



Aplicación de técnicas de minería de datos en un modelo predictivo orientado a analizar el efecto de la pandemia de COVID-19 sobre las rentas de los miembros de cooperativas.

Application of data mining techniques in a predictive model aimed at analyzing the effect of the COVID-19 pandemic on the income of cooperative members.

Alexander Palacios Zurita ^{1,2,*}, Rosa Giler Zambrano ², Dayron Rumbaut Rangel ¹

¹ Universidad Bolivariana del Ecuador, Guayas, Ecuador

² Escuela Superior Politécnica Agropecuaria de Manabí Manuel Félix López, Manabí, Ecuador

* Autor para correspondencia : Alexander Palacios Zurita, Correo electrónico: apalacios@espm.edu.ec ;

Abstracto: La crisis sanitaria provocada por la COVID-19 ocasionó una fuerte alteración socioeconómica que redujo los ingresos de numerosos hogares, haciendo imprescindible reconocer los factores moduladores en poblaciones vulnerables. El propósito de este estudio fue diseñar un modelo predictivo, basado en minería de datos, con el fin de identificar los determinantes socioeconómicos, demográficos y de bienestar que inciden en el nivel de ingresos de 1456 miembros pertenecientes a tres cooperativas de Manabí, Ecuador, en los periodos previos, durante y posteriores al confinamiento. Para el análisis, se utilizó el Proceso Estándar Interindustrial para Minería de Datos (CRISP-DM). Los datos recolectados mediante encuestas fueron divididos en conjuntos de entrenamiento y validación, aplicándose técnicas de sobremuestreo en las clases con menor representación para equilibrar la distribución. Posteriormente, se entrenaron diversos algoritmos de clasificación supervisada, evaluando su rendimiento comparativo. Los resultados mostraron un desempeño altamente satisfactorio: las redes neuronales lograron precisiones cercanas al 90% y valores de área bajo la curva ROC superiores a 0,95 en el conjunto de prueba. Se constató que el confinamiento impactó de forma negativa, reflejándose en un incremento de socios ubicados en la categoría de ingresos más bajos. Entre los factores con mayor peso en la predicción se destacaron la formación académica, la condición laboral y la percepción de bienestar. Los hallazgos constituyen una herramienta práctica para que las cooperativas identifiquen perfiles de mayor vulnerabilidad y elaboren estrategias de apoyo focalizadas. Asimismo, la validación de los modelos mediante el método TOPSIS Neutrosófico y la métrica de Distancia de Hausdorff reforzó su solidez, al integrar la indeterminación y la incertidumbre en el análisis de los factores socioeconómicos. De esta manera, se incrementa la confiabilidad de los resultados en escenarios de crisis, confirmando la pertinencia de la minería de datos para el estudio de la vulnerabilidad económica.

Palabras clave: Minería De Datos, Modelado Predictivo, Cooperativas De Crédito, Factores Socioeconómicos, Aprendizaje Automático.

Abstract: The health crisis caused by COVID-19 caused a significant socioeconomic disruption that reduced the income of numerous households, making it essential to identify the modulating factors in vulnerable populations. The purpose of this study was to design a predictive model, based on data mining, to identify the socioeconomic, demographic, and well-being determinants that influence the income level of 1,456 members belonging to three cooperatives in Manabí, Ecuador, during the periods before, during, and after the lockdown. The Cross-Industry Standard Process for Data Mining (CRISP-DM) was used for the analysis. The data collected through surveys were divided into training and validation sets, and oversampling techniques were applied to the underrepresented classes to balance the distribution. Subsequently, various supervised classification algorithms were trained, evaluating their comparative performance. The results showed highly satisfactory performance: the neural networks achieved accuracies close to 90% and ROC curve areas above 0.95 in the test set. It was found that the lockdown had a negative impact, reflected in an increase in members in the lowest-income category. Among the factors with the greatest weight in the prediction were academic education, employment status, and perception of well-being. The findings constitute a practical tool for cooperatives to identify more vulnerable profiles and develop targeted support strategies. Furthermore, the validation of the models using the Neutrosophic TOPSIS method and the Hausdorff Distance metric reinforced their robustness by integrating indeterminacy and uncertainty into the analysis of socioeconomic factors. This increases the reliability of the results in crisis scenarios, confirming the relevance of data mining for the study of economic vulnerability.

Keywords: Data Mining, Predictive Modeling, Credit Unions, Socioeconomic Factors, Machine Learning.

1. Introducción

La pandemia de COVID-19 causó una disrupción socioeconómica sin precedentes a nivel mundial, afectando significativamente los ingresos y la estabilidad financiera de millones de hogares [1,2,3]. En el contexto ecuatoriano, los socios de las cooperativas de crédito, muchos de ellos pertenecientes a sectores vulnerables o a la economía popular y solidaria, enfrentaron desafíos económicos particulares, como lo demuestran los estudios sobre el impacto a nivel de hogar [4]. Esta situación subraya la necesidad crucial de identificar los factores que influyeron en la variación de sus niveles de ingresos durante el confinamiento.

La minería de datos, definida como el proceso de organizar, analizar y extraer información valiosa de grandes conjuntos de datos [5,6], se perfila como una metodología clave. Los autores distinguen entre tareas descriptivas, que clasifican los datos basándose en incidentes históricos, y tareas predictivas, que utilizan datos históricos para predecir resultados futuros. Técnicas como la clasificación, la regresión y los árboles de decisión son fundamentales en aplicaciones financieras para la predicción de riesgos o la fidelización de clientes [5]. Otros estudios destacan la alta capacidad de aprendizaje, la inteligencia y la precisión de la tecnología de minería de datos, considerándola esencial para procesar información compleja y transformarla en herramientas aplicables en diversos campos [6].

Sin embargo, a pesar del potencial de las herramientas analíticas avanzadas, el problema central para abordar la situación de los ingresos de los miembros de la COAC durante la crisis de la COVID-19 reside en la necesidad de un enfoque metodológico que permita identificar con precisión cuantitativa los factores específicos que modularon estos niveles de ingresos y que facilite el desarrollo de modelos predictivos robustos para este escenario. El análisis de la dinámica económica en contextos de crisis sanitarias constituye una base esencial para promover la estabilidad social y la recuperación económica [3]. A pesar de los avances del aprendizaje automático en economía y finanzas, y su ventaja sobre los modelos estadísticos tradicionales [7], persiste una brecha en la aplicación de metodologías de minería de datos dirigidas al estudio de factores socioeconómicos, demográficos y de bienestar a nivel individual durante las emergencias sanitarias. Si bien existen aplicaciones de minería de datos en el sector cooperativo ecuatoriano, como el trabajo de Mejía Vanegas



[8] para predecir la morosidad crediticia, y estudios que utilizan aprendizaje automático para analizar el impacto socioeconómico del COVID-19 a nivel regional o macro [7, 1, 9, 3], la mayoría se han centrado en variables agregadas o enfoques no supervisados, dejando un vacío en la construcción de modelos predictivos supervisados para la admisión individual en este contexto específico.

Los enfoques tradicionales a menudo no logran captar patrones complejos ni construir los modelos predictivos robustos necesarios. Esto limita la capacidad de las cooperativas de ahorro y crédito para anticipar situaciones de vulnerabilidad económica entre sus socios y diseñar estrategias de apoyo efectivas. Como señala Winston, el nivel socioeconómico representa un factor determinante en la investigación sobre el impacto de las pandemias [10]. En esta línea, el objetivo general de este trabajo es desarrollar un modelo predictivo basado en técnicas de minería de datos para identificar los factores socioeconómicos, demográficos y de percepción del bienestar asociados con el nivel de ingresos de los socios de tres cooperativas de ahorro y crédito del segmento 1 en la provincia de Manabí, a partir de una muestra de 1456 encuestados, durante el período de confinamiento por la COVID-19, lo que resalta la importancia de este enfoque para la toma de decisiones informada en el sector cooperativo. Además, los resultados del modelo predictivo se validan mediante el método TOPSIS Neutrosófico y la Distancia de Hausdorff, que incorpora la lógica neutrosófica para evaluar la robustez de las predicciones y la relevancia de los factores identificados, considerando la incertidumbre y la indeterminación inherentes a los datos socioeconómicos en contextos de crisis.

Este artículo se organiza de la siguiente manera: comienza con la presentación del contexto de la pandemia y los problemas que justifican la investigación. Posteriormente, se revisa la literatura relevante sobre análisis económico en crisis y la aplicación de la minería de datos. A continuación, se detalla la metodología empleada, incluyendo la descripción de los datos recopilados de las tres cooperativas, el preprocesamiento, los algoritmos de aprendizaje automático utilizados por el software Orange Data Mining y las métricas de evaluación. A continuación, se presentan y discuten los resultados obtenidos del análisis comparativo de los modelos predictivos y la identificación de los factores más influyentes en el nivel de ingresos. Posteriormente, se describe la validación de los resultados utilizando el método TOPSIS Neutrosófico y la Distancia de Hausdorff para confirmar la consistencia de las predicciones y los factores determinantes bajo un marco neutrosófico. Finalmente, se presentan las conclusiones del estudio, junto con las limitaciones y las perspectivas de futuras investigaciones en esta línea.

2. Materiales y métodos

Esta investigación se desarrolló bajo un enfoque cuantitativo, basado en la recopilación y el análisis de datos estructurados para identificar patrones, relaciones y construir modelos predictivos. Siguiendo el proceso de investigación científica [11], una vez planteado el problema y revisada la literatura, se derivaron hipótesis específicas que se probaron mediante un diseño no experimental, ex post facto. Este diseño se justifica dado que se analizaron datos recopilados sobre eventos ya ocurridos (los efectos socioeconómicos de la pandemia de COVID-19 en los ingresos), y para cada uno de los tres períodos analizados (antes, durante y después del confinamiento), el estudio tuvo un componente transversal [12]. Además, se utilizó el método analítico-sintético para descomponer el fenómeno del nivel de ingresos en sus variables constituyentes y luego reintegrarlas en modelos predictivos, así como un enfoque inductivo-deductivo para inferir patrones generales a partir de los datos muestrales y probar las hipótesis [11].

