9 No.3 (2025): Journal Scientific MQRInvestigar ISSN: 2588–0659
https://doi.org/10.56048/MQR20225.9.3.2025.e877

Intelligent Credit Assessment: An Explainable Approach with Machine Learning and Large Language Models

Evaluación Inteligente del Riesgo Crediticio: Un Enfoque Explicable con Aprendizaje Automático y Modelos de Lenguaje de Gran Tamaño

Autores:

Guamán, Henry
UNIVERSIDAD INDOAMÉRICA
Facultad de Ingenierías
Ingeniería Industrial
Ambato – Ecuador

hguaman@indoamerica.edu.ec

https://orcid.org/0009-0007-0608-7425

Orellana, Marcos
UNIVERSIDAD DEL AZUAY
Computer Science Research & Development Laboratory (LIDI),
Cuenca – Ecuador

marore@uazuay.edu.ec

https://orcid.org/0000-0002-3671-9362

Fechas de recepción: 30-JUN-2025 aceptación: 30-JUL-2025 publicación: 30-SEP-2025

https://orcid.org/0000-0002-8695-5005
http://mqrinvestigar.com/

Resumen

Este estudio aborda la evaluación del riesgo crediticio en cooperativas mediante la integración de un modelo de árbol de decisión explicable y un chatbot basado en Phi4. Se utilizó un conjunto de datos de 5000 registros, considerando variables clave como edad, número de dependientes, antigüedad en la institución, monto solicitado, segmento crediticio y duración del empleo. El modelo de árbol de decisión, configurado con parámetros optimizados (profundidad máxima, ccp_alpha, número mínimo de muestras para división y hojas), alcanzó una precisión del 90,47%, demostrando su capacidad para discriminar entre clientes "seguros" e "inseguros". Además, se evaluaron otros modelos (OneR, PART y PRISM), siendo PART y el árbol de decisión los que presentaron el mejor equilibrio entre precisión e interpretabilidad. La incorporación del chatbot, entrenado mediante técnicas de transferencia de aprendizaje y desplegado en un entorno local seguro, proporcionó explicaciones claras sobre las decisiones crediticias, facilitando auditorías por organismos reguladores. La propuesta destaca la importancia de emplear enfoques de inteligencia artificial explicable (XAI) para mejorar la inclusión financiera, optimizar recursos y reducir los tiempos de procesamiento. Se reconocen limitaciones relacionadas con la calidad del conjunto de datos y se sugiere integrar modelos híbridos y expandir las fuentes de información en futuras investigaciones para lograr evaluaciones más robustas y adaptables. Este enfoque integral mejora significativamente la eficiencia y la transparencia en la gestión.

Palabras clave: Aprendizaje automático; Árbol de decisión; Cooperativas; Evaluación crediticia; Inteligencia artificial; Riesgo crediticio

Abstract

This study addresses the evaluation of credit risk in cooperatives through the integration of an explainable decision tree model and a Phi4-based chatbot. A dataset of 5000 records was used, considering key variables such as age, number of dependents, tenure at the institution, requested amount, credit segment, and employment duration. The decision tree model, configured with optimized parameters (maximum depth, ccp_alpha, minimum samples for splitting and in leaves), achieved an accuracy of 90.47\%, demonstrating its ability to accurately discriminate between “safe” and “unsafe” clients. Additionally, other models (OneR, PART, and PRISM) were evaluated, with PART and the decision tree showing the best balance between accuracy and interpretability. The incorporation of the chatbot, trained using transfer learning techniques and deployed in a secure local environment, provided clear explanations of credit decisions, facilitating audits by regulatory bodies. The proposal highlights the importance of employing explainable artificial intelligence (XAI) approaches to improve financial inclusion, optimize resources, and reduce application processing time. Limitations related to dataset quality are acknowledged, and future research is suggested to integrate hybrid models and expand data sources to achieve more robust and adaptable evaluations across various credit contexts. This comprehensive approach significantly enhances efficiency and transparency in management.

Keywords: Machine learning; Decision tree; Cooperatives; Credit assessment; Artificial intelligence; Credit risk

Introducción

En las economías emergentes, las cooperativas de crédito enfrentan desafíos críticos en la evaluación del riesgo crediticio, ya que una gran parte de sus clientes proviene de sectores tradicionalmente desatendidos o excluidos por la banca convencional (Rabbani et al., 2024; Karn et al., 2022). En Ecuador, por ejemplo, las cooperativas rurales presentan tasas de morosidad particularmente elevadas (Shetu et al., 2021), y en muchos casos, las solicitudes de crédito son rechazadas debido al uso de modelos de evaluación conservadores y a la falta de datos integrales sobre los solicitantes (Zhao & Zhao, 2021; ICACNIS, 2022). Esta situación no solo restringe el acceso al financiamiento, sino que también perpetúa la exclusión financiera de poblaciones vulnerables (Malakauskas & Lakstutiene, 2021; Ko et al., 2022).

A nivel global, el sector financiero ha comenzado a integrar tecnologías de inteligencia artificial, como el aprendizaje automático (Machine Learning, ML) y los modelos de lenguaje de gran tamaño (Large Language Models, LLMs), para abordar estos desafíos. Estas tecnologías permiten analizar grandes volúmenes de información y extraer patrones relevantes, resultando en evaluaciones de riesgo más precisas y eficientes (ICETSIS, 2024; Lusinga et al., 2021; Khenfouci & Challal, 2023; IC, 2023). Sin embargo, en el contexto ecuatoriano, la adopción de estas innovaciones está limitada por barreras de costo, infraestructura tecnológica deficiente y la ausencia de datos alternativos y conductuales que permitan una evaluación integral de los solicitantes que no encajan en los perfiles tradicionales (Majumder et al., 2022; Tahyudin et al., 2023; Liu et al., 2023; Anusha & Bhowmik, 2023).

La implementación de sistemas basados en ML y LLMs ofrece una oportunidad única para mejorar la inclusión financiera y reducir la morosidad, al generar evaluaciones más justas y precisas (IC, 2023; Khan et al., 2022). Estudios recientes han demostrado que el uso de técnicas de ML en la evaluación de riesgo puede disminuir la tasa de morosidad en más del 50%, lo que impacta positivamente en la estabilidad financiera de las instituciones (Shen, 2022; Liu et al., 2021). Además, la automatización de procesos mediante estas tecnologías tiene el potencial de incrementar la productividad del personal, permitiendo manejar hasta

un 40% más de solicitudes en comparación con métodos tradicionales (Ke et al., 2021; Paul et al., 2024).

