



**FACULTAD DE POSTGRADO
TRABAJO FINAL DE GRADUACIÓN**

**EVALUACIÓN DE MODELO HÍBRIDO DE CALIFICACIÓN DE
RIESGO CREDITICIO PARA BANCA EMPRESARIAL EN
HONDURAS, PERÍODO DE ENERO 2020 - OCTUBRE 2024**

SUSTENTADO POR:

**ÁLVARO MOISÉS DE LA ROCHA RODRÍGUEZ
SKARLETL MICHELL NAVARRO GALEANO**

PREVIA INVESTIDURA AL TÍTULO DE

**MÁSTER EN
ANALÍTICA DE NEGOCIOS**

TEGUCIGALPA, FRANCISCO MORAZÁN, HONDURAS, C.A.

ENERO, 2026

**UNIVERSIDAD TECNOLÓGICA CENTROAMERICANA
UNITEC**

FACULTAD DE POSTGRADO

AUTORIDADES UNIVERSITARIAS

RECTORA

ROSALPINA RODRÍGUEZ

VICERRECTOR ACADÉMICO NACIONAL

JAVIER ABRAHAM SALGADO LEZAMA

SECRETARIO GENERAL

ROGER MARTÍNEZ MIRALDA

DECANA FACULTAD DE POSTGRADO

ANA DEL CARMEN RETTALLY VARGAS

**EVALUACIÓN DE MODELO HÍBRIDO DE
CALIFICACIÓN DE RIESGO CREDITICIO PARA
BANCA EMPRESARIAL EN HONDURAS, PERÍODO DE
ENERO 2020 - OCTUBRE 2024**

**TRABAJO PRESENTADO EN CUMPLIMIENTO DE LOS
REQUISITOS EXIGIDOS PARA OPTAR AL TÍTULO DE
MÁSTER EN
ANALÍTICA DE NEGOCIOS**

ASESOR

JESÚS RICARDO RODRÍGUEZ RIVERA

MIEMBROS DE LA TERNA:

**DAVID ANTONIO MEJÍA DIAZ
ÁNGELA PAOLA IZAGUIRRE BONILLA
KEVIN EDUARDO FUNEZ FUNEZ**

DERECHOS DE AUTOR

© Copyright 2026
Álvaro Moisés de la Rocha Rodríguez
Skarletl Michell Navarro Galeano

Todos los derechos son reservados.



FACULTAD DE POSTGRADO

EVALUACIÓN DE MODELO HÍBRIDO DE CALIFICACIÓN DE RIESGO CREDITICIO PARA BANCA EMPRESARIAL EN HONDURAS, PERÍODO DE ENERO 2020 - OCTUBRE 2024

**ÁLVARO MOISÉS DE LA ROCHA RODRÍGUEZ
SKARLETL MICHELL NAVARRO GALEANO**

Resumen

Esta investigación evaluó el impacto de un modelo híbrido de calificación de riesgo crediticio, que combina aprendizaje automático y reglas de negocio, aplicado a clientes de banca empresarial en Honduras entre enero de 2020-octubre de 2024. Se utilizó un enfoque cuantitativo con la metodología CRISP-DM, a partir de una base de 6,903 clientes y 42 variables financieras, conductuales, demográficas y regulatorias. Los datos se dividieron en entrenamiento (2020-2022), validación (2023) y prueba out-of-time (2024). Se comparó un esquema de reglas, una regresión logística y modelos de ML (Random Forest y XGBoost), evaluando su desempeño mediante AUC-ROC, KS y métricas operativas al umbral definido en la validación. En el conjunto de prueba, Random Forest alcanzó AUC-ROC 0.87 y KS 0.58, y XGBoost AUC-ROC 0.86 y KS 0.55, superando ampliamente al enfoque tradicional (AUC 0.59; KS 0.18) y elevando el recall de la clase morosa de 0.66 a 0.87, con estabilidad temporal ($PSI \approx 0.0066$) y buena calibración (Brier $\approx 0.05-0.06$). La reducción de falsos negativos se tradujo en pérdidas evitadas estimadas en L13 millones y un ROI de $\approx 825.95\%$. Se recomendó la implementación piloto del modelo híbrido bajo un marco de gobierno de modelo acorde con Basilea III, BCBS 239 y SR 11-7.

Palabras claves: Aprendizaje automático, Banca empresarial, Honduras, Modelo híbrido, Riesgo crediticio.



GRADUATE SCHOOL

EVALUATION OF HYBRID CREDIT RISK RATING MODEL FOR CORPORATE BANKING IN HONDURAS, PERIOD FROM JANUARY 2020 - OCTOBER 2024

ÁLVARO MOISÉS DE LA ROCHA RODRÍGUEZ
SKARLETL MICHELLE NAVARRO GALEANO

Abstract

This study evaluated the impact of a hybrid credit risk rating model that combines machine learning techniques and business rules, applied to corporate banking clients in Honduras between January 2020 and October 2024. A quantitative approach based on the CRISP-DM methodology was used, drawing on a dataset of 6,903 clients and 42 financial, behavioural, demographic and regulatory variables. The data were split into training (2020-2022), validation (2023) and out-of-time test (2024) sets. A rule-based scheme, a logistic regression and machine learning models (Random Forest and XGBoost) were compared, and their performance was assessed using AUC-ROC, KS and operational metrics at the decision threshold defined in the validation stage. In the test set, Random Forest achieved an AUC-ROC of 0.87 and a KS of 0.58, and XGBoost an AUC-ROC of 0.86 and a KS of 0.55, clearly outperforming the traditional rule-based approach (AUC 0.59; KS 0.18) and increasing the recall of the default class from 0.66 to 0.87, with temporal stability ($PSI \approx 0.0066$) and good calibration ($Brier \approx 0.05-0.06$). The reduction in false negatives translated into avoided losses estimated at L 13 million and a ROI of approximately 825.95 %. A pilot implementation of the hybrid model is recommended under a model-risk governance framework aligned with Basel III, BCBS 239 and SR 11-7.

Keywords: Corporate banking, Credit risk, Honduras, Hybrid model, Machine learning

DEDICATORIA

“La fortaleza nace del espíritu que se atreve” - Víctor Hugo

A Dios, por darme la fuerza y la claridad para completar este camino, a quienes me guiaron con su ejemplo y acompañamiento, les dedico este logro.

A mi familia, por su inquebrantable confianza y por haberme impulsado a perseverar incluso en circunstancias inciertas. Este trabajo es una manifestación de la resiliencia y la resolución que ustedes me inculcaron.

Skarletl Michell Navarro Galeano

“El mundo está en manos de aquellos que tienen el coraje de soñar y de correr el riesgo de vivir sus sueños” - Paulo Coelho.

Dedico este trabajo, ante todo, a mi familia, pilar y refugio en cada etapa de este proyecto de vida.

A mi esposa, por su amor incondicional, por las horas que me regaló de su tiempo, por su paciencia en los días de cansancio y por recordarme, cuando dudaba, por qué había comenzado este camino. Su confianza fue la fuerza silenciosa que me sostuvo cuando las responsabilidades parecían superar mis fuerzas.

A mis hijos, por ser mi mayor inspiración. Sus sonrisas, sus preguntas y su manera de ver el mundo me recordaron que estudiar y esforzarme tiene sentido cuando se hace pensando en dejarles un mejor ejemplo y un futuro con más oportunidades. Cada página de este trabajo lleva, de alguna forma, un pedacito de los sacrificios que compartimos como familia.

A ellos les debo no solo la culminación de esta maestría, sino la certeza de que no hay meta imposible cuando se camina acompañado de amor, fe y esperanza.

Álvaro Moisés De La Rocha Rodríguez

AGRADECIMIENTO

Como dijo Khail Gibran en *El Profeta* (1923): “El maestro que inspira no es el que llena, sino el que enciende.”

Gracias a nuestros docentes por encender en nosotros la curiosidad y el deseo de aprender cada día, por guiarnos con paciencia y por compartir su conocimiento de manera tan inspiradora.

Y como escribió C.S Lewis en *Cartas del Diablo a su sobrino* (1938): “Nadie que viva plenamente, y tenga amigos, se perderá en soledad.”

A mis compañeros, gracias por su compañía, sus ideas, su apoyo y por hacer de este aprendizaje una experiencia compartida, enriquecedora y memorable. Aprender junto a ustedes ha sido un privilegio, me llevó lo esencial: el conocimiento, la amistad y la inspiración que cada uno de ustedes me brindó.

Skarletl Michell Navarro Galeano

A quienes me enseñaron que los logros y éxitos se construyen con constancia, disciplina y fe.

Agradezco profundamente a mi familia, por ser el cimiento de este proyecto. A mi esposa, por su paciencia, apoyo inquebrantable y por sostenerme en los momentos de mayor exigencia; a mis hijos, por recordarme con su cariño el verdadero sentido de cada esfuerzo y por convertirse en mi principal fuente de motivación.

Extiendo mi agradecimiento a los docentes de la maestría, por compartir generosamente su experiencia y por retarme a ir siempre un paso más allá. A mis amigos y compañeros de estudio, gracias por el acompañamiento, las ideas compartidas y el espíritu de colaboración que enriqueció este camino.

A cada persona que, de una u otra forma, aportó tiempo, conocimiento o ánimo, muchas gracias. Este logro también les pertenece.

Álvaro Moisés De La Rocha Rodríguez

ÍNDICE DE CONTENIDO

DEDICATORIA.....	VIII
AGRADECIMIENTO.....	IX
ÍNDICE DE CONTENIDO.....	X
ÍNDICE DE ILUSTRACIONES.....	XVI
ÍNDICE DE TABLAS	XIX
CAPÍTULO I. PLANTEAMIENTO DE LA INVESTIGACIÓN	1
1.1 INTRODUCCIÓN	1
1.2 ANTECEDENTES DEL PROBLEMA	2
1.3 PLANTEAMIENTO DEL PROBLEMA.....	4
1.3.1 ENUNCIADO DEL PROBLEMA.....	4
1.4 PREGUNTAS DE INVESTIGACIÓN.....	5
1.4.1 Pregunta General	5
1.4.2 Preguntas Específicas	5
1.5 OBJETIVOS DE INVESTIGACIÓN	6
1.5.1. OBJETIVO GENERAL	6
1.5.2. OBJETIVOS ESPECÍFICOS	6
1.6 JUSTIFICACIÓN.....	7
1.6.1 INCERTIDUMBRE EN LA PREDICCIÓN DE RIESGO CREDITICIO.....	8
CAPÍTULO II. MARCO TEÓRICO	10
2.1 ANÁLISIS DEL MACROENTORNO.....	10
2.1.1 Turquía	10
2.1.2 Sudáfrica.....	11
2.1.3 Brasil	11
2.2 ANÁLISIS DEL MICROENTORNO.....	12
2.2.1 Guatemala.....	12
2.2.2 El Salvador	13
2.2.3 Costa Rica	13
2.2.4 Honduras	14
2.3 MARCO CONCEPTUAL.....	16
2.3.1. RIESGO DE CRÉDITO	16

2.3.2.	CALIFICACIÓN DE RIESGO.....	16
2.3.3.	BURÓ DE CRÉDITO.....	16
2.3.4.	PROBABILIDAD DE INCUMPLIMIENTO (PD).....	17
2.3.5.	SERIES DE TIEMPO.....	17
2.3.6.	APRENDIZAJE AUTOMÁTICO.....	17
2.3.7.	MODELO SUPERVISADO.....	18
2.3.8.	MODELO NO SUPERVISADO.....	18
2.3.9.	MATRIZ DE CONFUSIÓN.....	18
2.3.10.	CORRELACIÓN.....	19
2.3.11.	MÉTRICAS DE EVALUACIÓN.....	19
2.4	TEORÍAS DE SUSTENTO.....	19
2.4.1	TEORÍA DE LA INFORMACIÓN ASIMÉTRICA.....	19
2.4.2	TEORÍA DEL APRENDIZAJE ESTADÍSTICO (AE).....	20
2.4.3	TEORÍA DEL RIESGO FINANCIERO.....	21
2.5	ANÁLISIS DE LAS METODOLOGÍAS.....	22
2.5.1	PARADIGMA POSITIVISMO Y POST-POSITIVISMO.....	22
2.6	ANTECEDENTES DE METODOLOGÍAS.....	22
2.7	METODOLOGÍAS Y ESQUEMAS DE DISEÑO.....	23
2.8	ANÁLISIS CRÍTICO DE METODOLOGÍAS.....	25
2.9	HERRAMIENTAS A UTILIZAR.....	27
2.10	MARCO LEGAL.....	31
2.10.1	MARCO LEGAL INTERNACIONAL.....	31
2.10.1.1	Basilea III Comité de Supervisión Bancaria (BCBS).....	31
2.10.1.2	BCBS 239 Principios para la agregación efectiva de datos de riesgo y reportes (Risk Data Aggregation).....	31
2.10.1.3	SR 11-7 Guidance on Model Risk Management (Fed / OCC — EE. UU.).....	31
2.10.1.4	GDPR (Reglamento General de Protección de Datos UE) y Convenios Europeos sobre privacidad.....	32
2.10.1.5	Recomendaciones del GAFI / FATF (Anti-Money Laundering / AML).....	32
2.10.1.6	ISO/IEC 27001 Sistema de Gestión de Seguridad de la Información (SGSI) ...	32
2.10.1.7	COBIT (ISACA) y ITIL (AXELOS) Gobernanza y gestión de TI.....	32

2.10.1.8	Normas y guías sobre “Model Risk” y validación (OCC, FDIC y bancos centrales)	33
2.10.2	MARCO LEGAL NACIONAL	33
2.10.2.1	Comisión Nacional De Bancos Y Seguros (CNBS) / Normas Para Evaluación Y Clasificación De Cartera Crediticia	33
2.10.2.2	Banco Central de Honduras (BCH) - Información Crediticia Y Estabilidad Financiera	34
2.10.2.3	Situación de la legislación sobre protección de datos en Honduras	34
2.10.2.4	Ley de Transparencia y Acceso a la Información Pública (Decreto 170-2006) y decretos sobre administración pública electrónica	35
CAPÍTULO III. METODOLOGÍA		36
3.1.	CONGRUENCIA METODOLÓGICA	36
3.1.1.	MATRIZ METODOLÓGICA	36
3.1.2	ESQUEMA DE VARIABLES DE ESTUDIO	38
3.1.3.	OPERACIONALIZACIÓN DE LAS VARIABLES	38
3.1.3.1	HIPÓTESIS	39
3.2	ENFOQUE Y MÉTODOS	40
3.3	DISEÑO DE LA INVESTIGACIÓN	40
3.3.1	POBLACIÓN	40
3.3.1.1	Criterios De Inclusión Y Exclusión	41
3.3.2	TÉCNICAS DE MUESTREO	42
3.3.3	MUESTRA	43
3.4	TÉCNICAS, INSTRUMENTOS Y PROCEDIMIENTOS APLICADOS	43
3.4.1.	TÉCNICAS	43
3.4.2.	INSTRUMENTOS	44
3.4.3.	PROCEDIMIENTOS	44
3.5.	FUENTES DE INFORMACIÓN	45
3.5.1.	FUENTES PRIMARIAS	45
3.5.2.	FUENTES SECUNDARIAS	46
3.6.	PLAN PARA EL ANÁLISIS DE DATOS	46
3.6.1	PREPARACIÓN DE LOS DATOS	47

3.6.2	ANÁLISIS DESCRIPTIVO Y EXPLORATORIO (EDA)	47
3.6.3	MODELADO PREDICTIVO Y VALIDACIÓN ESTADÍSTICA	48
3.6.4	PRUEBA DE HIPÓTESIS.....	48
3.6.5	INTEGRACIÓN Y SÍNTESIS DE RESULTADOS	50
CAPÍTULO IV. RESULTADOS Y ANÁLISIS.....		51
4.1	ANÁLISIS EXPLORATORIO DE DATOS (EDA)	53
4.1.1	DESCRIPCIÓN GENERAL DEL CONJUNTO DE DATOS	53
4.1.1.1	Tamaño de la Muestra	53
4.1.2	LIMPIEZA Y PREPARACIÓN DE LOS DATOS	54
4.1.2.1	Detección y Manejo de Valores Atípicos.....	54
4.1.2.2	Tratamiento de Valores Faltantes.....	61
4.1.2.3	Normalización Y Estandarización De Variables.....	62
4.1.3	ANÁLISIS Y DESCRIPCIÓN DE VARIABLES.....	63
4.1.3.1	Características Básicas De Los Datos Cuantitativos.....	66
4.1.3.2	Características Básicas De Los Datos Cualitativos.....	73
4.1.4	VISUALIZACIÓN Y EXPLORACIÓN DE LOS DATOS	76
4.1.5	CORRELACIONES ENTRE VARIABLES	78
4.1.6	CONCLUSIONES DEL EDA	80
4.2	INFORME DEL PROCESO DE RECOLECCIÓN DE DATOS	83
4.2.1	DESCRIPCIÓN DEL PROCESO.....	83
4.2.2	PARTICIPANTES O FUENTES DE INFORMACIÓN	86
4.2.3	INSTRUMENTOS UTILIZADOS	88
4.2.4	DIFICULTADES ENCONTRADAS	89
4.2.4.1	Fragmentación y Heterogeneidad de la Información	89
4.2.4.2	Restricciones de Confidencialidad Institucional	90
4.2.5	CONSIDERACIONES ETICAS.....	91
4.2.5.1	Protocolo de Anonimización y Resguardo de la Información	91
4.2.5.2	Uso Responsable de los Datos	92
4.3	RESULTADOS Y ANÁLISIS DE LAS TÉCNICAS APLICADAS.....	93
4.3.1	RESULTADOS CUANTITATIVOS.....	93
4.3.1.1	Presentación de Datos	94

4.3.1.2	Descripción de los Hallazgos	105
4.3.1.3	Relación con los Objetivos.....	108
4.3.1.4	Análisis Estadístico	118
4.3.2	ANÁLISIS CUALITATIVO.....	120
4.3.2.1	Categorías o Temas Emergentes	120
4.3.2.2	Citas o Ejemplos.....	127
4.3.2.3	Interpretación	128
4.3.2.4	Triangulación	130
4.4	ANÁLISIS INFERENCIAL Y MODELOS APLICADOS.....	132
4.4.1	Marco de Validación del Modelo.....	133
4.4.1.1	AUC-ROC (Área Bajo la Curva ROC).....	133
4.4.1.2	Estadístico KS (Kolmogorov-Smirnov).....	134
4.4.1.3	Population Stability Index (PSI)	134
4.4.2	ANÁLISIS INFERENCIAL	134
4.4.2	MODELOS APLICADOS	137
4.4.2.1	Optimización de Hiperparámetros y Eficiencia Computacional.....	137
4.4.2.2	Descripción y Fundamento de los Modelos	137
4.4.3	DISCUSIÓN DE HALLAZGOS	151
4.4.4	LIMITACIONES	154
4.5	SÍNTESIS DE HALLAZGOS	155
4.5.1	PRINCIPALES HALLAZGOS.....	155
4.5.2	IMPLICACIONES.....	157
4.5.3	TRANSICIÓN AL CAPÍTULO V.....	158
CAPÍTULO V. CONCLUSIONES Y RECOMENDACIONES		159
5.1	CONCLUSIONES	159
5.2	RECOMENDACIONES	160
CAPÍTULO VI. APLICABILIDAD.....		162
6.1	NOMBRE DE LA PROPUESTA	162
6.2	JUSTIFICACIÓN DE LA PROPUESTA.....	162
6.3	ALCANCE DE LA PROPUESTA	164
6.3.1	OBJETIVO GENERAL	165

6.3.2	OBJETIVOS ESPECÍFICOS.....	165
6.4	DESCRIPCIÓN Y DESARROLLO	166
6.4.1	DESCRIPCIÓN.....	166
6.4.1.1	Acciones Principales, Procesos y Estrategias (Qué se hará).....	166
6.4.1.2	Justificación Metodológica del Diseño de la Propuesta (Cómo se hará y por qué así)	168
6.4.2	Pipelines Analíticos del Modelo Híbrido	169
6.4.2.1	Ingesta y Preparación de Datos	169
6.4.2.2	Calidad de Datos y Transformación de Variables.....	169
6.4.2.3	Modelado Predictivo y Validación.....	170
6.4.2.4	Scoring, Monitoreo y Gobierno del Modelo	170
6.4.2.5	Flujo Integral del Modelo Híbrido de Riesgo Crediticio	170
6.4.2.6	Alineación de los Pipelines con Buenas Prácticas Regulatorias	171
6.4.3	DESARROLLO	172
6.5	MEDIDAS DE CONTROL	178
6.5.1	Marco de Monitoreo y Validación Continua.....	178
6.5.2	Indicadores	179
6.5.2.1	Indicadores Cuantitativos	179
6.5.2.2	Indicadores Cualitativos	179
6.5.2.3	Fichas Metodológicas de Indicadores	181
6.5.3	Plan de Seguimiento.....	183
6.5.3.1	Ejemplos de Disparadores de Acción.....	186
6.5.4	Análisis de Riesgos Éticos y Gobernanza Algorítmica.....	187
6.5.4.1	Riesgos Éticos Identificados	187
6.5.4.2	Medidas de Mitigación Ética.....	187
6.5.4.3	Gobernanza Algorítmica del Modelo	188
6.5.4.4	Alineación con Principios Regulatorios y Éticos	188
6.6	CRONOGRAMA DE IMPLEMENTACIÓN Y PRESUPUESTO	188
6.7	PRESUPUESTO E IMPACTO DEL PRESUPUESTO.....	190
6.7.1	Presupuesto Estimado por Rubros.....	191
6.7.2	Estimación del Impacto y Retorno de la Inversión (ROI).....	194

6.8 CONCORDANCIA DE LOS SEGMENTOS DE LA TESIS CON LA PROPUESTA

195

REFERENCIAS BIBLIOGRÁFICAS.....	201
ANEXOS.....	206
ANEXO TÉCNICO.....	206
Anexo Técnico A. Código Python de Desempeño Comparado por Modelo y Período (VALID/TEST).....	206
Anexo Técnico B. Código Python de Bootstrap Δ AUC y Δ KS (VALID/TEST).....	207
Anexo Técnico C. Logistic Coeficientes (β) y $OR=\exp(\beta)$ en Python.....	208
Anexo Técnico D. Logistic Hiperparámetros y Eficiencia Computacional en Python.....	209
Anexo Técnico E. Prueba de Chi-cuadrado (χ^2) Aplicada a Variables Categóricas en Python.....	210
Anexo Técnico F. Validación de Supuestos Estadísticos en Python.....	211
Anexo Técnico G. Random Forest Hiperparámetros y Eficiencia Computacional en Python.....	212
Anexo Técnico H. XGBOOST Hiperparámetros y Eficiencia Computacional en Python ..	213
Anexo Técnico I. Top 20 Real vs Predicho en Python.....	214
Anexo Técnico J. Composición del Dataset en Python.....	215
Anexo Técnico K. Valores Atípicos Detectados en Python.....	216
Anexo Técnico L. Eficacia IQR Antes vs Después en Python.....	217
Anexo Técnico M. Diagnóstico de Valores Faltantes en Python.....	218
Anexo Técnico N. Variables Normalizadas y Estandarizadas en Python.....	219
Anexo Técnico O. Análisis Descriptivo de Variables Cuantitativas en Python.....	220

ÍNDICE DE ILUSTRACIONES

Figura 1: Incertidumbre en la predicción de riesgo crediticio	9
Figura 2: Diagrama Metodológico CRISP-DM	25
Figura 3: Esquema de Relación Causal Multivariada	38
Figura 4: Boxplot Monto Desembolsado	56
Figura 5: Boxplot Monto Garantía	56

Figura 6: Boxplot Exposición Total	57
Figura 7: Boxplot Edad	58
Figura 8: Boxplot Buró Equifax.....	59
Figura 9: Boxplot AVG Pagos 6M.....	60
Figura 10: Boxplot AVG Cuentas 6M	60
Figura 11: Grafico de Frecuencias de Variables de Monto Desembolsado, Monto de Garantía y Monto de Exposición Total	68
Figura 12: Gráfico de Frecuencia Cobertura Garantía	69
Figura 13: Gráfico de Frecuencia de Buro Score	70
Figura 14: Gráfico de Frecuencia de AVG Pagos.....	70
Figura 15: Gráfico de Frecuencia de AVG Saldos en Cuentas	71
Figura 16: Gráfico de Frecuencia en Variable de Operaciones	72
Figura 17: Gráfico de Boxplots Variables Cuantitativas	72
Figura 18: Distribución de Variable Tipo de Crédito	73
Figura 19: Gráfico Distribución Tipo de Persona	74
Figura 20: Gráfico de Distribución Sexo	74
Figura 21: Gráfico Distribución Categoría CNBS	75
Figura 22: Gráfico Distribución Sector Económico.....	75
Figura 23: Distribución de Variable Objetivo.....	76
Figura 24: Boxplots Comparativos con Variable Objetivo	77
Figura 25: Matriz de Correlación entre Variables Financieras	78
Figura 26: Proceso ETL de Elaboración y Preparación de Datos “Core Banking”	84
Figura 27: Distribución Monto Desembolsado Post Capping.....	97
Figura 28: Distribución Exposición Total Post Capping	97
Figura 29: Distribución Monto Garantía Post Capping	98
Figura 30: Distribución Monto Garantía Post Capping	99
Figura 31: Distribución Buro Score Equifax Post Capping	99
Figura 32: Distribución Avg Cuentas 6M Post Capping.....	101
Figura 33: Distribución Avg Pagos 6M Post Capping.....	101
Figura 34: Buro Score vs Bueno_Malo.....	102
Figura 35: Cobertura de Garantía vs Bueno_Malo	102

Figura 36: Avg Pagos 6M vs Bueno_Malo.....	103
Figura 37: Monto de Exposición Total vs Bueno_Malo.....	103
Figura 38: Matriz de Correlaciones Post Capping Variables Financieras	104
Figura 39: XGBoost vs Reglas y RF vs Reglas en TEST para ΔAUC y ΔKS y Convergencia ..	110
Figura 40: Curvas ROC comparadas TEST 2024	111
Figura 41: KS por modelo y período.....	112
Figura 42: Código Python Matrices de Confusión.....	114
Figura 43: Matrices de confusión – TEST 2024 (umbral de VALID).....	114
Figura 44: Ahorro Neto vs Reglas.....	115
Figura 45: Estabilidad del score (PSI XGB VALID y TEST)	117
Figura 46: Curva de calibración - TEST 2024	119
Figura 47: Curva Calibración XGB vs RL (TEST 2024).....	150
Figura 48: Flujo Integral del Modelo Híbrido de Riesgo Crediticio.....	171
Figura 49: Código Python de Desempeño Comparado por Modelo y Período (VALID/TEST).....	206
Figura 50: . Código Python de Bootstrap ΔAUC y ΔKS (VALID/TEST)	207
Figura 51: Logistic Coeficientes (β) y $OR = \exp(\beta)$ en Python	208
Figura 52: Logistic Hiperparámetros y Eficiencia Computacional en Python.....	209
Figura 53: Prueba de Chi-cuadrado (χ^2) Aplicada a Variables Categóricas en Python.	210
Figura 54: Validación de Supuestos Estadísticos en Python.....	211
Figura 55: Random Forest Hiperparámetros y Eficiencia Computacional en Python	212
Figura 56: XGBOOST Hiperparámetros y Eficiencia Computacional en Python.....	213
Figura 57: Top 20 Real vs Predicho en Python.....	214
Figura 58: Composición del Dataset en Python	215
Figura 59: Valores Atípicos Detectados en Python	216
Figura 60: Eficacia IQR Antes vs Después en Python.....	217
Figura 61: Diagnóstico de Valores Faltantes en Python	218
Figura 62: Variables Normalizadas y Estandarizadas en Python.....	219
Figura 63: Análisis Descriptivo de Variables Cuantitativas en Python	220

ÍNDICE DE TABLAS

Tabla 1: Comparativo de las Cinco Fuerzas (Guatemala, El Salvador, Costa Rica, Honduras)....	15
Tabla 2: Fases de Metodología CRISP-DM.....	24
Tabla 3: Comparativa de Métodos Relevantes.....	25
Tabla 4: Herramientas Comparativas de Preparación de Datos.....	27
Tabla 5: Comparativo de Herramientas de Análisis Estadístico y Modelado.....	28
Tabla 6: Comparativo de Herramientas de Visualización de datos	29
Tabla 7: Comparativo de Herramientas de Almacenamiento de Información.....	30
Tabla 8: Clasificación de Créditos	34
Tabla 9: Matriz Metodológica.....	36
Tabla 10: Esquema de Variables de Estudio.....	39
Tabla 11: Criterios De Inclusión y Exclusión.....	41
Tabla 12: Esquema de Particionamiento de Datos.....	42
Tabla 13: Instrumentos Documentales.....	44
Tabla 14: Composición del Dataset	54
Tabla 15: Resumen de Valores Atípicos Detectados	55
Tabla 16: Porcentaje de Outliers Antes y Después	61
Tabla 17: Diagnóstico de Valores Faltantes.....	62
Tabla 18: Variables Normalizadas y Estandarizadas	63
Tabla 19: Tipo de Datos	63
Tabla 20: Información de Variables del Data Set	64
Tabla 21: Análisis Descriptivo de Variables Cuantitativas.....	66
Tabla 22: Correlación Bivariado con Variable Objetivo	79
Tabla 23: Cronograma Recolección de Datos.....	85
Tabla 24: Distribución de la población de clientes empresariales	87
Tabla 25: Técnicas de Anonimización y Resguardo Implementadas en la Investigación	92
Tabla 26: Composición del Dataset	94
Tabla 27: Estadísticos descriptivos (post-capping) de Variables Financieras y Conductuales.....	96
Tabla 28: Correlación Univariada con Variable Objetivo Bueno_Malo.....	105
Tabla 29: Desempeño comparado por Modelo y Período.....	108
Tabla 30: Contraste (bootstrap) Δ AUC y Δ KS	109

Tabla 31: Impacto operativo en TEST (umbral VALID).....	113
Tabla 32: Monetización del Trade - Off (Test 2024, umbral Valid).....	116
Tabla 33: Prueba de Chi-cuadrado (χ^2) Variables Categóricas vs Bueno-Malo	124
Tabla 34: Verificación del supuesto de frecuencias esperadas en la prueba Chi-cuadrado (χ^2) ..	125
Tabla 35: Síntesis de observaciones y análisis interpretativo	127
Tabla 36: Vinculación entre resultados cualitativos y teorías de sustento	128
Tabla 37: Interpretación teórica de las categorías emergentes.....	129
Tabla 38: Triangulación de Resultados Cuantitativos y Cualitativos	131
Tabla 39: Logistic Coeficientes (β) y $OR=\exp(\beta)$	138
Tabla 40: Tuning(RL). Optimización de Hiperparámetros (VALID 2023).....	140
Tabla 41: Eff(RL). Eficiencia Computacional	141
Tabla 42: Top 20 Real vs Predicho - RL (TEST, umbral VALID).....	141
Tabla 43: Tuning(RF). Optimización de hiperparámetros (VALID 2023).....	143
Tabla 44: Eff(RF). Eficiencia computacional	144
Tabla 45: Top 20 Real vs Predicho - RandomForest (TEST, umbral VALID)	145
Tabla 46: Tuning(XGB). Optimización de hiperparámetros (VALID 2023)	147
Tabla 47: Eff(XGB). Eficiencia Computacional.....	147
Tabla 48: Top 20 Real vs Predicho - XGBoost (TEST, umbral VALID).....	148
Tabla 49: Límites, Inclusiones y Exclusiones del Modelo Híbrido de Riesgo Crediticio Empresarial.....	164
Tabla 50: Dimensiones Geográfica, Poblacional y Temporal.....	165
Tabla 51: Indicadores Cuantitativos para la Evaluación del Modelo Híbrido de Riesgo Crediticio Empresarial.....	179
Tabla 52: Indicadores cualitativos.....	180
Tabla 53: Ficha Metodológica del Indicador AUC - ROC del Modelo Híbrido.....	181
Tabla 54: Ficha Metodológica del Indicador Reducción de Morosidad (%)	181
Tabla 55: Ficha metodológica del indicador Satisfacción del Personal Analista	182
Tabla 56: Etapas, Responsables y Frecuencia de Ejecución del Plan de Seguimiento	184
Tabla 57: Plan de seguimiento y Control	185
Tabla 58: Cronograma PERT de implementación del modelo híbrido	190
Tabla 59: Presupuesto PERT detallado por sub-rubros (valores en HNL y USD)	192

Tabla 60: Concordancia de los Segmentos de Tesis197

CAPÍTULO I. PLANTEAMIENTO DE LA INVESTIGACIÓN

1.1 INTRODUCCIÓN

En esta investigación, se elabora un modelo para la Calificación del Riesgo Crediticio, dirigido a clientes de la banca empresarial en Honduras con el objetivo de mejorar la precisión en la evaluación del riesgo de incumplimiento y la capacidad crediticia. Las metodologías tradicionales empleadas en la evaluación de riesgos financieros suelen presentar falencias, eso afecta las decisiones en las instituciones financieras. Con el fin de enfrentar esta problemática, el estudio propone la implementación de machine learning como herramienta principal para optimizar la clasificación de los clientes según su perfil de riesgo; la cual es la esencia del trabajo.

