



Vigilada Mineducación

MODELOS DE RIESGO CREDITICIO PARA REDUCIR EL RIESGO DE  
CRÉDITO EN LA CARTERA HIPOTECARIA DEL BANCO UNIÓN  
Credit risk models to reduce credit risk in Banco Unión's mortgage portfolio

ANA PATRICIA CÓRDOBA RUIZ

Tesis de grado

Asesor

Brayan Ricardo Rojas Ormaza

UNIVERSIDAD EAFIT  
ESCUELA DE ECONOMÍA Y FINANZAS  
MAESTRÍA EN ADMINISTRACIÓN FINANCIERA - MAF  
CALI  
2025

## TABLA DE CONTENIDO

INTRODUCCIÓN .....	9
Datos utilizados .....	11
Esquemas investigativos .....	11
Técnicas matemáticas y estadísticas .....	12
PLANTEAMIENTO DEL PROBLEMA.....	14
JUSTIFICACIÓN.....	17
OBJETIVOS.....	18
General .....	18
Específicos .....	18
MARCO TEÓRICO Y MARCO CONCEPTUAL .....	19
Riesgo de crédito en la cartera hipotecaria .....	19
Gestión de provisiones.....	20
Cómo se evalúa el riesgo crediticio.....	21
Modelos de riesgo crediticio como herramientas de clasificación .....	23
Modelos estadísticos para la clasificación del incumplimiento .....	24
Enfoques no lineales y técnicas de minería de datos.....	24
DISEÑO METODOLÓGICO .....	26
Enfoque y diseño de la investigación .....	26
Fuente de datos y población de estudio.....	26
Modelación y definición de variables.....	27
Plataforma tecnológica y entorno de análisis .....	30

Preparación y limpieza de los datos.....	30
Depuración inicial del conjunto de datos .....	30
Iteraciones de especificación (proceso de búsqueda).....	34
Modelos de elección discreta con variable dependiente binaria: aplicación al riesgo de default.....	36
Definición de la variable objetivo .....	39
Tratamiento del desbalance de clases .....	39
Balanceo de la muestra .....	40
Estrategia de modelamiento y selección de algoritmos.....	40
Modelos de clasificación evaluados .....	41
Criterios de evaluación del desempeño de los modelos .....	42
Elección de métricas de evaluación.....	42
Desbalanceo de clases.....	43
Supuestos del modelo Logit y riesgos de multicolinealidad .....	44
<b>DESARROLLO DEL TRABAJO .....</b>	<b>46</b>
Revisión de la literatura sobre los modelos de riesgo crediticio .....	46
Selección de los modelos aptos para el Banco Unión.....	49
Modelo de Green y Hendershott - Análisis de regresión múltiple .....	50
Identificación de las variables relevantes .....	50
Exploración de variables propias del mercado hipotecario colombiano.....	57
Variables de la cartera de crédito hipotecario del Banco Unión.....	61
<b>RESULTADOS.....</b>	<b>67</b>
Resultados de los modelos estimados .....	67
Comparación y selección del modelo .....	69

Selección final del modelo y consideraciones operativas .....	69
Resultados del modelo Logit.....	70
Resultados de la modelación en KNIME mediante AutoML.....	71
Comparación integrada de resultados .....	72
Validez y capacidad predictiva del modelo.....	74
Interpretación de los efectos marginales.....	74
CONCLUSIONES .....	80
RECOMENDACIONES.....	82
REFERENCIAS .....	83
ANEXOS.....	87
Anexo A. Estimación modelo Logit.....	87
Anexo B. Resultados detallados de la estimación del modelo Logit.....	88
Anexo C. Estructura gráfica del Árbol de Decisión para la clasificación de riesgo .....	89
Anexo D. Capacidad predictiva del modelo.....	90
Anexo E. Cálculos AIC y BIC .....	91
Anexo F. Resultados del modelo Random Forest.....	92
Anexo G. Modelo logístico por default.....	93
Anexo H. Efecto marginal de un año adicional de edad sobre la probabilidad de default .....	94
Anexo I. Efectos marginales para diferentes valores de edad .....	95

## LISTA DE FIGURAS

Figura 1. Categorías de referencia .....	33
Figura 2. Categoría nivel de estudios .....	33
Figura 3. Categoría tipo de vivienda .....	34
Figura 4. Categoría Proyecto VIS .....	34
Figura 5. Conjunto final de covariables utilizadas en el modelo .....	36
Figura 6. Flujo de preparación y balanceo de datos en KNIME .....	40
Figura 7. Selección de modelos para entrenar .....	41
Figura 8. Variación trimestral del PIB.....	51
Figura 9. IPC Colombia 2015-2025.....	53
Figura 10. Tasa efectiva anual. Comportamiento histórico (2015-2025).....	56
Figura 11. Comportamiento del IPVN (2015-2025).....	59
Figura 12. Comportamiento de la cartera de crédito hipotecario del Banco Unión	62
Figura 13. Comportamiento de la cartera de crédito hipotecario, tasa efectiva e ICVH .....	63
Figura 14. Modelos de clasificación ranqueados por la métrica de precisión .....	67
Figura 15. Resultados detallados de la estimación del modelo Logit.....	70
Figura 16. Estructura gráfica del Árbol de Decisión para la clasificación de riesgo .....	71
Figura 17. Efectos marginales para diferentes valores de edad .....	77
Figura 18. Márgenes predictivos con intervalos de confianza del 95 %.....	77
Figura 19. Muestra del área bajo la curva ROC del modelo logístico final.....	78

## LISTA DE TABLAS

Tabla 1. Variables que serán utilizadas .....	27
Tabla 2. Análisis anual de la cartera de crédito hipotecario vencida del Banco Unión .....	63
Tabla 3. Clasificación nivel de riesgo de la cartera de crédito hipotecario vencida del Banco Unión .....	65
Tabla 4. Descripción de la calificación del riesgo.....	65

## RESUMEN

La cartera hipotecaria del Banco Unión ha aumentado su exposición al riesgo debido a la incertidumbre económica pospandemia, evidenciada en niveles más altos de desempleo, presiones inflacionarias y crecimientos continuos en las tasas de interés. El presente estudio examina los modelos de riesgo crediticio diseñados para mejorar la capacidad predictiva del deterioro y para apoyar la toma de decisiones en la administración de carteras hipotecarias, con base en los datos históricos entregados por el Banco Unión, en los cuales se encuentra información sobre las hipotecas vigentes y deterioradas. Dichos datos históricos incluyen variables macroeconómicas, rasgos específicos del crédito y características sociodemográficas y financieras de los prestatarios. Para examinar la probabilidad de incumplimiento se estimó un modelo Logit binario desde una perspectiva metodológica que analiza la incidencia marginal de factores esenciales como el nivel de ingresos, la tasa de interés, las condiciones macroeconómicas y el plazo del crédito. Se utilizó además un modelo de Árbol de Decisión aplicado en la plataforma KNIME, a través de métodos AutoML, después del balanceo muestral. La capacidad de predicción de los modelos se midió mediante métricas estándar del sector financiero como el área bajo la curva ROC (AUC), la sensibilidad (recall) y la precisión. Se observó que el modelo Logit proporcionó una sólida estructura explicativa de los factores determinantes del riesgo crediticio, a diferencia del Árbol de Decisión, que tuvo una sensibilidad mayor para detectar clientes deteriorados y conservar altos niveles de precisión. Si se combinan ambas perspectivas, el Banco Unión obtiene un instrumento analítico más sólido para prever el deterioro de la cartera, mejorar la segmentación de los clientes y consolidar las estrategias de monitoreo y mitigación del riesgo crediticio en un contexto económico variable, alineando así los estándares estadísticos con criterios operativos y regulatorios.

**Palabras clave:** riesgo crediticio, cartera hipotecaria, modelo Logit, Árbol de Decisión, minería de datos, KNIME.

## ABSTRACT

Banco Unión's mortgage portfolio has increased its exposure to risk as a result of post-pandemic economic uncertainty, evidenced by higher unemployment levels, inflationary pressures, and sustained increases in interest rates. Within this context, the study examines credit risk models designed to enhance the predictive capacity of portfolio deterioration and to support decision-making in the management of mortgage portfolios. The analysis relies on a historical database provided by Banco Unión, which contains information on performing and impaired mortgage loans. This dataset includes macroeconomic variables, credit-specific attributes, and the sociodemographic and financial characteristics of borrowers. From a methodological perspective, a binary Logit model was estimated to assess the probability of default, analyzing the marginal impact of key factors such as income level, interest rate, macroeconomic conditions, and loan maturity. In addition, a Decision Tree model was implemented on the KNIME platform using AutoML techniques, following sample balancing procedures. The predictive performance of the models was evaluated using standard metrics in the financial sector: the area under the ROC curve (AUC), sensitivity (recall), and precision. The results show that the Logit model provides a strong explanatory framework for the determinants of credit risk, while the Decision Tree exhibits greater sensitivity in identifying impaired borrowers while maintaining high levels of precision. By combining both perspectives, Banco Unión gains a more robust analytical tool to anticipate portfolio deterioration, improve customer segmentation, and strengthen credit risk monitoring and mitigation strategies in a changing economic environment, thereby aligning statistical standards with operational and regulatory criteria.

**Keywords:** credit risk, mortgage portfolio, Logit model, decision tree, data mining, KNIME.

## INTRODUCCIÓN

La gestión interna de las entidades financieras requiere, en particular en momentos de inestabilidad económica, instrumentos que posibiliten prever deterioros y robustecer el manejo interno. Esto es necesario para la dinámica del riesgo crediticio en la cartera hipotecaria. Según los análisis de Urrea-Ríos y Piraján (2020), la calidad de las obligaciones hipotecarias se vio alterada y el comportamiento de pago de los hogares cambió significativamente en Colombia debido a las consecuencias que trajo consigo la contracción económica después de la pandemia. Como las entidades bancarias vieron un aumento en la fragilidad de sus carteras, se hace esencial el análisis estricto de modelos predictivos para garantizar la sostenibilidad financiera y proteger la exposición crediticia.

En respuesta a este contexto, las instituciones financieras implementaron medidas de alivio para respaldar a los prestatarios y reprogramar sus pagos, como sugiere González (2021). No obstante, la situación económica ha experimentado cambios, y es esencial evaluar y proponer modelos que mitiguen el riesgo crediticio en este nuevo escenario, donde los alivios previos pueden haber evolucionado. La persistente incertidumbre económica plantea riesgos adicionales para el sector hipotecario y, por lo tanto, es necesario ajustar continuamente la evaluación de riesgos y proponer modelos actualizados que sean efectivos en su gestión.

El Banco Unión, como un agente importante en el sistema financiero, ha tenido que ajustarse a una situación en la que los tipos de interés, el desempleo y la disminución de ingresos ejercen presión sobre las probabilidades de incumplimiento. En esta situación, se hace necesario determinar cuál modelo de riesgo crediticio posibilita entender de manera más efectiva la estructura de riesgos de su cartera hipotecaria y, simultáneamente, brinde una base empírica robusta para tomar decisiones estratégicas.

De acuerdo con los hallazgos de Martínez (2019), la evaluación de modelos específicos es fundamental para mitigar el riesgo de crédito en el sector hipotecario, y su implementación es esencial para garantizar la estabilidad y eficiencia del mercado hipotecario, beneficiando tanto a las instituciones financieras como a los prestatarios.

El enfoque central de este trabajo se dirige a la evaluación de diversos modelos de riesgo crediticio, específicamente en el contexto del Banco Unión y su dinámica en el mercado hipotecario colombiano, ya que este, como entidad representativa de dicho sector, se enfrenta al desafío de gestionar el riesgo asociado a los créditos hipotecarios en un entorno económico incierto y dinámico. Para lograrlo, se proponen metodologías que implican la revisión detallada de modelos existentes,

incluyendo las series temporales y los modelos econométricos, por mencionar solo algunos.

La elección de esta metodología se sustenta en la idea fundamental de que, como señala Gómez (2020) en sus investigaciones previas, la revisión de modelos y su aplicabilidad práctica son esenciales para el éxito en la gestión del riesgo crediticio. La adaptación de estos modelos al contexto específico del Banco Unión permitirá una toma de decisiones más informada y una gestión más efectiva del riesgo asociado a los créditos hipotecarios.

El impacto esperado de esta investigación es directamente relevante para el Banco Unión al proporcionarle herramientas y enfoques específicos, pues este estudio tiene como objetivo mejorar la gestión del riesgo crediticio. Como destaca Rodríguez (2021), esta mejora no solo beneficia a las instituciones financieras al optimizar sus procesos internos, sino que tiene implicaciones positivas para los prestatarios al promover condiciones crediticias más seguras y sostenibles.

Esta investigación se enfoca en dos direcciones complementarias que son, en primer lugar, la estimación de un modelo Logit tradicional, que es muy utilizado para predecir el incumplimiento, y, en segundo lugar, la aplicación de un modelo fundamentado en la minería de datos a través de la plataforma KNIME, que incorpora métodos de aprendizaje automático y clasificación. Esta combinación posibilita la comparación de enfoques computacionales y estadísticos con el objetivo de determinar cuál de estos métodos proporciona una mayor capacidad para predecir y operar dentro del Banco Unión.

La presente investigación representa un valioso aporte al ámbito financiero al proponer estrategias y modelos específicos para mitigar el riesgo crediticio en el sector hipotecario colombiano, con un enfoque puntual en el Banco Unión.

Aunque el Banco Unión ha creado procedimientos clásicos de evaluación y seguimiento del riesgo de crédito que cumplen con las pautas regulatorias actuales, estas metodologías tienen limitaciones porque no incluyen de manera sistemática modelos predictivos avanzados que combinen variables macroeconómicas, rasgos del crédito y cualidades del prestatario al mismo tiempo.

Esta carencia limita la habilidad de prever con más exactitud el deterioro de la cartera hipotecaria, particularmente en contextos económicos muy volátiles, y disminuye las posibilidades de análisis para segmentar a los clientes, mejorar la distribución de provisiones y reforzar las decisiones estratégicas sobre la gestión del riesgo.

La investigación actual se sitúa en este vacío metodológico y sugiere, como una opción sustentada empíricamente para optimizar la capacidad de pronóstico y

operativa del Banco Unión en la gestión de su cartera hipotecaria, el análisis y la incorporación de modelos estadísticos y técnicas de minería de datos.

### **Datos utilizados**

La investigación se apoya en una amplia gama de fuentes de datos que incluyen datos históricos de cartera hipotecaria del Banco Unión, tasas de interés hipotecarias, indicadores económicos y datos demográficos. Todos ellos se recopilan del Banco Unión, asegurando su confiabilidad y calidad.

La elección de utilizar datos de cartera hipotecaria del Banco Unión es fundamental, ya que permite un análisis detallado de la calidad y el riesgo crediticio en el contexto de las carteras hipotecarias de esta institución financiera. Los datos sobre las tasas de interés hipotecarias son esenciales para comprender cómo las condiciones del mercado afectan la asequibilidad de las hipotecas y, por lo tanto, el riesgo crediticio.

Además, la inclusión de indicadores económicos y datos demográficos enriquece el análisis, pues ayuda a evaluar cómo las condiciones macroeconómicas y las características de los prestatarios influyen en el riesgo crediticio. Esta amplia gama de datos proporciona una base sólida para el análisis de la relación entre las dinámicas del mercado hipotecario y el riesgo crediticio en las carteras hipotecarias del Banco Unión.

### **Esquemas investigativos**

En esta investigación se aplican diversos esquemas investigativos para analizar en profundidad la relación entre las dinámicas del mercado hipotecario y el riesgo crediticio en la cartera del Banco Unión. Estos esquemas incluyen análisis de tendencias, análisis de regresión y modelos econométricos.

El análisis de tendencias permite identificar patrones y cambios a lo largo del tiempo en variables clave relacionadas con la cartera hipotecaria. El análisis de regresión se utiliza para cuantificar la relación entre las variables independientes, como las tasas de interés hipotecarias y la calidad crediticia, y los modelos econométricos avanzados ayudan a evaluar la influencia de múltiples variables; además, son útiles para predecir el riesgo crediticio futuro.

Estos esquemas investigativos se han utilizado con éxito en investigaciones similares (Gujarati y Porter, 2009) y proporcionan un enfoque riguroso y cuantitativo para evaluar la relación entre las dinámicas del mercado hipotecario y el riesgo crediticio en la cartera del Banco Unión.

## **Técnicas matemáticas y estadísticas**

El presente estudio se apoya en técnicas matemáticas y estadísticas avanzadas para evaluar el impacto de las variables del mercado hipotecario en el riesgo crediticio de la cartera del Banco Unión, y para lograrlo se utilizan herramientas econométricas que ayudan a estimar los efectos y las probabilidades asociadas a las variables del mercado hipotecario en relación con el riesgo crediticio. Dichas técnicas han demostrado ser efectivas en investigaciones financieras similares (Wooldridge, 2010) y permiten una evaluación precisa y cuantitativa de los factores que influyen en el riesgo crediticio.

A continuación, se profundiza en cada una de ellas:

### **Análisis de series temporales**

Se realizó un análisis detallado de las tendencias y los patrones a lo largo del tiempo en variables relacionadas con la cartera hipotecaria del Banco Unión. Esta metodología se basa en la aplicación de técnicas estadísticas avanzadas, como el análisis de autocorrelación y las proyecciones de series temporales para identificar relaciones subyacentes. Ejemplos de autores que han utilizado este enfoque en el contexto financiero incluyen a Kothari (2004), en su investigación sobre el comportamiento de las tasas de interés hipotecarias a lo largo del tiempo.

Un estudio destacado que aplicó el análisis de series temporales es el trabajo de Dell'Araccia *et al.* (2012a), quienes utilizaron datos de series temporales para analizar la relación entre las tasas de interés, el apalancamiento y el riesgo bancario.

Los autores observaron cómo las tasas de interés influían en el riesgo crediticio y cómo el apalancamiento de los bancos estaba relacionado con su capacidad para asumir riesgos. El análisis de series temporales permitió identificar patrones y tendencias en estas variables a lo largo del tiempo. Los resultados indicaron que las tasas de interés más altas aumentaban el riesgo crediticio y que un alto apalancamiento aumentaba la exposición a pérdidas crediticias.

### **Modelos econométricos**

Los modelos econométricos avanzados se emplearán para evaluar la influencia de múltiples variables en el riesgo crediticio y para predecir el riesgo crediticio futuro en la cartera hipotecaria del Banco Unión. Se utilizaron herramientas econométricas como los modelos de regresión econométrica y los modelos de series temporales para estimar los efectos y las probabilidades asociadas a las variables del mercado hipotecario en relación con el riesgo crediticio.

Ejemplos de investigaciones que han utilizado estos enfoques incluyen el trabajo de Dell'Araccia *et al.* (2012a) sobre las tasas de interés y el riesgo bancario.

### **Modelos de regresión múltiple**

Estos son los modelos econométricos más frecuentes para el cálculo de la gestión de riesgo crediticio. Dicha técnica cuantifica la relación entre múltiples variables independientes y la variable dependiente del riesgo crediticio en la cartera hipotecaria.

En esta investigación se aplica un análisis de regresión múltiple para determinar factores como las tasas de interés hipotecarias, las características demográficas de los prestatarios y otros indicadores económicos que influyen en el riesgo crediticio.

Ejemplos de autores relevantes incluyen a Dell'Araccia *et al.* (2012a) en su estudio sobre las tasas de interés, el apalancamiento y la toma de riesgos bancarios.

Por su parte, el estudio de Merton (1974) se destacó por su uso de modelos de regresión múltiple en la evaluación del riesgo crediticio en bonos corporativos. Merton propuso un modelo que consideraba la estructura de las tasas de interés, la calidad crediticia de las empresas emisoras y otros factores para evaluar el riesgo de incumplimiento. Los modelos de regresión múltiple permitieron cuantificar la influencia de distintas variables en el riesgo crediticio de los bonos corporativos, lo que ayudó a los inversionistas y emisores a tomar decisiones más informadas.

### **Modelo SPSS**

Es ampliamente utilizado en el análisis de riesgo crediticio, pues especifica la probabilidad de que una variable cualitativa tome un valor particular y se utiliza comúnmente en situaciones donde la variable dependiente es binaria. Un ejemplo relevante es el estudio de Khandani *et al.* (2010), en el que se aplicó un modelo Logit<sup>1</sup> para evaluar el riesgo crediticio en préstamos al consumo.

Este estudio utilizó variables como puntajes de crédito, ingresos, historial crediticio y características demográficas de los prestatarios. El modelo Logit permitió calcular la probabilidad de incumplimiento de los prestatarios y clasificarlos en categorías de riesgo. Los resultados mostraron que dicho modelo era eficaz para identificar a los prestatarios de alto riesgo y contribuyó a una mejor gestión del riesgo crediticio.

---

<sup>1</sup> Un modelo Logit, también conocido como regresión logística en finanzas, es una herramienta estadística que se utiliza para predecir la probabilidad de que ocurra un evento binario, como el impago de un crédito o la aprobación de un préstamo.

## PLANTEAMIENTO DEL PROBLEMA

La pandemia de covid-19 provocó que la economía colombiana sufriera una caída significativa, siendo esta una de las más pronunciadas en décadas. Datos proporcionados por el Banco de la República (2022) revelan una disminución del 7 % en el Producto Interno Bruto (PIB) durante el año 2020. La capacidad de pago de los hogares se vio afectada por la disminución del PIB y el incremento del desempleo, lo que produjo un aumento en la morosidad de los créditos hipotecarios. Estas modificaciones revelaron las debilidades en los procesos tradicionales de evaluación crediticia y demostraron que se necesitan nuevos modelos que representen las condiciones actuales del mercado.

La situación económica afectó directamente a las familias colombianas, que experimentaron un aumento en las tasas de desempleo y una reducción de ingresos. Según el Departamento Administrativo Nacional de Estadística (DANE, 2021), la tasa de desempleo alcanzó el 14 % en el 2020, subrayando la magnitud de los desafíos económicos que enfrenta la población. Este escenario ha impulsado un aumento en el incumplimiento de las obligaciones financieras, destacando los créditos hipotecarios como uno de los sectores más afectados, con prestatarios que encuentran dificultades para cumplir con sus compromisos financieros.

Ante este panorama desafiante, las entidades financieras implementaron medidas de alivio, como la reprogramación de cuotas y la prórroga de plazos para préstamos hipotecarios, con el propósito de mitigar el impacto económico directo sobre los prestatarios (González, 2021). Sin embargo, es importante señalar que estas medidas han evolucionado y, actualmente, las entidades financieras ya no están ofreciendo estas ayudas. La persistente incertidumbre económica crea un escenario volátil en el mercado financiero, planteando desafíos continuos en el sector hipotecario.

El comportamiento de la cartera hipotecaria del Banco Unión sigue siendo afectado por la incertidumbre económica. El incremento de las tasas de interés hace que las obligaciones financieras sean más costosas, mientras que la reducción del ingreso disponible impacta directamente en el cumplimiento de los deudores. Pese a que se implementaron medidas de alivio transitorias, su eliminación evidenció la necesidad de mecanismos más precisos para predecir el riesgo.

Dicha incertidumbre se presenta como uno de los principales riesgos para la cartera hipotecaria. Según el Informe de Estabilidad Financiera del Banco de la República (2022), la posibilidad de un aumento en las tasas de interés, consecuencia directa del problema mencionado, encarece los préstamos hipotecarios y afecta la demanda de viviendas. Este escenario intensifica el riesgo de incumplimiento, ya

que los prestatarios podrían enfrentar dificultades adicionales para cumplir con sus obligaciones si su situación económica se deteriora aún más.

El Banco Unión necesita instrumentos que puedan prever deterioros en el futuro, reconocer perfiles más vulnerables y adecuar a tiempo sus políticas de concesión y monitoreo. La falta de un modelo optimizado reduce la habilidad para prevenir riesgos y deteriora la fortaleza de la cartera hipotecaria.

En este contexto, el presente estudio se configura como un esfuerzo para abordar y comprender los desafíos económicos planteados. La investigación, centrada en el Banco Unión, tiene como objetivo utilizar datos históricos de la cartera hipotecaria proporcionados por el propio banco. Aplicando modelos estadísticos de series de tiempo se busca identificar y cuantificar los riesgos asociados a la incertidumbre económica, logrando una evaluación más precisa y predictiva de los posibles efectos en la cartera hipotecaria.