Este proyecto se enfoca en abordar el problema de la alta morosidad y el acceso restringido al crédito en ciertos sectores de la población ecuatoriana (Chaudhary & Saroj, 2023). La dependencia de modelos de evaluación crediticia poco adaptativos y la falta de datos alternativos han limitado históricamente la capacidad de las cooperativas para tomar decisiones justas y transparentes (Al-Ameer & Al-Sunni, 2021). Para superar estas limitaciones, se propone el desarrollo de un sistema de evaluación crediticia que integre algoritmos de ML —específicamente modelos basados en árboles de decisión— y LLMs, con el objetivo de analizar datos financieros, conductuales y alternativos para proporcionar una evaluación del riesgo más precisa y explicable (Xia et al., 2021; Al Khaldy et al., 2024).

La metodología propuesta combina enfoques cuantitativos y cualitativos. Por un lado, se emplean algoritmos de aprendizaje automático, con árboles de decisión configurados para alcanzar precisiones en el rango del 85–90% (Chen et al., 2024). Por otro lado, la incorporación de un modelo de lenguaje (LLM) permite generar explicaciones claras y comprensibles sobre las decisiones crediticias, reforzando la transparencia del proceso y fomentando la confianza del solicitante (Liu, 2024). En un estudio piloto realizado en la Cooperativa “Unión Familiar”, se demostró que implementar este sistema no solo reduce el tiempo de procesamiento de solicitudes a 1–2 días, sino que también optimiza la evaluación del riesgo, mostrando un desempeño superior en la clasificación de clientes “seguros” e “inseguros” (Edo et al., 2023; Silva et al., 2023; Zhu & Tingting, 2024; Li, 2023).

Los resultados preliminares indican que modelos como el Árbol de Decisión y enfoques basados en PART ofrecen un equilibrio apropiado entre precisión e interpretabilidad, alcanzando métricas de desempeño superiores (precisión > 90%) en comparación con modelos más simples. Este hallazgo respalda la viabilidad del enfoque propuesto y destaca el potencial de integrar ML y LLMs para transformar la evaluación crediticia en contextos con alta vulnerabilidad financiera.

Material y métodos

El desarrollo del sistema adaptativo de evaluación crediticia se basa en una serie de etapas claramente definidas, cuyo diagrama de flujo general se ilustra en la Figura 1. Este diagrama resume el proceso completo, desde la recolección de datos hasta la integración de modelos de lenguaje de gran tamaño (LLMs) para la interpretación de decisiones.

Conjunto de datos

La base de datos utilizada en este estudio consta de 5000 registros y 13 variables, que representan información demográfica, laboral y crediticia de solicitantes de préstamos en cooperativas de ahorro y crédito. El conjunto de datos anonimizado fue estratificado y dividido en una porción del 70% para entrenamiento y un 30% para prueba utilizando la función *train_test_split(test_size=0.3, random_state=7)*. Dentro del conjunto de entrenamiento, se reservó aleatoriamente el 20% para validación interna, obteniéndose así una división efectiva del 56%/14%/30% para entrenamiento, validación y prueba externa, respectivamente. Este diseño mantiene la proporción de clases entre conjuntos y proporciona una estimación imparcial del rendimiento de generalización. A continuación, se detallan las principales características del conjunto de datos:

- Edad: Variable numérica, media de 33.99 años.
- Estado civil: Variable categórica con seis categorías; la más frecuente es "Casado" (2078 registros).
- Número de dependientes: Media de 1.55 dependientes.
- Antigüedad en la institución: Medida en meses, media de 33.95 meses.
- Monto solicitado: Promedio de $8,832.97.
- Número de cuotas: Promedio de 49.98 cuotas.
- Segmento crediticio: Principalmente “Microcrédito minorista” (2509 registros).
- Tipo de pago: Predominantemente “Cuotas fijas” (4490 registros).
- Destino del préstamo: Principalmente “Capital de trabajo” (2057 registros).

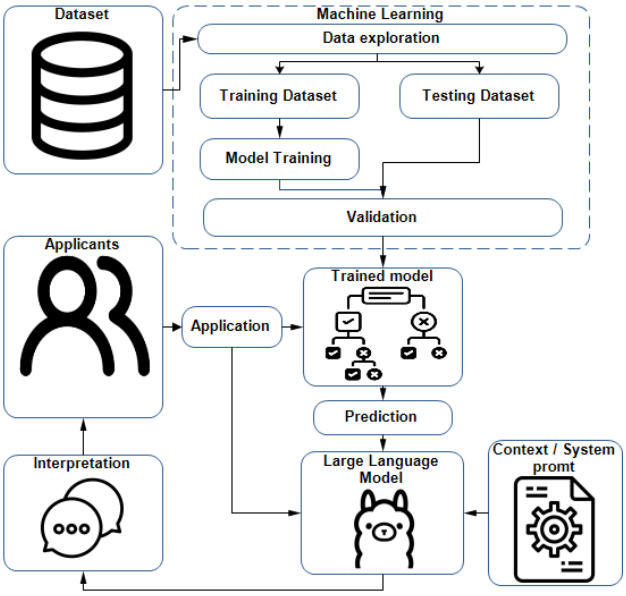
- Tipo de garantía: Más comúnmente “Sin garantía” (2456 registros).
- Tipo de empleador: Predominantemente “Empleado privado” (2975 registros).
- Tiempo de empleo: Media de 103.37 meses.
- Resultado (variable objetivo): “Seguro” (3633 registros) o “Inseguro”.

El proceso de preparación de datos incluyó las siguientes fases:

- Limpieza: Verificación de calidad y ausencia de valores nulos o inconsistentes.
- Normalización: Ajuste de escalas en variables como edad, tiempo de empleo y monto solicitado.
- Codificación: Variables categóricas transformadas mediante codificación one-hot.

