



Maestría en

CIENCIA DE DATOS Y MÁQUINAS DE APRENDIZAJE CON MENCIÓN EN INTELIGENCIA ARTIFICIAL

Trabajo previo a la obtención de título de Magister en Ciencia de Datos y Máquinas de Aprendizaje

AUTORES:

Alvaro Gabriel Benítez Bravo
Jairo Fernando Parra Andrade
Stalyn Alexander Nogales Shagñay
Geovanny Rodolfo Baños Noriega

TUTORES:

Andrés Roberto Navas Castellanos
Gabriela Estefania Chiliquinga Jiménez

TEMA

Predicción del incumplimiento de pagos en cooperativas de
ahorro y crédito mediante modelos explicables.

Certificación de autoría

Nosotros, **Stalyn Alexander Nogales Shagñay, Jairo Fernando Parra Andrade, Geovanny Rodolfo Baños Noriega, Álvaro Gabriel Benítez Bravo**, declaramos bajo juramento que el trabajo aquí descrito es de nuestra autoría; que no ha sido presentado anteriormente para ningún grado o calificación profesional y que se ha consultado la bibliografía detallada.

Cedemos nuestros derechos de propiedad intelectual a la Universidad Internacional del Ecuador (UIDE), para que sea publicado y divulgado en internet, según lo establecido en la Ley de Propiedad Intelectual, su reglamento y demás disposiciones legales.



Firma del graduando
Stalyn Alexander Nogales Shagñay



Firma del graduando
Jairo Fernando Parra Andrade



Firma del graduando
Geovanny Rodolfo Baños Noriega



Firma del graduando
Alvaro Gabriel Benitez Bravo

Autorización de Derechos de Propiedad Intelectual

Nosotros **Stalyn Alexander Nogales Shagñay, Jairo Fernando Parra Andrade, Geovanny Rodolfo Baños Noriega, Alvaro Gabriel Benitez Bravo**, en calidad de autores del trabajo de investigación titulado ***Predicción del incumplimiento de pagos en cooperativas de ahorro y crédito mediante modelos explicables***, autorizamos a la Universidad Internacional del Ecuador (UIDE) para hacer uso de todos los contenidos que nos pertenecen o de parte de los que contiene esta obra, con fines estrictamente académicos o de investigación. Los derechos que, como autores nos corresponden, lo establecido en los artículos 5, 6, 8, 19 y demás pertinentes de la Ley de Propiedad Intelectual y su Reglamento en Ecuador.

D. M. Quito, junio 2026



Validar únicamente en FirmaEC.
Firmado electrónicamente por:
**STALYN ALEXANDER
NOGALES SHAGÑAY**

**Firma del graduando
Stalyn Alexander Nogales Shagñay**



Validar únicamente en FirmaEC.
Firmado electrónicamente por:
**JAIRO FERNANDO
PARRA ANDRADE**

**Firma del graduando
Jairo Fernando Parra Andrade**



Validar únicamente en FirmaEC.
Firmado electrónicamente por:
**GEOVANNY RODOLFO
BAÑOS NORIEGA**

**Firma del graduando
Geovanny Rodolfo Baños Noriega**

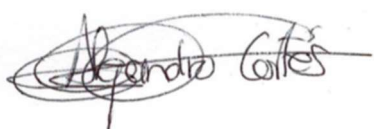


Validar únicamente en FirmaEC.
Firmado electrónicamente por:
**ALVARO GABRIEL
BENITEZ BRAVO**

**Firma del graduando
Alvaro Gabriel Benitez Bravo**

Aprobación de dirección y coordinación del programa

Nosotros, **Director Alejandro Cortés López de EIG y Coordinadora Karla Estefanía Mora Cajas de UIDE**, declaramos que: **Stalyn Alexander Nogales Shagñay, Jairo Fernando Parra Andrade, Geovanny Rodolfo Baños Noriega, Alvaro Gabriel Benitez Bravo**, son los autores exclusivos de la presente investigación y que ésta es original, auténtica y personal de ellos.



Alejandro Cortés López
Director/a de la
Maestría en Ciencia de datos y máquinas
de aprendizaje con mención en Inteligencia
Artificial



Karla Estefanía Mora Cajas
Coordinador/a de la
Maestría en Ciencia de datos y máquinas
de aprendizaje con mención en Inteligencia
Artificial

DEDICATORIA

Dedico el presente trabajo de investigación a mi familia, por su apoyo incondicional, comprensión y paciencia a lo largo de este proceso académico, Motivación y confianza nos impulsaron a continuar y culminar nuestros objetivos profesionales.

Del mismo modo dedicamos esta tesis a todas las personas que impulsan el poder del conocimiento y la educación como la herramienta más amplia de transformación social pues este trabajo representa el esfuerzo el compromiso y la constancia para mejorar de manera continua y desarrollarnos de manera profesional.

Finalmente, esta investigación va dedicada a quiénes desde las áreas académicas y profesionales promueven con responsabilidad el uso de la ciencia de datos y de la inteligencia artificial para la toma de decisiones con carácter ético transparente y coherente promoviendo la contribución del bienestar y la sostenibilidad de las instituciones financieras dentro del sector cooperativista.

AGRADECIMIENTOS

En primera instancia expresamos nuestra profunda gratitud a Dios, por acompañarnos y guiarnos durante todo este proceso de formación académica. Su fortaleza sabiduría y providencia nos permitieron afrontar cada desafío con perseverancia, responsabilidad y compromiso, haciendo posible la culminación de esta investigación.

Extendemos nuestro sincero agradecimiento a la universidad y a los docentes de maestría quienes con su experiencia dedicación y calidad humana contribuyeron significativamente a nuestra formación académica y profesional, sus enseñanzas orientaciones y aportes fortalecieron nuestras capacidades investigativas y nos motivaron a desarrollar este trabajo con rigor científico, ética y responsabilidad. De manera especial, expresamos nuestro reconocimiento al tutor de tesis por su valioso acompañamiento y orientación permanente sus observaciones oportunas durante esta etapa sus recomendaciones y los conocimientos transmitidos fueron fundamentales para consolidar este trabajo.

Asimismo, agradecemos a la cooperativa de ahorro y crédito por la apertura institucional, la colaboración brindada y el acceso a la información necesaria para el desarrollo de esta investigación, su disposición fue determinante para alcanzar los objetivos y generar resultados que aporten al fortalecimiento de la gestión del riesgo crediticio.

Finalmente, expresamos nuestro agradecimiento a todas las personas e instituciones que de manera directa o indirecta, contribuyeron con su apoyo conocimientos colaboración y motivación durante el desarrollo de este trabajo, cada aporte recibido presentó un valioso incentivo para culminar esta investigación con responsabilidad con el firme propósito de generar conocimiento útil en el ámbito académico y el sector cooperativo.

RESUMEN

La presente investigación tuvo como propósito desarrollar y validar un modelo predictivo explicable para estimar el incumplimiento de pagos en una cooperativa de ahorro y crédito ecuatoriana, con el objetivo de fortalecer la gestión del riesgo crediticio mediante la aplicación de técnicas de aprendizaje automático e inteligencia artificial explicable.

Se adoptó un enfoque cuantitativo, de carácter aplicado, con alcance predictivo y diseño no experimental, utilizando una base de datos consolidada compuesta por 75.783 operaciones crediticias correspondientes a 46.249 socios, en la cual se identificó una tasa de mora igual o superior a treinta días del 23,47 %. El procesamiento de la información comprendió etapas de depuración, transformación e ingeniería de características, incorporando variables demográficas, financieras, crediticias e históricas relacionadas con el perfil del socio, las condiciones del crédito, la capacidad de pago, el nivel de endeudamiento y el comportamiento crediticio, posteriormente, se implementaron y compararon los algoritmos de Regresión Logística, Random Forest y XGBoost, cuyo desempeño fue evaluado mediante las métricas de exactitud (accuracy), precisión (precision), sensibilidad (recall), F1-score y área bajo la curva ROC (AUC-ROC).

Resultados obtenidos evidenciaron que el modelo XGBoost presentó el mejor desempeño predictivo entre algunos algoritmos evaluados; este modelo alcanzó una exactitud del 84,30%, una precisión de 73,87%, un F1-score de 0,8797, indicadores que reflejan una mayor capacidad para diferenciar las operaciones con riesgo de incumplimiento de aquellas que presentan un comportamiento crediticio favorable, superando el rendimiento de los demás modelos analizados.

Adicionalmente la aplicación de las técnicas SHAP y LIME permitió interpretar de manera transparente el funcionamiento del modelo, identificando las variables con mayor incidencia en las predicciones; entre los factores más relevantes se encontraron la antigüedad del crédito, la condición laboral del socio, el historial de riesgo severo, las alertas crediticias previas, la existencia de préstamos vencidos, el valor de la cuota y la relación entre la cuota y los ingresos.

Los hallazgos del estudio permiten concluir que la integración de los modelos de aprendizaje automático con técnicas de inteligencia artificial explicable constituye una herramienta metodológica estratégica y robusta para fortalecer la evaluación del riesgo crediticio en las cooperativas de ahorro y crédito.

De este modo, la combinación de un alto desempeño predictivo con mecanismos de interpretabilidad no solo mejora la estimación de la probabilidad de incumplimiento de las obligaciones financieras, sino que también facilita la comprensión de los factores que sustentan cada predicción; esta capacidad de explicar el proceso de decisión incrementa la transparencia, la trazabilidad y la confianza en los resultados obtenidos, favoreciendo una gestión del riesgo crediticio más objetiva, confiable y fundamentada.

Palabras clave: incumplimiento de pagos, riesgo crediticio, cooperativas de ahorro y crédito, aprendizaje automático, XGBoost, SHAP, LIME, inteligencia artificial explicable.

ABSTRACT

The purpose of this study was to develop and validate an explainable predictive model to estimate loan default in an Ecuadorian savings and credit cooperative, with the aim of strengthening credit risk management through the application of machine learning and explainable artificial intelligence techniques.

A quantitative, applied research approach was adopted, with a predictive scope and a non-experimental design. The study used a consolidated dataset comprising 75,783 loan transactions corresponding to 46,249 cooperative members, in which a delinquency rate of 23.47% for loans overdue by 30 days or more was identified. Data processing included data cleaning, transformation, and feature engineering, incorporating demographic, financial, credit, and historical variables related to members' profiles, loan characteristics, repayment capacity, indebtedness level, and credit behavior. Subsequently, Logistic Regression, Random Forest, and XGBoost algorithms were implemented and compared. Their performance was evaluated using accuracy, precision, recall, F1-score, and the area under the receiver operating characteristic curve (AUC–ROC).

The results demonstrated that XGBoost achieved the best predictive performance among the evaluated algorithms. The model attained an accuracy of 84.30%, a precision of 73.87%, an F1-score of 59.93%, and an AUC–ROC of 0.8797, indicating a superior ability to distinguish between loan transactions with a high probability of default and those with favorable repayment behavior, outperforming the other models analyzed.

Furthermore, the application of SHAP (SHapley Additive Explanations) and LIME (Local Interpretable Model-agnostic Explanations) enabled a transparent interpretation of the

model's predictions by identifying the variables with the greatest influence on the decision-making process. The most relevant factors included loan age, member employment status, history of severe credit risk, previous credit alerts, outstanding overdue loans, installment amount, and the installment-to-income ratio.

The findings indicate that integrating machine learning models with explainable artificial intelligence techniques provides a robust and strategic methodological framework for enhancing credit risk assessment in savings and credit cooperatives. Moreover, combining high predictive performance with interpretability mechanisms not only improves the estimation of default probability but also facilitates a deeper understanding of the factors underlying each prediction. This capability enhances transparency, traceability, and confidence in the analytical process, thereby supporting more objective, reliable, and evidence-based credit risk management.

Keywords: payment default, credit risk, savings and credit cooperatives, machine learning, XGBoost, SHAP, LIME, explainable artificial intelligence.

TABLA DE CONTENIDOS (Índice)

CAPITULO 1	18
1 INTRODUCCIÓN	18
1.1. Definición del Proyecto.....	18
1.2 Formulación del problema	25
1.2 Objetivos	26
1.3 Justificación e importancia del trabajo de investigación	27
1.4 Perfil de la organización: Nombre, actividades, mercados servidos y principales cifras.....	28
CAPÍTULO 2	37
2. Revisión de literatura	37
2.1. Introducción.....	37
2.2. Conceptos fundamentales.....	38
2.3. Estado del arte	47
2.4. Marco teórico	51
CAPÍTULO 3	57
3.1. Enfoque, tipo y diseño de la investigación.....	57
3.2. Contexto, población, muestra y unidad de análisis	58
3.3. Variable dependiente y variables independientes	59

3.4. Proceso de extracción, transformación y carga de datos.....	61
3.5. Construcción del historial crediticio y prevención de fuga de información	62
3.6. Análisis exploratorio de datos	63
3.7. División de datos y estrategia de validación	64
3.8. Tratamiento del desbalance de clases.....	66
3.9. Modelos de aprendizaje automático	67
3.10. Evaluación, calibración y selección del modelo	68
3.11. Explicabilidad de los modelos.....	69
3.12. Validación de negocio, ética y reproducibilidad	69
CAPÍTULO 4.....	72
4. Análisis de resultados	72
4.1. Caracterización del conjunto de datos.....	72
4.2. Análisis exploratorio de la mora crediticia.....	74
4.3. Resultados de los modelos de clasificación	78
4.4. Análisis de matrices de confusión	79
4.5. Selección de XGBoost como modelo final	80
4.6. Explicabilidad del modelo XGBoost.....	96
4.7. Validación operativa y calibración de probabilidades.....	100
4.8. Discusión de resultados.....	102

4.9. Encuesta con expertos	104
CAPÍTULO 5	109
5. Conclusiones y recomendaciones.....	109
5.1 Conclusiones	109
5.2 Recomendaciones	110
Referencias Bibliográficas	111
ANEXOS	116

LISTA DE TABLAS (Índice de tablas)

<i>Tabla 1. Ubicación de sus actividades de negocio.....</i>	<i>32</i>
<i>Tabla 2. Variables para estimación de cumplimiento</i>	<i>60</i>
<i>Tabla 3. División de datos por socio.....</i>	<i>65</i>

LISTA DE FIGURAS (Índice de figuras)

<i>Figura 1. Ficha Informativa de directivos y representante legal</i>	<i>31</i>
<i>Figura 2. Certificado registro único de contribuyentes</i>	<i>32</i>
<i>Figura 3. Indicadores Financieros</i>	<i>35</i>
<i>Figura 4. Datos - variables históricas</i>	<i>74</i>
<i>Figura 5. Matriz de correlación</i>	<i>74</i>
<i>Figura 6. Tasa de mora por producto</i>	<i>75</i>
<i>Figura 7. Tasa de mora por edad</i>	<i>76</i>
<i>Figura 8. Mora Binaria</i>	<i>77</i>
<i>Figura 9. Diferencia entre modelos</i>	<i>81</i>
<i>Figura 10. Análisis de modelos</i>	<i>82</i>
<i>Figura 11. Matriz de confusión y regresión logística</i>	<i>83</i>
<i>Figura 12. Matriz de confusión - Random Forest</i>	<i>84</i>
<i>Figura 13. Matriz de confusión - XG Boost</i>	<i>85</i>
<i>Figura 14. Curva ROC Regresión Logística</i>	<i>86</i>
<i>Figura 15. Curva ROC Random Forest</i>	<i>87</i>
<i>Figura 16. Curva ROC - XG Boost</i>	<i>88</i>
<i>Figura 17. Variables - XGBoost</i>	<i>89</i>
<i>Figura 18. Histórico de predicción</i>	<i>91</i>
<i>Figura 19. SHAP Individual</i>	<i>92</i>
<i>Figura 20. LIME individual</i>	<i>93</i>
<i>Figura 21. Probabilidad socios riesgo bajo</i>	<i>94</i>
<i>Figura 22. SHAP Individual</i>	<i>95</i>
<i>Figura 23. LIME Individual</i>	<i>95</i>
<i>Figura 24. SHAP GLOBAL</i>	<i>98</i>

<i>Figura 25. Variables Influyentes.....</i>	<i>99</i>
<i>Figura 26 Encuestas – Expertos- Cargos</i>	<i>104</i>
<i>Figura 27. Experiencia.....</i>	<i>105</i>
<i>Figura 28. Pregunta 1</i>	<i>105</i>
<i>Figura 29. Pregunta 2</i>	<i>106</i>
<i>Figura 30. Pregunta 3</i>	<i>107</i>
<i>Figura 31. Pregunta 4</i>	<i>108</i>
<i>Figura 32. Pregunta 5</i>	<i>108</i>

CAPITULO 1

1 INTRODUCCIÓN

1.1. Definición del Proyecto

La transformación digital del sistema financiero ha impulsado la adopción de tecnologías basadas en inteligencia artificial (IA) y aprendizaje automático (Machine Learning, ML) como herramientas estratégicas para fortalecer la gestión del riesgo crediticio. Tomando en consideración un entorno caracterizado por una creciente competencia además de las nuevas normativas de regulación y de la disponibilidad masiva del uso de datos, donde las instituciones financieras requieren adoptar modelos analíticos que le permitan anticipar el comportamiento de los usuarios y clientes con altos niveles de precisión. (Superintendencia de Economía Popular y Solidaria, 2024)

En particular, las cooperativas de ahorro y crédito enfrentan el desafío de mantener la sostenibilidad financiera mientras cumplen su función social de inclusión económica, lo que exige optimizar los procesos de evaluación crediticia mediante metodologías predictivas que reduzcan la incertidumbre asociada al incumplimiento de pagos. En este escenario, la ciencia de datos se consolida como un elemento clave para generar conocimiento útil que fortalezca la toma de decisiones basada en evidencia (European Banking Authority, 2020)

El riesgo de crédito constituye uno de los principales factores que comprometen la estabilidad de las instituciones financieras, debido a que el incumplimiento de las obligaciones adquiridas por los prestatarios impacta directamente sobre la liquidez, solvencia y rentabilidad institucional. Tradicionalmente, la evaluación del riesgo crediticio se ha sustentado en modelos estadísticos como la regresión logística y en metodologías de calificación basadas en indicadores

financieros; sin embargo, el incremento del volumen, variedad y complejidad de la información ha evidenciado limitaciones para capturar relaciones no lineales entre las variables que influyen en la morosidad. En consecuencia, el aprendizaje automático ha demostrado un desempeño superior en tareas de clasificación y predicción del riesgo crediticio al identificar patrones complejos imposibles de detectar mediante técnicas convencionales, incrementando significativamente la capacidad predictiva de las entidades financieras (Lessmann et al., 2015; Baesens, 2014; Bishop, 2006).

No obstante, el elevado rendimiento alcanzado por algoritmos como Random Forest, Gradient Boosting o XGBoost ha venido acompañado de un importante desafío relacionado con la interpretabilidad de sus resultados; estos modelos suelen operar como sistemas de “caja negra”, lo que dificulta interpretar los criterios que sustentan una predicción específica. Está limitada capacidad de explicación restringe su aplicación en entornos regulados, donde la transparencia, la trazabilidad y la dirección de cuentas representan principios esenciales para garantizar la confiabilidad y legitimidad de las decisiones.

Ante esta problemática, la Inteligencia Artificial Explicable (Explainable Artificial Intelligence, XAI) ha surgido como enfoque innovador que busca desarrollar modelos capaces de combinar un elevado desempeño predictivo con mecanismos de interpretación claros y comprensibles. Este paradigma permite explicar cómo y en qué medida cada variable influye en las decisiones generadas por el algoritmo, favoreciendo una mayor transparencia y comprensión de los resultados

Herramientas como SHAP (SHapley Additive Explanations) y LIME (Local Interpretable Model-Agnostic Explanations) facilitan interpretar tanto el comportamiento global como las

predicciones individuales, fortaleciendo la confianza de analistas, auditores y organismos reguladores en los modelos de inteligencia artificial (Molnar, 2024; Lundberg & Lee, 2017; Ribeiro et al., 2016).

1.3. Justificación e importancia del trabajo de investigación

En el contexto ecuatoriano, las cooperativas de ahorro y crédito constituyen uno de los pilares del sistema financiero popular y solidario, desempeñando un papel fundamental en la inclusión financiera y en el desarrollo económico territorial. Sin embargo, el incremento de la cartera vencida y la necesidad de fortalecer los mecanismos de administración del riesgo crediticio demandan soluciones tecnológicas capaces de identificar oportunamente las operaciones con alta probabilidad de incumplimiento. (Banco Central del Ecuador, 2025) (Superintendencia de Economía Popular y Solidaria, 2025)

En este contexto, la presente investigación propone el desarrollo y validación de un modelo predictivo explicable para la estimación del riesgo de incumplimiento crediticio, utilizando información histórica consolidada de operaciones de una cooperativa de ahorro y crédito ecuatoriana.

La investigación integra un proceso metodológico estructurado que comprende la preparación y depuración de datos, la ingeniería de características el entrenamiento y evaluación de modelos de aprendizaje automático, así como la incorporación de técnicas de inteligencia artificial explicable, que permiten interpretar las predicciones generadas e identificar los factores con mayor incidencia del riesgo de mora; con este enfoque se busca generar un modelo analítico que combine la capacidad predictiva, interpretabilidad, proporcionando una herramienta de apoyo que permiten fortalecer los procesos de evaluación y gestión del riesgo crediticio, sin

reemplazar el criterio técnico de los responsables de la toma de decisiones. (European Commission, 2024)

La relevancia científica radica en la contribución al campo de la inteligencia artificial aplicada a la gestión del riesgo financiero particularmente con la integración de modelos predictivos que poseen mecanismos de explicabilidad y favorece la transparencia la trazabilidad y la comprensión de las decisiones generadas por los algoritmos, permitiendo que el estudio aporte evidencia empírica sobre la aplicabilidad de estas técnicas en el contexto cooperativista del Ecuador promoviendo un desarrollo de modelos analíticos alineados con los principios de ética, gobernanza de datos confiabilidad y responsabilidad en el uso de la IA.

De esta manera, la investigación contribuye al fortalecimiento del conocimiento científico en el ámbito de la analítica financiera y proporciona una base metodológica que podrá servir de referencia para futuras investigaciones y para el diseño de estrategias institucionales orientadas a optimizar la gestión preventiva del riesgo crediticio.