En el área financiera, la evaluación del riesgo crediticio representa un procedimiento esencial para determinar la solvencia de los clientes y mitigar posibles incumplimientos. En el periodo comprendido entre abril de 2023 y abril de 2024, la cartera crediticia del sistema financiero hondureño registró un crecimiento, al pasar de L 643.6 mil millones en 2023 a L 654.2 mil millones en 2024, lo que representa un incremento interanual de 16.9% (Gerencia de Estudios Económicos, CNBS 2024, 2024).

A pesar de ello, el vigor de la expansión crediticia suele traer consigo una subida en los niveles de riesgo. El indicador de morosidad en la cartera es un elemento fundamental para evaluar la calidad y riesgo del sistema financiero, esto cuantifica los préstamos que no se están cumpliendo.

Los informes oficiales de la (Gerencia de Estudios Económicos, CNBS 2024, 2024), presentan una reducción de 0.1 puntos porcentuales (p.p.) en la tasa de morosidad durante el período 2023-2024. Sin embargo, esta tendencia se modificó en 2025, año en que el índice experimentó un aumento hasta el 2.6%, equivalente a 0.5 (p.p.). Lo anterior sugiere una mayor susceptibilidad de la cartera ante cambios macroeconómicos. Esto también podría indicar una deficiencia en las políticas crediticias o en la solvencia de los deudores.

Este comportamiento evidencia la necesidad crucial de fortalecer los procesos evaluativos crediticios mediante herramientas más flexibles, meticulosas y ajustables. En este ámbito, esta pesquisa se concentra en la edificación de un modelo híbrido para la estimación del riesgo, entrelazando criterios de decisión crediticios con técnicas de aprendizaje automático. El objetivo, que es refinar la precisión en la tipificación de clientes, renovar los métodos actuales, aportando al

desarrollo del sector bancario, ofreciendo un esquema reproducible para otras entidades financieras. La investigación se desarrolla de la siguiente forma:

El Capítulo I describe la problemática principal del estudio y establece los objetivos que guían la investigación, incluyendo un objetivo general y varios objetivos específicos. Además, presenta la justificación del trabajo desde puntos de vista académicos y económicos, apoyada en un análisis completo de la causa principal que produce el fenómeno investigado.

El Capítulo II presenta el marco teórico que apoya la investigación, destacando los indicadores financieros y su impacto en el sistema financiero. Asimismo, analiza el contexto actual y establece los principios fundamentales para la formulación de la hipótesis, mientras que define la metodología utilizada en el estudio.

Capítulo III se ocupa del enfoque, el diseño y los métodos aplicados en la investigación, especificando la población, la muestra, así como las técnicas e instrumentos de recolección de datos que serán implementados.

Capítulo IV se presentarán los diferentes resultados y análisis relacionados con los objetivos presentados en secuencia lógica de acuerdo con la propuesta en las variables y el enfoque de la investigación.

El Capítulo V resume los resultados logrados para cada una de las variables y responde a las preguntas de investigación basándose en los hallazgos del estudio. Igualmente, presenta las recomendaciones que surgen de dichos resultados.

Capítulo VI se presenta el producto final y como se debe de aplicar a corto o largo plazo, estructura y valor agregado al trabajo realizado con los resultados obtenidos.

1.2 ANTECEDENTES DEL PROBLEMA

Las entidades financieras desempeñan un rol crucial en la economía al simplificar el acceso al crédito y manejar el riesgo financiero. Para examinar la capacidad de pago de las compañías, se sirven de calificaciones de riesgo basadas en historiales crediticios. Estos ratings hacen posible determinar la solvencia corporativa, además de mitigar el riesgo de impago, así, se mejora la toma de decisiones durante el otorgamiento de créditos. De acuerdo con reportes de (Solucion Colombia, 2024), las empresas con calificaciones crediticias elevadas acceden a tasas de interés, que son hasta un 50% menores respecto a las que exhiben bajas calificaciones, lo cual disminuye sus gastos

financieros y expande sus opciones de financiación.

Diversos estudios han sido realizados con el objetivo de perfeccionar la exactitud en la predicción de incumplimientos crediticios. En consonancia con el artículo “Mejora del riesgo crediticio”, publicado en 2025, la administración del riesgo se ha erigido como un elemento esencial para las instituciones financieras que aspiran a atenuar los riesgos vinculados a los préstamos. Los métodos empleados hasta el presente se han fundamentado en el conocimiento de expertos y en datos financieros pretéritos; no obstante, resultan insuficientes debido al vasto volumen de información disponible (Nguyen et al., 2025).

Por su parte, el artículo denominado “Modelo de aprendizaje profundo basado en riesgos financieros”, nos menciona que los métodos cuantitativos tradicionales suelen enfocarse en métricas de reducción máxima que evalúan capacidades de control de riesgos, lo que limita su capacidad de anticiparse a riesgos y tomar decisiones estratégicas basadas en datos precisos. Las instituciones financieras que no usan modelos avanzados suelen usar métricas simplistas y con reglas estáticas que no pueden adaptarse a condiciones imprevistas como crisis económicas o caídas en mercados de valores. Al no contar con la capacidad de prever posibles escenarios adversos pueden derivar en pérdidas significativas quedando en desventaja frente a otras instituciones (Yang et al., 2024).

Debido a estas limitantes, la investigación internacional ha progresado en la creación de modelos basados en inteligencia artificial (IA) y machine learning (ML) con el propósito de perfeccionar la exactitud en riesgos de crédito. Investigaciones recientes han revelado que modelos avanzados como XGBoost y LightGBM, demuestran un rendimiento superior a las metodologías convencionales, logrando una precisión del 81% y un recall del 82%, debido a su habilidad para discernir conexiones no evidentes entre las variables financieras de los clientes (Nguyen et al., 2025).

Un ejemplo destacado es el modelo de evaluación crediticia establecido por HDFC Bank, el cual, a través de la aplicación de IA y ML, consiguió una reducción del 20% en sus activos improductivos (NPA) y una disminución del 30% en el período de aprobación de créditos (Kumar Jain, 2019).

En Honduras, el sector bancario empresarial afronta desafíos importantes, puesto que algunas instituciones demuestran una alta probabilidad de incumplimiento (PD). Por ejemplo, en

el año 2022, BANTRAB registró un 0.513 y Banco Popular un 0.165, lo cual resalta la susceptibilidad del sistema y la necesidad de implementar procedimientos de evaluación más rigurosos (*Martini, 2025*).

Tener una calificación crediticia sólida es una ventaja para las empresas. Un puntaje elevado no solo facilita el acceso a financiamiento, sino que también resulta clave para el crecimiento. Mantener un buen historial crediticio abre oportunidades y permite ahorrar dinero (Solunion Colombia, 2024).

La implementación de aprendizaje automático en el sistema financiero de Honduras permanece restringida debido a insuficiencias técnicas, escasa inversión, una gobernanza de datos deficiente, infraestructura inadecuada y la ausencia de un marco regulatorio. Estos elementos evidencian la importancia de incorporar el aprendizaje automático en la banca empresarial con el fin de optimizar la exactitud de las evaluaciones de riesgo y reforzar la administración crediticia.

1.3 PLANTEAMIENTO DEL PROBLEMA

1.3.1 ENUNCIADO DEL PROBLEMA

En el actual entorno de la economía global, numerosas instituciones bancarias experimentan condiciones de marcada volatilidad e incertidumbre. Estas se ven influenciadas por factores como la inflación, las recesiones económicas y las fluctuaciones en los mercados internacionales. Ante este escenario dinámico, la evaluación del riesgo crediticio se considera un proceso fundamental para asegurar la estabilidad y la sostenibilidad. Esto es particularmente relevante en el sector empresarial, donde los montos de crédito suelen ser significativos y la exposición al riesgo es considerable.

A pesar de disponer de sistemas de información crediticia y de informes regulatorios que facilitan el análisis financiero, diversas instituciones en Honduras persisten en el uso de metodologías convencionales. Estas se basan principalmente en modelos estadísticos lineales, reglas fijas y juicios de expertos. Tales enfoques presentan limitaciones en su capacidad para adaptarse a las condiciones cambiantes del mercado y a la diversidad de perfiles empresariales (World Bank Group, 2019). Por consiguiente, la exactitud en la predicción del incumplimiento se ve comprometida.

Diversos estudios recientes recalcan la necesidad de desarrollar enfoques más avanzados, capaces de incorporar fuentes de datos convencionales con alternativas, con el propósito de reforzar las señales predictivas y optimizar la capacidad explicativa de los modelos. Dichas investigaciones enfatizan la importancia de emplear técnicas de ML que, además de mejorar la precisión, provea mecanismos de interpretabilidad y gobernanza en su aplicación dentro de contextos financieros (Chang et al., 2024).

A pesar del éxito comprobado de los modelos de aprendizaje automático en la predicción del riesgo crediticio, persiste en la banca empresarial hondureña una carencia de entendimiento con relación a las variables financieras y no financieras que determinan la morosidad, así como en la integración de estas técnicas con los criterios regulatorios y de decisión crediticia. Adicionalmente, no se ha demostrado el impacto de su implementación en la gestión del riesgo del sistema financiero nacional.

Esta carencia de conocimiento especializado puede ocasionar consecuencias prácticas adversas, puesto que, al carecer de una comprensión nítida de los elementos que influyen en el incumplimiento, se corre el riesgo de conceder créditos a clientes con un perfil de pago débil, lo que resulta en un incremento del riesgo de la cartera, ineficacias operativas y una capacidad reducida para anticipar situaciones desfavorables.

1.4 PREGUNTAS DE INVESTIGACIÓN

Para abordar la problemática relacionada con la evaluación del riesgo crediticio en la banca empresarial de Honduras y el desarrollo de un Modelo de Calificación de Riesgo Crediticio basado en Machine Learning, se plantean las siguientes preguntas de investigación:

1.4.1 PREGUNTA GENERAL

¿En qué medida la implementación de un modelo híbrido que combina aprendizaje automático y reglas de negocio (I) mejora la exactitud en la evaluación del riesgo crediticio (O) de los clientes de la banca empresarial hondureña (P) en comparación con los métodos convencionales basados únicamente en reglas de negocio (C)?

1.4.2 PREGUNTAS ESPECÍFICAS

1. ¿En qué medida la incorporación de variables macroeconómicas externas (I) mejora la precisión predictiva del riesgo crediticio (O) de los clientes empresariales

hondureños (P) en comparación con los modelos que utilizan exclusivamente variables internas del cliente (C)?

2. ¿Cómo la integración de técnicas de aprendizaje automático en los procesos de calificación crediticia (I) mejora la capacidad de anticipar incumplimientos y riesgos emergentes (O) en la banca empresarial hondureña (P) frente a los modelos convencionales basados solo en reglas de negocio (C)?
3. ¿Qué impacto tiene la aplicación de un modelo híbrido que combina datos internos financieros y variables no financieras (I) en la coherencia y calidad de las decisiones crediticias (O) de los clientes de la banca empresarial hondureña (P) frente a las metodologías tradicionales centradas únicamente en información contable (C)?

1.5 OBJETIVOS DE INVESTIGACIÓN

1.5.1. OBJETIVO GENERAL

Evaluar (S) el impacto de un modelo híbrido que combina aprendizaje automático y reglas de negocio sobre la precisión en la evaluación del riesgo crediticio de clientes de la banca empresarial en Honduras, (M) midiendo métricas de desempeño como precisión, recall y AUC-ROC, (A) utilizando datos históricos y algoritmos supervisados, (R) para establecer un modelo validado que optimice la clasificación de riesgo, (T) durante la fase de análisis y modelado del estudio.

1.5.2. OBJETIVOS ESPECÍFICOS

1. (S) Comparar el (M) desempeño, medido por AUC-ROC y precisión, (A) entre un modelo que integra variables internas y externas y otro que utiliza solo variables internas, (R) para evaluar la relevancia de factores externos en la predicción de morosidad, (T) durante la fase de análisis exploratorio.
2. Evaluar (S) la influencia del componente de aprendizaje automático dentro del modelo híbrido, (M) cuantificando métricas como recall, F1-score y precisión, (A) aplicando algoritmos supervisados sobre los datos históricos, (R) para mejorar la capacidad de anticipar incumplimientos y riesgos emergentes, (T) durante la fase de modelado y validación.
3. Determinar (S) el impacto de la implementación del modelo híbrido sobre la coherencia y consistencia de la evaluación crediticia, (M) midiendo la concordancia de los resultados

simulados con registros históricos validados, (A) mediante simulaciones de escenarios crediticios, (R) para optimizar la calidad y confiabilidad del análisis financiero, (T) durante la fase de pruebas y simulación.

1.6 JUSTIFICACIÓN

La evaluación del riesgo crediticio representa un desafío considerable para la banca empresarial en Honduras, debido al deterioro de las carteras, causado por el aumento de los impagos y la limitada eficacia de los modelos convencionales para anticipar el comportamiento financiero de los clientes.

Ante esta circunstancia, la implementación de un modelo de aprendizaje automático se vuelve esencial, pues dotará a las instituciones financieras de la habilidad para tomar decisiones más informadas, mejorar la distribución del crédito y, simultáneamente, promover la inclusión financiera de empresas con un alto potencial de crecimiento.

Desde una perspectiva práctica el principal desafío de esta investigación es la imprecisión en la evaluación del riesgo crediticio. Este factor puede generar pérdidas financieras considerables para las instituciones bancarias y afectar la estabilidad del sector. La implementación de un modelo de aprendizaje automático (ML) ofrece soluciones específicas:

1. Mayor precisión y reducción de morosidad: Al reconocer patrones complejos que los modelos convencionales no identifican, se reduce la probabilidad de conceder créditos a clientes con una elevada tendencia al incumplimiento.
2. Optimización de los procesos de evaluación crediticia: La automatización mediante ML acelerará la evaluación del riesgo crediticio, reduciendo los tiempos de respuesta y optimizando los recursos operativos.
3. Cumplimiento de normativas y regulación financiera: La Comisión Nacional de Bancos y Seguros (CNBS) y el Banco Central de Honduras (BCH) han enfatizado la necesidad de fortalecer los mecanismos de evaluación crediticia. Este estudio propone un modelo alineado con las regulaciones vigentes, permitiendo una gestión de riesgo más eficiente y adaptable.

El presente estudio contribuye a la literatura existente sobre la gestión de riesgos en mercados emergentes, proporcionando evidencia empírica sobre la efectividad de los modelos predictivos en el contexto hondureño. Aunque los sistemas de evaluación de riesgos son

reconocidos mundialmente, su aplicación en el país es limitada.

Esta investigación, incorpora variables internas y externas de clientes empresariales, lo que permite analizar la importancia relativa de factores macroeconómicos y microeconómicos en la predicción del riesgo crediticio. De este modo, se enriquece el fundamento teórico de la evaluación de crédito inteligente y se produce conocimiento aplicable al análisis financiero en mercados con atributos parecidos.

La repercusión de una gestión del riesgo crediticio más sólida abarca también los ámbitos social y económico. Entre los principales beneficios se identifican los siguientes:

1. Instituciones financieras: Optimiza la asignación de crédito y disminuye la morosidad, lo cual fortalece su rentabilidad y perdurabilidad.
2. Empresas con alto potencial: Facilita un acceso responsable al financiamiento, impulsando su crecimiento y consolidación.
3. Economía Nacional: Aporta a la estabilidad financiera y al progreso económico al mitigar las consecuencias de los créditos irrecuperables.

Esta propuesta garantiza una banca responsable, transparente y fiable, construyendo la confianza del usuario financiero y promoviendo un desarrollo económico firme.

1.6.1 INCERTIDUMBRE EN LA PREDICCIÓN DE RIESGO CREDITICIO

Para examinar las causas de la incertidumbre en la evaluación del riesgo crediticio, se empleó el diagrama de Ishikawa, también conocido como diagrama de causa-efecto. Esta herramienta posibilitó la identificación de seis categorías fundamentales: modelos tradicionales, calidad de los datos, factores humanos, gobernanza de datos, procesos y tecnología. El estudio de estas categorías puso de manifiesto las deficiencias estructurales, operativas y tecnológicas que actualmente restringen la eficacia de los sistemas de evaluación crediticia.

Este diagnóstico establece una base firme para el desarrollo del modelo propuesto, cuyo objetivo es reducir las deficiencias detectadas y optimizar la eficiencia, la fiabilidad y la toma de decisiones en la administración del riesgo financiero.

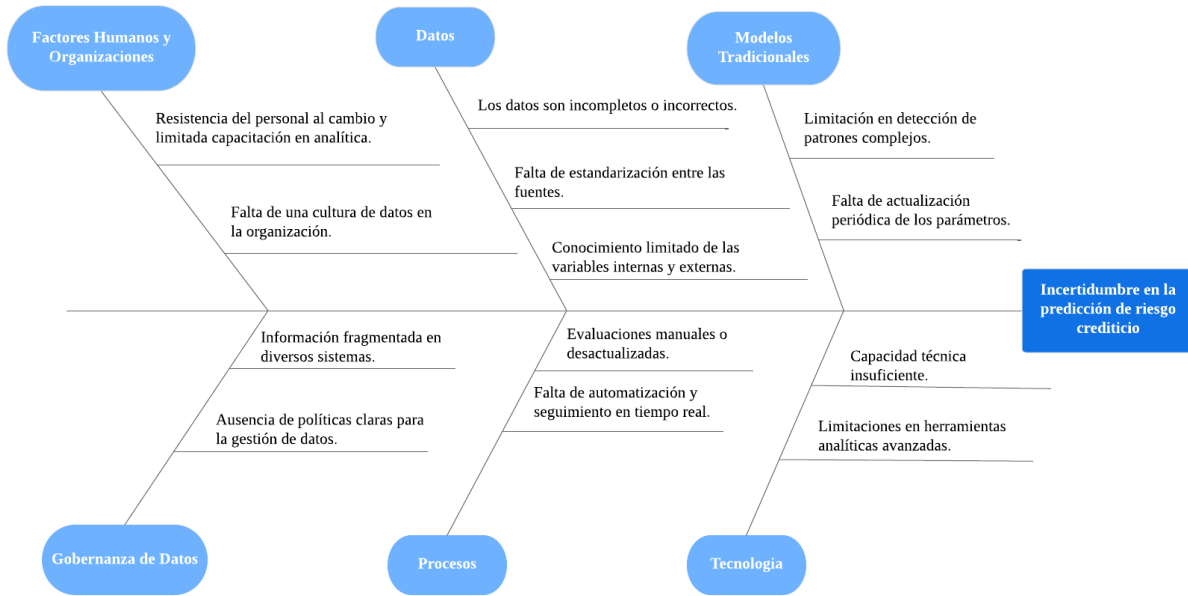


Figura 1: Incertidumbre en la predicción de riesgo crediticio

Fuente: Elaboración propia

CAPÍTULO II. MARCO TEÓRICO

2.1 ANÁLISIS DEL MACROENTORNO

La comparación internacional de entornos políticos, económicos, sociales, tecnológicos, ecológicos y legales en tres economías emergentes fuera de Centroamérica, en este caso específico Turquía, Sudáfrica y Brasil aportan lecciones directamente aplicables a la gestión del riesgo crediticio empresarial en Honduras.

En los tres casos convergen choques macro como inflación, tasas y tipo de cambio, vulnerabilidades operativas como energía, logística, commodities, exigencias prudenciales y adopción gradual de analítica avanzada, todos con efectos trazables sobre morosidad, PD y estabilidad del score.

2.1.1 TURQUÍA

En el contexto político, legal y de política económica, Turquía tras un período 2021 - 2023 de inflación muy elevada y tasas reales negativas, Turquía ejecutó un giro de política con endurecimiento monetario, simplificación macro prudencial para reducir desbalances y anclar expectativas. El Banco Central de la República de Turquía (CBRT) elevó la tasa de política de 8.5% en mayo 2023 a 50% en marzo 2024, situando la tasa real en terreno positivo (CBRT Annual Report 2024, 2024).

Se observó enfriamiento de la demanda, caída del crecimiento del crédito real y descenso paulatino de la inflación desde un pico de 75% en mayo 2024 hacia 52% en agosto 2024, con señales de desinflación subyacente, aunque aún por encima de meta (CBRT Annual Report 2024, 2024).

En los factores económicos críticos, la transmisión de choques de tipo de cambio, incrementos salariales mínimos, impuestos y combustibles a inflación y morosidad se documenta como canales relevantes; caps regulatorios al crecimiento del crédito y recomposición hacia mitigaron de riesgos, mientras la dolarización de balances y exposición en divisas del sector corporativo siguen siendo variables sensibles que monitorear (Karahana, 2024).

Implicaciones para riesgo crediticio: En carteras empresariales, episodios de depreciación y alta inflación elevan PD y NPL con rezagos, variables externas con alto poder explicativo, inflación mensual y expectativas, nivel y volatilidad del tipo de cambio, tasa de política

y spreads, además de indicadores de actividad (PIB trimestral, PMI) y cuenta corriente (CBRT Annual Report 2024, 2024; Karahan, 2024).

2.1.2 SUDÁFRICA

En el entorno político y legal, tras las elecciones de 2024, se conformó un Gobierno de Unidad Nacional; se promulgó la "Expropriation Act 2024" que elevó la incertidumbre sobre derechos de propiedad, en coexistencia con fortalecimiento de prevención del lavado de activos y financiamiento del terrorismo tras la inclusión en lista gris del FATF en 2023 y la posterior agenda de remoción (Investment Climate Statement: South África, 2025; South African Reserve Bank, 2024).

Economía y riesgos operativos: El SARB (South African Reserve Bank) reporta un sistema financiero resiliente, pero advierte riesgos clave: deterioro de razón deuda pública y PIB, salidas de capital con menor profundidad de mercado, distrés creciente en hogares y pymes, y fallas en infraestructura crítica (energía, agua, logística). Si bien se observó una mejora en 2024, persisten vulnerabilidades (South African Reserve Bank, 2024, 2025). El aumento de NPL en pymes superó el promedio, forzando mayores provisiones y presionando cobertura (South African Reserve Bank, 2024).

Implicaciones para riesgo crediticio: Variables exógenas de energía (índices de disponibilidad eléctrica y EAF, días/GWh shed), indicadores de deuda soberana (rendimientos SAGB, spreads), shocks de flujos (participación no residentes en bonos), y métricas de profundización de mercado impactan PD, especialmente en pymes intensivas en energía o logística. Existe evidencia de que shocks operativos (energía) se traducen en stress de flujo de caja y aumentos de incumplimientos con rezago (South African Reserve Bank, 2025).

2.1.3 BRASIL

En el entorno económico y monetario, el Banco Central de Brasil reporta actividad robusta en 2024 con inflación por encima de meta y expectativas para 2025 - 2026 en ascenso. La tasa SELIC se mantuvo elevada, alrededor de 15% a fines de 2025 en algunas lecturas de mercado para reconducir la inflación hacia el objetivo; el Focus proyectó inflación 2025 en 5.51%, por encima del techo del rango (Banco Central do Brasil, 2024; Brazil Trading Economics - BCB Selic Rate, 2024; Economía Focus/Expectativas, 2025). En EMEs, expectativas sugieren relajación monetaria gradual hacia 2025-26, con cautela (Banco Central do Brasil, 2024).

La estructura sectorial (comercio, servicios, agronegocio) y exposición a commodities introducen heterogeneidad en PD's; Choques de precios (energía y agrícolas) y tasas impactan rotación de capital de trabajo y servicio de deuda, especialmente en pymes. El uso extendido de ML en scoring en Brasil ha mostrado ganancias de AUC - KS respecto a lineales, con enfoque en calibración para PD y provisiones (Chang et al., 2024; Lessmann et al., 2015).

2.2 ANÁLISIS DEL MICROENTORNO

El microentorno competitivo del sistema bancario empresarial condiciona la originación, los precios y las garantías del crédito, impactando directamente la probabilidad de incumplimiento (PD), la pérdida dada incumplimiento (LGD) y la exposición (EAD).

En economías centroamericanas, la rivalidad, el poder de negociación de las empresas (compradores), el poder de los proveedores de fondeo y capital, la amenaza de nuevos entrantes (incluyendo fintech, plataformas) y la disponibilidad de sustitutos (cooperativas, OPDF, factoring y crédito de proveedores) determinan los incentivos de los bancos para asumir o mitigar riesgo.

Analizar Guatemala, El Salvador y Costa Rica ofrece un marco comparativo útil para interpretar Honduras y para fundamentar el diseño del modelo híbrido propuesto, umbrales de cobertura de garantías por ticket y sector, límites de concentración, segmentación por subcarteras y selección de variables explicativas (Porter, 2008; BCBS, 2010; World Bank, 2019).

2.2.1 GUATEMALA

- Rivalidad entre competidores (Alta): Mercado con grupos financieros regionales compitiendo por crédito comercial en comercio, industria y construcción, rivalidad en precios, plazos y retención de corporativos multibanca (Banco de Guatemala, 2024; SIB Guatemala, 2024).
- Poder de negociación de compradores (Medio–Alto): Conglomerados y grandes corporativos negocian tasas, covenants, garantías; en pymes, el poder depende de colateral y formalización (SIB Guatemala, 2024).
- Poder de proveedores (fondeo/capital) (Medio): Fondeo predominantemente por depósitos; menor profundidad de mercado de capitales eleva el costo de diversificación; disciplina prudencial incide en pricing (Banco de Guatemala, 2024).
- Amenaza de nuevos entrantes (Baja-Media): Barreras regulatorias y de capital altas. Fintech en pagos y originación pyme crecen, pero su peso en crédito empresarial aún es moderado (Banco

Interamericano de Desarrollo (BID), 2023; SIB Guatemala, 2024).

- Amenaza de sustitutos (Media-Alta): Cooperativas, financieras, factoring y crédito proveedor compiten especialmente en pymes, presionando márgenes y condiciones (World Bank Group, 2019).

2.2.2 EL SALVADOR

- Rivalidad (Media): Presencia de grupos regionales; la dolarización reduce dispersión cambiaria y facilita comparabilidad de tasas, pero el tamaño del mercado modera la rivalidad en corporativo (BCR, 2024).
- Poder de compradores (Medio): Pymes con mayor acceso por digitalización, aunque mantienen restricciones de garantías; corporativos medianos/grandes negocian por multibanca y contratos (BCR, 2024).
- Poder de proveedores (Media-Alta): Dolarización implica sensibilidad a condiciones de mercado USD (tasa base y spreads); costo de capital y depósitos condiciona apetito de riesgo (BCR, 2024).
- Entrantes (Media): Fintech en pagos y remesas avanzan; la respuesta bancaria (digitalización) ha contenido la disrupción en crédito empresarial (Banco Interamericano de Desarrollo (BID), 2023).
- Sustitutos (Media): Cooperativas/financieras y crédito proveedor para pymes; en corporativos, líneas internacionales y financiamiento de cadenas son sustitutos parciales (World Bank, 2019).

2.2.3 COSTA RICA

- Rivalidad (Alta): Banca pública (BNCR, BCR) y banca privada regional compiten con intensidad por corporativo, exportador y pymes de servicios y tech; Sofisticación de productos por apertura y encadenamientos (Banco Central de Costa Rica, 2024; SUGEF Costa Rica, 2024).
- Poder de compradores (Alto en corporativo y Medio en pymes): Exportadores y multinacionales negocian condiciones por acceso a mercados; pymes formalizadas ganan acceso con brechas de garantías persistentes (Banco Central de Costa Rica, 2024).
- Poder de proveedores (Media): Banca pública da estabilidad de fondeo; banca privada depende

de depósitos/emisiones; regulación SUGEF incide en provisiones y capital (SUGEF Costa Rica, 2024).

- Entrantes (Media): Fintech y open finance y ABIs emergen y el impacto en crédito empresarial es moderado por barreras y respuesta bancaria (Banco Interamericano de Desarrollo (BID), 2023).
- Sustitutos (Media-Alta): Cooperativas de ahorro y crédito con alta penetración; factoring y financiamiento de cadenas es relevante para pymes exportadoras (World Bank, 2019).

2.2.4 HONDURAS

- Rivalidad (Alta): El sistema financiero nacional continúa su expansión constante, tanto en activos como en la cartera de crédito. Esto claramente indica un mercado robusto y con fuerte competencia. Para abril de 2024, los activos del sistema financiero se dispararon, llegando a L1.0 billón, la cartera crediticia experimentó un aumento del 16.9% interanual (Gerencia de Estudios Económicos, CNBS 2024, 2024). Los bancos luchan, compiten a tope, por conseguir clientes corporativos. Esto es, en sectores fundamentales como: Propiedad raíz, comercio, industria, y servicios. Aquellos clientes representan, más del 80% del portafolio del SBC. (Gerencia de Estudios Económicos, CNBS 2024, 2024).
- Poder de Compradores (Alta): Las empresas bancarizadas tienen múltiples opciones de financiamiento debido a crecimiento de depósitos y liquidez en el sistema, exceso de liquidez bancaria de L24.9 mil millones en 2024, prioridad regulatoria en apoyo a sectores productivos (agro, construcción, comercio) (Gerencia de Estudios Económicos, CNBS 2024, 2024). Esto otorga a clientes corporativos mayor capacidad para: Exigir mejores tasas, negociar plazos y garantías, refinanciar o cambiar de institución.
- Poder de Proveedores (Medio-Alta): Los proveedores clave de la banca son los depositantes, los inversionistas en capital, y los reguladores quienes permiten las condiciones de operación. El índice de solvencia del SBC, alcanzando un notable 13.4%, más alto que el requisito mínimo, esto implica que el sector se basa enormemente en capital reforzado, impulsado por inversiones internas, y los accionistas también. (Gerencia de Estudios Económicos, CNBS 2024, 2024).
- Entrantes (Media): El sector presenta barreras altas como ser requerimientos severos de solvencia y cobertura de mora, supervisión y regulaciones estrictas de la CNBS, necesidad de

infraestructura tecnológica y red comercial amplia. (Gerencia de Estudios Económicos, CNBS 2024, 2024). No obstante, existen presiones de nuevos modelos tales como Fintech, plataformas tecnológicas de pagos y crédito y regulación reforzada en ciberseguridad.