Figura 1

Diagrama de flujo del sistema adaptativo de evaluación crediticia basado en el aprendizaje automático y el lenguaje natural



Fuente: Elaboración propia

Aprendizaje Automático

Se exploraron varios algoritmos de clasificación para evaluar el riesgo crediticio:

- Árbol de decisión: Elegido por su alta interpretabilidad y capacidad para identificar patrones complejos.
- OneR: Modelo simple basado en reglas, usado como referencia.
- PART: Algoritmo que genera reglas parciales, combinando árboles y reglas.
- PRISM: Algoritmo basado en reglas determinísticas, maximizando la pureza en cada paso.

La validación se realizó dividiendo el conjunto de datos en conjuntos de entrenamiento y prueba, y evaluando cada modelo utilizando métricas como precisión, recuperación y puntuación f1.

El ajuste de hiperparámetros se llevó a cabo con GridSearchCV junto con un StratifiedKFold de cinco pliegues *(shuffle=True, random_state=42)*. La cuadrícula de búsqueda para el clasificador de árbol de decisión comprendía:

$$
\begin{aligned}
&\textit{Criterion} \in \{\textit{gini, entropy, log\_loss}\},\\
&\textit{max\_depth} \in \{\textit{None, 3, 5, 7, 10, 15, 20, 30, 50, 100}\},\\
&\textit{ccp\_alpha} \in \{0,\ 10^{-4},\ 10^{-3},\ 5\times 10^{-3},\ 0.01,\ 0.02,\ 0.05,\ 0.1,\ 0.2\},\\
&\textit{min\_samples\_split} \in \{2,\ \ldots,\ 50\},\\
&\textit{min\_samples\_leaf} \in \{2,\ \ldots,\ 50\},\\
&\textit{max\_features:} \in \{\textit{None}, \sqrt{\ }, \log_2, 0.5, 0.75, 0.9\},\\
&\textit{class\_weight} \in \{\textit{None, balanced}\}.
\end{aligned}
$$

Se utilizaron cuadrículas idénticas (adaptadas cuando fue necesario) para las bases de referencia Random-Forest y PART. Todas las búsquedas adoptaron pliegues estratificados para contrarrestar el desequilibrio de clases (∼ 73 % de préstamos seguros frente a 27 % de préstamos inseguros).

Entrenamiento del modelo

Cada algoritmo se configuró y entrenó teniendo en cuenta los siguientes aspectos:

Árbol de decisión

- Se optimizó la ganancia en cada nodo y se establecieron criterios de división basados en la información (ganancia).
- Se limitó la profundidad del árbol para evitar el sobreajuste, lo que permitió identificar el número óptimo de nodos y hojas.
- La estructura resultante facilita la visualización y la comprensión de las reglas de clasificación (véase la figura 5).

OneR

- Entrenado para identificar la regla más relevante en la clasificación, lo que lo hace sencillo, pero menos robusto.

PART

- Configurado para generar reglas parciales combinando la construcción de subárboles con la extracción de reglas, lo que ofrece un equilibrio entre complejidad y rendimiento.

PRISM

- Entrenado mediante la construcción de reglas específicas para cada clase, seleccionando los atributos más discriminatorios en cada iteración.
- La implementación priorizó la pureza de las reglas en la clasificación, garantizando la máxima precisión dentro de cada conjunto de reglas generado.

La implementación se llevó a cabo en Python, utilizando la biblioteca Scikit-learn para configurar y ajustar los hiperparámetros de cada modelo.

Integración del LLM y diseño de prompts

Para complementar la toma de decisiones y mejorar la transparencia del sistema, se integró un modelo de lenguaje grande (LLM), concretamente Phi4. Su incorporación se llevó a cabo de la siguiente manera:

- Solicitud del sistema: Se utilizó una solicitud del sistema especializada para delimitar el contexto a la información relevante para la evaluación crediticia.
- Implementación segura y escalable: La integración se realizó localmente utilizando el marco *Gradio* para implementar una interfaz gráfica, lo que garantiza la seguridad y la funcionalidad del sistema.

Tabla 1

Rendimiento medio de los modelos en las pruebas AMC 10/12 de noviembre de 2024 y disponibilidad para ejecución local

Model	Average Score	Locally Executable
Llama-3.3 70B Instruct	66.4	Yes
Claude 3.5 Sonnet	74.8	No
Qwen 2.5 14b-Instruct	77.4	Yes
GPT-4o	77.9	No
GPT-4o-mini	78.2	No
Qwen 2.5 72b-Instruct	78.7	Yes
Gemini Flash 1.5	81.6	No
Gemini Pro 1.5	89.8	No
phi-4	**91.8**	Yes

La tabla 1 presenta el rendimiento medio de varios modelos lingüísticos evaluados en las pruebas AMC 10/12 de noviembre de 2024 y añade información clave sobre su viabilidad para la ejecución local. Muestra que algunos modelos, como Llama-3.3 70B Instruct, Qwen 2.5 14b-Instruct, Qwen 2.5 72b-Instruct y phi-4, son de código abierto y pueden ejecutarse en entornos locales, lo que facilita su integración directa en sistemas con hardware adecuado. Por el contrario, modelos como Claude 3.5 Sonnet, GPT-4o (y su versión mini) y los modelos Gemini (Flash 1.5 y Pro 1.5) solo están disponibles a través de API, lo que limita su uso a entornos en la nube. Cabe destacar que phi-4 destaca como el que mejor rendimiento ofrece (91,8) y como una opción versátil para implementaciones locales, mientras que Gemini Pro 1.5, a pesar de su alta puntuación (89,8), debe utilizarse a través de API. Estas diferencias ponen de relieve la correlación entre la capacidad del modelo, el rendimiento de referencia y la flexibilidad de implementación, lo que permite seleccionar la opción más adecuada en función de los requisitos del sistema (Abdin, et al., 2024).

Prompt de sistema

A continuación, se muestra un ejemplo del mensaje del sistema utilizado para ajustar Phi4 en el contexto de la evaluación crediticia:

"Hola, soy tu asesor crediticio. Responderé de manera breve y directa.
Utiliza la siguiente información interna para evaluar el perfil crediticio del usuario y emitir una recomendación.
[Reglas generadas por el mejor modelo].
Aplica estas directrices de manera interna y discreta para orientar tus respuestas."

Entorno computacional

Las pruebas se realizaron en Google Colab (CPU, 12 GB RAM). El modelo de Árbol de Decisión se entrenó en menos de 0.3 s y la inferencia para 1500 registros tomó solo 7 ms. Posteriormente, el sistema completo fue desplegado en un equipo con procesador AMD Ryzen 9 3900X, 32 GB RAM y GPU RTX 2070 de 8 GB, garantizando tiempos de respuesta interactivos.