Desde una perspectiva práctica, los resultados permitirán optimizar la evaluación del riesgo crediticio, fortalecer la detección temprana de operaciones susceptibles de incumplimiento y mejorar la asignación de recursos institucionales destinados a la prevención de la morosidad, además de que, la metodología propuesta constituye una referencia para futuras investigaciones orientadas a la implementación de modelos explicables en organizaciones financieras y otros sectores donde la confianza, la gobernanza de datos y la transparencia algorítmica representan factores esenciales para la transformación digital y la innovación basada en inteligencia artificial. (Organisation for Economic Co-operation and Development, 2022)

El papel estratégico que desempeñan las cooperativas de ahorro y crédito dentro del sistema financiero popular y solidario del Ecuador, al constituirse como instituciones que promueven la intermediación financiera como un enfoque orientado al desarrollo social y económico de sus asociados, dónde su actividad permite ampliar el acceso al financiamiento para personas, familias, emprendedores micro empresas que, en muchos casos, enfrentan limitaciones para acceder a los servicios ofrecidos por la banca tradicional, y dónde el sector cooperativista conlleva al fortalecimiento de la inclusión financiera impulsando el desarrollo de economías locales y apoyando la generación de actividades productivas mediante la oferta de productos crediticios adaptado de las necesidades de los diferentes segmentos de la población.

Debido a la relevancia de su función económica y social, resulta indispensable que las cooperativas implementen mecanismos deficientes de evaluación y administración del riesgo crediticio que les permita preservar la calidad de su cartera, garantizar su sostenibilidad financiera y mantener la confianza en sus socios y los organismos de control.

Sin embargo, la actividad crediticia implica una exposición permanente al riesgo de incumplimiento de pagos, el cual representa uno de los factores que mayor impacto genera sobre la liquidez, solvencia y sostenibilidad financiera de estas instituciones.

En este contexto se encuentra la Cooperativa de Ahorro y Crédito Riobamba Ltda., que es una institución financiera ecuatoriana fundada en 1978, cuya misión Está orientada a promover el desarrollo económico de sus socios mediante la oferta de servicios de ahorro, inversión y crédito. Como entidad perteneciente al segmento uno del sistema financiero popular y solidario, gestiona un volumen significativo de operaciones crediticias, lo que hace necesario fortalecer de forma permanente los procesos de evaluación, seguimiento y administración del riesgo

crediticio, haciendo de estas acciones el fundamento para la preservación de la calidad de la cartera, además de apoyar la adecuada toma de decisiones y contribuir a la estabilidad y sostenibilidad institucional.

En los últimos años, el sistema financiero ecuatoriano ha operado en un contexto marcado por diversas transformaciones económicas y sociales que han afectado la capacidad de pago de los prestatarios; entre los principales factores se encuentran la desaceleración económica, la elevada informalidad laboral, la inestabilidad de los ingresos familiares, el incremento del costo de vida y la incertidumbre financiera. Estas condiciones han favorecido el aumento de los niveles de morosidad en las instituciones financieras, lo que plantea nuevos retos para el fortalecimiento de los mecanismos de gestión y control del riesgo crediticio.

Esta realidad ha puesto de manifiesto la necesidad de fortalecer los mecanismos de evaluación utilizados durante el proceso de otorgamiento y seguimiento de los créditos con el propósito de anticipar oportunamente el riesgo de incumplimiento y preservar la calidad de la cartera.

En este escenario los modelos tradicionales de evaluación crediticia, fundamentados principalmente en las reglas de negocio, indicadores financieros y el criterio técnico de los analistas continúan siendo herramientas de amplia aplicación. Sin embargo, la creciente complejidad del comportamiento financiero de los clientes ha evidenciado que estos enfoques presentan limitaciones para identificar relaciones no lineales, interacciones complejas y patrones ocultos entre las múltiples variables que influyen el riesgo de mora; como consecuencia, resulta pertinente incorporar metodologías analíticas que permitan aprovechar el potencial de los datos históricos para mejorar la precisión de las estimaciones y fortalecer la toma de decisiones.

Como respuesta a esta necesidad, la ciencia de datos y el aprendizaje automático (machine learning) han adquirido un papel más relevante en el sector financiero al proporcionar herramientas capaces de procesar grandes volúmenes de información y desarrollar modelos predictivos con alta capacidad para identificar patrones de comportamiento asociados al riesgo crediticio.

Los algoritmos Random Forest y XGBoost, han demostrado un desempeño destacado en la clasificación de problemas al permitir estimar con mayor precisión la probabilidad de incumplimiento de las operaciones crediticias, su aplicación constituye una alternativa que complementa los procedimientos tradicionalistas de análisis y proporciona información objetiva para apoyar la gestión integral del riesgo crediticio.

No obstante, el uso de modelos predictivos de alta complejidad también plantea importantes desafíos relacionados con la transparencia y la interpretabilidad de sus resultados, en muchos casos, estos algoritmos generan predicciones como un elevado nivel de precisión pero sin ofrecer una explicación clara sobre los factores que influyen cada decisión, situación que dificulta su comprensión y validación por parte de analistas, directivos y organismos de control dentro del ámbito financiero, donde las decisiones deben sustentarse con criterios técnicos, éticos y verificables requiriendo de manera indispensable que las herramientas de inteligencia artificial proporcionan resultados que puedan ser interpretados, auditados y justificados de manera objetiva.

1.2 Formulación del problema

Problema general

¿Cómo desarrollar un modelo predictivo explicable, basado en técnicas de ciencia de datos y aprendizaje automático, que permita estimar el incumplimiento de pagos en las operaciones de crédito de la “Cooperativa de Ahorro y Crédito Riobamba Ltda.”, ¿fortaleciendo la gestión del riesgo crediticio y la toma de decisiones institucionales?

Problemas específicos

¿Cuáles son las variables crediticias y financieras que influyen de manera significativa en el incumplimiento de pagos en la “Cooperativa de Ahorro y Crédito Riobamba Ltda.”?

¿Qué modelos de aprendizaje automático presentan un mejor desempeño predictivo para identificar el incumplimiento de pagos en las operaciones de crédito?

¿Cómo pueden explicarse los resultados de los modelos predictivos para facilitar su interpretación y aplicación en la gestión del riesgo crediticio?

1.2.3 Naturaleza o tipo de proyecto

La naturaleza del proyecto es de tipo Tecnológico e investigativo con un enfoque aplicado orientado al desarrollo de un modelo predictivo explicable que permite identificar el incumplimiento de los pagos en los préstamos de la Cooperativa de Ahorro y Crédito Riobamba Ltda.

El proyecto busca a partir de un análisis de la información crediticia y financiera disponible, desarrollar un modelo predictivo utilizando técnicas de ciencia de datos con el fin de apoyar la gestión crediticia de dicha institución

1.3 Objetivos

1.3.1 Objetivo general

Desarrollar un modelo predictivo explicable que permita identificar el incumplimiento de pagos en las operaciones de crédito en la “Cooperativa de Ahorro y Crédito Riobamba Ltda.”, utilizando técnicas de ciencia de datos y aprendizaje automático, a partir del análisis de información crediticia y financiera disponible, con el fin de apoyar la gestión del riesgo crediticio y la toma de decisiones institucionales.

1.3.2 Objetivo específico

Analizar y preprocesar la base de datos de operaciones de crédito y la información financiera de la “Cooperativa de Ahorro y Crédito Riobamba Ltda.”, identificando las variables más relevantes asociadas al incumplimiento de pagos mediante técnicas de análisis exploratorio de datos.

Construir y evaluar modelos de aprendizaje automático para la predicción del incumplimiento de pagos, comparando su desempeño mediante métricas estadísticas adecuadas que permitan seleccionar el modelo más eficiente y robusto.

Aplicar técnicas de explicabilidad de modelos (XAI) para interpretar los factores que influyen en el incumplimiento de pagos, proporcionando explicaciones claras y comprensibles que faciliten la gestión del riesgo crediticio en la cooperativa de ahorro y crédito.

1.4 Justificación e importancia de la investigación

1.4.1 Justificación teórica

De una perspectiva teórica, esta investigación amplía el conocimiento existente sobre la aplicación de modelos de inteligencia artificial explicable en la evaluación del riesgo crediticio dentro del sector cooperativo, para ello, integra fundamentos de la ciencia de datos el aprendizaje automático y las técnicas de explicabilidad de modelos, proporcionando evidencia empírica que demuestra su utilidad para comprender los factores asociados al incumplimiento en los pagos en escenarios financieros reales.

1.4.2 Justificación práctica

Desde una perspectiva práctica, los resultados de esta investigación proporcionan a la cooperativa de ahorro y crédito una herramienta predictiva que contribuirá a identificar la manera oportuna las operaciones como ya probabilidad de incumplimiento, además de esto permitirá optimizar los procesos de evaluación del riesgo crediticio fortaleciendo la toma de decisiones y la mejor gestión preventiva de cartera mediante la incorporación de técnicas de explicabilidad las que facilitarán la comprensión de los resultados por parte de los analistas y directivos, al ofrecer una interpretación clara y fundamentada de los factores que influyen en cada predicción.

1.4.3 Justificación metodológica

Ámbito metodológico, la investigación plantea un enfoque cuantitativo de carácter aplicado sustentado en el análisis de datos reales provenientes de operaciones de crédito y en la implementación de modelos de aprendizaje automático con capacidades de explicabilidad, esta metodología constituye un referente para futuras investigaciones orientadas al desarrollo de

soluciones predictivas en cooperativas de ahorro y crédito, así como en otras instituciones del sector financiero y de las microfinanzas que busquen fortalecer la gestión del riesgo mediante el uso de inteligencia artificial explicable.

1.4.4 Alcance y delimitación

La investigación se enfoca en el análisis de las operaciones de crédito correspondientes a la “Cooperativa de Ahorro y Crédito Riobamba Ltda.” Utilizando datos históricos crediticios de los préstamos correspondientes al año 2025. El estudio se delimita al uso de técnicas de aprendizaje automático supervisado para la predicción del incumplimiento de pagos y al análisis de explicabilidad de los modelos seleccionados, sin considerar variables externas de carácter macroeconómico.

1.4.5 Hipótesis de la investigación

H1: Es posible predecir el incumplimiento de pagos en la Cooperativa de Ahorro y Crédito Riobamba Ltda., mediante modelos de aprendizaje automático con un nivel de precisión superior a los métodos tradicionales.

H2: Las técnicas de explicabilidad de modelos permiten identificar con claridad los factores que influyen en el incumplimiento de pagos en las operaciones de crédito.

1.5 Perfil de la organización: Nombre, actividades, mercados servidos y principales cifras

Nombre de la empresa

Cooperativa de Ahorro y Crédito Riobamba Ltda.

Misión, visión, valores

Misión. - Impulsamos el desarrollo socioeconómico de nuestros asociados, a través de soluciones financieras y sociales, que permitan un cambio en su calidad de vida, con un equipo humano comprometido que promueve la economía popular y solidaria. (COOP. RIOBAMBA LTDA, 2026)

Visión. - Ser una institución financiera inclusiva y humana, referente a nivel nacional, comprometida con los valores y principios cooperativos, impulsando el bienestar de nuestros socios y la comunidad. (COOP. RIOBAMBA LTDA, 2026)

Valores. -

- Solidaridad: Relación de apoyo mutuo
- Transparencia: Información clara, veraz y oportuna.
- Honestidad: Conducta decente.
- Equidad: Justicia según el grado de aporte.
- Trabajo en Equipo: Colaboración a los demás.
- Compromiso Social: Socorro y cooperación con la colectividad.

Actividades, marcas, productos y servicios

Actividades de recepción de depósitos y/o similares cercanos de depósitos y la concesión de créditos o préstamos de fondos. La concesión de crédito puede adoptar diversas formas, como préstamos, hipotecas, transacciones con tarjetas de crédito, etc. esas actividades son realizadas en general por instituciones monetarias diferentes de los bancos centrales, como: cooperativas de crédito y actividades de giro postal. (SRI, s.f.)

Ubicación de la sede

La entidad tiene su oficina principal en la ciudad de Riobamba,

Ubicación de las operaciones

La cooperativa de ahorro y crédito cuenta con 14 oficinas operativas:

- Oficina Matriz
- Oficina Norte
- Oficina Sur
- Oficina Condamine
- Oficina Davalos
- Oficina Guano
- Oficina Cumanda
- Oficina Alausi
- Oficina Chinchí
- Oficina Cuenca centro
- Oficina Cuenca arenal
- Oficina Quito Magdalena
- Oficina Quito condado
- Oficina Quito Guamaní

Propiedad y forma jurídica

Figura 1. Ficha Informativa de directivos y representante legal



Fecha de Generación de Documento: 21/enero/2026

FICHA INFORMATIVA DE DIRECTIVOS Y REPRESENTANTE LEGAL DE ENTIDADES DEL SECTOR FINANCIERO POPULAR Y SOLIDARIO

DATOS DE LA ENTIDAD

SECTOR:	COOPERATIVA DE AHORRO Y CREDITO
RUC:	0690045389001
RAZÓN SOCIAL:	COOPERATIVA DE AHORRO Y CREDITO RIOBAMBA LTDA
ESTADO A LA FECHA DE CONSULTA:	ACTIVA

DIRECTIVOS Y REPRESENTANTE LEGAL A LA FECHA DE CONSULTA:

REPRESENTANTE LEGAL (GERENTE):	MORALES MOROCHO SEGUNDO PEDRO
PRESIDENTE:	ZAMBRANO GARZON ERNESTO
SECRETARIO:	GUARACA MATUTE ELSA SUSANA
PRESIDENTE CONSEJO DE VIGILANCIA:	GUAMAN BAYAS VICTORIA
NOMBRE GERENTE SUBROGANTE:	CALERO RIVAS MARCELO SANTIAGO

El presente documento es una FICHA INFORMATIVA que **no constituye certificación de ningún tipo.**

Nota. El gráfico muestra una ficha informativa de directivos y representantes de la entidad emitida 21 enero del 2026

La cooperativa de Ahorro y Crédito está legalmente constituida ante todos los entes de supervisión y control normativos en las superintendencias como la SEPS, SRI, etc.

Figura 2. Certificado registro único de contribuyentes

SRI		Certificado Registro Único de Contribuyentes	
Razón Social COOPERATIVA DE AHORRO Y CREDITO RIOBAMBA LTDA.		Número RUC 0690045389001	
Representante legal • MORALES MOROCHO SEGUNDO PEDRO			
Estado ACTIVO	Régimen GENERAL	Artesano No registra	
Fecha de registro 15/05/1984	Fecha de actualización 01/12/2025	Inicio de actividades 12/10/1978	
Fecha de constitución 12/10/1978	Reinicio de actividades No registra	Cese de actividades No registra	
Jurisdicción ZONA 3 / CHIMBORAZO / RIOBAMBA		Obligado a llevar contabilidad SI	
Tipo SOCIEDADES	Agente de retención SI	Contribuyente especial SI	
Obligado a reportar a la UAFE SI			
Domicilio tributario Ubicación geográfica Provincia: CHIMBORAZO Cantón: RIOBAMBA Parroquia: LIZARZABURU Dirección Calle: 10 DE AGOSTO Intersección: COLON Edificio: COOPERATIVA DE AHORRO RIOBAMBA Referencia: A UNA CUADRA DEL MERCADO LA MERCED			
Medios de contacto Email: riobamba@cooprio.fin.ec Apartado postal: 000501 Celular: 0998109867			

Nota. El gráfico muestra un certificado de Registro único de contribuyente emitido por el SRI

Mercados servidos o ubicación de sus actividades de negocio

La Cooperativa de Ahorro y Crédito Riobamba Ltda, es una financiera social y solidaria líder en el centro del país, que atiende a toda la provincia de Chimborazo, Azuay y Pichincha siendo su zona de influencia directa Riobamba, Alausí, Chunchi, Cumandá, Guano, Quito, Cuenca.

Tabla 1. Ubicación de sus actividades de negocio

OFICINA	DIRECCION	TELEFONO / MAIL
Oficina Matriz	Dir: 10 DE AGOSTO S/N Y COLÓN	Teléfono: (03) 2960-270 mail: riobamba@cooprio.fin.ec
Oficina Norte	AV. DANIEL LEÓN BORJA Y LAS 44 (ESQ	Telefono: (03) 2961-250

		mail: norte@cooprio.fin.ec
Oficina Sur	AV. LEOPOLDO FREIRE Y BUCAREST	Telefono: (03) 2626-228 / 2626-904
		mail: sur@cooprio.fin.ec
Oficina Dávalos	PICHINCHA 30-54 Y NUEVA YORK	Telefono: (03) 2940-513
		mail: davalos@cooprio.fin.ec
Oficina Condamine	ESMERALDAS 28-62 Y CARABOBO	Telefono: (03) 2960-080
		mail: condamine@cooprio.fin.ec
Sede Administrativa	9 DE OCTUBRE Y 24 DE MAYO	Telefono: (03) 2928-096 / 2928-157 / 2928-129
Oficina Guano	GARCIA MORENO 06-45 Y COLON	Telefono: (03) 2900-112 / 2900-131
		mail: guano@cooprio.fin.ec
Oficina Alausi	AV. 5 DE JUNIO 43 Y RICAURTE	Telefono: (03) 2930-074 / 2931-154
		mail: alausi@cooprio.fin.ec
Oficina Chunchi	SIMON BOLIVAR 873 Y CAPITAN RICAURTE	Telefono: (03) 2936-049
		mail: chunchi@cooprio.fin.ec
Oficina Cumanda	AV. 9 DE OCTUBRE Y ABDON CALDERON	Telefono: (03) 2326-450
		mail: cumanda@cooprio.fin.ec
Oficina Quito Magdalena	AV. MARISCAL SUCRE S/N Y AV. TENIENTE HUGO ORTIZ	Telefono: (02) 2643-600
		mail: quito@cooprio.fin.ec
Oficina Quito Condado	AV. LA PRENSA N-70-121 Y PABLO PICASSO	Telefono: (02) 2590283
		mail: condado.quito@cooprio.fin.ec
Oficina Quito Guamani	PEDRO VICENTE MALDONADO / S52D / LEONIDAS DUBLES	Telefono: (02) 3651454
		mail: guamani.quito@cooprio.fin.ec
Oficina Cuenca Arenal	AV. DE LAS AMERICAS S/N Y GENERAL ESCANDON	Telefono: (07) 2855-212 / 2856-109
		mail: cuenca@cooprio.fin.ec
Oficina Cuenca Centro	SUCRE 584 Y HERMANO MIGUEL	Telefono: (07) 2835-099
		mail: cuenca_centro@cooprio.fin.ec

Nota. La tabla muestra las ubicaciones de sus actividades de negocio

En todas las oficinas operativas de la institución se brindan servicios de ahorro y crédito orientados a diversas actividades en favor del desarrollo de la microempresa y del

emprendimiento. Las actividades en las que trabaja directamente son la agricultura, el transporte, el microcrédito, el emprendimiento, el consumo, entre otras del sector financiero cooperativo.

Tamaño de la organización

La Cooperativa de Ahorro y Crédito Riobamba se encuentra ubicada en el segmento uno del sistema cooperativo nacional y al cierre del ejercicio económico alcanza \$620 millones en activos fruto de la intermediación financiera con recursos propios que garantizan la estabilidad financiera.

Información sobre empleados y otros trabajadores

La cooperativa está integrada con profesionales jóvenes con un promedio de edad de 30 años, actualmente consta con 280 empleados que intervienen en diferentes áreas de la cooperativa y gestionan eficientemente los recursos confiados a esta institución. Cabe indicar que todo el personal consta con niveles de estudio de tercer nivel y como caso particular el 80% está alcanzando maestrías es decir estudios de cuarto nivel.

No existe tercerización laboral por lo que el sentido de pertenencia es muy alto generando espacios de trabajo y convivencia eficientes.

Procesos claves relacionados con el objetivo propuesto

Al ser una institución financiera depende de la demanda de préstamos que garanticen el rendimiento y crecimiento de la institución para ellos los procesos de colocación como de recuperación son claves, siendo el proceso de cobranza uno de los principales para el control de la morosidad, se busca mediante este proyecto darles elementos y herramientas que ayuden en su gestión controlando la mora y bajando los índices de socios impagos.

Se busca generar escenarios proyectados en base a Datos e información cruzada y analizada, para crear algoritmos que ayuden al control y disminución de este índice de morosidad y baje el índice de impago y por ende el índice de riesgo, mejorando significativamente los indicadores.

Cifras, ratios y números que definan a la empresa

La cooperativa, dentro de sus principales indicadores con corte a diciembre de 2025, muestra una fortaleza que se encuentra reflejada en los siguientes indicadores:

Figura 3. Indicadores Financieros



Nota. Indicadores financieros de la “Cooperativa de Ahorro y Crédito Riobamba Ltda.”, enero 2026.

Modelo de Negocio

La “Cooperativa de Ahorro y Crédito Riobamba Ltda.” utiliza un modelo financiero en la modalidad de asociado, en el que cada miembro realiza un aporte al capital social y mediante dicho aporte, forma parte sustancial de la cooperativa. Estos certificados le permiten al socio

acceder a créditos, ahorros e inversiones, así como a beneficios como el seguro exequial, el fondo mortuario y la atención médica.

Grupos de interés internos y externos

Al ser una institución autónoma y regidos por normativa y política interna dentro de un marco legal macro emitido por la SEPS, se indica que los socios son los principales accionistas en la cooperativa y quienes en base a las asambleas forman consejos de administración y vigilancia los mismo que gestionan y hacen seguimiento a todas las actividades administrativas y financieras de la cooperativa; no hay grupos externos que influyan en la misma dando la posibilidad que sus asociados se desarrollen y crezcan con su institución buscando la igualdad y equidad social.

CAPÍTULO 2

2. Revisión de literatura

2.1. Introducción

El riesgo de incumplimiento crediticio constituye una de las principales preocupaciones de las entidades financieras, ya que incide de manera directa en la solvencia, la liquidez, la sostenibilidad y proporcionalmente, en la calidad de la cartera de las organizaciones; así, en las cooperativas de ahorro y crédito, la gestión adecuada del riesgo crediticio resulta especialmente relevante, ya que estas entidades canalizan los recursos de sus socios hacia actividades de consumo, emprendimiento, microempresa y producción. (Campuzano Camacho, 2026)

Este escenario, la capacidad de identificar anticipadamente las operaciones con mayor probabilidad de incumplimiento constituye un elemento estratégico para la gestión del riesgo crediticio, la detección temprana de potenciales de eventos de mora permite implementar medidas preventivas, optimizar los procesos de evaluación y seguimiento de las solicitudes de crédito, así como diseñar estrategias de monitoreo y cobranza más eficientes, orientadas a reducir el deterioro de la cartera y fortalecer la sostenibilidad financiera de la institución.