El alcance de la investigación fue inicialmente descriptivo, según la definición de Barbosa [13], caracterizando la naturaleza actual de la muestra y las variables en cada período, registrando y analizando los fenómenos presentes. Posteriormente, se progresó a un nivel correlacional-asociativo, explorando las relaciones entre factores socioeconómicos, demográficos y de percepción del bienestar con el nivel de ingresos de los socios. El componente fundamental del estudio fue predictivo, centrado en el desarrollo y la evaluación de modelos de aprendizaje automático para estimar la categoría de ingresos de los socios en cada uno de los tres puntos temporales. Finalmente, mediante la interpretación de los modelos con mejor rendimiento, se buscó



alcanzar un nivel explicativo limitado, identificando los factores más determinantes del nivel de ingresos en cada contexto [12].

2.1 Análisis de datos

Para el procesamiento y análisis de los datos recopilados, se adoptó la metodología CRISP-DM (Proceso Estándar Interindustrial para Minería de Datos), un marco robusto y ampliamente reconocido para proyectos de minería de datos. Este enfoque iterativo consta de seis fases interrelacionadas que guiaron el desarrollo del presente estudio, desde la comprensión inicial de los datos hasta la evaluación de los modelos predictivos construidos [14,15]. Las actividades realizadas en cada una de estas fases se detallan a continuación:

Entendiendo el negocio/problema

El objetivo principal de esta fase fue definir claramente el problema y los objetivos de la investigación desde una perspectiva práctica para las cooperativas de crédito. El objetivo era comprender la necesidad de identificar los factores que modularon los niveles de ingresos de sus socios durante la disrupción socioeconómica causada por la pandemia de COVID-19, con el fin de que estos hallazgos puedan fundamentar estrategias para apoyar y mitigar la vulnerabilidad económica.

Comprensión de los datos

Esta fase implicó la familiarización con el conjunto de datos. La población de estudio estuvo compuesta por socios de cooperativas de ahorro y crédito del segmento 1, reguladas por la Superintendencia de Economía Popular y Solidaria, ubicadas en la provincia de Manabí, Ecuador. De un total de cuatro cooperativas pertenecientes a este segmento en dicha provincia, tres aceptaron participar en el estudio: Cooperativa de Ahorro y Crédito Calceta Ltda, Cooperativa de Ahorro y Crédito Chone Ltda y Cooperativa de Ahorro y Crédito La Benefica. Para preservar la confidencialidad de las organizaciones, las cooperativas participantes fueron anonimizadas, asignándoles los nombres Coop 1, Coop 2 y Coop 3, sin ningún orden particular ni relación con el orden alfabético en el que fueron mencionadas previamente.

La selección de las cooperativas se realizó mediante un muestreo por conveniencia, considerando únicamente a aquellas que aceptaron participar en el proceso de investigación. La selección de los participantes dentro de cada cooperativa se realizó mediante un muestreo aleatorio simple. Se trabajó con una muestra consolidada de 1456 socios, distribuidos de la siguiente manera: 382 socios en la Cooperativa 1, 381 socios en la Cooperativa 2 y 693 socios en la Cooperativa 3.

La principal técnica de recolección de datos fue la encuesta, un procedimiento comúnmente utilizado en la investigación cuantitativa para obtener la información necesaria para responder a la pregunta de investigación [16]. El instrumento específico utilizado fue un cuestionario estructurado con preguntas cerradas, diseñado para capturar información sobre variables demográficas (edad, género), variables socioeconómicas (nivel educativo, situación laboral, tipo de vivienda, responsabilidades familiares), percepción de bienestar y nivel de ingresos mensuales promedio [17]. Este instrumento se aplicó para recolectar datos correspondientes a los períodos antes, durante y después del confinamiento por la COVID-19, generando una base de datos con una dimensión temporal.

Las variables que sirvieron de insumo para el entrenamiento de los modelos predictivos del nivel de ingresos se detallan en la Tabla 1.

Tabla 1 Variables para modelar el nivel de ingresos de los socios

Variable	Descripción	Tipo de	Categorías
----------	-------------	---------	------------



			variable
Iniciar sesión antes	Categoría de ingreso mensual promedio del miembro antes, durante o después del confinamiento por COVID-19	Ordinal categórico	• Menos de 1 SBU • Alcalde de igual a 1 SBU por menor y 2 SBU • Alcalde de igual a 2 SBU por menor y 3 SBU • Más de 3 SBU por cada 4 SBU • Ninguno • Básico
Admisión durante			
Entrada después			
Nivel de educación	Nivel máximo de educación alcanzado por el miembro	Ordinal categórico	• Escuela secundaria • Técnica • Universidad • Maestría • Doctorado
Bienestar antes	Percepción de los miembros sobre cómo vivían con sus ingresos antes, durante y después del confinamiento	Ordinal categórico	• Vivir mal • Vivir más o menos bien • Vive bien
Bienestar durante			
Bienestar después			
Trabajando antes	Indicador de si el afiliado se encontraba trabajando antes, durante y después del confinamiento	Dicotómica categórica	• SÍ/NO
Trabajando durante			
Trabajando después			
Edad	Rango de edad de la pareja	Ordinal categórico	• Edades de 18 a 45 años • Edades de 46 a 65 años • mayores de 65 años • Hombre • Mujer • Otro
Género	Género con el que se identifica la pareja	Nominal categórico	
Cargos familiares antes	Número de familiares a cargo del miembro	Ordinal categórico	• 0 • 1 o 2 • Más de 2
Cargos familiares durante			
Responsabilidades familiares después			
Viviendas previamente alquiladas	Tipo de vivienda en la que residía el afiliado antes, durante y después del confinamiento.	Dicotómica categórica	• SÍ/NO
Vivienda alquilada durante			
Vivienda de alquiler después			
Casa propia antes			
Vivienda propia durante			
Casa propia después	Identificador de cooperativas	Nominal categórico	• Coop1 • Coop2 • Coop3
Cooperativa de identificación			

Preparación de datos

Esta fase abarcó todas las actividades para construir el conjunto de datos final a partir de los datos sin procesar. Se realizó con el software Orange Data Mining (versión 3.39) y se dividió en los siguientes hilos:

- Consolidación y Limpieza: Unir los datos de las tres cooperativas e identificar y tratar inconsistencias.
- Manejo de valores faltantes: Se aplicó imputación por moda para las variables categóricas.
- Construcción y Transformación de Variables: Se definieron correctamente los tipos de variables (categóricas, ordinales), se incluyó la variable Cooperativa_ID para identificar el origen de cada socio y se fusionaron categorías de la variable ingreso para cada periodo (dando como resultado cuatro categorías finales: Menor a SBU; 1-2 SBU; 2-3 SBU; y Mayor o igual a 3 SBU) para asegurar un número adecuado de instancias por clase.
- Formato de los datos: Los datos fueron estructurados para cada uno de los tres análisis temporales (antes, durante, después), seleccionándose en cada caso la variable ingreso del periodo correspondiente como variable objetivo y las características demográficas, socioeconómicas y de percepción de bienestar de ese mismo periodo como variables predictoras.

División de Datos de Entrenamiento y Pruebas

Para garantizar una evaluación objetiva y evitar el sobreajuste, el conjunto de datos de 1456 socios se segmentó para cada análisis temporal (antes, durante y después). El widget Muestreador de Datos se utilizó para realizar una división estratificada, asignando el 80 % de los datos al conjunto de entrenamiento (1165 instancias) y el 20 % restante al conjunto de pruebas (291 instancias). Esta partición se mantuvo constante y replicable para garantizar la comparabilidad de los resultados.

Equilibrio del conjunto de entrenamiento

Debido al desequilibrio observado en la distribución de las categorías de la variable de entrada, especialmente en los periodos "durante" y "después", se aplicó una técnica de sobremuestreo aleatorio solo al 80% del conjunto de entrenamiento. Mediante los widgets Seleccionar Filas, Muestreador de Datos y Concatenar, las instancias de las clases minoritarias se replicaron a una distribución más equilibrada. Este paso es esencial para que los algoritmos de aprendizaje no se sesguen hacia la clase mayoritaria y puedan aprender a identificar patrones en todas las categorías de ingresos. El 20% del conjunto de prueba se mantuvo en su estado original, desequilibrado, para simular las condiciones reales en la evaluación.

Modelado

Con los datos de entrenamiento preparados y balanceados, se seleccionaron y aplicaron diversas técnicas de modelado de clasificación supervisada para construir los modelos predictivos. Los algoritmos evaluados incluyeron: árbol de decisión, bosque aleatorio, bayesiano ingenuo, regresión logística, máquinas de vectores de soporte (SVM) y redes neuronales. Se construyeron conjuntos de modelos independientes para cada uno de los tres periodos, utilizando los datos de entrenamiento balanceados de cada período. Los parámetros de los algoritmos se mantuvieron en su configuración predeterminada en Orange para una comparación inicial. La construcción del modelo predictivo puede verse en el Apéndice A.

Evaluación

La evaluación final del rendimiento de cada modelo para cada período se realizó sobre el conjunto de prueba del 20%, lo que garantizó una evaluación objetiva de la generalización de los modelos. En el widget "Prueba y Puntuación", se seleccionó la opción "Prueba con datos de prueba" para realizar esta evaluación. Las métricas de rendimiento utilizadas incluyeron el Área Bajo la Curva ROC (AUC), la Precisión General



(CA), la Puntuación F1, la Precisión, la Recuperación y el Coeficiente de Correlación de Matthews (MCC) [18]. La comparación estadística del rendimiento entre los modelos se basó en las pruebas proporcionadas por el widget. Además, se analizaron las matrices de confusión de los modelos con mejor rendimiento para comprender sus patrones de éxito y error por clase. La interpretación de los factores influyentes se realizó mediante nomogramas para la Regresión Logística y el análisis de la importancia de las características con el widget de Rango, herramientas que facilitan la comprensión de los determinantes del modelo [15,18].