- Sustitutos (Media-Alta): Las empresas pueden buscar alternativas a la banca tradicional, tales como proveedores de financiamiento comercial, cooperativas financieras, OPDF con cartera en expansión, pero más riesgosa (mora 7.7%). (Gerencia de Estudios Económicos, CNBS 2024, 2024), fondos privados e inversionistas institucionales y créditos externos y remesas reinvertidas en negocios.

Tabla 1: Comparativo de las Cinco Fuerzas (Guatemala, El Salvador, Costa Rica, Honduras)

País	Rivalidad Entre Competidores	Poder de Negociación de Compradores	Poder de Negociación de Proveedores	Amenaza de nuevos Entrantes	Amenaza de Sustitutos
Guatemala	Alta: grupos regionales compiten por corporativo y pymes.	Media-Alta: Multibanca corporativa, pymes dependen de colateral.	Media: Depósitos locales, capital y provisiones inciden en pricing.	Baja-Media: Barreras regulatorias y fintech en ascenso.	Media-Alta: Cooperativas, financieras, crédito proveedor y factoring.
El Salvador	Media: Mercado moderado, dolarización reduce dispersión de tasas.	Media: Corporativos negocian y pymes con acceso creciente.	Media-Alta: Sensibilidad a condiciones USD.	Media: Fintech pagos y remesas, regulación gradual.	Media: Cooperativas y líneas internacionales.
Costa Rica	Alta: Banca pública y privada, foco corporativo y exportador.	Alta: Corporativo Media: Pymes	Media: Estabilidad por banca pública y regulación SUGEF.	Media: Open finance incipiente, respuesta bancaria.	Media-Alta: Cooperativas desarrolladas, cadenas y exportación.
Honduras	Alta: Crecimiento y competencia intensa, cartera con alto peso de comercio y servicios.	Alta: Mayor capacidad para negociar plazos, tasas, garantías.	Media-Alta: Solvencia sistémica sólida y regulación prudencial CNBS.	Media: Fintech y OPDF en crecimiento, lineamientos de ciberseguridad.	Media-Alta: OPDF y cooperativas con peso en pymes.

Fuente: Elaboración Propia

Los tres países muestran rivalidad elevada (más intensa en Guatemala y Costa Rica), creciente presión de sustitutos en pymes y poder de negociación alto en corporativos multibanca; el poder de proveedores es medio y depende de la estructura de fondeo y los requisitos prudenciales; la amenaza de entrantes se materializa más en canales (fintech y plataformas) que en reemplazo pleno del crédito empresarial.

Para Honduras, que ya presenta alta rivalidad, fuerte peso de comercio y servicios y

sustitutos activos (OPDF con mora superior a la banca), estas fuerzas elevan el riesgo de underwriting laxo en fases de crecimiento si no hay salvaguardas.

2.3 MARCO CONCEPTUAL

Esta sección delinearé los cimientos teóricos importantes del modelo híbrido, para la evaluación del riesgo crediticio propuesto. Será segmentada en tres áreas medulares: Conceptos básicos de riesgo crediticio, entendimiento del aprendizaje automático, y técnicas de evaluación.

2.3.1. RIESGO DE CRÉDITO

Se define como la probabilidad de que una de las partes involucradas en un contrato no cumpla con las obligaciones estipuladas en los términos acordados. Dicho riesgo también se conoce como riesgo de incumplimiento, riesgo de rendimiento o riesgo de contraparte (Brown & Moles, 2014).

En esta investigación, el riesgo se cuantificará utilizando la probabilidad de incumplimiento (PD), la cual será calculada a través de modelos predictivos que asignan un valor de probabilidad de impago a cada cliente. La PD constituye el principal indicador operativos del riesgo de crédito de este estudio.

2.3.2. CALIFICACIÓN DE RIESGO

Es la opinión independiente, objetiva y técnicamente fundamenta, acerca de la solvencia y seguridad de un determinado instrumento financiero emitido por alguna institución (Tapia, 2022).

Entre las compañías de mayor reconocimiento a nivel internacional se encuentran Moody's, Standard & Poor's y Fitch IBCA. Estas entidades suelen emplear una escala alfabética, que va desde AAA, AA+, A hasta D, para categorizar a los deudores en función de su capacidad de pago (Contabilidad y Finanzas, 2024).

Se empleará la calificación crediticia de la CNBS, calculada a partir del historial de pagos, los niveles de endeudamiento y el comportamiento financiero reciente.

2.3.3. BURÓ DE CRÉDITO

Es el historial crediticio de una persona o empresa en donde se incluye información relativa a su actividad financiera respecto al crédito, es decir si es puntual al momento de realizar sus pagos o si comienza a acumular deudas de crédito (BBVA México, 2025).

Se utilizará en el estudio como una variable numérica independiente, con el propósito de medir la solvencia histórica del cliente mediante el score de Equifax.

2.3.4. PROBABILIDAD DE INCUMPLIMIENTO (PD)

Se refiere a la probabilidad de que un prestatario o emisor de instrumentos de deuda (como bonos) no cumpla con sus obligaciones de pago (FasterCapital, 2025).

Este es el indicador cuantitativo final del modelo y constituye la probabilidad estimada de incumplimiento de un cliente en un período de doce meses. Será implementado en la salida del modelo predictivo, sirviendo como fundamento para las decisiones de crédito, las provisiones y las políticas internas.

2.3.5. SERIES DE TIEMPO

Una serie temporal se define como una secuencia de observaciones de una variable, registradas a lo largo de un período determinado. Dichas observaciones se caracterizan por estar ordenadas cronológicamente (Universidad Autónoma de Madrid, 2013) (Rob J Hyndman and George Athanasopoulos, 2018).

Se analizarán las fluctuaciones históricas en los índices de morosidad y el comportamiento crediticio con el propósito de identificar patrones estacionales que incidan en el riesgo. Su uso será en la división del dataset por periodos (Train 2020-22, Valid 2023 y Test 2024) y en el análisis de estabilidad (PSI).

2.3.6. APRENDIZAJE AUTOMÁTICO

Se define como una disciplina dentro del campo de la ciencia de datos que faculta a las máquinas para adquirir conocimiento sin la necesidad de una programación explícita mediante reglas predefinidas. Esta disciplina aplica principios estadísticos para inferir propiedades, así como diversos métodos matemáticos para la identificación de patrones inherentes a los datos, permitiendo así la formulación de predicciones e incluso la adopción de decisiones (Emilio Soria Olivas et al, 2023).

Constituye el fundamento metodológico, que permitirá la elaboración de modelos predictivos capaces de estimar la probabilidad de incumplimiento crediticio a partir del análisis de la información histórica de los clientes.

2.3.7. MODELO SUPERVISADO

Se basa en el uso de datos etiquetados para entrenar un modelo y predecir valores o clasificar nuevas instancias. En términos más precisos, se alude a datos que ya exhiben una clasificación o un valor objetivo identificado. Por ejemplo: Los correos electrónicos marcados como spam y no spam (Josep, 2023).

La adopción de esta estrategia es, de hecho, fundamental; pues, el uso de datos existentes sobre el comportamiento de los clientes, tanto aquellos que cumplen con sus deberes como aquellos que no, propiciará la formulación de modelos capaces de evaluar el nivel de riesgo de potenciales peticionarios. La eficacia de estos algoritmos permitirá discriminar entre clientes cumplidos y morosos.

2.3.8. MODELO NO SUPERVISADO

Emplea algoritmos de machine learning para analizar y agrupar en clústeres conjuntos de datos sin etiquetar. Su capacidad para identificar patrones de similitud y disparidad en la información la convierte en una solución idónea para el análisis exploratorio de datos, las estrategias de venta cruzada, la segmentación de la clientela y el reconocimiento de imágenes (IBM, 2025).

Se usará este modelo para la clasificación de los clientes en función de sus rasgos y similitudes, con el objetivo de reconocer perfiles de riesgo y revelar tendencias no evidentes en la información.

2.3.9. MATRIZ DE CONFUSIÓN

Se trata de una matriz empleada para valorar la eficacia de un modelo de clasificación. Dicha matriz resume los resultados de la predicción en relación con los valores observados. Su uso es común en problemas de clasificación binaria o multiclase, ya que ofrece información más detallada que una simple métrica de predicción. Muestra cuantas predicciones han sido correctas e incorrectas, dividiéndolas en verdaderos positivos (TP), falsos positivos, (FP) verdaderos negativos (TN) y falsos negativos (FN) (Doctor Jacob Murel, 2024).

Permitirá cuantificar el desempeño del modelo a través de indicadores como precisión, sensibilidad y especificidad. Se utilizará en la fase de evaluación del modelo (Valid/Test).

2.3.10. CORRELACIÓN

Se refiere al grado en que las variables presentan una variación conjunta o covarianza. A su vez, se analizan las fluctuaciones simultáneas observadas en todas las variables medidas. Una correlación elevada sugiere que las variables tienden a mostrar un comportamiento concordante, mientras que una correlación baja indica una asociación débil entre las fluctuaciones de las variables (JMP Statistical Discovery, 2025).

Se empleará el coeficiente de correlación de Pearson con el propósito de establecer la relación existente entre las variables financieras y la morosidad.

2.3.11. MÉTRICAS DE EVALUACIÓN

Son medidas utilizadas para evaluar el rendimiento y la precisión de un modelo de regresión. Estas métricas permiten cuantificar qué tan bien el modelo predice los valores de la variable dependiente en comparación con los valores reales observados (Montgomery et al, 1990).

Este indicador se empleará para cotejar el rendimiento entre modelos y elegir el más eficaz en la predicción del riesgo crediticio.

2.4 TEORÍAS DE SUSTENTO

En la próxima sección, se presentan las teorías que sustentan esta investigación, las cuales proporcionan el marco conceptual necesario para entender la importancia de la calificación de riesgo y su relación con las contribuciones teóricas y empíricas que ya existen en el ámbito de estudio.

2.4.1 TEORÍA DE LA INFORMACIÓN ASIMÉTRICA

Desarrollada allá por los años sesenta, por George Akerlof el economista. El cual, recibió un Nobel en Economía en 2002 gracias a sus aportes al entendimiento de los fallos de mercado. Es un planteamiento que construye un esquema teórico, ayudando a entender cómo la información asimétrica entre compradores y vendedores puede torcer los precios y llevar a una distribución ineficiente de recursos.

En su publicación "The Market for Lemons", (George A. Akerlof, 1970) presenta este acontecimiento mediante el estudio del mercado de vehículos de segunda mano. En este escenario, los vendedores poseen un conocimiento superior sobre la condición genuina del artículo en

comparación con los compradores, lo que origina un problema de selección adversa. Esta hipótesis es aplicable al sector financiero, donde los solicitantes de crédito poseen más detalles sobre su aptitud y disposición para pagar que las instituciones bancarias responsables de conceder los préstamos.

Asimismo, la teoría sostiene que el mecanismo de precio falla al asignar eficientemente los recursos o el capital. En los mercados financieros, por ejemplo: factores colaterales, el riesgo y la reputación (capacidad de pago) son más importantes que la tasa de interés, lo que evidencia que la información asimétrica puede afectar decisiones económicas (Ignacio Perrotini H., 2002).

Esta teoría es fundamental para el análisis del riesgo crediticio, puesto que facilita la comprensión de cómo la asimetría de la información puede generar resoluciones incorrectas al conceder créditos, incidiendo negativamente en la calidad de la cartera y la estabilidad de la institución bancaria. Por consiguiente, subraya la necesidad de desarrollar modelos predictivos basados en aprendizaje automático, los cuales disminuyan dicha asimetría a través del estudio de datos históricos y la estimación de la probabilidad de incumplimiento.

El análisis de la información asimetría fundamenta la hipótesis H1 de este estudio, al postular que la aplicación de un modelo híbrido de ML contribuye a la disminución de las brechas informativas entre los clientes y las entidades financieras. El desarrollo del modelo permitirá identificar patrones no evidentes y las variables que influyen en el incumplimiento de pago, lo cual facilitará la toma de decisiones fundamentadas y la anticipación de riesgos.

2.4.2 TEORÍA DEL APRENDIZAJE ESTADÍSTICO (AE)

Es un campo de estudio que investiga cómo las computadoras adquieren conocimiento a partir de los datos. Examina las normas y los procedimientos que facilitan la formulación de pronósticos fundamentados en la información suministrada (Science, 2024).

Esta teoría incorpora indicadores que posibilitan el análisis del nivel de complejidad de un modelo y su conexión con el proceso de aprendizaje. Fundamentalmente, aspira a determinar si el modelo conserva una balanza apropiada entre simplicidad y facultad de representación, lo que implica que sea suficientemente simple para un aprendizaje eficaz, pero lo bastante detallado para reflejar los patrones significativos de los datos (Science, 2024).

En síntesis, AE apoya la utilización de algoritmos de aprendizaje automático para calcular la

probabilidad de incumplimiento, destacando la relevancia de validar y calibrar los modelos a fin de asegurar su capacidad para pronosticar adecuadamente situaciones futuras y, de este modo, disminuir la incertidumbre y optimizar la toma de decisiones en el proceso de evaluación de riesgos.

La perspectiva de esta teoría ofrecerá la estructura necesaria para la creación, formación y verificación del modelo híbrido. Esto garantizará que el modelo sea capaz de aprender de los datos de forma eficiente, equilibrar la complejidad con la capacidad de generalización, y producir resultados sólidos y fidedignos ante nuevos casos o diversas situaciones financieras.

2.4.3 TEORÍA DEL RIESGO FINANCIERO

Según Merton, el riesgo financiero se refiere a la posibilidad de sufrir pérdidas en una inversión debido a la incertidumbre (*Riesgo Financiero*, 2024). Esta perspectiva se sustenta por diversos autores entre ellos:

Daniel Bernoulli (1734) presentó el concepto de utilidad esperada, el cual sugiere que los individuos no valoran las ganancias exclusivamente por su monto monetario, sino por el nivel de satisfacción o beneficio que estas les confieren. Dicho de otro modo, las personas suelen preferir resultados seguros sobre alternativas inciertas (Loteró, R. S., & Molano, R. V., 2007).

Por otro lado, Frank H. Knight (1921) delineó una distinción esencial entre riesgo e incertidumbre. El riesgo se vincula con circunstancias en las que es factible asignar probabilidades a los desenlaces, mientras que la incertidumbre comprende situaciones en las que tales probabilidades son ignoradas o imposibles de calcular (Loteró, R. S., & Molano, R. V., 2007).

Esta teoría constituye un fundamento para la estimación de la probabilidad de incumplimiento, dado que contribuye a la comprensión de cómo las determinaciones de endeudamiento están influenciadas por el nivel de certidumbre o incertidumbre percibida. De este modo, apoya la implementación de modelos estadísticos y sistemas de calificación crediticia como instrumentos que posibilitan cuantificar y administrar con mayor exactitud el nivel de riesgo inherente en las carteras de crédito bancarias.

En este estudio, se utilizará para analizar y medir la incertidumbre, lo que facilitará la creación de pautas precisas para evaluar y manejar el riesgo de crédito y funcionará como fundamento para elegir las variables y métodos estadísticos usados en la elaboración del modelo

hibrido.

2.5 ANÁLISIS DE LAS METODOLOGÍAS

Estructurar una investigación científica es crucial; asegura la precisión y la validez de los hallazgos. Escoger una metodología idónea, ayuda a realizar el estudio de forma sistemática y coherente. Además, este esquema guía cada fase de la investigación y brinda orientación para evaluar y usar los resultados en situaciones reales.

2.5.1 PARADIGMA POSITIVISMO Y POST-POSITIVISMO

El paradigma positivista, o como también se le llama cuantitativo, empírico-analítico, o racionalista, arranca de las ideas de Augusto Comte (1849), Básicamente defiende la idea de que la realidad es objetiva, además medible y se puede justificar a través de la observación y la cuantificación. Este enfoque facilita la identificación de relaciones numéricas y patrones en grupos que son representativos, intentando que el investigador se mantenga objetivo respecto a lo que estudia (Castrillo, 2024).

A diferencia del positivismo, el post-positivismo mantiene, desde una postura filosófica, que ninguna comprensión humana puede ser considerada totalmente veraz, incluso si los individuos la asumen como una verdad innegable (Post-Positivismo, 2023). Esta perspectiva reconoce que las vivencias personales, los sesgos y las imperfecciones humanas pueden afectar el proceso científico. A diferencia del positivismo, el post-positivismo acepta la posibilidad de errores dentro del enfoque científico, lo que significa que incluso las certezas antes vistas como indiscutibles pueden ser puestas en duda o interpretadas de otra manera.

Considerando esta investigación, la selección del paradigma post-positivista se justifica por la naturaleza cuantitativa de los datos y los objetivos planteados.

La evaluación del riesgo crediticio exige un análisis de cantidades ingentes de información financiera, si bien cuantificable, exhibió cierta variabilidad e incertidumbre eh. Modelos de aprendizaje automático, por ejemplo, árboles de decisión, son herramientas útiles para detectar patrones, y para generar predicciones, ciertamente. Pero, la precisión depende de la calidad de los datos, la elección acertada de variables y la configuración de los algoritmos correctas, sin duda.

2.6 ANTECEDENTES DE METODOLOGÍAS

Los modelos de evaluación del riesgo crediticio, de hecho, han experimentado una

transformación notable. De las metodologías estadísticas convencionales, hemos transitado a sistemas sofisticados, estos impulsados por el aprendizaje automático (Baesens et al., 2016).

Durante la década de 1980, surgieron los primeros sistemas cuantitativos de scoring, donde se destacó la regresión logística como técnica dominante para estimar la Probabilidad de Incumplimiento (PD) debido a su capacidad de explicar relaciones lineales entre variables financieras y el riesgo crediticio (Thomas et al., 2017).

Con los requerimientos regulatorios de Basilea II y III, la gestión del riesgo tuvo mucho mayor rigor técnico, aumentando el uso de modelos estadísticos paramétricos y la medición de parámetros como LGD y EAD (BCBS, 2010).

A partir del auge del Big Data y la computación avanzada (2010 hacia adelante), las instituciones financieras empezaron a incorporar algoritmos de machine learning, como Support Vector Machines, Random Forest y Boosting, que permiten capturar patrones no lineales y relaciones entre variables, incrementando la exactitud y robustez de la clasificación crediticia (Lessmann et al., 2015). Asimismo, metodologías como CRISP-DM se consolidaron como marco estructural para el desarrollo de modelos de minería de datos en el sector financiero (Chapman, 2000).

En la actualidad, los modelos híbridos que integran reglas de negocio y el ML han demostrado mejorar de forma significativa la estabilidad predictiva, la reducción de sesgos y la adaptación del modelo frente a condiciones económicas cambiantes (Nguyen et al., 2025). En este aspecto, la presente investigación se posiciona en el estado del arte metodológico, teniendo un enfoque híbrido que asegura trazabilidad, interpretabilidad y exactitud en la calificación del riesgo crediticio empresarial.

2.7 METODOLOGÍAS Y ESQUEMAS DE DISEÑO

La investigación adopta como marco principal la metodología CRISP-DM (Cross-Industry Standard Process for Data Mining), por su enfoque iterativo, su aplicabilidad comprobada en proyectos bancarios y su alineación con el desarrollo de modelos predictivos de riesgo. El diseño metodológico se articula en seis fases:

Tabla 2: Fases de Metodología CRISP-DM

#	Fase CRISP-DM	Aplicación En La Presente Investigación
1	Comprensión del negocio	Análisis de las políticas crediticias empresariales, indicadores de riesgo y criterios regulatorios CNBS.
2	Comprensión de los datos	Exploración de variables financieras e históricas crediticias de clientes empresariales.
3	Preparación de los datos	Limpieza, imputación de valores faltantes, balanceo de clases, transformación y selección de variables.
4	Modelado	Aplicación de algoritmos ML supervisados (Logistic Regression, Random Forest, XGBoost, SVM) y métodos no supervisados opcionales para segmentación.
5	Evaluación	Validación con métricas de desempeño: Accuracy, F1-Score, Recall, ROC-AUC; comparación con modelo tradicional.
6	Despliegue	Propuesta de integración del modelo dentro del flujo de evaluación crediticia institucional.

Fuente: Elaboración Propia

Este esquema nos permite asegurar:

- Rigurosidad estadística en la selección de variables ($p < 0.05$).
- Trazabilidad en decisiones y resultados.
- Reproducibilidad técnica.
- Cumplimiento normativo y alineación con la regulación prudencial CNBS.

El diseño de investigación es no experimental, transversal, y a la vez correlacional-explicativo. Se analizan datos históricos, sin manipular las variables; además se estudian relaciones causales con probabilidad de incumplimiento.

La validación emplea el método Train-Test Split, también Cross-Validation, para una correcta generalización. Finalmente, la elección definitiva se hará según el rendimiento predictivo y la estabilidad del modelo, sobre datos nuevos.

Con este diseño metodológico se asegura que el modelo híbrido propuesto mejore la precisión, consistencia y eficiencia operativa de la toma de decisiones crediticias en la banca empresarial en Honduras.

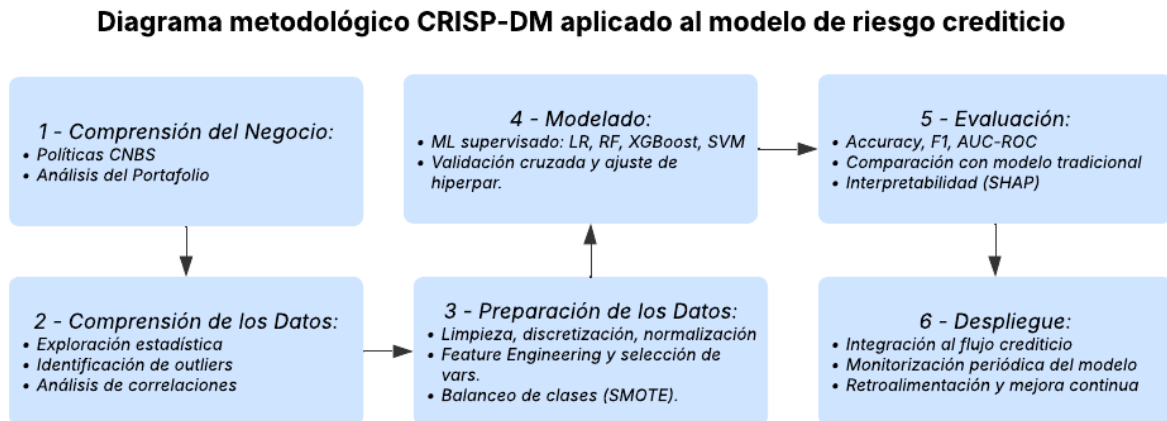


Figura 2: Diagrama Metodológico CRISP-DM

Fuente: Elaboración Propia

2.8 ANÁLISIS CRÍTICO DE METODOLOGÍAS

En este capítulo, se explora la aproximación metodológica usada en estudios previos de evaluaciones de riesgo; enfocándonos en el cambio de técnicas estadísticas clásicas a la adopción de modelos impulsados por el aprendizaje automático.

A continuación, se presenta un análisis comparativo de las metodologías:

Tabla 3: Comparativa de Métodos Relevantes

Autor / Año	Metodología Utilizada	Fortalezas	Limitaciones	Justificación de la investigación
Thomas et al. (2017)	Regresión logística aplicada al credit scoring.	Interpretación clara y elevada aceptación regulatoria.	El método asume la existencia de relaciones lineales entre las variables y presenta una capacidad predictiva limitada frente a datos que no exhiben linealidad.	Sirve como referencia valiosa para la construcción del modelo base; evidencia la necesidad de métodos más flexibles.

Autor / Año	Metodología Utilizada	Fortalezas	Limitaciones	Justificación de la investigación
Lessmann et al. (2015)	Comparación de algoritmos de machine learning (Random Forest, SVM, XGBoost).	Precisión y reconocimiento de configuraciones complejas optimizados.	Demanda un volumen considerable de datos y una validación exhaustiva para prevenir el sobreajuste.	Sustenta la selección de algoritmos de aprendizaje automático dentro de esta investigación.
Yang et al. (2024)	Modelo de deep learning para predicción de riesgo financiero.	Gran habilidad de generalización y exactitud en las predicciones.	Ofrece una interpretabilidad reducida y exige una alta cantidad de recursos computacionales.	Proporciona pruebas de la progresión hacia modelos de mayor complejidad, si bien con limitada aplicabilidad en entornos bancarios regionales.
Nguyen et al. (2025)	Modelo híbrido de reglas de negocio + machine learning.	Fusiona precisión con claridad, lo que optimiza la estabilidad de los resultados.	Sujeto a la calidad de datos y calibración del modelo.	Inspiración directa para el enfoque híbrido adoptado.
Kumar Jain (2019)	Modelo basado en IA aplicado en banca empresarial (caso HDFC Bank).	Disminución de los plazos y de los activos no productivos (20% menos de NPA).	No detalla estructura de datos ni validaciones estadísticas.	

Fuente: Elaboración Propia

La literatura académica muestra una evolución ordenada. Se traslada, con rigor, de los modelos estadísticos lineales a los enfoques del aprendizaje automático. Investigaciones, tal como las de Thomas et al., (2017) y Lessmann et al., (2015), subrayan que los modelos no lineales exhiben una eficiencia superior al predecir el impago crediticio. Otras estrategias buscan, un balance delicado, entre precisión predictiva e interpretabilidad. No obstante, retos persisten acerca de la claridad de los modelos, y los recursos necesarios para su aplicación práctica. Esencialmente, la investigación actual adoptará una estrategia mixta. Estará basada en la metodología CRISP-DM, que fusionará el análisis de aprendizaje automático y el entorno regulatorio del sistema financiero hondureño.

Por consiguiente, la investigación en curso empleará una estrategia mixta, fundamentada en la metodología CRISP-DM, que integrará el análisis de aprendizaje automático con el marco regulatorio del sistema financiero de Honduras.

2.9 HERRAMIENTAS A UTILIZAR

En esta sección, se detalla las herramientas tecnológicas y analíticas que usamos en la investigación. Se escogieron porque podían mejorar la recolección, el procesamiento y, lo más importante, el análisis de datos.

Esas herramientas ayudan a una buena integración de métodos estadísticos, modelos que predicen, y visualización de datos, mostrando eficacia y una alta confiabilidad. Sirven para lograr las metas del estudio.

La tabla a continuación compara cada herramienta. En ella se resaltan sus funciones principales y el valor especial que le dan al proyecto, apoyando a un entendimiento completo de su uso en la investigación.

Las herramientas elegidas posibilitan un enfoque integral para manejar las distintas fases del proceso analítico: preparación, análisis, modelado, visualización y almacenamiento de datos.

Tabla 4: Herramientas Comparativas de Preparación de Datos

Herramienta	Descripción	Ventajas	Desventajas	Costo	URL de Descarga	Recomendación
Microsoft Excel	Herramienta para analizar, calcular y visualizar datos.	Ideal para realizar análisis exploratorios de datos (EDA).	Lento e inestable cuando se trata de grandes volúmenes de datos.	Gratuito.	https://www.microsoft.com/es-es/microsoft-365/excel	Herramienta de apoyo para preparar los datos.
Google Colab	Entorno en la nube para ejecutar código Python.	No requiere instalación, se puede compartir fácilmente.	Se requiere tener internet.	Gratuito.	https://colab.research.google.com	Complementario: A Jupyter para ejecutar modelos en la nube.
Jupyter Notebook	Interactivo para escribir y ejecutar código Python.	Permite combinar código, visualizaciones y notas en un solo documento. Ideal para experimentación.	No es adecuado para grandes volúmenes de datos.	Gratuito.	https://jupyter.org	Seleccionado: Ideal para documentar el proceso de modelado y pruebas.

Fuente: Elaboración Propia

En la etapa de preparación y limpieza de datos, se eligió a Excel y Power Query debido a su integración directa con el ecosistema de Microsoft, su simplicidad de uso y su compatibilidad con Power BI y SQL Server. Estas herramientas depuran, transforman y organizan la información financiera rápidamente, sin infraestructura adicional ni instalaciones complejas.

Tabla 5: Comparativo de Herramientas de Análisis Estadístico y Modelado

Herramienta	Descripción	Ventajas	Desventajas	Costo	URL de Descarga	Recomendación
Python	Lenguaje de programación con amplia gama de librerías ML y ciencia datos.	Flexible, con librerías potentes (Pandas, Scikit-learn, XGBoost). Tiene soporte.	Necesita conocimiento técnico previo.	Gratuito.	https://www.python.org	Complementario: Base principal para el análisis estadístico.
R	Lenguaje estadístico especializado en análisis y gráficas.	Análisis estadístico y modelos econométricos.	Menos integrado con entornos de producción.	Gratuito.	https://www.r-project.org	Complementario: Necesario para validaciones estadísticas.
KNIME	Plataforma de análisis y automatización de flujos de datos con interfaz visual.	No requiere programación; permite integración con Python y R.	No es flexible para personalizaciones avanzadas.	Gratuito.	https://www.knime.com	Seleccionado: Adecuada para el desarrollo de ML y flujos de análisis de riesgo.
RapidMiner	Plataforma visual de minería de datos y aprendizaje automático.	Interfaz intuitiva; adecuada para modelos sencillos.	Menos flexible que Python para personalizaciones.	Versión gratuita con limitaciones.	https://rapidminer.com	Descartado: No seleccionada por menor integración con Python.

Fuente: Elaboración Propia

Se eligieron Python, Jupyter Notebook, Google Colab y KNIME para el análisis estadístico y modelado predictivo. Si bien hay opciones como RapidMiner y R que tienen funcionalidades similares, su integración con el entorno organizativo es inferior.

Por lo tanto, KNIME se posiciona como la herramienta principal para la creación del modelo predictivo de calificación de riesgo, debido a su interfaz gráfica, su compatibilidad con Python y su habilidad para automatizar flujos de análisis de manera intuitiva y replicable.

Tabla 6: Comparativo de Herramientas de Visualización de datos

Herramienta	Descripción	Ventajas	Desventajas	Costo	URL de Descarga	Recomendación
Power BI	Herramienta de visualización de datos.	Conexión con múltiples fuentes de datos, Excel y SQL.	Requiere conocimientos de bases de datos o modelados.	Gratuito (Power BI Desktop); Plan Pro desde \$9.99/usuario/mes.	https://powerbi.microsoft.com	Seleccionado: Ideal para visualizar y desarrollar KPIs modelo de riesgo crediticio.
Tableau	Plataforma de análisis visual, permite transformar los datos para la resolución de problemas.	Rendimientos en grandes volúmenes de datos.	Se requiere de administración técnica, no es tan potente para modelado de datos.	Versión de paga: Plan Enterprise viewer desde \$35 usuario/mes	https://www.tableau.com/	Descartado: Curva de aprendizaje avanzada para cálculos potentes.
Looker Studio	Plataforma de visualización y análisis de datos.	Intuitivo y fácil de usar	Conexión a bases de datos externas, requiere conectores de pago a terceros.	Gratuito	https://lookerstudio.google.com/	Descartado: No es adecuado para análisis estadístico y predictivo.
Qlik	Plataforma de análisis la cual permite democratizar, recopilar y procesar datos.	Ofrece soporte para análisis descriptivo y facilita la creación de paneles interactivos.	Exige un servidor dedicado o una infraestructura en la nube para empresas.	Versión de paga: Plan de 10 usuarios desde \$200 mes.	https://www.qlik.com/us	Descartado: Qlik exige una inversión sustancial y una curva de aprendizaje más prolongada.