Paralelamente, la transformación digital del sector financiero ha generado un crecimiento significativo en la disponibilidad de información proveniente de sistemas transaccionales, historiales crediticios, registro de pagos, saldos, productos financieros y variables sociodemográficas de socios; este incremento en la disponibilidad de datos ha favorecido la incorporación de metodologías basadas en la ciencia de datos y aprendizaje automático, las cuales complementan los enfoques tradicionales de evaluación crediticia mediante el análisis de grandes volúmenes de información.

A diferencia de los métodos convencionales, sustentados principalmente en reglas predefinidas y análisis manuales, los sistemas de algoritmos de aprendizaje automático manejan la capacidad para identificar patrones complejos, interacciones entre múltiples variables y relaciones no lineales, mejorando la estimación del riesgo de cumplimiento y proporcionando información más precisa para apoyar la toma de decisiones en las instituciones financieras.

Sin embargo, el uso de modelos predictivos en decisiones crediticias no debe limitarse a obtener altos indicadores de desempeño estadístico. En un entorno financiero, es indispensable que las predicciones sean confiables, auditables, comprensibles y coherentes con la información disponible al momento de otorgar un crédito. Por esta razón, la presente investigación integra modelos de clasificación supervisada con técnicas de inteligencia artificial explicable, a fin de estimar la probabilidad de incumplimiento y, al mismo tiempo, identificar los factores que sustentan cada predicción.

2.2. Conceptos fundamentales

2.2.1. Riesgo crediticio e incumplimiento de pagos

El riesgo crediticio puede entenderse como la posibilidad de que un prestatario no cumpla total o parcialmente con las obligaciones financieras adquiridas frente a una entidad; este riesgo se materializa mediante el análisis de atrasos, mora, reestructuraciones, refinanciamientos, deterioro de la cartera o pérdida definitiva de recursos. La gestión del riesgo crediticio comprende la identificación, medición, monitoreo y control de factores que pueden afectar la capacidad de pago del deudor.

El incumplimiento de las obligaciones crediticias constituye la materialización del riesgo crediticio este incumplimiento se produce cuando el prestatario no cumple de manera parcial o

total, con los pagos establecidos en las condiciones del crédito dentro de los plazos acordados, este comportamiento puede estar asociado a factores económicos financieros laborales o personales que afectan la capacidad de pago del deudor, así como a circunstancias que inciden en su voluntad de cumplirlos adquiridos, si bien es cierto los criterios para determinar el incumplimiento pueden variar de acuerdo con las políticas internas de cada institución financiera y normativa legal vigente, la mora superior a un periodo previamente definido es uno de los indicadores más utilizados para identificar operaciones con un nivel mayor de riesgo.

En este estudio el incumplimiento se operacionaliza mediante la variable: MORA - BINARIA, Asignando el valor de uno a las operaciones de crédito cuya mora máxima es igual o superior a 30 días y el valor de cero a aquellas que no alcanzan dicho umbral.

La estimación anticipada de la probabilidad de cumplimiento constituye entonces un elemento fundamental para fortalecer la gestión del riesgo crediticio y mejorar la calidad de las decisiones institucionales lo que va más allá de determinar la aprobación o rechazo de una solicitud de crédito por lo que los modelos predictivos permiten entonces clasificar las operaciones según su nivel de riesgo, establecer condiciones acordes con el perfil del solicitante, implementar mecanismos de alerta temprana y orientar las estrategias de seguimiento, acompañamiento y recuperación de cartera.

Desde esta perspectiva, la predicción del riesgo de mora se convierte en una herramienta de apoyo para optimizar la administración de cartera crediticia, reducir las pérdidas potenciales y contribuir a la sostenibilidad financiera en las instituciones, en concordancia con este enfoque la Superintendencia de Economía Popular y Solidaria (SEPS 2024), Él reconoce que la gestión integral del riesgo constituye un componente estratégico para garantizar la estabilidad, solvencia

y continuidad operativa de las entidades que conforman el Sistema Financiero Popular y Solidario del Ecuador

2.2.2. Cooperativas de Ahorro y Crédito

Las cooperativas de ahorro y crédito forman parte del sector de la economía popular y solidaria, cuya propuesta base se centra en la intermediación financiera, bajo principios de cooperación, participación y servicio a sus socios; estas organizaciones desempeñan un papel importante en la inclusión financiera, especialmente en sectores donde el acceso al crédito bancario tradicional puede ser limitado, además, su actividad crediticia abarca productos de consumo, microcrédito, emprendimiento, vivienda y actividades productivas.

La gestión de la cartera crediticia en las cooperativas de ahorro y crédito requiere un análisis integral de diversos factores que pueden incidir en la capacidad de pago de los socios, entre éstos se encuentran variables de carácter económico, financiero, social y demográfico, tales como el nivel de ingresos, la estabilidad laboral, las cargas familiares, los montos solicitados, los plazos de créditos, el valor de las cuotas, el nivel de endeudamiento, el historial de pagos y las características del producto financiero.

La interacción de estos factores influye en el comportamiento crediticio de los prestatarios y en consecuencia, en la probabilidad de que una operación incurra en mora, por esta razón el desarrollo de los modelos predictivos debe sustentarse en información institucional confiable, completa y de calidad que refleje las particularidades de la cartera de cada cooperativa, asimismo, resulta fundamental que estos modelos sean diseñados y validados considerando el contexto operativo y las características propias de la institución, evitando la

adopción indiscriminada de metodologías desarrolladas para otras realidades financieras que podrían limitar la precisión y confiabilidad de las predicciones obtenidas.

2.2.3. Ciencia de datos y aprendizaje automático

La ciencia de datos es un campo interdisciplinario que integra estadística, programación, conocimiento del dominio y métodos computacionales para extraer información útil a partir de datos. En el ámbito financiero, la ciencia de datos permite transformar registros operativos en indicadores que respaldan la gestión de riesgos, la segmentación de clientes, la detección de fraude, la evaluación crediticia y la optimización de procesos. (Superintendencia de Economía Popular y Solidaria, 2024)

El aprendizaje automático (machine learning), constituye una de las principales áreas de desarrollo de inteligencia artificial y se fundamenta en la creación de algoritmos capaces de identificar patrones y relaciones a partir de datos históricos, a diferencia de enfoques tradicionales, en las que las reglas de decisión son definidas de manera explícita por un experto y los modelos de aprendizaje automático construyen su conocimiento mediante un proceso de entrenamiento, durante el cual ajustan sus parámetros utilizando ejemplos previamente observados, permitiendo reconocer el comportamiento complejo y generar predicciones sobre nuevas observaciones con un alto grado de precisión. (Superintendencia de Economía Popular y Solidaria, 2024)

En el ámbito de la gestión del riesgo crediticio, el aprendizaje supervisado representa una de las metodologías más utilizadas para la predicción de incumplimientos y obligaciones financieras permitiendo utilizar datos históricos para diferenciar entre créditos que fueron pagados oportunamente y aquellos que incurrieron en mora, con esta información y a manera de

aprendizaje se convierte en una herramienta analítica que complementa los procesos tradicionales de evaluación proporcionando evidencia cuantitativa para fortalecer la toma de decisiones y mejorar la gestión preventiva del riesgo crediticio.

La literatura reciente señala que los modelos de aprendizaje automático son especialmente útiles en problemas crediticios porque pueden procesar un gran número de variables, detectar relaciones no lineales y capturar interacciones complejas entre las características del solicitante y de la operación. No obstante, su desempeño depende de la calidad del conjunto de datos, la correcta definición de la variable objetivo, la prevención de fuga de información, el tratamiento del desbalance de clases y la validación temporal de los resultados (Bussmann et al., 2021; Shrestha et al., 2022).

2.2.4. Clasificación supervisada

La clasificación supervisada es una técnica de aprendizaje automático en la que el algoritmo aprende a asignar una observación a una clase previamente definida. En la presente investigación, las clases corresponden a cumplimiento e incumplimiento de pagos. Cada préstamo se representa mediante variables predictoras, tales como edad, ingresos, plazo, cuota estimada, relación cuota-ingreso, relación deuda-ingreso, producto crediticio e historial de mora.

El desarrollo de un modelo de clasificación requiere organizar la información en 3 conjuntos de datos entrenamiento validación y prueba, el conjunto de entrenamiento se emplea para que el algoritmo aprenda patrones presentes en los datos y ajuste sus parámetros, posteriormente el conjunto de validación permite optimizar la configuración del modelo definiendo los hiperparámetros y los umbrales de decisión más adecuados, y finalmente, el conjunto de pruebas se utiliza para evaluar el rendimiento del modelo utilizado mediante el uso

de información que no participó en el proceso de aprendizaje, lo que permite estimar su calidad de generalización. (Stanford University, 2024)

En el ámbito del riesgo crediticio, esta división debe realizarse con especial cuidado, considerando la secuencia temporal de los datos y la posible relación entre las observaciones, por ejemplo, cuando un socio registra varias operaciones de crédito, es indispensable evitar que estas se distribuyan simultáneamente entre los conjuntos de entrenamiento y prueba, ya que se podría generar una sobreestimación del desempeño real del modelo al introducir información relacionada en ambas fases de evaluación. (Scikit-learn Developers, 2025)

2.2.5. Regresión Logística

La Regresión Logística es un modelo estadístico ampliamente utilizado para problemas de clasificación binaria, su objetivo es estimar la probabilidad de que una observación pertenezca a una clase, mediante una función logística que transforma una combinación lineal de variables predictoras en un valor comprendido entre cero y uno. En riesgo crediticio, este modelo ha sido utilizado históricamente por su interpretabilidad, facilidad de implementación y capacidad para comunicar el efecto de cada variable mediante coeficientes y razones de probabilidad.

A pesar de sus ventajas, la Regresión Logística presenta limitaciones cuando las relaciones entre las variables y el incumplimiento son altamente no lineales o cuando existen interacciones complejas entre características financieras, demográficas e históricas. Por esta razón, se utiliza en esta investigación como línea de base para comparar su desempeño con modelos de ensamble más flexibles, como Random Forest y XGBoost.

2.2.6. Random Forest

Random Forest es un algoritmo de ensamblaje basado en la construcción de múltiples árboles de decisión. Cada árbol se entrena utilizando una muestra aleatoria de observaciones y un subconjunto aleatorio de variables. La predicción final se obtiene mediante votación mayoritaria en clasificación o mediante un promedio en regresión; este mecanismo reduce la varianza de un árbol individual y disminuye el riesgo de sobreajuste.

Para los problemas de incumplimiento crediticio, Random Forest es relevante porque puede modelar relaciones no lineales, interacciones entre variables y patrones heterogéneos sin requerir una especificación funcional previa; además, es relativamente robusto frente a variables correlacionadas, valores atípicos y combinaciones de características numéricas y categóricas transformadas. La evidencia reciente indica que los métodos de ensamble presentan ventajas frente a modelos lineales cuando el comportamiento de incumplimiento depende de múltiples factores interrelacionados (Kou et al., 2021; Yang et al., 2025).

Random Forest también permite estimar la importancia de las variables, lo que facilita identificar qué características contribuyen con mayor fuerza a la predicción; sin embargo, la importancia convencional basada en la impureza puede estar sesgada cuando hay variables correlacionadas o de alta cardinalidad. Por ello, resulta conveniente complementar el análisis con importancia por permutación y técnicas de explicabilidad como SHAP.

2.2.7. XGBoost

XGBoost es un algoritmo de gradient boosting que construye árboles de decisión secuencialmente, cada nuevo árbol explora para corregir los errores cometidos por los árboles anteriores, optimizando una función de pérdida y aplicando mecanismos de regularización para

controlar la complejidad del modelo. Esta arquitectura le permite alcanzar altos niveles de desempeño predictivo en conjuntos de datos tabulares.

En la predicción de incumplimiento, XGBoost puede capturar patrones complejos y lograr una alta capacidad discriminativa. No obstante, su uso requiere un ajuste cuidadoso de hiperparámetros, el control del sobreajuste, el tratamiento del desequilibrio de clases y la calibración de probabilidades. Un AUC elevado no garantiza, por sí solo, que el modelo sea el más útil para la gestión crediticia, ya que deben analizarse también su recall precisión, F1-score, matriz de confusión y capacidad de identificar préstamos realmente riesgosos en el umbral operativo definido.

2.2.8. Desbalance de clases

El desbalance de clases ocurre cuando una categoría posee una frecuencia considerablemente menor que otra. En la predicción de mora, la clase de incumplimiento suele ser minoritaria frente a la clase de cumplimiento. Esta situación puede llevar a los modelos a favorecer la clase mayoritaria, lo que se traduce en una alta exactitud, pero en una baja capacidad para identificar los casos de mora.

Para abordar este problema, se utilizan técnicas como la ponderación de clases, el sobremuestreo de la clase minoritaria, el submuestreo de la clase mayoritaria y el ajuste de umbrales de clasificación. En el presente estudio, la tasa de mora de 23,47 % evidencia un desbalance moderado, por lo que resulta necesario priorizar métricas sensibles a la clase minoritaria, como recall precisión, F1-score y curva precision-recall. La literatura recomienda que el tratamiento del desbalance se aplique exclusivamente sobre los datos de entrenamiento

para evitar contaminación de los conjuntos de validación y prueba (Buda et al., 2018; Krawczyk, 2021).

2.2.9. Evaluación de modelos de clasificación

La evaluación de un modelo de clasificación requiere utilizar múltiples métricas. La exactitud representa la proporción total de predicciones correctas, pero puede resultar engañosa en conjuntos desbalanceados. La precisión mide la proporción de casos clasificados como incumplidos que realmente incurrieron en mora. El recall o sensibilidad representa la proporción de préstamos morosos que fueron correctamente detectados por el modelo; el F1-Score combina precisión y recall mediante su media armónica. (Scikit-learn Developers, 2025)

Por otro lado, la matriz de confusión permite analizar los verdaderos positivos, los verdaderos negativos, los falsos positivos y los falsos negativos, donde, en el contexto crediticio, los falsos negativos tienen especial importancia porque representan operaciones que el modelo consideró de bajo riesgo, pero que posteriormente incurrieron en mora. El AUC-ROC mide la capacidad del modelo para distinguir entre las clases a través de diferentes umbrales, mientras que la curva precisión-recall resulta particularmente útil cuando existe desbalance de clases. (McDermott, 2024)

2.2.10. Inteligencia artificial explicable

La inteligencia artificial explicable, conocida como XAI, comprende métodos y técnicas orientados a hacer comprensibles las decisiones generadas por modelos de inteligencia artificial. En entornos financieros, la explicabilidad es necesaria para que los analistas, directivos, auditores y usuarios comprendan los factores que influyen en una predicción y puedan evaluar si el modelo utiliza criterios razonables, consistentes y no discriminatorios. (Černevičienė, 2024)

Las técnicas de explicabilidad pueden ser globales o locales. Las explicaciones globales describen el comportamiento general del modelo e identifican las variables más influyentes en el conjunto de datos. Las explicaciones locales permiten conocer por qué un préstamo específico fue clasificado con un determinado nivel de riesgo. SHAP se basa en valores de Shapley, provenientes de la teoría de juegos, y permite asignar a cada variable una contribución positiva o negativa sobre la predicción. LIME, por su parte, aproxima localmente el comportamiento de un modelo complejo mediante un modelo interpretable más simple (Misheva et al., 2021; Phillips et al., 2021).

La explicabilidad no reemplaza la validación estadística, ni la supervisión humana; su función es complementar el desempeño predictivo mediante la transparencia, la trazabilidad y la posibilidad de revisión; en decisiones de crédito, una explicación adecuada permite justificar alertas, detectar variables problemáticas, verificar la coherencia con las políticas institucionales y mejorar la aceptación del modelo por parte de los usuarios internos.

2.3. Estado del arte

La literatura reciente sobre la predicción de incumplimiento muestra una transición de los modelos estadísticos tradicionales hacia los algoritmos de aprendizaje automático y los métodos de ensamble. La Regresión Logística continúa siendo utilizada como línea base por su interpretabilidad; sin embargo, diversos estudios reportan que Random Forest, XGBoost, LightGBM y otros modelos basados en árboles pueden alcanzar un mejor desempeño cuando los datos contienen relaciones no lineales, variables heterogéneas e interacciones complejas.

Uno de los aspectos que ha adquirido mayor relevancia en los últimos años es la necesidad de equilibrar el desempeño predictivo con la interpretabilidad de los modelos, en este

sentido Bussmann et al. (2021) analizaron El papel de la inteligencia artificial explicable en la gestión del riesgo crediticio y demostraron que las técnicas SHAP y LIME Permiten comprender las razones que sustentan las predicciones generadas por modelos complejos los autores evidencias que estas metodologías facilitan la identificación de variables con mayor influencia sobre cada decisión individual, fortaleciendo la transparencia y la confianza en los sistemas automatizados, asimismo sostienen que la incorporación de mecanismos de explicabilidad constituyen un requisito fundamental para el uso responsable de la inteligencia artificial, particularmente en sectores altamente regulados, como el financiero donde las decisiones pueden generar importantes repercusiones económicas y sociales.

Desde una perspectiva más amplia Kou et al. (2021) realizaron una revisión De los principales modelos de aprendizaje automático aplicados al riesgo crediticio y concluyeron que los algoritmos de ensamble presentan una capacidad predictiva superior frente a los métodos de estadísticos tradicionales, sin embargo destacan que el rendimiento de estos modelos no depende únicamente del algoritmo seleccionado sino también de la calidad de la información utilizada durante el entrenamiento, de una adecuada selección de variables, del tratamiento del desbalance entre las clases, de la optimización de hiperparámetros y de la aplicación de procedimientos rigurosos de validación, los autores señalan además que la evaluación del desempeño debe trascender indicadores globales como el área bajo la curva ROC (AUC) , incorporando métricas adicionales que permitan valorar la capacidad del modelo para identificar correctamente los casos de mayor probabilidad de incumplimiento.

Por su parte, Shrestha et al. (2022) profundizaron los factores metodológicos que influyen en la capacidad de generalización de los modelos predictivos utilizados en evaluación crediticia,

sus resultados muestran que la ingeniería de características y la validación temporal representan elementos determinantes para obtener modelos robustos y aplicables en escenarios reales; en particular advierten que la utilización de particiones aleatorias de datos puede generar estimaciones excesivamente optimistas cuando existen variaciones temporales en el comportamiento de los prestatarios ó cuándo se incorporan variables que contienen información posterior al evento de incumplimiento fenómeno conocido como Data Leakage De información, este planteamiento resulta especialmente pertinente para la investigación debido a que el modelo fue estructurado considerando cada operación crediticia comunidad de análisis y priorizando la utilización exclusiva de información disponible al momento de la concesión del crédito.

Las investigaciones más recientes continúan confirmando la superioridad de los métodos de ensamble en la predicción del riesgo crediticio, Yang et al. (2025) compararon el desempeño de la regresión logística, Random Forest, XGBoost y LightGBM, Utilizando diferentes estrategias de procesamiento de ingeniería de variables, los resultados evidenciaron que los algoritmos basados en árboles alcanzaron mejores niveles de precisión y discriminación, particularmente cuando se implementaron técnicas para el tratamiento de desbalance de clases y la optimización de características predictoras, además el estudio incorporó herramientas de inteligencia artificial explicable específicamente SHAP, para identificar el grado de contribución de cada variable al resultado final concluyendo que la combinación entre alta capacidad predictiva y explicabilidad constituye una condición indispensable para favorecer la adopción de estos modelos dentro de las instituciones financieras.

En la misma línea, Golec y AlabdulJalil (2025) desarrollaron una revisión sistemática sobre la aplicación de enfoques interpretables de inteligencia artificial en la evaluación del riesgo

crediticio, los autores identifican una tendencia creciente hacia el desarrollo de modelos que, además de ofrecer elevados niveles de precisión, incorpore mecanismos de transparencia, gobernanza algorítmica y evaluación de posibles sesgos, su análisis destaca que la confiabilidad de los sistemas de inteligencia artificial depende no sólo de la capacidad predictiva, sino también, de la posibilidad de justificar técnicamente las decisiones generadas especialmente cuando éstas inciden en el acceso de las personas a servicios y productos financieros.

En el contexto ecuatoriano, la administración del riesgo crediticio en las cooperativas de ahorro y crédito se encuentra regulada por la superintendencia de economía popular y solidaria (SEPS), organismo que promueve la implementación de sistemas integrales para la gestión del riesgo, el monitoreo permanente de la cartera y el fortalecimiento de los mecanismos de control sobre la exposición crediticia; a pesar de estos avances regulatorios, la producción científica nacional realizada con la aplicación de modelos explicables de aprendizaje automático utilizando bases de datos reales provenientes de cooperativas, continúa siendo limitada, esta situación evidencia la oportunidad de investigación orientada a generar conocimiento aplicado que contribuya en el fortalecimiento de procesos de evaluación crediticia dentro del sector financiero popular y solidario.

De este modo se puede sintetizar que en el análisis de la literatura especializada se ha permitido identificar cuatro tendencias que orientan el desarrollo de la investigación contemporánea en riesgo crediticio.

En primer lugar, se observa una creciente adopción de algoritmos de ensamble debido a su mayor capacidad para modelar relaciones complejas y mejorar la precisión predictiva.

En segundo lugar, existe un consenso respecto a la importancia de prevenir la fuga de información y utilizar esquemas de validación temporal que reflejen de manera realista el comportamiento futuro de la cartera.

En tercer lugar, la evaluación de los modelos debe fundamentarse en un conjunto amplio de indicadores de desempeño que permitan valorar adecuadamente su comportamiento frente al desbalance de clases característico del riesgo crediticio.

Y finalmente la incorporación de técnicas de inteligencia artificial explicable se consolida como un componente esencial para garantizar la transparencia y trazabilidad de la confianza de los sistemas de apoyo en la toma de decisiones financieras favoreciendo su adopción en entornos regulados fortaleciendo la gobernanza de los modelos predictivos.

2.4. Marco teórico

2.4.1. Teoría de gestión del riesgo crediticio

La gestión del riesgo crediticio se fundamenta en la identificación de factores que pueden afectar la recuperación de los recursos otorgados. Desde esta perspectiva, la evaluación de crédito debe considerar la capacidad de pago, el comportamiento histórico, el nivel de endeudamiento, las características del producto y las condiciones económicas del socio. La predicción de incumplimiento se integra a esta gestión como una herramienta de alerta temprana que permite estimar la probabilidad de deterioro de una operación antes de que se materialice la mora.