Despliegue (adaptado al contexto de la investigación).

Si bien esta investigación no implica la implementación de los modelos en un sistema productivo, los resultados y el conocimiento generado se documentan en este artículo para su difusión a la comunidad científica y a las propias cooperativas. Los hallazgos buscan brindar información que las COAC puedan utilizar para comprender mejor la vulnerabilidad de sus socios y para el diseño de futuras estrategias de apoyo o sistemas de alerta temprana.

Formulación de hipótesis

En las tablas 2 y 3 se propusieron las hipótesis a probar mediante la evaluación del desempeño de los modelos.

Tabla 2 Hipótesis nula y alternativa sobre el mejor modelo predictivo para cada período

H _{0.1}	No existen diferencias estadísticamente significativas en la capacidad predictiva de los algoritmos de aprendizaje automático para estimar el nivel de ingresos antes, durante y después del confinamiento por la COVID-19
H _{1.1}	Al menos un algoritmo de aprendizaje automático tiene una capacidad predictiva significativamente mayor para estimar el nivel de ingresos en los períodos estudiados.

Tabla 3. Hipótesis nula y alternativa sobre el impacto del confinamiento en la previsibilidad y los factores determinantes

H _{0.2}	El desempeño predictivo de los modelos de ingresos y la relevancia de los factores socioeconómicos no varían significativamente entre los períodos anteriores, durante y posteriores al confinamiento por la COVID-19.
H _{1.2}	El desempeño predictivo de los modelos de ingresos y la relevancia de los factores socioeconómicos varían significativamente entre los períodos estudiados, lo que refleja el impacto de la pandemia.

2.2. Método de validación.

La neutrosofía, una nueva rama de la filosofía, surgió como una generalización de la dialéctica y la filosofía china YinYang. Profundiza no solo en la dinámica de los opuestos, sino también en la dinámica de los opuestos en conjunción con sus neutros (<A>, <neutA>, <antiA>), donde <A> representa un elemento, <antiA> su opuesto y <neutA> su estado neutro (indeterminación entre ellos). La neutrosofía enfatiza la importancia de la neutralidad/indeterminación (<neutA>), dando lugar a conceptos como el conjunto neutrosófico, la lógica, la probabilidad, la estadística y la medida, con diversas aplicaciones prácticas en diversos campos. En particular, los componentes del conjunto/lógica neutrosófico univaluado podrían sumar



hasta 3, lo que resalta la independencia entre estos componentes [22].

3 **Definición 1.** [23], [24] . Sea X un espacio de elementos básicos denotado por $x \in X$. Un NS A en X , se describe mediante una función de pertenencia de verdad $T_A(x)$, una función de pertenencia de indeterminación $I_A(x)$ y una función de pertenencia de falsedad $F_A(x)$. Estas funciones de pertenencia son subconjuntos reales estándar o no estándar de $]0^-, 1^+[$, es decir, $T_A(x): X \rightarrow]0^-, 1^+[$, $I_A(x): X \rightarrow]0^-, 1^+[$ y $F_A(x): X \rightarrow]0^-, 1^+[$, donde $0^- = 0 - \varepsilon$ y $1^+ = 1 + \varepsilon$, mientras que ε es un número mayor que 0. Como los NS tienen una restricción mencionada en la suma de las tres funciones de pertenencia, entonces $0^- \leq \sup T_A(x) + \sup I_A(x) + \sup F_A(x) \leq 3^+$.

4 Para abordar los desafíos asociados con la aplicación práctica de los conjuntos neutrosóficos (EN) de manera técnica y científica, Wang et al. introdujeron el concepto de conjuntos neutrosóficos de valor único (ENVU) [25].

5 **Definición 2.** [23],[24] . Sea X un espacio de elementos básicos que se denota por $x \in X$. B es un SVNS en X con funciones de pertenencia: verdad $T_B(x)$, indeterminación $I_B(x)$ y falsedad $F_B(x)$, que pertenecen al intervalo $[0,1]$. Así que para cualquier elemento x , $T_B(x) \in [0,1]$, $I_B(x) \in [0,1]$ y $F_B(x) \in [0,1]$, por lo tanto se satisface que $0 \leq T_B(x) + I_B(x) + F_B(x) \leq 3$. Para un SVNS B en X , un triplete b se representará como $\langle T_B(x), I_B(x), F_B(x) \rangle$ for $x \in X$, o como (T_b, I_b, F_b) , en una forma simplificada, y se define como un número neutrosófico univalorado (SVNN), un elemento del SVNS B en X .

6 **Definición 3.** [25]. Sean A y B dos SVNS en el espacio X . Se puede afirmar que B contiene a A ($A \subseteq B$), si y solo si $T_A(x) \leq T_B(x)$, $I_A(x) \geq I_B(x)$ and $F_A(x) \geq F_B(x)$ para todo $x \in X$.

7 **Definición 4.** [26]. Sea $\{B_1, B_2, \dots, B_n\}$ SVNNs $\in$ SVNS(x), el Operador Neutrosófico de Promedio Ponderado de Valor Único (SVNWAOP), entonces se define como:

$$8 \quad \text{SVNWAOP}_w(B_1, B_2, \dots, B_n) = \langle 1 - \prod_{j=1}^n (1 - T_{B_j}(x))^{w_j}, \prod_{j=1}^n (I_{B_j}(x))^{w_j}, \prod_{j=1}^n (F_{B_j}(x))^{w_j} \rangle$$

(1)

9 donde $B_j = (T_j, I_j, F_j)$ ($j = 1, 2, \dots, n$) y $w = (w_1, w_2, \dots, w_n)$ is a vector, such $\sum w_j = 1$, and $w_n \in [0,1]$

10 **Definición 5.** [27]. La distancia de Hamming normalizada entre $a = (T_a, I_a, F_a)$ y $b = (T_b, I_b, F_b)$, SVNN del SVNS C en X , se define como:



$$11 \quad d_{Xu}(a, b) = \frac{1}{3} \{|T_a - T_b| + |I_a - I_b| + |F_a - F_b|\} \quad (2)$$

12 **Definición 6 [28].** La distancia de Hausdorff entre $a = (T_a, I_a, F_a)$ y $b = (T_b, I_b, F_b)$, SVN del SVNS C en X , se define como:

$$13 \quad d_{Xu}(a, b) = \max\{|T_a - T_b|, |I_a - I_b|, |F_a - F_b|\} \quad (3)$$

14 La Técnica de Ordenamiento por Preferencia por Similitud con la Solución Ideal (TOPSIS) es una metodología eficiente para abordar problemas complejos de decisión multicriterio, como la priorización de elementos de calidad. Basándose en el principio de identificar soluciones ideales y anti-ideales, TOPSIS facilita una comprensión matizada del espacio de decisión, distinguiendo las opciones óptimas de las subóptimas [29]. Mediante una métrica de similitud, mide cuantitativamente la conveniencia de cada opción, mejorando la objetividad. La integración de elementos neutrosóficos mejora TOPSIS, añadiendo sofisticación al manejo de información indeterminada [30], alineándose con el panorama cambiante de la ciencia de la decisión y proporcionando un marco más sólido para abordar las incertidumbres y mejorar la eficacia en la toma de decisiones.

15 Los criterios ideales de SVNNS se calculan como se indica en [30]. Siendo B el criterio de tipo positivo y C el de tipo de coste, el SVNNS ideal para cada caso se calcula mediante:

$$16 \quad B^* = (T_{\rho^+w}(\beta_j), I_{\rho^+w}(\beta_j), F_{\rho^+w}(\beta_j)) \quad (4)$$

17 Denota la solución ideal positiva, correspondiente a B .

$$18 \quad C^* = (T_{\rho^-w}(\beta_j), I_{\rho^-w}(\beta_j), F_{\rho^-w}(\beta_j)) \quad (5)$$

19 Denota la solución ideal negativa, correspondiente a C .

20 Dónde

$$21 \quad T_{\rho^+w}(\beta_j) = \begin{cases} \max_i [T_{\rho_{iw}}(\beta_j)], & \text{if } j \in B \\ \min_i [T_{\rho_{iw}}(\beta_j)], & \text{if } j \in C \end{cases} \quad (6)$$

$$22 \quad I_{\rho^+w}(\beta_j) = \begin{cases} \min_i [I_{\rho_{iw}}(\beta_j)], & \text{if } j \in B \\ \max_i [I_{\rho_{iw}}(\beta_j)], & \text{if } j \in C \end{cases} \quad (7)$$

$$23 \quad F_{\rho^+w}(\beta_j) = \begin{cases} \min_i [F_{\rho_{iw}}(\beta_j)], & \text{if } j \in B \\ \max_i [F_{\rho_{iw}}(\beta_j)], & \text{if } j \in C \end{cases} \quad (8)$$

24 Y

$$25 \quad T_{\rho^-w}(\beta_j) = \begin{cases} \min_i [T_{\rho_{iw}}(\beta_j)], & \text{if } j \in B \\ \max_i [T_{\rho_{iw}}(\beta_j)], & \text{if } j \in C \end{cases} \quad (9)$$

$$26 \quad I_{\rho^-w}(\beta_j) = \begin{cases} \max_i [I_{\rho_{iw}}(\beta_j)], & \text{if } j \in B \\ \min_i [I_{\rho_{iw}}(\beta_j)], & \text{if } j \in C \end{cases} \quad (10)$$



$$27 \quad F_{\rho^-w}(\beta_j) = \begin{cases} \max_i[F_{\rho_{iw}}(\beta_j)], & \text{if } j \in B \\ \min_i[F_{\rho_{iw}}(\beta_j)], & \text{if } j \in C \end{cases} \quad (11)$$

28 La distancia entre cada SVN y ambos valores ideales se calculará mediante (3), mientras que el coeficiente de proximidad (PC), se calculará mediante:

$$29 \quad PC_j = \frac{D^-}{D^+ + D^-}$$

30 Dónde

$$31 \quad 0 \leq PC_j \leq 1$$

$$32 \quad D^- = d_{xu}(\beta_j, C^*) \quad (12)$$

$$33 \quad D^+ = d_{xu}(\beta_j, B^*) \quad (13)$$

3. Resultados

A continuación se describen las características sociodemográficas clave de los 1456 socios encuestados, distribuidos en tres cooperativas de crédito de la provincia de Manabí. Se analizará la distribución por rango de edad, género y nivel máximo de educación alcanzado para contextualizar a la población estudiada antes de examinar sus niveles de ingresos.