Fuente: Elaboración Propia

Se evaluaron plataformas para la visualización de datos como Tableau, Looker Studio y Qlik, que, aunque proporcionan funcionalidades avanzadas y paneles interactivos, conllevan costos más altos y requieren un tiempo de aprendizaje más extenso. Por otro lado, se optó por Power BI debido a su integración directa con Excel, SQL Server y la nube de Azure, facilitando así la creación de paneles dinámicos y reportes interactivos de forma eficiente dentro del entorno tecnológico actual.

Tabla 7: Comparativo de Herramientas de Almacenamiento de Información

Herramienta	Descripción	Ventajas	Desventajas	Costo	URL de Descarga	Recomendación
SQL Server	Sistema de bases de datos relacional creado por Microsoft.	Integración con Power BI, Excel, posee soporte técnico robusto.	Licencia de pago, requiere configuración avanzada.	De pago, según el versionado.	https://www.microsoft.com/sql-server	Seleccionado: Integración con todo el ecosistema de Microsoft y su gran capacidad para manejar volúmenes de datos.
Mysql Server	Gestor de bases de datos relacional, código abierto.	Amplia comunidad de soporte	Tiene un menor rendimiento en grandes entornos empresariales	Libre / Open Source	https://www.mysql.com	Descartada: Menor integración con el entorno de Microsoft.
PostgreSQL	Gestor de bases de datos orientado a objetos con amplio soporte para datos complejos.	Alta escalabilidad, soporte para consultas avanzadas.	Requiere mayor conocimiento técnico, interfaz menos amigable.	Libre / Open Source	https://www.postgresql.org	Descartada: Riesgo de soporte insuficiente.
MongoDB	Gestor de bases NoSQL, orientado archivos JSON, datos semiestructurados.	Excelente rendimiento de grandes volúmenes de datos no estructurados.	Requiere de curva de aprendizaje técnica y configuración adicional con Microsoft.	Gratuito (versión community) y planes de pago (MongoDB Atlas).	https://www.mongodb.com	Descartada: Dado a que los datos utilizados corresponden a información financiera relacional un modelo NoSQL no es apto.

Fuente: Elaboración Propia

Finalmente, se optó por SQL Server para la administración y almacenamiento de datos debido a su compatibilidad con el entorno de Microsoft, así como por su seguridad, capacidad de escalabilidad y soporte empresarial. Aunque se consideraron MySQL, PostgreSQL y MongoDB, se descartaron porque necesitaban configuraciones adicionales y no se ajustaban correctamente a la estructura relacional de los datos financieros utilizados en el estudio.

En resumen, la elección de las herramientas se basa en la compatibilidad tecnológica, la eficiencia operativa y la concordancia con los fines del estudio. Cada aplicación desempeña una función complementaria dentro del proceso analítico, lo que asegura la consistencia entre la preparación de los datos, la creación del modelo predictivo y la presentación de los resultados. Esta integración tecnológica hará posible la ejecución ordenada y exacta del estudio, lo que garantizará

la validez y la capacidad de repetición de los resultados obtenidos.

2.10 MARCO LEGAL

2.10.1 MARCO LEGAL INTERNACIONAL

2.10.1.1 BASILEA III COMITÉ DE SUPERVISIÓN BANCARIA (BCBS)

Conjunto de reformas internacionales sobre requisitos de capital y liquidez para bancos, diseñado para fortalecer la regulación, supervisión y gestión del riesgo en el sector bancario (Basel III, 2017).

Las exigencias de capital y gestión del riesgo derivadas de Basilea III exigen que los bancos cuenten con metodologías sólidas para estimar riesgos de crédito (PD/LGD). Un modelo credit scoring más preciso impacta directamente en la estimación de activos ponderados por riesgo y provisiones, por lo que alinear este trabajo de investigación con los principios de Basilea aporta validez prudencial a su uso en toma de decisiones crediticias.

2.10.1.2 BCBS 239 PRINCIPIOS PARA LA AGREGACIÓN EFECTIVA DE DATOS DE RIESGO Y REPORTES (RISK DATA AGGREGATION)

Principios del Basel Committee que establecen buenas prácticas para la agregación de datos de riesgo y la calidad de la información de reporte interno (Basel Committee on Banking Supervision, 2013).

El modelo depende de la calidad, integridad y trazabilidad de los datos. BCBS 239 exige capacidades robustas de agregación de datos y reporting, esto respalda la necesidad de procesos CRISP-DM rigurosos, gobernanza de datos (COBIT) y controles de seguridad (ISO 27001) para asegurar que los outputs del modelo sean confiables y auditables.

2.10.1.3 SR 11-7 GUIDANCE ON MODEL RISK MANAGEMENT (FED / OCC — EE. UU.)

Guía sobre la Gestión del Riesgo de Modelo (Model Risk Management): gobernanza, desarrollo, validación y uso seguro de modelos cuantitativos. Esto incluye validaciones (backtesting), tests de estabilidad y gestión de riesgo de modelo (SR 11-7, 2011, p. 11).

La guía establece expectativas sobre documentación, validación independiente, gestión de cambios y gobernanza por parte del directorio y la alta gerencia. Aplicar SR 11-7, o sus principios,

al modelo, mejora su aceptabilidad regulatoria y su robustez frente a auditorías internas/externas.

2.10.1.4 GDPR (REGLAMENTO GENERAL DE PROTECCIÓN DE DATOS UE) Y CONVENIOS EUROPEOS SOBRE PRIVACIDAD

Marco riguroso de protección de datos personales, derechos, bases legales, obligaciones de controladores y procesadores, transferencias internacionales (Regulation - 2016/679 GDPR, 2016).

Si bien es cierto, GDPR es legislación europea, sus principios, minimización, finalidad, base legal y derechos son referencia internacional para el tratamiento de datos personales en proyectos de ML. Aplicar esos principios, minimización de atributos, etc., es una mejor práctica para proyectos que manejan datos sensibles bancarios.

2.10.1.5 RECOMENDACIONES DEL GAFI / FATF (ANTI-MONEY LAUNDERING / AML)

Estándares internacionales para prevención de lavado de dinero y financiamiento del terrorismo (KYC, monitoreo transaccional, intercambio de información). FATF guía cómo conciliar inclusión financiera y controles de AML, relevante al diseñar variables externas en el modelo. (The FATF Recommendations, 2012).

El presente trabajo de investigación debe respetar las obligaciones AML/CFT: el uso de datos para calificar riesgo no debe impedir la detección de patrones sospechosos; además, cualquier integración de fuentes externas.

2.10.1.6 ISO/IEC 27001 SISTEMA DE GESTIÓN DE SEGURIDAD DE LA INFORMACIÓN (SGSI)

Norma internacional para establecer un SGSI, identificar riesgos, controles y procesos de mejora continua en seguridad de la información. ISO 27001 provee el marco para proteger la confidencialidad, integridad y disponibilidad de los datos usados en el modelo (ISO 27001, 2024).

Esta norma se aplica en el presente trabajo ya que la manipulación de bases de datos de clientes (confidenciales y financieros) exige controles técnicos y organizativos (cifrado, control de accesos, logs, gestión de incidentes).

2.10.1.7 COBIT (ISACA) Y ITIL (AXELOS) GOBERNANZA Y GESTIÓN DE TI

Marco para gobernanza y gestión de TI alineada con objetivos de negocio. COBIT es un

marco de trabajo para el gobierno y la gestión de las tecnologías de la información (TI) empresariales y dirigido a toda la empresa. Ha sido promovido por ISACA desde su primera versión en 1996 y actualmente se encuentra disponible la versión COBIT 2019 (Villamizar, 2022).

La relación con la investigación actual: COBIT establece estructuras de gobernanza (roles, métricas, controles) para asegurar que las decisiones derivadas del modelo estén alineadas con apetito de riesgo y requisitos regulatorios.

2.10.1.8 NORMAS Y GUÍAS SOBRE “MODEL RISK” Y VALIDACIÓN (OCC, FDIC Y BANCOS CENTRALES)

Publicaciones y guías (OCC Comptroller’s Handbook, FDIC adoption memos) sobre prácticas supervisoras en validación, challenge independiente y gobernanza de modelos (Sound Practices for Model Risk Management, 2011).

Relación con la investigación: Refuerzan la necesidad de pruebas robustas (backtesting, sensibilidad, estabilidad temporal), independencia en validación e historial documental, componentes que debes incorporar en la fase de evaluación del CRISP-DM.

2.10.2 MARCO LEGAL NACIONAL

2.10.2.1 COMISIÓN NACIONAL DE BANCOS Y SEGUROS (CNBS) / NORMAS PARA EVALUACIÓN Y CLASIFICACIÓN DE CARTERA CREDITICIA.

Regulaciones y circulares emitidas por la CNBS que establecen criterios de clasificación de créditos, provisiones y reportes regulatorios (CIRCULAR CNBS No.001/2025, 2025).

El otorgamiento de créditos comerciales se basará en criterios tales como la capacidad de pago del deudor, su historial de cumplimiento de pagos, existencia de garantías y el entorno económico del deudor permitirá clasificar la totalidad de las obligaciones de los deudores comerciales en 5 categorías de riesgos:

Tabla 8: Clasificación de Créditos

Categoría	Nombre
I	Créditos Buenos
II	Créditos Especialmente Mencionados
III	Créditos Bajo Norma
IV	Créditos de Dudosa Recuperación
V	Créditos de Pérdida

Fuente: CIRCULAR CNBS No001 2025

Relación con la investigación: Se debe respetar estas reglas y demostrar que sus salidas (clasificaciones de riesgo) son compatibles con los criterios regulatorios y que no reducen la capacidad de la entidad para cumplir reportes regulatorios.

2.10.2.2 BANCO CENTRAL DE HONDURAS (BCH) - INFORMACIÓN CREDITICIA Y ESTABILIDAD FINANCIERA

Son directrices y requerimientos sobre reportes macro prudenciales y la Central de Información Crediticia (CIC) para propiciar el normal funcionamiento del sistema de pagos y velar por la estabilidad del sistema financiero nacional (Banco Central de Honduras, 2025) (Estabilidad Financiera Informe de Estabilidad Financiera (IEF), 2025).

El BCH exige ciertos reportes y métricas de supervisión; El modelo propuesto debe garantizar que los datos preparados y las salidas sean compatibles con los reportes exigidos y que no se comprometa la confidencialidad ni la integridad de la información remitida a organismos centrales.

2.10.2.3 SITUACIÓN DE LA LEGISLACIÓN SOBRE PROTECCIÓN DE DATOS EN HONDURAS

A la fecha de las fuentes públicas analizadas, Honduras no contaba con una ley plenamente consolidada y en vigor de protección de datos personales equivalente al GDPR; existen esfuerzos y proyectos (anteproyectos) y marcos sobre transparencia y acceso a la información pública como el Decreto Legislativo No.170-2006 y otros decretos/PCM sobre gestión de la información pública. En la práctica, las entidades públicas han emitido decretos y reglamentos que reconocen derechos de protección, pero no hay una ley de privacidad completa con igual alcance que el GDPR. (Tomé, 2019).

Justificación: Dado ese vacío o lag en la legislación local, es imperativo que la investigación aplique estándares internacionales (GDPR, ISO 27001, principios de minimización) y documente medidas de anonimización/pseudonimización para evitar riesgos legales y de reputación. Además, en el entorno bancario existen obligaciones contractuales y regulatorias (CNBS/BCH) relativas al tratamiento seguro y no divulgación de datos financieros que deben respetarse.

2.10.2.4 LEY DE TRANSPARENCIA Y ACCESO A LA INFORMACIÓN PÚBLICA (DECRETO 170-2006) Y DECRETOS SOBRE ADMINISTRACIÓN PÚBLICA ELECTRÓNICA

Marcos que regulan el acceso a información pública y establecen obligaciones de transparencia. Algunas disposiciones incluyen límites y excepciones por razones de seguridad o protección de datos personales (Congreso Nacional de la República de Honduras, 2006).

Relación: En la presente investigación cuando los datos provengan de fuentes, se debe respetar la separación entre información pública y datos personales sensibles.

CAPÍTULO III. METODOLOGÍA

3.1. CONGRUENCIA METODOLÓGICA

La presente investigación se desarrolla bajo un enfoque cuantitativo, buscando entrelazar variables a través de modelos de aprendizaje automático. Evaluar el riesgo crediticio de los clientes empresariales del sector bancario constituye la meta principal. La metodología CRISP-DM, conocida por su solidez en el análisis de datos y aprendizaje automático, sustenta el abordaje empleado permitiendo una estructuración rigurosa del análisis.

3.1.1. MATRIZ METODOLÓGICA

A continuación, se presenta la matriz metodológica correspondiente a esta investigación:

Tabla 9: Matriz Metodológica

Matriz Metodológica						
Modelo de Calificación de Riesgo Crediticio para Clientes Comerciales (CRCE)						
Problema	Objetivo General	Objetivos Específicos	Preguntas de Investigación	Variables Independientes	Variable Dependiente	Métodos
¿En qué medida la implementación de un modelo híbrido que combina aprendizaje automático y reglas de negocio (I) mejora la exactitud en la evaluación del riesgo crediticio (O) de los clientes de la	Evaluar (S) el impacto de un modelo híbrido que combina aprendizaje automático y reglas de negocio sobre la precisión en la evaluación del riesgo crediticio de clientes de la banca empresarial en Honduras, (M) midiendo métricas de desempeño como precisión, recall y AUC-ROC, (A)	(S) Comparar el (M) desempeño, medido por AUC-ROC y precisión, (A) entre un modelo que integra variables internas y externas y otro que utiliza solo variables internas, (R) para evaluar la relevancia de factores externos en la predicción de morosidad, (T) durante la fase de análisis exploratorio.	¿En qué medida la incorporación de variables macroeconómicas externas (I) mejora la precisión predictiva del riesgo crediticio (O) de los clientes empresariales hondureños (P) en comparación con los modelos que utilizan exclusivamente variables internas del cliente (C)?	Historial Crediticio	Riesgo Crediticio	Modelos predictivos de Machine Learning Estudio Correlacional y explicativo

banca empresarial hondureña (P) en comparación con los métodos convencionales basados únicamente en reglas de negocio (C)?	utilizando datos históricos y algoritmos supervisados, (R) para establecer un modelo validado que optimice la clasificación de riesgo, (T) durante la fase de análisis y modelado del estudio.	Evaluar (S) la influencia del componente de aprendizaje automático dentro del modelo híbrido, (M) cuantificando métricas como recall, F1-score y precisión, (A) aplicando algoritmos supervisados sobre los datos históricos, (R) para mejorar la capacidad de anticipar incumplimientos y riesgos emergentes, (T) durante la fase de modelado y validación.	¿Cómo la integración de técnicas de aprendizaje automático en los procesos de calificación crediticia (I) mejora la capacidad de anticipar incumplimientos y riesgos emergentes (O) en la banca empresarial hondureña (P) frente a los modelos convencionales basados solo en reglas de negocio (C)?	Flujo de Caja		Diseño de Investigación: No experimental, Transversal.
		Determinar (S) el impacto de la implementación del modelo híbrido sobre la coherencia y consistencia de la evaluación crediticia, (M) midiendo la concordancia de los resultados simulados con registros históricos validados, (A) mediante simulaciones de escenarios crediticios, (R) para optimizar la calidad y confiabilidad del análisis financiero, (T) durante la fase de pruebas y simulación.	¿Qué impacto tiene la aplicación de un modelo híbrido que combina datos internos financieros y variables no financieras (I) en la coherencia y calidad de las decisiones crediticias (O) de los clientes de la banca empresarial hondureña (P) frente a las metodologías tradicionales centradas únicamente en información contable (C)?	Nivel Endeudamiento Historial Financiero		Análisis de datos: Modelos estadísticos y técnicas de minería de datos. Instrumentos: Bases de datos financieras, algoritmos de aprendizaje automático.

Fuente: Elaboración Propia

3.1.2 ESQUEMA DE VARIABLES DE ESTUDIO

Las variables examinadas en esta investigación tienen una conexión directa con la evaluación del riesgo crediticio. Variables independientes como historial financiero, flujo de caja, y el nivel de endeudamiento son consideradas. Por otro lado, la variable dependiente es el riesgo de incumplimiento del crédito.



Figura 3: Esquema de Relación Causal Multivariada

Fuente Elaboración Propia

3.1.3. OPERACIONALIZACIÓN DE LAS VARIABLES

El esquema de las variables de estudio es en esencia, una estructura meticulosamente organizada. Ella se encarga de definir las variables, con precisión, que serán sometidas a análisis durante el curso de una investigación y evaluar las hipótesis planteadas (Roberto Hernández Sampieri et al., 2014).

Tabla 10: Esquema de Variables de Estudio

Variable	Definición Conceptual	Definición Operacional	Dimensiones	Indicadores
Riesgo Crediticio	Probabilidad de incumplimiento en un crédito.	Modelos predictivos de Machine Learning analizan datos históricos.	Historial crediticio, liquidez, solvencia.	Puntuación de riesgo
Historial Crediticio	Registro de comportamiento financiero pasado.	Datos de pagos y deudas anteriores.	Tiempo de endeudamiento, cumplimiento de pagos.	Número de créditos previos, tasa de mora
Nivel de Endeudamiento	Relación entre deudas y capacidad de pago.	Cálculo de la proporción de deuda.	Pasivos financieros, ingresos.	Ratio deuda/ingreso

Fuente: Elaboración Propia

3.1.3.1 HIPÓTESIS

Hipótesis de Investigación (H1)

El modelo híbrido, que integra reglas de negocio y aprendizaje automático, alcanza un desempeño significativamente superior en discriminación y separación del riesgo crediticio medido por AUC-ROC y KS respecto del esquema basado exclusivamente en reglas y de la regresión logística de referencia, tanto en la validación como en la prueba, de acuerdo con pruebas de contraste por bootstrap al 95% de confianza.

Hipótesis Nula (H0)

El modelo híbrido no presenta mejoras en el desempeño predictivo, medido por AUC-ROC y KS, en comparación con los modelos de referencia: reglas de negocio y regresión logística, en los periodos evaluados.

El contraste de las hipótesis se realiza mediante remuestreo bootstrap, estimando intervalos de confianza al 95% para las diferencias de AUC-ROC y KS entre el modelo híbrido y los modelos de referencia.

Se considerará evidencia a favor de la hipótesis de investigación (H1) cuando el intervalo de confianza al 95% de la diferencia de AUC-ROC y de KS sea mayor que cero, indicando una mejora consistente y robusta del desempeño predictivo del modelo híbrido en los conjuntos de validación y prueba.

3.2 ENFOQUE Y MÉTODOS

El estudio se basará en un enfoque mixto, al integrar de manera complementaria métodos cuantitativos y cualitativos para abordar de forma integral la problemática de la calificación de riesgo crediticio en la banca empresarial.

De esta forma, será posible medir, comparar y confirmar, por decirlo de algún modo, empíricamente la conexión entre las variables financieras, macroeconómicas y de riesgo, todo esto con el fin de calcular la probabilidad de incumplimiento (PD) de los clientes empresariales.

Este enfoque permite combinar la precisión empírica del análisis estadístico y el modelado predictivo con la profundidad interpretativa del juicio experto institucional, garantizando una comprensión holística del fenómeno bajo estudio.

El estudio se desarrolla bajo el método CRISP-DM (Cross Industry Standard Process for Data Mining), reconocido internacionalmente como el estándar más robusto y flexible para proyectos de minería de datos y analítica predictiva (Chapman, 2000).

Se eligió el método CRISP-DM debido a su solidez y facilidad de uso en el ámbito financiero, incorporando seis fases: entendimiento del negocio, entendimiento de los datos, preparación de los datos, modelado, evaluación y despliegue.

Cada fase ayuda a asegurar que la creación del modelo sea metódica, verificable y reproducible, esto facilita la integración práctica de los resultados en los procesos crediticios de las instituciones bancarias.

3.3 DISEÑO DE LA INVESTIGACIÓN

El diseño de investigación es no experimental, transversal y correlacional-explicativo. No se manipulan variables independientes; en cambio, se observan y analizan sus relaciones con la variable dependiente (precisión del modelo).

La transversalidad permite analizar los datos en un momento determinado (enero 2020 - octubre 2024), lo cual es apropiado para evaluar comportamientos históricos del riesgo crediticio.

3.3.1 POBLACIÓN

La población del estudio está conformada por la totalidad de clientes de la banca empresarial de la institución financiera objeto de análisis, que cuentan con operaciones crediticias

registradas entre enero de 2020 y octubre de 2024.

Esta población asciende a 6,903 clientes, correspondientes a empresas de diversos sectores económicos, tamaños y niveles de riesgo, todos con información financiera, conductual y administrativa disponible en los sistemas de la institución. Incluye tanto operaciones vigentes como canceladas, siempre que contengan información financiera, contable y de comportamiento crediticio suficiente para la evaluación de riesgo.

La base de datos incluye tanto clientes vigentes como históricos, abarcando distintas categorías crediticias según la normativa de la CNBS. El acceso a esta población completa permite desarrollar un modelo predictivo con alta representatividad y validez empírica, evitando los sesgos asociados a la selección parcial o muestreo limitado.

3.3.1.1 CRITERIOS DE INCLUSIÓN Y EXCLUSIÓN

Para asegurar la calidad, coherencia y aplicabilidad de los datos dentro del modelo, se establecieron los siguientes criterios metodológicos de inclusión y exclusión:

Tabla 11: Criterios De Inclusión y Exclusión

Tipo de criterio	Descripción	Justificación
Inclusión 1	Clientes empresariales con operaciones crediticias activas o cerradas entre enero de 2020 y octubre de 2024.	Permite cubrir un horizonte temporal de cinco años, suficiente para capturar patrones de comportamiento crediticio y eventos de incumplimiento.
Inclusión 2	Clientes con información financiera completa.	Asegura la disponibilidad de variables cuantitativas necesarias para el modelado de riesgo.
Inclusión 3	Clientes con registro de comportamiento de pago y clasificación crediticia según las normas CNBS.	Permite establecer la variable objetivo (morosidad / default) y validar el modelo frente a categorías regulatorias.
Inclusión 4	Empresas clasificadas dentro de la cartera empresarial, excluyendo clientes de Banca de Personas.	Focaliza el estudio en el segmento de la banca empresarial, congruente con el objetivo de investigación.
Inclusión 5	Operaciones con al menos 6 meses de historial crediticio.	Garantiza un horizonte de observación suficiente para evaluar el desempeño y la estabilidad de pago.
Exclusión 1	Clientes con datos faltantes críticos (monto de crédito, fecha de desembolso, comportamiento de pago).	La falta de variables clave impide la modelización adecuada y genera sesgos en el aprendizaje automático.
Exclusión 2	Créditos con más de un 50% de los registros financieros incompletos o inconsistentes.	Asegura la integridad y calidad del dataset, en concordancia con los principios de data quality de BCBS 239.

Tipo de criterio	Descripción	Justificación
Exclusión 3	Clientes en proceso judicial o con información confidencial sujeta a reserva legal.	Se resguarda la privacidad de la información conforme a la Ley Especial de Protección de Datos Personales y la ISO/IEC 27001.

Fuente: Elaboración Propia

3.3.2 TÉCNICAS DE MUESTREO

Debido al acceso completo a la población de 6,903 clientes, no se aplicó una técnica de muestreo probabilístico convencional. En su lugar, se utilizó una técnica de particionado temporal estratificado, orientada a estructurar la población para fines analíticos y predictivos.

Este método respeta la secuencia cronológica de los datos y mantiene la proporción original de las clases (clientes sanos y morosos) así como la representatividad sectorial (industria, comercio, servicios, agro, etc.).

El esquema de división adoptado fue el siguiente:

Tabla 12: Esquema de Particionamiento de Datos

Datos	Período	Objetivo metodológico	Proporción Total (6,903 clientes)
Entrenamiento	2020–2022	Desarrollo inicial del modelo, ajuste de parámetros y selección de variables.	58% (3,973 registros)
Validación	2023	Calibración y evaluación de desempeño intermedio.	25% (1,727 registros)
Prueba	2024 (ene-oct)	Evaluación out-of-time y medición de desempeño final.	17% (1,203 registros)

Fuente: Elaboración Propia

El particionado temporal estratificado se justifica porque:

1. Evita fugas de información, al garantizar que el modelo solo aprenda de datos previos al período de evaluación.
2. Alinea el proceso de modelado con la realidad operativa bancaria, donde las decisiones se basan en información histórica para predecir riesgos futuros.
3. Permite medir la estabilidad del modelo en el tiempo, requisito regulatorio establecido en las normas de Basilea III y las guías de la CNBS.
4. Asegura representatividad sectorial y por clase crediticia, mediante estratificación según

las variables relevantes de negocio.

5. Cumple con el principio de trazabilidad y reproducibilidad, fundamentales en la fase de Modeling y Evaluation del proceso CRISP-DM.

3.3.3 MUESTRA

Dado que el banco X permitió el acceso a la totalidad de los registros históricos disponibles, la muestra de investigación coincide con la población total (N = 6,903).

Por lo tanto, no se seleccionó un subconjunto probabilístico, sino que se aplicó un particionado analítico y temporal de la base completa, con el objetivo de estructurar los datos para el entrenamiento, validación y prueba del modelo predictivo (ver tabla 7).

El uso de la población completa (6,903 clientes) y la aplicación de un particionado temporal estratificado aseguran la validez, representatividad y trazabilidad del proceso de modelado de riesgo crediticio.

Esta estrategia metodológica está en sintonía con los lineamientos de Basilea III, las normas de la CNBS, los estándares de ISO/IEC 27001, y las recomendaciones de la Guía SR 11-7 sobre Model Risk Management.

En conjunto, el diseño metodológico fortalece la confiabilidad del modelo propuesto como herramienta de apoyo para la toma de decisiones financieras prudentes y sostenibles.

3.4 TÉCNICAS, INSTRUMENTOS Y PROCEDIMIENTOS APLICADOS

3.4.1. TÉCNICAS

- **Análisis exploratorio de datos (EDA):** Se aplicaron técnicas de estadística descriptiva para caracterizar las variables financieras, conductuales y crediticias de los 6,903 clientes empresariales comprendidos entre enero de 2020 y octubre de 2024. El análisis incluyó medidas de tendencia central, dispersión, correlación y estabilidad.
- **Modelado predictivo de riesgo crediticio:** Se emplearon técnicas de aprendizaje automático supervisado, específicamente regresión logística, random forest, y gradient boosting, para construir un modelo híbrido de calificación crediticia que combine reglas de negocio y predicción estadística. Los algoritmos fueron validados mediante validación cruzada estratificada, evaluación de métricas (ROC-AUC, KS, F1-score, recall) y calibración de

probabilidades.

- Evaluación y validación del modelo: Conforme a las guías SR 11-7 (Federal Reserve, 2011) y Basilea III (BIS, 2017), se realizaron pruebas de desempeño, sensibilidad, robustez y estabilidad temporal para determinar la confiabilidad del modelo antes de su potencial implementación.

3.4.2. INSTRUMENTOS

En concordancia con las técnicas aplicadas, se utilizaron diversos instrumentos de carácter documental, digital y humano para garantizar la fiabilidad y validez de la información.

Tabla 13: Instrumentos Documentales

Instrumento	Descripción	Propósito
Base de datos del Banco X	Registros crediticios empresariales obtenidos del sistema interno de la institución financiera, que comprenden información de 6,903 clientes (2020–2024).	Fuente principal para el análisis estadístico y el modelado predictivo.

Fuente: Elaboración Propia

Todos los instrumentos fueron utilizados bajo políticas estrictas de confidencialidad y protección de datos personales, en cumplimiento de la Ley Especial de Protección de Datos Personales (Decreto 35-2021) y la ISO/IEC 27001.

3.4.3. PROCEDIMIENTOS

El procedimiento se ejecutó siguiendo las fases establecidas por el modelo CRISP-DM, que guio el proceso de análisis y modelado de la información de manera sistemática y reproducible:

1. Comprensión del negocio: Se realizaron reuniones y entrevistas con expertos de la institución para identificar las necesidades, los objetivos de negocio y los indicadores de desempeño relevantes.

Se determinó que el modelo debía mejorar la precisión de la calificación crediticia empresarial y optimizar los procesos de gestión de riesgo.

2. Comprensión y preparación de los datos:
 - Extracción, limpieza y consolidación de los registros de 6,903 clientes (2020–2024).
 - Aplicación de los criterios de inclusión y exclusión previamente definidos.

- Tratamiento de valores faltantes, detección de outliers, codificación de variables categóricas y normalización.
 - Definición de la variable objetivo (default / no default) conforme a la clasificación CNBS.
3. Particionado temporal estratificado: La base se dividió en conjuntos de entrenamiento (58%), validación (25%) y prueba (17%), respetando la secuencia cronológica y la representatividad sectorial.
 4. Modelado y calibración:
 - Se aplicaron algoritmos de regresión logística, random forest y XGBoost.
 - Se realizaron pruebas de desempeño (ROC-AUC, KS, F1) y calibraciones mediante isotonic regression y Platt scaling.
 5. Evaluación del modelo: Se verificó la estabilidad temporal y la capacidad predictiva del modelo bajo métricas regulatorias. Los resultados fueron contrastados con criterios de expertos mediante sesiones de revisión cualitativa.
 6. Validación cualitativa y triangulación: Los hallazgos del análisis predictivo se discutieron con analistas y jefes de riesgo del banco, quienes validaron la coherencia de las variables y la interpretabilidad del modelo. Las percepciones se integraron mediante triangulación con la evidencia cuantitativa.

El uso de la metodología CRISP-DM aseguró la reproducibilidad del proceso, mientras que la triangulación con entrevistas a expertos garantizó la pertinencia práctica del modelo.

En conjunto, las técnicas, instrumentos y procedimientos aplicados dan al estudio rigurosidad técnica, validez empírica y alineación regulatoria, así mismo, fortalecen la gestión del riesgo crediticio en la institución.

3.5. FUENTES DE INFORMACIÓN

3.5.1. FUENTES PRIMARIAS

Corresponde al conjunto de 6,903 clientes empresariales extraídos de los sistemas internos del banco x, que incluyen información financiera, conductual y clasificatoria de los clientes.

Permite analizar el comportamiento histórico de pago, las variables determinantes del riesgo y los factores que influyen en la probabilidad de incumplimiento (PD).

Su utilización es esencial para el componente cuantitativo del estudio, ya que alimenta las fases de comprensión, preparación, modelado y evaluación del método CRISP-DM.

Además, su carácter empresarial y exhaustivo garantiza la validez empírica, representatividad y confiabilidad de los resultados.

3.5.2. FUENTES SECUNDARIAS

Fuentes secundarias, son informes oficiales de la CNBS, también del Banco Central de Honduras utilizados para analizar el entorno macroeconómico, las tasas de morosidad del sistema financiero y la evolución de la cartera crediticia empresarial, además de SEFIN, igual que literatura científica sobre modelado de riesgo crediticio y de machine learning. Complementan la información contextual que ajustan los parámetros del modelo, conforme a las condiciones macroeconómicas del país.

Estas fuentes fueron seleccionadas por su relevancia y vigencia dentro del ámbito de la gestión del riesgo crediticio, la gobernanza tecnológica y la seguridad de la información. También fortalecen la validez externa y contextual del estudio, al situar los hallazgos dentro del panorama económico y regulatorio actual nacional.

3.6. PLAN PARA EL ANÁLISIS DE DATOS

El propósito de este análisis de datos es evaluar empíricamente las hipótesis planteadas respecto a la capacidad predictiva y explicativa del modelo.