El modelo predictivo desarrollado no tiene como finalidad reemplazar el criterio profesional de los asesores ni las políticas de otorgamiento de crédito establecidas por la

institución, su propósito es servir como una herramienta de apoyo que, a partir del análisis de datos históricos aporte una estimación objetiva del riesgo de incumplimiento, de esta manera, sus resultados pueden complementar la toma de decisiones al facilitar La definición de estrategias de seguimiento, la solicitud de información adicional, el ajuste de las condiciones crediticias y la implementación de acciones preventivas orientadas a fortalecer la gestión del riesgo de la cartera.

2.4.2. Teoría de clasificación y probabilidad de incumplimiento

Desde la teoría de clasificación supervisada, cada préstamo se representa mediante un vector de características predictoras y una variable objetivo que identifica si ocurrió o no el incumplimiento. El modelo aprende una función que estima la probabilidad condicional de mora, expresada como:

$$P(\text{MORA_BINARIA} = 1 \mid X)$$

Donde X representa el conjunto de variables del socio, del préstamo, de la capacidad de pago e historial crediticio.

La clasificación final de cada solicitud de crédito se determinará a partir de un umbral de decisión aplicado a la probabilidad de incumplimiento, estimada por el modelo, donde, cuando la probabilidad calculada supera dicho valor de referencia, la operación se identifica como de mayor riesgo; en caso contrario, se considera de menor riesgo. Este procedimiento permite transformar una predicción probabilística en una herramienta práctica que apoya el análisis crediticio y fortalece el proceso de toma de decisiones en la cooperativa.

La definición del umbral de clasificación requiere un análisis técnico que responda a los objetivos y políticas de gestión del riesgo de la institución, por lo que no debe establecerse de

manera arbitraria, su selección debe considerar el impacto de los falsos positivos y los falsos negativos, el nivel de riesgo de la cooperativa está dispuesto a asumir y su capacidad para gestionar las alertas generadas, mientras que un umbral más bajo favorece la identificación temprana de posibles incumplimientos, uno más elevado reduce las alertas innecesarias, aunque puede incrementar la probabilidad de que algunos créditos de riesgo sean clasificados como operaciones seguras.

2.4.3. Teoría de ensamble y Random Forest

La teoría de ensamble sostiene que la combinación de múltiples modelos puede producir predicciones más estables y precisas que un único modelo individual. Random Forest aplica este principio mediante la construcción de numerosos árboles de decisión entrenados sobre muestras bootstrap y subconjuntos aleatorios de variables.

En el algoritmo Random Forest, cada árbol de decisión realiza una predicción de manera independiente y la clasificación final se obtiene mediante un proceso de votación mayoritaria entre todos los árboles que conforman el modelo. Mediante este enfoque se reduce la variabilidad de las predicciones individuales y se mejora la estabilidad de los resultados al combinar múltiples estimaciones en una única decisión con mayor capacidad de generalización.

Una de las principales fortalezas de Random Forest en la predicción del incumplimiento crediticio radica en su capacidad para identificar relaciones no lineales e interrelaciones complejas entre variables que describen el perfil del solicitante, en la práctica, el riesgo de mora no suele depender de un único factor, sino de la combinación de ciertas características como ingresos insuficientes una elevada relación entre la cuota y la capacidad de pago, antecedentes de incumplimiento, o un alto nivel de endeudamiento a diferencia de los modelos lineales, este

algoritmo puede detectar automáticamente estos patrones sin necesidad de establecer previamente evacuaciones o reglas específicas lo que favorece una representación más realista del comportamiento crediticio. (Su, 2021)

2.4.4. Teoría de gradient boosting y XGBoost

Gradient boosting es una técnica de ensamble secuencial en la cual los modelos se construyen de manera iterativa para corregir los errores de los modelos anteriores. XGBoost incorpora regularización, optimización computacional y control de complejidad, lo que lo convierte en un algoritmo competitivo para datos tabulares. En la predicción de mora, XGBoost puede alcanzar una elevada capacidad discriminativa al identificar patrones complejos entre las variables.

Aunque el valor del AUC constituye un indicador importante para comparar el desempeño de los modelos predictivos no es el único criterio que debe considerarse en la gestión del riesgo crediticio. Un modelo puede alcanzar una elevada capacidad de discriminación y al mismo tiempo presentar un nivel traducido de recall, lo que implica una menor capacidad para identificar los créditos que efectivamente incurrirán en incumplimiento, en consecuencia, la evaluación de algoritmos como XGBoost, debe realizarse de manera integral, contrastando los resultados obtenidos por Random Forest e incorporando métricas complementarias, proceso de calibración probabilidades de un análisis detallado de falsos negativos, con el fin de seleccionar el modelo que ofrezca un equilibrio adecuado entre precisión, capacidad predictiva y utilidad para la toma de decisiones institucionales.

2.4.5. Teoría de explicabilidad y transparencia algorítmica

La explicabilidad algorítmica se fundamenta en la necesidad de comprender las razones que sustentan una predicción. En modelos complejos, como Random Forest y XGBoost, las decisiones se generan a partir de múltiples árboles y relaciones internas que no son fácilmente interpretables de manera directa. Las técnicas XAI permiten traducir estas relaciones en explicaciones comprensibles para los usuarios.

La técnica SHAP se fundamenta en los valores de Shapley para cuantificar la influencia que cada variable ejerce sobre la predicción generada por el modelo. Cuando la contribución de una variable es positiva, esta incrementa la probabilidad estimada de incumplimiento; en cambio, una contribución negativa disminuye dicho riesgo, esta metodología facilita tanto el análisis global del modelo, al identificar los factores con mayor incidencia de las predicciones, como el análisis individual de cada operación crediticia, permitiendo explicar las razones que sustentan el nivel de riesgo asignado a un préstamo específico.

De manera complementaria, LIME genera modelos locales simplificados que aproximan al comportamiento del algoritmo alrededor de una predicción completa, favoreciendo la comprensión e interpretación de las decisiones emitidas por modelos de aprendizaje automático.

En la presente investigación, la explicabilidad cumple tres funciones

- 1) Permite validar que el modelo utilice factores coherentes con la lógica crediticia.
- 2) Facilita la comunicación de los resultados a analistas y directivos.
- 3) contribuye a detectar posibles sesgos, variables irrelevantes o patrones que podrían conducir a decisiones injustificadas.

2.4.6. Modelo conceptual de la investigación

El modelo conceptual de la investigación plantea que la probabilidad de incumplimiento de pagos depende de la interacción entre cuatro dimensiones principales: perfil del socio, condiciones del préstamo, capacidad de pago e historial crediticio.

El perfil del socio se caracteriza mediante variables demográficas y aspectos relacionados con su vínculo con la cooperativa por su parte las características del crédito comprenden elementos como el monto otorgado, el plazo de financiamiento, el tipo de producto, la tasa de interés y la periodicidad de los pagos, la evaluación de la capacidad de pago considera indicadores como los ingresos del solicitante el valor estimado de la cuota y los niveles del endeudamiento, mientras que, el historial crediticio incorpora información referente a cumplimiento previo, alertas de riesgo, obligaciones vencidas y procesos de recuperación de cartera. (Siddique, 2024)

En conjunto estas dimensiones conforman el conjunto de variables de entrada utilizado para el entrenamiento y la evaluación de los modelos de aprendizaje automático, a partir de esta información, algoritmos como la Regresión Logística, Random Forest XGBoost estiman la probabilidad de incumplimiento de cada operación crediticio, posteriormente, las técnicas de inteligencia artificial explicable como SHAP y LIME, permiten identificar la influencia de cada variable con el comportamiento del modelo y ofrecer una explicación comprensible de las predicciones realizadas para cada préstamo,

.El resultado final es una clasificación de riesgo que puede ser utilizada como apoyo para la gestión crediticia, el seguimiento preventivo y la toma de decisiones basada en evidencia.

CAPÍTULO 3

3.1. Enfoque, tipo y diseño de la investigación

La presente investigación se desarrolló bajo un enfoque cuantitativo, debido a que utiliza datos históricos de operaciones de crédito, variables financieras, sociodemográficas y de comportamiento crediticio para identificar patrones asociados al incumplimiento de pagos.

El análisis se fundamenta en técnicas de estadística descriptiva, de procesamiento de datos y de aprendizaje automático, con el propósito de estimar la probabilidad de que una operación crediticia alcance una mora igual o superior a 30 días.

El estudio es de tipo aplicado y tecnológico, ya que busca:

Resolver una necesidad concreta de la “Cooperativa de Ahorro y Crédito Riobamba Ltda.”,

Fortalecer la gestión del riesgo crediticio mediante una herramienta predictiva que permita identificar de manera temprana las operaciones con mayor probabilidad de incumplimiento.

Alcance descriptivo, correlacional y predictivo.

Es descriptivo porque caracteriza las operaciones de crédito y el comportamiento de la mora; correlacional porque examina las relaciones entre variables financieras, crediticias y demográficas; y

Predictivo porque compara modelos de aprendizaje automático para anticipar el incumplimiento de pagos.

El diseño de la investigación es no experimental, longitudinal y retrospectivo.

No experimental, porque las variables no fueron manipuladas por los investigadores; longitudinal, porque se analizaron registros crediticios generados durante el período de estudio; y retrospectivo, porque los datos utilizados corresponden a operaciones ya registradas en la base institucional.

La investigación se delimitó al análisis de préstamos otorgados posteriores al proceso de migración llevado a finales del año 2024 e inicios del 2025, utilizando una fecha de corte fija del 09 de junio de 2026 para estandarizar los cálculos temporales y asegurar la consistencia entre el entrenamiento y la predicción individual.

3.2. Contexto, población, muestra y unidad de análisis

La investigación se desarrolló con información de la Cooperativa de Ahorro y Crédito Riobamba Ltda., obtenida desde su sistema transaccional institucional almacenado en una base de datos Oracle. La población estuvo conformada por las operaciones de crédito disponibles durante el período de estudio. Se utilizó una muestra censal, debido a que se consideraron todas las operaciones que cumplían con los criterios mínimos de calidad, disponibilidad y consistencia de información.

La base original estuvo conformada por 75.783 préstamos, correspondientes a 46.249 socios y 44 variables iniciales. Posteriormente, luego de la aplicación del proceso de extracción, transformación y carga, se generó una base analítica con 71 variables.

La tasa de mora observada, definida mediante el criterio de mora igual o superior a treinta días, fue de 23,47 %. Esta proporción evidencia que el incumplimiento representa una clase

minoritaria, pero con una frecuencia suficiente para entrenar y evaluar modelos de clasificación supervisada.

La unidad de análisis seleccionada fue el préstamo; esta decisión metodológica resulta fundamental, debido a que, el objetivo de la investigación es predecir el riesgo de incumplimiento desde el momento de otorgamiento de una operación crediticia.

El enfoque anterior, a nivel de cuota, podía incorporar información relacionada con vencimientos o pagos posteriores, lo que generaba el riesgo de que el modelo aprendiera patrones temporales y no condiciones reales de riesgo crediticio; por ello, la consolidación de una fila por préstamo permite que cada observación represente una decisión crediticia individual y que las variables utilizadas estén disponibles al momento de su otorgamiento.

3.3. Variable dependiente y variables independientes

La variable dependiente de la investigación se denominó MORA_BINARIA.

Esta variable se definió como una clasificación dicotómica del comportamiento de pago de cada préstamo: se asignó el valor 1 a las operaciones que registraron una mora máxima de 30 días o más, y el valor 0 a las que no alcanzaron dicho umbral.

Esta definición permitió transformar el problema del riesgo crediticio en una tarea de clasificación binaria, lo que facilitó el entrenamiento y la comparación de los modelos de aprendizaje automático.

Valor 1: préstamo con mora máxima igual o superior a 30 días.

Valor 0: préstamo sin mora o con mora inferior a 30 días.

Las variables independientes se organizaron en grupos relacionados con el perfil del socio, las condiciones de la operación crediticia, la capacidad de pago, el endeudamiento y el historial crediticio. Entre las variables relacionadas con el socio se incluyeron la edad, el sexo, las cargas familiares, la condición de discapacidad, la ciudad de residencia y la antigüedad como socio. En cuanto a las condiciones del préstamo, se consideraron el producto crediticio, el monto, el plazo, la tasa de interés, la frecuencia de pago y el valor estimado de la cuota.

Asimismo, se construyeron variables derivadas para representar la capacidad financiera y el nivel de exposición al riesgo, tales como la relación cuota-ingreso, la relación deuda-ingreso, indicadores de insuficiencia de ingresos, antigüedad del crédito y alertas de préstamo. Finalmente, se incorporaron variables de historial crediticio, entre ellas el número de préstamos previos, préstamos con alerta, préstamos morosos, vencidos, judiciales, prejudiciales, castigados y el indicador de riesgo histórico severo. Estas variables permiten representar la trayectoria crediticia previa del socio y aportan información relevante para la estimación del incumplimiento futuro.

Tabla 2. Variables para estimación de cumplimiento

Grupo	Ejemplos de variables
Perfil del socio	Edad, sexo, cargas familiares, discapacidad, ciudad
Capacidad económica	Ingresos, cuota estimada, ratio cuota/ingreso, ratio deuda/ingreso
Características del préstamo	Producto, monto, plazo, tasa de interés, frecuencia de pago

Historial crediticio	Total, de préstamos previos, alertas, morosidad, vencidos, judiciales, castigados
Variables derivadas	Antigüedad como socio, edad del crédito, riesgo histórico severo, indicadores de ingresos insuficientes

Nota. Información elaborada a partir de datos institucionales

3.4. Proceso de extracción, transformación y carga de datos

El proceso de preparación de datos se estructuró mediante una metodología ETL, correspondiente a las etapas de extracción, transformación y carga. En la fase de extracción, se obtuvo la información de la base de datos Oracle de la cooperativa y se almacenó en archivos Parquet. Este formato fue seleccionado por su eficiencia en el almacenamiento y el procesamiento de datos tabulares de gran volumen.

La extracción se orientó a consolidar una observación por préstamo, integrando:

- información del socio,
- características de la operación,
- pagos,
- saldos,
- productos
- antecedentes crediticios.

Durante la fase de transformación, se ejecutaron acciones de normalización de variables categóricas, conversión de campos de fecha al formato adecuado, tratamiento de valores faltantes, imputación de ingresos mediante reglas estadísticas, corrección de la información

geográfica cuando fue posible y eliminación de variables de baja calidad o con alto porcentaje de valores ausentes. Además, se revisaron duplicados, inconsistencias en identificadores y registros que no cumplían con los criterios mínimos de calidad para el análisis.

Como parte de la ingeniería de características, se generaron variables derivadas relacionadas con la edad del socio, la antigüedad como socio, el plazo del crédito, el valor estimado de la cuota, los ratios de cuota-ingreso y deuda-ingreso, así como indicadores de ingresos insuficientes. Estas variables permitieron transformar los datos operativos originales en indicadores analíticos capaces de representar la capacidad de pago y la exposición financiera de los socios.

El ETL incluyó normalización, conversión de fechas, tratamiento de faltantes, imputación de ingresos, corrección geográfica y eliminación de variables de baja calidad.

3.5. Construcción del historial crediticio y prevención de fuga de información

La construcción del historial crediticio se realizó utilizando únicamente información correspondiente a préstamos anteriores al crédito que se desea predecir. Para cada operación se calcularon indicadores históricos, tales como el total de préstamos previos, préstamos con alerta, préstamos morosos, vencidos, judiciales, prejudiciales y castigados. También se generó la variable de riesgo histórico severo, la cual permite identificar antecedentes relevantes de deterioro crediticio.

La prevención de fuga de información constituyó una condición metodológica prioritaria. En este sentido, se excluyeron variables que revelaban directa o indirectamente el resultado futuro del préstamo, tales como días de mora posteriores, pagos efectuados después del

otorgamiento, estados finales de la operación, calificaciones posteriores y cualquier información generada una vez iniciado el comportamiento de incumplimiento. De igual manera, se utilizó una fecha de corte fija para el cálculo de variables temporales, evitando diferencias entre la información utilizada durante el entrenamiento y aquella utilizada en la predicción individual.

La validación de las variables históricas incluyó la verificación de que el préstamo analizado no estuviera incluido en su propio historial. Esta medida es necesaria para evitar que el modelo incorpore información del resultado que pretende predecir. La metodología implementada busca garantizar que el sistema aprenda patrones de riesgo existentes antes de la concesión del crédito y no señale posteriores asociadas al incumplimiento.

3.6. Análisis exploratorio de datos

El análisis exploratorio de datos consistió en comprender la estructura de la base de datos, la distribución de la variable objetivo y las relaciones entre las variables predictoras. Se analizaron la frecuencia de préstamos con y sin mora, la tasa de incumplimiento por producto crediticio, el comportamiento de la mora según historial de alerta y la distribución de variables financieras como ingresos, deuda inicial, saldo actual, valor de cuota, plazo y ratios de endeudamiento.

La elaboración de una matriz de correlación para identificar relaciones lineales entre variables numéricas permitió detectar grupos de variables relacionadas con el saldo, la deuda, las cuotas, los pagos y el historial crediticio. La identificación de correlaciones elevadas resulta relevante para revisar posibles redundancias y controlar la multicolinealidad, especialmente en el modelo de Regresión Logística.

En el caso de Random Forest, la presencia de variables correlacionadas tiene un efecto menos crítico, ya que el algoritmo construye múltiples árboles a partir de subconjuntos aleatorios de observaciones y variables.

El análisis exploratorio permitió, además, identificar la necesidad de utilizar métricas adicionales a la exactitud. Dado que la tasa de mora fue de 23,47 %, un modelo que predijera mayoritariamente la clase sin mora podría mostrar una exactitud aparentemente aceptable, pero tendría poca utilidad para identificar créditos riesgosos. Por esta razón, se priorizaron métricas como precision, recall, F1-score, AUC-ROC y matriz de confusión.

3.7. División de datos y estrategia de validación

La división de los datos se realizó considerando la dependencia entre las operaciones de un mismo socio, dado que un socio puede tener más de un préstamo; por ello, se utilizó una estrategia de agrupamiento mediante GroupShuffleSplit, tomando como variable de agrupación el identificador del socio NSOCIO. Esta técnica permite evitar que préstamos pertenecientes al mismo socio se ubiquen simultáneamente en los conjuntos de entrenamiento y de prueba, lo que reduce el riesgo de sobreestimación del desempeño debido a la similitud entre las observaciones.

En la implementación desarrollada, la base de datos se dividió en dos conjuntos: entrenamiento y prueba.

El conjunto de entrenamiento representó aproximadamente el 70 % de los datos y fue utilizado para ajustar los modelos de Regresión Logística, Random Forest y XGBoost.

El conjunto de prueba representó aproximadamente el 30 % restante y se reservó para evaluar el desempeño final de los modelos mediante métricas como exactitud, precisión, recall, F1-score, AUC-ROC y la matriz de confusión.

La separación se realizó después del proceso ETL y antes del entrenamiento de los modelos, garantizando que la evaluación se efectuara sobre observaciones no utilizadas durante el ajuste del modelo. En esta versión no se implementaron validaciones temporales ni un conjunto de validación independiente para la optimización de hiperparámetros ni para la selección de umbrales. Por tanto, la validación temporal se plantea como una mejora futura, ya que permitiría entrenar los modelos con préstamos otorgados en períodos anteriores y evaluarlos con préstamos posteriores, lo que simularía mejor el escenario real de uso del modelo en nuevas operaciones crediticias.

1. **División por socio:** Al utilizar GroupShuffleSplit por NSOCIO, impide que préstamos del mismo socio aparezcan en entrenamiento y prueba.

Tabla 3. División de datos por socio

Conjunto	Proporción	Finalidad
Entrenamiento	70 %	Ajustar los modelos de aprendizaje automático
Prueba	30 %	Evaluar el desempeño final con datos no utilizados durante el entrenamiento.

Nota. Información elaborada a partir de datos institucionales

3.8. Tratamiento del desbalance de clases

La tasa de mora de 23,47 % evidencia la presencia de un desbalance moderado entre las clases de cumplimiento e incumplimiento, debido a que los préstamos sin mora representan la clase mayoritaria y los préstamos con mora constituyen la clase minoritaria. Por esta razón, la evaluación de los modelos no se basó únicamente en la exactitud, sino también en métricas más sensibles al comportamiento de la clase minoritaria, como la precisión, el recall, el F1-score, el AUC-ROC y la matriz de confusión.

Para reducir el efecto del desbalance en los modelos, se aplicó ponderación de clases en los algoritmos que lo permiten directamente dentro de su configuración. En la Regresión Logística y en Random Forest se utilizó el parámetro `class_weight="balanced"`, con el propósito de asignar mayor peso relativo a la clase minoritaria durante el entrenamiento. Esta estrategia busca evitar que los modelos favorezcan excesivamente la clase de préstamos sin mora y mejorar la identificación de operaciones con posible incumplimiento.

En el caso de XGBoost, en la versión implementada no se aplicó una técnica explícita de balanceo como `scale_pos_weight` ni sobremuestreo mediante SMOTE. El modelo fue entrenado con la distribución original de los datos, evaluando su desempeño mediante métricas complementarias a la exactitud. Esta decisión permite comparar el comportamiento de XGBoost frente a los demás modelos bajo la distribución real de la cartera.

No se aplicaron técnicas de sobremuestreo o submuestreo antes de dividir los datos, con el fin de evitar contaminación entre los conjuntos de entrenamiento y prueba. Por tanto, el tratamiento del desbalance se limitó a la ponderación de clases en Regresión Logística y Random

Forest, mientras que la comparación final consideró especialmente la capacidad de cada modelo para detectar préstamos en mora y controlar las alertas incorrectas.

3.9. Modelos de aprendizaje automático

Se evaluaron tres algoritmos de clasificación supervisada: Regresión Logística, Random Forest y XGBoost. La Regresión Logística se utilizó como modelo base, debido a su simplicidad, interpretabilidad y utilidad para establecer una línea de comparación inicial. Sin embargo, su capacidad para representar relaciones complejas entre variables es limitada, especialmente cuando el incumplimiento depende de interacciones no lineales entre ingresos, endeudamiento, historial crediticio y características del préstamo.

Random Forest se incorporó como modelo de ensamble comparativo, debido a su capacidad para identificar patrones no lineales, interacciones complejas y relaciones entre múltiples variables sin requerir una especificación manual de dichas interacciones; este algoritmo genera múltiples árboles de decisión utilizando muestras aleatorias de observaciones y subconjuntos de variables, lo que reduce el riesgo de sobreajuste y mejora la estabilidad de las predicciones; además, presenta tolerancia frente a variables correlacionadas y puede manejar adecuadamente conjuntos de datos con características heterogéneas.