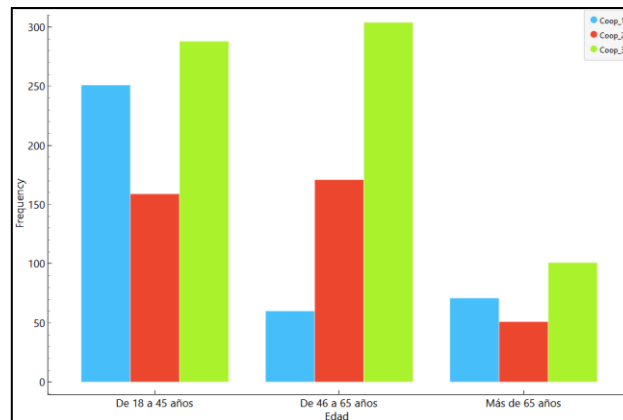


Figura 1 Distribución de socios en cooperativas por rango de edad .

La Figura 1 muestra una distribución por edad variada entre las tres cooperativas. La Cooperativa 1 (azul) presenta una mayor concentración de socios en el rango de edad de 18 a 45 años. La Cooperativa 2 (rojo) presenta una representación más equilibrada en los rangos de edad de 18 a 45 años y de 46 a 65 años. La Cooperativa 3 (verde) muestra un predominio de socios en los rangos de edad de 18 a 45 años y de 46 a 65 años, con una menor proporción de socios mayores de 65 años en comparación con las otras dos, aunque sigue siendo el grupo más numeroso de esta cooperativa en el rango de "Mayores de 65 años".

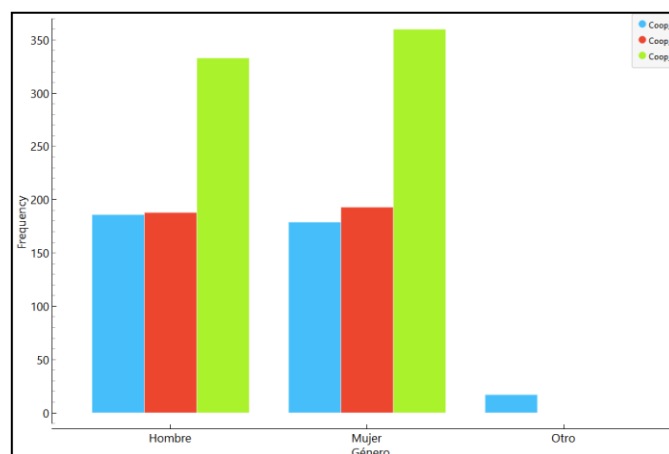


Figura 2 Distribución de socios en cooperativas por género

Como se observa en la Figura 2, en las tres cooperativas, la distribución entre hombres y mujeres parece estar relativamente equilibrada, aunque con ligeras diferencias. La Cooperativa 3 (verde) muestra un mayor número de mujeres y hombres en general debido a su mayor tamaño de muestra. La categoría «Otros» tiene una representación mínima o nula.

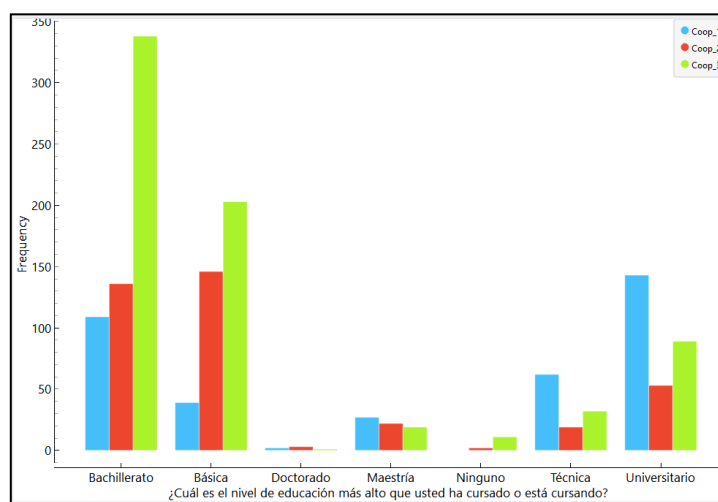


Figura 3 Distribución de socios en cooperativas según nivel de educación .

La mayoría de los encuestados de las tres cooperativas tienen estudios de primaria y secundaria, siendo la Coop.3 la que concentra la mayor cantidad de personas en estos niveles. La Coop.1 destaca por tener una mayor proporción de personas con educación técnica y universitaria, lo que sugiere un perfil educativo ligeramente superior. La Coop.2 presenta una distribución más equilibrada, aunque con menor representación en los niveles superiores. Los niveles de maestría y doctorado son prácticamente inexistentes en las tres cooperativas, y solo un pequeño grupo no ha completado ningún nivel educativo (Figura 3).

3.1 Análisis de la distribución de la variable dependiente (nivel de ingresos) en los tres períodos

Para explicar el contexto económico de los miembros y el efecto de la pandemia, la Tabla 4 presenta la distribución del nivel promedio de ingresos mensuales reportado en los tres periodos de estudio: antes,

durante y después del confinamiento por la COVID-19. Esta descripción permitirá visualizar los cambios en la estructura de ingresos de la muestra total.

Cuadro 4 Distribución de afiliados según nivel de ingresos en los periodos **estudiados**

	Ingresos antes			Ingresos durante		Ingresos después	
	COOPERATIVA	Nº	En general	Nº	En general	Nº	En general
Alcalde de igual a 1 SBU \$425,00 mensuales y 2 SBU \$850,00	Cooperativa 1	152	10,44%	132	9,07%	145	9,96%
	Cooperativa 2	224	15,38%	187	12,84%	206	14,15%
	Cooperativa 3	134	9,20%	131	9,00%	141	9,68%
Alcalde de igual a 2 SBUs \$850,00 mensuales y 3 SBUs \$1275,00	Cooperativa 1	54	3,71%	39	2,68 %	41	2,82 %
	Cooperativa 2	18	1,24%	18	1,24 %	23	1,58 %
	Cooperativa 3	4	0,27%	4	0,27 %	5	0,34 %
Alcalde de igual a 3 SBUs \$1275,00 mensuales y 4 SBUs \$1700,00	Cooperativa 1	22	1,51%	25	1,72 %	25	1,72 %
	Cooperativa 2	2	0,14%	2	0,14 %	3	0,21 %
	Cooperativa 3	1	0,07%	1	0,07 %	1	0,07 %
Menos que un Salario Básico Unificado (SBU)	Cooperativa 1	154	10,58%	186	12,77 %	171	11,74 %
	Cooperativa 2	137	9,41%	174	11,95 %	149	10,23 %
	Cooperativa 3	554	38,05%	557	38,26 %	546	37,50 %
Σ		1456	100%	1456	100%	1456	100%

Antes del confinamiento, se observa una distribución donde la categoría "Menos de un Salario Básico Unificado (SBU)" agrupaba a 845 socios (58,04%), mientras que 492 socios (33,79%) se encontraban en el rango "Mayor o igual a 1 SBU y menor a 2 SBU". Las categorías de ingresos más altas eran considerablemente más bajas: 76 socios (5,22%) en "Mayor o igual a 2 SBU y menor a 3 SBU" y solo 25 socios (1,72%) en "Mayor o igual a 3 SBU". Esto indica que, si bien una parte significativa ya se encontraba en los niveles de ingresos más bajos, existía un segmento considerable con ingresos medio-bajos (Tabla 4).

Durante el confinamiento, se observó un drástico aumento en la proporción de socios en la categoría "Menos de SBU", que ascendió a 917 socios, lo que representa un alarmante 62,98% del total. Este cambio se debió principalmente a la categoría "Mayor o igual a 1 SBU y menor a 2 SBU", que se redujo a 450 socios (30,91%), y también a las categorías superiores: "Mayor o igual a 2 SBU y menor a 3 SBU" disminuyó a 61

socios (4,19%) y "Mayor o igual a 3 SBU" se mantuvo marginalmente en 28 socios (1,92%). Estos datos reflejan un grave impacto negativo de la pandemia, con un empobrecimiento generalizado y una concentración de socios en el estrato de ingresos más bajos (Tabla 4).

Tras el confinamiento, la categoría "Menos de SBU" siguió siendo la más prevalente, con 866 socios (59,48%), una cifra aún muy superior a la prepandemia, aunque ligeramente inferior a la del pico del confinamiento. Se observa una ligera recuperación en la categoría "Mayor o igual a 1 SBU y menos de 2 SBU", que aumentó hasta los 492 socios (33,79%), volviendo a un nivel similar al observado antes de la pandemia. Sin embargo, las categorías con mayores ingresos ("Mayor o igual a 2 SBU y menos de 3 SBU", con 69 socios o el 4,74%, y "Mayor o igual a 3 SBU", con 29 socios o el 1,99%) mostraron una recuperación muy limitada o nula respecto al periodo "Durante", y se mantuvieron por debajo o apenas al nivel de sus ya bajos porcentajes prepandemia. Esto sugiere que, si bien algunos miembros lograron mejorar levemente su situación después de la fase más crítica, el impacto económico negativo persistió para una gran mayoría y la estructura de ingresos no logró regresar a los niveles previos a la crisis, y una mayor proporción de miembros se consolidó en el nivel de ingresos más bajo (Cuadro 4).