La relevancia de esta investigación radica en su capacidad para ofrecer soluciones sólidas a un problema complejo y multifacético. A través del análisis de ejemplos internacionales y la revisión de prácticas en instituciones financieras similares se busca proporcionar una visión holística de la problemática de la cartera hipotecaria en un contexto de incertidumbre económica. La urgencia de comprender y mitigar los riesgos asociados a esta incertidumbre se convierte en un componente esencial para preservar la estabilidad financiera y garantizar las condiciones crediticias sostenibles en el sector hipotecario colombiano.

Esta investigación también contribuirá al fortalecimiento de la capacidad del Banco Unión para afrontar desafíos económicos imprevistos y, a su vez, promoverá una gestión del riesgo más efectiva en el ámbito hipotecario. En vista de la complejidad del escenario actual, se propone brindar una perspectiva integral que no solo identifique los desafíos, sino que también ofrezca herramientas prácticas y estrategias específicas para mitigar el riesgo crediticio en un entorno de incertidumbre económica.

Así las cosas, el presente texto emerge como un aporte al entendimiento y la gestión de la cartera hipotecaria en un escenario económico incierto y destaca la necesidad de estrategias adaptativas y proactivas para garantizar la estabilidad financiera y el bienestar de los prestatarios en el mercado hipotecario colombiano, específicamente en el ámbito del Banco Unión.

De esa forma, la combinación de factores económicos adversos a nivel macro, la sensibilidad del mercado hipotecario y las restricciones de los métodos tradicionales para evaluar el crédito muestran que es necesario disponer de instrumentos más exactos para prever el deterioro de la cartera y guiar decisiones dentro del Banco Unión en un tiempo apropiado.

Esta circunstancia presenta una brecha que la investigación intenta abordar y conduce a la formulación de la pregunta principal que orientará el estudio: en un contexto de incertidumbre económica, ¿cuál modelo de riesgo de crédito posibilita prever con mayor eficacia el deterioro de la cartera hipotecaria del Banco Unión y colaborar en la disminución del riesgo crediticio?

## JUSTIFICACIÓN

La situación económica de Colombia, después de la pandemia del covid-19, ha aumentado los peligros vinculados a la actividad crediticia, especialmente en el sector hipotecario. Esto requiere un análisis detallado de las estrategias empleadas para su manejo. Dada esta situación, el presente estudio se justifica, ya que es necesario reforzar la habilidad de las entidades financieras para prever y reducir el riesgo crediticio en situaciones de elevada incertidumbre a nivel macroeconómico, lo cual es esencial para mantener la estabilidad financiera y asegurar el flujo del crédito.

Mishkin (2011) argumenta, desde un punto de vista teórico, que una gestión apropiada del riesgo crediticio es fundamental para la estabilidad del sistema financiero, sobre todo en tiempos de tensión económica. Esta idea es relevante para el Banco Unión, pues su cartera hipotecaria supone una parte significativa de la cartera total y su estabilidad a lo largo del tiempo, demostrada en tasas de morosidad bajas y rentabilidad constante, se enfrenta a nuevos retos por el incremento del desempleo, los tipos de interés y la inflación.

En este escenario, es necesario, de manera estratégica, modificar las prácticas tradicionales de gestión del riesgo. Según Malhotra y Malhotra (2015), la habilidad para valorar y manejar el riesgo de manera eficaz no solo determina la estabilidad, sino también la sostenibilidad de las entidades financieras a largo plazo. Además, Laeven y Valencia (2012) enfatizan que los riesgos crediticios se ven aumentados debido a los entornos macroeconómicos volátiles, especialmente en carteras de largo plazo como la hipotecaria.

Desde el punto de vista metodológico, la utilización de modelos estadísticos sofisticados es una opción relevante para optimizar la capacidad predictiva del deterioro crediticio. Jorion (2007) enfatiza lo útil que resultan estos modelos para prever y cuantificar riesgos económicos en situaciones inciertas, a la vez que Shiller (2012) destaca la importancia de que el campo financiero avance hacia perspectivas más prácticas y adaptadas al contexto.

Por lo tanto, esta investigación tiene sentido no solo por su contribución práctica a la gestión del riesgo crediticio del Banco Unión, sino también por su aporte al conocimiento aplicado en finanzas al analizar modelos predictivos dentro del particular escenario del mercado hipotecario de Colombia.

## **OBJETIVOS**

### **General**

Evaluar un modelo de riesgo crediticio óptimo para reducir el riesgo de crédito en la cartera hipotecaria del Banco Unión.

### **Específicos**

- Realizar una revisión de la literatura sobre los modelos de riesgo crediticio efectivos para anticipar y gestionar el riesgo de crédito en la cartera hipotecaria.
- Evaluar modelos de riesgo crediticio para identificar el enfoque más efectivo en la gestión de riesgos en la cartera hipotecaria del Banco Unión.
- Desarrollar recomendaciones precisas para la implementación de estrategias de gestión de riesgos que consideren tanto el modelo de riesgo crediticio elegido y las condiciones cambiantes del mercado.

## MARCO TEÓRICO Y MARCO CONCEPTUAL

El análisis del riesgo crediticio ha adquirido una relevancia creciente en el sistema financiero, impulsado por la necesidad de anticipar el comportamiento de pago de los prestatarios y reducir las pérdidas derivadas del incumplimiento. En el caso de la cartera hipotecaria, la literatura coincide en que el riesgo crediticio se ve determinado por una interacción compleja entre factores macroeconómicos, como el desempleo, el costo financiero de la deuda y el crecimiento económico y características propias del prestatario, asociadas a su nivel de ingresos, endeudamiento, tipo de garantía e historial crediticio. Esta combinación de elementos internos y externos exige enfoques analíticos que permitan distinguir adecuadamente entre prestatarios con alta y baja probabilidad de incumplimiento.

### **Riesgo de crédito en la cartera hipotecaria**

La definición propuesta por Jorion (2007) conceptualiza el riesgo de crédito como la probabilidad de que un prestatario incumpla sus obligaciones contractuales, estableciendo una base teórica sólida para el análisis del default en carteras hipotecarias. Desde esta perspectiva, el riesgo crediticio se entiende como un evento discreto, en el cual el prestatario cumple o no cumple, lo que sitúa el problema en el ámbito de la clasificación más que en el de la predicción continua.

Smith (2010) profundiza en los determinantes del riesgo crediticio al resaltar la calidad crediticia del prestatario como uno de los factores centrales del incumplimiento. Su análisis evidencia que historiales financieros sólidos se asocian sistemáticamente con menores probabilidades de default, lo que refuerza la necesidad de modelos que permitan clasificar a los clientes según su perfil de riesgo a lo largo del ciclo de vida del crédito.

Desde una óptica complementaria, Black (2015) analiza el impacto de las tasas de interés sobre el riesgo de crédito hipotecario, señalando que los incrementos en estas tasas pueden deteriorar la capacidad de pago de los prestatarios y elevar la probabilidad de incumplimiento. Este efecto resulta especialmente relevante en créditos de largo plazo, donde pequeñas variaciones en las condiciones financieras pueden modificar sustancialmente el perfil de riesgo del deudor.

La dinámica del mercado inmobiliario constituye otro determinante relevante del riesgo crediticio. Dell'Ariccia *et al.* (2012b) destacan que las fluctuaciones en el valor de los inmuebles, así como la formación y corrección de burbujas inmobiliarias, inciden directamente en la calidad de la cartera hipotecaria. Estas dinámicas refuerzan la necesidad de modelos de riesgo que integren variables macroeconómicas y del mercado inmobiliario en la clasificación del default.

En el contexto colombiano, Meneses y Macuacé (2011) aportan evidencia empírica sobre la relación entre las características crediticias de los prestatarios y su probabilidad de incumplimiento, confirmando que historiales financieros sólidos y condiciones crediticias favorables reducen, de una manera significativa, el riesgo de default. Los autores también analizan la influencia de las tasas de interés y del mercado inmobiliario local, subrayando la importancia de contextualizar los modelos de riesgo crediticio a las particularidades del entorno económico colombiano.

El Banco de la República (2023) complementa este enfoque al analizar la evolución reciente del mercado inmobiliario y del crédito hipotecario en Colombia, identificando fuentes potenciales de riesgo asociadas a la desaceleración económica, la inflación elevada y el aumento de las tasas de interés. Estos factores macroeconómicos refuerzan la relevancia de contar con herramientas analíticas que permitan anticipar el deterioro de la cartera hipotecaria en escenarios de mayor volatilidad.

### **Gestión de provisiones**

La gestión de provisiones, como elemento esencial en la gestión del riesgo crediticio, se ha convertido en un componente estratégico fundamental para preservar la solidez financiera de las instituciones. Este enfoque, según Merton (1974), cumple con las obligaciones regulatorias y se erige como una herramienta proactiva para anticipar y mitigar las pérdidas derivadas del riesgo crediticio, especialmente en el contexto de las carteras hipotecarias.

En el análisis de la literatura académica se evidencia la obra de Kealhofer (2010) como un faro guía en el desarrollo de enfoques y modelos que les permiten a las instituciones calcular provisiones de manera precisa. Esta investigación no solo cumple con las exigencias regulatorias, sino que propone métodos avanzados para anticipar pérdidas, brindando a las instituciones financieras una ventaja estratégica en la gestión del riesgo crediticio.

Estudios adicionales, como los de Bluhm *et al.* (2002), profundizan en la gestión de provisiones desde una perspectiva que abarca diferentes tipos de préstamos. Estos autores argumentan que la gestión eficiente de provisiones consiste en cumplir con las regulaciones, de comprender la dinámica subyacente de los diferentes activos financieros y ajustar las reservas en consecuencia.

Al considerar específicamente las carteras hipotecarias, autores como Phillips y Siems (2008) han explorado cómo la gestión de provisiones puede adaptarse a la dinámica única de los préstamos para vivienda. En sus investigaciones, destacan la importancia de evaluar no solo la calidad crediticia de los prestatarios, sino también la dinámica del mercado inmobiliario y las tasas de interés para determinar provisiones adecuadas.

Este enfoque holístico, respaldado por la investigación de autores diversos, subraya la gestión de provisiones como un instrumento multifacético que va más allá del simple cumplimiento normativo. Los estudios de autores como Becker *et al.* (2012) ofrecen perspectivas adicionales al explorar cómo la gestión de provisiones puede adaptarse a entornos económicos variables, anticipando de manera proactiva las fluctuaciones del mercado y mejorando la capacidad de las instituciones para enfrentar eventos inesperados.

La gestión de provisiones en el ámbito de las carteras hipotecarias emerge como un campo de estudio rico y complejo, respaldado por la obra de diversos autores. Estos estudios no solo validan la importancia de cumplir con las regulaciones, sino que también proporcionan herramientas y enfoques avanzados para anticipar y mitigar pérdidas, consolidando así la solidez financiera de las instituciones en un entorno económico cambiante.

### **Cómo se evalúa el riesgo crediticio**

La Superintendencia Financiera de Colombia (1995) ha establecido normativas y directrices fundamentales para la evaluación del riesgo crediticio en diversas categorías de cartera, incluyendo, de manera específica, la cartera hipotecaria. Este proceso de evaluación se lleva a cabo mediante la implementación de un Sistema de Administración del Riesgo Crediticio (SARC), el cual debe ser adoptado por todas las entidades financieras.

En términos generales, el SARC combina políticas y procedimientos que abarquen la gestión integral del riesgo crediticio, contemplando aspectos como la estructura organizacional, los límites de exposición crediticia y pérdida tolerada, los criterios para la aprobación de créditos y el manejo de garantías, así como el monitoreo, el control y la constitución de provisiones. Además, este sistema aborda aspectos cruciales como el capital económico, la recuperación de cartera y las políticas relacionadas con las bases de datos que respaldan el SARC.

En concordancia con estas directrices, el capítulo II de la normativa detalla los principios y criterios generales que rigen la evaluación del riesgo crediticio, específicamente en el contexto de la cartera de créditos, incluyendo la cartera hipotecaria. La superintendencia enfatiza la importancia de que las entidades financieras adopten y apliquen de manera efectiva el SARC para garantizar una gestión integral y efectiva del riesgo crediticio. Este enfoque incluye la implementación de modelos para la estimación y cuantificación de pérdidas esperadas, así como el establecimiento de un sistema de provisiones y procesos de control interno (Superintendencia Financiera de Colombia, 1995).

Es fundamental destacar que esta normativa proporciona un marco sólido y estructurado para la gestión del riesgo crediticio en el ámbito financiero colombiano,

asegurando un enfoque sistemático y proactivo por parte de las entidades financieras en la administración de su cartera, especialmente la hipotecaria.

Desde un enfoque internacional, el Marco de Basilea es la normativa fundamental centrada en el riesgo crediticio, y representa el conjunto integral de estándares establecidos por el Comité de Basilea sobre Supervisión Bancaria (BCBS), la entidad líder en la formulación de normas globales para la regulación prudencial de instituciones bancarias (Banco de Pagos Internacionales, 2019). La adopción plena de estos estándares por parte de los miembros del BCBS implica su implementación y aplicación a nivel internacional en jurisdicciones con bancos activos globalmente.

El Marco de Basilea III, en especial, surge como respuesta directa a la crisis financiera del 2007-2009, representando un conjunto de medidas internacionalmente acordadas. Su propósito principal radica en fortalecer la regulación, supervisión y gestión del riesgo en el sector bancario. Estos estándares, siendo mínimos aplicables a bancos con presencia internacional, requieren el compromiso de los miembros para su implementación en sus respectivas jurisdicciones nacionales o regionales dentro de los plazos establecidos por el comité (Banco de Pagos Internacionales, 2019).

Este marco regulatorio más riguroso impone requisitos más estrictos para la gestión del riesgo crediticio. Entre sus disposiciones clave se incluyen la introducción de un coeficiente de apalancamiento mínimo, una relación de cobertura de liquidez y un marco para el riesgo de contraparte central. Además, Basilea III establece nuevos requisitos de capital para los bancos, que comprenden un colchón de conservación de capital y un colchón de capital anticíclico. En conjunto, estas medidas buscan fortalecer la solidez y estabilidad del sistema financiero global mediante una gestión más robusta y eficaz del riesgo crediticio en las instituciones bancarias.

Además del Marco de Basilea existen otras normas internacionales para evaluar el riesgo crediticio. Algunas de ellas son:

- ISO 31000: proporciona un marco para la gestión del riesgo en cualquier tipo de organización. La norma se centra en la identificación, evaluación y tratamiento del riesgo, y se aplica a todos los tipos de riesgo, incluyendo el riesgo crediticio 1.
- COSO ERM: proporciona un enfoque integral para la gestión del riesgo empresarial. El marco se centra en la identificación, evaluación y gestión del riesgo en toda la organización, y se aplica a todos los tipos de riesgo, incluyendo el riesgo crediticio
- Norma Internacional de Información Financiera 9 (NIIF 9): se relaciona principalmente con las normativas y prácticas contables que orientan la

evaluación y presentación de activos financieros en los estados financieros de las entidades. Fue emitida por el Consejo de Normas Internacionales de Contabilidad.

De acuerdo con el Consejo Técnico de la Contaduría Pública (2014), esta norma se basa en lo siguiente:

- **Objetivo:** establece principios para el reconocimiento, la medición, presentación y revelación de activos financieros y pasivos financieros. Uno de los aspectos centrales abordados por esta norma es la evaluación y gestión de riesgos crediticios asociados a estos activos financieros.
- **Enfoque en riesgo crediticio:** introduce un modelo basado en pérdidas esperadas para el reconocimiento de pérdidas crediticias. Este enfoque busca anticipar las pérdidas crediticias a lo largo de la vida del instrumento financiero, en lugar de esperar a que ocurra un incumplimiento. Se identifica el riesgo crediticio desde el momento inicial de reconocimiento del activo financiero.
- **Categorías de activos financieros:** la norma clasifica los activos financieros en varias categorías, como aquellos al valor razonable con cambios en resultados, instrumentos de deuda con cambios en otro resultado integral y activos financieros al costo amortizado. Cada categoría tiene implicaciones específicas para la evaluación del riesgo crediticio.
- **Divulgación:** la NIIF 9 también establece requerimientos detallados para la divulgación de información relacionada con riesgos crediticios. Esto incluye información sobre la metodología utilizada para evaluar el riesgo crediticio, las características crediticias de los activos financieros y las pérdidas crediticias esperadas.

La NIIF 9 se posiciona como un marco contable integral que guía a las entidades financieras en la evaluación y presentación de activos financieros, incluyendo una consideración detallada del riesgo crediticio. La adopción de esta norma busca mejorar la relevancia y la representación fiel de la situación financiera de las entidades al proporcionar información más anticipada sobre las pérdidas crediticias esperadas.

### **Modelos de riesgo crediticio como herramientas de clasificación**

Los modelos de riesgo crediticio constituyen el núcleo analítico para la gestión del riesgo financiero en las entidades bancarias. Altman (1968) establece los fundamentos conceptuales de estos modelos al definirlos como instrumentos

orientados a estimar la probabilidad de incumplimiento de los prestatarios, introduciendo una lógica de clasificación que distingue entre clientes solventes y clientes en riesgo.

La evolución de estos modelos ha estado marcada por el tránsito desde enfoques basados en reglas y puntajes simples hacia modelos estadísticos y probabilísticos más sofisticados. Khandani *et al.* (2010) destacan que este desarrollo responde a la necesidad de mejorar la capacidad predictiva y de clasificación en contextos donde el comportamiento crediticio presenta patrones complejos y no lineales.

En el ámbito específico de la cartera hipotecaria, Dell’Ariccia *et al.* (2012b) subrayan que los modelos de riesgo deben adaptarse a las características particulares de este tipo de créditos, incorporando variables relacionadas con el mercado inmobiliario, las condiciones macroeconómicas y la estructura del financiamiento. La selección del modelo adecuado resulta determinante, ya que una clasificación incorrecta puede derivar en subestimaciones o sobreestimaciones del riesgo, con efectos directos sobre la gestión de la cartera y la estabilidad financiera de la institución.

### **Modelos estadísticos para la clasificación del incumplimiento**

Desde una perspectiva estadística, el análisis del riesgo crediticio se formula principalmente como un problema de clasificación binaria, en el que el prestatario se encuentra en estado de cumplimiento o de incumplimiento. En este contexto, los modelos logísticos han adquirido un papel central debido a su capacidad para estimar probabilidades asociadas a eventos dicotómicos.

El modelo Logit se consolida como uno de los enfoques más utilizados en la literatura para el análisis del default. Su estructura teórica permite modelar la relación entre la probabilidad de incumplimiento y un conjunto de variables explicativas, garantizando que las probabilidades estimadas se mantengan dentro del intervalo lógico entre 0 y 1. Lucas y Klaassen (2006) destacan que esta especificación facilita el análisis del impacto marginal de factores macroeconómicos y financieros sobre la probabilidad de deterioro crediticio, aportando tanto capacidad predictiva como interpretabilidad.

### **Enfoques no lineales y técnicas de minería de datos**

El avance de la capacidad computacional ha impulsado el desarrollo de técnicas de minería de datos y aprendizaje automático aplicadas al riesgo crediticio. Estos enfoques complementan los modelos estadísticos tradicionales al permitir la identificación de patrones no lineales, interacciones complejas entre variables y estructuras jerárquicas de decisión, especialmente útiles en contextos donde los datos presentan alta dimensionalidad o desbalance entre clases.

Los modelos basados en árboles de decisión destacan dentro de este conjunto por su capacidad para clasificar prestatarios de manera transparente, generando reglas explícitas que facilitan la interpretación de los resultados. Desde el punto de vista teórico, estos modelos resultan particularmente valiosos en entornos regulatorios, donde la explicación de las decisiones de clasificación constituye un requisito fundamental para la validación y el seguimiento de los modelos de riesgo crediticio.

En el caso de las carteras hipotecarias, la capacidad de los árboles de decisión para capturar relaciones no lineales entre las variables macroeconómicas, las características del crédito y los atributos del prestatario refuerza su utilidad como complemento a los modelos logísticos, contribuyendo a una clasificación más precisa del default sin sacrificar la claridad interpretativa.

## DISEÑO METODOLÓGICO

### **Enfoque y diseño de la investigación**

La presente investigación adopta un enfoque cuantitativo, de tipo aplicado, orientado a la evaluación de modelos de riesgo crediticio para la clasificación de clientes hipotecarios en estados de cumplimiento e incumplimiento. El diseño es no experimental y de carácter explicativo-predictivo (Hair *et al.*, 2019), dado que se analizan datos históricos de la cartera hipotecaria del Banco Unión con el propósito de estimar la probabilidad de default y apoyar la toma de decisiones en la gestión del riesgo crediticio.

Además, se incorpora un enfoque descriptivo que permite analizar detalladamente la calidad de la cartera hipotecaria, lo que contribuye a una comprensión integral de la situación (Gujarati y Porter, 2009).

### **Fuente de datos y población de estudio**

La base de datos utilizada en la investigación fue proporcionada por el Banco Unión y contiene información histórica de créditos hipotecarios vigentes y deteriorados. La población de estudio está conformada por los prestatarios de la cartera hipotecaria del Banco durante el período analizado, incorporando variables macroeconómicas, características propias del crédito y atributos financieros y sociodemográficos de los deudores.

Esta investigación se basa en un conjunto de datos compuesto por 1782 observaciones de créditos hipotecarios otorgados por el Banco Unión. Del total, 1674 registros mantuvieron el estado de cumplimiento y 108 incurrieron en default; es decir, se registraron como créditos en mora. La tasa de incumplimiento en la base original, comparativamente, es del 6,06 %, mientras que la de créditos no deteriorados es del 93,94 %. Esta distribución muestra una particularidad estructural de las carteras hipotecarias, donde el incumplimiento es un evento raro pero muy significativo para la administración del riesgo crediticio.

La notable asimetría entre las clases supone retos significativos para la evaluación y estimación de modelos de clasificación, pues puede generar prejuicios a favor de la clase mayoritaria y restringir la habilidad de los modelos para reconocer correctamente a los prestatarios con una probabilidad más alta de incumplimiento. Bajo estas circunstancias, la comprensión explícita de la estructura del dataset resulta fundamental para interpretar de manera correcta los indicadores de desempeño y justificar las decisiones metodológicas adoptadas en etapas posteriores del análisis.

## Modelación y definición de variables

La selección de los modelos estadísticos propuestos se basa en la naturaleza específica de los datos de la cartera hipotecaria del Banco Unión y los objetivos de la investigación.

### Modelación

#### Paso 1: Definición de variables

Tabla 1. Variables que serán utilizadas

Variable	Descripción	Lo que mide	Razón teórica para incluirla
Ingresos mensuales del cliente	Nivel de ingresos del solicitante del crédito	Capacidad de pago del cliente	Los ingresos afectan la capacidad del cliente para cumplir con sus pagos. Un ingreso mayor reduce la probabilidad de default.
Tasa de interés del crédito	Tasa que el banco cobra sobre el monto prestado	Costo del crédito para el solicitante	Una tasa de interés más alta incrementa la carga financiera del cliente, lo que podría aumentar la probabilidad de impago. Un monto mayor puede implicar un mayor riesgo para el Banco y una mayor probabilidad de incumplimiento si el cliente no tiene capacidad de pago suficiente.
Monto del crédito solicitado	Valor del préstamo solicitado por el cliente	Nivel de endeudamiento asumido	Un plazo más largo puede aumentar el riesgo de impago debido a la mayor incertidumbre futura sobre los ingresos del cliente.
Plazo del crédito (meses)	Número de meses en los que se pagará el crédito	Duración del compromiso financiero	La edad puede influir en la estabilidad laboral y en el comportamiento financiero. Por ejemplo, personas muy jóvenes o
Edad del cliente	Edad actual del solicitante	Ciclo de vida y estabilidad financiera	

Variable	Descripción	Lo que mide	Razón teórica para incluirla
Historial crediticio ( <i>score</i> o número de créditos previos)	Información sobre el comportamiento de pago pasado	Fiabilidad del cliente como pagador	muy mayores pueden presentar más riesgos Un buen historial crediticio reduce la probabilidad de impago; un mal historial la incrementa
Relación deuda/ingresos	Proporción entre las deudas actuales y los ingresos mensuales	Nivel de apalancamiento del cliente	Una mayor relación deuda/ingresos indica menor capacidad de pago disponible y, por tanto, mayor riesgo de default

Fuente: elaboración propia.

El proceso de búsqueda de la especificación del modelo para estimar los determinantes de la probabilidad de default se desarrolló de manera iterativa, combinando criterios teóricos del análisis del riesgo crediticio con procedimientos empíricos de depuración y contraste estadístico. El objetivo fue obtener un modelo parsimonioso, estable y coherente con la lógica económica subyacente al comportamiento de pago de los deudores.