XGBoost se utilizó como modelo candidato de alto rendimiento, ya que su enfoque de gradient boosting permite mejorar progresivamente la capacidad predictiva mediante la corrección de los errores cometidos por los árboles anteriores. Su evaluación no se basó exclusivamente en la exactitud ni en el AUC-ROC, sino también en la precisión, el recall, el F1-score, la matriz de confusión y la utilidad operativa de sus predicciones para la gestión del riesgo

crediticio. En la implementación desarrollada, los tres modelos se entrenaron y se compararon bajo el mismo flujo de preprocesamiento y evaluación.

3.10. Evaluación, calibración y selección del modelo

La evaluación de los modelos se realizó mediante las métricas de exactitud, precisión, recall, F1-score, AUC-ROC y matriz de confusión. La exactitud permitió conocer el porcentaje total de clasificaciones correctas; la precisión indicó qué proporción de préstamos clasificados como riesgosos realmente incurrió en mora; el recall midió la capacidad del modelo para identificar préstamos que efectivamente presentaron incumplimiento; y el F1-score integró la precisión y el recall en una sola medida de equilibrio.

En el contexto de riesgo crediticio, los falsos negativos son especialmente relevantes, debido a que representan préstamos clasificados como de bajo riesgo que posteriormente alcanzan una mora significativa. Por esta razón, la selección del modelo no se fundamentó únicamente en la exactitud o en el AUC-ROC, sino en un análisis integral de las métricas, la matriz de confusión y la capacidad de detectar operaciones riesgosas sin generar un volumen excesivo de alertas incorrectas.

En la versión implementada, los modelos generaron probabilidades de incumplimiento mediante `predict_proba`, las cuales fueron utilizadas para calcular el AUC-ROC y apoyar la clasificación del riesgo. No obstante, como mejora para una implementación operativa, se recomienda calibrar dichas probabilidades mediante técnicas como calibración isotónica o sigmoide, por ejemplo, con `CalibratedClassifierCV`. Esta calibración permitiría que una probabilidad estimada, como 60 %, tenga una correspondencia más cercana con la frecuencia real de incumplimiento observada en la cartera. Posteriormente, los niveles de riesgo bajo, medio

y alto deberían definirse a partir de probabilidades calibradas, la distribución real de la mora, el apetito de riesgo institucional y la capacidad operativa del área de crédito y cobranza.

3.11. Explicabilidad de los modelos

La investigación incorpora técnicas de inteligencia artificial explicable con el propósito de interpretar los factores que influyen en las predicciones. Se propone utilizar SHAP como herramienta principal para analizar la importancia global de las variables y explicar casos individuales. A nivel global, SHAP permite identificar cuáles variables contribuyen con mayor frecuencia al incremento o reducción del riesgo de incumplimiento. A nivel individual, permite explicar por qué un préstamo específico fue clasificado con una determinada probabilidad de mora.

Este análisis permitió identificar las variables con mayor contribución en la clasificación del riesgo de incumplimiento, tales como características del producto crediticio, línea de crédito, historial previo del socio, número de cuotas, endeudamiento y condiciones asociadas al perfil financiero del solicitante. La interpretación de estas variables facilita la revisión de la coherencia del modelo con la lógica del riesgo crediticio y permite detectar posibles variables que requieran revisión técnica por su alta influencia en la predicción.

3.12. Validación de negocio, ética y reproducibilidad

Como parte de la validación de negocio, se contrastaron los resultados del modelo con el criterio de personal del área de crédito, considerando la experiencia de oficiales y jefes de crédito en la evaluación y seguimiento de operaciones crediticias. Esta revisión permitió valorar la coherencia de la clasificación de riesgo, la razonabilidad de las probabilidades de

incumplimiento estimadas y la utilidad del modelo como herramienta de apoyo para la gestión preventiva del riesgo crediticio. Los resultados de la encuesta aplicada evidenciaron una valoración favorable, ya que los participantes consideraron que las predicciones generadas por el sistema son consistentes con su experiencia profesional y pueden aportar al proceso de evaluación crediticia.

La investigación consideró principios de privacidad, seguridad y uso responsable de datos durante la preparación y modelado de la información. Los identificadores directos de los socios, como número de socio, identificación, nombre, número de préstamo y número de operación, fueron excluidos del conjunto de variables predictoras, evitando que el modelo aprenda patrones asociados a datos personales y no a factores reales de riesgo. Además, se excluyeron variables que podían revelar directamente el resultado futuro del préstamo, tales como días máximos de mora, cuotas en mora, mora pagada, estado del préstamo, calificaciones posteriores y saldo actual.

El modelo fue concebido como una herramienta de apoyo a la decisión humana y no como un mecanismo automático de aprobación o rechazo de créditos. Por ello, sus resultados deben ser interpretados por los analistas de crédito junto con la información financiera, documental y normativa utilizada en el proceso institucional de evaluación. Esta orientación permite fortalecer la transparencia, reducir decisiones injustificadas y mantener la supervisión humana en la gestión del riesgo crediticio.

La reproducibilidad del proceso se fortaleció mediante el uso de Python y la documentación de parámetros relevantes del experimento, como la fecha de corte, el umbral de mora de treinta días, las variables utilizadas, la división de datos, las semillas aleatorias, los

hiperparámetros de los modelos y las métricas obtenidas. Además, los modelos entrenados, matrices de confusión, curvas ROC, metadatos y resultados de métricas fueron almacenados en archivos, lo que permite repetir el experimento, auditar los resultados y actualizar el modelo de manera controlada.

CAPÍTULO 4

4. Análisis de resultados

El presente capítulo expone los resultados obtenidos durante el desarrollo de los modelos de aprendizaje automático orientados a predecir el incumplimiento de pagos en operaciones crediticias de una cooperativa de ahorro y crédito ecuatoriana. El análisis se realizó a nivel de préstamo, debido a que esta unidad permite estimar el riesgo de cada operación crediticia de manera individual y reduce el riesgo de duplicidad que podría generarse al trabajar a nivel de cuotas. Para la construcción del modelo se excluyeron como variables predictoras aquellos campos que revelaban directamente el resultado del préstamo, tales como días máximos de mora, cuotas en mora, mora pagada, estado del préstamo, calificaciones posteriores y saldo actual.

La investigación comparó tres modelos de clasificación supervisada: Regresión Logística, Random Forest y XGBoost. La evaluación se efectuó mediante las métricas de exactitud, precisión, recall, F1-score, AUC-ROC y matriz de confusión. La selección del modelo final se basó en un análisis integral de estas métricas, priorizando la capacidad discriminativa, la precisión de las alertas de riesgo y el equilibrio entre la identificación de incumplimientos y la reducción de clasificaciones incorrectas.

4.1. Caracterización del conjunto de datos

La base analítica estuvo conformada por 75.783 operaciones crediticias correspondientes a 46.249 socios. La información fue consolidada a nivel de préstamo y procesada mediante un procedimiento ETL que permitió transformar 44 variables iniciales en 71 variables analíticas. Las variables incluyeron información relacionada con el perfil del socio, características del crédito, capacidad de pago, endeudamiento e historial crediticio. Estos valores coinciden con los

resultados generados en la aplicación, donde se reportan 75.783 préstamos, 46.249 socios, 44 variables originales, 71 variables luego del ETL y una tasa de mora de 23,47 %.

Indicadores del Dataset por Préstamo

Préstamos: 75.783

Socios: 46.249

Variables: 71

La variable objetivo fue denominada MORA_BINARIA y se definió con el valor de uno para préstamos que registraron una mora máxima igual o superior a treinta días, y cero para operaciones que no alcanzaron dicho umbral. La tasa de incumplimiento identificada fue de 23,47 %, lo que evidencia una distribución desbalanceada, aunque adecuada para el desarrollo de modelos de clasificación supervisada. La proporción de préstamos con mora justifica el uso de métricas adicionales a la exactitud, dado que un modelo podría clasificar correctamente la mayoría de los préstamos sin mora y, aun así, presentar una capacidad limitada para detectar los casos de incumplimiento.

El proceso de preparación de datos incluyó depuración de registros, tratamiento de valores faltantes, conversión de fechas, normalización de variables categóricas, creación de ratios financieros y construcción de variables históricas. Entre las variables derivadas se incluyeron la edad del socio, antigüedad como socio, cuota estimada, relación cuota-ingreso, relación deuda-ingreso, indicadores de ingresos insuficientes, préstamos previos, préstamos morosos, préstamos vencidos, alertas crediticias, préstamos judiciales, prejudiciales y castigados.

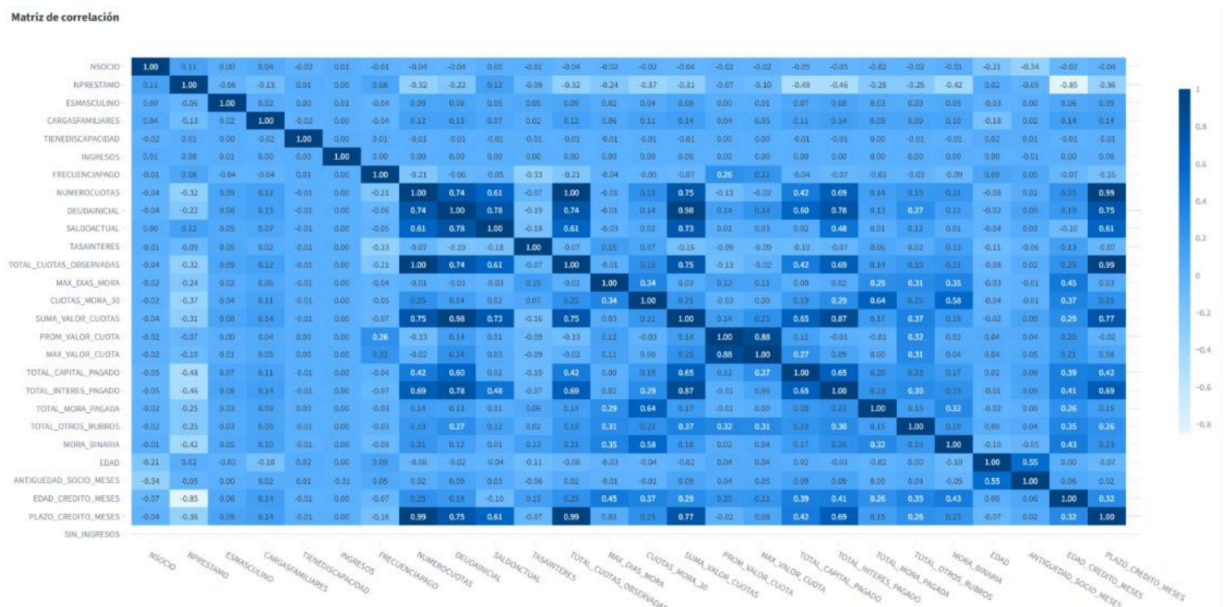
Figura 4. Datos - variables históricas

	Columna	Tipo	Nulos	% Nulos
NSOCIO	NSOCIO	int64	0	0
IDENTIFICACION	IDENTIFICACION	str	0	0
NOMBREUNIDO	NOMBREUNIDO	str	0	0
NPRESTAMO	NPRESTAMO	int64	0	0
NUMEROPRESTAMO	NUMEROPRESTAMO	str	0	0
FECHAINGRESOCIENTE	FECHAINGRESOCIENTE	datetime64[us]	0	0
CIUDADINGRESOCIENTE	CIUDADINGRESOCIENTE	str	0	0
OFICINACIENTE	OFICINACIENTE	str	0	0
TIPOIDENTIFICACION	TIPOIDENTIFICACION	str	0	0

4.2. Análisis exploratorio de la mora crediticia

El análisis exploratorio permitió identificar que el incumplimiento de pagos no depende de una única característica, sino de la interacción entre condiciones personales, financieras y crediticias.

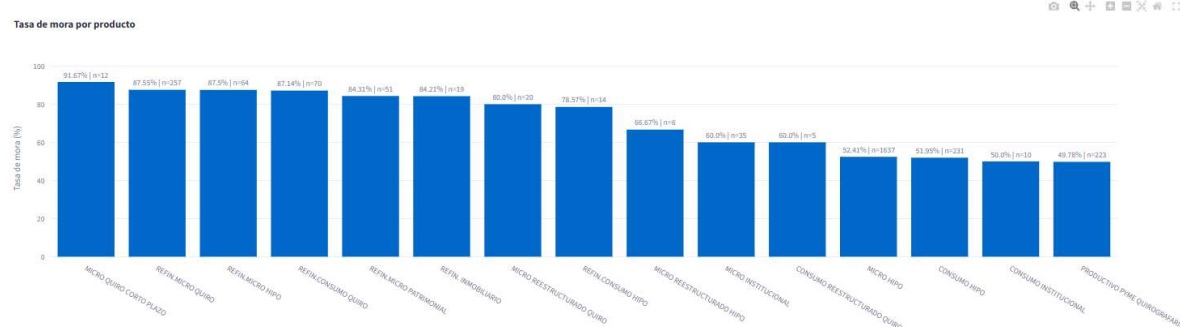
Figura 5. Matriz de correlación



Los préstamos con mayor exposición al riesgo presentaron patrones asociados a niveles elevados de endeudamiento, cuotas altas en relación con los ingresos, historial de alertas y antecedentes de mora. Estos resultados respaldan la necesidad de utilizar modelos capaces de identificar relaciones no lineales e interacciones complejas entre variables.

Figura 6. Tasa de mora por producto

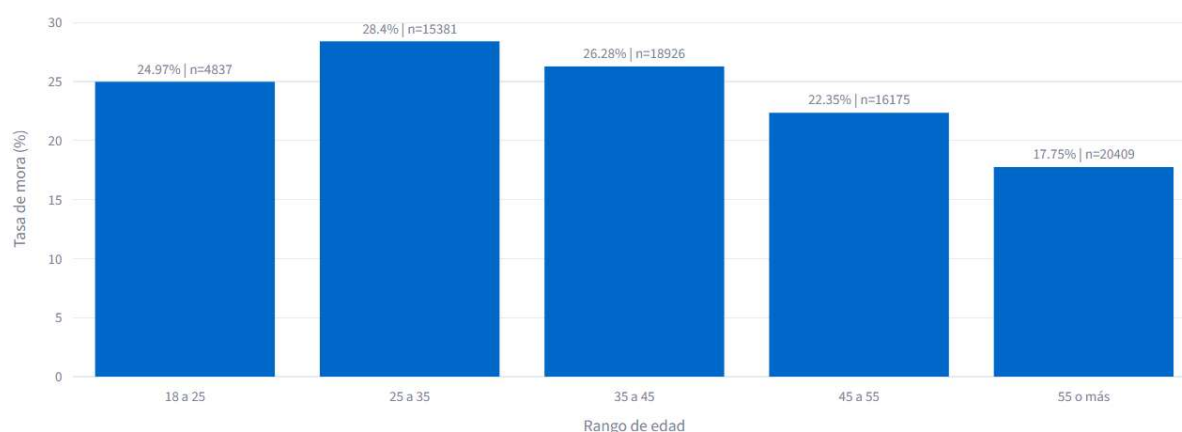
Tasa de mora por producto



La figura presenta la tasa de mora por producto, considerando las categorías con mayor proporción de incumplimiento luego de omitir productos asociados al proceso de migración del sistema. Esta depuración se realizó porque, durante la migración, no se trasladaron de forma completa los préstamos cancelados previamente, sino principalmente operaciones con saldos, atrasos o pendientes de pago, lo que podía generar tasas de mora artificialmente elevadas.

En el gráfico se observa que productos como Micro Quiro Corto Plazo, Refin Micro Quiro, Refin Micro Hipo y Refin Consumo Quiro presentan las tasas más altas dentro de los productos analizados. Asimismo, la etiqueta n permite identificar el número de préstamos considerados por producto, lo cual facilita interpretar los porcentajes de acuerdo con el volumen de observaciones.

Figura 7. Tasa de mora por edad



La figura evidencia que la tasa de mora varía según el rango de edad de los socios. El grupo de 25 a 35 años registra la mayor proporción de incumplimiento, con 28,40 %, seguido por el rango de 35 a 45 años, con 26,28 %, y el grupo de 18 a 25 años, con 24,97 %. En contraste, los socios de 45 a 55 años presentan una tasa de mora de 22,35 %, mientras que el grupo de 55 o más años registra el nivel más bajo, con 17,75 %.

Estos resultados muestran una tendencia decreciente de la mora a medida que aumenta la edad, lo que sugiere que los socios más jóvenes pueden presentar una mayor exposición al incumplimiento. No obstante, la edad debe interpretarse como una variable complementaria, ya que el riesgo crediticio depende también del historial del socio, el endeudamiento, la cuota del préstamo y las características de la operación.

Figura 8. Mora Binaria



La distribución de la variable MORA_BINARIA mostró que aproximadamente uno de cada cuatro préstamos alcanzó una mora igual o superior a treinta días. Este resultado confirma que la mora constituye un fenómeno relevante para la gestión de cartera y que la cooperativa requiere mecanismos de identificación temprana. Asimismo, el análisis de correlación evidenció asociaciones entre variables financieras como monto, cuota, plazo, deuda e ingresos, lo cual demuestra la existencia de relaciones complejas que no pueden ser explicadas únicamente mediante modelos lineales.

La construcción de variables de historial crediticio permitió incorporar información relevante sobre el comportamiento previo de los socios. Variables como préstamos morosos, vencidos, judiciales, prejudiciales y castigados permitieron representar antecedentes que pueden incrementar la probabilidad de incumplimiento. Para garantizar la validez metodológica, estas variables fueron construidas considerando el historial previo del socio y excluyendo del entrenamiento aquellas variables que revelaban directamente el resultado futuro del préstamo evaluado.

4.3. Resultados de los modelos de clasificación

La Regresión Logística fue utilizada como modelo base por su simplicidad e interpretabilidad. Sin embargo, los resultados evidenciaron limitaciones en su capacidad para representar relaciones complejas entre las variables predictoras y el incumplimiento. El modelo obtuvo una exactitud de 34,75 %, precisión de 25,18 %, recall de 91,44 %, F1-score de 39,49 % y AUC-ROC de 0,6580. Aunque el recall fue alto, la baja precisión indica que una proporción considerable de los préstamos clasificados como riesgosos no incurrió realmente en mora. Esto generó un número elevado de alertas falsas y limitó su utilidad operativa.

El modelo Random Forest presentó un desempeño superior al de la Regresión Logística. Alcanzó una exactitud de 74,86 %, precisión de 47,51 %, recall de 76,08 %, F1-score de 58,49 % y AUC-ROC de 0,8335. Estos resultados evidencian que Random Forest logró identificar una proporción importante de préstamos en mora y, al mismo tiempo, redujo el número de falsas alertas respecto del modelo lineal. Su desempeño confirma la utilidad de los modelos de ensamble para representar relaciones no lineales entre variables financieras, crediticias e históricas.

XGBoost obtuvo el mejor desempeño global entre los modelos evaluados. El modelo alcanzó una exactitud de 84,30 %, precisión de 73,87 %, recall de 50,41 %, F1-score de 59,93 % y AUC-ROC de 0,8797. La exactitud obtenida indica que XGBoost clasificó correctamente una mayor proporción de operaciones crediticias en comparación con los otros modelos. La precisión de 73,87 % demuestra que una proporción elevada de los préstamos identificados como riesgosos efectivamente alcanzó una mora significativa. Asimismo, el AUC-ROC de 0,8797 evidencia una alta capacidad del modelo para discriminar entre préstamos con y sin incumplimiento.

La comparación de los resultados permite observar que XGBoost superó a Random Forest en exactitud, precisión, F1-score y AUC-ROC. Aunque Random Forest obtuvo un recall superior, XGBoost generó menos falsas alertas y presentó una capacidad discriminativa más elevada. En un escenario institucional, esta característica resulta relevante porque permite focalizar los recursos de seguimiento, evaluación y cobranza en operaciones que presentan una probabilidad más alta de incumplimiento. No obstante, debido a que XGBoost detecta menos casos reales de mora que Random Forest, se recomienda ajustar el umbral de decisión según el apetito de riesgo de la cooperativa.

4.4. Análisis de matrices de confusión

La matriz de confusión de la Regresión Logística mostró que el modelo identificó un número elevado de préstamos en mora, pero también clasificó incorrectamente una gran cantidad de operaciones cumplidas como riesgosas. Esta situación explica su bajo nivel de precisión y limita su aplicación práctica, debido a que la cooperativa tendría que destinar recursos a revisar un volumen excesivo de alertas que no necesariamente representan riesgo real.

Random Forest presentó una matriz de confusión más equilibrada. El modelo identificó correctamente 4.007 préstamos con mora y 12.931 préstamos sin mora. Asimismo, generó 4.427 falsos positivos y 1.260 falsos negativos. Estos resultados reflejan una capacidad importante para detectar incumplimientos, aunque todavía genera un número considerable de alertas que deben ser revisadas por el área de crédito.

XGBoost presentó la menor cantidad de falsos positivos entre los modelos de ensamble evaluados. El modelo identificó correctamente 2.655 préstamos con mora y 16.419 préstamos sin mora, mientras que generó 939 falsos positivos y 2.612 falsos negativos. Esta configuración

muestra que XGBoost es más selectivo al clasificar una operación como riesgosa. Por tanto, cuando el modelo emite una alerta de alto riesgo, existe una mayor probabilidad de que dicha alerta corresponda a un préstamo que efectivamente presentará incumplimiento.

La reducción de falsos positivos en XGBoost evidencia que el modelo genera alertas de riesgo más precisas, es decir, clasifica como riesgosas menos operaciones que realmente no presentaron mora. Esto representa una ventaja frente a Random Forest, que detecta más préstamos morosos, pero también produce un mayor número de alertas incorrectas. Sin embargo, XGBoost registra una mayor cantidad de falsos negativos, lo que indica que algunos préstamos que sí incurrieron en mora fueron clasificados inicialmente como no riesgosos. En conjunto, estos resultados muestran un equilibrio diferente entre ambos modelos: Random Forest prioriza la detección de más casos de mora, mientras que XGBoost prioriza la calidad de las alertas y la reducción de clasificaciones incorrectas. Por esta razón, XGBoost se considera el modelo con mejor desempeño global, aunque su interpretación debe considerar que no identifica todos los casos reales de incumplimiento.

4.5. Selección de XGBoost como modelo final

La selección de XGBoost como modelo final se fundamenta en su desempeño integral. El modelo obtuvo la mayor exactitud, la mayor precisión, el mayor F1-score y el mayor AUC-ROC entre los algoritmos evaluados. Estos resultados demuestran que XGBoost presenta una mayor capacidad para distinguir entre préstamos con y sin incumplimiento, además de reducir de forma importante las alertas falsas frente a los demás modelos.