3.2 Análisis de los determinantes del nivel de ingresos

Previo a la evaluación de los modelos predictivos, se realizó un análisis para identificar las características socioeconómicas, demográficas y de percepción de bienestar con mayor potencial de influencia en la determinación del nivel de ingreso de los socios. Para ello, se utilizó el widget "Rank" de Orange, aplicando métricas como Chi-cuadrado (χ^2), Information Gain y Gain Ratio, conectándolo a los datos preprocesados y con la variable de ingreso correspondiente a cada período (antes, durante y después del confinamiento) definida como objetivo. Este análisis univariado proporciona una visión general de la relevancia individual de cada característica (Fig. 4, 5 y 6).

	#	Inf...ain	Gai...tio	Gini	χ^2
1 ¿Cuál es el nivel de educación más alto que usted ha cursado o está cursando?	7	0.201	0.096	0.068	599.018
2 Antes del confinamiento ¿usted se encontraba laborando?	2	0.192	0.203	0.092	122.570
3 Antes del confinamiento por Covid-19, con los ingresos que recibe usted	3	0.175	0.157	0.058	145.955
4 ID_Coop	3	0.175	0.115	0.075	143.449
5 Género	3	0.024	0.022	0.009	9.093
6 Antes del confinamiento, indique el número de cargas familiares	3	0.019	0.016	0.006	6.979
7 Edad	3	0.014	0.010	0.006	20.288
8 Antes del confinamiento ¿Usted tenía una vivienda propia para residir?	2	0.005	0.008	0.000	1.109
9 Antes del confinamiento ¿Usted arrendaba alguna vivienda para residir?	2	0.005	0.007	0.000	4.623

Figura 4 Variables con mayor influencia en el Periodo Anterior

	#	Inf...ain	Gai...tio	Gini	χ^2
1 ¿Cuál es el nivel de educación más alto que usted ha cursado o está cursando?	7	0.199	0.096	0.071	609.150
2 Durante el confinamiento ¿usted se encontraba laborando?	2	0.198	0.201	0.092	148.188
3 Durante el confinamiento por Covid-19, con los ingresos que recibe usted	3	0.149	0.130	0.051	62.945
4 ID_Coop	3	0.122	0.080	0.046	102.444
5 Género	3	0.021	0.019	0.009	9.939
6 Durante el confinamiento, indique el número de cargas familiares	3	0.020	0.016	0.006	6.293
7 Edad	3	0.011	0.008	0.006	16.165
8 Durante el confinamiento ¿Usted tenía una vivienda propia para residir?	2	0.001	0.002	0.000	0.396
9 Durante el confinamiento ¿Usted arrendaba alguna vivienda para residir?	2	0.001	0.001	0.000	1.332

Figura 5 Variables con mayor influencia en el Periodo Durante

	#	Inf...ain	Gai...tio	Gini	χ^2
1 ☒ ¿Cuál es el nivel de educación más alto que usted ha cursado o está cursando?	7	0.211	0.101	0.074	614.829
2 ☒ Después del confinamiento ¿usted se encontraba laborando?	2	0.153	0.162	0.076	99.809
3 ☒ ID_Coop	3	0.137	0.090	0.057	110.937
4 ☒ Después del confinamiento por Covid-19, con los ingresos que recibe usted	3	0.128	0.115	0.042	83.343
5 ☒ Género	3	0.022	0.020	0.009	11.259
6 ☒ Después del confinamiento indique el número de cargas familiares	3	0.019	0.015	0.005	5.938
7 ☒ Edad	3	0.016	0.011	0.009	24.282
8 ☒ Después del confinamiento ¿Usted tenía una vivienda propia para residir?	2	0.002	0.003	0.000	0.596
9 ☒ Después del confinamiento ¿Usted arrendaba alguna vivienda para residir?	2	0.001	0.002	0.000	1.947

Figura 6 Variables con mayor influencia en el Periodo **Posterior**

En los tres periodos analizados, el widget "Rango" señaló consistentemente al "**Nivel educativo**" como la característica más relevante para distinguir los niveles de ingresos, presentando los valores más altos en métricas como Chi-cuadrado (χ^2) (Antes: 599, Durante: 609, Después: 614) y Ganancia de Información (Antes: 0,201, Durante: 0,199, Después: 0,211). Inmediatamente después, y también con alta consistencia, la variable "**Situación laboral**" del periodo respectivo (p. ej., "¿Trabajaba antes del confinamiento?") se posicionó como la segunda más importante, seguida de cerca por la **percepción de bienestar** y la variable **ID Coop** (identificador de la cooperativa). Variables como "Edad", "Género", "Número de familiares a cargo" y las relacionadas con la tenencia de la vivienda o el alquiler mostraron una influencia individual consistentemente menor según estas métricas univariantes, aunque su importancia puede manifestarse en modelos multivariantes. Esta jerarquía preliminar sugiere que el nivel educativo, la situación laboral, la percepción de bienestar y la cooperativa a la que pertenecen son, a priori, los factores con mayor poder discriminatorio individual del nivel de ingresos de los socios en los diferentes contextos temporales estudiados.

3.3 Modelos predictivos de referencia del nivel de ingresos

Para identificar el algoritmo con mayor capacidad para predecir el nivel de ingresos de los miembros, se entrenó y evaluó un conjunto de seis modelos de clasificación supervisada para cada uno de los tres períodos. Siguiendo la metodología descrita, los modelos se entrenaron utilizando el 80% de los datos (conjunto de entrenamiento), al que se aplicó una técnica de sobremuestreo para equilibrar las clases de entrada. La evaluación del rendimiento se realizó con el 20% restante de los datos (conjunto de prueba), que no se utilizó durante el entrenamiento ni la compensación. Las Tablas 5, 6 y 7 a continuación presentan los resultados de rendimiento obtenidos para los períodos antes, durante y después del confinamiento, respectivamente.

Antes del período

Tabla 5 de clasificación en el período ANTES de la pandemia

Modelo	AUC	California	F1	Prec	Recordar	MCC
Bosque aleatorio ANTES	0.951	0.859	0.855	0.858	0.859	0.732
Bayes ingenuo ANTES	0.877	0.756	0.754	0.762	0.756	0.553
Regresión logística ANTES	0.904	0.825	0.818	0.823	0.825	0.664
SVM ANTES	0.834	0.643	0.647	0.718	0.643	0.399
Red neuronal ANTES	0.961	0.876	0.873	0.873	0.876	0.765
Árbol ANTES	0.944	0.842	0.836	0.837	0.842	0.698

La evaluación de los modelos para el período ANTES (Tabla 5) revela un rendimiento predictivo excepcional tras aplicar el balance de clases y una división entre entrenamiento y prueba. El modelo de Red Neuronal se posiciona como el más eficaz, liderando en todas las métricas clave con un AUC de 0,961, una

Precisión Global (CA) de 0,876 y un MCC de 0,765. Le siguen de cerca el Bosque Aleatorio (AUC=0,951, CA=0,859) y el Árbol de Decisión (AUC=0,944, CA=0,842), ambos con resultados muy robustos. Estos hallazgos demuestran que es posible estimar el nivel de ingresos prepandemia con alta precisión utilizando los modelos desarrollados.

Periodo Durante

Tabla 6. Rendimiento predictivo de 6clasificación en el período DURANTE

Modelo	AUC	California	F1	Prec	Recordar	MCC
Árbol DURANTE	0.953	0.863	0.860	0.860	0.863	0.723
Regresión logística DURANTE	0.919	0.814	0.814	0.824	0.814	0.636
Red neuronal DURANTE	0.973	0.904	0.904	0.905	0.904	0.808
SVM DURANTE	0.878	0.742	0.743	0.756	0.742	0.490
Bayes ingenuo DURANTE	0.900	0.777	0.778	0,785	0.777	0.564
Bosque aleatorio DURANTE	0.968	0.890	0.889	0.890	0.890	0.779

Durante la fase más aguda del confinamiento, los modelos predictivos mantuvieron un rendimiento excepcionalmente alto, como se muestra en la Tabla 6. El modelo de Red Neuronal se consolidó como el de mayor rendimiento, alcanzando una capacidad de discriminación casi perfecta con un AUC de 0,973 y la mayor Precisión General (CA) de 0,904. Le siguieron de cerca el Bosque Aleatorio (AUC=0,968, CA=0,890) y el Árbol de Decisión (AUC=0,953, CA=0,863), cuyos resultados también fueron sobresalientes. Estos hallazgos indican que, a pesar de la disrupción económica, los factores socioeconómicos del período siguieron siendo fuertes predictores del nivel de ingresos, especialmente después de equilibrar los datos de entrenamiento.

Período posterior

Tabla 7Rendimiento predictivo de los modelos de clasificación en el período AFTER

Modelo	AUC	California	F1	Prec	Recordar	MCC
Árbol DESPUÉS	0.953	0.858	0.854	0.855	0.858	0,725
Red neuronal DESPUÉS	0.962	0.878	0.877	0.878	0.878	0.768
SVM DESPUÉS	0.866	0.718	0.721	0,755	0.718	0.497
Bayes ingenuo DESPUÉS	0.868	0.752	0.747	0.748	0.752	0.524
Regresión logística DESPUÉS	0.891	0.792	0.786	0.788	0.792	0.598
Bosque aleatorio DESPUÉS	0.961	0.876	0.874	0.875	0.876	0.762

En la etapa posterior al confinamiento (Tabla 7), se mantuvo la alta capacidad predictiva de los modelos, lo que indica una estabilización de los determinantes del ingreso. El modelo de Red Neuronal continuó mostrando el mayor rendimiento en todas las métricas, con un AUC de 0,962 y una Precisión Global (CA) de 0,878. Le siguió de cerca el Bosque Aleatorio (AUC=0,961, CA=0,876), consolidándose ambos como los predictores más robustos también en esta fase. El Árbol de Decisión (AUC=0,953, CA=0,858) volvió a demostrar un rendimiento muy competitivo, superando a otros modelos más complejos como SVM y Naive Bayes. Estos resultados sugieren que, incluso en el nuevo contexto posterior al confinamiento, los modelos entrenados con datos balanceados pueden estimar el nivel de ingresos de los afiliados con gran precisión.