Inicialmente, se partió de un conjunto amplio de variables socioeconómicas y crediticias que incluían información sobre las características del titular del crédito (género, edad, estado civil, nivel educativo, tipo de vivienda, ingreso mensual) y sobre las condiciones del préstamo (monto, plazo, tasa de interés, antigüedad del crédito, participación en proyectos VIS, entre otras).

Se decidió excluir las variables de *mora\_maxima* y *calificacion\_riesgo\_bu*, dado que ambas se encuentran directamente relacionadas con la construcción de la variable dependiente *default*; el indicador de mora máxima define la condición de incumplimiento y la calificación de riesgo se deriva del historial de mora del cliente.

Su inclusión habría generado problemas de endogeneidad, ya que estas variables contienen información posterior o simultánea al evento de default, sesgando los coeficientes e inflando artificialmente el poder explicativo del modelo.

Las variables continuas fueron transformadas cuando la distribución presentaba fuerte asimetría o cuando se buscaba interpretar efectos en términos relativos. Así,

se utilizaron los logaritmos del ingreso, del monto del crédito y de la tasa de interés efectiva anual, lo que permitió interpretar los coeficientes como elasticidades y reducir la influencia de valores extremos. En el caso de la edad, se incluyó además su término cuadrático ( $edad^2$ ) con el fin de capturar posibles efectos no lineales asociados al ciclo de vida. Este enfoque reconoce que la propensión al incumplimiento puede variar según la etapa vital, ya que los prestatarios jóvenes suelen enfrentar mayor inestabilidad laboral, mientras que los de edad media exhiben ingresos más estables y menor riesgo, y los mayores pueden experimentar nuevamente vulnerabilidad por retiro o reducción de ingresos. La inclusión del término cuadrático permite modelar esta relación curvilínea y estimar empíricamente el punto en el que cambia la pendiente del efecto de la edad sobre la probabilidad de default.

Para las variables categóricas como estado civil, nivel educativo, tipo de vivienda y proyecto VIS se construyeron variables *dummy* que representan cada categoría posible, dejando siempre una como base de comparación. Esta estrategia se adoptó en lugar de utilizar las variables categóricas originales, dado que permite capturar diferencias específicas entre grupos sin imponer relaciones lineales u ordinales que no son conceptualmente válidas. El uso de *dummies* también facilita la interpretación de los efectos marginales y la comparación de probabilidades de incumplimiento entre categorías. En casos de categorías con muy baja frecuencia (por ejemplo, Magíster o Viudo), se consideró su agrupación o exclusión para evitar inestabilidad en la estimación.

La selección final de variables se basó en criterios de relevancia teórica, significación estadística y ausencia de colinealidad excesiva. Se realizaron pruebas de diagnóstico mediante el factor de inflación de la varianza (VIF), la comparación de criterios de información (AIC y BIC) y la evaluación del ajuste global (pruebas de bondad de ajuste y curvas ROC). Asimismo, se estimaron modelos alternativos bajo especificaciones Logit y KNIME para comprobar la solidez de los signos y la magnitud de los coeficientes. En todos los casos, los resultados se interpretaron en términos de efectos marginales promedio, lo que facilita su lectura en probabilidades y no solo en razón de momios.

Este proceso permitió construir una especificación sólida sustentada tanto en la teoría del riesgo crediticio como en la evidencia empírica. El modelo resultante integra variables *ex ante* que reflejan las condiciones socioeconómicas del deudor y las características del crédito, evitando el uso de información derivada del propio incumplimiento. De esta forma, las estimaciones obtenidas permiten analizar de manera coherente los factores asociados al default y constituyen una base confiable para la evaluación de riesgo y el diseño de políticas de crédito más precisas.

El objetivo del ejercicio fue identificar los determinantes de la probabilidad de default de los créditos otorgados por el Banco a partir de información disponible al momento del desembolso.

Esto implica estimar un modelo probabilístico donde la variable dependiente es una medida binaria de incumplimiento (default) y las covariables representan atributos individuales, socioeconómicos y crediticios.

### **Plataforma tecnológica y entorno de análisis**

El Banco Unión cuenta con la plataforma de minería de datos KNIME, un instrumento que posibilita la creación de varios modelos de riesgo mediante la integración de nodos para modelar, analizar y transformar. Esta plataforma permite la automatización de procesos y la selección de modelos a través de flujos de trabajo visuales, lo que ayuda al análisis de diversos algoritmos y la elección del que tenga un rendimiento superior.

La utilización de KNIME resulta adecuada para este estudio, dado que permite integrar procesos de preparación de datos, entrenamiento de modelos y evaluación del desempeño dentro de un mismo entorno analítico.

### **Preparación y limpieza de los datos**

Se llevaron a cabo diversos procesos fundamentales para poner en marcha el flujo en KNIME. Primero se cargó la base de datos proporcionada por el Banco, utilizando métodos de depuración que comprendían la conversión de variables a texto cuando era necesario y la supresión de espacios o símbolos que impedían una adecuada lectura del modelo.

Estos procedimientos permitieron garantizar la consistencia de la información y la correcta interpretación de las variables por parte de los algoritmos de clasificación.

### **Depuración inicial del conjunto de datos**

La base original incluía un conjunto amplio de variables relacionadas con el crédito y las características del cliente, las cuales se detallan a continuación.

id\_credito, fecha\_desembolso, género, edad, rango\_edad, ingreso\_mensual, rango\_salario\_smlv, estado\_civil, nivel\_estudio, tipo\_vivienda, país, proyecto\_vis, monto\_crédito, plazo\_meses, tasa\_interés\_e\_a, calificación\_riesgo\_bu, mora\_máxima, default, antigüedad\_crédito\_meses, además de transformaciones y variables auxiliares (edad\_cuadrado, log\_ingreso, log\_monto, prob\_default, etc.).

Durante la depuración se realizaron los siguientes pasos:

- Revisión de consistencia para la detección de valores extremos o negativos en ingreso, monto y tasas; tratamiento de datos faltantes.

- Transformación de variables categóricas en *dummies* para permitir su inclusión en el modelo sin imponer supuestos de orden o linealidad.

También se hizo una elección conceptual de variables explicativas, así:

- Se partió de un universo amplio de posibles determinantes, que se fue depurando con base en tres criterios complementarios
- Se tuvo en cuenta la relevancia teórica que supone la coherencia económica y comportamental con la probabilidad de impago.

Se realizó una medición adecuada con base en la disponibilidad y confiabilidad de los datos, teniendo en cuenta:

- La no endogeneidad; es decir, la exclusión de variables que, por construcción, contienen información posterior al evento o derivada del propio default.

Bajo estos criterios se excluyeron dos variables clave, que son:

- *mora\_máxima*: constituye una medida directa del comportamiento de pago y es parte de la definición del default. Su inclusión generaría una dependencia mecánica con la variable dependiente.
- *calificación\_riesgo\_bu*: aunque es útil operacionalmente, incorpora información del historial de mora y comportamiento reciente del crédito. Usarla para explicar el default introduce endogeneidad y circularidad.

En consecuencia, el modelo se restringe a variables observables en INGRESAR FECHA DE CONSULTA DE LOS DATOS.

### **Transformaciones y tratamiento de variables continuas**

Para las variables monetarias (*ingreso\_mensual*, *monto\_crédito*, *tasa\_interés\_e\_a*), se aplicaron transformaciones logarítmicas con el fin de:

- Reducir asimetrías y heterocedasticidad.
- Interpretar los coeficientes en términos de semielasticidades.

De ahí surgen las variables *log\_ingreso*, *log\_monto* y *log\_tasa*.

La variable edad se incorporó junto con su término cuadrático (*edad\_cuadrado*) para capturar no linealidades en el ciclo de vida, dado que económicamente el comportamiento de pago tiende a variar con la edad de la siguiente manera:

- En edades tempranas, los ingresos inestables o la menor experiencia financiera aumentan el riesgo.
- En edades medias, la estabilidad laboral reduce la probabilidad de impago.
- En edades altas, la transición a retiro o disminución de ingresos puede incrementar nuevamente el riesgo.

El término cuadrático permite que el modelo identifique empíricamente ese patrón curvo sin imponerlo *a priori*.

La variable antigüedad del crédito se mantuvo como indicadora de madurez del producto, siendo que una mayor antigüedad puede reflejar experiencia crediticia acumulada o, alternativamente, mayor exposición temporal al riesgo.

### **Tratamiento de variables categóricas y creación de dummies**

Las variables estado civil, nivel educativo, tipo de vivienda y participación en proyectos VIS se trataron mediante codificación dummy (0/1).

La decisión de crear variables binarias en lugar de usar categorías directamente se basa en:

- Evitar imponer un orden ficticio entre categorías nominales.
- Permitir estimar efectos marginales independientes de cada grupo.
- Facilitar la interpretación de los coeficientes respecto a una categoría base (omitida para evitar multicolinealidad perfecta).

Por ejemplo, en el caso de *estado\_civil* se generaron seis indicadores (*d\_estado\_civil1* a *d\_estado\_civil6*) que identifican categorías como casado, divorciado, separado, soltero, unión libre y viudo.

Al estimar el modelo, una de ellas (por ejemplo, soltero) se omite y actúa como categoría de referencia (figuras 1, 2, 3 y 4).

estado_civil	Freq.	Percent	Cum.
Casado	664	37.26	37.26
Divorciado	10	0.56	37.82
Separado	68	3.82	41.64
Soltero	822	46.13	87.77
Union Libre	205	11.50	99.27
Viudo	13	0.73	100.00
Total	1,782	100.00	

Figura 1. Categorías de referencia

Fuente: elaboración propia con el *software* Stata.

Este procedimiento se replicó para nivel\_estudio, tipo\_vivienda y proyecto\_vis, permitiendo capturar diferencias en el comportamiento de pago entre los grupos socioeconómicos sin introducir la colinealidad estructural.

nivel_estudio	Freq.	Percent	Cum.
Bachiller	690	38.72	38.72
Especialización	43	2.41	41.13
Magister	2	0.11	41.25
Primaria	14	0.79	42.03
Profesional	563	31.59	73.63
Técnico	307	17.23	90.85
Tecnólogo	163	9.15	100.00
Total	1,782	100.00	

Figura 2. Categoría nivel de estudios

Fuente: elaboración propia con el *software* Stata.

tipo_vivien da	Freq.	Percent	Cum.
Alquilada	173	9.71	9.71
Familiar	1,577	88.50	98.20
Propia	32	1.80	100.00
Total	1,782	100.00	

Figura 3. Categoría tipo de vivienda

Fuente: elaboración propia con el *software* Stata.

proyecto_vi s	Freq.	Percent	Cum.
NO	1,475	82.77	82.77
SI	307	17.23	100.00
Total	1,782	100.00	

Figura 4. Categoría Proyecto VIS

Fuente: elaboración propia con el *software* Stata.

### **Iteraciones de especificación (proceso de búsqueda)**

El proceso de modelamiento fue iterativo y guiado por evidencia, combinando la validación empírica, el criterio económico y las pruebas de especificación. No se trató de un camino lineal, sino de sucesivos ajustes sobre la base de los resultados obtenidos en cada ronda de estimación.

### **Primera etapa: modelo exploratorio (amplio)**

Se incluyó el conjunto completo de covariables con potencial explicativo, tanto continuas como categóricas. El objetivo era observar direcciones de efecto, magnitudes y estabilidad de los signos.

En esta fase se detectaron:

- Variables con baja significancia estadística y alta colinealidad (por ejemplo, rango de edad y edad, o rango salarial e ingreso).

- Algunos problemas de ajuste global (pseudo  $R^2$  bajos y criterios AIC/BIC poco favorables).

Estos hallazgos motivaron una primera depuración empírica.

### **Segunda etapa: simplificación estructural**

Se eliminaron las variables redundantes y se consolidaron categorías con pocas observaciones. También se compararon versiones alternativas del modelo: con tasas nominales vs. logarítmicas, y con o sin logaritmo del monto.

El uso de logaritmos mostró mejor comportamiento en los criterios de información y en la estabilidad de los coeficientes. Durante esta etapa se comprobó que las variables con mayor contribución al ajuste eran las relacionadas con la edad, el ingreso, el plazo y la antigüedad del crédito.

### **Tercera etapa: solidez y consistencia funcional**

Se probaron las dos formas de especificación, Logit y KNIME, manteniendo el mismo conjunto de covariables.

Los signos y las significancias fueron consistentes entre ambos, lo que reforzó la validez estructural del modelo.

La inclusión del término  $edad^2$  mejoró los indicadores de ajuste (mayor pseudo  $R^2$  de McFadden y menor BIC), confirmando la pertinencia de una relación no lineal entre edad y riesgo de impago.

Asimismo, se evaluaron interacciones entre variables (por ejemplo, ingreso y plazo), pero no se mantuvieron en la versión final por no aportar una mejora significativa en el ajuste ni claridad interpretativa.

### **Cuarta etapa: especificación final (modelo parsimonioso y estable)**

El modelo final se definió con las variables que mostraron:

- Signo estable y sentido económico coherente.
- Contribución estadísticamente significativa al 10 % y al 5 %.
- Comportamiento robusto en ambas especificaciones (Logit y KNIME).

De esta forma, el conjunto final de covariables incluyó:

- Continuas: edad,  $edad^2$ , log\_ingreso, antigüedad\_crédito\_meses, plazo\_meses.
- Categóricas (dummies): tipo de vivienda y nivel educativo (figura 5).

prob_default	Pr(default)	float	%9.0g
d_estado_civil1	estado_civil==Casado	byte	%8.0g
d_estado_civil2	estado_civil==Divorciado	byte	%8.0g
d_estado_civil3	estado_civil==Separado	byte	%8.0g
d_estado_civil4	estado_civil==Soltero	byte	%8.0g
d_estado_civil5	estado_civil==Union Libre	byte	%8.0g
d_estado_civil6	estado_civil==Viudo	byte	%8.0g
d_nivel_estudio1	nivel_estudio==Bachiller	byte	%8.0g
d_nivel_estudio2	nivel_estudio==Especialización	byte	%8.0g
d_nivel_estudio3	nivel_estudio==Magister	byte	%8.0g
d_nivel_estudio4	nivel_estudio==Primaria	byte	%8.0g
d_nivel_estudio5	nivel_estudio==Profesional	byte	%8.0g
d_nivel_estudio6	nivel_estudio==Técnico	byte	%8.0g
d_nivel_estudio7	nivel_estudio==Tecnólogo	byte	%8.0g
d_tipo_vivienda1	tipo_vivienda==Alquilada	byte	%8.0g
d_tipo_vivienda2	tipo_vivienda==Familiar	byte	%8.0g
d_tipo_vivienda3	tipo_vivienda==Propia	byte	%8.0g
d_proyecto_vis1	proyecto_vis==NO	byte	%8.0g
d_proyecto_vis2	proyecto_vis==SI	byte	%8.0g
loq_tasa		float	%9.0g

Figura 5. Conjunto final de covariables utilizadas en el modelo

Fuente: elaboración propia con el software Stata.

## Paso 2: Ejecución del modelo

### Modelos de elección discreta con variable dependiente binaria: aplicación al riesgo de default

A continuación, se describe el enfoque metodológico empleado para analizar los determinantes del riesgo de default en la cartera de crédito del Banco. Dado que el evento de interés que es el posible incumplimiento o default de un cliente, es de naturaleza binaria, es decir, ocurre o no ocurre, se utilizan modelos de elección discreta con variable dependiente binaria, específicamente los modelos Logit y KNIME, pues permiten estimar la probabilidad de ocurrencia del evento en función de un conjunto de características observables tanto del cliente como del crédito y del entorno macroeconómico. A diferencia de una regresión lineal tradicional, los modelos binarios garantizan que las probabilidades estimadas se encuentren

acotadas entre 0 y 1, lo cual es una condición fundamental cuando se analizan eventos de tipo probabilístico.

Los modelos de elección discreta parten del supuesto de que los individuos (o, en este caso, los prestatarios) toman decisiones o experimentan resultados en función de una utilidad o propensión latente que no puede observarse directamente. Esta variable latente, denotada como ( $y_i^*$ ), se define como una combinación lineal de las variables explicativas más un término de error:

Fórmula 1:

$$y_i^* = X_i' \beta + \varepsilon_i.$$

Donde  $X_i$  representa el vector de características observadas del individuo ( $i$ ),  $\beta$  es el vector de parámetros a estimar y  $\varepsilon_i$  es un término de error aleatorio.

La diferencia entre el modelo Logit y el aplicado mediante la plataforma KNIME se encuentra en la manera en que cada enfoque representa la relación entre las variables y la probabilidad de deterioro. El modelo Logit parte de una estructura paramétrica en la que el término de error  $\varepsilon_i$  se asume con una distribución logística estándar, lo que permite modelar la probabilidad del evento mediante una función sigmoide estrictamente definida por los parámetros estimados. Esta especificación obliga a que la relación entre las variables explicativas y la probabilidad sea no lineal y se encuentre en el rango entre 0 y 1, garantizando estabilidad interpretativa y consistencia estadística.

En contraste, los modelos implementados en KNIME, especialmente el Árbol de Decisión seleccionado en esta investigación, no dependen de supuestos explícitos sobre la distribución del término de error. En lugar de una formulación probabilística paramétrica, el algoritmo construye sucesivas particiones del conjunto de datos basadas en reglas de decisión optimizadas para separar de manera más efectiva a los clientes deteriorados de los no deteriorados. Cada partición se genera evaluando cuál variable y qué punto de corte maximizan la pureza de los grupos resultantes, lo que permite obtener relaciones no lineales y posibles interacciones entre variables sin imponer una forma funcional predeterminada.

Mientras el modelo Logit vincula la probabilidad del evento a una estructura estadística que depende de parámetros definidos y efectos marginales interpretables, el KNIME clasifica a los individuos según patrones detectados directamente en los datos, permitiendo identificar rutas de decisión específicas que revelan comportamientos diferenciados dentro de la cartera. Esta flexibilidad hace que el árbol pueda adaptarse mejor a escenarios en los que las interacciones entre variables no siguen patrones lineales, ampliando así la capacidad predictiva del análisis.

Los parámetros ( $\beta$ ) se estiman mediante el método de Máxima Verosimilitud (MLE, por sus siglas en inglés). Este procedimiento busca los valores de  $\beta$  que maximizan la probabilidad conjunta de observar los resultados de la muestra dada la especificación del modelo.

El proceso de estimación produce estimadores consistentes y asintóticamente eficientes bajo los supuestos estándar del modelo. Sin embargo, los coeficientes estimados no representan directamente el cambio en la probabilidad de default, ya que los modelos son no lineales. Por ello, se calculan los efectos marginales, que miden el cambio en la probabilidad asociada a una variación infinitesimal en una variable explicativa, manteniendo las demás constantes.

Los efectos marginales pueden evaluarse en el punto medio de las variables o promediarse sobre toda la muestra, en cuyo caso se denominan efectos marginales promedio (AME).

En el contexto del análisis crediticio, la variable dependiente  $y_i$  representa si un cliente ha incurrido en default (1) o no (0). Las variables explicativas  $X_i$  pueden incluir factores financieros del cliente (ingresos, nivel de endeudamiento, historial de pago), características del préstamo (monto, plazo, tipo de garantía) y variables macroeconómicas (tasa de desempleo, tasa de interés de política, crecimiento del PIB, entre otras).

El signo y la magnitud de los efectos marginales permiten identificar los determinantes más relevantes del riesgo de incumplimiento. Por ejemplo, un coeficiente negativo asociado al ingreso indicaría que mayores ingresos reducen la probabilidad de default, mientras que un coeficiente positivo del endeudamiento sugeriría que un mayor apalancamiento incrementa el riesgo.

Una vez estimados el modelo Logit y el modelo complementario implementado mediante la plataforma KNIME, se procedió a evaluar su validez y su capacidad predictiva mediante un conjunto de pruebas diseñadas para medir la calidad del ajuste y el desempeño clasificatorio. En el caso del modelo Logit, se calcularon los criterios de información (AIC y BIC) y el logaritmo de la verosimilitud, con el propósito de determinar la consistencia interna del modelo y verificar si la especificación propuesta resulta adecuada frente a posibles alternativas. Estos indicadores permiten valorar el equilibrio entre ajuste y complejidad, lo que asegura que la estimación no dependa de un exceso de parámetros y mantenga una estructura interpretable y eficiente.

Luego, el desempeño predictivo del Logit se analizó mediante métricas operativas como la tasa de clasificación y la curva ROC, cuya área bajo la curva (AUC) permite

cuantificar la capacidad del modelo para distinguir entre clientes que presentan deterioro y aquellos que se mantienen al día en sus obligaciones.

En paralelo, el modelo desarrollado en KNIME fue evaluado a través de métricas diseñadas específicamente para algoritmos de clasificación como el recall, la precisión y la matriz de confusión. Estas medidas permiten valorar la sensibilidad del modelo para identificar clientes con probabilidad de incumplimiento y su capacidad para evitar clasificaciones erróneas que impacten negativamente las decisiones de crédito. Además, se examinó el comportamiento del algoritmo en función de la estructura del árbol generado, lo que permitió observar la coherencia de las reglas de decisión y la estabilidad de los nodos relevantes. Este proceso de validación proporcionó una comparación equilibrada entre el enfoque paramétrico del Logit y el enfoque no paramétrico aplicado en KNIME, fortaleciendo la solidez de los resultados obtenidos.

Además, se estimaron los efectos marginales promedio para cada variable explicativa, con el objetivo de interpretar de manera directa la magnitud e incidencia de los factores sobre la probabilidad de incumplimiento. También se exploró la relación no lineal de la edad y las posibles interacciones con las variables de ingreso, para verificar la estabilidad de la especificación.

Estas etapas permitieron validar la coherencia interna del modelo y asegurar que los resultados fueran consistentes tanto en el enfoque Logit como en el KNIME.

### **Definición de la variable objetivo**

Se creó la variable objetivo llamada IBM = Default, que posibilita diferenciar a los clientes que cumplen con sus obligaciones de aquellos que tienen deterioro.

La definición de esta variable permitió estructurar el problema de investigación como una tarea de clasificación binaria, coherente con los modelos de riesgo crediticio empleados en la industria financiera.

### **Tratamiento del desbalance de clases**

Fue necesario llevar a cabo un proceso de balanceo debido a que la base de datos mostraba un desbalance importante, con una cantidad inferior de clientes clasificados como malos. Este ajuste asegura que los modelos sean capaces de aprender correctamente las características vinculadas con el deterioro, evitando inclinaciones hacia la clase más numerosa y perfeccionando la exactitud de la clasificación.

Para reducir el impacto del desbalance estructural de la base original y optimizar la habilidad de aprendizaje de los algoritmos de clasificación que se han puesto en

marcha en la plataforma KNIME, se llevó a cabo un proceso de balanceo de clases sobre el conjunto de entrenamiento. En consecuencia, de este proceso, la base balanceada se compone de 5346 observaciones; 5022 de ellas son créditos en situación de cumplimiento y 324 créditos en default.

Esta modificación aumenta significativamente la representación de la clase minoritaria, lo que posibilita que los modelos reconozcan patrones relacionados con el deterioro del crédito sin ser dominados por la clase mayoritaria. Aunque la distribución final no refleja la proporción real que se observa en el portafolio, su propósito es puramente analítico y está enfocado en entrenar el modelo, siguiendo los procedimientos comunes de análisis de riesgo crediticio, donde el objetivo principal es aumentar la capacidad discriminativa frente a eventos poco frecuentes.

El tratamiento del desbalance resulta fundamental en estudios de riesgo crediticio, ya que una clasificación sesgada puede subestimar la probabilidad de incumplimiento y afectar la gestión del riesgo.

### Balanceo de la muestra

La base mostraba un desequilibrio importante, a causa de que había una menor cantidad de clientes deteriorados en comparación con los que tenían un buen comportamiento. Con el fin de asegurar que los modelos aprendieran correctamente, se llevó a cabo un proceso de balanceo de la muestra para disminuir el sesgo hacia la clase predominante y optimizar la habilidad para identificar casos deteriorados.

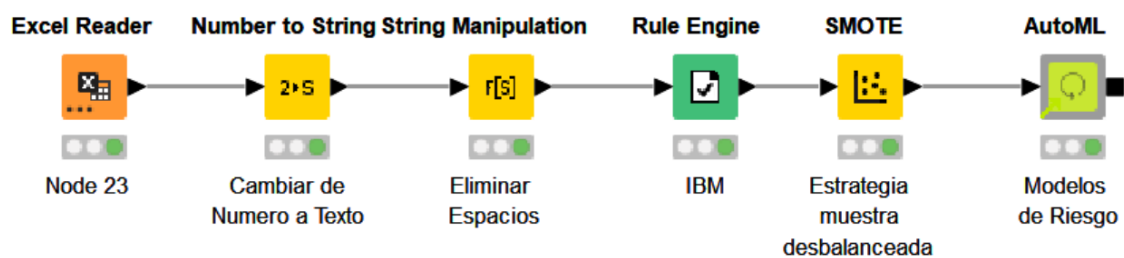


Figura 6. Flujo de preparación y balanceo de datos en KNIME

Fuente: herramienta minería de datos KNIME.