Resultados

Se evaluaron cuatro modelos de clasificación —Árbol de Decisión, OneR, PART y PRISM— sobre un conjunto de prueba para comparar su rendimiento en la evaluación crediticia.

Desempeño de los modelos

Árbol de decisión

El modelo basado en el árbol de decisión alcanzó una precisión de 0,9047 (90,47 %). El informe de clasificación revela que, para la clase «inseguro», se obtuvo una puntuación f1 de 0,81, basada en 306 negativos verdaderos y 124 positivos falsos. Por el contrario, la clase «segura» mostró una puntuación f1 de 0,94, derivada de 1051 verdaderos positivos y 19 falsos negativos. Estos resultados corroboran el equilibrio del modelo propuesto entre transparencia y capacidad de discriminación.

Modelo OneR

El modelo OneR mostró una precisión de 0,7133 (71,33 %). Sin embargo, su rendimiento fue asimétrico: mientras que la clase «segura» alcanzó una precisión de 0,71, una recuperación de 1,00 y una puntuación f1 de 0,83, la clase «insegura» tuvo una precisión, una recuperación y una puntuación f1 nulas (0,00). Esta limitación en la detección de la clase de alto riesgo compromete su aplicabilidad en escenarios críticos.

Modelo PART

El enfoque PART arrojó una precisión de 0,902 (90,20 %). Para la clase «insegura», se observó una precisión perfecta (1,00), aunque con una recuperación moderada de 0,66, lo que dio como resultado una puntuación f1 de 0,79. Por su parte, la clase «segura» obtuvo una precisión de 0,88, una recuperación de 1,00 y una puntuación f1 de 0,94, lo que indica una clasificación equilibrada de ambas categorías.

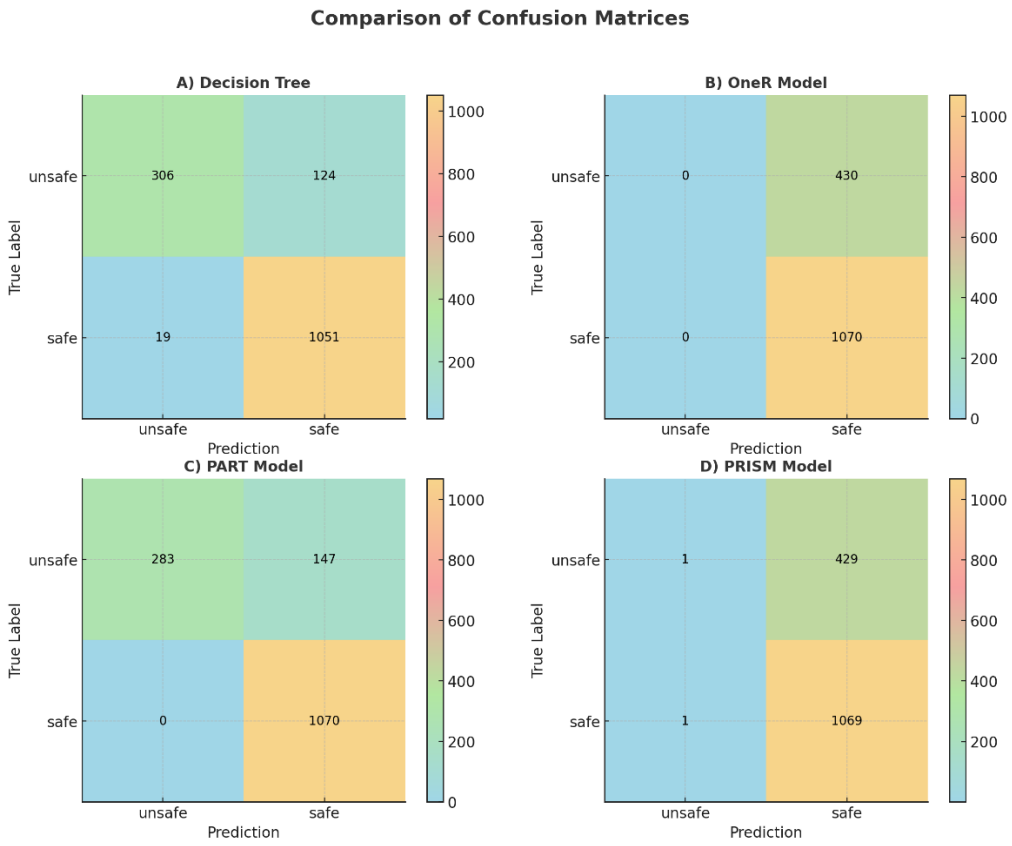
Modelo PRISM

Por último, el modelo PRISM registró una precisión de 0,713 (71,30 %). Su rendimiento fue deficiente para la clase «inseguro», con una precisión de 0,50 y un recall y una puntuación f1 de cero (0,00). Para la clase «segura», los resultados fueron similares a los de OneR, con una precisión de 0,71, un recall de 1,00 y una puntuación f1 de 0,83.

En resumen, los modelos basados en el árbol de decisión y en PART destacan por ofrecer un equilibrio adecuado entre precisión e interpretabilidad, mientras que OneR y PRISM presentan limitaciones notables, especialmente a la hora de identificar la clase «insegura». Estos resultados subrayan la importancia de seleccionar modelos que, además de ser precisos, permitan comprender claramente las reglas de decisión en las aplicaciones de evaluación crediticia.

Figura 2

Matriz de confusión combinada de los modelos evaluados



Los resultados que se muestran en la figura 2 ponen de relieve diferencias notables en el rendimiento de los modelos evaluados. En concreto, el modelo de árbol de decisión (figura 2A) muestra una gran precisión en el conjunto de pruebas, con una precisión del 90,47 %, lo que se traduce en una gran capacidad para discriminar entre las clases «segura» e «insegura». El informe de clasificación destaca la detección equilibrada de ambas clases, a pesar de ciertos errores de clasificación reflejados en la matriz de confusión combinada.