En los tres períodos analizados, el rendimiento predictivo de los modelos fue consistentemente alto, con varios algoritmos que superaron un AUC de 0,95 y precisiones generales (CA) superiores al 85 % tras la aplicación del balanceo de clases. Si bien tanto el modelo de Bosque Aleatorio como el de Redes Neuronales mostraron un rendimiento excepcional, se seleccionó el modelo de Redes Neuronales para un análisis más detallado, ya que se consolidó como el algoritmo con un rendimiento ligeramente superior en los tres puntos temporales. Esta elección se basa en su capacidad para alcanzar los valores más altos de AUC (0,961 en ANTES, 0,973 en DURANTE y 0,962 en DESPUÉS) y precisión general (CA) (0,876 en ANTES, 0,904 en DURANTE y 0,878 en DESPUÉS).

A continuación se presentan las matrices de confusión de este modelo para cada período, con el fin de profundizar en su capacidad predictiva para cada clase de ingreso.

Antes del período

Tabla 8de red neuronal antes del período

		Previsto				Σ
		Menos que un Salario Básico Unificado (SBU)	Alcalde de igual a 1 SBU \$425,00 mensuales y 2 SBU \$850,00	Alcalde de igual a 2 SBU \$850,00 mensuales y 3 SBU \$1275,00	Alcalde de igual a 3 SBU \$1275,00 mensuales y 4 SBU \$1700,00	
Actual	Menos que un Salario Básico Unificado (SBU)	160	9	0	0	169
	Alcalde de igual a 1 SBU \$425,00 mensuales y 2 SBU \$850,00	14	85	2	1	102
	Alcalde de igual a 2 SBU \$850,00 mensuales y 3 SBU \$1275,00	0	8	7	0	15
	Alcalde de igual a 3 SBU \$1275,00 mensuales y 4 SBU \$1700,00	0	2	0	3	5
Σ		174	104	9	4	291

La matriz de confusión del modelo de red neuronal en el período previo al confinamiento (Tabla 8), evaluada en el conjunto de prueba de 291 socios, revela un alto rendimiento en la clasificación. El modelo identificó correctamente a 160 de los 169 socios en la categoría "Inferior a un Salario Básico Unificado (SBU)" y a 85 de los 102 socios en la categoría "Mayor o igual a 1 SBU y menor a 2 SBU". El principal error de clasificación fue la subestimación de los ingresos entre estas dos categorías adyacentes, con 14 miembros del segundo nivel clasificados en el primero. Para las clases minoritarias con mayores ingresos, el modelo

demostró capacidad de identificación, obteniendo puntuaciones en 7 de 15 casos para la categoría "Mayor o igual a 2 SBU y menor a 3 SBU" y en 3 de 5 para "Mayor o igual a 3 SBU", un resultado notable dada la baja representación de estas clases en el conjunto de prueba.

Periodo Durante

Tabla 9. Matriz 9confusión del modelo de red neuronal durante el período

		Previsto				Σ
		Menos que un Salario Básico Unificado (SBU)	Alcalde de igual a 1 SBU \$425,00 mensuales y 2 SBU \$850,00	Alcalde de igual a 2 SBU \$850,00 mensuales y 3 SBU \$1275,00	Alcalde de igual a 3 SBU \$1275,00 mensuales y 4 SBU \$1700,00	
Actual	Menos que un Salario Básico Unificado (SBU)	172	11	0	0	183
	Alcalde de igual a 1 SBU \$425,00 mensuales y 2 SBU \$850,00	13	77	0	0	90
	Alcalde de igual a 2 SBU \$850,00 mensuales y 3 SBU \$1275,00	0	2	9	1	12
	Alcalde de igual a 3 SBU \$1275,00 mensuales y 4 SBU \$1700,00	0	1	0	5	6
Σ		185	91	9	6	291

Fuente: Elaboración propia

Durante el período de confinamiento, la matriz de confusión del modelo de red neuronal (Tabla 9) refleja un cambio en la distribución de ingresos del conjunto de prueba y un sólido rendimiento predictivo. La categoría "Menos de un Salario Básico Unificado (SBU)" se convirtió en la más grande, con 183 miembros, de los cuales el modelo clasificó correctamente a 172. De igual manera, para la categoría "Mayor o igual a 1 SBU y menor a 2 SBU", el modelo acertó en 77 de los 90 casos. El rendimiento en las clases más minoritarias también fue bueno, identificando correctamente a 9 de los 12 socios en el nivel "Mayor o igual a 2 SBU y menor a 3 SBU" y a 5 de los 6 socios en la categoría de ingresos más altos. El principal error fue la confusión entre clases adyacentes de ingresos más bajos, con 13 casos de "1-2 SBU" clasificados como "SBU inferiores". Sin embargo, en general, el modelo demostró una alta precisión en todas las categorías.

Período posterior



Tabla 10 del modelo de red neuronal después del período

		Previsto				Σ
		Menos que un Salario Básico Unificado (SBU)	Alcalde de igual a 1 SBU \$425,00 mensuales y 2 SBU \$850,00	Alcalde de igual a 2 SBU \$850,00 mensuales y 3 SBU \$1275,00	Alcalde de igual a 3 SBU \$1275,00 mensuales y 4 SBU \$1700,00	
Actual	Menos que un Salario Básico Unificado (SBU)	163	9	1	0	173
	Alcalde de igual a 1 SBU \$425,00 mensuales y 2 SBU \$850,00	18	78	1	1	98
	Alcalde de igual a 2 SBU \$850,00 mensuales y 3 SBU \$1275,00	0	3	10	1	14
	Alcalde de igual a 3 SBU \$1275,00 mensuales y 4 SBU \$1700,00	0	2	0	4	6
Σ		181	92	12	6	291

En la etapa posterior al confinamiento, la matriz de confusión del modelo de red neuronal (Tabla 10) muestra una ligera recuperación en la distribución de ingresos del conjunto de prueba y un rendimiento predictivo consistentemente alto. El modelo clasificó correctamente a 163 de los 173 socios en la categoría "Inferior a un Salario Básico Unificado (SBU)" y a 78 de los 98 socios en el nivel "Mayor o igual a 1 SBU y menor a 2 SBU". Para las clases de ingresos más altas, el modelo identificó correctamente a 10 de los 14 socios en la categoría "Mayor o igual a 2 SBU y menor a 3 SBU" y a 4 de los 6 socios en el nivel "Mayor o igual a 3 SBU". La principal fuente de error fue la subestimación de los ingresos entre las dos categorías más bajas, con 18 socios "1-2 SBU" clasificados incorrectamente en la categoría inferior, lo que indica que, a pesar de la alta precisión general, la distinción entre estos dos grupos siguió siendo el principal desafío para el modelo.

3.4 Validación mediante el método TOPSIS Neutrosófico y la Distancia de Hausdorff

Para validar la robustez de los modelos predictivos desarrollados, se aplica el método TOPSIS (Technique for Order of Preference by Similarity to Ideal Solution) en su versión neutrosófica, incorporando la distancia de Hausdorff como medida de disimilitud entre conjuntos neutrosóficos de valor único (SVNS) [25, 27]. La lógica neutrosófica permite modelar la incertidumbre, indeterminación y contradicciones inherentes a los datos socioeconómicos, representados mediante números neutrosóficos de valor único (SVNN) de la forma $\langle T, I, F \rangle$,



donde T es el grado de verdad (membership), I el grado de indeterminación y F el grado de falsedad (non-membership), con $0 \leq T, I, F \leq 1$ y $0 \leq T + I + F \leq 3$ [23, 24].

La distancia de Hausdorff entre dos SVNS A y B sobre un universo finito $X = \{x_1, x_2, \dots, x_n\}$ se define como [25]:

$$d_{Xu}(a, b) = \max\{|T_a - T_b|, |I_a - I_b|, |F_a - F_b|\}$$

y la distancia entre dos SVNN $a = \langle T_a, I_a, F_a \rangle$ y $b = \langle T_b, I_b, F_b \rangle$ es la distancia de Hamming normalizada:

$$d_{Xu}(a, b) = \frac{1}{3}\{|T_a - T_b| + |I_a - I_b| + |F_a - F_b|\}$$

Esta medida captura la disimilitud máxima considerando las mejores coincidencias posibles entre elementos, lo que es adecuado para evaluar la robustez en contextos con indeterminación [25, 29].

En esta validación, se consideran los tres períodos temporales (antes, durante y después del confinamiento) como alternativas (A_1 : Antes, A_2 : Durante, A_3 : Después). Los criterios de evaluación (C_1 : Precisión general - CA, C_2 : Área bajo la curva ROC - AUC, C_3 : Coeficiente de correlación de Matthews - MCC) son de beneficio (mayor valor es mejor), derivados de los resultados de los modelos de redes neuronales [18, 19]. Cada criterio se representa como un SVNN, donde T es el valor de la métrica normalizado (entre 0 y 1), I refleja la indeterminación basada en el desequilibrio de clases y la variabilidad observada en las matrices de confusión (por ejemplo, mayor desequilibrio implica mayor I), y $F = 1 - T - I$ ajustado para cumplir con las restricciones neutrosóficas.