### Estrategia de modelamiento y selección de algoritmos

En este estudio se creó un modelo de clasificación empleando el nodo AutoML, que lleva a cabo varios algoritmos para definir cuál brinda una clasificación más precisa entre clientes de calidad alta y baja. La técnica supervisada de esta técnica posibilita

el entrenamiento de los modelos utilizando datos históricos de la cartera hipotecaria, adaptándose a la estructura auténtica del portafolio del Banco Unión.

A través de este enfoque se evaluaron distintos algoritmos de clasificación, priorizando aquellos compatibles con la naturaleza de los datos disponibles y con los requerimientos operativos y regulatorios de la entidad financiera.

### **Modelos de clasificación evaluados**

A partir del proceso de AutoML se seleccionaron tres modelos de clasificación para su evaluación comparativa, debido a su amplia utilización en el análisis del riesgo crediticio y a la disponibilidad de las variables incluidas en la base de datos; estos son:

- Regresión logística
- Árbol de Decisión
- Random Forest

Los tres modelos representan distintos niveles de complejidad y permiten contrastar enfoques probabilísticos y no lineales en la clasificación del incumplimiento.

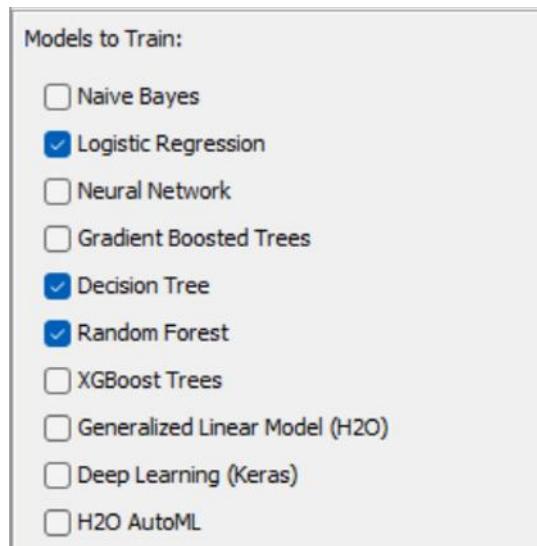


Figura 7. Selección de modelos para entrenar en la configuración del nodo AutoML

Fuente: herramienta minería de datos KNIME.

## **Criterios de evaluación del desempeño de los modelos**

Cada uno de estos modelos fue evaluado, especialmente en lo que respecta a su precisión y recall, porque son indicadores relevantes en el análisis del riesgo crediticio. La precisión señala la fiabilidad de la clasificación cuando el modelo identifica a un cliente como riesgoso, mientras que el recall permite determinar qué proporción de los clientes deteriorados fue correctamente identificada.

La elección de estas métricas responde a la necesidad de equilibrar la detección efectiva del incumplimiento con la minimización de errores de clasificación que puedan afectar a clientes con buen comportamiento crediticio.

## **Elección de métricas de evaluación**

Para valorar el rendimiento de los modelos de clasificación en el análisis del riesgo crediticio, se necesitan métricas que posibiliten diferenciar correctamente entre prestatarios cumplidos e incumplidos, más allá del porcentaje total de aciertos. En esta línea, la curva ROC (Receiver Operating Characteristic) es una herramienta esencial porque ilustra gráficamente la relación entre la tasa de verdaderos positivos y la tasa de falsos positivos a diferentes niveles de clasificación. Es particularmente relevante su uso en cuestiones de riesgo crediticio, en las cuales el propósito principal es identificar de manera temprana a aquellos clientes que tienen una mayor posibilidad de incumplir, sin aumentar la tasa de errores en la clasificación sobre los clientes solventes (Hosmer *et al.*, 2013).

El área bajo la curva ROC (AUC) resume esta relación en una medición a escala que refleja el grado de discriminación general del modelo. Una buena capacidad de separación entre clientes deteriorados y no deteriorados señala un valor de AUC que esté cerca de 1, en cambio, un valor cercano a 0,5 indica que el desempeño es similar al de una clasificación aleatoria. Debido a que es independiente del umbral de corte y bastante sólido ante distribuciones desbalanceadas de la variable objetivo, el AUC se ha establecido como un estándar en la industria financiera para validar modelos de scoring crediticio (Thomas *et al.*, 2017).

Adicionalmente, el estadístico Kolmogorov-Smirnov (KS) posibilita calcular la distancia máxima entre las distribuciones acumuladas de las probabilidades estimadas para los grupos de cumplimiento e incumplimiento. La banca emplea este indicador de manera extendida por su interpretación sencilla: un KS alto muestra que los dos grupos están más distanciados, lo cual significa una mejor clasificación del riesgo crediticio. De acuerdo con Siddiqi (2016), las cifras de KS que sobrepasan el 30 % se consideran, por lo general, aceptables en modelos de riesgo crediticio; sin embargo, su interpretación debe contextualizarse en función del tipo de cartera y las condiciones económicas del entorno.

El coeficiente de Gini, que tiene una estrecha relación con el AUC, se calcula como una transformación lineal de este último. Se utiliza frecuentemente en procesos internos de validación y regulación de modelos. Su mayor beneficio consiste en que posibilita comparar modelos y periodos temporales, lo que hace posible analizar la estabilidad y la coherencia del poder discriminatorio del sistema de clasificación. En el ámbito de las carteras hipotecarias, la utilización conjunta de AUC, KS y Gini ofrece una valoración del rendimiento, armonizando estándares estadísticos con prácticas establecidas en la administración del riesgo crediticio (Basel Committee on Banking Supervision, 2017).

### **Desbalanceo de clases**

El análisis del riesgo crediticio presenta con frecuencia el dilema de la desproporción entre clases, pues los sucesos de incumplimiento tienden a constituir una parte pequeña del total de observaciones en la cartera. En situaciones reales de bancos, la mayor parte de los clientes satisface sus responsabilidades, creando bases de datos donde la mayoría está compuesta por prestatarios que no están deteriorados y donde los casos de incumplimiento son un subconjunto más pequeño. Esta característica estructural tiene implicaciones importantes en términos de método, porque puede llevar a modelos que parecen ser precisos pero que, en la práctica, no son capaces de identificar adecuadamente a los clientes con mayor riesgo (He y García, 2009).

Si hay un desequilibrio considerable, las métricas convencionales, como el porcentaje global de clasificación correcta, tienden a exagerar el rendimiento del modelo porque un clasificador puede lograr altos niveles de precisión simplemente asignando la mayor parte de las observaciones a la clase predominante. En el campo del riesgo crediticio, esta conducta es especialmente problemática porque el costo económico de no detectar a un cliente que va a caer en default suele ser mucho más alto que el de clasificar incorrectamente a un cliente solvente como riesgoso. Por esta razón, es fundamental que la evaluación del desempeño se centre en métricas que son sensibles a la clase minoritaria, como el recall, la curva ROC y el AUC. Estas métricas posibilitan una valoración más realista de la habilidad del modelo para identificar eventos de incumplimiento (Fawcett, 2006).

El desequilibrio entre los clientes cumplidos y los deteriorados se trató en este estudio usando métodos de equilibrado de la base de datos en la etapa de modelación con KNIME, especialmente al entrenar los algoritmos de clasificación que fueron evaluados a través del nodo AutoML. Esta técnica logró que los modelos aumentaran su habilidad para captar patrones vinculados con la clase minoritaria y, de esta manera, disminuir el sesgo hacia la categorización de los clientes no deteriorados. El balanceo ayuda a que el algoritmo investigue más a fondo las características de los prestatarios con riesgo de incumplimiento, aumentando su

sensibilidad sin disminuir significativamente la precisión total (Kuhn y Johnson, 2013).

Sin embargo, es fundamental señalar que el balanceo de clases no erradica completamente las restricciones propias de la rareza del evento ni asegura de manera automática un rendimiento óptimo fuera de la muestra. Por lo tanto, se analiza con precaución el resultado obtenido, reconociendo que la fortaleza del modelo depende de la calidad del balanceo y de la estabilidad de las relaciones entre las variables explicativas y el comportamiento de pago. En este contexto, la utilización adicional del modelo Logit, calculado con la base de datos original, posibilita comparar la coherencia económica de los factores determinantes del default con los patrones encontrados por los algoritmos de clasificación, lo cual fortalece el análisis del estudio.

### **Supuestos del modelo Logit y riesgos de multicolinealidad**

El modelo Logit se basa en una serie de supuestos que, a pesar de ser menos limitantes que los de la regresión lineal clásica, son esenciales para asegurar que los resultados sean estadísticamente válidos y tengan una interpretación económica. En primer lugar, se supone que la variable dependiente es dicotómica y que las observaciones son independientes entre sí. En este estudio, esto se cumple al modelar el evento de default como una variable binaria determinada a nivel individual del crédito. Además, el modelo supone que la probabilidad de incumplimiento y las variables explicativas están relacionadas mediante una función logística. Esto significa que existe una relación no lineal entre las covariables y la probabilidad estimada, pero sí lineal entre los parámetros del modelo (Wooldridge, 2010).

La adecuada especificación funcional del modelo es otra suposición importante. En este contexto, se añadieron transformaciones logarítmicas a las variables monetarias y un término cuadrático para la edad, con el objetivo de captar potenciales no linealidades y prevenir errores de especificación que podrían alterar los coeficientes estimados. La incorporación de estas modificaciones se basa tanto en principios empíricos como en fundamentos teóricos del análisis del comportamiento financiero, entendiendo que el riesgo de incumplimiento no es afectado de forma estrictamente lineal por los efectos de ingreso, monto del crédito y ciclo de vida (Gujarati y Porter, 2009).

Un elemento clave para calcular el modelo Logit es que no exista una multicolinealidad grave entre las variables explicativas. La multicolinealidad no sesga los coeficientes, pero aumenta su varianza, disminuye la exactitud de las estimaciones y complica la detección del impacto individual de cada covariable. Cuando se introducen juntas variables muy correlacionadas, como el rango de edad y la edad misma, o el ingreso y la franja salarial, por ejemplo, este problema es

común en situaciones de riesgo crediticio. Para reducir este riesgo, se llevó a cabo la selección definitiva de variables mediante una combinación de criterios teóricos y pruebas empíricas de diagnóstico, suprimiendo las variables redundantes y dando prioridad a las que contaban con más información y coherencia económica.

Se realizó el diagnóstico de multicolinealidad a través de la determinación del Factor de Inflación de la Varianza (VIF), usando como referencia los umbrales que suelen ser aceptados en la literatura empírica. Valores altos de VIF señalaron la necesidad de eliminar o consolidar variables que proporcionaban información redundante, asegurando así la estabilidad de los coeficientes y la capacidad de interpretación del modelo. Este método permitió mantener una especificación parsimoniosa, previniendo que el ajuste del modelo se inflara de manera artificial a expensas de perder solidez estadística (Kutner *et al.*, 2005).

Además, se examinó la existencia de endogeneidad potencial, sobre todo en variables asociadas con el comportamiento histórico de pago. Siguiendo los métodos de análisis causal, se descartaron las covariables que contienen información obtenida directa o indirectamente del evento de default en sí, como la calificación interna de riesgo o la mora máxima. Su inclusión habría infringido el supuesto de exogeneidad, lo que habría generado circularidad explicativa y puesto en peligro la validez de los resultados. Así, el modelo Logit estimado se basa únicamente en datos disponibles antes del incumplimiento, lo que aumenta su fiabilidad analítica y su valor como herramienta para evaluar el riesgo crediticio.

## DESARROLLO DEL TRABAJO

### Revisión de la literatura sobre los modelos de riesgo crediticio

Dado lo anterior, en el marco de esta investigación se ha llevado a cabo una revisión de los modelos de riesgo crediticio más relevantes y más utilizados en la industria financiera. Esta revisión se ha realizado con el objetivo de obtener una comprensión profunda de las metodologías existentes y de identificar aquellas que mejor se adapten al contexto específico de la cartera hipotecaria del Banco Unión.

Se inició este proceso mediante la consulta y el análisis detallado de trabajos fundamentales que han sentado las bases teóricas para la evaluación del riesgo crediticio. Entre ellos, se destaca la influyente investigación de Altman (1968), cuyo modelo Z-Score ha sido un referente en la evaluación del riesgo financiero de las empresas. Este enfoque clásico ha proporcionado valiosas perspectivas para comprender los elementos clave en la determinación del riesgo crediticio.

Además, se ha prestado especial atención a investigaciones más contemporáneas que exploran modelos basados en técnicas de aprendizaje automático. En este sentido, los trabajos de Khandani *et al.* (2010) han sido fundamentales para entender cómo las herramientas avanzadas de análisis de datos pueden mejorar la predicción del riesgo crediticio. La aplicación de algoritmos de aprendizaje automático ha representado un avance significativo en la capacidad de anticipar y gestionar riesgos de manera más precisa y dinámica.

A continuación, se detallan los modelos de riesgo crediticio más relevantes y que pueden aplicarse a una entidad financiera como el Banco Unión.

#### 1. Modelo Z-Score de Edward Altman

- Metodología: Altman (1968) desarrolló el Z-Score en su estudio “Financial Ratios, Discriminant Analysis, and the Prediction of Corporate Bankruptcy”. Este modelo se basa en el análisis de las proporciones financieras clave de una empresa, incluyendo el capital de trabajo, el patrimonio, los activos y las ganancias. Utiliza la técnica de análisis discriminante para asignar una puntuación a las empresas y predecir la probabilidad de quiebra.
- Resultados: el modelo Z-Score ha demostrado ser eficaz para predecir la quiebra empresarial, al proporcionar una evaluación cuantitativa del riesgo de insolvencia. Ha sido utilizado en la evaluación del riesgo crediticio de empresas.

## **2. Modelos econométricos de John Campbell y Glenn Loury**

- Metodología: Campbell y Loury (1991) utilizaron modelos econométricos en su investigación para evaluar el riesgo crediticio de prestatarios. Analizaron variables como ingresos, historial crediticio y otras características personales para predecir la probabilidad de incumplimiento crediticio.
- Resultados: estos modelos econométricos han sido útiles para predecir el riesgo crediticio en carteras de préstamos personales, brindando una comprensión más profunda de las relaciones entre las variables personales y la probabilidad de incumplimiento.

## **3. Modelo Logit de Lucas y Klaassen**

- Metodología: Lucas y Klaassen (2006) aplicaron el modelo Logit en su estudio para evaluar el riesgo crediticio en las carteras hipotecarias. Utilizaron variables como el historial crediticio, ingresos y otros factores para predecir el incumplimiento crediticio. El modelo Logit es una técnica de regresión logística que evalúa la relación entre un conjunto de variables independientes y una variable dependiente binaria (en este caso, incumplimiento o no incumplimiento).
- Resultados: el modelo Logit ha demostrado ser útil en la predicción del incumplimiento de préstamos hipotecarios, permitiendo una evaluación cuantitativa de la probabilidad de incumplimiento.

## **4. Modelos de clasificación aplicados mediante minería de datos: enfoque KNIME**

- Metodología: Berthold *et al.* (2009) desarrollaron el KNIME, que fue creado como una plataforma modular para la minería de datos y facilita la creación de flujos de trabajo para labores de clasificación por medio de la integración visual de nodos. Los investigadores que utilizan esta metodología organizan el proceso en etapas sucesivas como la adición y limpieza de datos, la selección de atributos importantes, el entrenamiento de clasificadores como máquinas de soporte vectorial, los Árboles de Decisión, las redes neuronales o regresión logística y la validación a través de técnicas como particiones hold-out o validación cruzada. La plataforma posibilita la experimentación comparativa al habilitar el ajuste de variables, la implementación de transformaciones anteriores y la evaluación del desempeño bajo distintas configuraciones, sin requerir programación directa.
- Resultados: los modelos que se utilizan en KNIME tienden a tener un rendimiento competitivo cuando se trata de pronosticar eventos dicotómicos, pues la herramienta posibilita mejorar la calidad del conjunto de datos y

escoger algoritmos que tengan un equilibrio adecuado entre la capacidad de generalización y precisión. El empleo de KNIME en investigaciones aplicadas ha posibilitado la detección de patrones coherentes y el aumento del porcentaje de aciertos con respecto a métodos manuales o menos sistemáticos, lo cual respalda su efectividad para clasificar personas y objetos basándose en criterios establecidos anteriormente dentro de un ambiente reproducible y claro.

### **5. Análisis de regresión múltiple de Richard Green y Patric Hendershott**

- Metodología: Green y Hendershott (1996) realizaron un análisis de regresión múltiple para evaluar la influencia de los factores macroeconómicos y las variables del mercado inmobiliario en el riesgo crediticio. Su enfoque se centró en la relación entre la evolución del mercado de bienes raíces y el incumplimiento de hipotecas.
- Resultados: su investigación reveló la importancia de considerar factores macroeconómicos y las tendencias del mercado inmobiliario al evaluar el riesgo crediticio en las carteras hipotecarias.

### **6. Regresión logística de Andrea Heuson y Gary Smith**

- Metodología: Heuson y Smith (1995) utilizaron el modelo de regresión logística para evaluar la probabilidad de incumplimiento de préstamos hipotecarios, considerando variables como el historial crediticio, el tamaño del préstamo y las tasas de interés.
- Resultados: su investigación proporcionó información valiosa sobre la probabilidad de incumplimiento de préstamos hipotecarios y destacó la influencia de diversas variables en esta probabilidad.

### **7. Modelo Score de William R. Scott**

- Metodología: el modelo Score de Scott (1965) se basa en la teoría de puntuación y utiliza una variedad de variables como historial crediticio, ingresos y otros factores, para asignarles una puntuación a los prestatarios. Cuanto más alta sea la puntuación menor será el riesgo crediticio. Este modelo se ha utilizado en la evaluación de préstamos hipotecarios y otros tipos de crédito.
- Resultados: ha demostrado ser efectivo en la clasificación de prestatarios según su riesgo crediticio, lo que ayuda a las instituciones financieras a tomar decisiones informadas.

## **8. Modelo de regresión de supervivencia**

- Metodología: este modelo, desarrollado por Kleinbaum y Klein (2012) se utiliza para evaluar el riesgo crediticio a lo largo del tiempo, considerando que los prestatarios pueden incumplir en momentos diferentes. Utiliza datos de supervivencia y tiene en cuenta la duración hasta que se produce el incumplimiento. Es útil en situaciones en las que se necesita comprender cómo cambia el riesgo crediticio con el tiempo.
- Resultados: ha proporcionado información valiosa sobre cómo evoluciona el riesgo crediticio en las carteras de préstamos a lo largo del tiempo, lo que permite una gestión más efectiva del riesgo.

## **9. Modelo de aprendizaje automático (Machine Learning)**

- Metodología: los modelos de aprendizaje automático propuestos por Hastie, Tibshirani y Friedman (2009), como las redes neuronales, los bosques aleatorios y las máquinas de soporte vectorial, se utilizan para evaluar el riesgo crediticio al analizar una gran cantidad de datos, incluyendo información de crédito, historial crediticio, ingresos y más. Estos modelos pueden detectar patrones complejos y no lineales en los datos.
- Resultados: ha demostrado ser efectivo en la predicción del riesgo crediticio, y se utiliza en la industria financiera para mejorar la precisión en la evaluación del riesgo.

En la etapa de desarrollo de este proyecto fue necesario sumergirse en la tarea de evaluar cada modelo de riesgo crediticio identificado. La meta central era determinar la aplicabilidad y adecuación de estos modelos al contexto singular de la cartera hipotecaria del Banco Unión. Para llevar a cabo dicha tarea, se dirigió la atención hacia las particularidades del mercado colombiano y las características únicas que definen la cartera hipotecaria de esta institución financiera.

### **Selección de los modelos aptos para el Banco Unión**

La revisión de la literatura académica ha sido esencial, especialmente al considerar la adaptación de modelos al contexto específico del mercado hipotecario. Los autores clave que influyeron en la elección y el desarrollo de los modelos son Green y Hendershott (1996), cuyo trabajo se centra en el análisis de regresión múltiple.

## **Modelo de Green y Hendershott - Análisis de regresión múltiple**

El enfoque de Green y Hendershott en el análisis de regresión múltiple ha sido importante para comprender las relaciones entre las diversas variables que afectan el riesgo crediticio en el ámbito hipotecario. Su investigación destaca la importancia de considerar las particularidades del entorno tanto interno como externo al implementar modelos de riesgo.

El modelo de regresión múltiple permite examinar la contribución relativa de diferentes variables explicativas, lo que resulta fundamental para identificar los factores clave que inciden en el riesgo crediticio. La investigación de Green y Hendershott (1996) ha proporcionado una base sólida para la elección de un modelo que aborde de manera integral y explicativa las complejidades asociadas con el riesgo crediticio en la cartera hipotecaria del Banco Unión.

Por lo tanto, se procede a mostrar el desarrollo de la aplicación de estos modelos a la situación actual de la entidad financiera en estudio. Para lograrlo, se realizaron los siguientes pasos:

### **Identificación de las variables relevantes**

La identificación de las variables relevantes para el modelo de riesgo crediticio en la cartera hipotecaria del Banco Unión se llevó a cabo con un enfoque preciso y detallado, considerando los aspectos internos y externos del contexto colombiano. En la siguiente sección se presenta un desarrollo más específico de este proceso:

- Análisis de la información macroeconómica

De forma paralela, se llevó a cabo un análisis de la información macroeconómica para comprender el contexto económico en el que opera el Banco Unión. Entre las variables macroeconómicas consideradas, el PIB emergió como un factor de gran relevancia.

El PIB es un indicador importante que refleja la salud general de la economía colombiana. Este indicador, que representa el valor total de todos los bienes y servicios producidos en el país, se convirtió en un elemento central para entender la estabilidad económica y su impacto en el riesgo crediticio. Dado que el PIB influye directamente en factores como el empleo y los ingresos, su inclusión en el modelo permitirá capturar las variaciones macroeconómicas que podrían afectar la capacidad de los prestatarios para cumplir con sus obligaciones crediticias. El análisis detallado del PIB proporcionará una visión más precisa de cómo los cambios en la actividad económica pueden traducirse en riesgos crediticios específicos para la cartera hipotecaria del Banco Unión.

En la figura 8 se observa la variación del PIB de forma trimestral desde el año 2015 al tercer trimestre del año 2025.

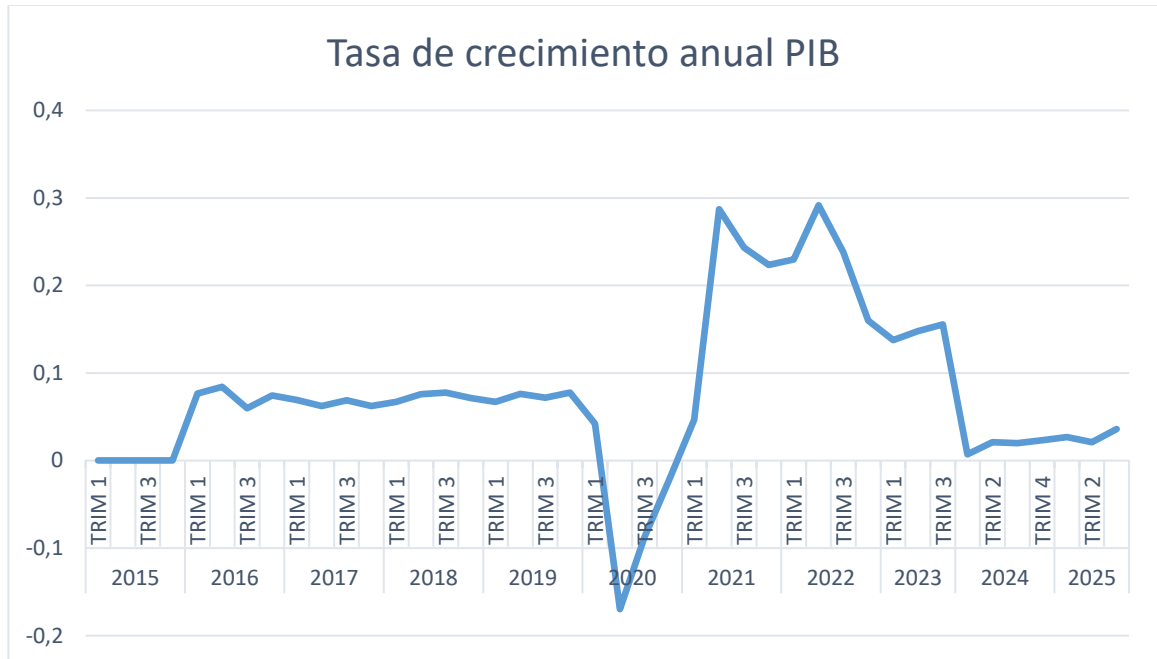


Figura 8. Variación trimestral del PIB años 2015-tercer trimestre de 2025 en Colombia

Fuente: elaboración propia con base en el DANE (2025).