El AUC-ROC de 0,8797 indica que el modelo posee una capacidad discriminativa alta. En términos prácticos, esto significa que XGBoost logra asignar mayores probabilidades de

riesgo a préstamos que posteriormente presentan mora y menores probabilidades a aquellos que cumplen con sus obligaciones. Esta capacidad resulta relevante para clasificar operaciones según niveles de riesgo y apoyar el análisis preventivo de la cartera crediticia.

La precisión de 73,87 % representa una ventaja importante en la calidad de las alertas generadas. Esto significa que, de los préstamos clasificados como riesgosos por XGBoost, aproximadamente siete de cada diez efectivamente presentaron mora. En comparación con Random Forest, XGBoost genera menos falsas alertas, por lo que sus predicciones positivas son más confiables para identificar operaciones con mayor probabilidad de incumplimiento.

Aunque XGBoost presentó un recall inferior al de Random Forest, este resultado evidencia una diferencia en el comportamiento de ambos modelos. Random Forest identifica una mayor cantidad de préstamos morosos, pero también genera más alertas incorrectas. En cambio, XGBoost es más selectivo al clasificar una operación como riesgosa, lo que mejora la precisión de sus alertas, aunque deja sin identificar algunos casos reales de mora. Por esta razón, XGBoost se considera el modelo con mejor desempeño global, al presentar un mejor equilibrio entre capacidad discriminativa, precisión de las alertas y reducción de clasificaciones incorrectas

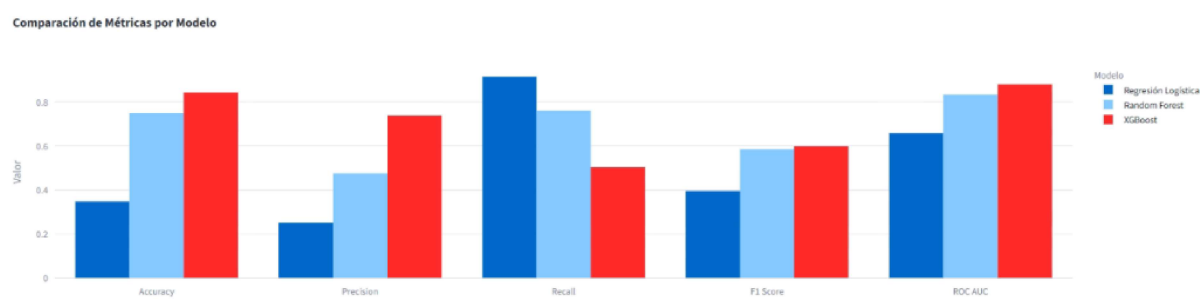
Figura 9. Diferencia entre modelos

	Modelo	Accuracy	Precision	Recall	F1 Score	ROCAUC	VN	FP	FN	VP
0	Regresión Logística	0.3475	0.2518	0.9144	0.3949	0.658	3047	14311	451	4816
1	Random Forest	0.7486	0.4751	0.7608	0.5849	0.8335	12931	4427	1260	4007
2	XGBoost	0.843	0.7387	0.5041	0.5993	0.8797	16419	939	2612	2655

Los resultados del entrenamiento muestran diferencias relevantes entre los modelos evaluados. La Regresión Logística obtuvo el desempeño más limitado, con una exactitud de 34,75 %, precisión de 25,18 %, recall de 91,44 %, F1-score de 39,49 % y AUC-ROC de 0,6580.

Aunque presentó un recall elevado, su baja precisión evidencia que clasificó como riesgosos muchos préstamos que realmente no incurrieron en mora, generando un alto número de falsas alertas. Por su parte, Random Forest mejoró de manera importante el desempeño general, alcanzando una exactitud de 74,86 %, precisión de 47,51 %, recall de 76,08 %, F1-score de 58,49 % y AUC-ROC de 0,8335, destacándose por su mayor capacidad para identificar préstamos que efectivamente presentaron incumplimiento. Finalmente, XGBoost obtuvo el mejor desempeño global, con una exactitud de 84,30 %, precisión de 73,87 %, recall de 50,41 %, F1-score de 59,93 % y AUC-ROC de 0,8797. Estos resultados evidencian que XGBoost presentó la mayor capacidad discriminativa y generó alertas de riesgo más precisas. Sin embargo, su menor recall frente a Random Forest indica que el modelo es más selectivo al clasificar operaciones como riesgosas, por lo que identifica menos casos reales de mora, pero reduce de forma importante las clasificaciones incorrectas. En conjunto, XGBoost se considera el modelo con mejor desempeño global por su equilibrio entre exactitud, precisión, F1-score y capacidad discriminativa.

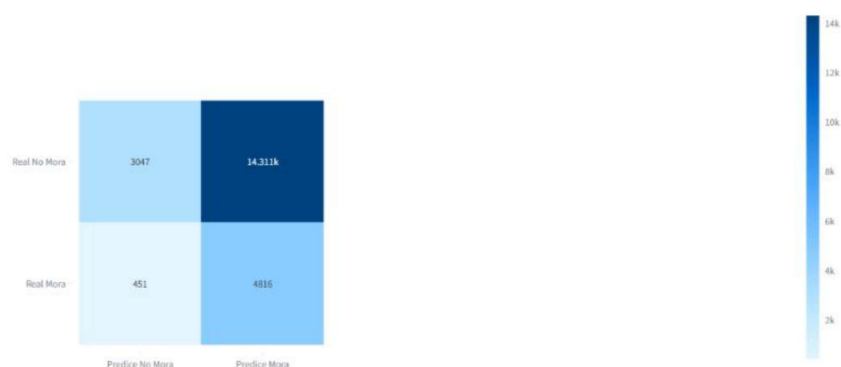
Figura 10. Análisis de modelos



La matriz de confusión de la Regresión Logística muestra que el modelo identificó correctamente 4.816 préstamos con mora, pero solo clasificó adecuadamente 3.047 préstamos sin mora. Además, generó 14.311 falsos positivos, es decir, un número elevado de operaciones cumplidas que fueron clasificadas erróneamente como riesgosas, y 451 falsos negativos. Esto explica su alto recall, pues detecta gran parte de los casos reales de mora, pero también su baja precisión y exactitud, debido a que emite demasiadas alertas incorrectas. Por tanto, la Regresión Logística resulta útil como modelo de referencia, pero presenta limitaciones para su aplicación operativa frente a Random Forest y XGBoost.

Figura 11. Matriz de confusión y regresión logística

Matriz de Confusión - Regresión Logística



La matriz de confusión de Random Forest evidencia que el modelo clasificó correctamente 12.931 préstamos sin mora y 4.007 préstamos con mora, demostrando una capacidad importante para identificar operaciones riesgosas. Sin embargo, registró 4.427 falsos positivos, es decir, préstamos que no incumplieron pero fueron clasificados como morosos, y 1.260 falsos negativos, correspondientes a créditos con mora real que el modelo no detectó. En comparación con XGBoost, Random Forest genera más alertas falsas, pero identifica una mayor

cantidad de casos reales de incumplimiento; por ello, constituye una alternativa útil cuando la cooperativa prioriza la detección preventiva de créditos riesgosos.

Figura 12. Matriz de confusión - Random Forest

Matriz de Confusión - Random Forest



La matriz de confusión de XGBoost muestra que el modelo clasificó correctamente 16.419 préstamos sin mora y 2.655 préstamos con mora, lo que evidencia una elevada capacidad para reconocer operaciones de cumplimiento y explica su alta exactitud y precisión.

Asimismo, generó 939 falsos positivos, es decir, préstamos sin mora clasificados como riesgosos, y 2.612 falsos negativos, correspondientes a operaciones que realmente incurrieron en mora pero fueron predichas como no riesgosas. Por tanto, aunque XGBoost presenta un desempeño global superior, la cantidad de falsos negativos confirma la necesidad de calibrar las probabilidades y ajustar el umbral de decisión para mejorar la detección temprana de créditos con riesgo de incumplimiento.

Figura 13. Matriz de confusión - XG Boost

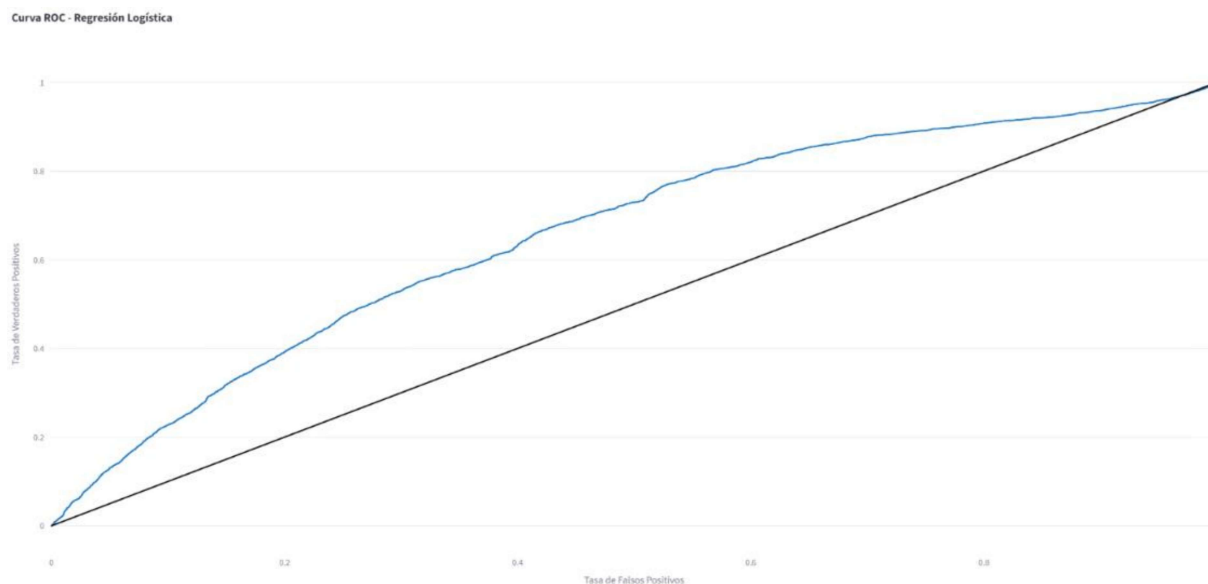
Matriz de Confusión - XGBoost



La curva ROC de la Regresión Logística se mantiene por encima de la línea diagonal de referencia, aunque con una separación limitada, lo que indica una capacidad discriminativa moderada-baja para diferenciar entre préstamos que incumplen y aquellos que cumplen sus obligaciones.

Aunque el modelo logra identificar ciertos patrones asociados a la mora, su desempeño es inferior al de los modelos de ensamble evaluados. Este resultado es consistente con su AUC-ROC de 0,6580, valor menor al obtenido por Random Forest y XGBoost. Por ello, la Regresión Logística se considera útil como modelo base de comparación, pero no como la alternativa más adecuada para la predicción operativa del incumplimiento crediticio.

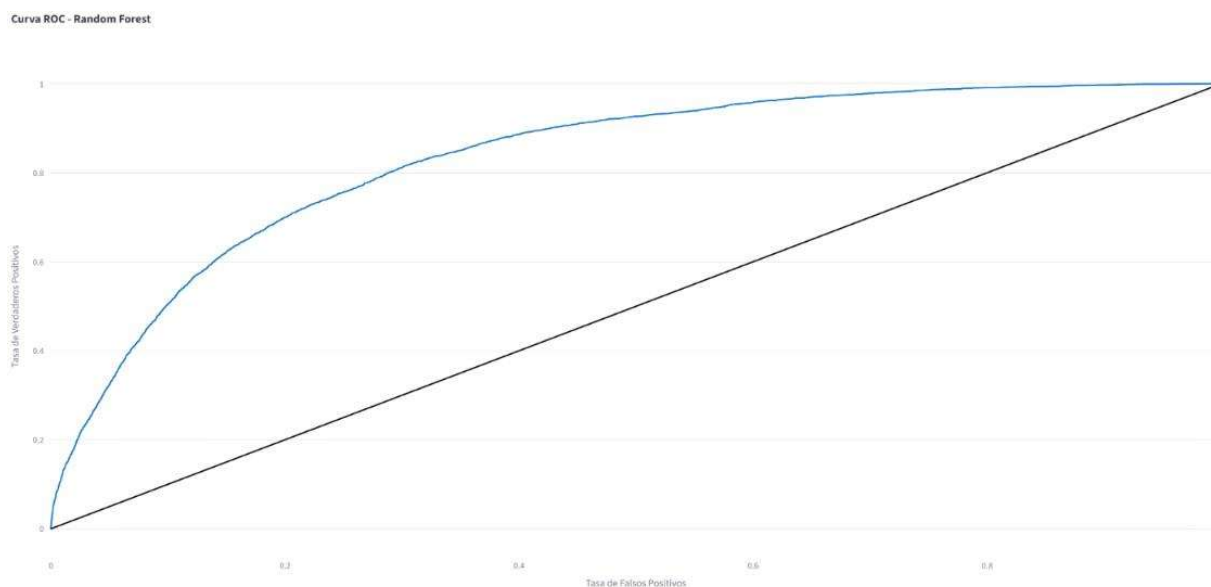
Figura 14. Curva ROC Regresión Logística



La curva ROC de Random Forest se ubica de forma consistente por encima de la línea diagonal de referencia, lo que evidencia una adecuada capacidad para distinguir entre préstamos con incumplimiento y préstamos sin mora.

La curva presenta un incremento sostenido de la tasa de verdaderos positivos a medida que aumenta la tasa de falsos positivos, comportamiento coherente con un AUC-ROC de 0,8335. Aunque este resultado es inferior al obtenido por XGBoost, confirma que Random Forest posee un desempeño discriminativo alto y constituye un modelo robusto para la identificación de riesgo crediticio, especialmente por su mayor capacidad de detección de casos reales de mora.

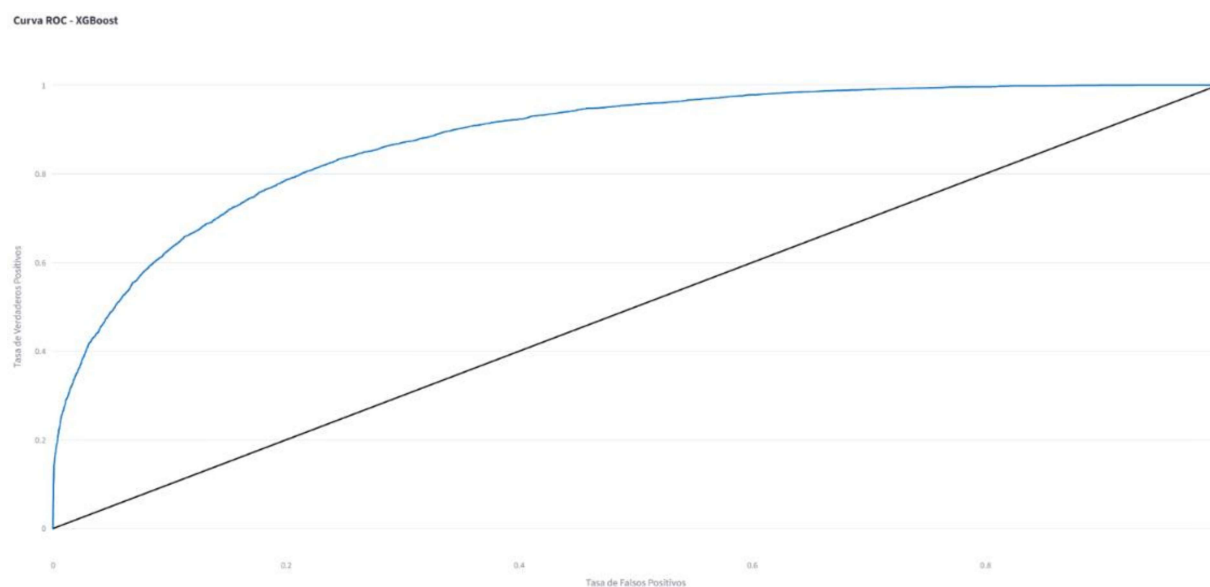
Figura 15. Curva ROC Random Forest



La curva ROC del modelo XGBoost se mantiene claramente por encima de la línea diagonal de referencia, lo que demuestra que el modelo posee una adecuada capacidad para diferenciar entre préstamos que incumplen y aquellos que no presentan mora.

La elevación de la curva en los primeros valores de la tasa de falsos positivos indica que el algoritmo logra identificar casos de incumplimiento manteniendo un nivel controlado de clasificaciones erróneas. Este comportamiento es coherente con el AUC-ROC obtenido de 0,8797, valor que confirma un desempeño discriminativo alto y respalda la selección de XGBoost como el modelo con mejor rendimiento global para la predicción de incumplimiento crediticio.

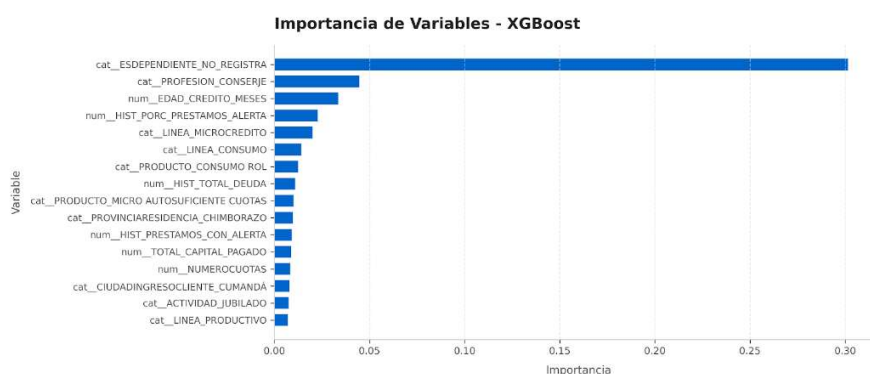
Figura 16. Curva ROC - XG Boost



El análisis de importancia de variables del modelo XGBoost evidencia que las variables con mayor contribución predictiva se relacionan principalmente con la línea de crédito, la actividad económica del socio, la ciudad de ingreso del cliente, el número de cuotas, el historial de alertas, la deuda histórica y el producto crediticio. Entre las variables más relevantes se identificaron categorías como:

Figura 17. Variables - XGBoost**Top 30 variables más importantes - XGBoost**

	Variable	Importancia
538	cat_ESDEPENDIENTE_NO_REGISTRA	0.3015
256	cat_PROFESION_CONSERJE	0.0447
536	cat_ESDEPENDIENTE_0	0.0379
17	num_EDAD_CREDITO_MESES	0.0336
537	cat_ESDEPENDIENTE_1	0.0312
34	num_HIST_PORC_PRESTAMOS_ALERTA	0.0228
14	num_TOTAL_OTROS_RUBROS	0.0200
1539	cat_LINEA_MICROCREDITO	0.0200
179	cat_TIPOVIVIENDA_PROPIA NO HIPOTECADA	0.0177
1533	cat_LINEA_CONSUMO	0.0142



Estos resultados muestran que el modelo no basa su predicción en una sola variable, sino en la combinación de características asociadas al tipo de crédito, condiciones de la operación, comportamiento histórico del socio y variables relacionadas con su perfil económico. En particular, la presencia de variables de historial, como préstamos con alerta, porcentaje histórico de préstamos con alerta y deuda histórica, evidencia que los antecedentes crediticios tienen un peso relevante en la estimación del riesgo de incumplimiento. Asimismo, las variables relacionadas con línea y producto crediticio permiten diferenciar patrones de mora según las características propias de cada operación.

La importancia de variables fue utilizada como una herramienta de interpretación global del modelo, permitiendo identificar los factores que más aportan a la clasificación realizada por XGBoost. Este análisis contribuye a verificar que el modelo considere elementos coherentes con la lógica del riesgo crediticio, como historial previo, endeudamiento, número de cuotas, tipo de producto y características económicas del socio.

Como parte de la validación con personal del área de crédito, se utilizó la pestaña de Predicción Individual del sistema desarrollado. Esta pantalla permitió consultar el historial crediticio de un socio, ingresar las condiciones de un nuevo préstamo y obtener como resultado la probabilidad estimada de incumplimiento y su clasificación de riesgo. Los resultados generados en esta funcionalidad sirvieron como insumo para que los oficiales y jefes de crédito evaluaran, mediante la encuesta aplicada, la coherencia de la predicción, la utilidad del modelo y su correspondencia con la experiencia profesional en el análisis crediticio.

SOCIO CON RIESGO ALTO

La simulación del nuevo préstamo evidencia una probabilidad de incumplimiento de 77,07 %, por lo que el modelo lo clasifica como de riesgo alto y predice que la operación podría incumplir. Este resultado se explica principalmente por el historial del socio, que registra un préstamo previo con alerta y un crédito castigado, configurando un indicador de riesgo histórico severo; por ello, se recomienda una revisión crediticia reforzada antes de aprobar la operación.