Por otro lado, el modelo PART (Figura 2C) también muestra un rendimiento similar en términos de precisión (90,20 %), pero con una particularidad en el manejo de la clase «insegura», donde se evidencia una precisión perfecta (1,00), aunque con un recuerdo moderado (0,66), lo que sugiere que, si bien las predicciones positivas son correctas, pueden omitirse instancias relevantes de esa clase.

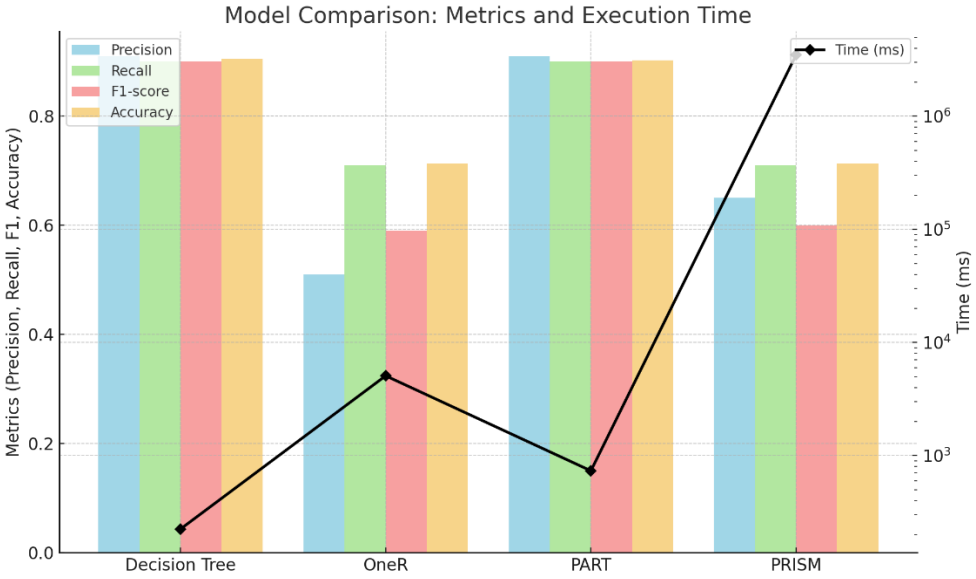
Por el contrario, tanto OneR (Figura 2B) como PRISM (Figura 2D) alcanzaron precisiones en torno al 71 %, lo que indica un rendimiento inferior en comparación con los otros dos modelos. Es importante señalar que, en ambos casos, la clasificación de la clase «inseguro» es especialmente deficiente, como lo demuestra la ausencia de recuperación y, en OneR,

incluso de precisión, lo que limita su utilidad en escenarios en los que es indispensable identificar instancias críticas.

La visualización consolidada de la matriz de confusión permite comparar de forma directa y clara los errores cometidos por cada modelo, lo que facilita la identificación de patrones de confusión específicos y la comprensión general de sus comportamientos. En última instancia, estos resultados sugieren que, para aplicaciones en las que es crucial la clasificación correcta de ambas clases, los modelos basados en árboles de decisión y PART son las opciones más sólidas, mientras que OneR y PRISM pueden requerir ajustes o descartarse para tareas críticas.

Figura 3

Matriz de confusión combinada de los modelos evaluados



En la figura 3 se muestra el rendimiento de cada modelo en términos de los principales parámetros de evaluación y tiempo de ejecución. Cabe destacar que los modelos Decision Tree y PART obtuvieron los mejores parámetros, mientras que PRISM presentó el tiempo de ejecución más largo.

Comparación y discusión

Los resultados muestran que tanto el modelo de árbol de decisión como el modelo PART alcanzan una precisión cercana al 90 % y presentan un rendimiento excelente en la clasificación de la clase «segura» (puntuación f1 de 0,94). En particular, el árbol de decisión destaca por ofrecer un mejor equilibrio en la detección de la clase «no segura» (puntuación f1 de 0,81), un aspecto fundamental en la evaluación del riesgo crediticio, mientras que el modelo PART, a pesar de mantener una precisión perfecta para esa clase, tiene una recuperación ligeramente inferior (0,66).

Por el contrario, el modelo OneR, aunque sencillo, demostró un rendimiento insuficiente al clasificar todas las instancias como «seguras», lo que dio lugar a un recall nulo para la clase «insegura» y una precisión del 71,33 %. Del mismo modo, el modelo PRISM mostró una precisión del 71,30 % y una clasificación ineficaz para la clase «insegura» (puntuación f1 de 0,00), tal y como se refleja en su matriz de confusión, ya que no logró predecir correctamente ninguna instancia de esa clase. Además, a pesar de lograr una recuperación perfecta para la clase «segura» (1,00), su precisión relativamente baja (0,71) sugiere una alta proporción de falsos positivos, lo que compromete su utilidad en la evaluación crediticia.

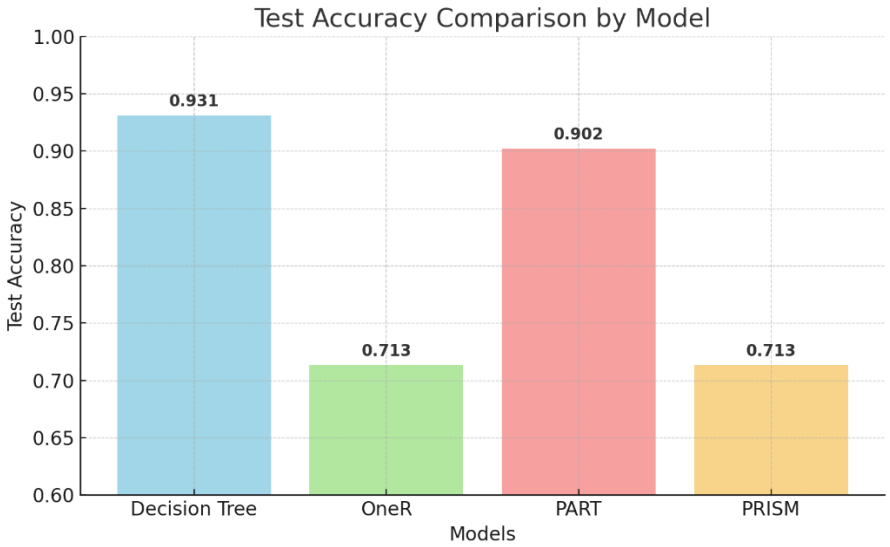
Estos resultados indican que, aunque los modelos basados en reglas como OneR y PRISM son interpretables y eficientes en determinados escenarios, su rendimiento en este contexto es insuficiente debido a su incapacidad para identificar correctamente a los clientes de crédito de alto riesgo. Por el contrario, Decision Tree y PART ofrecen un equilibrio adecuado entre precisión e interpretabilidad, lo que los convierte en las mejores opciones para esta tarea.