Obsérvese que la variación trimestral del PIB en Colombia muestra una serie de patrones notables durante el período 2015-Tercer trimestre de 2025. Es importante destacar que los datos del año 2020 reflejan el impacto significativo que tuvo la pandemia de covid-19 en la economía. Los otros movimientos se detallan así:

- Crecimiento sostenido: durante los primeros tres años del período analizado, Colombia experimentó un crecimiento económico constante, con tasas trimestrales que oscilaron entre el 6,0 % y el 8,4 %. Estos números sugieren una economía en expansión, respaldada por factores como la estabilidad política y las condiciones favorables para la inversión.
- Impacto de la pandemia: el año 2020 presenta una marcada contracción económica en los dos primeros trimestres, con tasas de -17,0 % y -8,7 %, respectivamente. Este declive abrupto coincide con las restricciones impuestas para contener la propagación del covid-19. La economía se

contrajo de manera significativa durante este período, reflejando el impacto global de la pandemia en las actividades económicas.

- Recuperación pospandemia (2020-2021): a medida que se flexibilizaron las restricciones y se implementaron medidas de estímulo económico, se observa una recuperación en el tercer y cuarto trimestre del 2020, con tasas de -2,1% y 4,6 %, respectivamente. Este patrón de recuperación se consolida en el 2021 con tasas que van desde el 22,3 % en el cuarto trimestre del 2020 hasta el 24,3 % en el tercer trimestre del 2021. Estos indicadores sugieren una resiliencia económica y una capacidad de adaptación positiva en la pospandemia.
- Estabilización y moderación del crecimiento (2022-2023): a partir del 2022 se observa una estabilización y moderación en las tasas de crecimiento, indicando una fase de consolidación económica. Las tasas trimestrales oscilan entre el 13,8 % y el 29,2 %, mostrando una economía que busca mantener un crecimiento sólido, pero más controlado.
- Año 2024-2025: para el año 2024 en adelante, se muestra una desaceleración respecto a los años anteriores, pasando de tasas cercanas al 15 % en 2022-2023 alrededor del 2 % anual en el 2024, lo que refleja un crecimiento moderado, con estabilidad tras la recuperación de la pospandemia.

En general, el análisis revela la influencia significativa de eventos externos, como la pandemia, en la economía colombiana. La capacidad de recuperación y adaptación a las circunstancias desafiantes se destaca en la segunda mitad del 2020 y a lo largo del 2021. La estabilización observada en los años 2022 y 2024 sugiere un retorno a las tasas de crecimiento más sostenibles después de las fluctuaciones causadas por la crisis sanitaria global.

La variación trimestral del PIB en estos períodos incide directamente en el riesgo crediticio de las entidades financieras. Durante períodos de crecimiento económico (2016-2018) se observa una disminución del riesgo crediticio. Por el contrario, en el 2020 la contracción económica generó un aumento significativo en este riesgo.

La gestión efectiva del riesgo durante estos períodos desafiantes implica políticas internas sólidas y flexibilidad para enfrentar las condiciones cambiantes. En síntesis, la dinámica económica impacta directamente en el riesgo crediticio, subrayando la necesidad de estrategias ágiles y adaptativas por parte del Banco Unión.

- Inflación - IPC

En el contexto de las complejas políticas económicas que guían el desarrollo de un país, la gestión efectiva de las variables macroeconómicas, con especial énfasis en

la política monetaria, emerge como un elemento fundamental. Al centrarse en el control de la inflación mediante herramientas como el encaje bancario y la intervención en el mercado cambiario, la política monetaria no solo busca objetivos económicos, sino que también incide directamente en la sociedad al reducir la incertidumbre y propiciar mejores condiciones para la inversión.

Mantener una inflación baja se traduce en beneficios tangibles para la sociedad, generando confianza en los agentes económicos, lo que resulta esencial para tomar decisiones a largo plazo como la creación de empresas, la inversión en educación y la adquisición de viviendas.

Este contexto resalta la importancia de la inflación controlada como un indicador de confianza y un mecanismo defensivo para sectores vulnerables, al proporcionar flexibilidad frente a cambios en los precios y facilitar una adaptación ágil a las condiciones económicas.

La figura 9 presenta la variación año corrido del IPC (sin alimentos) en Colombia en los últimos años, ilustrando la evolución de esta variable clave. El análisis visual proporciona una perspectiva integral de los precios en el país, sirviendo como referencia para evaluar la efectividad de las políticas implementadas y su impacto en la toma de decisiones económicas a nivel individual y empresarial.

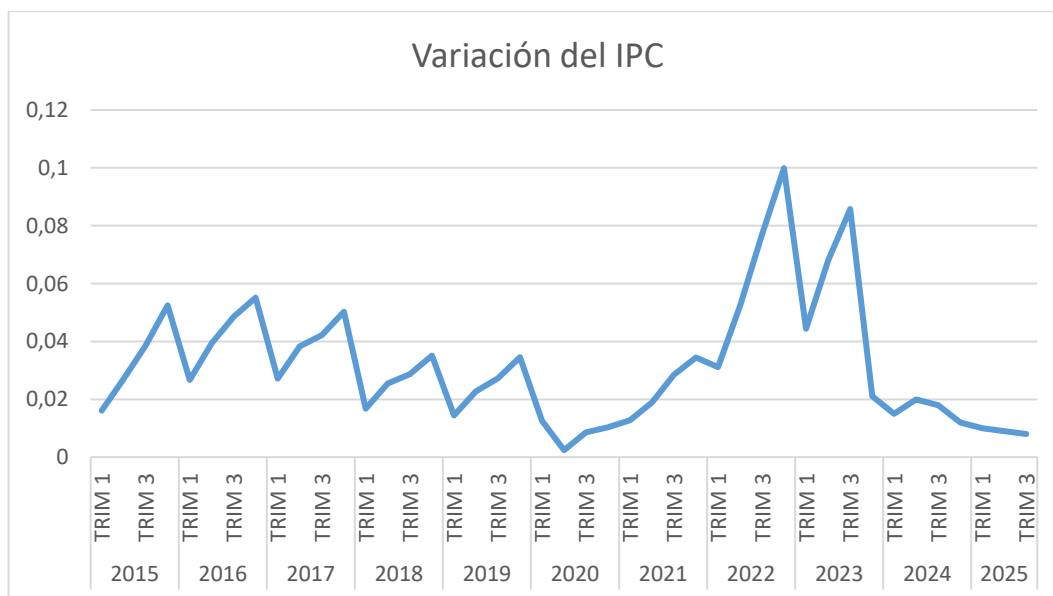


Figura 9. IPC Colombia 2015-2025

Fuente: elaboración propia con base en el DANE (2025).

Con respecto al análisis detallado de la variación año corrido del IPC en Colombia, la figura 9 revela patrones significativos. La variación anual del IPC en Colombia ha tenido una tendencia con altibajos desde el 2015 hasta el último trimestre del 2016, cuando alcanzó un valor máximo de 6,48%. Esto se debió principalmente al fenómeno de El Niño, que afectó la producción agrícola y el suministro de energía, y al aumento de los precios internacionales del petróleo.

A partir del año 2017, la variación anual del IPC empezó a disminuir, debido a la menor presión de los precios de los alimentos, la reducción de las tasas de interés y la apreciación del peso frente al dólar. En el 2020, la variación anual del IPC fue de 1,75 %, la más baja desde 1955, como consecuencia de la pandemia de covid-19, que generó una caída de la demanda y una menor actividad económica.

En el 2021, la variación anual del IPC se situó en 4,58 %, superando el rango meta del Banco de la República, que era del 3 % +/- 1 %. Esto se explicó por el aumento de los precios de los alimentos, el transporte, la vivienda y los servicios públicos, asociado a la recuperación de la demanda interna, los efectos del paro nacional, los choques climáticos y la depreciación del peso. En el año 2022, la variación anual del IPC se disparó a 12,22 %, la más alta desde 1992, debido a la crisis energética, que provocó un racionamiento de la electricidad y un alza de las tarifas, y a la escasez de algunos productos básicos, como el gas, el arroz y el azúcar.

En el 2023, la variación anual del IPC se moderó a 10,48 %, pero aún se mantuvo en niveles elevados, reflejando las dificultades para normalizar el abastecimiento y la persistencia de las presiones inflacionarias. Para el año 2024, y hasta el tercer trimestre del año 2025, se ha visto una disminución del IPC, volviendo a niveles razonables y usuales en Colombia.

De igual forma, este comportamiento del IPC muestra una serie de tendencias que tienen implicaciones significativas para la gestión del riesgo crediticio, especialmente en entidades financieras como el Banco Unión. El aumento sostenido que tuvo después de la pandemia, pudo tener impactos directos en la gestión del riesgo crediticio. En entidades financieras como el Banco Unión, el riesgo crediticio está intrínsecamente vinculado a las condiciones macroeconómicas, y la inflación es un indicador clave. Tasas de inflación más altas pueden generar presiones sobre los costos y los márgenes financieros, afectando la capacidad de los prestatarios para cumplir con sus obligaciones.

Además, la inflación puede influir en la capacidad adquisitiva de los clientes, lo que impacta directamente su capacidad para pagar préstamos. En un entorno de inflación creciente, los costos de vida aumentan, lo que podría llevar a un deterioro en la calidad crediticia de los prestatarios y aumentar el riesgo de morosidad.

En el caso específico del Banco Unión, es esencial considerar cómo estas fluctuaciones en las tasas de inflación afectan a los sectores económicos a los que la entidad está expuesta. Por ejemplo, si el Banco Unión tiene una cartera significativa en sectores sensibles a la inflación, como bienes raíces o consumo, el aumento de la inflación podría tener un impacto negativo en la capacidad de pago de los prestatarios de estos sectores.

La entidad debe ajustar sus estrategias de gestión del riesgo crediticio en respuesta a este escenario. Esto podría incluir la revisión de las tasas de interés, el establecimiento de límites de exposición en sectores específicos y una evaluación más rigurosa de la capacidad de pago de los prestatarios en un entorno inflacionario. Posteriormente, se evidencia el impacto de todo ello en el comportamiento de la cartera de la entidad.

- Tasa de interés efectiva (TIE)

La tasa de interés efectiva (TIE), en el contexto económico colombiano, emerge como una variable de gran relevancia, cuyo comportamiento y fluctuaciones ejercen una influencia significativa en diversos aspectos del sistema financiero, la economía en general y, de manera destacada, en el sector inmobiliario. La comprensión detallada de esta variable es esencial para anticipar sus efectos y diseñar estrategias que mitiguen posibles impactos negativos.

La TIE se calcula utilizando la fórmula de equivalencia financiera, que tiene en cuenta los diversos componentes asociados al préstamo. La fórmula general es la siguiente:

Fórmula 2:

$$1TIE = (1 + ni)^n - 1$$

Donde:

- TIE: es la Tasa de Interés Efectiva.
- i: es la tasa nominal de interés anual.
- n: es el número de períodos de capitalización por año.

Este análisis no solo se limita a la observación de las variaciones numéricas, sino que se adentra en la comprensión de los eventos económicos y las políticas que han incidido en su evolución (Levy, 2012).

La influencia de la tasa de interés efectiva se manifiesta en múltiples dimensiones. Desde la perspectiva financiera, impacta directamente en los costos de endeudamiento para individuos y empresas, condicionando las decisiones de inversión, consumo y financiamiento.

En el contexto específico del sector inmobiliario, las TIE desempeñan un papel importante al determinar la accesibilidad a créditos hipotecarios y, por ende, influyen en la demanda de viviendas y en la estabilidad del mercado. Además, las variaciones en las tasas de interés pueden impactar la actividad constructiva, los precios de la vivienda y la dinámica del mercado hipotecario (Montoriol, 2023). A continuación, se presenta el comportamiento histórico de la tasa de interés efectiva de Colombia desde el 2010 hasta el 2023 (figura 10):

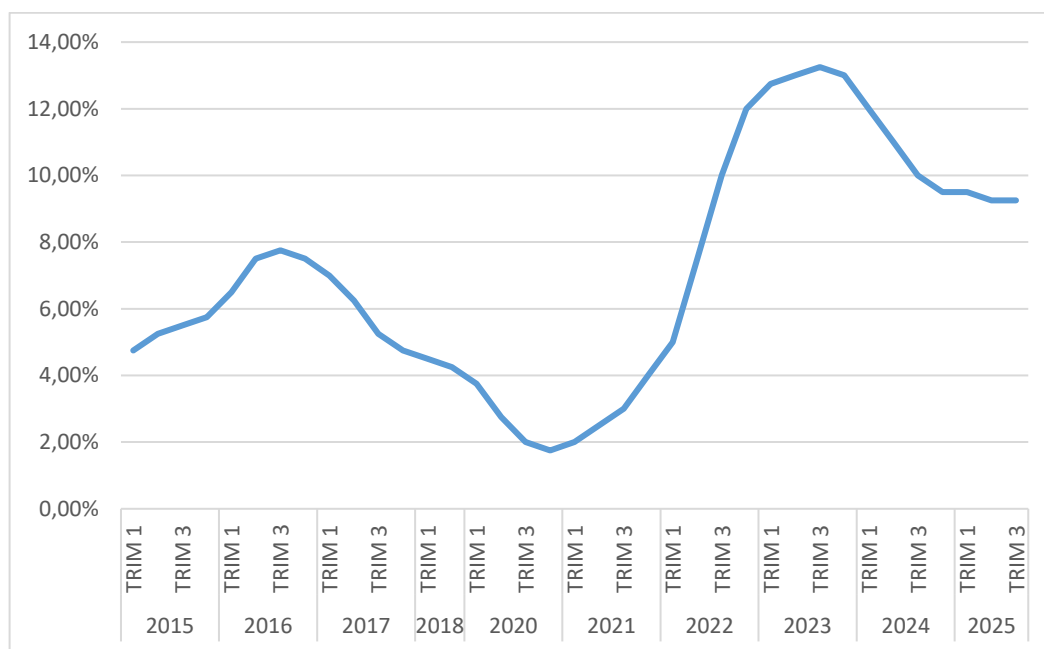


Figura 10. Tasa efectiva anual. Comportamiento histórico (2015-2025)

Fuente: elaboración propia con base en el DANE (2025).

La Tasa de Interés Efectiva (TIE) en Colombia ha mostrado variaciones significativas durante el periodo examinado. En febrero de 2016 alcanzó un 6,00 %, generando preocupación sobre la inflación, y aumentó hasta un máximo del 7,75 % en agosto de 2016. Posteriormente, experimentó una disminución gradual hasta situarse en el 7,50 % en diciembre de 2016. De 2017 a 2019 la tasa mantuvo una tendencia a la baja, alcanzando un mínimo del 4,75 % en noviembre de 2017, seguido de estabilidad hasta marzo de 2020, cuando registró un 3,75 % debido a las medidas de flexibilización monetaria ante la pandemia de covid-19. A partir de mayo de 2020 se implementaron reducciones graduales, llegando al mínimo del 1,75 % en septiembre de 2020. Sin embargo, desde octubre de 2021 hasta marzo de 2023 se observó un cambio en la tendencia con aumentos sucesivos, alcanzando

un 13,25 %. Estas variaciones reflejan la respuesta de las autoridades económicas a las condiciones macroeconómicas y eventos como la pandemia, con implicaciones directas en el sector inmobiliario, ya que las tasas de interés afectan los costos de financiamiento para la adquisición de viviendas, influyendo en la demanda, los precios y la actividad en el mercado inmobiliario.

- TIE hipotecaria

En el contexto específico de los préstamos hipotecarios, esta fórmula se ajusta para incluir los costos adicionales asociados. La expresión más detallada sería la que se plantea en la siguiente fórmula:

Fórmula 3:

$$TIE = (1 + ni)^n - 1 + \text{costos adicionales}$$

Los costos adicionales pueden incluir comisiones, seguros y otros cargos que forman parte de los costos totales del préstamo.

El proceso de cálculo implica obtener la tasa nominal de interés anual del préstamo, identificar el número de períodos de capitalización por año y sumar los costos adicionales. Posteriormente, estos valores se incorporan en la fórmula de la TIE para obtener el resultado final.

Es importante destacar que, dado que los costos adicionales pueden variar entre diferentes préstamos y entidades financieras, la TIE proporciona una medida más precisa de los costos totales para el prestatario.

### **Exploración de variables propias del mercado hipotecario colombiano**

En la exploración de variables propias del mercado hipotecario colombiano se aborda una comprensión detallada del entorno inmobiliario. La obtención de datos confiables y actuales se llevó a cabo mediante la revisión de informes especializados en el mercado inmobiliario colombiano, proporcionando una base sólida para la identificación de tendencias y comportamientos del sector.

La identificación de tendencias y comportamientos en dicho sector revela una serie de dinámicas significativas que influyen directamente en el riesgo crediticio en el ámbito hipotecario. Con base en datos y cifras actuales, se destacan las siguientes tendencias:

Crecimiento sostenido del sector inmobiliario: en los últimos años, Colombia ha experimentado un crecimiento sostenido en el sector inmobiliario. La demanda de viviendas ha sido impulsada por factores como el crecimiento demográfico, la estabilidad económica y programas gubernamentales de subsidios a la vivienda.

Este crecimiento constante es fundamental para evaluar la estabilidad y la solidez del mercado inmobiliario.

Variaciones en los precios de las viviendas: los precios de las viviendas en Colombia han experimentado variaciones notables en diferentes regiones del país. Mientras que algunas áreas metropolitanas han registrado aumentos significativos, otras han mantenido una estabilidad relativa. Este fenómeno puede atribuirse a la oferta y demanda específicas de cada región, así como a factores económicos y sociales particulares.

Influencia de programas gubernamentales: la implementación de programas gubernamentales, como subsidios a la vivienda y políticas de financiamiento accesibles, ha tenido un impacto directo en el comportamiento del sector inmobiliario. Estos programas buscan fomentar la adquisición de viviendas y mejorar el acceso al crédito hipotecario, lo que puede influir en la dinámica del riesgo crediticio.

Adaptación a nuevas modalidades de trabajo: la pandemia de covid-19 ha impulsado cambios en las preferencias y necesidades de vivienda. La adopción del trabajo remoto ha llevado a una mayor flexibilidad en la elección de ubicación para residir. Esto ha generado una reevaluación de la demanda de viviendas, con un interés creciente en áreas suburbanas y ciudades intermedias.

Impacto en las tasas de interés: las fluctuaciones en las tasas de interés, particularmente las tasas hipotecarias, desempeñan un papel importante en el comportamiento del sector inmobiliario. Tasas más bajas pueden estimular la demanda al hacer el financiamiento más accesible, mientras que tasas más altas pueden tener el efecto contrario.

Desarrollo de proyectos inmobiliarios: el desarrollo de proyectos inmobiliarios, tanto residenciales como comerciales, es un indicador clave de la salud del sector. La planificación y ejecución exitosa de nuevos proyectos pueden ser un signo de confianza y dinamismo en el mercado.

De igual manera, la oferta y demanda de viviendas representan pilares fundamentales en la dinámica del mercado inmobiliario. La interrelación entre estos dos factores no solo incide en la estructura del sector, sino que también proporciona *insights* para entender las tendencias y proyecciones a largo plazo. En este contexto, se explorará la oferta y demanda de viviendas en Colombia, abordando su comportamiento general con cifras y datos estadísticos, y llevando a cabo un análisis específico del Índice de Precios a la Vivienda Nueva (IPVN).

El IPVN se erige como un indicador clave para evaluar las variaciones en los precios de la vivienda. Este índice, que refleja las fluctuaciones en el costo de la vivienda

nueva en el mercado, proporciona información valiosa sobre la estabilidad y la dinámica del sector inmobiliario.

El comportamiento del IPVN es un reflejo directo de las condiciones del mercado. Incrementos sostenidos pueden indicar una demanda robusta y limitaciones en la oferta, mientras que disminuciones podrían sugerir un exceso de oferta o cambios en las preferencias del consumidor. Por lo tanto, el análisis detallado de este indicador permite discernir las fuerzas que impulsan el mercado inmobiliario y anticipar posibles ajustes en la estrategia de inversión.

En la figura 11 se detalla el comportamiento del IPVN desde el año 2015 hasta el tercer trimestre del año 2025.

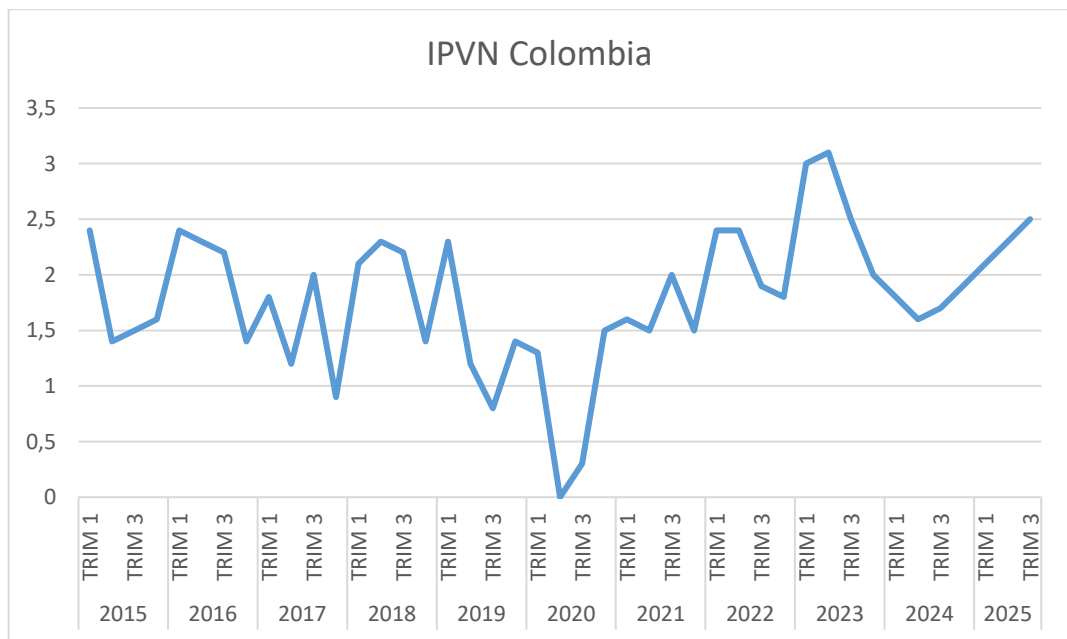


Figura 11. Comportamiento del IPVN (2015-2025)

Fuente: elaboración propia con base en el DANE (2025).

El análisis del IPVN proporciona una visión integral del comportamiento del mercado inmobiliario en Colombia. En los primeros años se evidenció un crecimiento sostenido, seguido de fluctuaciones que sugieren sensibilidad a factores económicos y políticos. El año 2020, marcado por la pandemia, influyó en una contracción temporal del IPVN, reflejando la incertidumbre económica.

A partir del 2021 se observa una recuperación gradual, indicando la resiliencia del sector. El año 2022 presenta cierta volatilidad, pero en general el mercado mantiene

un crecimiento moderado. En el primer semestre de 2023 se destaca un crecimiento continuo del IPVN, señalando la robustez del mercado inmobiliario. Posteriormente, en el año 2024 hubo un declive y en el año 2025 vuelve a recuperarse de forma gradual.

Este comportamiento del IPVN tiene implicaciones directas en el riesgo de cartera. Variaciones significativas pueden afectar la capacidad de los prestatarios para cumplir con sus obligaciones hipotecarias, aumentando el riesgo de cartera del Banco Unión. La pandemia, por ejemplo, ilustra cómo eventos externos pueden generar cambios bruscos en el IPVN, afectando la estabilidad del mercado y el riesgo crediticio.

El análisis del IPVN proporciona una herramienta clave para evaluar el riesgo crediticio en el sector hipotecario. La capacidad de anticipar tendencias y entender la relación entre el IPVN y la cartera vencida le permite al Banco Unión tomar decisiones informadas para mitigar riesgos y mantener la solidez de su cartera crediticia.