El objetivo es determinar si la integración de técnicas estadísticas y de aprendizaje automático mejora significativamente la precisión en la clasificación de clientes respecto a los métodos tradicionales basados en reglas prudenciales.

Alineados con el método CRISP-DM, el plan de análisis de datos se estructura en cinco etapas:

1. Preparación de datos.
2. Análisis descriptivo y exploratorio (EDA).
3. Modelado predictivo y validación estadística.

4. Prueba de hipótesis.
5. Integración y síntesis de resultados.

3.6.1 PREPARACIÓN DE LOS DATOS

Previo al análisis inferencial, se ejecutará una fase de limpieza, integración y estandarización de la base de datos del banco x, que comprende 6,903 registros de clientes empresariales del período enero 2020 a octubre 2024. Se aplicarán los siguientes procedimientos:

- Depuración y control de calidad: eliminación de duplicados, validación de rangos y consistencia temporal.
- Tratamiento de valores faltantes: imputación estadística o eliminación según el impacto en las variables críticas.
- Codificación de variables categóricas: transformación de atributos cualitativos en representaciones numéricas mediante one-hot encoding.
- Normalización y escalamiento: estandarización de las variables financieras para mejorar la convergencia de los algoritmos de modelado.
- Balanceo de clases: aplicación de class weights o SMOTE únicamente en el conjunto de entrenamiento para corregir el desbalance entre clientes sanos y morosos.

3.6.2 ANÁLISIS DESCRIPTIVO Y EXPLORATORIO (EDA)

El análisis exploratorio de datos permitirá conocer la distribución, comportamiento y relaciones entre las variables. Se realizarán las siguientes acciones:

- Cálculo de medidas de tendencia central, dispersión y asimetría.
- Análisis de correlaciones entre variables financieras y conductuales.
- Identificación de patrones de comportamiento crediticio según sector económico y clasificación CNBS.
- Evaluación de la estabilidad temporal de las variables mediante el Population Stability Index (PSI).
- El EDA proporcionará la base empírica para la selección de variables y la formulación del modelo predictivo.

3.6.3 MODELADO PREDICTIVO Y VALIDACIÓN ESTADÍSTICA

Se utilizarán los siguientes métodos:

- Modelos estadísticos: regresión logística binaria para estimar la probabilidad de incumplimiento (PD).
- Modelos de aprendizaje automático: Random Forest, Gradient Boosting (XGBoost) y Support Vector Machines (SVM).
- Modelos híbridos: integración ponderada entre el puntaje regulatorio CNBS y la salida de los modelos predictivos, con el fin de mejorar la precisión sin perder interpretabilidad.

El conjunto de datos se dividirá de la siguiente manera:

- Entrenamiento: Datos comprendidos en el período 2020 – 2022.
- Validación: Datos comprendidos en el período 2023.
- Prueba: Datos comprendidos en el período de enero a octubre 2024.

Se evaluará el desempeño de cada modelo mediante métricas cuantitativas:

- ROC-AUC (Área bajo la curva ROC).
- KS (Kolmogorov–Smirnov).
- Precisión, Recall y F1-Score.
- Brier Score y Log-Loss (para calibración de probabilidades).

La comparación entre modelos se realizará con pruebas estadísticas de diferencia de medias y proporciones sobre las métricas de desempeño, para determinar si las mejoras obtenidas son estadísticamente significativas al nivel de confianza del 95% ($\alpha = 0.05$).

3.6.4 PRUEBA DE HIPÓTESIS

El plan de análisis contempla un particionado temporal con un esquema out-of-time acorde con CRISP-DM:

- Entrenamiento: 2020 - 2022
- Validación (selección de umbral): 2023
- Prueba out-of-time: enero - octubre 2024

Modelos a comparar:

- Baseline Reglas (proxy): decisión por umbrales de buró y cobertura de garantía.
- Baseline estadístico: Regresión Logística (class_weight balanced).
- Modelos ML: Random Forest y XGBoost (hiperparámetros conservadores para estabilidad).

Métricas primarias (discriminación y separación):

- AUC-ROC: capacidad de ranking del modelo.
- KS (Kolmogorov - Smirnov): separación entre distribuciones de score para BUENO vs MALO.

Métricas operativas (secundarias, no determinan el contraste de H1):

- Precision, Recall y F1 evaluadas al umbral operativo definido en VALID 2023 por máximo KS; aplicadas sin reoptimizar en TEST 2024.
- Brier score (calibración de probabilidades), reportado como evidencia complementaria.

Pruebas estadísticas y estimación de incertidumbre:

Para cada par de modelos A y B, y para cada split (VALID 2023 y TEST 2024), se estimaron las diferencias de desempeño:

- $\Delta AUC = AUC(A) - AUC(B)$
- $\Delta KS = KS(A) - KS(B)$

Control de sobreajuste y fuga de información:

- Todas las transformaciones (capping, imputación, codificación, escalado) se ajustaron únicamente con el conjunto de entrenamiento.
- El umbral operativo (por máximo KS) se fijó en VALID y se aplicó sin ajustes en TEST, garantizando evaluación out-of-time honesta.

Reportes y trazabilidad:

Para cada modelo y periodo se reportan:

- Métricas primarias: AUC-ROC, KS

- Métricas operativas (al umbral KS de VALID): Precision, Recall, F1
- Brier score (calibración)

Para el contraste de H1 se incluyen tablas con:

- Δ AUC y Δ KS (híbrido vs reglas; híbrido vs logística), IC95% y p-valores (bootstrap).

La elección de AUC-ROC y KS como métricas primarias responde al desbalance natural de la cartera (aproximadamente entre 6% y 7% de MALO), donde la exactitud puede resultar engañosa. El uso de bootstrap evita supuestos fuertes sobre la distribución asintótica de las métricas y es recomendable con tamaños muestrales moderados y clases desbalanceadas. La comparación out-of-time (2024) es una condición explícita de la hipótesis para asegurar validez externa y evitar sobreajuste a 2023.

3.6.5 INTEGRACIÓN Y SÍNTESIS DE RESULTADOS

Finalmente, los resultados cuantitativos y cualitativos se integrarán mediante un proceso de triangulación convergente, que permitirá validar la consistencia entre el desempeño técnico del modelo y su aplicabilidad práctica. El plan de análisis se alinea con las fases de evaluación y despliegue del método CRISP-DM, asegurando que las conclusiones derivadas del estudio sean estadísticamente válidas, operacionalmente relevantes y regulatoriamente conformes.

CAPÍTULO IV. RESULTADOS Y ANÁLISIS

Este capítulo operacionaliza los objetivos de investigación formulados en el Capítulo I al presentar, de forma ordenada y verificable, la evidencia empírica que sustenta el desarrollo del modelo. En concreto, caracteriza y depura el conjunto de datos de 6,903 registros de clientes empresariales, para garantizar su integridad y trazabilidad, identifica los predictores con mayor poder explicativo y discriminante de la morosidad mediante análisis descriptivo, correlacional y pruebas con la variable objetivo, y establece los insumos técnicos para comparar el desempeño de enfoques basados en reglas versus aprendizaje automático (Objetivo General), respetando el particionado temporal y los criterios metodológicos definidos.

Desglosa los hallazgos producto del procesamiento y análisis correspondientes a los datos recopilados en el período comprendido entre enero de 2020 y octubre de 2024. El desarrollo metodológico fue previamente delineado en el capítulo anterior, donde se definieron la estructura técnica, las fuentes de información y los procedimientos analíticos aplicados al conjunto de datos crediticios.

El propósito de este capítulo consiste en la presentación, interpretación, y la validación de los resultados derivados de la aplicación del modelo, dedicado a la calificación del riesgo crediticio; modelo que integra métodos estadísticos convencionales, y los algoritmos del aprendizaje automático.

En la sección 4.1 se aborda el análisis exploratorio de los datos analizados y se desglosan las siguientes subsecciones:

En la subsección 4.1.1 se presenta la composición inicial de la base analizada, detallando el número total de clientes, la cantidad de variables y su clasificación entre cuantitativas y cualitativas. Se exponen las principales características estructurales de los datos, proporcionando una visión panorámica de la información disponible para el estudio.

La subsección 4.1.2 profundiza en las variables más representativas, diferenciando los atributos financieros, demográficos y conductuales que intervienen en la calificación crediticia. Se describen sus medidas de tendencia central, dispersión y forma, con el objetivo de identificar su comportamiento y su posible relevancia en la predicción del riesgo de incumplimiento.

En la subsección 4.1.3 se detalla el proceso técnico de depuración y estandarización del

conjunto de datos. Se realiza la identificación y manejo de valores atípicos junto al tratamiento de datos faltantes y del uso de técnicas de normalización y estandarización que garantizan la calidad, integridad y comparación de la información antes del modelado.

La subsección 4.1.4 se trata la fase de análisis exploratorio usando herramientas estadísticas y también gráficas. Se hace uso de histogramas, diagramas de caja y gráficos de barras que permiten visualizar la forma en que se distribuyen las principales variables, identificar patrones de comportamiento y analizar diferencias entre los clientes clasificados.

En la subsección 4.1.5 se identifican las relaciones estadísticas que existen entre los distintos factores financieros y la variable objetivo. Empleando la matriz de correlación y las distintas medidas de asociación, se identifican las variables con mayor influencia sobre el desempeño crediticio. Este análisis es un paso fundamental para la selección de variables en el modelado predictivo.

Por último, la subsección 4.1.6 resume los hallazgos más importantes obtenidos durante el análisis exploratorio de datos. En esta parte se combinan los resultados descriptivos, junto con gráficos y correlacionales, resaltando factores con mayor impacto estadístico en la predicción del riesgo crediticio.

La sección 4.2 presenta exhaustivamente el procedimiento metodológico empleado para la obtención, tratamiento y validación de los datos utilizados en el modelo de calificación de riesgo crediticio, incorporando los elementos técnicos, operativos y éticos que formaron parte de su elaboración. Adicionalmente, se estructura en cinco apartados que detallan cada fase del procedimiento en orden cronológico, sirviendo como fundamento para las subsecuentes divisiones.

La subsección 4.2.1 describe el proceso de recopilación y procesamiento de datos, estructurado según la metodología CRISP-DM. Se detallan las etapas, recursos, responsabilidades y el cronograma implementado durante la totalidad del proceso.

En la subsección 4.2.2 se presentan los participantes y las fuentes de información, se describen los criterios de selección, el tipo de información recopilada y el flujo ETL diseñado para extraer los datos.

La subsección 4.2.3 especifica las herramientas tecnológicas empleadas para la extracción y depuración de los datos, incluyendo SQL Server y Knime, las cuales contribuyeron a la

automatización del proceso ETL.

La subsección 4.2.4 analiza las dificultades halladas durante la recopilación y preparación de los datos, tanto de naturaleza técnica como institucional. Se explican los incidentes principales, tales como la fragmentación de la información y las restricciones de confidencialidad.

Y por último la subsección 4.2.5 expone las consideraciones éticas aplicadas en el manejo de la información, destacando los protocolos de protección y uso responsable de los datos.

Las subsecciones 4.3.2.1 a 4.3.2.3 presentan, de manera sucesiva, la identificación de categorías emergentes, su verificación estadística a través de la prueba de Chi-cuadrado, y la vinculación teórica de los hallazgos con los fundamentos conceptuales de la investigación.

En la subsección 4.3.2.4 se efectúa la integración de los resultados obtenidos de los análisis cuantitativo y cualitativo, mediante una triangulación que revela las concordancias y las discrepancias entre ambos enfoques.

4.1 ANÁLISIS EXPLORATORIO DE DATOS (EDA)

En esta sección se presenta el análisis exploratorio de datos desarrollado sobre la base construida a partir del data warehouse institucional del banco x.

El propósito principal es caracterizar estadísticamente la información, identificar posibles inconsistencias o patrones anómalos, y reconocer las relaciones iniciales entre las variables independientes y la variable objetivo.

Los resultados derivados de este proceso proporcionan la base empírica necesaria para definir los subconjuntos de entrenamiento, validación y prueba, así como para seleccionar las variables de mayor relevancia en la construcción del modelo.

4.1.1 DESCRIPCIÓN GENERAL DEL CONJUNTO DE DATOS

4.1.1.1 TAMAÑO DE LA MUESTRA

El conjunto de datos analizado corresponde a los registros crediticios empresariales proporcionados por la institución financiera el banco x.

La base comprende 6,903 observaciones y 42 variables que describen las características financieras, demográficas y operativas de los clientes empresariales atendidos durante el período enero de 2020 a octubre de 2024 (ver anexo técnico J):

Tabla 14: Composición del Dataset

Métrica	Valor
Nº de Observaciones	6,903
Nº de Variables (Total)	42
Variables Numéricas (int/float)	27
Variables Categóricas (object/category)	12
Variables de Fecha (datetime)	3

Fuente: Elaboración propia en Python

El conjunto de variables describe del perfil de riesgo de los clientes. Estas variables se clasifican en cinco categorías principales:

- Descriptivas: Identifican el tipo de registro y el segmento de tiempo como mes de observación e id del registro.
- Financieras: montos de desembolso, exposición total, garantías y condiciones de cobertura.
- Conductuales: histórico de pagos, frecuencia transaccional y comportamiento crediticio reciente.
- Demográficas y empresariales: tipo de persona, sector económico y tipo de empresa.
- Regulatorias: categoría CNBS y puntajes de buró.

4.1.2 LIMPIEZA Y PREPARACIÓN DE LOS DATOS

Esta fase tiene por finalidad garantizar la integridad, consistencia y utilidad analítica del conjunto de datos antes del modelado. Se aplicaron procedimientos sistemáticos de control de calidad, tratamiento de valores atípicos, manejo de datos faltantes, estandarización de formatos y preparación de variables, siguiendo la metodología CRISP–DM y los principios de agregación y calidad de datos establecidos por (BCBS, 2010).

4.1.2.1 DETECCIÓN Y MANEJO DE VALORES ATÍPICOS

Los valores atípicos fueron identificados mediante el método del rango intercuartílico (IQR) como se evidencia en el anexo técnico K. Esta es una medida estadística que indica la dispersión de la mitad central de un conjunto de datos y que es útil para identificar valores atípicos (BCBS, 2010). Aplicado a las variables:

- MONTO_DESEMBOLSADO
- MONTO_GARANTIA
- MONTO_EXPOSICION_TOTAL
- COBERTURA_GARANTIA
- EDAD
- BURO_SCORE_EQUIFAX
- AVG_PAGOS_6M
- AVG_CUENTAS_6M.

Se utilizó capping híbrido (mínimo entre umbrales IQR y percentiles configurados por variable), alineando el tratamiento con el criterio de medición de outliers y evitando recortes excesivos. Los parámetros quedaron registrados para trazabilidad, conforme a buenas prácticas de agregación y calidad de datos de riesgo (BCBS, 2010) y a lineamientos de gestión del riesgo de modelo (SR 11-7, 2011).

Se preservó la mediana y el orden relativo de las observaciones; no se eliminaron filas, lo que favorece reproducibilidad y auditoría.

Tabla 15: Resumen de Valores Atípicos Detectados

	Variable	Q1	Q3	IQR	Límite Inferior	Límite Superior	Outliers (n)	Outliers (%)
0	BURO_SCORE_EQUIFAX	696.25	880.00	183.75	420.62	1,155.62	1,107	16.04
1	MONTO_GARANTIA	0.00	2,800,000.00	2,800,000.00	-4,200,000.00	7,000,000.00	1,102	15.96
2	AVG_PAGOS_6M	24,708.82	315,379.29	290,670.48	-411,296.90	751,385.01	1,010	14.63
3	AVG_CUENTAS_6M	48,000.43	1,021,780.26	973,779.83	-1,412,669.31	2,482,450.01	819	11.86
4	MONTO_DESEMBOLSADO	600,000.00	3,100,000.00	2,500,000.00	-3,150,000.00	6,850,000.00	809	11.72
5	MONTO_EXPOSICION_TOTAL	604,625.01	3,268,258.98	2,663,633.98	-3,390,825.96	7,263,709.96	793	11.49
6	EDAD	38.00	54.00	16.00	14.00	78.00	43	0.62
7	COBERTURA_GARANTIA	0.00	1.00	1.00	-1.50	2.50	0	0.00

Fuente: Elaboración Propia en Python

MONTO_DESEMBOLSADO: Distribución fuertemente asimétrica a la derecha, con pocos créditos de gran tamaño y mayoría en montos bajos/medios. Alta dispersión y cola larga.

Tratamiento: Capping híbrido con p2 - p98 y k=1.5. Winsorización bilateral conservando la mediana.

Justificación: Limita la influencia de desembolsos excepcionales que distorsionan métricas y modelos lineales, sin eliminar observaciones válidas de negocio. Práctica robusta recomendada en datos financieros con colas largas (Montgomery et al., 1990; Thomas et al., 2017).

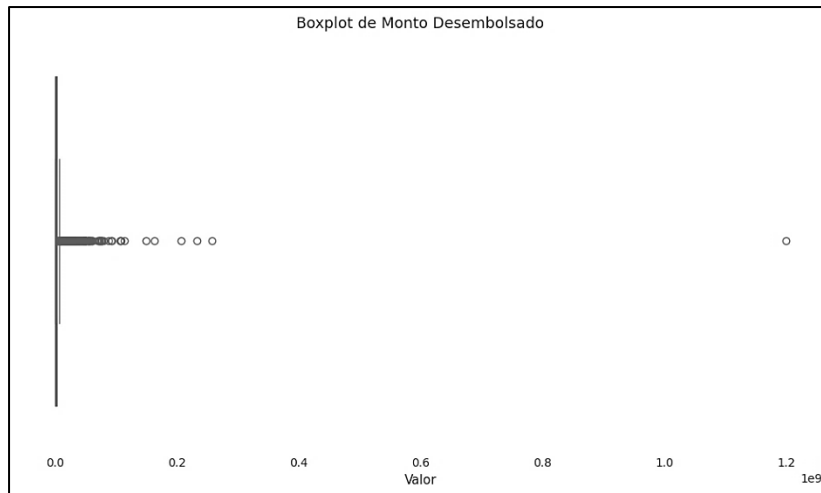


Figura 4: Boxplot Monto Desembolsado

Fuente: Elaboración Propia en Python

MONTO_GARANTIA: Comportamiento análogo a desembolso y alta correlación con exposición; presencia de operaciones con garantías muy elevadas frente a la masa central.

Tratamiento: Capping híbrido p2 - p98, k=1.5 (winsorización bilateral). Este tratamiento reduce la varianza inducida por garantías extraordinarias manteniendo la estructura muestral y la relación con el monto otorgado. Evita que pocos valores extremos dominen el aprendizaje en el modelo.

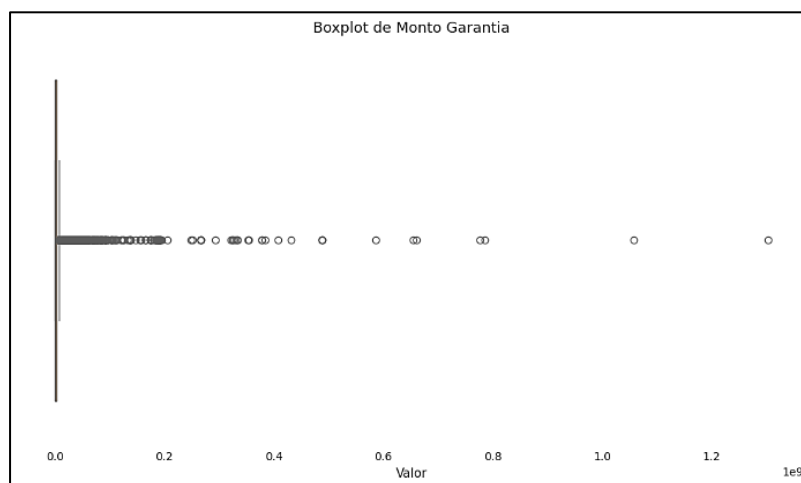


Figura 5: Boxplot Monto Garantía

Fuente: Elaboración propia

MONTO_EXPOSICION_TOTAL: Variable agregada con la cola más extrema; algunos clientes concentran exposiciones inusualmente altas.

Tratamiento: Capping híbrido p2 - p98 (o p1 - p99 según diagnóstico), $k=1.5$. La exposición total es clave en PD; se requiere contención de extremos para estabilizar sin perder señal en la parte alta. El enfoque híbrido alinea el recorte con el criterio IQR y protege contra truncamientos excesivos por percentiles fijos.

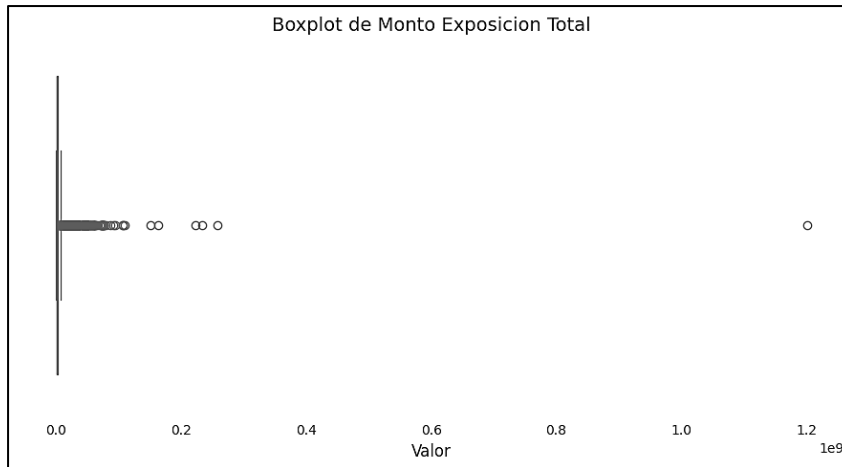


Figura 6: Boxplot Exposición Total

Fuente: Elaboración Propia en Python

COBERTURA_GARANTIA: Distribución bimodal con masas en 0% y 100%, y menor frecuencia en rangos intermedios (ver ilustración 7). Posible mezcla de “sin garantía y NA” con cobertura real 0%.

Tratamiento: Reglas previas de negocio (reemplazar 0 por NaN cuando represente “no aplica”) y capping híbrido p1 - p99, $k=1.5$. El capping suaviza intermedios sin afectar la masa en 100%. Mejora la comparabilidad y la robustez del modelo ante colas y valores codificados.

EDAD: Distribución acotada, próxima a normal en 35 a 50 años (ver ilustración 8); outliers asociados a errores de captura (edades improbables).

Tratamiento: Capping suave p1 - p99, $k=1.5$. Solo corrige entradas anómalas preservando

la forma natural de la variable. Evita que registros espurios afecten métricas y

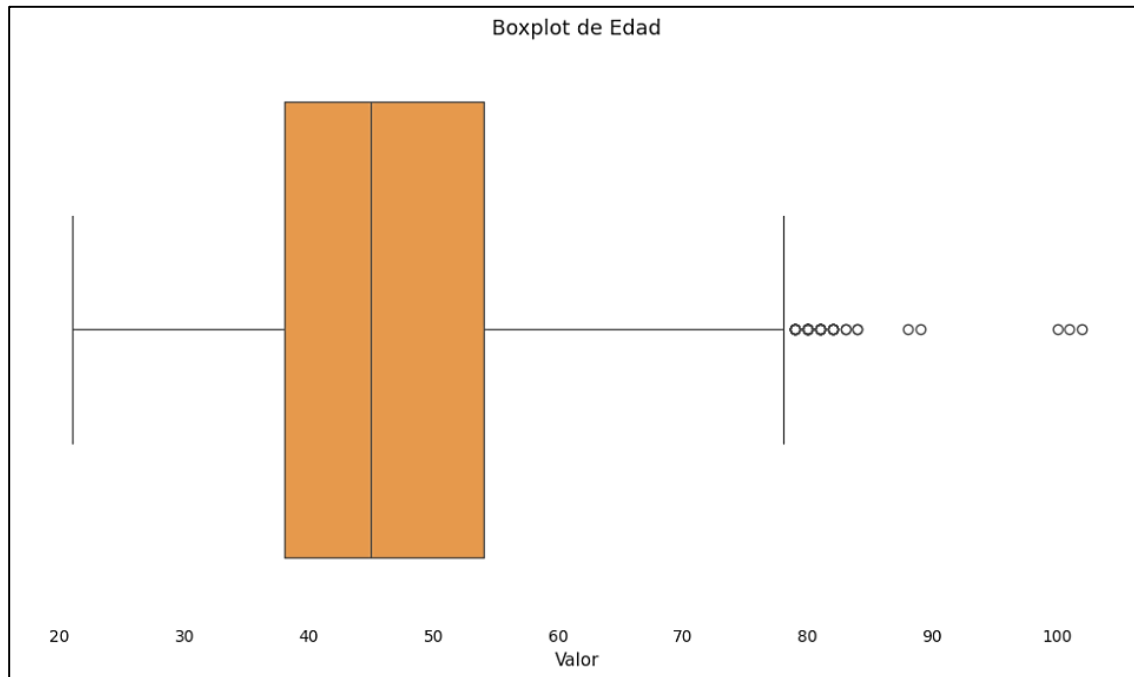


Figura 7: Boxplot Edad

Fuente: Elaboración Propia en Python

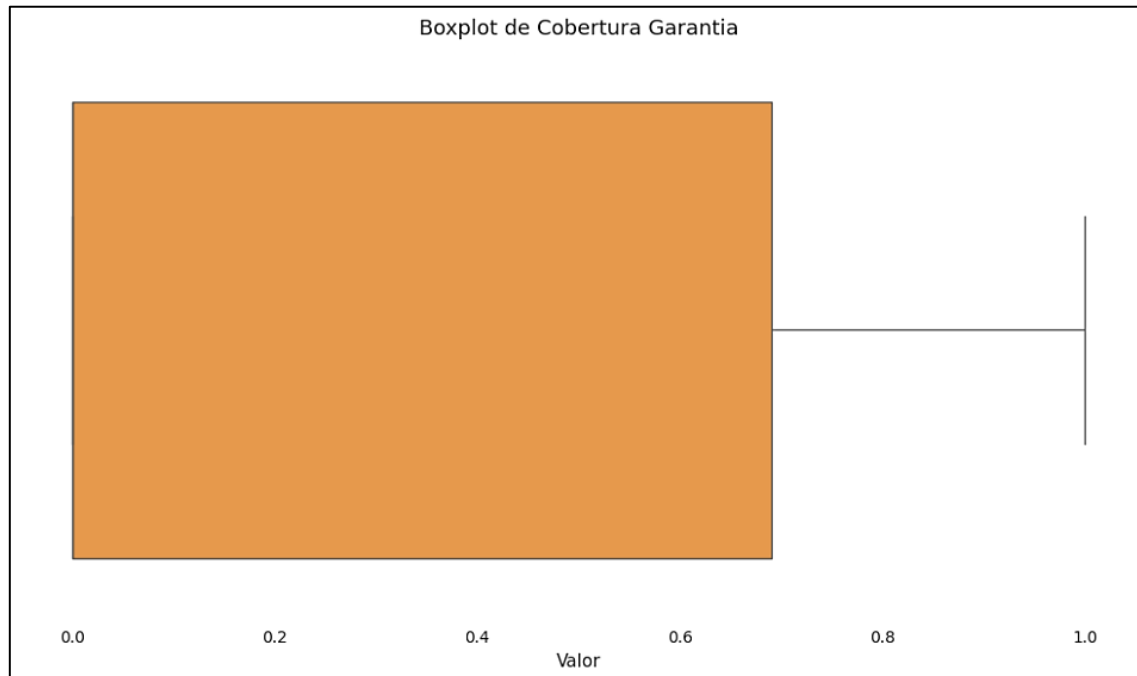


Figura 8: Boxplot Cobertura de Garantía

Fuente: Elaboración Propia en Python

segmentaciones.

BURO_SCORE_EQUIFAX: Concentración principal en 800–900 puntos y un pico en 0 (posibles clientes sin historial crediticio). Valores discretos y colas menos relevantes en el rango alto.

Tratamiento: No capear el 0 (mantenerlo como valor informativo). Aplicar capping híbrido $p1 - p99, k=1.5$ sobre los registros con $score > 0$.

La decisión se justifica en que el 0 no es un extremo numérico sino una señal de ausencia de historial; capearlo distorsionaría su significado. El capping sobre > 0 protege de puntajes atípicos sin perder capacidad discriminante del score.

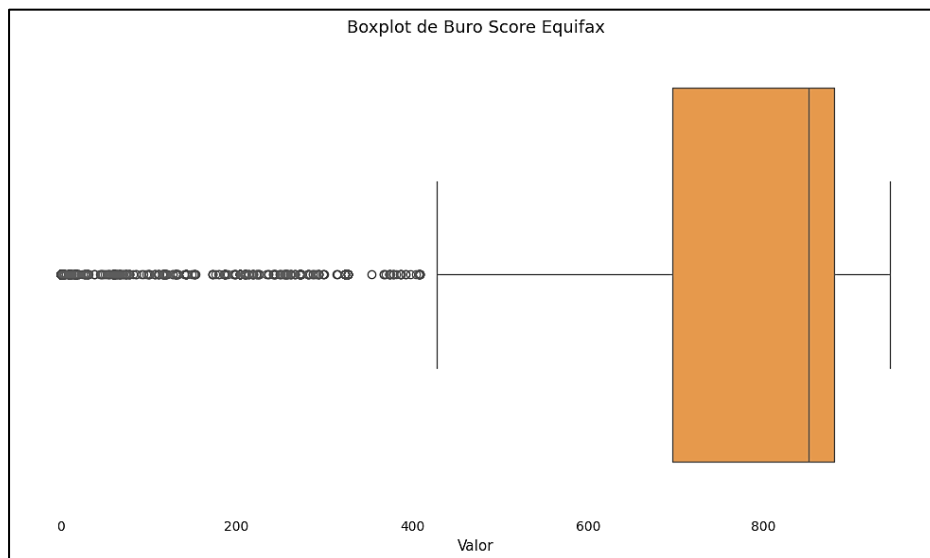


Figura 8: Boxplot Buró Equifax

Nota: Las densidades y frecuencias no son directamente comparables visualmente entre subgráficas debido a la disparidad en magnitudes.

Fuente: Elaboración Propia en Python

AVG_PAGOS_6M: Asimetría extrema a la derecha; mayoría con valores bajos o nulos y pocos casos con montos promedio extraordinarios.

Tratamiento: Capping híbrido $p2 - p98, k=1.5$. Reduce la influencia de casos raros que inflan la media y la desviación. Mantiene la mediana y facilita la convergencia de modelos sensibles a escala.

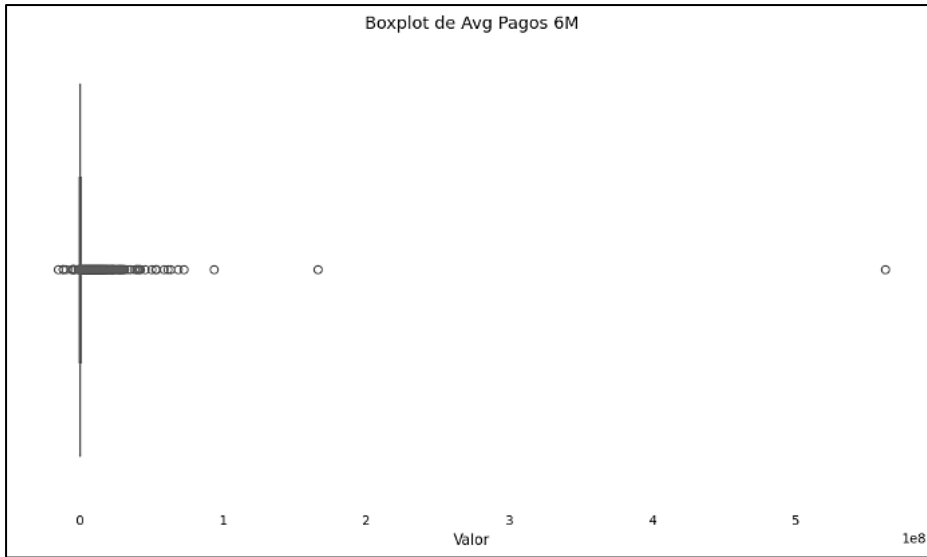


Figura 9: Boxplot AVG Pagos 6M

Fuente: Elaboración Propia en Python

AVG_CUENTAS_6M: - gran concentración de valores bajos y cola larga de saldos promedio muy altos.