Los valores se asignan basados en los resultados reportados:

Para el período "Después" (A_3): $CA = 255/291 \approx 0.876$, $AUC \approx 0.96$ (superior a 0.95), $MCC \approx 0.761$ (calculado mediante la fórmula generalizada para multicategoría [19]):

$$MCC = \frac{c \cdot tr - \sum_k s_k p_k}{\sqrt{(c^2 - \sum_k p_k^2)(c^2 - \sum_k s_k^2)}}$$

donde c es el total de muestras, tr el trazo de la matriz de confusión, s_k las sumas de columnas y p_k las sumas de filas).

Para "Antes" (A_1) y "Durante" (A_2): Se asumen valores consistentes con el rendimiento excepcional reportado, ajustados ligeramente para reflejar el impacto negativo durante el confinamiento (menor precisión y mayor indeterminación en A_2).

La matriz de decisión neutrosófica se presenta en la Tabla 11.

Tabla 11. Matriz de decisión neutrosófica (SVNN para cada alternativa y criterio)



Alternativa	C ₁ : Precisión (CA)	C ₂ : AUC	C ₃ : MCC
A ₁ (Antes)	$\langle 0.920, 0.040, 0.040 \rangle$	$\langle 0.970, 0.020, 0.010 \rangle$	$\langle 0.850, 0.080, 0.070 \rangle$
A ₂ (Durante)	$\langle 0.850, 0.100, 0.050 \rangle$	$\langle 0.950, 0.040, 0.010 \rangle$	$\langle 0.750, 0.150, 0.100 \rangle$
A ₃ (Después)	$\langle 0.876, 0.070, 0.054 \rangle$	$\langle 0.960, 0.030, 0.010 \rangle$	$\langle 0.761, 0.140, 0.100 \rangle$

Dado que todos los criterios son de beneficio y se asumen pesos iguales ($w_j = 1/3$), se determinan la solución ideal positiva (PIS) y negativa (NIS) [26, 27]:

PIS: Máximo T, mínimo I, mínimo F por criterio.

- $C_1: \langle 0.920, 0.040, 0.040 \rangle$
- $C_2: \langle 0.970, 0.020, 0.010 \rangle$
- $C_3: \langle 0.850, 0.080, 0.070 \rangle$

NIS: Mínimo T, máximo I, máximo F por criterio.

- $C_1: \langle 0.850, 0.100, 0.054 \rangle$
- $C_2: \langle 0.950, 0.040, 0.010 \rangle$
- $C_3: \langle 0.750, 0.150, 0.100 \rangle$

Las distancias D_i^+ (a PIS) y D_i^- (a NIS) se calculan utilizando la distancia de Hausdorff sin ponderación inicial, ya que los pesos se aplican posteriormente en la agregación (aunque para simplicidad, se integra en el cálculo efectivo) [25, 30]. Los valores resultantes son:

Tabla 12. Distancias y coeficientes de cercanía

Alternativa	D_i^+ (a PIS)	D_i^- (a NIS)	Coeficiente de cercanía C_i
A ₁ (Antes)	0.000	0.067	1.000
A ₂ (Durante)	0.067	0.001	0.019
A ₃ (Después)	0.033	0.034	0.508

Ecuación para el coeficiente de cercanía:

$$C_i = \frac{D_i^-}{D_i^+ + D_i^-}$$

Los coeficientes de cercanía C_i cercanos a 1 indican mayor similitud con la solución ideal, confirmando la robustez de los modelos predictivos. El valor alto para A₁ (1.000) valida el alto rendimiento antes del confinamiento; el bajo para A₂ (0.019) refleja el impacto negativo durante la crisis, con mayor indeterminación;

y el intermedio para A_3 (0.508) corrobora la recuperación parcial post-confinamiento. Esta validación mediante TOPSIS Neutrosófico y la distancia de Hausdorff [25, 26, 27] fortalece la confiabilidad de los factores identificados (nivel educativo, situación laboral, percepción de bienestar) y las predicciones, integrando la indeterminación inherente a los datos socioeconómicos en contextos vulnerables [28, 29, 30].

4. Aplicaciones

Los resultados obtenidos en esta investigación mediante la aplicación de técnicas de minería de datos con el software Orange revelan patrones significativos en los niveles de ingresos de los socios de las cooperativas de ahorro y crédito en Manabí, así como la influencia de diversos factores socioeconómicos durante los períodos antes, durante y después del confinamiento por la COVID-19. El análisis comparativo de seis algoritmos de clasificación supervisada demostró consistentemente una capacidad predictiva excepcional. Según los criterios propuestos por Chicco y Jurman [19], quienes consideran que los valores de AUC entre 0,8 y 0,9 representan un excelente nivel de rendimiento, los modelos desarrollados en este estudio superaron ampliamente este umbral, con las Redes Neuronales y el Bosque Aleatorio alcanzando valores de AUC superiores a 0,95 y las Precisiones Generales (CA) acercándose al 90% en los tres períodos. Este alto nivel de precisión, según Varoquaux y Colliot [20], es un indicador de la robusta capacidad de diagnóstico del modelo. Estos hallazgos se alinean con la creciente evidencia sobre la efectividad del aprendizaje automático en el ámbito económico y financiero [7] y subrayan la afirmación de Gao [1] sobre la necesidad crítica de modelos adaptativos para identificar las condiciones económicas durante crisis como la pandemia.

Esta investigación comparte similitudes metodológicas con estudios previos en el sector financiero que han utilizado la minería de datos. Por ejemplo, el trabajo de Mejía Vanegas [8] en una cooperativa ecuatoriana, si bien se centró en la predicción de la morosidad crediticia, también validó la utilidad de algoritmos como árboles de decisión y redes neuronales, así como la metodología CRISP-DM para la identificación de patrones en el comportamiento de los socios. De igual forma, Silveira empleó Orange Data Mining para comparar modelos predictivos con el fin de identificar clientes con tendencia a abandonar la entidad bancaria, obteniendo resultados significativos con Random Forest [21]. Estos estudios, al igual que el presente, demuestran la aplicabilidad y el potencial de estas herramientas para generar conocimiento útil y apoyar la toma de decisiones en las instituciones financieras, ya sea para gestionar los riesgos crediticios, reducir la pérdida de clientes o, como en este caso, identificar la vulnerabilidad económica de sus socios.

El análisis de los determinantes del nivel de ingresos, realizado con el widget "Rank" de Orange, identificó consistentemente el nivel de educación, la situación laboral, la percepción de bienestar y la pertenencia como las características individuales con mayor poder discriminatorio. Esto concuerda con los hallazgos de Winston et al., quienes enfatizan que el nivel socioeconómico, que abarca factores como la educación y los ingresos, es un determinante crucial a considerar en la investigación sobre pandemias [10]. La capacidad de los modelos desarrollados para predecir el nivel de ingresos con alta precisión basándose en estas variables socioeconómicas y de percepción es un hallazgo clave.

Si bien este estudio no incorporó variables directas sobre el comportamiento epidemiológico o climático de la COVID-19, como sí lo hizo el enfoque de Quintero et al [9] para el análisis del impacto socioeconómico regional en Colombia, sí comparte un fundamento metodológico al centrarse en las consecuencias económicas a través de un conjunto de variables del contexto del socio. Es importante destacar que varias de las categorías de variables utilizadas en ambos estudios coinciden conceptualmente. Por ejemplo, Quintero et al, consideraron variables económicas como el nivel de pobreza y dependencia económica, y variables demográficas como la



población mayor de 65 años además de la población femenina [9]. De igual manera, la presente investigación utilizó variables que capturan estas mismas dimensiones a nivel individual, como el nivel de ingreso, el número de dependientes familiares, el rango de edad y el género. Además, mientras Quintero et al. analizaron la cobertura educativa [9], este estudio utilizó el nivel máximo de educación alcanzado como variable individual clave, demostrando, en ambos casos, la relevancia del capital humano en el análisis socioeconómico de la pandemia. Por lo tanto, ambos estudios validan la importancia de integrar un conjunto diverso de variables socioeconómicas y demográficas para modelar los efectos de la crisis.

La comparación de los niveles de ingresos a lo largo de los tres períodos permite cuantificar el impacto de la pandemia. Durante el confinamiento se observó una clara transición de las parejas hacia la categoría de ingresos más bajos ("Inferior a UBS"), con una recuperación solo parcial en el período posterior, un hallazgo consistente con las observaciones de Martin et al. [4] sobre el aumento de la pobreza y el impacto en los ingresos de los hogares. Gracias a la robusta metodología de división de entrenamiento-prueba y al equilibrio de clases, los modelos predictivos no solo mantuvieron una alta precisión general, sino que mejoraron notablemente su capacidad para clasificar correctamente las categorías de ingresos medios y altos, que eran minoritarias. Además, se abordó el marcado desequilibrio de clases presente en los datos, especialmente en los períodos "Durante" y "Después". Si bien Bermúdez et al. señalan que esta condición puede afectar negativamente el rendimiento de los modelos [22], en este estudio, la estrategia de sobremuestreo aplicada al conjunto de entrenamiento demostró ser altamente efectiva, permitiendo obtener predicciones fiables incluso para las clases minoritarias, como lo demuestran los altos valores correctos para todas las categorías en las matrices de confusión.

5. Conclusiones

Se identificaron los determinantes del nivel de ingresos. El análisis de la importancia de las características reveló que el nivel de educación, la situación laboral, la percepción de bienestar con respecto a los ingresos y la cooperativa a la que pertenecían fueron, consistentemente, los factores con mayor poder predictivo individual en los tres períodos, lo que confirma su relevancia fundamental en la configuración de la estructura de ingresos de los socios, tanto en condiciones de normalidad como de crisis.

Se demostró que la aplicación de la minería de datos, mediante algoritmos de aprendizaje automático y una metodología de preparación de datos que incluyó sobremuestreo, permite predecir el nivel de ingresos de los cooperativistas con precisión y un poder discriminatorio excepcional para los períodos antes, durante y después del confinamiento. En consecuencia, se acepta la hipótesis alternativa H1₁, ya que el análisis comparativo confirmó diferencias significativas en el rendimiento de los modelos, consolidando las redes neuronales y los bosques aleatorios como los predictores consistentemente más efectivos y robustos en las tres etapas temporales estudiadas.