- Nivel de riesgo de crediticio

El nivel de riesgo crediticio se refiere a la evaluación y medición de la probabilidad de que un prestatario incumpla con sus obligaciones de pago. Es una medida utilizada por las instituciones financieras y prestamistas para evaluar la solvencia y capacidad crediticia de los individuos, las empresas u otras entidades que solicitan préstamos. La evaluación de este riesgo implica analizar diversos factores como el historial crediticio, los ingresos, las deudas existentes y otros indicadores financieros relevantes.

En términos prácticos, el riesgo crediticio se calcula mediante modelos y metodologías que consideran estos factores para asignar una calificación o clasificación al prestatario. Estas calificaciones suelen expresarse en términos de categorías de riesgo, que van desde bajo riesgo hasta alto riesgo. Los modelos utilizados pueden incluir análisis de puntajes crediticios, modelos estadísticos y otros enfoques cuantitativos.

La evaluación del riesgo crediticio es esencial para que las entidades financieras tomen decisiones informadas sobre la aprobación de préstamos, establezcan tasas de interés adecuadas y gestionen de manera efectiva sus carteras crediticias. Además, ayuda a los prestatarios al proporcionar una comprensión clara de las condiciones y los términos bajo los cuales se les otorgará el crédito.

## **Variables de la cartera de crédito hipotecario del Banco Unión**

En el análisis del comportamiento de la cartera de crédito hipotecario del Banco Unión es necesario examinar, de manera integral, el desempeño a lo largo del tiempo. Este proceso implica la revisión detallada de cifras y datos estadísticos que ofrecen una visión panorámica de la dinámica de esta cartera. El presente estudio se adentrará en la interpretación de dichas cifras, destacando patrones, tendencias y puntos de inflexión relevantes que puedan influir en la toma de decisiones estratégicas. Sin embargo, lo más importante es que aquí se definen las variables de la cartera de crédito hipotecario del Banco Unión que entraran en el modelo que deseamos procesar. Estas son:

**Cartera vencida crédito hipotecario Banco Unión:** refleja el saldo total, es decir, toda la cartera vencida de la entidad, por trimestre correspondiente a la deuda o saldo pendiente del Banco Unión. Aquí se incluyen los intereses por mora y el castigo de cartera. Esta medida se calcula considerando los días de mora, específicamente en intervalos de 30, 60, 180 y 360 días.

**Tasa efectiva hipotecaria del Banco Unión:** se calculó tomando el promedio del interés efectivo hipotecario aplicado a los créditos con cartera vencida.

En la figura 12 se presentan los datos de la cartera de crédito hipotecario del Banco Unión y la tasa efectiva hipotecaria promedio trimestral desde el 2015 hasta octubre del 2025.

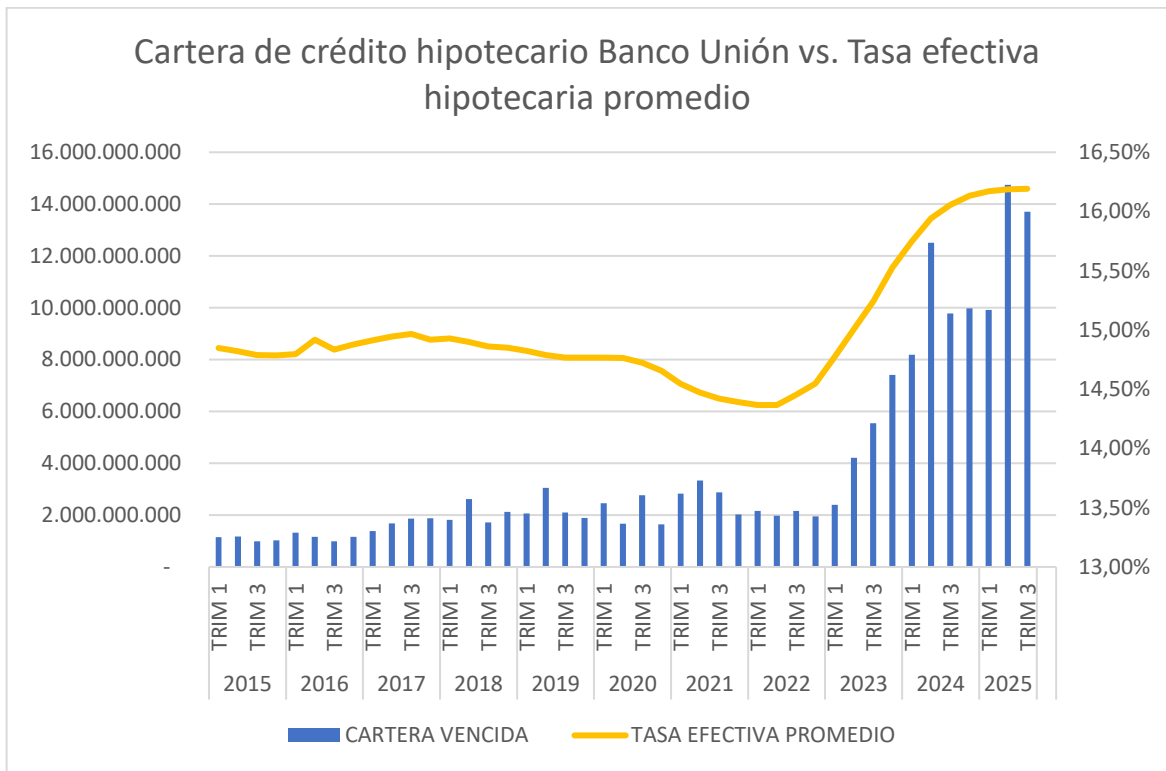


Figura 12. Comportamiento de la cartera de crédito hipotecario del Banco Unión (trimestral vs. tasa efectiva promedio)

Fuente: elaboración propia.

En términos trimestrales, se observa que la cartera vencida y la tasa efectiva promedio han experimentado variaciones importantes. Por ejemplo, en el primer trimestre del 2015 la cartera vencida fue de \$ 1 147 413 218, con una tasa efectiva del 14,8476 %, mientras que en el tercer trimestre del 2025 la cartera vencida aumentó notablemente a \$ 13 702 622 013, con una tasa efectiva del 16,1918 %.

Índice de Cartera Vencida (ICVH): este índice, según la clasificación establecida por la Superintendencia Financiera de Colombia, es la categorización de la cartera vencida en el período de uno a cuatro meses y se limita al valor de las cuotas que se han vencido. Sin embargo, a partir de la quinta cuota vencida se contempla el saldo total adeudado. La representación de este índice se expresa como un porcentaje, y su cálculo implica considerar tanto las cuotas vencidas como el saldo global pendiente.

En la tabla 2 se observa el cálculo del ICVH de forma anual. Este enfoque proporciona una perspectiva más detallada de la situación de la cartera vencida, permitiendo una evaluación general de los riesgos asociados.

Tabla 2. Análisis anual de la cartera de crédito hipotecario vencida del Banco Unión

Año	Tasa efectiva prom. (%)	Saldo SBI cuadrado (\$)	Cartera vencida (\$)	ICVH (%)
31/12/2015	14,79	34 796 492 195	1 027 192 863	2,95
31/12/2016	14,82	42 133 190 050	1 166 087 768	2,77
31/12/2017	14,92	51 207 701 557	1 876 783 850	3,67
31/12/2018	14,85	63 806 369 782	2 121 726 052	3,33
31/12/2019	14,77	78 675 251 159	1 891 065 263	2,40
31/12/2020	14,66	87 274 251 709	1 647 744 492	1,89
31/12/2021	14,39	102 266 285 676	2 029 765 142	1,98
31/12/2022	14,55	115 615 725 592	1 956 622 811	1,69
31/12/2023	15,36	138 440 449 342	6 485 040 217	4,68
31/12/2024	16,13	184 711 802 020	9 975 994 589	5,40
31/10/2025	16,17	211 934 681 915	12 862 039 083	6,34

Fuente: elaboración propia.

A continuación, en la figura 13 se observa el comportamiento de las variables “tasa efectiva” e “ICVH” y su impacto en el comportamiento de la cartera de la entidad Banco Unión.

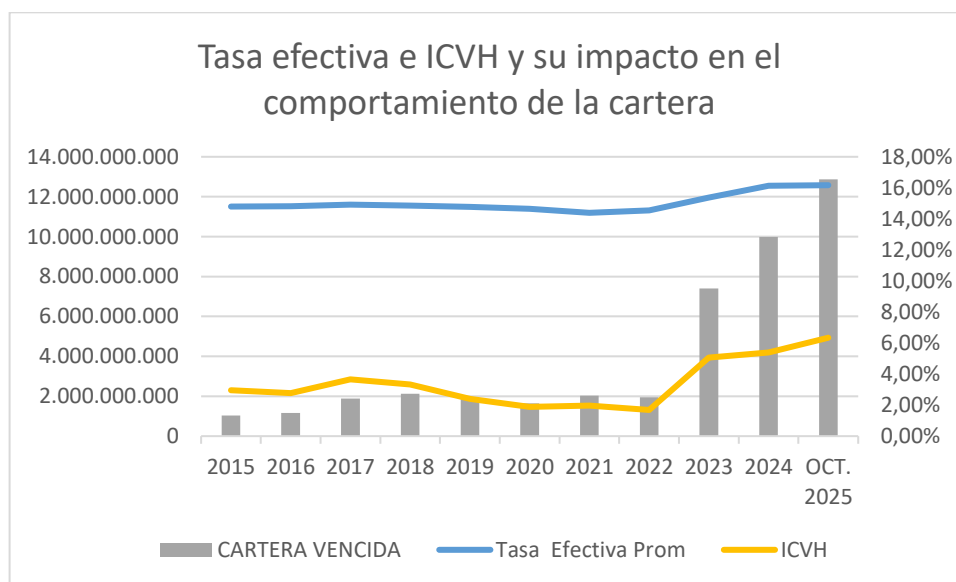


Figura 13. Comportamiento de la cartera de crédito hipotecario, tasa efectiva e ICVH

Fuente: elaboración propia.

En concreto, se resalta que a partir de 2022 la Tasa Efectiva Promedio ha crecido, al tiempo que el ICVH y el volumen de Cartera Vencida han tenido un incremento alarmante y muy marcado. En octubre de 2025, el ICVH aumenta hasta más del 6,0 %, luego de estar en alrededor del 2,0 %. El aumento del riesgo y el volumen de la cartera morosa indica que las elevadas tasas de interés, impulsadas por la política del Banco de la República para luchar contra la inflación, han impactado de manera importante y negativa la capacidad de pago de los prestatarios, lo que ha hecho crecer el riesgo crediticio en Colombia. Es importante analizar con mayor detalle los factores que contribuyen a este aumento, considerando la temporalidad de las tasas de interés y su impacto en el comportamiento de los prestatarios.

La cartera vencida reportada en los datos abarca tanto la morosidad como las pérdidas efectivas por incumplimiento. Es importante destacar que estos valores reflejan no solo las provisiones establecidas como medida preventiva ante posibles pérdidas futuras, sino también los castigos, que son las pérdidas reales registradas cuando un préstamo se considera incobrable. De esta manera, la cifra de cartera vencida proporciona una visión integral de las exposiciones al riesgo crediticio, incluyendo tanto las estimaciones preventivas como las pérdidas reales incurridas.

Este análisis trimestral y anual proporciona una comprensión integral de la dinámica de la cartera de crédito hipotecario del Banco Unión. Estos hallazgos son fundamentales para la evaluación del riesgo crediticio y la implementación de estrategias que fortalezcan la gestión de la cartera en el contexto cambiante del mercado financiero.

Nivel de riesgo crediticio para el Banco Unión: se calcula mediante diversos enfoques y modelos que consideran diferentes variables financieras y crediticias. A continuación, se presenta la forma en la que el Banco Unión calcula actualmente el riesgo crediticio.

- Análisis de puntajes crediticios:
  - Variables consideradas: historial crediticio, días en mora e ingresos.
  - Proceso: cada variable se pondera según su importancia y se asignan puntajes. Estos puntajes se suman para obtener un puntaje crediticio total.
  - Resultado: en la tabla 3 se observa la clasificación de las categorías de riesgo que el Banco Unión le da al prestatario:

Tabla 3. Clasificación nivel de riesgo de la cartera de crédito hipotecario vencida del Banco Unión

<b>Calif. riesgo</b>	<b>Días de vencimiento</b>	<b>Nivel de riesgo</b>
A	1 (0-30)	Muy bajo (10 %)
	2 (>30 <= 60)	
B	3 (>60 <= 90)	Bajo (20 %)
	4 (>90 <= 180)	
C	5 (>180 <= 270)	Medio (30 %)
D	6 (>270 <= 360)	Alto (40 %)
E	9 (>720)	Muy alto (50 %)

Fuente: elaboración propia.

Tabla 4. Descripción de la calificación del riesgo

<b>Calif. riesgo</b>	<b>Descripción</b>
A	Crédito con riesgo crediticio NORMAL. Créditos que reflejan una estructuración y atención apropiadas. Los estados financieros de las contrapartes o los flujos de caja del proyecto, así como la demás información crediticia, indican una capacidad de pago adecuada en términos del monto y origen de los ingresos con que cuentan las contrapartes para atender los pagos requeridos.
B	Crédito con riesgo ACEPTABLE. Créditos que están aceptablemente atendidos y protegidos, pero existen debilidades que potencialmente pueden afectar, transitoria o permanentemente, la capacidad de pago de la contraparte o los flujos de caja del proyecto, en forma tal que, de no ser corregidas oportunamente, llegarían a afectar el normal recaudo del crédito o contrato.
C	Créditos que presentan insuficiencias en la capacidad de pago de la contraparte o en los flujos de caja del proyecto, que comprometen el normal recaudo de la obligación en los términos convenidos.
D	Crédito de difícil cobro, con riesgo SIGNIFICATIVO. Créditos que presentan insuficiencias en la capacidad de pago de la contraparte o en los flujos de caja del proyecto, que comprometen significativamente el recaudo de la obligación en los términos convenidos.

<b>Calif. riesgo</b>	<b>Descripción</b>
E	Crédito IRRECUPERABLE. Crédito que se considera incobrable.

Fuente: elaboración propia.

## RESULTADOS

Los hallazgos de esta investigación posibilitan la evaluación del rendimiento de los modelos implementados para prever el deterioro crediticio en la cartera hipotecaria del Banco Unión. Para ello, primero se muestran los descubrimientos del modelo Logit y luego los resultados obtenidos a través de la plataforma KNIME, con algoritmos de clasificación que se han probado por medio del nodo AutoML.

### Resultados de los modelos estimados

A pesar de que el proceso AutoML posibilitó evaluar varios modelos, se eligieron tres algoritmos fundamentales debido a la disponibilidad de variables y al carácter de la información; estos son:

- Regresión Logística
- Árbol de Decisión
- Random Forest

Cada uno de estos modelos fue evaluado, especialmente en lo que respecta a su precisión y recall, porque son indicadores relevantes en el análisis del riesgo crediticio. La precisión señala la fiabilidad de la clasificación cuando el modelo clasifica a un cliente como arriesgado, mientras que el recall permite determinar qué porcentaje de clientes deteriorados fueron clasificados con acierto



Figura 14. Modelos de clasificación ranqueados por la métrica de precisión (KNIME)

Fuente: herramienta minería de datos KNIME.

Se examinaron las matrices de confusión y las curvas ROC asociadas a las especificaciones estimadas para enriquecer las métricas de recall y precisión, así como para brindar un análisis más profundo del rendimiento clasificatorio de los modelos. Estas herramientas posibilitan la división del rendimiento global en función de los aciertos y fallos de clasificación, lo cual es especialmente importante al examinar el riesgo crediticio, ya que los costos relacionados a las diferentes clases de error no son iguales.

La matriz de confusión del modelo Logit muestra una gran habilidad para clasificar bien a los clientes sin deterioro, lo que es consistente con el alto porcentaje global de aciertos. Sin embargo, se nota que no hay suficiente sensibilidad para detectar a los clientes en default, lo que se evidencia en una cantidad considerable de falsos negativos. Esta conducta está de acuerdo con la configuración del conjunto de datos original, que se distingue por tener una baja frecuencia relativa de incumplimiento. Además, corrobora que el Logit, en este caso, favorece la estabilidad explicativa y la clasificación adecuada de los clientes sanos por encima del descubrimiento anticipado de sucesos poco comunes.

La curva ROC del modelo Logit presenta un área bajo la curva (AUC) de aproximadamente 0,66, lo que señala una capacidad discriminativa moderada, superior a la aleatoriedad, aunque insuficiente para realizar una clasificación individual exacta. Este resultado indica que el modelo puede clasificar a los clientes de manera razonable en función de su nivel de riesgo, aunque tiene dificultades para distinguir sin ambigüedades entre prestatarios deteriorados y no deteriorados. Desde un enfoque aplicado, este rendimiento es apropiado como herramienta de segmentación general del riesgo y de explicación, pero no es suficiente como único instrumento de alerta temprana.

El Árbol de Decisión, por otro lado, estimado en la plataforma KNIME, muestra una matriz de confusión más balanceada en términos de precisión y sensibilidad. El modelo aumenta considerablemente la identificación de clientes en default al disminuir la cifra de falsos negativos, sin que se produzca un incremento significativo en los falsos positivos. Este hallazgo, que es consistente con los valores de recall más altos que se han visto, valida la habilidad del Árbol de Decisión para detectar patrones no lineales y agrupaciones particulares de variables vinculadas a la degradación crediticia.

La idea de que ambas perspectivas se complementan entre sí se fortalece al comparar juntas las matrices de confusión y las métricas ROC. El Árbol de Decisión contribuye a que se mejoren las habilidades operativas para detectar a los clientes

deteriorados, algo fundamental para la administración preventiva del riesgo crediticio en la cartera hipotecaria del Banco Unión. Por su parte, el Logit proporciona una estructura estadística que es interpretable y permite una lectura clara de los elementos determinantes del riesgo.

### **Comparación y selección del modelo**

A pesar de que el modelo Random Forest tuvo el recall más alto, la investigación determinó que el Árbol de Decisión es el más apropiado en términos operativos para el Banco Unión por su simplicidad, al ser explicado ante requerimientos regulatorios, por su estabilidad y su capacidad para ser interpretado. Asimismo, mantiene un balance sólido entre la detección de clientes en mal estado y la correcta clasificación de aquellos que no representan riesgo.

Se aconseja el uso del Árbol de Decisión como modelo complementario al Logit, ya que tiene la habilidad de identificar patrones complejos y no lineales dentro de los datos, lo que proporciona un mayor nivel de sensibilidad sin aumentar la cantidad de falsos positivos. Esta característica es fundamental para una entidad financiera que busca minimizar pérdidas por deterioro sin afectar a los clientes de buen comportamiento.

### **Selección final del modelo y consideraciones operativas**

La determinación de elegir el Árbol de Decisión como modelo alternativo al Logit no solo se basa en los indicadores cuantitativos obtenidos, sino que también se apoya en cuestiones operativas y metodológicas que están en línea con las necesidades reales del Banco Unión. Para la validación interna, el seguimiento de las decisiones y el cumplimiento de las normativas en los modelos de riesgo crediticio resulta fundamental, ya que los reguladores requieren que las entidades financieras tengan la capacidad de explicar con claridad la clasificación de los clientes y los elementos que definen su riesgo.

Aunque modelos más complejos como Random Forest pueden presentar mejoras en la sensibilidad gracias a su estructura, que se basa en varios árboles y procesos de agregación, esta complejidad hace que sea difícil interpretarlos directamente. En cambio, el Árbol de Decisión posibilita observar de forma jerárquica y estructurada los criterios que determinan si un cliente es bueno o malo, lo cual hace más fácil su implementación en equipos técnicos y departamentos administrativos.

El Árbol de Decisión, además, proporciona un beneficio extra al detectar interacciones no lineales entre variables que en numerosas ocasiones los modelos estadísticos convencionales no logran captar. En carteras hipotecarias, esta habilidad de adaptarse es particularmente beneficiosa, ya que el comportamiento del deudor puede verse afectado a la vez por alteraciones en las condiciones

macroeconómicas, en los ingresos del hogar, en el nivel de endeudamiento y en las propiedades del crédito.

La estabilidad del modelo es otro factor que respalda la elección. El Árbol de Decisión demostró resultados coherentes en la capacitación llevada a cabo con AutoML, conservando un balance apropiado entre la precisión y el recall. Esto significa que, sin aumentar la cantidad de falsos positivos, lo cual es esencial para no afectar el origen de clientes con buena conducta, identifica a los clientes deteriorados con mayor fiabilidad y reduce las omisiones que pueden resultar en pérdidas.

### Resultados del modelo Logit

El modelo Logit fue calculado a partir de las variables elegidas a lo largo del proceso metodológico, lo que permitió establecer la correlación entre elementos económicos, atributos del cliente, rasgos del crédito y la probabilidad de impago. Los coeficientes, en su mayoría, presentaron niveles apropiados de significancia, según la evaluación estadística del modelo, lo que señala que las variables incluidas aportan constantemente a la explicación del comportamiento de deterioro.

La investigación de los efectos marginales reveló que algunos elementos vinculados a la situación macroeconómica y a la situación financiera del deudor tienen un impacto significativo en las posibilidades de incurrir en mora. Además, la bondad de ajuste del modelo reveló indicadores satisfactorios, lo cual respalda su eficacia como instrumento analítico en el proceso de administración del riesgo crediticio del Banco Unión.

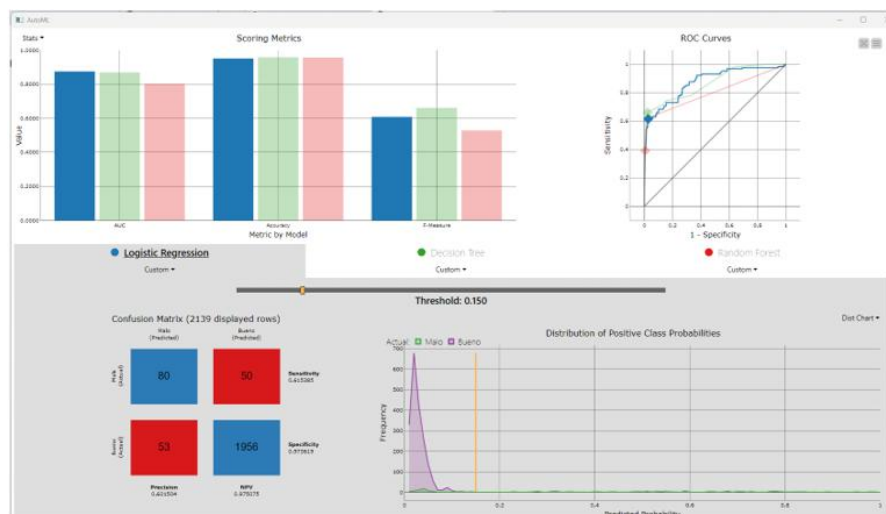


Figura 15. Resultados detallados de la estimación del modelo Logit

Fuente: herramienta minería de datos KNIME.

La capacidad del modelo para distinguir entre clientes con y sin deterioro, evaluada a través de la curva ROC, mostró un desempeño predictivo razonable. Aunque el modelo Logit contribuye a la interpretación y facilita entender cómo cada variable impacta, su capacidad predictiva se complementa con métodos más versátiles como los que se utilizan en la plataforma KNIME.

## Resultados de la modelación en KNIME mediante AutoML

La evaluación de los resultados de tres modelos principales fue posible gracias a que se implementó el nodo AutoML en la plataforma KNIME: Árbol de Decisión, Regresión Logística y Random Forest. Estos algoritmos fueron entrenados con una base de datos balanceada y se les aplicaron criterios de rendimiento basados en dos indicadores clave: precisión y recall.

Los resultados obtenidos fueron los siguientes:

- Regresión Logística: recall 60,15 %, precisión 97,51 %.
- Árbol de Decisión: recall 66,93 %, precisión 97,76 %.
- Random Forest: recall 80,95 %, precisión 96,19 %.