Figura 18. Histórico de predicción

Resumen histórico usado para la predicción

Total préstamos	Con alerta	Castigados	Judiciales
1	1	1	0
Prejudiciales	Morosos	Vencidos	Riesgo severo histórico
0	0	0	Sí

Datos del nuevo préstamo a simular

Producto: MICRO MIGRA F Línea: MICROCREDITO

Deuda inicial: 50000.00 Tasa de interés anual (%): 15.00 Número de cuotas: 80

Frecuencia de pago en días: 30

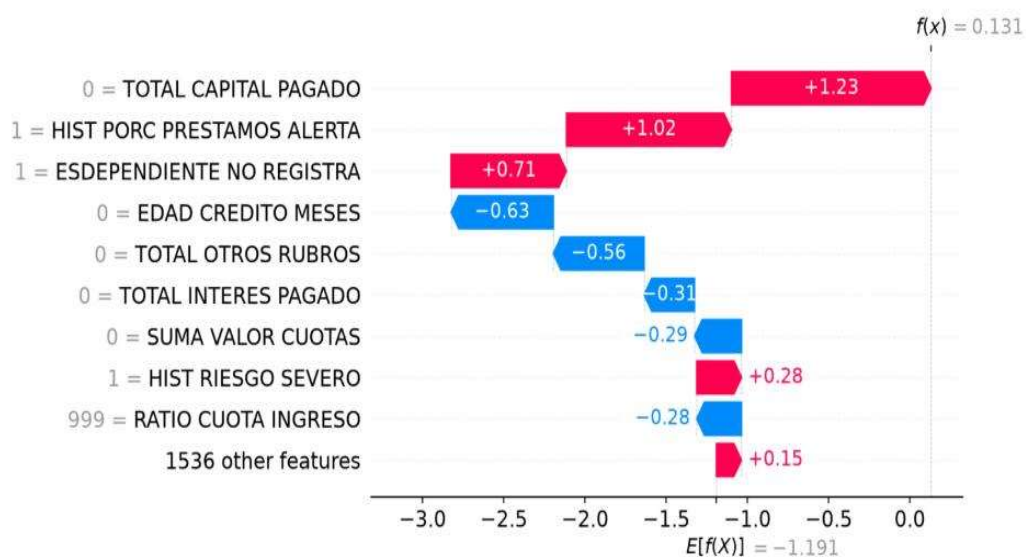
Valor cuota estimada: 992.33

[Predecir riesgo del préstamo](#)

Probabilidad de incumplimiento: 77.07% Nivel de riesgo: ALTO Predicción: Incumple

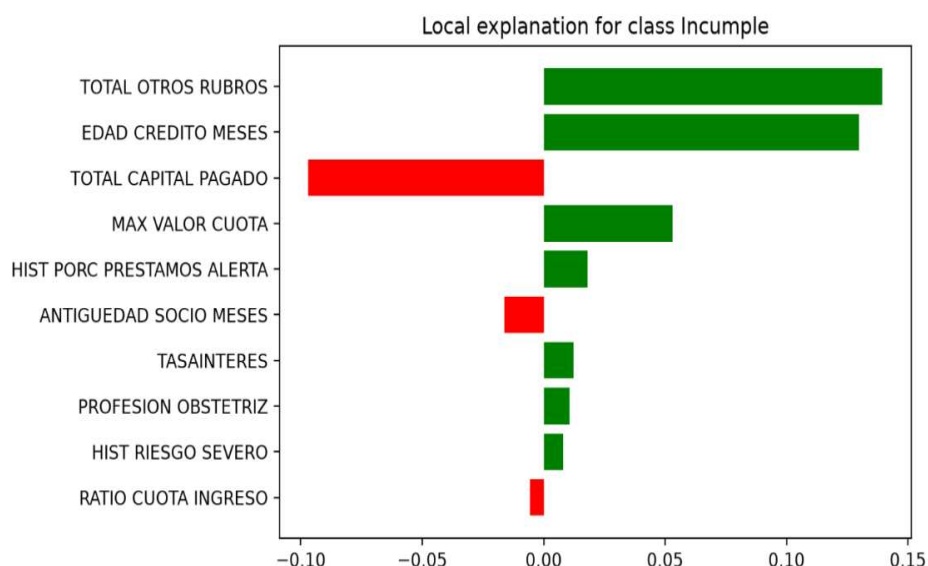
El análisis individual mediante SHAP permite interpretar con mayor detalle esta predicción. En este caso, las variables que más contribuyen a incrementar el riesgo estimado son TOTAL_CAPITAL_PAGADO, HIST_PORC_PRESTAMOS_ALERTA, ESDEPENDIENTE_NO_REGISTRA y HIST_RIESGO_SEVERO. Estos factores evidencian que el modelo asigna una importancia relevante al historial previo del socio, especialmente cuando existen alertas crediticias anteriores, señales de riesgo severo o condiciones laborales no registradas.

Figura 19. SHAP Individual



Por otra parte, variables como EDAD_CREDITO_MESES, TOTAL_OTROS_RUBROS, TOTAL_INTERES_PAGADO, SUMA_VALOR_CUOTAS y RATIO_CUOTA_INGRESO ejercen una contribución en sentido contrario, reduciendo parcialmente la estimación del riesgo. Sin embargo, el efecto conjunto de las variables que incrementan la probabilidad de incumplimiento es mayor, por lo que la clasificación final del modelo se mantiene en un nivel de riesgo alto.

Figura 20. LIME individual



En el caso analizado, LIME evidencia que variables como TOTAL_OTROS_RUBROS, EDAD_CREDITO_MESES, MAX_VALOR_CUOTA, HIST_PORC_PRESTAMOS_ALERTA, TASAINTERES, PROFESION_OBSTETRIZ e HIST_RIESGO_SEVERO contribuyen positivamente hacia la clase de incumplimiento. Esto indica que, para esta predicción individual, dichas variables empujan al modelo a clasificar la operación con mayor nivel de riesgo.

Por otro lado, variables como TOTAL_CAPITAL_PAGADO, ANTIGUEDAD_SOCIO_MESES y RATIO_CUOTA_INGRESO presentan una contribución negativa, es decir, reducen parcialmente la probabilidad estimada de incumplimiento. Sin embargo, el efecto conjunto de las variables que incrementan el riesgo es mayor, por lo que el modelo mantiene la clasificación final como riesgo alto.

SOCIO CON RIESGO BAJO

La simulación muestra una probabilidad de incumplimiento de 24,60 %, por lo que el modelo clasifica la operación como de riesgo bajo y predice que no incumplirá. Este resultado se relaciona con la ausencia de alertas, mora, créditos vencidos, judiciales, prejudiciales o castigados en el historial del socio, lo que refleja un comportamiento crediticio favorable.

Figura 21. Probabilidad socios riesgo bajo

Resumen histórico usado para la predicción

Total préstamos	Con alerta	Castigados	Judiciales
1	0	0	0
Prejudiciales	Morosos	Vencidos	Riesgo severo histórico
0	0	0	No

Datos del nuevo préstamo a simular

Producto: MICRO QUIRO Línea: MICROCREDITO

Deuda inicial: 50000,00 Tasa de interés anual (%): 15,00 Número de cuotas: 80

Frecuencia de pago en días: 30

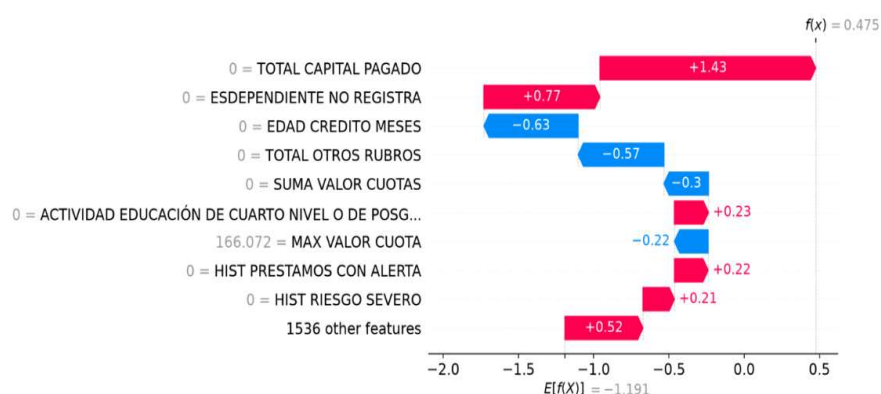
Valor cuota estimada: 992.33

Probabilidad de incumplimiento	Nivel de riesgo	Predicción
24.60%	BAJO	No incumple

El análisis individual mediante SHAP permite observar que la predicción no depende de un solo factor, sino del efecto combinado de varias variables. En este caso, algunas características del préstamo y del historial del socio contribuyen a reducir el riesgo estimado, mientras que otras variables ejercen un efecto en sentido contrario. Entre los factores que más influyen en la explicación individual se encuentran la edad del crédito, los otros rubros del préstamo, la suma del valor de las cuotas y el valor máximo de cuota, junto con variables

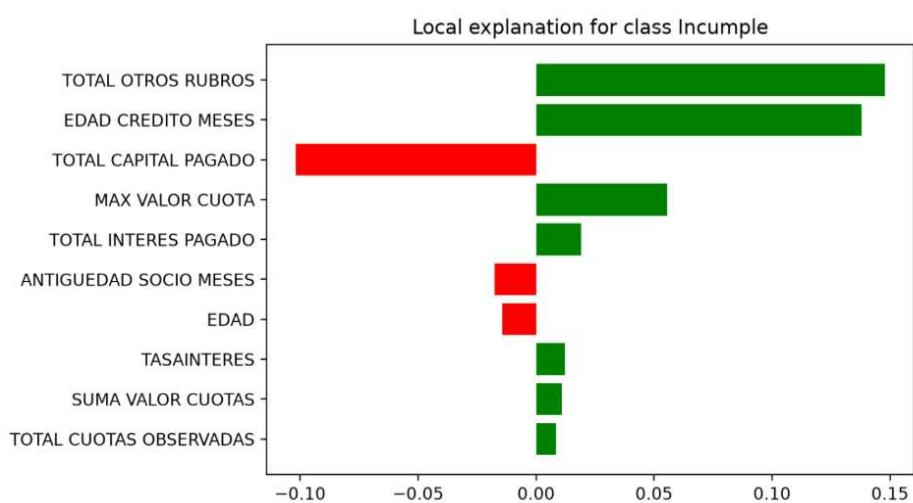
asociadas al historial del socio, como la ausencia de préstamos con alerta y de riesgo histórico severo.

Figura 22. SHAP Individual



En conjunto, el gráfico SHAP evidencia que, aunque existen variables con contribuciones positivas y negativas, el balance final de estos efectos mantiene la operación dentro de un nivel de riesgo bajo, coherente con el comportamiento crediticio favorable observado en el socio.

Figura 23. LIME Individual



En la explicación local se observa que variables como TOTAL_OTROS_RUBROS, EDAD_CREDITO_MESES, MAX_VALOR_CUOTA, TOTAL_INTERES_PAGADO, TASAINTERES, SUMA_VALOR_CUOTAS y TOTAL_CUOTAS_OBSERVADAS presentan una contribución positiva hacia la clase de incumplimiento. Esto significa que, para este caso individual, dichas variables aportan en cierta medida al incremento del riesgo estimado por el modelo.

Por otro lado, variables como TOTAL_CAPITAL_PAGADO, ANTIGUEDAD_SOCIO_MESES y EDAD presentan una contribución negativa, es decir, reducen la probabilidad de que el préstamo sea clasificado como incumplido. Estos factores ayudan a compensar parcialmente las variables que incrementan el riesgo.

4.6. Explicabilidad del modelo XGBoost

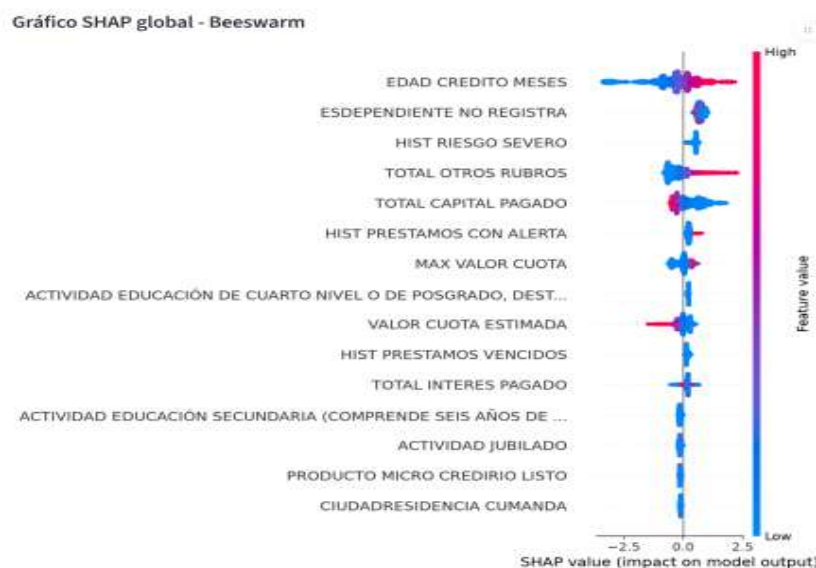
La selección de XGBoost como modelo final requirió incorporar mecanismos de explicabilidad que permitan comprender qué variables influyen en la estimación del incumplimiento de pagos. Para ello, se aplicó el método SHAP, el cual permite identificar la contribución de cada variable en la predicción generada por el modelo. Esta técnica fue utilizada a nivel global para conocer los factores más influyentes en el conjunto de datos y como soporte para interpretar la clasificación de riesgo generada por el sistema.

El análisis de importancia global según SHAP evidenció que las variables con mayor contribución en el modelo fueron EDAD_CREDITO_MESES, con una importancia de 0,7547; ESDEPENDIENTE_NO_REGISTRA, con 0,7207; HIST_RIESGO_SEVERO, con 0,5064; TOTAL_OTROS_RUBROS, con 0,4479; y TOTAL_CAPITAL_PAGADO, con 0,4393. Entre otras.

Estos resultados muestran que el modelo XGBoost considera principalmente variables asociadas con la antigüedad del crédito, la condición laboral registrada, el historial de riesgo del socio, los antecedentes de alerta, los préstamos vencidos, los valores de cuota y la relación entre cuota e ingreso. En particular, la presencia de variables como HIST_RIESGO_SEVERO, HIST_PRESTAMOS_CON_ALERTA, HIST_PRESTAMOS_VENCIDOS, HIST_PORC_PRESTAMOS_ALERTA e HIST_PRESTAMOS_MOROSOS evidencia que el comportamiento crediticio previo del socio tiene un peso relevante en la estimación del riesgo de incumplimiento.

Asimismo, variables como VALOR_CUOTA_ESTIMADA, MAX_VALOR_CUOTA y RATIO_CUOTA_INGRESO permiten interpretar la influencia de la capacidad de pago sobre el riesgo estimado. Esto resulta coherente con la lógica financiera, debido a que una mayor carga de cuota frente a los ingresos puede incrementar la probabilidad de incumplimiento. De igual forma, las variables relacionadas con el producto crediticio, la actividad económica y la ciudad de residencia permiten identificar diferencias en los patrones de riesgo según las características propias de cada operación y del perfil del socio.

Figura 24. SHAP GLOBAL



El gráfico *beeswarm* permite observar que `EDAD_CREDITO_MESES` presenta una dispersión amplia de valores SHAP, lo que indica que la antigüedad del crédito influye de manera importante en la clasificación del riesgo.

Asimismo, `ESDEPENDIENTE_NO_REGISTRA` muestra una contribución relevante, lo que evidencia que la ausencia o falta de registro de información laboral puede afectar la estimación del modelo. Por su parte, variables como `HIST_RIESGO_SEVERO` e `HIST_PRESTAMOS_CON_ALERTA` se relacionan con antecedentes crediticios que incrementan la probabilidad estimada de incumplimiento.

Figura 25. Variables Influyentes



El gráfico de barras confirma que las variables más influyentes fueron `EDAD_CREDITO_MESES`, `ESDEPENDIENTE_NO_REGISTRA` y `HIST_RIESGO_SEVERO`, seguidas por variables relacionadas con rubros del crédito, capital pagado, historial de alertas, valor máximo de cuota, cuota estimada y préstamos vencidos. Este resultado demuestra que el modelo no basa su predicción en una sola característica, sino en la combinación de factores temporales, financieros, históricos y sociodemográficos.

Finalmente, el análisis SHAP permitió fortalecer la transparencia del modelo, debido a que no solo se obtiene una probabilidad de incumplimiento, sino también una explicación de las variables que influyen en dicha estimación. Esta información facilita la interpretación de los resultados por parte del personal de crédito y permite utilizar el modelo como una herramienta de apoyo a la decisión, manteniendo la revisión humana dentro del proceso de evaluación crediticia.

Es importante señalar que los resultados obtenidos mediante SHAP y LIME en las predicciones individuales no necesariamente deben coincidir de forma exacta, debido a que ambas técnicas explican el modelo desde enfoques metodológicos diferentes. SHAP calcula la contribución de cada variable considerando la estructura interna del modelo XGBoost y permite obtener explicaciones más estables tanto a nivel global como individual. En cambio, LIME genera una aproximación local alrededor del caso evaluado, construyendo un modelo interpretable simple a partir de variaciones de la observación analizada.

Por esta razón, las variables identificadas y la magnitud de sus contribuciones pueden variar entre ambos métodos. Esta diferencia no invalida la predicción del modelo, sino que evidencia que cada técnica aporta una perspectiva distinta para interpretar el resultado. En la presente investigación, SHAP se utiliza como técnica principal de explicabilidad, mientras que LIME se incorpora como herramienta complementaria para contrastar y reforzar la interpretación de casos individuales

4.7. Validación operativa y calibración de probabilidades

Como parte de la validación operativa, se utilizó la pestaña de Predicción Individual del sistema desarrollado, la cual permitió consultar el historial crediticio de un socio, ingresar las condiciones de un nuevo préstamo y obtener una probabilidad estimada de incumplimiento junto con su clasificación de riesgo. Esta funcionalidad fue utilizada por oficiales y jefes de crédito para revisar casos individuales y valorar, mediante la encuesta aplicada, la coherencia de las predicciones generadas por el modelo.

Los resultados individuales permitieron clasificar las operaciones en niveles de riesgo bajo, medio y alto, de acuerdo con la probabilidad estimada de incumplimiento. Esta clasificación facilitó la interpretación práctica del modelo, ya que permitió diferenciar perfiles con antecedentes negativos, como alertas previas o créditos castigados, frente a socios con historial crediticio favorable. De esta manera, el modelo no solo entregó una predicción numérica, sino también una salida comprensible para apoyar el análisis crediticio.

La validación con personal del área de crédito permitió contrastar los resultados del modelo con el criterio profesional de quienes participan en la evaluación y seguimiento de operaciones crediticias. Esta revisión aportó evidencia sobre la utilidad práctica del sistema, debido a que los participantes evaluaron la correspondencia entre la clasificación de riesgo generada por el modelo y su experiencia en el análisis de cartera.

En cuanto a la calibración de probabilidades, en esta versión del modelo se utilizaron las probabilidades generadas por XGBoost para clasificar el nivel de riesgo y calcular métricas como el AUC-ROC. Sin embargo, para una implementación institucional más robusta, se considera necesario aplicar un proceso posterior de calibración mediante métodos como calibración isotónica o sigmoide, por ejemplo, con `CalibratedClassifierCV`. Este proceso permitiría que las probabilidades estimadas por el modelo tengan una correspondencia más cercana con la frecuencia real de incumplimiento observada en la cartera.

Una vez calibradas las probabilidades, la cooperativa podría ajustar los puntos de corte para los niveles de riesgo bajo, medio y alto con base en evidencia empírica, considerando la distribución de probabilidades, el comportamiento histórico de la mora y la capacidad operativa

del área de crédito y cobranza. Esto permitiría fortalecer el uso del modelo como herramienta de apoyo a la decisión, manteniendo la revisión humana dentro del proceso de evaluación crediticia.

4.8. Discusión de resultados

Los resultados obtenidos confirman que los modelos de aprendizaje automático pueden aportar valor a la gestión del riesgo crediticio en cooperativas de ahorro y crédito. La comparación realizada evidenció que los modelos de ensamble superaron a la Regresión Logística en capacidad discriminativa y desempeño general. Este resultado se explica porque el incumplimiento de pagos depende de relaciones complejas entre múltiples factores, tales como historial crediticio, endeudamiento, características del préstamo, condiciones del socio y comportamiento previo.

XGBoost se posicionó como el modelo con mejor desempeño global debido a sus resultados superiores en exactitud, precisión, F1-score y AUC-ROC. Su principal ventaja fue la capacidad de generar alertas de mayor calidad, reduciendo el número de falsos positivos y permitiendo identificar con mayor precisión las operaciones clasificadas como riesgosas. Esta característica resulta relevante en el contexto cooperativo, donde las áreas de crédito y cobranza requieren priorizar operaciones con mayor probabilidad de deterioro.

No obstante, los resultados también evidencian que XGBoost presentó un recall inferior al de Random Forest. Esto significa que, aunque sus alertas fueron más precisas, dejó sin identificar una parte de los préstamos que realmente incurrieron en mora. Por tanto, la interpretación del modelo debe considerar este equilibrio: Random Forest detecta más casos reales de incumplimiento, pero genera más alertas falsas; mientras que XGBoost reduce las clasificaciones incorrectas y presenta una mayor capacidad discriminativa global.

El análisis realizado permitió identificar que la consolidación de datos a nivel de préstamo constituye una alternativa metodológica adecuada para estimar el incumplimiento en operaciones crediticias individuales. La construcción de variables relacionadas con el perfil del socio, capacidad de pago, endeudamiento, características del crédito e historial crediticio permitió representar factores relevantes para la predicción de mora.

La comparación entre Regresión Logística, Random Forest y XGBoost evidenció que XGBoost obtuvo el mejor desempeño global. El modelo alcanzó una exactitud de 84,30 %, precisión de 73,87 %, F1-score de 59,93 % y AUC-ROC de 0,8797, superando a los demás modelos en capacidad discriminativa y precisión de alertas. Estos resultados justifican su selección como modelo final para la estimación del riesgo de incumplimiento.

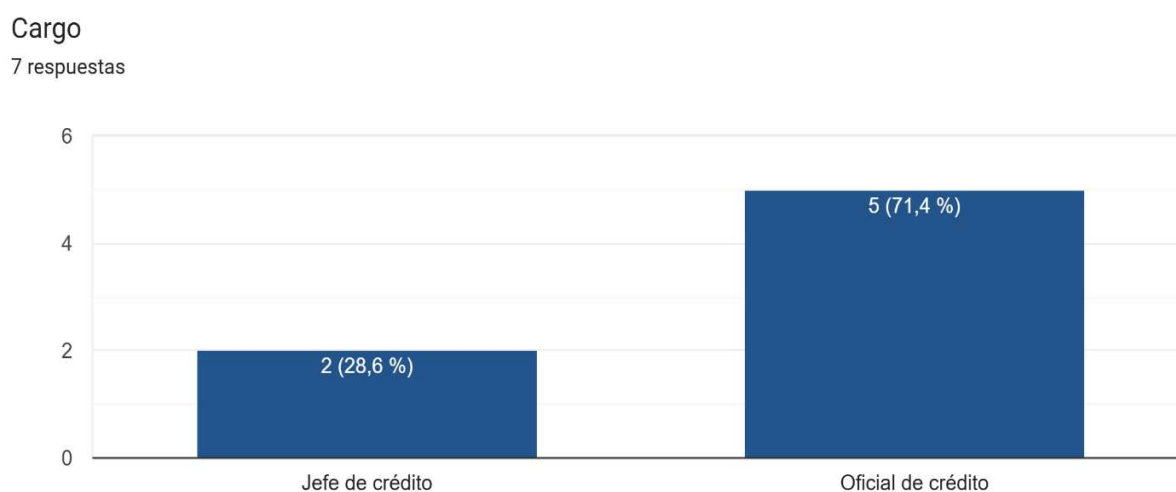
La incorporación de SHAP fortaleció la explicabilidad del modelo, ya que permitió identificar las variables con mayor influencia en la predicción del incumplimiento. Los resultados mostraron la importancia de factores como edad del crédito, condición laboral registrada, riesgo histórico severo, historial de alertas, préstamos vencidos, valor de cuota y relación cuota-ingreso. Esto aporta transparencia al modelo y permite que sus resultados sean interpretados por el personal de crédito como apoyo a la toma de decisiones.