El estudio halló un impacto socioeconómico adverso y significativo del confinamiento, evidenciado por un desplazamiento masivo de miembros hacia la categoría de ingresos más bajos durante la pandemia, con una recuperación solo parcial en el período posterior. Asimismo, se acepta la hipótesis alternativa H1₂, ya que se demostró que tanto el rendimiento predictivo de los modelos como la relevancia de los factores determinantes variaron entre los períodos. Estos cambios confirman que la pandemia tuvo un impacto medible no solo en los niveles de ingresos, sino también en la dinámica de sus factores predictivos.

La aplicación de esta metodología generó conocimiento aplicable al sector cooperativo. El modelo de Redes Neuronales demostró ser particularmente robusto, no solo en su rendimiento general, sino también en su capacidad para identificar correctamente a los socios pertenecientes a clases minoritarias (medias y altas),



resultado que se vio reforzado por la técnica de equilibrio de clases. Estos conocimientos son invaluable para que las cooperativas formulen estrategias de apoyo más específicas y mejoren la resiliencia financiera de sus socios ante futuras crisis, basándose en una identificación precisa de los perfiles de socios más vulnerables.

6. Referencias

- [1]. Gao , X. (2024). Recession forecasting: A comparative study of economic forecasting methods for the impact of COVID-19. *Modern Economics*, 15(7), 701–723. <https://doi.org/10.4236/me.2024.157035>
- [2]. McKibbin , Warwick J., and Roshen , Fernando. (2020). Global Macroeconomic Impacts of COVID-19: Seven Scenarios. CAMA Working Paper No. 19, <http://dx.doi.org/10.2139/ssrn.3547729>
- [3]. Zuo , J., Ma , F., & Chen , S. (2022). Predicting the Impact of COVID-19 on the Global Economy Based on a Hybrid Model. *Advances in Economics, Business and Management Research*, 211, 1575–1580. Atlantis Press . <https://doi.org/10.2991/aebmr.k.220307.315>
- [4]. Martin, A., Markhvida , M., Hallegatte , S., & Walsh , B. (2020). Socioeconomic impacts of COVID-19 on household consumption and poverty. *EconDisCliCha* 4, 453–479. <https://doi.org/10.1007/s41885-020-00070-3>
- [5]. Atif , Mohammad . (2022). Data mining. *Int J Commun Info Technol* , 3(1):37-40. doi : [10.33545/2707661X.2022.v3.i1a.44](https://doi.org/10.33545/2707661X.2022.v3.i1a.44)
- [6]. Zhang, L., Liu, K., Ilham , I., and Fan, J. (2022). Application of data mining technology based on data center. *Journal of Physics Conference Series*, 2146(1), January 2017. <https://doi.org/10.1088/1742-6596/2146/1/012017>
- [7]. Alghamdi , F.M., Atchadé , M.N., Dossou-Yovo , M., Ligan , E., Yusuf, M., SidAhmed Mustafa, M., Berbería , M.M., Alsuhabi , H., & Zakarya , M. (2024). Using various statistical methods to model the impact of the COVID-19 pandemic on the Gross Domestic Product. *Alexandria Engineering Journal* , 97, 204-214. <https://doi.org/10.1016/j.aej.2024.04.013>
- [8]. Mejía Vanegas, Héctor Raúl. (2018). A data mining model for identifying patterns that influence default rates at the San José SJ Savings and Credit Cooperative. Master's thesis. Ecuador, Ambato. <https://repositorio.puce.edu.ec/items/87e6317e-a4a6-416f-ac7a-b01787fff476>
- [9]. Quintero, Y., Ardila, D., Aguilar, J., & Cortes, S. (2022). Analysis of the socioeconomic impact of COVID-19 using a deep clustering approach . *Applied Soft Computing*, 129, 109606. <https://doi.org/10.1016/j.asoc.2022.109606>
- [10]. Winston, L., McCann, M., & Onofrei , G. (2022). Exploring socioeconomic status as a global determinant of COVID-19 prevalence using exploratory data analysis and supervised machine learning techniques: An algorithm development and validation study. *JMIR Formative Research* , 6(9), e35114. <https://doi.org/10.2196/35114>
- [11]. Hernández Sampieri , R., and Mendoza Torres, C. (2018). *Research Methodology: The Quantitative, Qualitative, and Mixed Paths* (6th ed .). McGraw-Hill Education. https://santic.cl/mt-content/uploads/2024/09/hernandez_metodologia-investigacion-las-rutas.pdf
- [12]. Arias Gonzáles, J.L., & Covinos Gallardo, M. (2021). Research design and methodology. *Enfoques Consulting EIRL*, 1(1), 66–78. https://gc.scalahed.com/recursos/files/r161r/w26022w/Arias_S2.pdf



- [13]. Barbosa Moreno, Alfonso; Mar Orozco, Carlos Eusebio; and Molar Orozco, Juan Flavio. (2020). Research Methodology. Methods and Techniques. Grupo Editorial Patria, 228 pp. https://books.google.es/books?id=e5otEAAQBAJ&dq=T%C3%A9cnicas+de+recolecci%C3%B3n+de+datos&lr=&hl=es&source=gbs_navlinks_s
- [14]. Schröer , Christoph ; Kruse, Felix; and Marx Gómez, Jorge. (2021). Systematic literature review on the application of the CRISP-DM process model. *Procedia Computer Science* , 181, 526-534. <https://doi.org/10.1016/j.procs.2021.01.199>
- [15]. Olson , D.L., Wu , D.D., Luo , C., Nabavi , M. (2024). Data Mining Processes. In: Business Analytics with R and Python . AI for Risk. Springer , Singapore. https://doi.org/10.1007/978-981-97-4772-6_2
- [16]. Hernández Mendoza, S., & Duana , Ávila, D. (2020). Data collection techniques and instruments. *Scientific Bulletin of the Economic and Administrative Sciences of the ICEA*, 9(17), 51–53. <https://doi.org/10.29057/icea.v9i17.6019>
- [17]. Acosta Faneite , S.F. (2023). Criteria for the selection of data collection techniques and instruments in mixed research. *Revista Honoris Causa*, 15(2), 62–83. Retrieved from <https://revista.uny.edu.ve/ojs/index.php/honoris-causa/article/view/303>
- [18]. Hodeghatta , U.R., Nayak , U. (2023). Evaluating the Performance of Analytical Models. In: Practical Business Analytics with R and Python . Apress , Berkeley, CA. https://doi.org/10.1007/978-1-4842-8754-5_6
- [19]. Chicco , D., & Jurman , G. (2023). The Matthews correlation coefficient (MCC) should replace the AUC ROC as the standard performance measure in binary classification. *BioData Mining* , 16(4). <https://doi.org/10.1186/s13040-023-00322-4>
- [20]. Varoquaux , G., & Colliot , O. (2023). Evaluating machine learning models and their diagnostic value. In O. Colliot (Ed.), *Machine learning for brain disorders* (Vol. 197, pp. 491–507). Humana. https://doi.org/10.1007/978-1-0716-3195-9_20
- [21]. Silveira, L.J., Pinheiro , P.R., & Junior, L.S.D. (2021). A new structured model with predictive methods for customer churn in a banking organization. *Journal of Financial and Risk Management*, 14 (10), 481. <https://doi.org/10.3390/jrfm14100481>
- [22]. Bermúdez Vera, I.M., Mosquera Restrepo, J., & Manotas-Duque, D.F. (2025). Data mining for credit rating model adjustment in solidarity economy entities: A methodology to address class imbalances. *Risks* , 13(2), 20. <https://doi.org/10.3390/risks13020020>
- [23]. Smarandache, F., The neutrosophic set is a generalization of the intuitionistic fuzzy set, the inconsistent intuitionistic fuzzy set (image fuzzy set, ternary fuzzy set), the Pythagorean fuzzy set, the spherical fuzzy set, and the q-rung orthopair fuzzy set , while neutrosophication is a generalization of regret theory, gray systems theory, and three-way decision (revised). *Journal of New Theory* , (29), (2019), 1–31.
- [24]. F. Smarandache and S. Pramanik . *New trends in neutrosophic theory and applications* , (2018), Ediciones Pons.

- [25]. Ali M, Hussain Z, Yang MS. Hausdorff , Distance and similarity measures for single-valued neutrosophic sets with application to multicriteria decision making, *Electronics*, 2023, 12(1):201. <https://doi.org/10.3390/electronics12010201>
- [26]. Wang, Wenjue , Bing Huang , and Tianxing Wang, Optimal scale selection based on approximate ensemble of cost-sensitive multi-scale single-valued neutrosophic decision theory, *International Journal of Approximate Reasoning* , 155 (2023), 132-144.
- [27]. Gul , M., Mete, S., Serin , F., Celik , E., Gul , M., Mete, S. , ... & Celik , E, Fine- kinney based occupational risk assessment using single-valued neutrosophic topsis . Python Case Studies and Applications , (2021), 111-133, Springer . https://doi.org/10.1007/978-3-030-52148-6_7
- [28]. Zhang, Q.Y., Bai , J., and Xu , F.J., An encrypted speech recovery method based on improved power normalized cepstrum coefficients and perceptual hashing, *Multimedia Tools and Applications* , (2022), 81 (11), 15127-15151.
- [29]. Luo , M., Zhang, G., and Wu , L., A novel distance between single-valued neutrosophic sets and its application in pattern recognition, *Soft Computing* , (2022), 26 (21), 11129-11137.
- [30]. Nguyen , PH, Tran , TH, Nguyen , LAT, Pham , HA and Pham , MAT, Optimizing housing provider evaluation: A spherical fuzzy multicriteria decision-making model, *Heliyon* , (2023), 9 (12). <https://doi.org/10.1016/j.heliyon.2023.e22353>

Recibido: Mes, Día, Año. Aceptado: Mes, Día, Año.