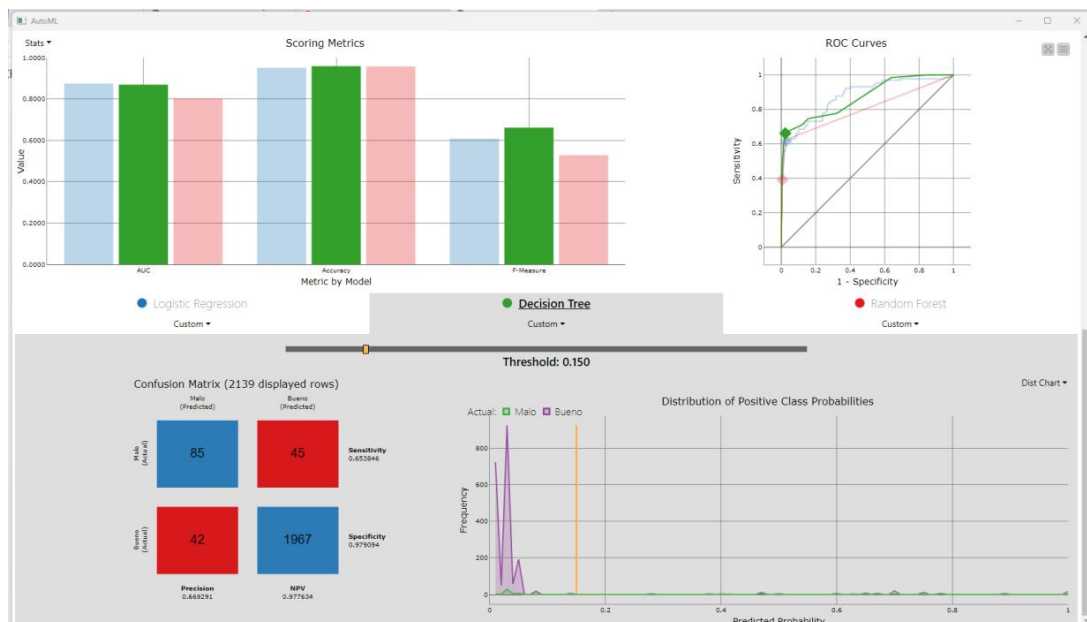


Figura 16. Estructura gráfica del Árbol de Decisión para la clasificación de riesgo

Fuente: herramienta minería de datos KNIME.

El modelo de Árbol de Decisión logró una identificación apropiada de clientes deteriorados sin incrementar los falsos positivos, lo que demuestra un desempeño estable. A pesar de que el Random Forest tuvo el recall más alto, su complejidad metodológica y su escasa interpretabilidad restringieron su idoneidad en el contexto operativo del Banco Unión.

El Árbol de Decisión creó una estructura que posibilitó el reconocimiento de caminos concretos para categorizar a los clientes como buenos o malos, lo cual simplifica su empleo en auditorías, procesos internos y revisiones regulatorias. Para asegurar la trazabilidad en la toma de decisiones, es esencial tener claridad en la segmentación.

### **Comparación integrada de resultados**

La comparación conjunta entre los modelos de KNIME y los resultados del Logit muestra que cada enfoque brinda beneficios complementarios. El Árbol de Decisión mejora la capacidad predictiva al identificar patrones no lineales y conexiones complejas en los datos, mientras que el Logit proporciona una interpretación estadística y claridad acerca del impacto de cada variable.

Esta complementariedad fortalece el análisis del riesgo crediticio, creando una estrategia más sólida y flexible para examinar la degradación de las hipotecas. El Banco Unión obtiene una herramienta completa que une la exactitud operativa con el rigor estadístico al adoptar los dos modelos de manera simultánea.

La implementación del modelo KNIME posibilitó un análisis más profundo de la conducta de la cartera hipotecaria, empleando una perspectiva de clasificación supervisada y el nodo AutoML para comparar diversos algoritmos y elegir el que demostrara un rendimiento más constante y acorde con las exigencias operativas del Banco Unión. Los patrones identificados por los modelos analizados presentaron una notable consistencia, sobre todo en cuanto al peso que tienen las variables asociadas con el ingreso del deudor, la naturaleza del crédito y la organización temporal de la obligación.

El flujo desarrollado en la plataforma, después de limpiar y estandarizar la base de datos, mostró que los predictores siguen una tendencia congruente con lo que se espera en un entorno de riesgo hipotecario: a niveles más altos de ingresos corresponden menores probabilidades de deterioro, y a créditos más antiguos les corresponde un leve incremento del riesgo. Esto indica que algunos patrones de comportamiento tardío pueden predecir un cambio hacia categorías más vulnerables.

El patrón que se observó entre la edad y el deterioro, estudiado a través de las divisiones realizadas por los algoritmos, mostró una respuesta no lineal delicada: los segmentos intermedios de edad evidenciaron un mejor desempeño crediticio,

mientras que los extremos suelen estar relacionados con un ligero aumento en el riesgo. El Árbol de Decisión, cuyo mecanismo particional facilita la visualización jerárquica de estos puntos críticos, capturó este comportamiento con claridad.

Además, la clase de vivienda mantiene un efecto moderado de protección, mientras que ciertos niveles altos de educación, especialmente tener un máster, se presentaron con un riesgo algo más alto. Esta última observación, que concuerda con lo hallado en el modelo Logit, puede ser entendida como un reflejo de perfiles de endeudamiento más exigentes o de dinámicas financieras específicas dentro de ese segmento, en lugar de ser una señal de fragilidad inherente.

El proceso AutoML mostró métricas muy parecidas entre los algoritmos que se compararon en términos de rendimiento global, lo cual demuestra que la estructura de la base de datos tiene relaciones suficientemente bien definidas para ser identificadas por diferentes métodos de clasificación. No obstante, el Árbol de Decisión sobresalió por brindar un balance apropiado entre precisión y recall, con una habilidad más alta para detectar clientes en deterioro sin incrementar los falsos positivos. A pesar de que el modelo Random Forest obtuvo un recall más alto, la ventaja no compensa las desventajas de su escasa interpretabilidad, en particular en un contexto regulado en el que es necesario poder rastrear las decisiones. La regresión logística, en cambio, presentó un comportamiento estable, aunque con menos capacidad para captar interacciones que no son lineales.

Se verificó que la selección definitiva del Árbol de Decisión no solo obedece a criterios estadísticos, sino también a aspectos metodológicos vinculados con la transparencia de las reglas, la estabilidad y la sencillez para implementar operativamente, gracias al análisis de las iteraciones en KNIME. El modelo se mantuvo coherente en el sentido y el peso de sus divisiones principales durante las fluctuaciones que ocurrieron a lo largo del entrenamiento, incluyendo el balance necesario para corregir la desproporción entre clientes buenos y malos. Esto refuerza su fortaleza frente a cambios lógicos en el conjunto de datos. Para el Banco Unión, esta estabilidad es particularmente relevante porque cualquier modelo de riesgo tiene que ser capaz de mantener su comportamiento en diversas condiciones y ciclos económicos.

La versión definitiva del modelo KNIME es una especificación coherente y parsimoniosa, en la que las variables mantienen tanto su significación operacional como su interpretación lógica. En consecuencia, la aplicación del Árbol de Decisión enriquece el diagnóstico proporcionado por el modelo Logit al ampliar su capacidad predictiva a través de una interpretación más flexible de las relaciones no lineales que se encuentran en la cartera. Los hallazgos obtenidos a través de KNIME, en general, proporcionan un respaldo firme para entender los factores que determinan el deterioro del crédito y favorecen la estructura analítica existente para manejar el riesgo hipotecario en la entidad.

## **Validez y capacidad predictiva del modelo**

La evaluación de la validez estadística del modelo Logit sugiere un ajuste adecuado, aunque con limitaciones en su poder discriminante. El porcentaje de casos correctamente clasificados (94 %) refleja una elevada capacidad para identificar a los clientes que no incurrieron en mora, pero una sensibilidad muy baja (0,93 %) indica dificultad para detectar a quienes efectivamente cayeron en default. Este patrón es característico de modelos aplicados a eventos poco frecuentes, donde la proporción de incumplimientos en la muestra es reducida.

La curva ROC muestra un área bajo la curva (AUC) de 0,66, lo que evidencia una capacidad predictiva moderada, superior al azar, pero insuficiente para un modelo de clasificación sólido.

La interpretación de estos resultados debe realizarse considerándose la estructura del conjunto de datos utilizado para el análisis. En la base original, el incumplimiento representa apenas el 6,06 % de las observaciones, lo que explica la elevada tasa de clasificación global combinada con una sensibilidad reducida del modelo Logit.

Este comportamiento es característico de contextos en los que el evento de interés es poco frecuente, ya que los modelos tienden a privilegiar la correcta identificación de la clase mayoritaria.

Por otro lado, los modelos entrenados en la base balanceada a través de KNIME presentaron un notable progreso en la identificación de clientes morosos (default), lo que corrobora que el balanceo de clases permitió una valoración más rigurosa y realista del potencial discriminante de los algoritmos de clasificación. Así, las disparidades que se observan entre los resultados del modelo Logit y los modelos de minería de datos no representan irregularidades en la metodología, sino respuestas previsibles a la configuración de los datos y a las metas complementarias de identificación temprana del riesgo y de explicación.

Los resultados validan la coherencia interna de la especificación, aunque advierten que su fortaleza radica más en la explicación de los determinantes del default que en la predicción individual de la ocurrencia del evento.

## **Interpretación de los efectos marginales**

Los resultados del modelo Logit revelan que el ingreso del deudor es el determinante más robusto y significativo del riesgo de impago. El efecto marginal del logaritmo del ingreso muestra que un incremento del 1 % en el ingreso reduce la probabilidad de default en 3,16 puntos porcentuales ( $p = 0,002$ ), lo cual es estadística y económicamente significativo. Esto implica que, por ejemplo, un

aumento del 1 % en el ingreso del cliente se traduce en una reducción de aproximadamente 3,2 puntos porcentuales en la probabilidad de incumplimiento.

Este hallazgo confirma la importancia fundamental de la capacidad de pago en la evaluación del riesgo crediticio y sugiere que las políticas sobre el origen económico deben mantener estándares estrictos respecto a la verificación y el análisis de ingresos de los solicitantes.

La edad del deudor presenta una relación no lineal con la probabilidad de default, evidenciada por los efectos marginales, tanto del término lineal como del cuadrático. El análisis sugiere una relación en forma de U, donde el riesgo de impago disminuye hasta aproximadamente los 50 años de edad y posteriormente comienza a aumentar.

Específicamente, cada año adicional de edad reduce la probabilidad de default en 0,58 puntos porcentuales ( $p = 0,083$ ), pero este efecto se modera y eventualmente se revierte debido al término cuadrático positivo (0,0000585,  $p = 0,086$ ). Este patrón indica que tanto los clientes muy jóvenes como los de mayor edad representan segmentos de mayor riesgo, posiblemente debido a menor estabilidad financiera en el primer caso y a potenciales vulnerabilidades económicas o de salud en el segundo.

Las características socioeconómicas adicionales muestran efectos diferenciados y reveladores sobre el riesgo crediticio. Los clientes que viven en vivienda familiar presentan una reducción de 2,60 puntos porcentuales en la probabilidad de default ( $p = 0,089$ ), sugiriendo que compartir vivienda con la familia puede implicar menores gastos de vivienda o redes de apoyo financiero informal que facilitan el cumplimiento de las obligaciones crediticias.

Por otro lado, un hallazgo revelador es que los clientes con nivel educativo de magíster presentan un incremento de 19,08 puntos porcentuales en la probabilidad de default ( $p = 0,018$ ). Este resultado, paradójico, podría explicarse por varios factores: primero, los profesionales con títulos de maestría suelen acceder a créditos de mayor monto que pueden comprometer excesivamente su capacidad de pago; segundo, este segmento podría tener mayor propensión al sobreendeudamiento debido a expectativas de ingresos futuros que no siempre se materializan, y, tercero, podrían estar más involucrados en emprendimientos o inversiones que conllevan mayor riesgo financiero.

Para finalizar, las características del crédito mismo también influyen en el riesgo de impago. La antigüedad del crédito muestra un efecto pequeño pero significativo, donde cada mes adicional incrementa la probabilidad de default en 0,07 puntos porcentuales ( $p = 0,034$ ), acumulando aproximadamente 0,84 puntos porcentuales por año de antigüedad.

Este resultado sugiere que los créditos más antiguos han estado expuestos a más choques económicos o deterioro en la situación financiera del deudor a lo largo del tiempo. Por su parte, el plazo del crédito presenta un efecto protector marginal reduciendo la probabilidad de default en 0,04 puntos porcentuales por cada mes adicional ( $p = 0,058$ ), lo cual es consistente con la hipótesis de que plazos más extensos facilitan el cumplimiento al reducir el monto de las cuotas periódicas y aliviar la presión sobre el flujo de caja mensual del deudor.

Estos resultados respaldan la validez estadística y la coherencia económica del modelo Logit estimado. Las variables sociodemográficas y crediticias incluidas capturan relaciones plausibles con el comportamiento de pago, y aunque el poder predictivo es limitado, el modelo ofrece una base sólida para comprender los factores estructurales asociados al riesgo de default dentro de la cartera analizada.

La figura 17 muestra cómo el efecto marginal de la variable “edad” cambia a medida que dicha variable toma diferentes valores. Esta figura indica la sensibilidad de la probabilidad estimada a pequeños cambios en una variable como la edad en diferentes puntos de esa variable.

	Delta-method		z	P> z	[95% Conf. Interval]	
	Margin	Std. Err.				
_at						
1	.4805906	.287864	1.67	0.095	-.0836125	1.044794
2	.4614018	.278257	1.66	0.097	-.0839719	1.006776
3	.4424	.2682386	1.65	0.099	-.083338	.968138
4	.4236283	.2578878	1.64	0.100	-.0818224	.9290791
5	.405127	.2472805	1.64	0.101	-.0795337	.8897878
6	.3869335	.2364887	1.64	0.102	-.0765759	.8504428
7	.3690818	.2255803	1.64	0.102	-.0730474	.811211
8	.3516031	.214618	1.64	0.101	-.0690404	.7722466
9	.3345253	.2036599	1.64	0.100	-.0646408	.7336914
10	.3178731	.1927591	1.65	0.099	-.0599278	.695674
11	.3016682	.1819637	1.66	0.097	-.054974	.6583105
12	.2859294	.1713172	1.67	0.095	-.0498461	.6217049
13	.2706724	.1608588	1.68	0.092	-.0446049	.5859498
14	.2559103	.1506234	1.70	0.089	-.0393061	.5511267
15	.2416533	.1406422	1.72	0.086	-.0340003	.5173069
16	.2279093	.1309427	1.74	0.082	-.0287336	.4845522
17	.2146836	.121549	1.77	0.077	-.0235482	.4529153
18	.2019791	.1124822	1.80	0.073	-.018482	.4224401
19	.1897967	.1037601	1.83	0.067	-.0135693	.3931627

20	.1781351	.0953976	1.87	0.062	-.0088408	.365111
21	.1669911	.087407	1.91	0.056	-.0043234	.3383056
22	.1563596	.0797973	1.96	0.050	-.0000403	.3127596
23	.146234	.0725754	2.01	0.044	.0039889	.2884792
24	.136606	.065745	2.08	0.038	.0077481	.2654638
25	.1274658	.0593077	2.15	0.032	.0112249	.2437067
26	.1188025	.0532623	2.23	0.026	.0144103	.2231947
27	.1106042	.0476057	2.32	0.020	.0172988	.2039096
28	.1028577	.0423325	2.43	0.015	.0198876	.1858278
29	.0955493	.0374355	2.55	0.011	.022177	.1689215
30	.0886643	.0329061	2.69	0.007	.0241695	.1531591
31	.0821878	.0287343	2.86	0.004	.0258696	.138506
32	.0761042	.0249093	3.06	0.002	.0272829	.1249255
33	.0703978	.0214197	3.29	0.001	.0284159	.1123797
34	.0650527	.0182544	3.56	0.000	.0292747	.1008306
35	.0600527	.0154027	3.90	0.000	.0298639	.0902415
36	.0553819	.012856	4.31	0.000	.0301846	.0805792
37	.0510245	.0106084	4.81	0.000	.0302325	.0718165
38	.0469647	.0086593	5.42	0.000	.0299927	.0639366
39	.043187	.0070165	6.16	0.000	.0294348	.0569391
40	.0396763	.0056992	6.96	0.000	.0285062	.0508465
41	.0364178	.0047372	7.69	0.000	.0271331	.0457025
42	.0333971	.0041552	8.04	0.000	.0252531	.0415411
43	.0306002	.0039352	7.78	0.000	.0228873	.0383131

Figura 17. Efectos marginales para diferentes valores de edad

Fuente: elaboración propia a través del software Stata.

En la figura 18 se observa un resultado típico de un modelo de regresión binaria como Logit, que es utilizado en el análisis de riesgo crediticio, y explica cómo la probabilidad de que una persona incumpla con sus obligaciones financieras (default) se puede relacionar con la variable “edad”.

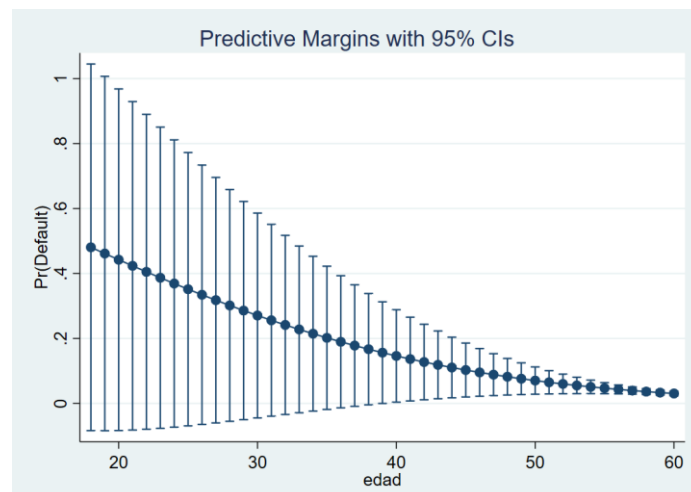
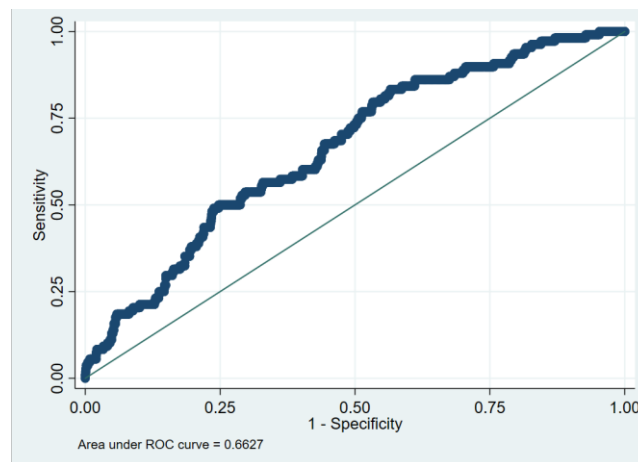


Figura 18. Márgenes predictivos con intervalos de confianza del 95 %

Fuente: elaboración propia a través del software Stata.

El modelo propuesto predice que los individuos más jóvenes tienen una probabilidad mucho más alta de entrar en default que los individuos mayores. Factores como un patrimonio más pequeño, una experiencia financiera inferior o unos ingresos estables más bajos podrían ser la causa de esta diferencia.

A continuación (figura 19), se muestra la Curva Característica Operativa del Receptor (ROC) del modelo de default analizado previamente.



Nota. Se observa un área bajo la curva del modelo final equivalente a 0,6627

Figura 19. Muestra del área bajo la curva ROC del modelo logístico final

Fuente: elaboración propia a través del software Stata.

El valor de 0,6627 es mayor que 0,50, pero se considera un poder predictivo condicionado para un modelo de clasificación. Sin embargo, es superior al azar, ya que señala que el modelo tiene cerca de un 66,27 % de probabilidad de diferenciar adecuadamente entre los prestatarios que incurrirán en default y aquellos que no lo harán.

El uso conjunto del algoritmo de Árbol de Decisión desarrollado en KNIME y el modelo Logit ofrece una visión más flexible y completa de los factores que inciden en el incumplimiento y de las dinámicas que definen la conducta de los morosos.

El modelo Logit demostró una estructura estadística sólida, lo que permitió reconocer conexiones importantes entre las variables financieras, las circunstancias macroeconómicas y la probabilidad de deterioro. Su valor se encuentra en la claridad de interpretación que proporciona, puesto que permite entender cómo los

cambios en indicadores clave alteran el riesgo crediticio. Esta habilidad analítica es importante en los procesos internos del Banco Unión, y la interpretación de resultados es un requisito esencial para evaluar las políticas de concesión, supervisión y recuperación.

Sin embargo, la naturaleza del mercado hipotecario y las recientes modificaciones en el panorama económico exigen modelos más adaptables que puedan identificar comportamientos no lineales y distintivos entre diferentes grupos de clientes. En esta línea, el modelo Árbol de Decisión que se emplea en KNIME posibilitó la identificación de vías de clasificación que muestran interacciones complicadas entre variables, las cuales no siempre son detectables a través de métodos convencionales.

Los hallazgos del Árbol de Decisión dejaron ver una sensibilidad más alta para identificar a los clientes que tienen posibilidades de deteriorarse. Este punto es particularmente importante para la administración de riesgos, puesto que dejar de lado a los clientes con posibilidades de caer en mora significa un gasto directo para la entidad y compromete la estabilidad del portafolio. A pesar de que el modelo Random Forest sobresalió en términos de recall, su escasa interpretabilidad y la dificultad para justificar decisiones en un ambiente regulado hicieron que fuera menos adecuado para el funcionamiento bancario.

La interpretación integrada de ambos modelos permite observar que las variables relacionadas con el perfil financiero del cliente y las condiciones económicas del periodo analizado tienen un peso considerable en la explicación del deterioro. Además, la identificación de patrones mediante el Árbol de Decisión revela que ciertos segmentos de clientes presentan mayor vulnerabilidad en escenarios de incremento en tasas de interés o reducción del ingreso disponible.

La combinación de los dos modelos crea una estrategia más sólida y flexible que posibilita prever riesgos con mayor exactitud, optimizar la segmentación de clientes, modificar las políticas crediticias y elaborar intervenciones a tiempo para disminuir el deterioro.

## CONCLUSIONES

Se puede concluir que la combinación de modelos de clasificación fundamentados en la minería de datos con métodos estadísticos tradicionales es una táctica eficaz para consolidar la gestión del riesgo crediticio en el portafolio hipotecario del Banco Unión, según lo que revelan los resultados conseguidos. El análisis que se ha llevado a cabo muestra que la entidad obtiene ventajas de métodos complementarios que explican las relaciones entre las variables financieras y aumentan la exactitud para identificar clientes con un alto riesgo de deterioro.

El modelo Logit evidenció que los rasgos del crédito, las características del deudor y los elementos macroeconómicos tienen un impacto notorio en el peligro de incumplimiento. La estructura interpretativa que poseen permitió entender la manera en que las fluctuaciones de estas variables influyen en la probabilidad de deterioro, lo cual proporciona insumos significativos para establecer y modificar las políticas de origen, de seguimiento y recuperación.

La utilización del modelo Árbol de Decisión, a través de KNIME, mejoró la habilidad predictiva del análisis, particularmente en la identificación de los clientes más vulnerables. La capacidad del modelo para reducir la omisión de casos deteriorados, un factor esencial para la estabilidad del portafolio hipotecario, se evidenció a través de su rendimiento, que combina adecuadamente el recall y la precisión. Asimismo, su interpretación sencilla favorece que se adopte en el contexto operativo del Banco, donde es importante tener criterios de clasificación claros.

Los dos modelos posibilitan una valoración más exhaustiva del riesgo crediticio. El Árbol de Decisión facilita la identificación de patrones complejos en los datos y, al hacerlo, proporciona más flexibilidad para clasificar el portafolio estudiado con mayor precisión. Por su parte, el Logit ayuda a comprender los elementos que inciden en el deterioro.

El primer objetivo se logró al examinar la literatura sobre modelos de riesgo crediticio, lo que ayudó al reconocimiento de los enfoques teóricos y metodológicos que se han usado a lo largo del tiempo para prever el deterioro y entender su importancia en el entorno hipotecario.

El segundo objetivo se cumplió mediante la evaluación del modelo Logit y de los algoritmos que se probaron en KNIME. Esto permitió establecer que la combinación de ambos métodos aumenta la capacidad predictiva del análisis y simplifica la clasificación de clientes con diferentes grados de riesgo.

El tercer objetivo se alcanzó al proponer sugerencias para poner en marcha estrategias de gestión del riesgo que incorporen los descubrimientos de ambos

modelos. Entre estas acciones se encuentran el fortalecimiento de los procesos de monitoreo, la implementación de criterios de segmentación fundados en el Árbol de Decisión y la modificación de las políticas crediticias según las variables que se han señalado como determinantes del deterioro.