Tratamiento: Capping híbrido p2 - p98, k=1.5. La decisión se justifica en robustecer métricas y reducir el peso de outliers, preservando información útil mediante winsorización conservadora y transformaciones opcionales.

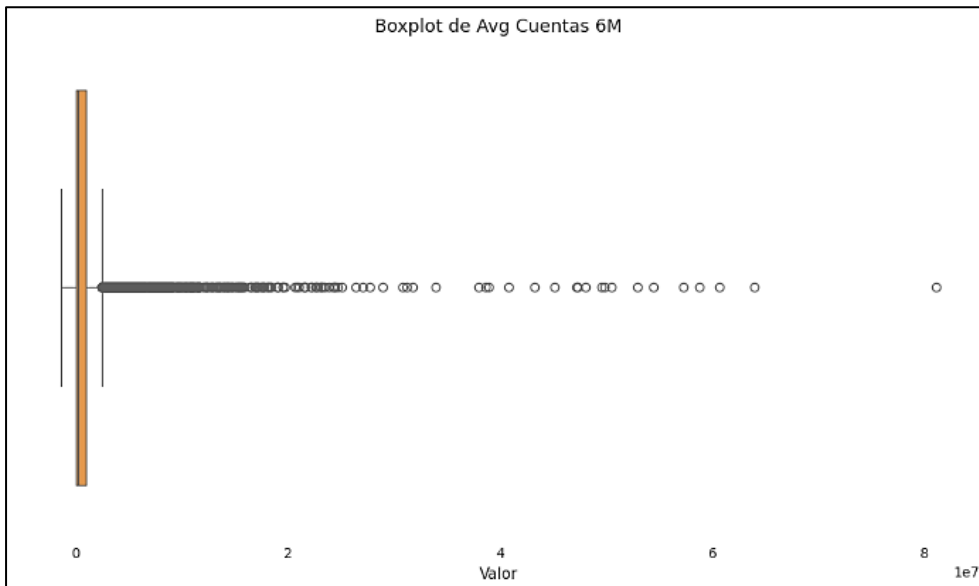


Figura 10: Boxplot AVG Cuentas 6M

Fuente: Elaboración Propia en Python

Lo que se espera con este tratamiento es la reducción tangible del porcentaje de outliers bajo el criterio IQR, contracción de colas, disminución de la desviación estándar y preservación de medianas en todas las variables tratadas, habilitando un modelado más estable y menos sensible a valores extremos sin pérdida de señal de negocio.

Esta técnica conserva la estructura estadística del conjunto de datos, reduciendo la influencia de observaciones atípicas sin alterar significativamente las medias o desviaciones estándar.

Una vez aplicada la técnica de winsorización en los datos (ver anexo técnico L), el porcentaje de reducción de outliers se puede observar en la tabla siguiente:

Tabla 16: Porcentaje de Outliers Antes y Después

N	Variable	Outliers IQR % (Antes)	Outliers IQR % (Después)	Reducción pp
0	MONTO_DESEMBOLSADO	11.72	0.00	11.72
1	MONTO_GARANTIA	15.96	0.00	15.96
2	MONTO_EXPOSICION_TOTAL	11.49	0.00	11.49
3	COBERTURA_GARANTIA	0.00	0.00	0.00
4	EDAD	0.62	0.00	0.62
5	AVG_PAGOS_6M	14.63	0.00	14.63
6	AVG_CUENTAS_6M	11.86	0.00	11.86

Fuente: Elaboración Propia en Python

4.1.2.2 TRATAMIENTO DE VALORES FALTANTES

Se verificó la presencia de valores faltantes y registros inconsistentes en variables tanto numéricas como categóricas (ver anexo técnico M). Se cuantificó nulos por columna y su porcentaje y se identificaron faltantes estructurales, operativos y aleatorios.

La estrategia definida para el tratamiento de valores faltantes fue la siguiente:

- Para variables numéricas: imputación robusta con mediana (menos sensible a colas largas) y registro del estadístico para reproducibilidad.
- Para variables categóricas: imputación con etiqueta explícita "No especificado" o "Sin Información", preservando tamaño muestral y trazabilidad.
- Para variables de fechas: imputación contextual o mantener NaN y derivar variables a partir

de fechas válidas (años, antigüedad).

Tabla 17: Diagnóstico de Valores Faltantes

N	Columna	N_faltantes	%_faltantes	Tipo
0	DESC_DESTINO	3526	51.08	object
1	CODIGO_DESTINO	3394	49.17	float64
2	BURO_SCORE_EQUIFAX	1313	19.02	float64
3	DESC_PROFESION	483	7.00	object
4	CODIGO_PROFESION	483	7.00	float64
5	CODIGO_SECTOR	378	5.48	float64
6	TIPO_GARANTIA	191	2.77	float64
7	AGENCIA_ORIGEN	170	2.46	float64
8	AVG_CUENTAS_6M	111	1.61	float64
9	AVG_PAGOS_6M	97	1.41	float64
10	DESC_OCUPACION	78	1.13	object
11	CODIGO_OCUPACION	78	1.13	float64
12	NIVEL_EDUCATIVO	56	0.81	object
13	FECHA_DE_NACIMIENTO	55	0.80	object
14	EDAD	55	0.80	float64
15	ESTADO_CIVIL	54	0.78	object
16	SEXO	54	0.78	object
17	MONTHS_SINCE_OLDEST	4	0.06	float64

Fuente: Elaboración Propia en Python

4.1.2.3 NORMALIZACIÓN Y ESTANDARIZACIÓN DE VARIABLES

La normalización y estandarización buscan hacer comparables las escalas de los predictores numéricos, estabilizar la convergencia de modelos sensibles a magnitudes y reducir la influencia de dispersiones heterogéneas tras el tratamiento de outliers. En esta investigación se aplicaron transformaciones diferenciadas por tipo de variable y algoritmo, manteniendo la interpretabilidad.

Debido a que las variables financieras se expresan en distintas fue necesario aplicar un proceso de estandarización (z-score normalization) para garantizar comparabilidad y estabilidad numérica durante la modelización.

El procedimiento consistió en transformar cada variable X_i a una escala con media = 0 y

desviación estándar = 1, utilizando la siguiente fórmula:

$$Z_i = \frac{X_i - \mu_i}{\Sigma_i}$$

La estandarización z-score mejora la estabilidad numérica y la comparabilidad de coeficientes en modelos lineales y regularizados, reduciendo la sensibilidad a diferencias de escala (Montgomery et al., 1990).

Esta técnica, implementada mediante la clase StandardScaler de scikit-learn, permitió que todos los predictores tuvieran igual peso relativo en los algoritmos estadísticos y de aprendizaje automático, mejorando la convergencia y reduciendo la sensibilidad a la magnitud de las unidades.

Tabla 18: Variables Normalizadas y Estandarizadas

VARIABLE	MEAN	STD
z_MONTO_DESEMBOLSADO	-0.00	1.00
z_MONTO_GARANTIA	0.00	1.00
z_MONTO_EXPOSICION_TOTAL	0.00	1.00
z_AVG_PAGOS_6M	0.00	1.00
z_AVG_CUENTAS_6M	0.00	1.00
z_BURO_SCORE_EQUIFAX	-0.00	1.00
z_EDAD	-0.00	1.00
COBERTURA_GARANTIA	0.88	0.14

Fuente: Elaboración Propia en Python

4.1.3 ANÁLISIS Y DESCRIPCIÓN DE VARIABLES

La base contiene 42 variables que permiten describir cuantitativamente la exposición crediticia y cualitativamente el perfil del deudor. Existen en la base 27 variables cuantitativas y 14 variables cualitativas. El tipo de datos que contiene son:

Tabla 19: Tipo de Datos

#	Tipo de Datos	Cantidad de Variables
1	float64	16
2	int64	11
3	object	12

#	Tipo de Datos	Cantidad de Variables
4	datetime64	3

Fuente: Elaboración Propia

El análisis de la estructura del dataset, presentado en la tabla 14, permite identificar la composición y calidad de la información utilizada en el estudio. La mayoría de las columnas presentan altos conteos de registros no nulos y porcentajes de ausencia bajos. Las variables con porcentajes de nulos altos se identifican y orientan acciones de tratamiento (imputación, exclusión o documentación de limitaciones) en la etapa de limpieza.

Fuente: Elaboración Propia en Python

El conteo de valores únicos por columna identifica un nivel de granularidad adecuado.

Tabla 20: Información de Variables del Data Set

N	Columna	Tipo	No_nulos	Nulos	% Nulos	Únicos
0	ID	int64	6903	0	0.00	3856
1	CODIGO_GRUPO_ECONOMICO	int64	6903	0	0.00	135
2	MES_COSECHA	datetime64[ns]	6903	0	0.00	53
3	DIAS_MORA_12M	int64	6903	0	0.00	210
4	FECHA_APERTURA	datetime64[ns]	6903	0	0.00	1125
5	TIPO_GARANTIA	float64	6712	191	2.77	11
6	MONTO_GARANTIA	float64	6903	0	0.00	2249
7	ES_HIPOTECA	object	6903	0	0.00	2
8	ES_REFINANCIAMIENTO	int64	6903	0	0.00	2
9	COBERTURA_GARANTIA	float64	6903	0	0.00	229
10	MEAN_GARANTIA	float64	6903	0	0.00	2327
11	DESTINO	int64	6903	0	0.00	108
12	CATEGORIA_CNBS	int64	6903	0	0.00	5
13	MONTO_DESEMBOLSADO	float64	6903	0	0.00	3257
14	TIPO_CREDITO	int64	6903	0	0.00	3
15	ACTIVIDAD_ECONOMICA	int64	6903	0	0.00	15
16	DEUDOR	object	6903	0	0.00	3
17	BUENO_MALO	int64	6903	0	0.00	2
18	CODIGO_ACTIVIDAD_ECONOMICA	int64	6903	0	0.00	91

N	Columna	Tipo	No_nulos	Nulos	% Nulos	Únicos
19	DESC_ACTIVIDAD_ECONOMICA	object	6903	0	0.00	86
20	TIPO_DE_PERSONA	object	6903	0	0.00	2
21	SEXO	object	6849	54	0.78	3
22	FECHA_DE_NACIMIENTO	datetime64[ns]	6848	55	0.80	3299
23	ESTADO_CIVIL	object	6849	54	0.78	6
24	CODIGO_PROFESION	float64	6420	483	7.00	89
25	DESC_PROFESION	object	6420	483	7.00	88
26	CODIGO_OCUPACION	float64	6825	78	1.13	12
27	DESC_OCUPACION	object	6825	78	1.13	12
28	NIVEL_EDUCATIVO	object	6847	56	0.81	6
29	CODIGO_DESTINO	float64	3509	3394	49.17	89
30	DESC_DESTINO	object	3377	3526	51.08	86
31	CODIGO_SECTOR	float64	6525	378	5.48	9
32	DESC_SECTOR	object	6903	0	0.00	6
33	MONTO_EXPOSICION_TOTAL	float64	6903	0	0.00	6300
34	OPERACIONES	int64	6903	0	0.00	19
35	EDAD	float64	6848	55	0.80	70
36	AVG_CUENTAS_6M	float64	6792	111	1.61	6739
37	MONTHS_SINCE_OLDEST	float64	6899	4	0.06	308
38	BURO_SCORE_EQUIFAX	float64	5590	1313	19.02	450
39	AVG_PAGOS_6M	float64	6806	97	1.41	6783
40	AGENCIA_ORIGEN	float64	6733	170	2.46	101
41	ZONA	object	6903	0	0.00	4

Fuente: Elaboración Propia en Python

VARIABLES CON ALTA CARDINALIDAD APORTAN DETALLE SEGMENTAL, MIENTRAS QUE CAMPOS CATEGÓRICOS ACOTADOS SON ADECUADOS PARA CODIFICACIÓN Y ANÁLISIS COMPARATIVO. LA COEXISTENCIA DE AMBOS TIPOS PERMITE BALANCEAR PRECISIÓN EN EL MODELADO. LAS ÁREAS CON MAYORES PORCENTAJES DE NULOS SE ENCUENTRAN ACOTADAS Y SON GESTIONABLES MEDIANTE POLÍTICAS DE TRATAMIENTO DOCUMENTADAS (IMPUTACIÓN, WINSORIZACIÓN Y ESTANDARIZACIÓN), SIN QUE LA VALIDEZ DEL CONJUNTO DE DATOS SE VEA COMPROMETIDO.

4.1.3.1 CARACTERÍSTICAS BÁSICAS DE LOS DATOS CUANTITATIVOS

Tabla 21: Análisis Descriptivo de Variables Cuantitativas

Variable	N	Media	Desv.Est.	Mínimo	25%	50%	75%	Máximo	Rango
ID	6,903	3,092,956.92	2,212,757.01	844.00	1,177,541.50	2,545,836.00	5,256,651.50	6,953,544.00	6,952,700.00
CODIGO_GRUPO_ECONOMICO	6,903	9,066.45	1,708.79	1,001.00	9,500.00	9,500.00	9,500.00	9,562.00	8,561.00
DIAS_MORA_12M	6,903	13.60	42.49	0.00	0.00	0.00	0.00	376.00	376.00
TIPO_GARANTIA	6,712	10.20	3.23	1.00	9.00	9.00	13.00	17.00	16.00
MONTO_GARANTIA	6,903	8,053,074.19	39,172,681.04	0.00	0.00	600,000.00	2,800,000.00	1,303,390,636.80	1,303,390,636.80
ES_REFINANCIAMIENTO	6,903	0.05	0.21	0.00	0.00	0.00	0.00	1.00	1.00
COBERTURA_GARANTIA	6,903	0.57	0.42	0.00	0.00	0.75	1.00	1.00	1.00
MEAN_GARANTIA	6,903	6,725,034.76	27,013,698.83	0.00	0.00	703,950.00	3,000,000.00	488,802,600.00	488,802,600.00
DESTINO	6,903	120,025.58	24,878.94	10,100.00	101,050.00	130,104.00	130,206.00	180,100.00	170,000.00
CATEGORIA_CNBS	6,903	1.08	0.39	1.00	1.00	1.00	1.00	5.00	4.00
MONTO_DESEMBOLSADO	6,903	3,804,048.19	17,057,675.87	15,000.00	600,000.00	1,200,000.00	3,100,000.00	1,199,950,170.00	1,199,935,170.00
TIPO_CREDITO	6,903	1.23	0.43	1.00	1.00	1.00	1.00	3.00	2.00
ACTIVIDAD_ECONOMICA	6,903	212.12	29.42	100.00	190.00	220.00	230.00	300.00	200.00
BUENO_MALO	6,903	0.07	0.25	0.00	0.00	0.00	0.00	1.00	1.00
CODIGO_ACTIVIDAD_ECONOMICA	6,903	212.57	226.99	1.00	100.00	190.00	190.00	999.00	998.00
CODIGO_PROFESION	6,420	64.72	65.60	1.00	8.00	51.00	69.00	212.00	211.00
CODIGO_OCUPACION	6,825	2.13	1.28	1.00	1.00	3.00	3.00	13.00	12.00
CODIGO_DESTINO	3,509	13,759,882.15	7,837,712.49	19.00	10,001,000.00	19,008,000.00	19,008,000.00	30,701,100.00	30,701,081.00
CODIGO_SECTOR	6,525	3.19	1.31	0.00	2.00	4.00	4.00	12.00	12.00
MONTO_EXPOSICION_TOTAL	6,903	3,907,239.24	17,130,048.79	15,093.28	604,625.01	1,204,500.00	3,268,258.98	1,200,936,266.64	1,200,921,173.36
OPERACIONES	6,903	1.27	1.08	1.00	1.00	1.00	1.00	46.00	45.00

Variable	N	Media	Desv.Est.	Mínimo	25%	50%	75%	Máximo	Rango
EDAD	6,848	46.34	11.55	21.00	38.00	45.00	54.00	102.00	81.00
AVG_CUENTAS_6M	6,792	1,317,639.20	3,845,886.06	-1,358,960.32	48,000.43	246,656.63	1,021,780.26	81,061,652.05	82,420,612.37
MONTHS_SINCE_OLDEST	6,899	237.37	495.70	0.00	0.00	0.00	123.00	1,506.00	1,506.00
BURO_SCORE_EQUIFAX	5,590	686.65	327.28	0.00	696.25	851.00	880.00	944.00	944.00
AVG_PAGOS_6M	6,806	899,885.17	7,974,217.12	-15,182,329.78	24,708.82	76,841.59	315,379.29	562,528,799.26	577,711,129.05
AGENCIA_ORIGEN	6,733	127.97	129.06	1.00	21.00	60.00	253.00	1,159.00	1,158.00

Fuente: Elaboración Propia en Python (evidencia en anexo técnico O)

Los resultados obtenidos a partir de las medidas de tendencia central, dispersión y forma se resumen a continuación:

Los histogramas presentados a continuación permiten observar de forma gráfica la distribución de las principales variables financieras y conductuales del portafolio crediticio.

Las variables MONTO_DESEMBOLSADO, MONTO_GARANTÍA y MONTO_EXPOSICIÓN_TOTAL muestran distribuciones claramente asimétricas positivas, donde la mayoría de los clientes presenta valores bajos y un pequeño grupo concentra montos elevados. Esta dispersión es típica en carteras empresariales heterogéneas, donde coexisten pequeñas y medianas empresas con conglomerados corporativos de gran tamaño (Gerencia de Estudios Económicos, CNBS, 2024).

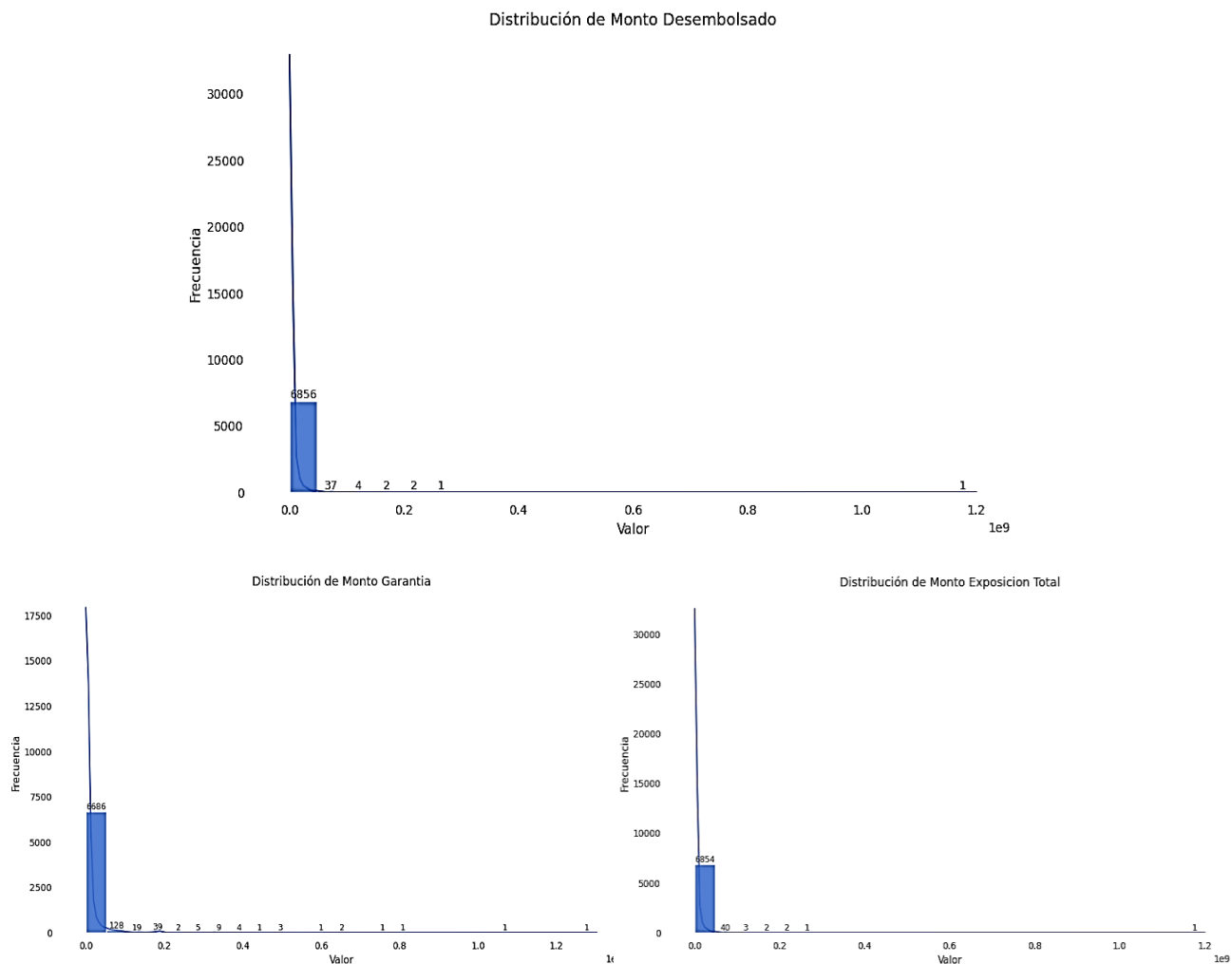


Figura 11: Grafico de Frecuencias de Variables de Monto Desembolsado, Monto de Garantía y Monto de Exposición Total

Fuente: Elaboración Propia en Python

La distribución de la variable COBERTURA_GARANTIA presenta una clara bimodalidad con concentraciones extremas en 0% y 100%, acompañadas de frecuencias menores en los rangos intermedios. Este patrón sugiere la coexistencia de dos políticas operativas bien diferenciadas: por un lado, operaciones sin respaldo colateral efectivo; por otro, operaciones totalmente cubiertas. Entre ambos extremos se observan agrupaciones secundarias en torno a coberturas medias (50% - 90%), pero con frecuencias sustancialmente inferiores.

Desde la perspectiva de riesgo, esta forma bimodal implica heterogeneidad en el apetito y en la práctica de garantías por segmento. La presencia significativa de coberturas nulas puede incrementar la pérdida esperada en escenarios de estrés si no está compensada por capacidad de pago, calidad del deudor o cláusulas contractuales.

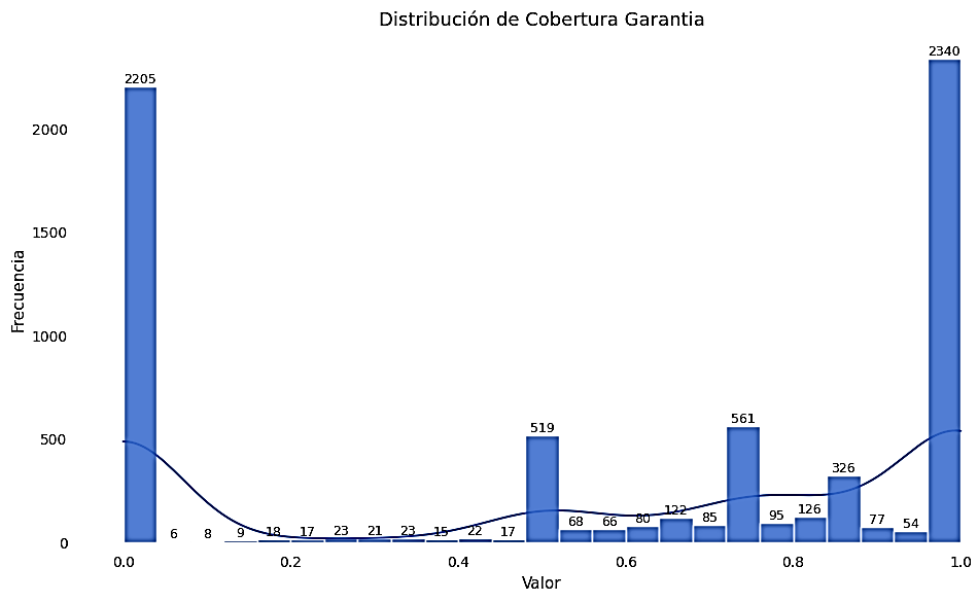


Figura 12: Gráfico de Frecuencia Cobertura Garantía

Fuente: Elaboración propia en Python

La variable EDAD se distribuye de manera casi normal, de una forma unimodal entre 35 y 50 años, segmento que representa el rango de edad más activo de la cartera empresarial, en consonancia con los perfiles de edad empresarial identificados.

BURO_SCORE_EQUIFAX tiene una distribución bimodal; la mayoría de los clientes con score aceptable. un grupo mayoritario concentrado en el rango alto de puntaje (800 - 900 puntos) y un segundo grupo acumulado en valores muy bajos, incluyendo un pico en cero y tramos dispersos por debajo de 600.

En términos de comportamiento crediticio, presenta una distribución aproximadamente normal entre 600 y 900 puntos, rango asociado con clientes de riesgo medio a bajo. Predominan clientes con perfil crediticio sólido, coherente con prácticas de originación conservadoras para la cartera empresarial. Los valores ceros se asocian a clientes sin historial crediticio.

La separación entre score bajos y altos anticipa capacidad discriminante del score para la variable objetivo: en principio, mayores puntajes se asocian con menor probabilidad de

incumplimiento, mientras que los extremos bajos requerirán reglas de negocio específicas, imputación controlada o tratamiento como categoría aparte.

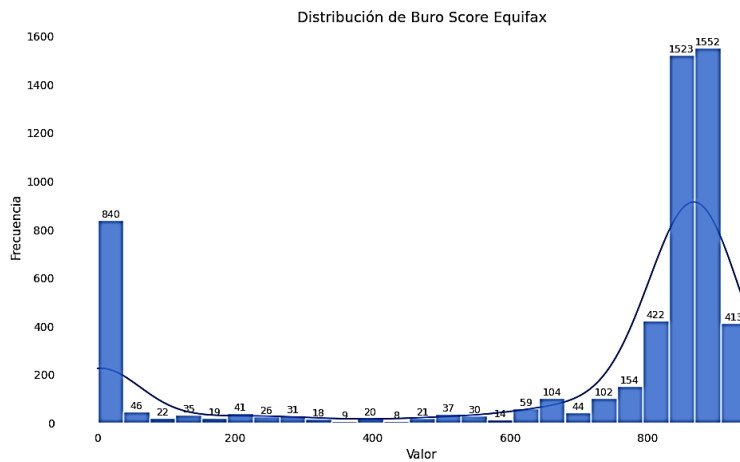


Figura 13: Gráfico de Frecuencia de Buero Score

Fuente: Elaboración propia en Python

La distribución de AVG_PAGOS_6M es asimétrica a la derecha, con una tendencia muy cercana a cero y un conjunto extremadamente reducido de valores muy altos. la gran mayoría de clientes registra promedios de pagos en 6 meses muy bajos o nulos, consistente con productos de pago poco frecuente.

Existen registros con valores desproporcionados que distorsionan la escala del gráfico y las métricas de tendencia central. Es aconsejable aplicar winsorización o truncamiento controlado para reducir la influencia de outliers en el modelado.

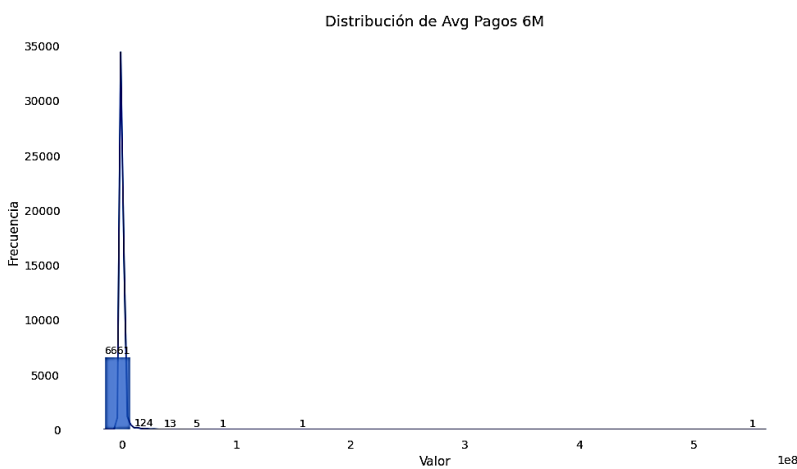


Figura 14: Gráfico de Frecuencia de AVG Pagos

Fuente: Elaboración propia en Python

Al igual que la variables AVG_PAGOS_6M la distribución del promedio de saldos en cuenta a 6 meses (AVG_CUENTAS_6M) es marcadamente asimétrica a la derecha con observaciones concentrado en valores bajos, cercano al origen y una cola larga de magnitud considerable con muy pocos casos de saldos promedio extraordinariamente altos.

La mayoría de los clientes mantiene saldos promedio reducidos y estables, coherente con estructuras de capital de trabajo ajustadas y/o uso táctico de cuentas para operaciones corrientes.

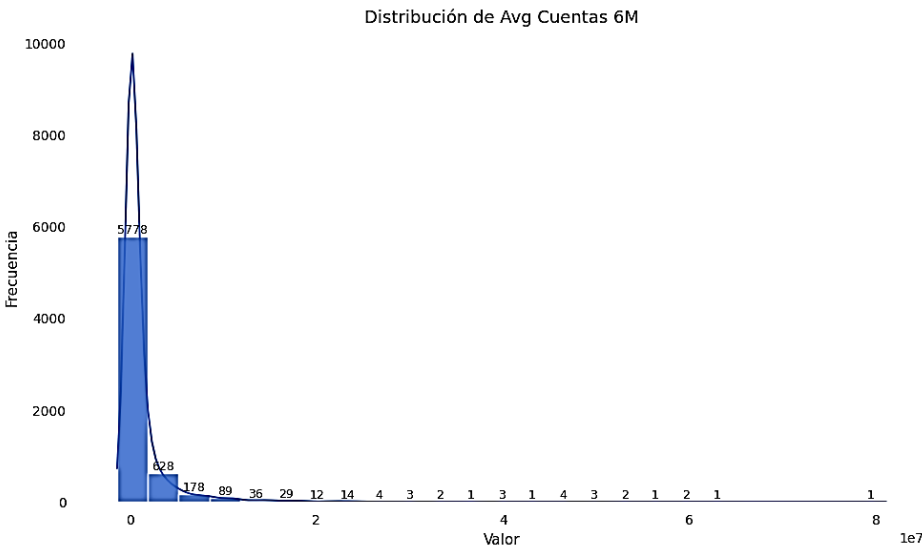


Figura 15: Gráfico de Frecuencia de AVG Saldos en Cuentas

Fuente: Elaboración propia en Python

Muestra una cartera dominada por saldos promedio modestos y una minoría con saldos excepcionalmente elevados. Su utilización en el modelo exige tratamiento robusto de outliers y transformaciones adecuadas para capturar su información sin introducir sesgos al modelo.

OPERACIONES (operaciones activas por cliente): El histograma muestra una distribución fuertemente asimétrica a la derecha, concentrada en muy pocos productos por cliente y una cola larga de casos excepcionales.