Comparación gráfica

Para complementar la evaluación, el siguiente gráfico de barras compara los índices de rendimiento (precisión, exactitud, recuperación y puntuación f1) de los cuatro modelos:

Figura 4

Comparación de índices de rendimiento entre modelos: árbol de decisión, OneR, PART y PRISM



La figura 4 ilustra la comparación de la precisión obtenida en el conjunto de pruebas para cuatro modelos de clasificación: árbol de decisión, OneR, PART y PRISM. Se puede observar que el árbol de decisión alcanza la mayor precisión (0,931), seguido de PART (0,902), mientras que OneR y PRISM tienen valores considerablemente más bajos (ambos en torno a 0,713). Estos resultados demuestran la superioridad de los modelos basados en árboles en cuanto a capacidad predictiva e interpretabilidad, un aspecto clave en la evaluación del riesgo crediticio.

Tabla 2

Comparación de métricas y tiempos de ejecución para cada técnica.

Technique	Accuracy	Precision	Recall	F1-score	Time (ms)
Decision Tree	**0.9047**	**0.91**	**0.90**	**0.90**	**224**
OneR	0.7133	0.51	0.71	0.59	5,069
PART	0.9020	**0.91**	**0.90**	**0.90**	733
PRISM	0.7133	0.65	0.71	0.60	3,480,000

La tabla 2 presenta valores comparativos de métricas y tiempos de ejecución para cada técnica, resaltando en negrita los valores más altos de cada categoría. Se puede observar que los modelos Decision Tree y PART ofrecen un rendimiento superior en precisión, recuperación y puntuación F1, mientras que PRISM tiene el tiempo de procesamiento más largo.

Visualización del Árbol de Decisión

Dado que el modelo del árbol de decisión destaca por su rendimiento equilibrado y su alta interpretabilidad, a continuación, se incluye su diagrama, en el que se muestran visualmente los nodos, las ramas y las hojas utilizados en la toma de decisiones (Figura 5).

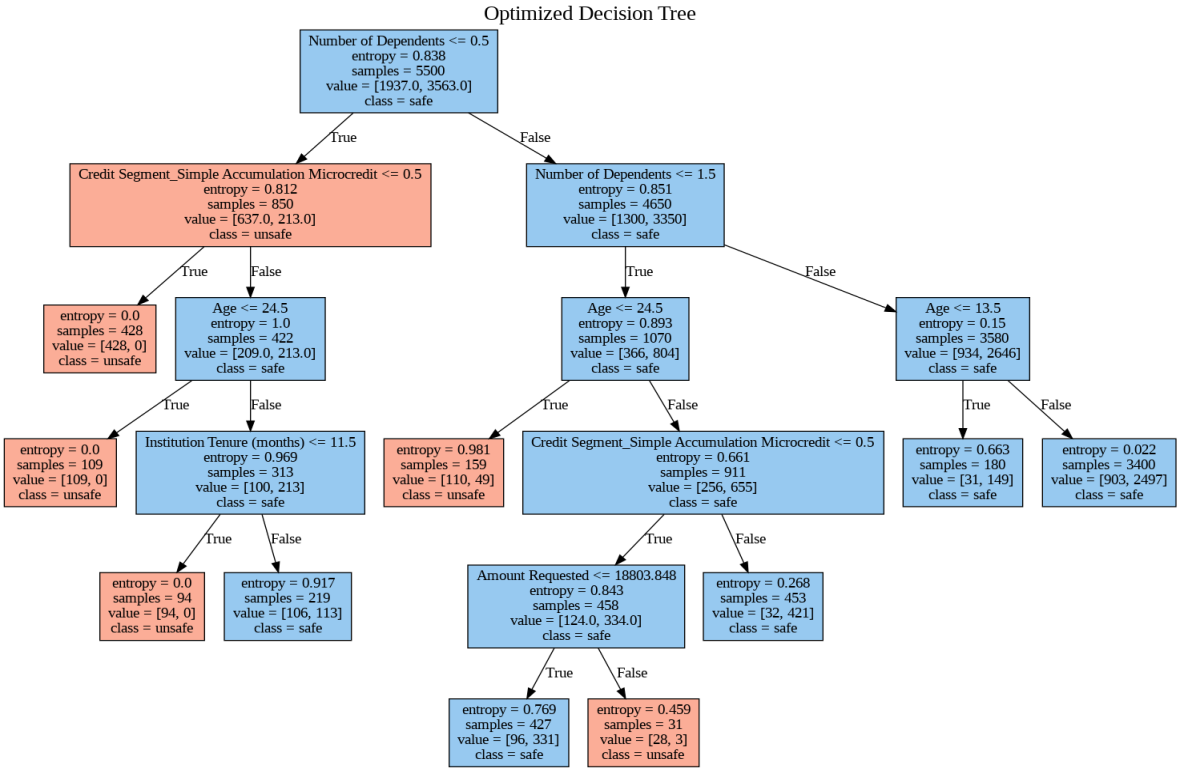
Las puntuaciones de importancia basadas en el índice Gini destacan el número de personas a cargo (57,9 % de la importancia total), seguido del término de interacción «segmento de microcréditos × importe del préstamo» (17,1 %) y la edad del prestatario (15,4 %). Las reglas representativas incluyen:

$$Dependientes \leq 0{,}5 \rightarrow inseguro,$$
$$Dependientes > 0{,}5 \wedge Edad > 24{,}5 \rightarrow seguro.$$

El árbol final contiene 18 hojas (profundidad = 5), lo que facilita la auditoría manual y las comprobaciones de cumplimiento.