El estudio evidencia que la implementación paralela de un modelo estadístico interpretativo y de otro predictivo, fundamentado en la minería de datos, proporciona al Banco Unión una herramienta sólida para prever riesgos, mejorar las decisiones estratégicas y propiciar una administración más eficiente de su cartera hipotecaria.

Por lo tanto, se cumplieron los objetivos específicos de la investigación y se estableció una base sólida para la implementación práctica de estrategias de gestión de riesgos en la cartera hipotecaria del Banco Unión. Este modelo integral contribuirá a la optimización del nivel de riesgo crediticio de la entidad financiera en un mercado hipotecario dinámico y desafiante.

## RECOMENDACIONES

Los resultados logrados indican que la valía del estudio no se restringe a estimar y comparar modelos, sino que también pueden ser implementados de manera gradual en la administración del riesgo crediticio del Banco Unión. En este contexto, la evidencia empírica posibilita trazar un plan de aplicación por etapas que combina el rigor estadístico con las exigencias regulatorias y operativas de la entidad.

Se sugiere que en una primera fase se adopten los modelos en un entorno piloto de validación interna, empleando datos recientes y funcionando simultáneamente con los sistemas existentes para evaluar y monitorear el riesgo. Esta fase posibilitaría la comparación del rendimiento del Logit y el Árbol de Decisión en situaciones reales, el análisis de la estabilidad de las métricas de clasificación y la modificación de los umbrales decisionales sin perjudicar el proceso formal de concesión y supervisión crediticia.

En una segunda etapa, el modelo Logit tiene la posibilidad de consolidarse como un instrumento estructural para el análisis explicativo y la segmentación del riesgo, lo que ayuda a definir perfiles crediticios y revisar las políticas de origen. El Árbol de Decisión, como sistema de alerta temprana enfocado en detectar clientes con mayor riesgo de deterioro, se puede sumar para fortalecer los procesos de seguimiento preventivo y la priorización de acciones posteriores.

Finalmente, en una tercera etapa, la combinación de los dos modelos daría lugar a una estrategia analítica más sólida. Esta integraría la capacidad no lineal y adaptativa del Árbol de Decisión con la interpretación estadística y económica del Logit. Esta fase supone convertir el modelo en un componente del Sistema de Administración del Riesgo Crediticio, con procesos de recalibración, valoración de desempeño y documentación, que se llevan a cabo periódicamente, cumpliendo así con las exigencias de trazabilidad, transparencia y control que los organismos reguladores establecen.

Esta perspectiva progresiva disminuye los peligros vinculados a la implementación de modelos cuantitativos, mejora su aceptación dentro de la organización y optimiza el uso de la información existente, lo que posibilita que el Banco Unión potencie su capacidad prospectiva ante el deterioro de la cartera hipotecaria en un contexto económico inestable.

## REFERENCIAS

- Altman, E. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The Journal of Finance*, 23(4), 589-609.
- Banco de la República (2022). Informe de estabilidad financiera 2021. <https://www.banrep.gov.co/es/informe-estabilidad-financiera>.
- Banco de la República (2023). Análisis de la cartera y del mercado inmobiliario en Colombia. Informe especial de estabilidad financiera - Primer semestre 2023. [https://repositorio.banrep.gov.co/bitstream/handle/20.500.12134/10663/Informe\\_Especial\\_vivienda\\_2023-I.pdf](https://repositorio.banrep.gov.co/bitstream/handle/20.500.12134/10663/Informe_Especial_vivienda_2023-I.pdf).
- Banco de Pagos Internacionales (2019). *Basilea III: marco regulador internacional para los bancos*. [https://www.bis.org/bcbs/basel3\\_es.htm](https://www.bis.org/bcbs/basel3_es.htm).
- Basel Committee on Banking Supervision (2017). *Sound practices: implications of fintech developments for banks and bank supervisors*. Bank for International Settlements. <https://www.bis.org/bcbs/publ/d415.pdf>.
- Becker, B., Ivashina, V. y Ostrovsky, M. (2012). Information as regulation. *The Review of Financial Studies*, 25(9), 2613-2649.
- Berthold, M., Cebron, N., Dill, F., Gabriel, T., Kötter, T., Meinl, T., ... y Wiswedel, B. (2009). KNIME. The Konstanz Information Miner: Version 2.0 and beyond. En C. Preisach *et al.* (Eds.), *Data analysis, machine learning and applications* (pp. 319-326). Springer.
- Black, T. (2015). *Real estate finance and investments*. McGraw-Hill Education.
- Bluhm, C., Overbeck, L. y Wagner, C. (2002). *An introduction to credit risk modeling*. CRC Press.
- Campbell, J. y Loury, G. (1991). A model of credit market equilibrium with imperfect information. En W. T. Barnett y K. J. Singleton (Eds.), *New approaches to monetary economics: Proceedings of the Second International Symposium in Economic Theory and Econometrics* (pp. 353-381). Cambridge University Press.
- Cañón, M. y Suárez, J. (2016). Riesgo de crédito en la cartera hipotecaria de Colombia (2006-2014). [https://ciencia.lasalle.edu.co/finanzas\\_comercio/23](https://ciencia.lasalle.edu.co/finanzas_comercio/23).
- Consejo Técnico de la Contaduría Pública (2014). Enmienda a la NIIF 9 instrumentos. <https://www.ctcp.gov.co/proyectos/contabilidad-e-informacion-financiera/documentos-discusion-publica/enmienda-a-la-niif-9-instrumentos/doc-ctcp-xl8w9-141>.

- Dell'Ariccia, G., Laeven, L. y Suárez, G. (2012a). Bank leverage and monetary policy's risk-taking channel: Evidence from the United States. *IMF Working Papers*, 12(245), 1-34.
- Dell'Ariccia, G., Igan, D. y Laeven, L. (2012b). Credit booms and lending standards: Evidence from the subprime mortgage market. *Journal of Money, Credit and Banking*, 44(2-3), 367-384.
- Departamento Administrativo Nacional de Estadística (DANE). (2025). Producto Interno Bruto (PIB) nacional trimestral. <https://www.dane.gov.co/index.php/estadisticas-por-tema/cuentas-nacionales/cuentas-nacionales-trimestrales/pib-informacion-tecnica>.
- Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, 27(8), 861-874. <https://doi.org/10.1016/j.patrec.2005.10.010>.
- Gómez, A. (2020). Revisión de modelos de riesgo crediticio: una perspectiva práctica. *Revista de Banca y Finanzas*, 18(1), 55-72.
- González, L. (2021). Medidas de alivio financiero en el sector hipotecario durante la pandemia. *Revista de Finanzas Aplicadas*, 11(2), 45-62.
- Green, R. y Hendershott, P. (1996). Age, housing demand, and real house prices. *Regional Science and Urban Economics*, (26), 465-480.
- Greene, W. (2003). *Econometric analysis*. Prentice Hall.
- Gujarati, D. y Porter, D. (2009). *Basic econometrics*. McGraw Hill.
- Hair, J., Black, W., Babin, B. y Anderson, R. (2019). *Multivariate data analysis*. Cengage.
- Hastie, T., Tibshirani, R. y Friedman, J. (2009). *The elements of statistical learning*. Springer.
- He, H. y García, E. (2009). Learning from imbalanced data. *IEEE Transactions on Knowledge and Data Engineering*, 21(9), 1263-1284. <https://doi.org/10.1109/TKDE.2008.239>.
- Heuson, A. y Smith, G. (1995). Default probabilities and interest rates on conventional and insured mortgages. *Journal of Housing Research*, 6(2), 271-294.
- Hosmer, D., Lemeshow, S. y Sturdivant, R. (2013). *Applied logistic regression*. John Wiley & Sons.
- Jorion, P. (2007). *Value at risk: The new benchmark for managing financial risk*. McGraw-Hill.

- Kealhofer, S. (2010). Contingent claims analysis of corporate capital structures: An empirical investigation. *Journal of Financial and Quantitative Analysis*, 45(3), 613-646.
- Khandani, A., Kim, A. y Lo, A. (2010). Consumer credit-risk. Models via machine-learning algorithms. *Journal of Banking & Finance*, 34(11), 2767-2787.
- Kleinbaum, D. y Klein, M. (2012). *Survival analysis: A self-learning text*. Springer.
- Kothari, C. (2004). *Research methodology: Methods and techniques*. New Age International Publishers.
- Kuhn, M. y Johnson, K. (2013). *Applied predictive modeling*. Springer.
- Kutner, M., Nachtsheim, C., Neter, J. y Li, W. (2005). *Applied linear statistical models*. McGraw-Hill.
- Laeven, L. y Valencia, F. (2012). Systemic banking crises database: An update. *IMF Economic Review*, 60(2), 229-255.
- Levy, N. (2012). Tasas de interés, demanda efectiva y crecimiento económico. *Economía UNAM*, 9(25), 74-93.
- Lucas, A. y Klaassen, P. (2006). Discrete versus continuous state switching models for portfolio credit risk. *Journal of Banking & Finance*, 30(1), 23-35.
- Malhotra, R. y Malhotra, D. (2015). *Handbook of research on strategic risk. Management in banking and finance*. IGI Global.
- Martínez, R. (2019). Modelos de riesgo crediticio: evaluación y aplicabilidad en el mercado hipotecario. *Journal of Financial Management*, 35(3), 123-140.
- Meneses, L. y Macuacé, R. (2011). Valoración y riesgo crediticio en Colombia. *Finanzas y Política Económica*, 3(2), 65-82. <https://revfinypolecon.ucatolica.edu.co/article/view/519>.
- Merton, R. (1974). On the pricing of corporate debt: The risk structure of interest rates. *The Journal of Finance*, 29(2), 449-470.
- Mishkin, F. S. (2011). *The economics of money, banking, and financial markets*. Pearson.
- Montoriol, J. (2023). El precio de la vivienda aceleró el paso en el 3 de 2023. <https://www.caixabankresearch.com/es/publicaciones/notas-breves-actualidad-economica-y-financiera/espana/precio-vivienda-acelero-paso-3t>.
- Phillips, R. y Siems, T. (2008). The role of banks in the transmission of monetary policy. <http://www.bostonfed.org/economic/ppdp/index.htm>.

- Rodríguez, E. (2021). Impacto de la gestión del riesgo crediticio en las instituciones financieras. *Journal of Banking and Finance*, 42(4), 301-318.
- Scott, W. (1965). A score model for predicting the creditworthiness of borrowers. *Journal of Finance*, 20(4), 651-665.
- Shiller, R. (2012). *Finance and the good society*. Princeton University Press.
- Siddiqi, N. (2016). *Intelligent credit scoring: Building and implementing better credit risk scorecards*. John Wiley & Sons.
- Smith, C. (2010). *Mortgage banking*. John Wiley & Sons.
- Superintendencia Financiera de Colombia (1995). Reglas relativas a la gestión del riesgo crediticio. <https://www.fasecolda.com/cms/wp-content/uploads/2019/08/ce100-1995-cap-ii.pdf>.
- Thomas, L., Crook, J. y Edelman, D. (2017). *Credit scoring and its applications*. SIAM. <https://doi.org/10.1137/1.9781611974560>.
- Urrea-Ríos, J. y Piraján, C. (2020). Impacto de la pandemia de covid-19 en la economía colombiana. *Revista de Economía del Rosario*, 23(2), 197-220.
- Wooldridge, J. (2010). *Introducción a la econometría. Un enfoque moderno*. Cengage.

## ANEXOS

### Anexo A. Estimación modelo Logit

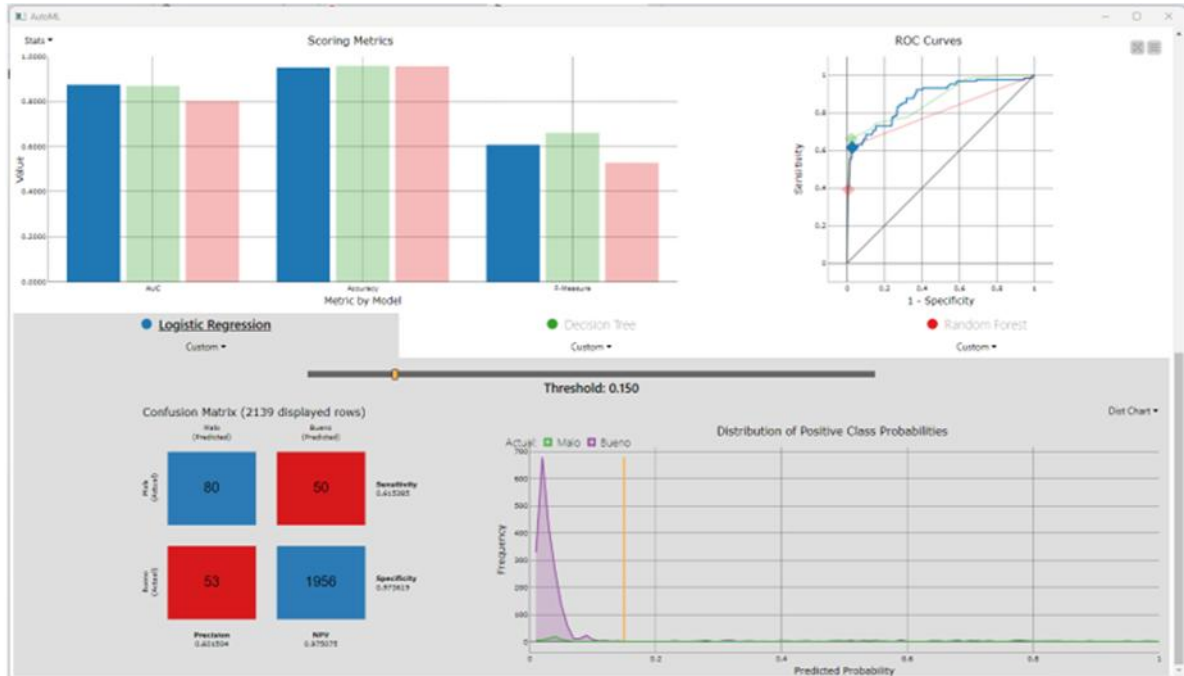
Logistic regression  Log likelihood = <b>-388.55927</b>	Number of obs = <b>1,782</b> LR chi2(7) = <b>37.73</b> Prob > chi2 = <b>0.0000</b> Pseudo R2 = <b>0.0463</b>
---	---

default	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
edad	-.1047332	.0601147	-1.74	0.081	-.2225559 .0130894
edad_cuadrado	.0010548	.000611	1.73	0.084	-.0001428 .0022524
log_ingreso	-.5692235	.1792536	-3.18	0.001	-.9205541 -.217893
d_tipo_vivi~2	-.4677324	.2738353	-1.71	0.088	-1.00444 .0689748
d_nivel_est~3	3.439272	1.444786	2.38	0.017	.6075437 6.271
antiguedad~s	.0117589	.0054841	2.14	0.032	.0010102 .0225075
plazo_meses	-.0079922	.0041781	-1.91	0.056	-.0161811 .0001966
_cons	9.943222	3.158458	3.15	0.002	3.752757 16.13369

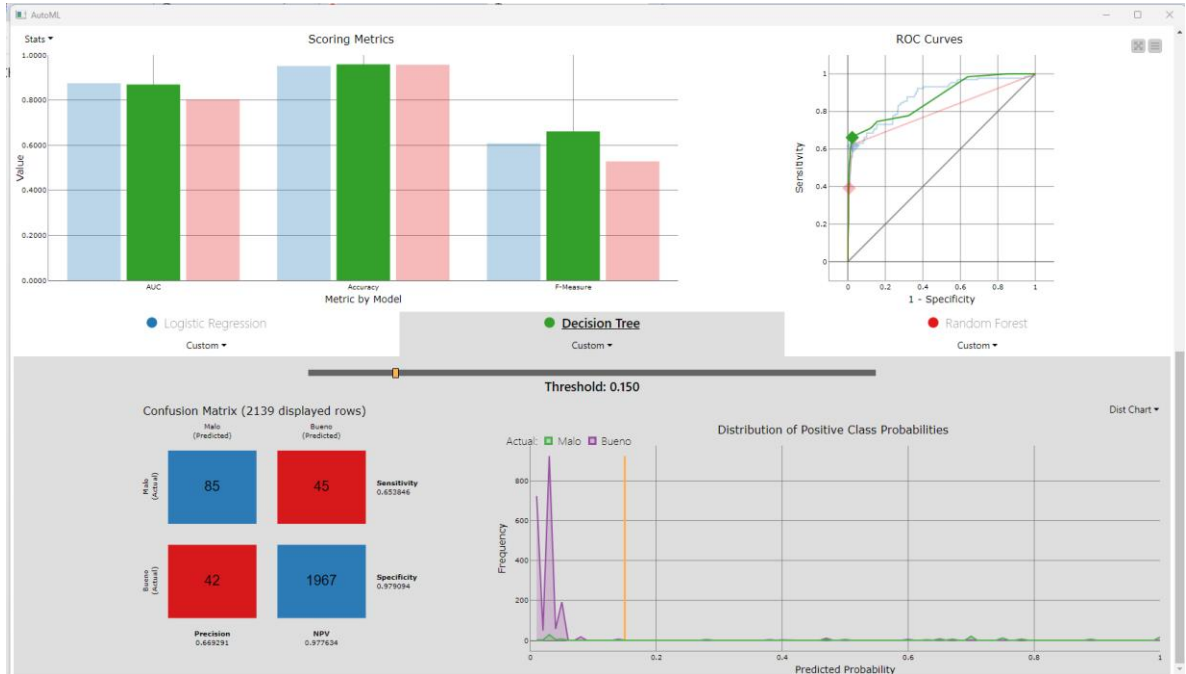
Fuente: elaboración propia a través del software Stata.

## Anexo B. Resultados detallados de la estimación del modelo Logit



Fuente: herramienta minería de datos KNIME.

## Anexo C. Estructura gráfica del Árbol de Decisión para la clasificación de riesgo



Fuente: herramienta minería de datos KNIME.

## Anexo D. Capacidad predictiva del modelo

Logistic model for default

Classified	True		Total
	D	~D	
+	<b>1</b>	<b>0</b>	<b>1</b>
-	<b>107</b>	<b>1674</b>	<b>1781</b>
Total	<b>108</b>	<b>1674</b>	<b>1782</b>

Classified + if predicted  $\Pr(D) \geq .5$   
 True D defined as default  $\neq 0$

Sensitivity	$\Pr(+ D)$	<b>0.93%</b>
Specificity	$\Pr(- \sim D)$	<b>100.00%</b>
Positive predictive value	$\Pr(D +)$	<b>100.00%</b>
Negative predictive value	$\Pr(\sim D -)$	<b>93.99%</b>
False + rate for true ~D	$\Pr(+ \sim D)$	<b>0.00%</b>
False - rate for true D	$\Pr(- D)$	<b>99.07%</b>
False + rate for classified +	$\Pr(\sim D +)$	<b>0.00%</b>
False - rate for classified -	$\Pr(D -)$	<b>6.01%</b>
Correctly classified		<b>94.00%</b>

Fuente: elaboración propia a través del software Stata.

## Anexo E. Cálculos AIC y BIC

N	1782	1782
Log-Lik	-388.559	-388.379
AIC	793.119	792.758
BIC	837.002	836.641

Standard errors in parentheses  
\* p<0.10, \*\* p<0.05, \*\*\* p<0.01

Fuente: elaboración propia a través del software Stata.

## Anexo F. Resultados del modelo Random Forest



Fuente: herramienta minería de datos KNIME.

## Anexo G. Modelo logístico por default

Logistic model for default

Classified	True		Total
	D	~D	
+	<b>1</b>	<b>0</b>	<b>1</b>
-	<b>107</b>	<b>1674</b>	<b>1781</b>
Total	<b>108</b>	<b>1674</b>	<b>1782</b>

Classified + if predicted Pr(D) >= .5  
True D defined as default != 0

Sensitivity	Pr( +  D)	<b>0.93%</b>
Specificity	Pr( -  ~D)	<b>100.00%</b>
Positive predictive value	Pr( D  +)	<b>100.00%</b>
Negative predictive value	Pr( ~D  -)	<b>93.99%</b>
False + rate for true ~D	Pr( +  ~D)	<b>0.00%</b>
False - rate for true D	Pr( -  D)	<b>99.07%</b>
False + rate for classified +	Pr( ~D  +)	<b>0.00%</b>
False - rate for classified -	Pr( D  -)	<b>6.01%</b>
Correctly classified		<b>94.00%</b>

Fuente: elaboración propia a través del software Stata.

**Anexo H. Efecto marginal de un año adicional de edad sobre la probabilidad de default**

	Delta-method					[95% Conf. Interval]	
	dy/dx	Std. Err.	z	P> z			
edad	-.0058115	.003356	-1.73	0.083	-.0123891	.0007661	
edad_cuadrado	.0000585	.0000341	1.72	0.086	-8.31e-06	.0001254	
log_ingreso	-.0315853	.0101677	-3.11	0.002	-.0515137	-.0116569	
d_tipo_vivi~2	-.0259537	.0152824	-1.70	0.089	-.0559067	.0039993	
d_nivel_est~3	.1908397	.0805235	2.37	0.018	.0330165	.3486629	
antiguedad_~s	.0006525	.0003077	2.12	0.034	.0000495	.0012555	
plazo_meses	-.0004435	.0002336	-1.90	0.058	-.0009014	.0000144	

Fuente: elaboración propia a través del software Stata.

## Anexo I. Efectos marginales para diferentes valores de edad

_at	Delta-method		z	P> z	[95% Conf. Interval]	
	Margin	Std. Err.				
1	.4805906	.287864	1.67	0.095	-.0836125	1.044794
2	.4614018	.278257	1.66	0.097	-.0839719	1.006776
3	.4424	.2682386	1.65	0.099	-.083338	.968138
4	.4236283	.2578878	1.64	0.100	-.0818224	.9290791
5	.405127	.2472805	1.64	0.101	-.0795337	.8897878
6	.3869335	.2364887	1.64	0.102	-.0765759	.8504428
7	.3690818	.2255803	1.64	0.102	-.0730474	.811211
8	.3516031	.214618	1.64	0.101	-.0690404	.7722466
9	.3345253	.2036599	1.64	0.100	-.0646408	.7336914
10	.3178731	.1927591	1.65	0.099	-.0599278	.695674
11	.3016682	.1819637	1.66	0.097	-.054974	.6583105
12	.2859294	.1713172	1.67	0.095	-.0498461	.6217049
13	.2706724	.1608588	1.68	0.092	-.0446049	.5859498
14	.2559103	.1506234	1.70	0.089	-.0393061	.5511267
15	.2416533	.1406422	1.72	0.086	-.0340003	.5173069
16	.2279093	.1309427	1.74	0.082	-.0287336	.4845522
17	.2146836	.121549	1.77	0.077	-.0235482	.4529153
18	.2019791	.1124822	1.80	0.073	-.018482	.4224401
19	.1897967	.1037601	1.83	0.067	-.0135693	.3931627
20	.1781351	.0953976	1.87	0.062	-.0088408	.365111
21	.1669911	.087407	1.91	0.056	-.0043234	.3383056
22	.1563596	.0797973	1.96	0.050	-.0000403	.3127596
23	.146234	.0725754	2.01	0.044	.0039889	.2884792
24	.136606	.065745	2.08	0.038	.0077481	.2654638
25	.1274658	.0593077	2.15	0.032	.0112249	.2437067
26	.1188025	.0532623	2.23	0.026	.0144103	.2231947
27	.1106042	.0476057	2.32	0.020	.0172988	.2039096
28	.1028577	.0423325	2.43	0.015	.0198876	.1858278
29	.0955493	.0374355	2.55	0.011	.022177	.1689215
30	.0886643	.0329061	2.69	0.007	.0241695	.1531591
31	.0821878	.0287343	2.86	0.004	.0258696	.138506
32	.0761042	.0249093	3.06	0.002	.0272829	.1249255
33	.0703978	.0214197	3.29	0.001	.0284159	.1123797
34	.0650527	.0182544	3.56	0.000	.0292747	.1008306
35	.0600527	.0154027	3.90	0.000	.0298639	.0902415
36	.0553819	.012856	4.31	0.000	.0301846	.0805792
37	.0510245	.0106084	4.81	0.000	.0302325	.0718165
38	.0469647	.0086593	5.42	0.000	.0299927	.0639366
39	.043187	.0070165	6.16	0.000	.0294348	.0569391
40	.0396763	.0056992	6.96	0.000	.0285062	.0508465
41	.0364178	.0047372	7.69	0.000	.0271331	.0457025
42	.0333971	.0041552	8.04	0.000	.0252531	.0415411
43	.0306002	.0039352	7.78	0.000	.0228873	.0383131

Fuente: elaboración propia a través del software Stata.