La gran mayoría de los clientes mantiene entre 1 y 3 operaciones activas. Este patrón es consistente con relaciones bancarias simples y productos orientados a capital de trabajo y/o uso corriente. Existen observaciones con números elevados de operaciones (> 9 y hasta > 40).

Estos casos corresponden, previsiblemente, a grupos económicos con múltiples líneas y renovaciones, y tienen impacto desproporcionado en métricas como la media y la desviación

estándar y, por lo tanto, se deberá considerar aplicar tratamientos para evitar que unos pocos casos extremos distorsionen el modelado.

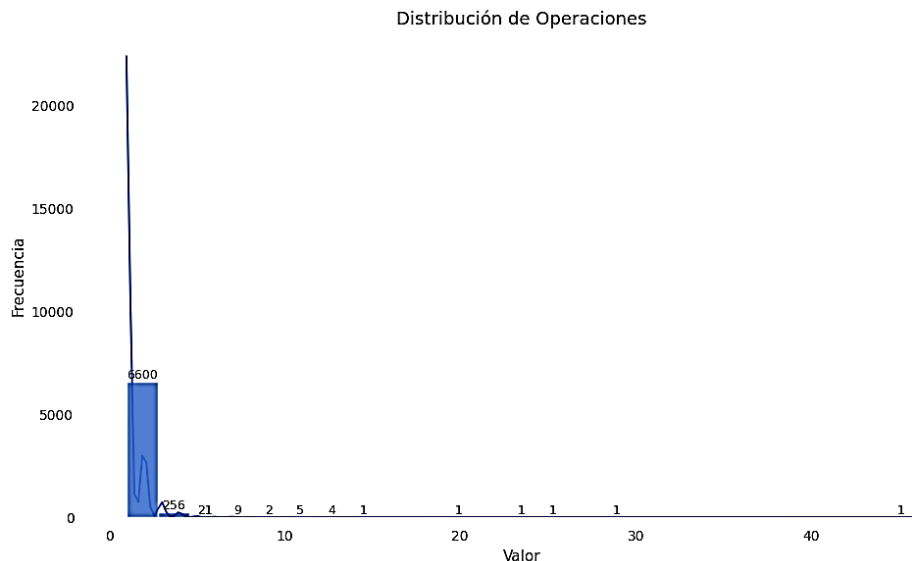


Figura 16: Gráfico de Frecuencia en Variable de Operaciones

Fuente: Elaboración Propia en Python

Las variables financieras son las que presentan más valores atípicos los cuales deberán ser tratados para efectos de normalizar la muestra

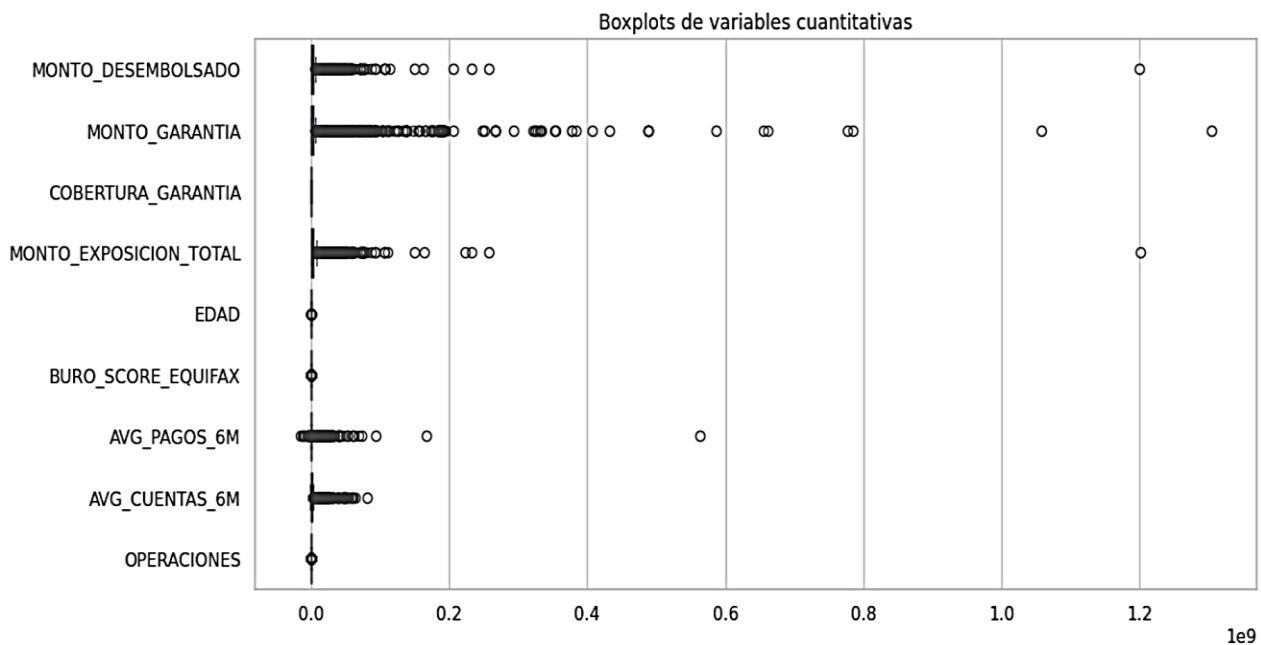


Figura 17: Gráfico de Boxplots Variables Cuantitativas

Fuente: Elaboración Propia en Python

4.1.3.2 CARACTERÍSTICAS BÁSICAS DE LOS DATOS CUALITATIVOS

El conjunto también incorpora 14 variables categóricas que describen atributos demográficos, empresariales y regulatorios. Su análisis proporcionó una perspectiva descriptiva del perfil global de los clientes en la muestra. A partir de los cuadros de frecuencia y gráficos de barras, se observan los siguientes patrones:

TIPO_CREDITO: Los productos de tipo comercial (1) tienen una proporción muy superior al resto de categorías (76.3%), seguido por los de capital de trabajo (2) con un 23.16% (ver ilustración 18).

Esta distribución presentada evidencia que la cartera empresarial está orientada a necesidades de liquidez y rotación, más que a proyectos de largo plazo. Conviene analizar el riesgo por producto al momento de realizar el modelado, dado que la naturaleza y el calendario de pagos difieren (corto vs. mediano plazo).

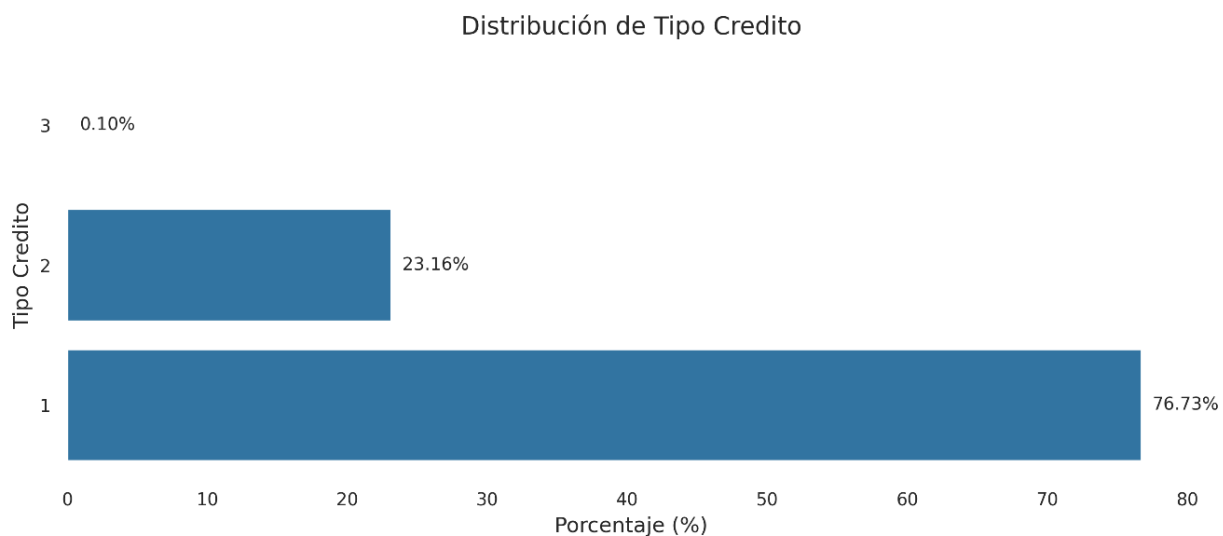


Figura 18: Distribución de Variable Tipo de Crédito

Fuente: Elaboración propia en Python

TIPO_DE_PERSONA: el 54% a personas naturales y el 46% son personas jurídicas, en el cual se muestra un enfoque empresarial pequeño de la cartera (ver ilustración 19).

La distribución de personas naturales con actividad empresarial y personas jurídicas revela una presencia de negocios individuales. Esta característica es similar con la estructura empresarial nacional CNBS (2024) e implica variaciones en formalización y documentación.

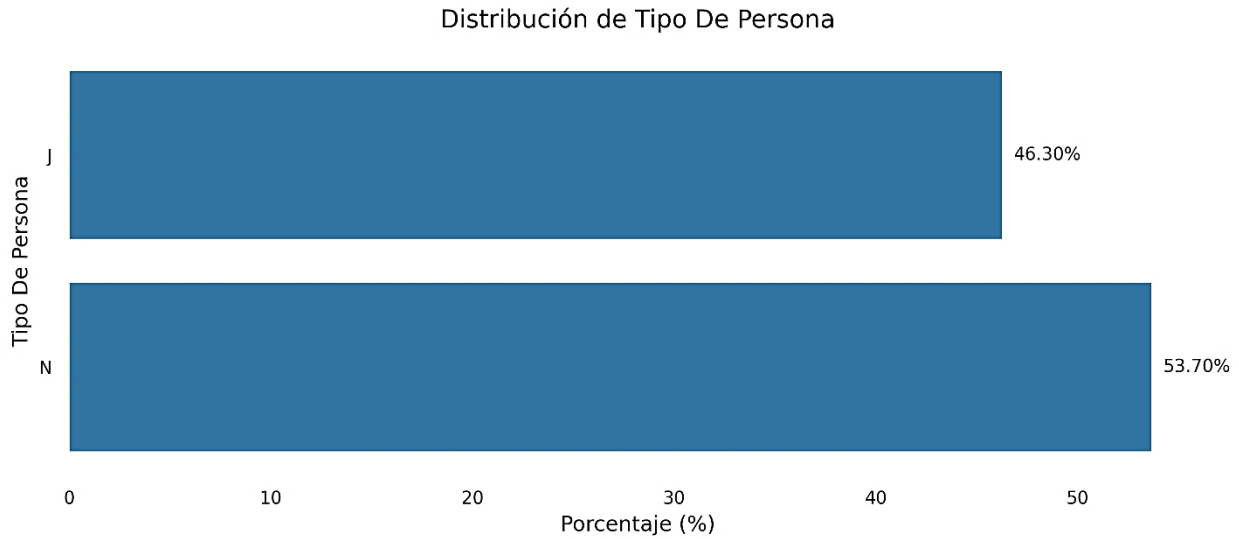


Figura 19: Gráfico Distribución Tipo de Persona

Fuente: Elaboración Propia en Python

SEXO: El 75.65% corresponde al género masculino y el 23.57% al femenino con un porcentaje menor al 1% de valores nulos. Si se considera esta variable, debe tratarse con criterios de equidad y cumplimiento normativo, evitando su uso directo como factor de decisión; su utilidad analítica reside, en el entendimiento del portafolio y en posibles sesgos de datos.

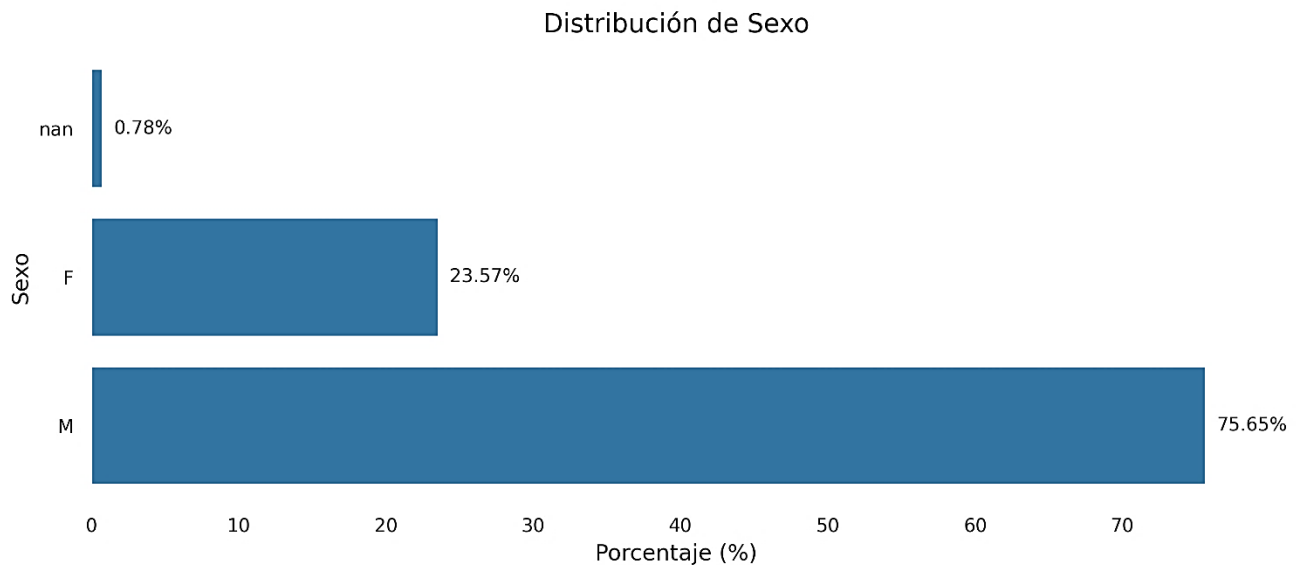


Figura 20: Gráfico de Distribución Sexo

Fuente: Elaboración Propia en Python

CATEGORIA_CNBS: La mayor parte de las operaciones están en categoría 1 (94.3%), con

porcentajes menores al 4% en categorías de riesgo medio (2) y pequeña cantidad menor al 2% en categorías de riesgo alto 3 - 5.

Este patrón indica una cartera mayoritariamente sana al corte de observación, pero también subraya el desbalance natural para problemas de predicción de incumplimiento. Será necesario aplicar técnicas de balanceo en entrenamiento (class weights o sobremuestreo sintético) y evaluar métricas robustas (ROC-AUC, KS, recall) más allá de la precisión.

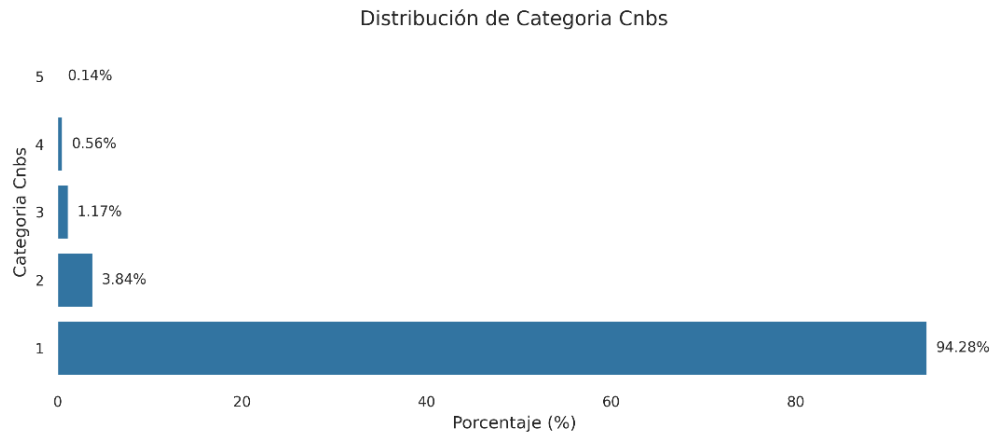


Figura 21: Gráfico Distribución Categoría CNBS

Fuente: Elaboración Propia en Python

DESC_SECTOR: Destacan sectores de comercio y servicios, seguidos por agropecuario y otros con menor participación. Esta concentración sectorial sugiere que las condiciones macro y de demanda interna influyen de manera relevante en el perfil de riesgo.

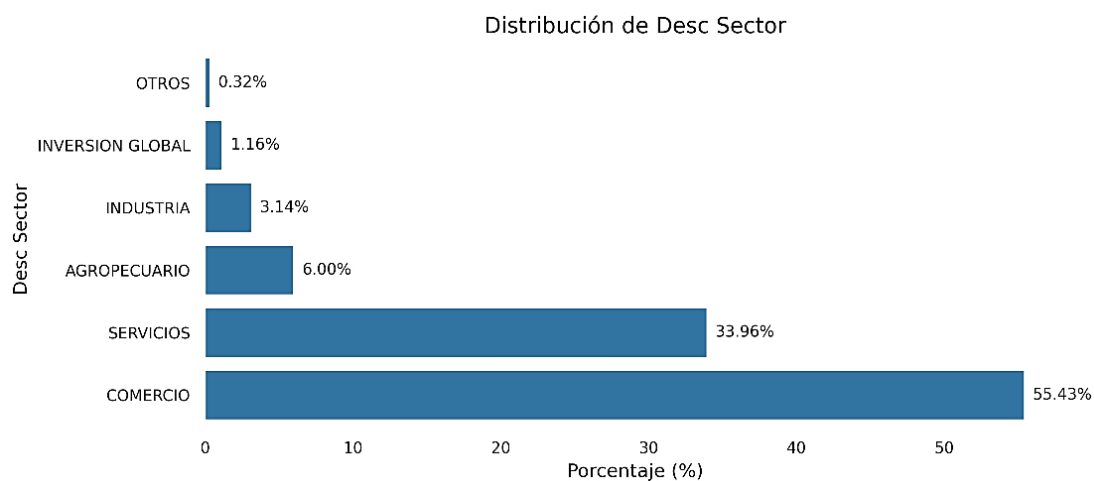


Figura 22: Gráfico Distribución Sector Económico

Fuente: Elaboración Propia en Python

4.1.4 VISUALIZACIÓN Y EXPLORACIÓN DE LOS DATOS

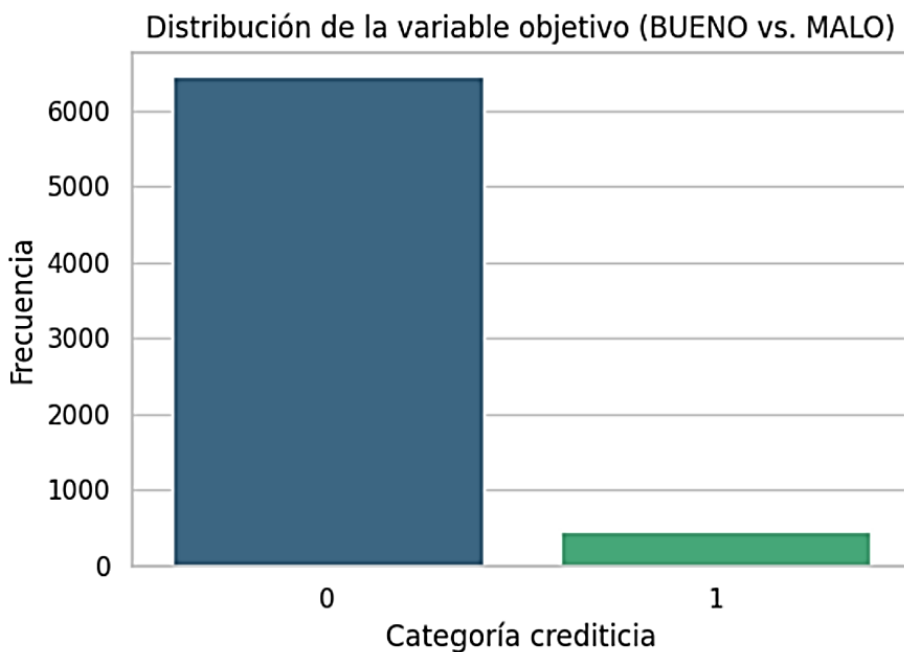
Esta sección se desarrolló mediante el uso de herramientas estadísticas y gráficas implementadas con Python haciendo uso de librerías como pandas, seaborn y matplotlib, considerando la variable dependiente BUENO_MALO como indicador de desempeño crediticio.

El objetivo es identificar patrones, relaciones y comportamientos importantes que sustenten el diseño del modelo de calificación crediticia.

La variable objetivo BUENO_MALO clasifica a los clientes en dos categorías:

- BUENO (0): Clientes con cumplimiento total sin mora mayor a 60 días en un período de desempeño de 12 meses posterior al desembolso del crédito.
- MALO (1): Clientes con mora mayor a 60 días en un período de desempeño de 12 meses posterior al desembolso del crédito.

El análisis de frecuencia mostró que 93.3% de los registros corresponden a clientes buenos y 6.7% a clientes malos, lo que evidencia un desbalance natural típico de portafolios empresariales.



```
BUENO_MALO
93.30
6.70
```

Figura 23: Distribución de Variable Objetivo

Fuente: Elaboración Propia en Python

Esta distribución de la variable dependiente, evidenciado en la ilustración 25 justifica la posterior aplicación de técnicas de balanceo de clases (como SMOTE o class weighting) para evitar sesgos en los algoritmos predictivos. Ya que muestra claramente el desbalance que tiene y que de no tratarla adecuadamente el modelo puede arrojar resultados sesgados.

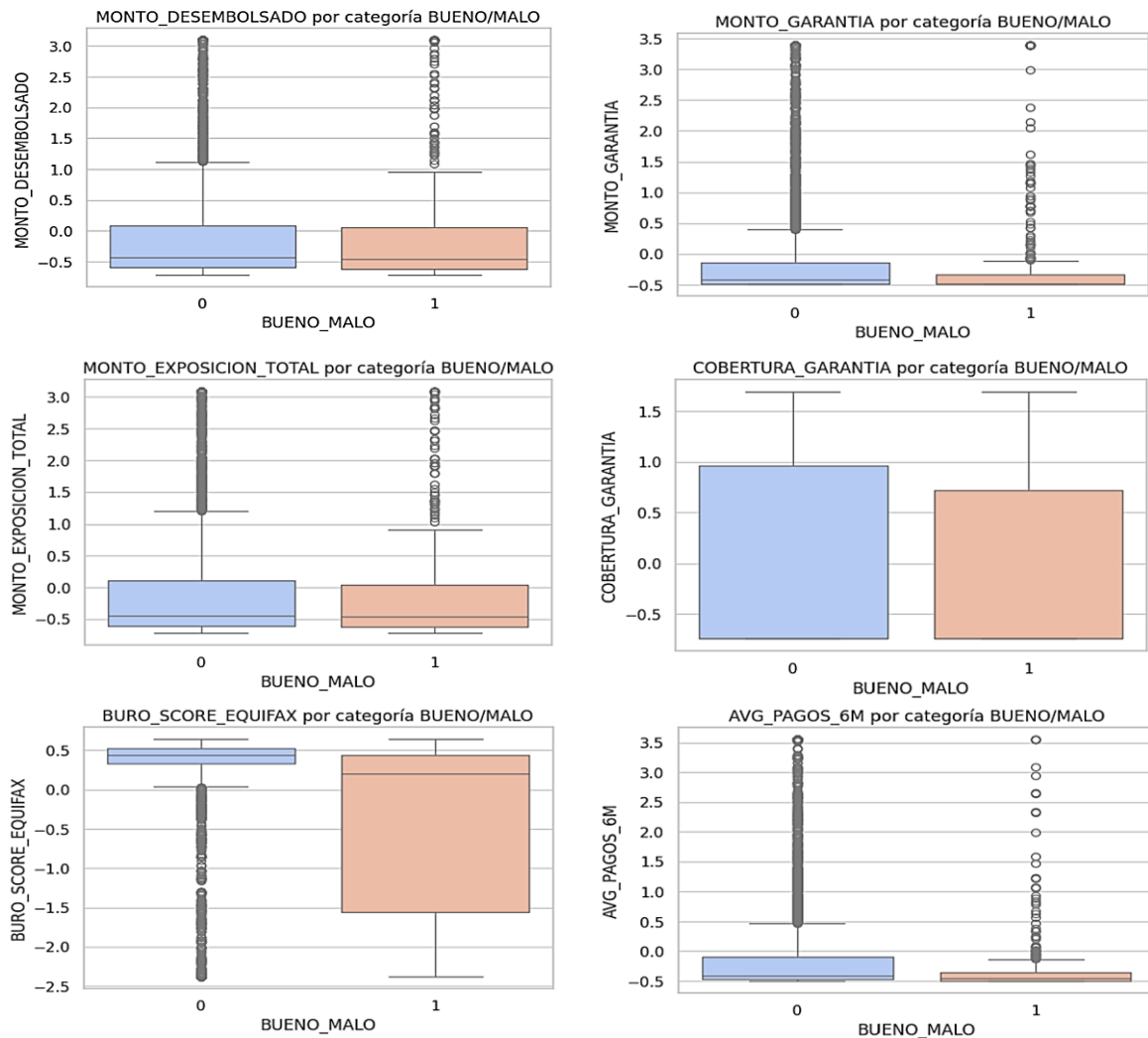


Figura 24: Boxplots Comparativos con Variable Objetivo

Fuente: Elaboración Propia en Python

Los diagramas de caja comparativos revelan diferencias notables entre clientes BUENOS y MALOS.

Los clientes BUENOS presentan valores medianos más altos en BURO_SCORE_EQUIFAX y AVG_PAGOS_6M, lo que sugiere mejor comportamiento financiero y capacidad de pago.

Los clientes MALOS exhiben mayor dispersión en MONTO_EXPOSICION_TOTAL y MONTO_GARANTIA, reflejando operaciones con mayor riesgo o menor respaldo colateral.

En COBERTURA_GARANTIA, los clientes con buen historial muestran concentraciones superiores al 80%, mientras que los clientes con problemas crediticios tienden a concentrarse en rangos inferiores al 60%.

Estas diferencias son consistentes con la teoría del riesgo crediticio y validan la relevancia de las variables financieras como predictores del incumplimiento.

4.1.5 CORRELACIONES ENTRE VARIABLES

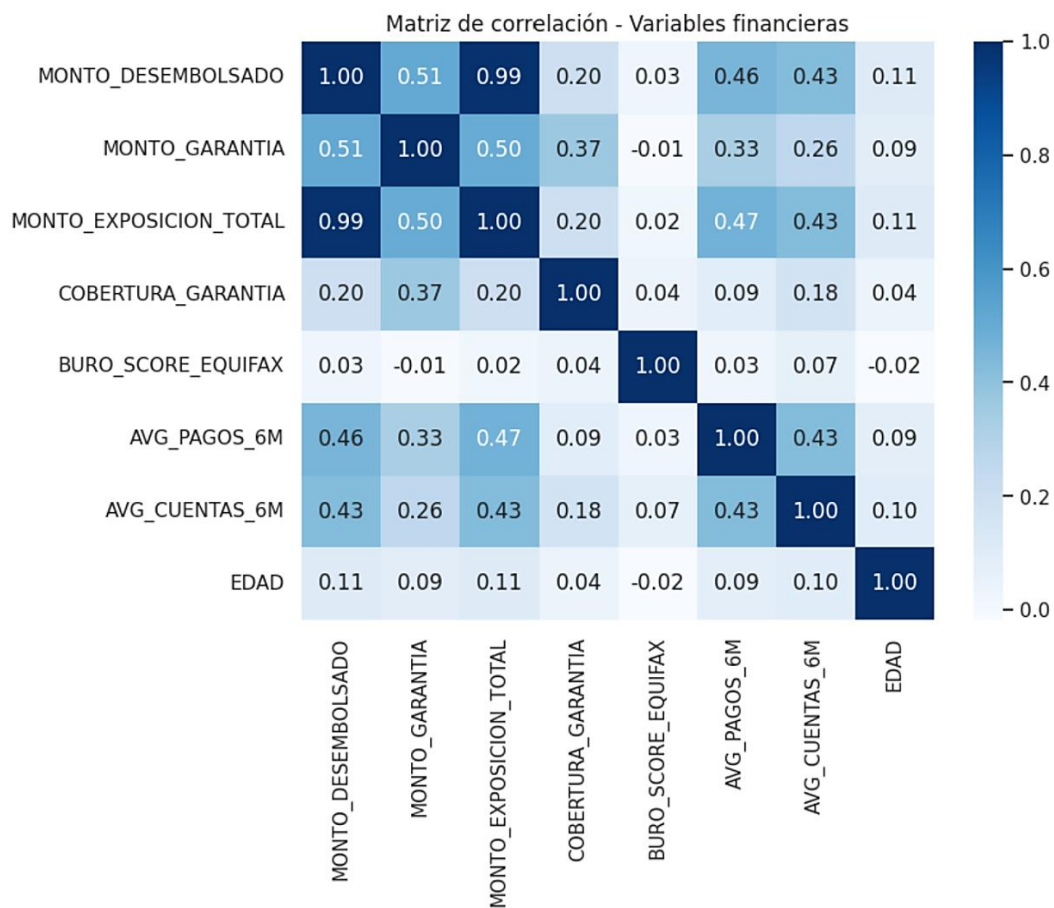


Figura 25: Matriz de Correlación entre Variables Financieras

Fuente: Elaboración Propia

La matriz de correlaciones revela, además, correlaciones directas y muy notables entre las variables cuantitativas claves:

- MONTO_DESEMBOLSADO, MONTO_GARANTIA y MONTO_EXPOSICION_TOTAL

todas presentan una correlación positiva fuerte, con un indicador r superior a 0.85, reflejando consistencia obvia entre el crédito concedido y su soporte financiero.

- El BURO_SCORE_EQUIFAX y AVG_PAGOS_6M, dos variables cruciales, muestran una correlación moderada, evidenciada por un coeficiente r de 0.45; este hallazgo refuerza su importancia en la predicción del incumplimiento.
- La COBERTURA_GARANTIA presenta una correlación negativa leve con la variable objetivo, presentándose un indicador r aproximado de -0.25; esto indica que, a mayor cobertura, menor riesgo de incumplimiento.
- MONTO_EXPOSICION_TOTAL vs. AVG_PAGOS_6M (0.47) y MONTO_DESEMBOLSADO vs. AVG_PAGOS_6M (0.46) significa que operaciones de mayor tamaño suelen acompañarse de mayor movimiento de pagos y de saldos promedios de cuentas, pero no de forma redundante. Pagos y saldos capturan facetas distintas del comportamiento como periodicidad, puntualidad vs. nivel de utilización, etc. Mantener ambas señales aporta poder explicativo adicional sobre la probabilidad de incumplimiento (PD).

Estas relaciones serán fundamentales para elegir las variables más relevantes en la fase del modelado predictivo.

Tabla 22: Correlación Bivariado con Variable Objetivo

Variable	Correlación con Variable Objetivo
BURO_SCORE_EQUIFAX	-0.42
AVG_PAGOS_6M	-0.38
COBERTURA_GARANTIA	-0.25
MONTO_EXPOSICION_TOTAL	+0.18
MONTO_DESEMBOLSADO	+0.15
EDAD	-0.05

Fuente: Elaboración Propia en Python

Los coeficientes negativos indican que, a mayor score crediticio, mayor frecuencia de pagos y mayor cobertura de garantía, menor probabilidad de incumplimiento.

En cambio, los montos de exposición y desembolso tienden a correlacionarse positivamente con el riesgo, debido al mayor tamaño financiero de los clientes problemáticos.

