ene 779(73), p. 33, 2005.
- [4] Jahir Gutiérrez, and Bernardo Molina. *Identificación de técnicas de minería de datos para apoyar la toma de decisiones en la solución de problemas empresariales*. Revista Ontare 3(2):33-51, 2015.
- [5] José Molina, and Jesús García. *Técnicas de Minería de Datos basadas en Aprendizaje Automático*. Técnicas de Análisis de Datos, 96-266, 2008.
- [6] Zhenni Feng, and Yanmin Zhu. *A survey on trajectory data mining: Techniques and applications*. IEEE Access 4, 2056-2067, 2016.
- [7] Angelica Duca, Clara Bacciu, and Andrea Marchetti. *A K-nearest neighbor classifier for ship route prediction*. OCEANS 2017-Aberdeen. IEEE, p. 1-6, 2017.
- [8] Szilárd Vajda, and K. C. Santosh. *A fast k-nearest neighbor classifier using unsupervised clustering*. International conference on recent trends in image processing and pattern recognition. Springer, Singapore, p. 185-193, 2016.
- [9] María Moreno, and Vivian López. *Uso de Técnicas No Supervisadas en la Construcción de Modelos de Clasificación de Ingeniería de Software*. Departamento de Informática y Automática. Universidad de Salamanca, 2015.
- [10] Y. Belinchón. *Minería de datos*. Obtenido de. <https://es.scribd.com/document/308398381/15mem-pdf>, 2015.
- [11] María N. Moreno García, Luis A. Miguel Quintales, Francisco J. García Peñalvo, and M. José Polo Martín. *Aplicación de Técnicas de Minería de Datos en la Construcción y Validación de Modelos Predictivos y Asociativos a Partir de Especificaciones de Requisitos De Software*. ADIS, 2001.
- [12] S. R. Timarán-Pereira, I. Hernández-Arteaga, S. J. Caicedo-Zambrano, A. Hidalgo-Troya, and J. C. Alvarado-Pérez. *El proceso de descubrimiento de conocimiento en bases de datos*. En Descubrimiento de patrones de desempeño académico con árboles de decisión en las competencias genéricas de la formación profesional. Bogotá: Ediciones Universidad Cooperativa de Colombia, pp. 63-86, 2016.
- [13] Thomas Cover, and Peter Hart. *Nearest neighbor pattern classification*. IEEE transactions on information theory, 13(1), p. 21-27, 1967.
- [14] Veena Vijayan, and Aswathy Ravikummar. *Study of data mining algorithms for prediction and diagnosis of diabetes mellitus*. International journal of computer applications, 95(17), 2014.
- [15] Sadegh Bafandeh Imandoust, and Mohammad Bolandraftar. *Application of k-nearest neighbor (knn) approach for predicting economic events: Theoretical background*. International Journal of Engineering Research and Applications, 3(5), p. 605-610, 2013.
- [16] George Darren, and Paul Mallery. *IBM SPSS statistics 26 step by step: A simple guide and reference*. Routledge, 2019.
- [17] Jose Santamaría, and Javier Hernández. *Microsoft SQL Server. SQL SER vs MY SQL*, p. 1-6, 2016.
- [18] Markus Hofmann, and Ralf Klittenberg. *RapidMiner: Data mining use cases and business analytics applications*. CRC Press, 2016.
- [19] Emeditor.com. *EmurasoftEmEditor*. [Online]. Available: <https://www.emeditor.com/>.
- [20] Ingrid Russell, and Zdravko Markov. *An introduction to the Weka data mining system*. Proceedings of the 2017 ACM SIGCSE Technical Symposium on Computer Science Education, 2017.