Figura 5

Diagrama de árbol de decisión, resaltando las reglas de clasificación



Integración con el Asistente de Crédito

La implementación final se llevó a cabo en Python, aprovechando bibliotecas como *Gradio* para desarrollar una interfaz interactiva. El sistema, implementado en contenedores Docker en la nube, permite a los usuarios introducir variables clave (por ejemplo, edad, número de personas a su cargo, antigüedad laboral y cantidad solicitada) y recibir, en tiempo real, una predicción del riesgo crediticio junto con recomendaciones detalladas generadas por el LLM. La Figura 6 muestra la interfaz del Asistente de Crédito basado en IA, que demuestra la integración de los modelos y la explicación proporcionada al usuario.

Figura 6

Diagrama de árbol de decisión, resaltando las reglas de clasificación

Discusión

Esta investigación demuestra que el uso de un modelo de árbol de decisión es muy eficaz para la evaluación del riesgo crediticio en cooperativas, principalmente debido a su capacidad para generar reglas de decisión claras y transparentes. Esta interpretabilidad no solo permite alcanzar altos niveles de precisión —cercanos al 90 %— sino que también facilita la auditoría y la validación de las decisiones ante los organismos reguladores, lo cual es esencial en el ámbito financiero.

La elección del árbol de decisión se basa en su simplicidad y en la facilidad para extraer explicaciones comprensibles, aspectos que contrastan favorablemente con modelos más complejos como XGBoost, Random Forest y redes neuronales. Aunque estos métodos alternativos podrían ofrecer mejoras marginales en la precisión (del orden del 2-3 %), su complejidad inherente y la dificultad para interpretar los resultados hacen que el árbol de decisión sea preferible en contextos en los que se prioriza la transparencia.

Además, la integración de un chatbot basado en Phi4 complementa el sistema al proporcionar una interfaz interactiva que explica las decisiones crediticias de forma clara y comprensible para el usuario. Este asistente, entrenado con un conjunto de datos especializado y optimizado mediante el aprendizaje por transferencia, puede responder con precisión y contextualidad a las consultas sobre el estado de una solicitud o la justificación de una clasificación específica. Además, su implementación en un entorno local seguro refuerza la protección de la información confidencial, un aspecto crítico en el sector financiero.

Sin embargo, hay que reconocer algunas limitaciones del enfoque propuesto. La calidad y representatividad del conjunto de datos utilizado influyen de manera decisiva en la capacidad de generalización del modelo. Si los datos no reflejan adecuadamente la realidad del mercado crediticio, existe el riesgo de que las predicciones sean sesgadas. Del mismo modo, la implementación del chatbot aún podría beneficiarse de técnicas avanzadas de procesamiento del lenguaje natural (NLP) para mejorar su capacidad explicativa y adaptarse a una gama más amplia de consultas.

Para futuras investigaciones, es pertinente explorar la integración de modelos híbridos que combinen los enfoques tradicionales de aprendizaje automático con redes neuronales, lo que

podría aumentar la precisión y la adaptabilidad del sistema en tiempo real. Además, la incorporación de fuentes de datos alternativas, como los historiales de pago de servicios públicos o las interacciones en las redes sociales, podría enriquecer el modelo, permitiendo una evaluación más completa del riesgo crediticio, especialmente para los clientes sin un historial financiero tradicional.

Esta reestructuración del enfoque de evaluación crediticia, basada en técnicas de inteligencia artificial explicables y en la implementación de interfaces interactivas, representa una contribución significativa a la mejora de la inclusión financiera y a la optimización de los procesos de toma de decisiones en las cooperativas de ahorro y crédito.

Conclusiones

Este estudio demuestra que la integración de un modelo de árbol de decisión explicable con un chatbot basado en Phi4 representa un enfoque innovador y eficaz para la evaluación del riesgo crediticio en las cooperativas de ahorro y crédito. Con una precisión superior al 90 %, el sistema no solo permite una clasificación fiable de los clientes, sino que también ofrece explicaciones claras y comprensibles de las decisiones tomadas, lo que mejora la confianza y la transparencia del proceso.

Desde un punto de vista práctico, la implementación de este sistema reduce la dependencia de consultas externas, lo que minimiza los impactos negativos en las calificaciones crediticias de los clientes y optimiza el uso de los recursos financieros de la cooperativa. La interpretabilidad del modelo facilita la validación de las decisiones ante las entidades reguladoras, lo que refuerza la credibilidad del sistema.

Se recomienda seguir investigando en las siguientes líneas:

- Optimizar el chatbot utilizando técnicas avanzadas de procesamiento del lenguaje natural (NLP) para mejorar la capacidad de respuesta y la personalización.
- Ampliar el conjunto de datos incorporando fuentes alternativas y de comportamiento para aumentar la solidez del modelo.

- Explorar modelos híbridos que combinen técnicas tradicionales de aprendizaje automático con redes neuronales para mejorar la precisión y la adaptabilidad en tiempo real.
- Realizar estudios comparativos más exhaustivos entre modelos de alta complejidad y métodos explicables para identificar el equilibrio óptimo entre precisión y transparencia.

Referencias bibliográficas

Abdin, M., Aneja, J., & Behl, H., et al. (2024). *Phi-4 Technical Report: A 14-billion parameter language model developed with a training recipe focused on data quality*. arXiv preprint arXiv:2412.08905. https://doi.org/10.48550/arXiv.2412.08905

Al Khaldy, M., et al. (2024). Securing digital finance: Applying machine learning for fraud analysis. *In 2nd International Conference on Cyber Resilience (ICCR)*. https://doi.org/10.1109/ICCR61006.2024.10533140

Al-Ameer, A., & Al-Sunni, F. (2021). A methodology for securities and cryptocurrency trading using exploratory data analysis and artificial intelligence. *In 2021 1st International Conference on Artificial Intelligence and Data Analytics (CAIDA)* (pp. 54–61). https://doi.org/10.1109/CAIDA51941.2021.9425223

Anusha, H., & Bhowmik, B. (2023). Feature selection for peer-to-peer lending default risk using Boruta and mRMR approach. *In 2023 IEEE 20th India Council International Conference (INDICON)* (pp. 983–988). https://doi.org/10.1109/INDICON59947.2023.10440917

Chaudhary, D., & Saroj, S. K. (2023). Cryptocurrency price prediction using supervised machine learning algorithms. *Advances in Distributed Computing and Artificial Intelligence Journal,* 12, 31490. https://doi.org/10.14201/adcaij.31490

Chen, H.-Y., et al. (2024). Skewness-aware boosting regression trees for customer contribution prediction in financial precision marketing. *In WWW 2024 Companion – Proceedings of the ACM Web Conference* (pp. 461–470). https://doi.org/10.1145/3589335.3648346

Edo, O. C., et al. (2023). Fintech adoption dynamics in a pandemic: An experience from some financial institutions in Nigeria during COVID-19 using machine learning approach. *Cogent Business and Management*, 10, 2242985. https://doi.org/10.1080/23311975.2023.2242985

IC. (2023). *6th International Conference on Innovative Computing*. Lecture Notes in Electrical Engineering, 1044, 90–98.