Finalmente, el modelo propuesto debe entenderse como una herramienta de apoyo a la decisión humana y no como un mecanismo automático de aprobación o rechazo de créditos. Para una implementación institucional más robusta, se plantea como mejora posterior la calibración de probabilidades, el ajuste de umbrales de riesgo y la validación temporal con nuevos préstamos, con el fin de verificar que el desempeño del modelo se mantenga en escenarios futuros de uso operativo.

4.9. Encuesta con expertos

La muestra estuvo conformada por siete colaboradores del área de crédito: cinco oficiales de crédito, equivalentes al 71,4 %, y dos jefes de crédito, correspondientes al 28,6 %. Esta distribución evidencia una mayor participación del personal operativo, cuya experiencia directa en la evaluación y seguimiento de préstamos aporta información relevante para validar la utilidad práctica del modelo predictivo.

Figura 26 Encuestas – Expertos- Cargos

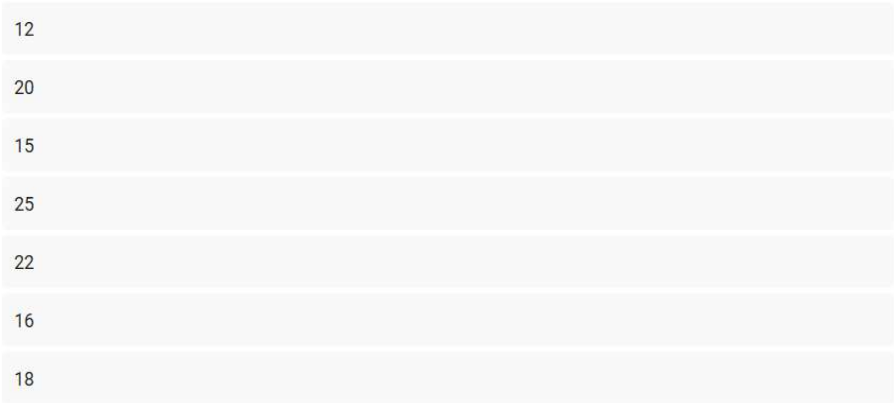


Los siete participantes presentan una experiencia profesional amplia en el área de crédito, con valores comprendidos entre 12 y 25 años. El promedio de experiencia es de aproximadamente 18 años, lo que evidencia que las respuestas fueron emitidas por personal con conocimiento sólido sobre evaluación crediticia, comportamiento de cartera y gestión del riesgo de incumplimiento.

Figura 27. Experiencia

Años de experiencia:

7 respuestas

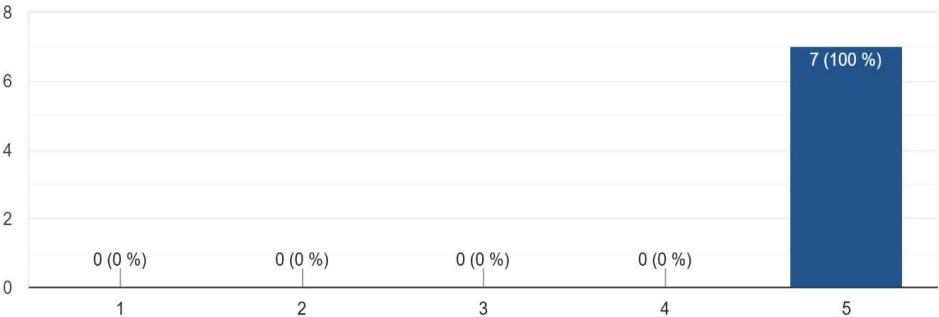


Los resultados evidencian una valoración totalmente favorable respecto de la coherencia de la clasificación de riesgo generada por el modelo. Los siete participantes, equivalentes al 100 % de la muestra, calificaron con 5 sobre 5 la correspondencia entre la predicción del sistema y su criterio profesional. Este resultado respalda la utilidad práctica del modelo como herramienta de apoyo para la evaluación crediticia y la gestión preventiva del riesgo.

Figura 28. Pregunta 1

¿La clasificación de riesgo generada por la predicción es coherente con su criterio profesional?

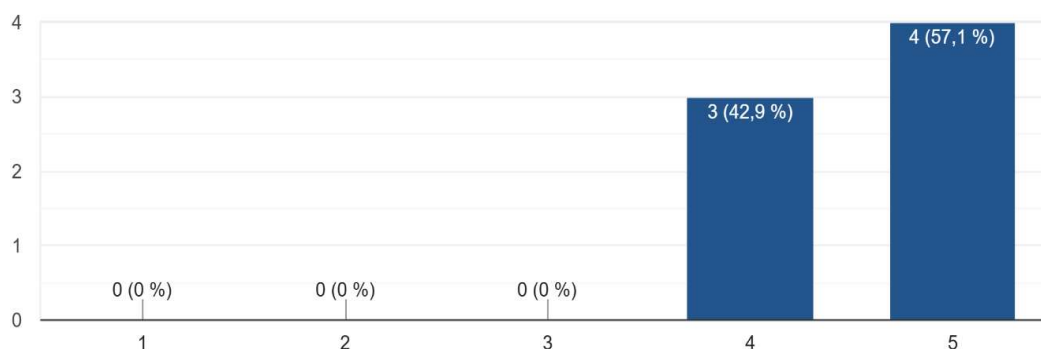
7 respuestas



Los resultados muestran una percepción favorable sobre la razonabilidad de la probabilidad de incumplimiento estimada por el modelo. Cuatro participantes, equivalentes al 57,1 %, calificaron este aspecto con 5 sobre 5, mientras que tres participantes, correspondientes al 42,9 %, otorgaron una calificación de 4 sobre 5. En conjunto, las valoraciones reflejan que el personal de crédito considera que las probabilidades generadas por el sistema son coherentes y útiles para apoyar la evaluación de riesgo.

Figura 29. Pregunta 2

¿Considera que la probabilidad de incumplimiento estimada por la predicción es razonable?
7 respuestas

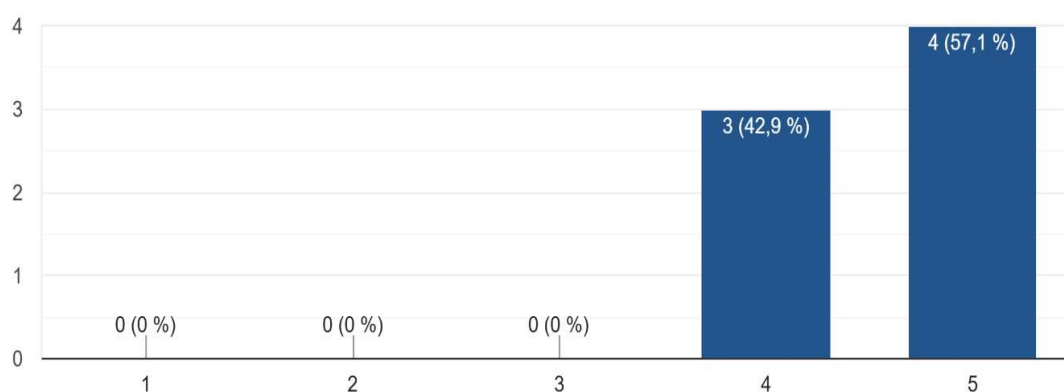


Los resultados reflejan una aceptación favorable del modelo como herramienta de apoyo para la evaluación del incumplimiento de pagos. Cuatro participantes, equivalentes al 57,1 %, calificaron esta utilidad con 5 sobre 5, mientras que tres participantes, correspondientes al 42,9 %, otorgaron una calificación de 4 sobre 5. En consecuencia, el personal de crédito considera que el modelo puede contribuir de manera efectiva a la toma de decisiones y al fortalecimiento de la gestión preventiva del riesgo crediticio.

Figura 30. Pregunta 3

¿Considera que el modelo puede ser una herramienta de apoyo para la evaluación del incumplimiento de pagos?

7 respuestas

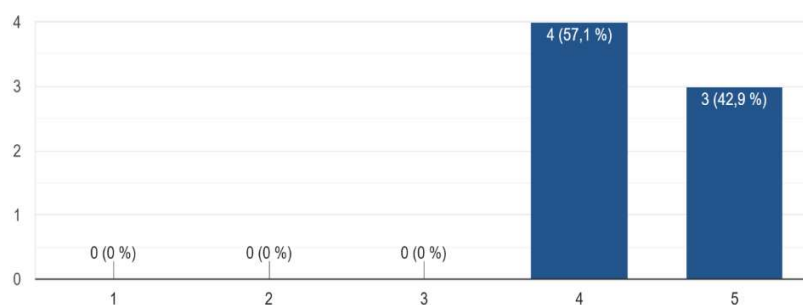


Los resultados muestran una valoración positiva sobre la utilidad de la información presentada por el modelo para la evaluación crediticia. Cuatro participantes, equivalentes al 57,1 %, calificaron este aspecto con 4 sobre 5, mientras que tres participantes, correspondientes al 42,9 %, otorgaron la calificación máxima de 5 sobre 5. En conjunto, estas respuestas evidencian que el sistema proporciona elementos relevantes y comprensibles para complementar el análisis de riesgo y apoyar la toma de decisiones crediticias.

Figura 31. Pregunta 4

¿La información presentada por el modelo aporta con elementos útiles para una evaluación crediticia?

7 respuestas

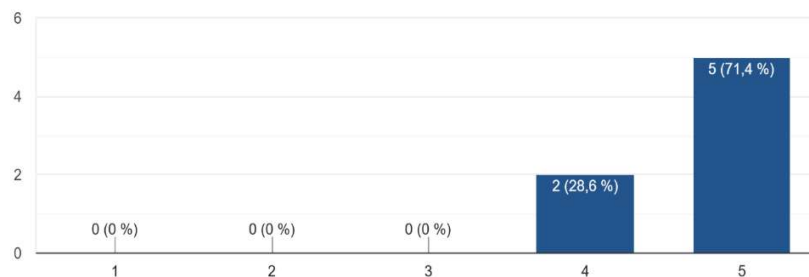


Los resultados evidencian una percepción favorable sobre la consistencia del modelo frente a los casos reales evaluados. Cinco participantes, equivalentes al 71,4 %, otorgaron una calificación de 5 sobre 5, mientras que dos participantes, correspondientes al 28,6 %, calificaron con 4 sobre 5. Estas valoraciones indican que el personal de crédito considera que las predicciones generadas por el modelo son coherentes con la realidad observada y pueden aportar valor a la gestión del riesgo crediticio.

Figura 32. Pregunta 5

¿Los resultados obtenidos por el modelo son consistentes con la realidad de los casos evaluados?

7 respuestas



CAPÍTULO 5

5. Conclusiones y recomendaciones

5.1 Conclusiones

El análisis y preprocesamiento de la información crediticia permitió consolidar una base de datos a nivel de préstamo, incorporando variables relacionadas con el perfil del socio, producto crediticio, capacidad de pago, endeudamiento e historial financiero. El análisis exploratorio evidenció que el incumplimiento se asocia con factores como alertas previas, antecedentes crediticios, condiciones laborales, edad del crédito y características del producto, confirmando que la mora responde a una combinación de variables y no a un único factor.

Los modelos de aprendizaje automático mostraron diferencias importantes en su desempeño. La Regresión Logística presentó una capacidad discriminativa limitada, con un AUC-ROC de 0,6580 y una precisión de 25,18 %. En cambio, los modelos de ensamble obtuvieron mejores resultados: Random Forest alcanzó un AUC-ROC de 0,8335 y XGBoost logró el mejor desempeño global, con un AUC-ROC de 0,8797.

XGBoost fue seleccionado como el modelo de mejor desempeño global, al obtener una exactitud de 84,30 %, precisión de 73,87 %, F1-score de 59,93 % y AUC-ROC de 0,8797. Estos resultados evidencian que el modelo logra clasificar con mayor precisión los préstamos con y sin incumplimiento, reduciendo significativamente las alertas incorrectas frente a los demás algoritmos evaluados. No obstante, debido a que su recall fue de 50,41 %, se recomienda calibrar las probabilidades y ajustar el umbral de decisión para mejorar la detección de préstamos con riesgo de mora.

La validación realizada con personal del área de crédito confirmó la utilidad práctica del modelo. Los siete participantes consideraron coherente la clasificación de riesgo con su criterio profesional, y la mayoría valoró favorablemente la consistencia de las predicciones frente a los casos evaluados. Además, la aplicación de SHAP permitió fortalecer la explicabilidad del modelo, identificando las variables con mayor influencia en la predicción del incumplimiento y facilitando que los resultados sean comprensibles y utilizables como apoyo a la evaluación crediticia y a la toma de decisiones preventivas.

5.2 Recomendaciones

Implementar XGBoost como modelo principal de apoyo a la gestión de riesgo crediticio, integrándolo en un tablero o sistema de consulta que permita estimar la probabilidad de incumplimiento usando información disponible según la fecha de corte.

Calibrar las probabilidades y ajustar el umbral de clasificación de XGBoost, debido a que su recall fue de 50,41 %. Esta medida permitirá reducir los falsos negativos y mejorar la identificación de préstamos que podrían incumplir, sin perder de forma significativa la precisión alcanzada.

Mantener y fortalecer el uso de SHAP en la interpretación global e individual del modelo, mostrando los factores que aumentan o disminuyen el riesgo, tales como historial de alertas, riesgo histórico severo, préstamos vencidos, capacidad de pago, endeudamiento, valor de cuota y características del préstamo. Esto fortalecerá la transparencia y facilitará la revisión por parte de oficiales y jefes de crédito.

Establecer un proceso periódico de monitoreo y actualización del modelo, incluyendo validación temporal, revisión de calidad de datos, control de sesgos, evaluación de métricas y contraste con casos reales. El modelo debe utilizarse como apoyo a la decisión humana y no como mecanismo automático de aprobación o rechazo de créditos.

Referencias Bibliográficas

- Baesens, B. (2014). *Analytics in a Big Data World: The Essential Guide to Data Science and Its Applications*. Wiley.
- Barredo Arrieta, A., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., García, S., Gil-López, S., Molina, D., Benjamins, R., Chatila, R., & Herrera, F. (2020). Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, 58, 82–115. <https://doi.org/10.1016/j.inffus.2019.12.012>
- Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer.
- Bussmann, N., Giudici, P., Marinelli, D., & Papenbrock, J. (2021). Explainable AI in credit risk management. *Computational Economics*, 57(1), 203–216. <https://doi.org/10.1007/s10614-020-10042-0>
- COOP. RIOBAMBA LTDA. (2026). *COOP. RIOBAMBA LTDA*.
https://www.cooprio.fin.ec/?page_id=864
- European Banking Authority. (2021). Report on Machine Learning for IRB Models.
<https://www.eba.europa.eu>

- European Commission. (2024). Artificial Intelligence Act (Regulation (EU) 2024/1689). <https://eur-lex.europa.eu>
- Géron, A. (2022). Hands-On Machine Learning with Scikit-Learn, Keras & TensorFlow (3rd ed.). O'Reilly Media.
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press.
<https://www.deeplearningbook.org>
- Golec, M., & AlabdulJalil, M. (2025). Interpretable LLMs for credit risk: A systematic review and taxonomy. arXiv. <https://arxiv.org/abs/2506.04290>
- Huyen, C. (2022). Designing Machine Learning Systems: An Iterative Process for Production-Ready Applications. O'Reilly Media.
- ISO/IEC. (2023). ISO/IEC 23894:2023 Information technology—Artificial intelligence—Guidance on risk management. International Organization for Standardization.
- Kou, G., Olgu Akdeniz, Ö., Dinçer, H., & Yüksel, S. (2021). Fintech investments in European banks: A hybrid IT2 fuzzy multidimensional decision-making approach. Financial Innovation, 7, Article 39. <https://doi.org/10.1186/s40854-021-00256-y>
- Krawczyk, B. (2021). Learning from imbalanced data: Open challenges and future directions. Progress in Artificial Intelligence, 10, 1–13. <https://doi.org/10.1007/s13748-020-00213-8>
- Lessmann, S., Baesens, B., Seow, H. V., & Thomas, L. C. (2015). Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research. European Journal of Operational Research, 247(1), 124–136. <https://doi.org/10.1016/j.ejor.2015.05.030>

- Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, 30.
<https://proceedings.neurips.cc/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html>
- Misheva, B. H., Osterrieder, J., Hirs, A., Kulkarni, O., & Lin, S. F. (2021). Explainable AI in credit risk management. *Frontiers in Artificial Intelligence*, 4, Article 661483.
<https://doi.org/10.3389/frai.2021.661483>
- Molnar, C. (2024). *Interpretable Machine Learning* (3rd ed.).
<https://christophm.github.io/interpretable-ml-book/>
- Murdoch, W. J., Singh, C., Kumbier, K., Abbasi-Asl, R., & Yu, B. (2019). Definitions, methods, and applications in interpretable machine learning. *Proceedings of the National Academy of Sciences*, 116(44), 22071–22080. <https://doi.org/10.1073/pnas.1900654116>
- OECD. (2024). *OECD Framework for the Classification of AI Systems*. <https://oecd.ai>
- Phillips, P. J., Hahn, C. A., Fontana, A. A., Yates, A. N., Greene, K., Broniatowski, D. A., & Przybocki, M. A. (2021). Four principles of explainable artificial intelligence (NISTIR 8312). National Institute of Standards and Technology. <https://doi.org/10.6028/NIST.IR.8312>
- Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). “Why Should I Trust You?”: Explaining the predictions of any classifier. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1135–1144.
<https://doi.org/10.1145/2939672.2939778>

Rudin, C., Chen, C., Chen, Z., Huang, H., Semenova, L., & Zhong, C. (2022). Interpretable machine learning: Fundamental principles and 10 grand challenges. *Statistics Surveys*, 16, 1–85.

<https://doi.org/10.1214/21-SS133>

Shrestha, Y. R., Ben-Menahem, S. M., & von Krogh, G. (2022). Organizational decision-making structures in the age of artificial intelligence. *California Management Review*, 64(2), 66–83.

<https://doi.org/10.1177/00081256211067746>

Superintendencia de Economía Popular y Solidaria. (2024). Informe de gestión y supervisión del sector financiero popular y solidario. <https://www.seps.gob.ec/>

Siddique, A. K. (2024). Advancing credit risk modelling with machine learning: A comprehensive review of the state-of-the-art. . *Advancing credit risk modelling with machine learning: A comprehensive review of the state-of-the-art.*, 137. Engineering Applications of Artificial Intelligence. doi: <https://doi.org/10.1016/j.engappai.2024.109>

SRI. (s.f.). *SRI EN LINEA*. Obtenido de <https://srienlinea.sri.gob.ec/sri-en-linea>

Su, X. S. (2021). Random Forest. En Wiley StatsRef: Statistics Reference Online. *Random Forest*. En *Wiley StatsRef: Statistics Reference Online*. Wiley. Obtenido de <https://doi.org/10.1002/9781118445112.stat08287>

Superintendencia de Economía Popular y Solidaria. (2024). Norma para la Administración Integral de Riesgos. SEPS. *Norma para la Administración Integral de Riesgos. SEPS*. Obtenido de https://www.seps.gob.ec/wp-content/uploads/RESOLUCIO%CC%81N-Nro.-SEPS-IGT-INR-INSESP-INGINT-2024-0122-REFORMA-NORMA-OFICIALES-CUMPLIMIENTO_firmado.pdf

Yang, S., Huang, Z., Xiao, W., & Shen, X. (2025). Interpretable credit default prediction with ensemble learning and SHAP. arXiv. <https://arxiv.org/abs/2505.20815>

Zhang, S., Tay, J., & Baiz, P. (2024). The effects of data imbalance under a federated learning approach for credit risk forecasting. arXiv. <https://arxiv.org/abs/2401.07234>

ANEXOS

ENCUESTA CON EXPERTOS

**RIOBAMBA Ltda.**
Cooperativa de Ahorro y Crédito

Sección 1 de 2

Ficha de validación

×

⋮

Evaluar si las predicciones individuales generadas de un Sistema de Predicción del Incumplimiento de Pagos para la Cooperativa de Ahorro y Crédito Riobamba Ltda.

Estimado(a) colaborador(a):

El presente instrumento tiene como objetivo validar la coherencia y utilidad de las predicciones generadas por un sistema de inteligencia artificial desarrollado para apoyar la identificación temprana del riesgo de mora en créditos.

Para esta investigación se considera que una cuota presenta incumplimiento cuando registra una mora igual o superior a 30 días.

A continuación, se validaran la coherencia de las predicciones generadas por el sistema mediante la evaluación de expertos del área de crédito utilizando casos reales de socios conocidos por ellos.

La información recopilada será utilizada únicamente con fines académicos.

PERFIL DEL EVALUADOR

Descripción (opcional)

Cargo *

Texto de respuesta corta

Años de experiencia: ★

Texto de respuesta corta

Sección 2 de 2

Caso 1

Recuerde que para esta investigación se considera incumplimiento de pago cuando una cuota presenta una mora igual o superior a 30 días.

Con base en la información mostrada por el sistema, responda las siguientes preguntas según su criterio profesional.

¿La clasificación de riesgo generada por la predicción es coherente con su criterio profesional?

1. Totalmente en desacuerdo
2. En desacuerdo
3. Neutral
4. De acuerdo
5. Totalmente de acuerdo

1 2 3 4 5

Totalmente en desacuerdo ☐ ☐ ☐ ☐ ☐ Totalmente de acuerdo

¿Considera que la probabilidad de incumplimiento estimada por la predicción es razonable?

1. Totalmente en desacuerdo
2. En desacuerdo
3. Neutral
4. De acuerdo
5. Totalmente de acuerdo

1 2 3 4 5

Totalmente en desacuerdo ☐ ☐ ☐ ☐ ☐ Totalmente de acuerdo

¿Considera que el modelo puede ser una herramienta de apoyo para la evaluación del incumplimiento de pagos?

1. Totalmente en desacuerdo
2. En desacuerdo
3. Neutral
4. De acuerdo
5. Totalmente de acuerdo

1 2 3 4 5

Totalmente en desacuerdo ☐ ☐ ☐ ☐ ☐ Totalmente de acuerdo