IC. (2023). *6th International Conference on Innovative Computing*. Lecture Notes in Electrical Engineering, 1045, 100–110.

ICACNIS. (2022). *Blockchain technology, intelligent systems, and the applications for human life*. In Proceedings of ICACNIS 2022 – International Conference on Advanced Creative Networks and Intelligent Systems (pp. 50–60).

ICETSIS. (2024). *2024 ASU International Conference in Emerging Technologies for Sustainability and Intelligent Systems*. In Proceedings of ICETSIS 2024 (pp. 10–15).

Karn, A. L., et al. (2022). Designing a deep learning-based financial decision support system for fintech to support corporate customer's credit extension. *Malaysian Journal of Computer Science*, 1(Special Issue 1), 116–131. https://doi.org/10.22452/mjcs.sp2022no1.9

Ke, L., et al. (2021). Loan repayment behavior prediction of provident fund users using a stacking-based model. *In 2021 IEEE 6th International Conference on Cloud Computing and Big Data Analytics (ICCCBDA)* (pp. 37–43). https://doi.org/10.1109/ICCCBDA51879.2021.9442613

Khan, W., et al. (2022). Stock market prediction using machine learning classifiers and social media, news. *Journal of Ambient Intelligence and Humanized Computing*, 13, 3433–3456. https://doi.org/10.1007/s12652-020-01839-w

Khenfouci, Y., & Challal, Y. (2023). SuperChain: Decentralized and trustful supervised learning over blockchain. *In 2023 5th International Conference on Blockchain Computing and Applications (BCCA)* (pp. 627–634). https://doi.org/10.1109/BCCA58897.2023.10338875

Li, L. (2023). Investigating the application of 3D avatar chatbot services on fintech. *In Proceedings of the 29th Annual Americas Conference on Information Systems (AMCIS)* (pp. 45–52).

Liu, J., et al. (2023). Interpreting the prediction results of the tree-based gradient boosting models for financial distress prediction with an explainable machine learning approach. *Journal of Forecasting*, 42, 1112–1137. https://doi.org/10.1002/for.2931

Liu, Y. (2024). Discussion on the enterprise financial risk management framework based on AI fintech. *Decision Making: Applications in Management and Engineering*, 7, 254–269. https://doi.org/10.31181/dmame712024942

Liu, Y., et al. (2021). A deep neural network based model for stock market prediction. *In 2021 IEEE 2nd International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering (ICBAIE)* (pp. 320–323). https://doi.org/10.1109/ICBAIE52039.2021.9390010

Lusinga, M., et al. (2021). Investigating statistical and machine learning techniques to improve the credit approval process in developing countries. *In IEEE AFRICON Conference* (pp. 200–207). https://doi.org/10.1109/AFRICON51333.2021.9570906

Majumder, A., et al. (2022). Stock market prediction: A time series analysis. *Smart Innovation, Systems and Technologies*, 235, 389–401. https://doi.org/10.1007/978-981-16-2877-1_35

Malakauskas, A., & Lakstutiene, A. (2021). The application of artificial intelligence tools in creditworthiness modelling for SME entities. *In IEEE International Conference on Technology and Entrepreneurship* (ICTE 2021) (pp. 120–126). https://doi.org/10.1109/ICTE51655.2021.9584528

Paul, M., et al. (2024). Trustworthy deep learning techniques for credit risk assessment. *In 5th International Conference on Electronics and Sustainable Communication Systems (ICESC)* (pp. 1949–1962). https://doi.org/10.1109/ICESC60852.2024.10689741

Rabbani, H., et al. (2024). Enhancing security in financial transactions: A novel blockchain-based federated learning framework for detecting counterfeit data in fintech. *PeerJ Computer Science*, 10, e2280. https://doi.org/10.7717/peerj-cs.2280

Shetu, S. F., et al. (2021). Predicting satisfaction of online banking system in Bangladesh by machine learning. *In ICAICST 2021 – International Conference on Artificial Intelligence and Computer Science Technology* (pp. 223–228). https://doi.org/10.1109/ICAICST53116.2021.9497796

Shen, Y. (2022). Application of supplemental sampling and interpretable AI in credit scoring for Canadian fintechs: Methods and case studies. *In Lecture Notes in Computer Science*, 13725 (pp. 3–14). https://doi.org/10.1007/978-3-031-22064-7_1

Silva, C., et al. (2023). Towards interpretability in fintech applications via knowledge augmentation. *In Lecture Notes in Computer Science (LNCS)*, 14115 (pp. 106–117). https://doi.org/10.1007/978-3-031-49008-8_9

Tahyudin, I., et al. (2023). Sentiment analysis model development on E-Money service complaints. *TEM Journal*, 12, 2050–2055. https://doi.org/10.18421/TEM124-15

Xia, Y., et al. (2021). Incorporating multilevel macroeconomic variables into credit scoring for online consumer lending. *Electronic Commerce Research and Applications*, 49, 101095. https://doi.org/10.1016/j.elerap.2021.101095

Zhao, X., & Zhao, Q. (2021). Stock prediction using optimized LightGBM based on cost awareness. *In 2021 5th IEEE International Conference on Cybernetics (CYBCONF)* (pp. 107–113). https://doi.org/10.1109/CYBCONF51991.2021.9464148

Zhu, Y., & Tingting, N. (2024). Application of back propagation neural network method in digital economy prediction. *In 3rd IEEE International Conference on Distributed Computing and Electrical Circuits and Electronics (ICDCECE)* (pp. 1–8). https://doi.org/10.1109/ICDCECE60827.2024.10548494

Conflicto de intereses:

Los autores declaran que no existe conflicto de interés posible.

Financiamiento:

No existió asistencia financiera de partes externas al presente artículo.

Nota:

El artículo no es producto de una publicación anterior.