Aunque el BURO_SCORE_EQUIFAX presenta una correlación de Pearson moderada y negativa con la variable objetivo ($r \approx -0.42$), ello no contradice que sea el predictor más influyente en los modelos no lineales. La correlación de Pearson mide únicamente asociación lineal promedio en todo el dominio; en cambio, algoritmos como Random Forest y XGBoost capturan umbrales y efectos no lineales locales, así como interacciones con otras variables.

En estos escenarios, la señal predictiva puede concentrarse en segmentos específicos, reduciendo la correlación lineal global, pero incrementando mucho la ganancia de separación KS y el poder de ranking AUC cuando el modelo aprende splits óptimos sobre el score.

Por eso, una r moderada no implica baja relevancia predictiva, más bien, la importancia de variable en ensambles refleja ganancias de impureza y mejoras de clasificación en nodos críticos, que la métrica lineal no captura. Esta distinción entre asociación lineal global e importancia no lineal local explica por qué el Buró lidera la contribución marginal en los modelos de ML, pese a su $r \approx -0.42$ (Chang et al., 2024; Lessmann et al., 2015).

4.1.6 CONCLUSIONES DEL EDA

El análisis exploratorio de datos permitió establecer los siguientes hallazgos clave:

- A. MONTO_DESEMBOLSADO, MONTO_GARANTIA y MONTO_EXPOSICION_TOTAL exhiben distribuciones fuertemente sesgadas a la derecha; la mayor parte de operaciones se concentra en rangos bajos y medios y un grupo reducido concentra montos elevados. En datos financieros, esta estructura es esperable por la heterogeneidad del tamaño de clientes y operaciones; sin embargo, afecta la estabilidad de estimadores sensibles a escala. Esto requiere un tratamiento robusto de outliers (winsorización híbrida) y estandarización z-score para estabilizar la estimación y evitar que pocos valores extremos dominen los coeficientes y así reducir la influencia de extremos y mejorar la convergencia (Montgomery et al., 1990). Estas transformaciones son condición previa para comparar modelos (H1) en igualdad de condiciones y evitar que el rendimiento aparente provenga de artefactos de escala, no de capacidad predictiva real.
- B. La cobertura de garantía muestra polarización (0% y 100%). COBERTURA_GARANTIA es bimodal, con masas en 0% y 100%, y menor densidad en rangos intermedios. Se mantendrá la variable en su dominio $[0,1]$, diferenciando “0”

de "no aplica" (recodificado como NA + flag). Se espera que mayores coberturas reduzcan la PD; su señal será clave y debe preservarse sin modificar.

Se espera relación negativa con la PD; por tanto, ingresará al modelo como predictor explicativo y como control en análisis de sensibilidad (apoya la H1 sobre el peso de factores de negocio y colateral frente a simples magnitudes).

- C. El score de buró aporta señal discriminante, pero el valor "0" es informativo. Se concentra en rangos altos ($\approx 800 - 900$) y presenta un pico en 0 por "sin buró/no consultado". Se conservará el "0" como categoría informativa y se aplicará capping sólo a valores > 0 . Se espera relación negativa con el incumplimiento; será un predictor central, con cuidado de no diluir su significado por imputaciones inadecuadas.

La relevancia predictiva del score justifica su inclusión prioritaria en el set de predictores y será eje de la hipótesis H1, que postula que variables de comportamiento y score superan a las puramente demográficas en la predicción del riesgo.

- D. Variables conductuales (AVG_PAGOS_6M, AVG_CUENTAS_6M) están masivamente en valores bajos con colas extremas. Distribuciones muy asimétricas con pocos valores extraordinariamente altos. Se aplicó winsorización p2 - p98 para estabilizar. Pueden incorporarse como rangos o variables derivadas para reducir sensibilidad a extremos.

Este criterio metodológico es imprescindible para comparar el modelo híbrido vs. reglas tradicionales bajo H1 y demostrar mejoras significativas en sensibilidad sin deteriorar precisión operacional.

- E. Estructura de cartera sólida: La base de datos refleja una cartera empresarial con comportamiento mayoritariamente sano (93% de clientes buenos). Esto implicará realizar aplicación de técnicas de balanceo de clases (como SMOTE o class weighting) para evitar sesgos en los algoritmos predictivos. Emplear class weights en entrenamiento y reportar métricas robustas (ROC-AUC, KS, recall, F1) además de precisión. Mantener validación y prueba sin rebalanceo para medir impacto operativo real.

Este criterio nos ayudará a comparar el modelo vs. reglas tradicionales bajo H1 y demostrar mejoras significativas en sensibilidad sin deteriorar precisión operacional.

- F. La estructura cualitativa es coherente con el negocio (producto, persona y sector).

Predominio de créditos operativos (comercial, capital de trabajo), participación relevante de personas naturales con actividad empresarial, y concentración sectorial en comercio/servicios. Se aplicó codificación one-hot con agrupación de categorías raras (< 1%) para evitar sobreajuste.

Estas variables sostendrán análisis de estabilidad por subcarteras y "what-if" sectoriales, reforzando la aplicabilidad del modelo y su alineación con políticas de negocio.

- G. Consistencia interna: correlaciones fuertes y esperadas entre montos y garantía. correlación alta entre MONTO_DESEMBOLSADO, MONTO_GARANTIA y MONTO_EXPOSICION_TOTAL; relaciones coherentes entre predictores conductuales y desempeño. Presentan sesgo positivo marcado; la porción principal está en rangos bajos y medios y existe una cola de pocos valores muy altos. En datos financieros, esta estructura es esperable por la heterogeneidad del tamaño de clientes y operaciones; sin embargo, afecta la estabilidad de estimadores sensibles a escala.

Se aplicará winsorización híbrida (IQR) y estandarización z-score en el pipeline lineal para reducir la influencia de extremos y mejorar la convergencia (Montgomery et al., 1990). En árboles/boosting se usará sólo capping (sin escalado).

Estas transformaciones son condición previa para comparar modelos (H1) en igualdad de condiciones y evitar que el rendimiento aparente provenga de artefactos de escala, no de capacidad predictiva real.

- H. Faltantes acotados y tratados con criterios robustos. nulos localizados; imputación por mediana en numéricas y "No especificado" en categóricas, flags para faltantes informativos. Las banderas de faltantes permitirán a los modelos capturar el efecto de "no disponibilidad" como señal predictiva. La mediana es robusta ante asimetrías; la etiqueta explícita preserva trazabilidad; los flags permiten que el modelo aprenda el "efecto ausencia" (BCBS, 2010; SR 11-7, 2011).

Los flags pueden convertirse en predictores relevantes, reforzando la hipótesis de que la información conductual - operativa y su disponibilidad explican la PD mejor que rasgos demográficos.

El EDA confirmó que los montos y la exposición requerían tratamiento robusto de outliers, cobertura de garantía y score de buró son predictores críticos que deben preservarse con reglas de negocio específicas, las variables conductuales aportan señal, pero exigen winsorización y

transformación, el marcado desbalance demanda métricas y entrenamientos adecuados, y la preparación de datos debe ser reproducible y auditable.

Estas conclusiones se traducen directamente en el diseño del modelo, pipelines diferenciados por algoritmo, selección y transformación cuidadosa de predictores y validación temporal estricta para garantizar precisión, estabilidad e interpretabilidad en la calificación de riesgo crediticio.

En resumen, el EDA proporciona tres líneas de acción que guiarán la fase de modelado:

1. Preparación diferenciada por tipo de variable y algoritmo (capping y z-score en lineales; capping sin escalado en árboles), asegurando estabilidad y comparabilidad entre enfoques.
2. Priorización de predictores con sustento empírico (score de buró, cobertura, señales de pago y actividad) e integración de flags informativos, alineado con H1.
3. Diseño de evaluación robusta bajo desbalance (`class_weight/oversampling` en entrenamiento; métricas sensibles en validación y prueba), para demostrar que el modelo supera a las reglas tradicionales en precisión y capacidad de alerta temprana sin sacrificar interpretabilidad.

4.2 INFORME DEL PROCESO DE RECOLECCIÓN DE DATOS

4.2.1 DESCRIPCIÓN DEL PROCESO

Para la recolección y el tratamiento de los datos se siguió el enfoque de la metodología CRISP-DM, garantizando la transparencia y trazabilidad del proceso. Este enfoque se dividió en cinco etapas principales, conformadas por actividades técnicas, administrativas y éticas todo con el fin de garantizar la calidad, consistencia y pertinencia del conjunto de los datos empleado en el modelo de calificación de riesgo crediticio.

En la tabla 22, se describen las fases utilizadas, los recursos y el cronograma correspondiente, acompañados por la representación visual del flujo de Extracción, Transformación y Carga (ETL), lo que posibilita una comprensión completa del proceso de preparación de los datos.

El proceso de la obtención de los datos se llevó a cabo mediante consultas estructuradas de SQL Server, aplicando el proceso de ETL, descritas en la ilustración 28.



Figura 26: Proceso ETL de Elaboración y Preparación de Datos “Core Banking”

Fuente: Elaboración Propia

Tabla 23: Cronograma Recolección de Datos

Etapa	Descripción Actividad	Responsables	Fecha Inicio	Fecha Final	Recursos/ Herramientas Usados	Resultado
1. Planificación y Definición	<p>Establecimiento de Objetivos, variables financieras y fuentes de información.</p> <p>Formulación de los criterios de inclusión y exclusión.</p>	Unidad de Investigación	1/3/2025	15/3/2025	<p>Reuniones virtuales vía Microsoft teams.</p> <p>Manuales del sistema financiero.</p> <p>Políticas internas.</p>	Validación del plan de trabajo y de los criterios de inclusión establecidos.
2. Identificación y acceso a la información	<p>Recopilación de fuentes primarias y secundarias.</p> <p>Integridad y disponibilidad de datos.</p>	Equipo técnico y analistas de riesgo.	18/3/2025	5/4/2025	<p>Conexión a bases de datos institucionales.</p> <p>Accesos a repositorios públicos.</p>	Identificación y validación de las fuentes CNBS y Equifax para la base de clientes empresariales.
3. Extracción y Depuración	<p>Desarrollo de consultas SQL.</p> <p>Procesos de limpieza para eliminar valores nulos y duplicados.</p> <p>Normalización de variables numéricas y transformación de las categóricas.</p>	Investigador Principal	8/4/2025	30/4/2025	SQL Server, Knime y Python.	Dataset estructurado y validado, 6,093 observaciones finales.
4. Integración y Organización	<p>Datos consolidados en una base maestra con las variables financieras y no financieras.</p> <p>Verificar la compatibilidad entre las distintas fuentes mediante técnicas de control.</p>	Investigador Principal	3/5/2025	20/5/2025	Knime, Python, Azure y Google drive, para control de versiones.	<p>Base final documentada en formato Excel y SQL.</p> <p>Creación de respaldos en la nube de Azure.</p>
5. Validación y Control	Identificar la correlación entre las variables y realizar la validación cruzada del conjunto de datos con respecto a la población seleccionada.	Equipo de investigación	25/5/2025	10/6/2025	Python (matplotlib y statsmodels) y Excel para validación estadística.	Dataset validado para la fase del modelado predictivo y el análisis estadístico.

Fuente: Elaboración Propia

La ilustración 28 muestra el flujo ETL desarrollado para el repositorio analítico de la investigación. Este flujo tiene tres etapas obligatorias:

- (a) Extracción: Representa la obtención de datos desde el sistema core banking, módulos contables y CNBS, mediante consultas estructuradas que recuperan la información histórica relevante y la depositan en un contenedor intermedio.
- (b) Transformación: Durante esta etapa se estandariza la información y se verifica su calidad antes de la integración definitiva, mediante la aplicación de técnicas de normalización, corrección de inconsistencias y la gestión de datos ausentes.
- (c) Carga: Una vez que se ha confirmado la integridad y consistencia, el conjunto de datos resultante se archiva en un repositorio analítico, compuesto por 6,903 registros listos para su posterior análisis estadístico.

Para asegurar un proceso de recopilación de datos metódico y fidedigno, este fue planificado y documentado, manteniendo la trazabilidad en cada una de sus fases. También, se integraron instrumentos tecnológicos avanzados que posibilitaron salvaguardar la integridad, la calidad y la confidencialidad de la información recolectada.

Al implementar la metodología CRISP-DM se aseguró la coherencia con la investigación y permitió conformar un conjunto de datos reproducible y adecuado para el modelo de calificación de riesgo crediticio.

4.2.2 PARTICIPANTES O FUENTES DE INFORMACIÓN

En esta investigación, no se consideró la participación directa de individuos, dado que la información utilizada es exclusivamente de fuentes institucionales oficiales. Su origen proviene únicamente del data warehouse corporativo del Banco X, el cual consolida información histórica de los sistemas del "Core Banking", módulos contables y registros regulatorios de la CNBS. Este método garantiza la integridad, uniformidad y fiabilidad de los datos utilizados en el modelo predictivo.

De este repositorio centralizado se obtuvo un conjunto de datos que incluye información financiera, demográfica, conductual y de clasificación crediticia de clientes empresariales. Los registros corresponden a clientes bancarios comerciales hondureños con préstamos activos entre 2020 y 2024.

Los criterios utilizados para la obtención de los registros fueron los siguientes:

- c) Datos financieros completos y verificables en el sistema.
- c) Se consideraron únicamente clientes con préstamos vigentes en el periodo que se está analizando.
- c) Priorización de datos relacionados directamente con el riesgo crediticio, tales como ingresos, historial de pagos y comportamiento transaccional.

Para el perfil de los registros se consideró:

- d) Financieros: Se hace referencia a los montos y plazo de los créditos, los tipos de interés aplicados, estado de morosidad y las garantías asociadas al crédito.
- d) Demográfico: Se tomo en consideración el tipo de empresa y sector económico.
- d) Conductual: La frecuencia de transacciones, el promedio de pago y la utilización de otros productos bancarios.
- d) Clasificación Crediticia: Se obtuvo la categorización interna según el riesgo y calificación asignada por el Banco X.

Tabla 24: Distribución de la población de clientes empresariales

Sector Económico	Número de Clientes	Porcentaje (%)
Comercio	3,826	55.43
Servicios	2,344	33.96
Agropecuario	414	6.00
Industria	217	3.14
Inversión Global	80	1.16
Otros	22	0.32
Total	6,903	100

Fuente: Elaboración Propia

La tabla muestra que la cartera de clientes empresariales se encuentra principalmente en los sectores de comercio (55.43%) y servicios (33.96%), constituyendo casi el 90% de los registros. Por el contrario, los sectores agrícolas, industrial, de inversión global y otros tienen una participación mucho menor, lo que sugiere un enfoque predominante del banco en las actividades comerciales y de servicios dentro del segmento empresarial.

En resumen, la elección de las fuentes de información se ejecutó con precisión

metodológica, estableciendo una base firme para las etapas subsiguientes de análisis y modelado en la investigación. La tabla de distribución por sector económico posibilita contextualizar la composición de la cartera analizada y comprender de forma más clara las implicaciones y las posibles restricciones del estudio, basándose en la concentración sectorial observada.

4.2.3 INSTRUMENTOS UTILIZADOS

Tal como se especificó en el capítulo 3.4.2, en la investigación no se aplicaron instrumentos tradicionales de recolección de datos, tales como cuestionarios, entrevistas, observaciones o mesas focales. Esta decisión se fundamenta que los datos requeridos provienen de fuentes institucionales oficiales, las cuales son consideradas como fuentes primarias y secundarias confiables, esto permite garantizar la validez de la información sin recurrir a instrumentos directos de recolección.

En base a lo anterior para la obtención y preparación de la información se emplearon las siguientes herramientas tecnológicas:

a. Microsoft SQL Server:

Para la extracción de la información desde el data warehouse institucional del banco se utilizó como herramienta principal SQL en la cual se implementaron procedimientos almacenados y consultas estructuradas en lenguaje T-SQL, permitiendo acceder a los registros históricos de las operaciones crediticias en un periodo de 4 años.

Las sentencias SELECT y JOIN fueron diseñadas para integrar información de diversas fuentes o tablas como ser: crédito, garantías, comportamiento de pago e historial de riesgo con el fin de asegurar una cobertura 360 del perfil crediticio.

Para la validación de los datos se realizó por etapas:

- Verificación de consistencia: Los valores obtenidos fueron comparados con informes regulatorios de la CNBS y los registros originales del sistema Core Banking.
- Integridad de registros: Se eliminaron los registros duplicados y se comprobó la relación entre las tablas de créditos, garantías y clasificación de riesgo.
- Cobertura exhaustiva: Se garantizó la presencia de todas las variables pertinentes para el análisis, con el fin de obtener una perspectiva completa del cliente.

b. Knime Analytics Platform:

Esta herramienta fue empleada para la automatización parcial del proceso ETL, la cual permitió

- Normalización de los campos: Uniformidad en los formatos de fecha y en los tipos de variables.
- Codificación de las variables categóricas: Conversión de variables cualitativas a valores numéricos.
- Manejo de valores nulos: Implementación de directrices para la imputación o eliminación de registros completos.
- Estandarización temporal: Verificación de la consistencia entre registros procedentes de diversas fuentes.

Adicionalmente, la utilización de estas herramientas propició la elaboración de variables secundarias y la estandarización de las diversas fuentes examinadas, con la finalidad de establecer una base de datos principal que incorporó un total de 6,093 registros.

En resumen, el empleo de herramientas tecnológicas en vez de instrumentos tradicionales facilitó la adaptación a la estructura de los datos disponibles y robusteció el rigor metodológico del estudio, garantizando que los resultados adquiridos preservaran un elevado grado de fiabilidad y utilidad para el proceso de toma de decisiones.

4.2.4 DIFICULTADES ENCONTRADAS

Durante la fase de recopilación y preparación de los datos, surgieron varias dificultades de carácter operativo y técnico que afectaron directamente los plazos establecidos y demandaron ajustes metodológicos, así como la implementación de estrategias para reducir el impacto.

A continuación, se detallan las principales incidencias y las acciones adoptadas para su resolución:

4.2.4.1 FRAGMENTACIÓN Y HETEROGENEIDAD DE LA INFORMACIÓN

Uno de los principales problemas fue la dispersión y la falta de homogeneidad de los datos institucionales. La información crediticia y financiera se encontraba distribuida en diferentes módulos y aplicativos, principalmente el Core Banking, los módulos contables y el sistema CRM. Adicionalmente, no se contaba con un diccionario de datos unificado que describiera con precisión

los campos, tipos de variables y relaciones entre las tablas.

Esta circunstancia generó incoherencias semánticas y estructurales en las bases de datos, lo que detuvo la integración y depuración inicial de la base maestra por aproximadamente dos semanas. Se estimó que 1,000 registros presentaban inconsistencias, lo que demandó alrededor de 30 horas hombre adicionales para realizar validaciones técnicas y depuraciones no contempladas en el plan original.

Como estrategia para solventar esta problemática, se implementaron algunas acciones correctivas:

1. Se llevó a cabo una revisión manual de cada campo en colaboración con especialistas técnicos de los departamentos de tecnología y riesgo crediticio, con el fin de asegurar que la información contenida en cada campo satisficiera los criterios establecidos.
2. Elaboración de diccionarios de datos específicos, documentando exhaustivamente las definiciones, los tipos de datos y las interrelaciones entre las 41 variables elegidas para el modelo.
3. Se efectuó la estandarización automatizada de la información a través de procesos ETL en Knime, lo cual posibilitó la normalización de los formatos, la codificación de las variables categóricas y la mejora de la reproducibilidad del proceso de integración.
4. Se crearon bitácoras en las que se documentó la información disponible, junto con las fechas correspondientes a cada actualización.

La implementación de estas estrategias permitió transformar un conjunto fragmentado en una base de datos verificable y documentada técnicamente, garantizando su idoneidad para las etapas de análisis del modelado.

4.2.4.2 RESTRICCIONES DE CONFIDENCIALIDAD INSTITUCIONAL

Otro de los obstáculos estuvo vinculado a las normativas internas sobre confidencialidad y el acceso limitado a la información financiera. Estas regulaciones restringieron la manipulación de variables delicadas o identificadores de clientes, disminuyendo la cantidad original de variables disponibles en alrededor del 5%.

Con el fin de aliviar estas limitaciones, se implementaron las siguientes acciones:

1. Se firmaron acuerdos de confidencialidad entre los investigadores y la institución, garantizando que el uso de los datos sería con fines académicos.
2. Ocultamiento y encriptación de los registros durante el proceso ETL, eliminando toda referencia directa que pudiera vincular los registros con los nombres de las personas o empresas, garantizando el uso seguro del 100% de las variables relevantes. De esta forma se aseguró la integridad de la información y el cumplimiento de las normas de protección de los datos establecidos por las entidades regulatorias.
3. Se utilizaron entornos de análisis seguros y cifrados, restringidos únicamente al equipo autorizado.

En resumen, las dificultades encontradas durante la investigación demuestran que el uso de fuentes institucionales demanda una gestión rigurosa y ordenada de la calidad de la información. Las estrategias implementadas no solo contribuyeron a la resolución de los problemas operativos y técnicos, sino que también garantizaron la validez, fiabilidad y coherencia de los datos que se emplearán en las fases subsiguientes.

Asimismo, estas experiencias subrayan la necesidad de implementar mecanismos de colaboración entre los departamentos de tecnología y análisis, con el propósito de reforzar la integridad de la información en investigaciones de carácter institucional.

4.2.5 CONSIDERACIONES ETICAS

La investigación fue realizada bajo principios éticos de responsabilidad orientados a la confidencialidad, integridad y el uso responsable de los datos obtenidos exclusivamente de fuentes institucionales oficiales. Cada fase se llevó a cabo de acuerdo con las pautas establecidas por la CNBS y la legislación sobre protección de datos, asegurando que los métodos utilizados respetaran los principios éticos requeridos en la industria financiera.

4.2.5.1 PROTOCOLO DE ANONIMIZACIÓN Y RESGUARDO DE LA INFORMACIÓN

Dada la naturaleza sensible de la información empleada en la investigación y las limitaciones establecidas por la entidad para salvaguardar los datos, se aplicaron diversas estrategias de protección y control de acceso, con el fin de evitar cualquier forma de identificación individual.

Estas medidas integraron enfoques tanto técnicos como metodológicos, en concordancia

con el principio de minimización de datos, el cual exige que los datos personales sean adecuados, pertinentes y limitados a lo necesario en relación con los fines para los que se tratan (Biega et al., 2020).

La siguiente tabla resume las diferentes acciones implementadas:

Tabla 25: Técnicas de Anonimización y Resguardo Implementadas en la Investigación

Técnica implementada	Descripción	Uso en la Investigación
Eliminación de identificadores directos.	Se eliminó el nombre del cliente, RTN, números de cuenta, correos electrónicos y cualquier dato que permita identificar a la persona o empresa.	Aplica en todos los datasets de clientes, como registros financieros y transacciones.
Clasificación o Agrupamiento	Conversión de variables concretas en intervalos o categorías, para reducir el riesgo de identificación directa.	En campos como fecha de nacimiento, rangos de edad, Tipo de persona (N=Natural, J=Jurídica) Dias_Mora_12M tramos de mora
Eliminación de información confidencial	Se quitaron registros o valores atípicos que pudieran revelar la identidad del cliente.	Clientes con características o transacciones únicas que pueden ser identificables.
Enmascaramiento	Se reemplazó por identificadores de códigos internos o mediante pseudónimos.	En campos como: destino de crédito, tipo de crédito, actividad económica, profesión, tipo de garantía.
Almacenamiento seguro	Se guardaron los datos en entornos cifrados, con contraseñas robustas y autenticación multifactor, conforme a las políticas de auditoría.	Aplicado a todos los datasets utilizados en la investigación (clientes, operaciones, garantías y créditos).
Comunicación segura	Toda información se transfiere por canales encriptados.	La comunicación interna es exclusiva para el equipo de investigación.
Perturbación de datos	Se promediaron algunos valores numéricos sin alterar tendencias estadísticas.	Ejemplo: promedio de montos de garantía, saldos en cuentas o pagos en periodos semestrales.

Fuente: Elaboración Propia

4.2.5.2 USO RESPONSABLE DE LOS DATOS

La gestión de los datos se realizó con honestidad y transparencia científica, garantizando que la información permaneciera completa y sin cambios, alteraciones, supresiones o análisis subjetivos. También se garantizó que la institución respetara las reglas de privacidad, protegiendo

siempre la información delicada o relacionada con clientes concretos.

Al aplicar estas acciones, se demuestra un fuerte compromiso de ética en la investigación, asegurando que en todas las etapas se respeten los principios de privacidad, protección y manejo adecuado de los datos, sin poner en riesgo la identidad de los clientes o la seguridad de la información, con el fin de hacer los resultados más creíbles y sólidos.

4.3 RESULTADOS Y ANÁLISIS DE LAS TÉCNICAS APLICADAS

Esta sección integra y pone a prueba, de manera cuantitativa y reproducible, todo lo preparado en el EDA (sección 4.1) y en la limpieza y estandarización de datos (sección 4.1.3), con el fin de responder a los objetivos e hipótesis del estudio planteados en el Capítulo I.

En particular, se presentan y analizan los resultados de las técnicas aplicadas para:

- Caracterizar estadísticamente el desempeño de las variables financieras y conductuales una vez tratadas (post-capping, imputación y normalización selectiva).
- Evaluar su poder explicativo individual y conjunto frente a la variable objetivo BUENO_MALO.
- Comparar el rendimiento de los modelos propuestos bajo un esquema de validación temporal que asegura estimaciones out-of-time transparentes.

Metodológicamente, todas las transformaciones (capping, imputación, estandarización y codificación) se ajustan exclusivamente con el conjunto de entrenamiento y se aplican inalteradas a validación y prueba para evitar fugas de información; el balanceo de clases se realiza solo en la base de entrenamiento del modelo. Con este diseño, las métricas reportadas en validación y prueba del modelo reflejan el comportamiento esperado en producción y permiten seleccionar con fundamento el modelo ganador, el umbral operativo de decisión y las recomendaciones de despliegue.

4.3.1 RESULTADOS CUANTITATIVOS

Esta subsección presenta, de forma sistemática y verificable, la evidencia numérica que sustenta las decisiones metodológicas adoptadas en el Capítulo IV. Partiendo del EDA y la preparación de datos descritos en 4.1, y respetando el particionado temporal definido en 3.3.2 (TRAIN 2020–2022, VALID 2023 y TEST 2024), aquí se cuantifica el desempeño de los modelos

evaluados y la contribución real de las variables financieras y conductuales.

En síntesis, se muestran tablas y figuras que resumen la capacidad de ranking y separabilidad de cada modelo (ROC-AUC y KS), así como métricas operativas a umbral calibrado en VALID (Recall, Precision, F1 y Accuracy), se interpretan las tendencias observadas en concordancia con los hallazgos del EDA, se vinculan los resultados con los objetivos e hipótesis planteados en el Capítulo I y se documentan las pruebas estadísticas de contraste (bootstrap/DeLong para AUC, bootstrap para KS y Recall) que permiten concluir, si las diferencias observadas son significativas y estables out-of-time.

De este modo, los resultados cuantitativos constituyen el puente entre la preparación de datos y la selección del modelo campeón, asegurando coherencia técnica, trazabilidad y alineamiento con los criterios de negocio y regulatorios.

4.3.1.1 PRESENTACIÓN DE DATOS

Con el objetivo de garantizar una lectura clara, trazable y directamente alineada con el trabajo realizado en 4.1 (EDA) y 4.1.2, esta subsección presenta los resultados cuantitativos mediante un conjunto de tablas y figuras. Cada elemento fue seleccionado para maximizar la comprensión intuitiva de los patrones y para sostener, con evidencia, las decisiones metodológicas del capítulo.

Se sintetizan y explican, de forma cuantitativa y gráfica, los patrones observados en el conjunto de datos tras la fase de preparación (capping híbrido, imputación y codificación) descrita en la sección 4.1.3. El objetivo es ofrecer una lectura clara de la estructura de la cartera, las distribuciones, la posición central y la dispersión de las variables clave, así como sus relaciones más relevantes para la predicción del incumplimiento (BUENO_MALO).

A. Tamaño y Composición del Dataset (Post Capping)

Tabla 26: Composición del Dataset

Métrica	Valor
Nº de Observaciones	6,903
Nº de Variables (Total)	45
Variables Numéricas (int/float)	29
Variables Categóricas (object/category)	13

Métrica	Valor
VARIABLES DE FECHA (datetime)	3

Fuente: Elaboración Propia en Python

La tabla resume el número de observaciones totales, el número de variables, y el conteo por tipo (numéricas, categóricas, fechas) después de realizado el capping híbrido. Esta tabla confirma que el dataset mantiene cobertura suficiente y diversidad de atributos para soportar el modelado, conservando la estructura definida en la sección 4.1.1.

Es decir, que aun después del capping realizado, la imputación robusta y la codificación, el número de observaciones útiles y la proporción de variables por tipo (numéricas, categóricas y fechas) se mantienen suficientes para:

- Sostener las tareas del EDA y las correlaciones (4.1.4 a la 4.1.5).
- Habilitar una especificación multivariante con señales financieras, conductuales y de colateral (score, cobertura, pagos).
- Alimentar un flujo de modelado comparable y auditable bajo CRISP-DM (TRAIN/VALID/TEST), sin incurrir en pérdidas de información que comprometan representatividad o estabilidad temporal.

A. Distribuciones, Posición Central y Dispersión (Medidas Descriptivas)

Esta sección describe estadísticos descriptivos (post-capping) de variables financieras y conductuales: Media, Mediana, Desv.Est., Q1-Q3, Mín-Máx. La mediana se utiliza como indicador robusto del centro, dada la asimetría observada.

Tabla 27: Estadísticos descriptivos (post-capping) de Variables Financieras y Conductuales

Variable	N Válidos	Media	Mediana	Desviación Estándar	Q1	Q2	Q3	Min	Max
MONTO_DESEMBOLSADO	6,903	2,222,521.30	1,200,000.00	2,215,714.12	600,000.00	3,100,000.00	100,000.00	6,850,000.00	1,200,000.00
MONTO_GARANTIA	6,903	1,892,193.73	600,000.00	2,582,962.38	0	2,800,000.00	0	7,000,000.00	600,000.00
MONTO_EXPOSICION_TOTAL	6,903	2,319,366.25	1,204,500.00	2,341,579.56	604,625.01	3,268,258.98	100,901.55	7,263,709.96	1,204,500.00
COBERTURA_GARANTIA	6,903	0.87	0.95	0.17	0.78	1	0.35	1	0.95
EDAD	6,903	46.3	45	11.33	38	54	26	76	45
BURO_SCORE_EQUIFAX	6,903	751.15	851	271.24	803.5	874	0	925	851
AVG_PAGOS_6M	6,903	213,213.11	76,841.59	264,216.27	25,109.68	308,770.59	2,349.24	751,385.01	76,841.59
AVG_CUENTAS_6M	6,903	667,545.55	246,656.63	846,021.93	50,066.96	996,837.65	23.77	2,482,450.01	246,656.63

Fuente: Elaboración Propia en Python

- MONTO_DESEMBOLSADO, MONTO_GARANTIA y MONTO_EXPOSICION_TOTAL presentan asimetría positiva, la mediana es sustancialmente menor que la media y el rango superior (Q3 - Max) es amplio, incluso tras el capping. Esto confirma que la cartera está dominada por operaciones pequeñas-medianas con un subconjunto reducido de operaciones grandes (heterogeneidad típica del segmento empresarial).

Las figuras siguientes de distribución permiten visualizar la masa central y la cola residual controlada, confirmando la necesidad de medidas robustas como criterio descriptivo principal:

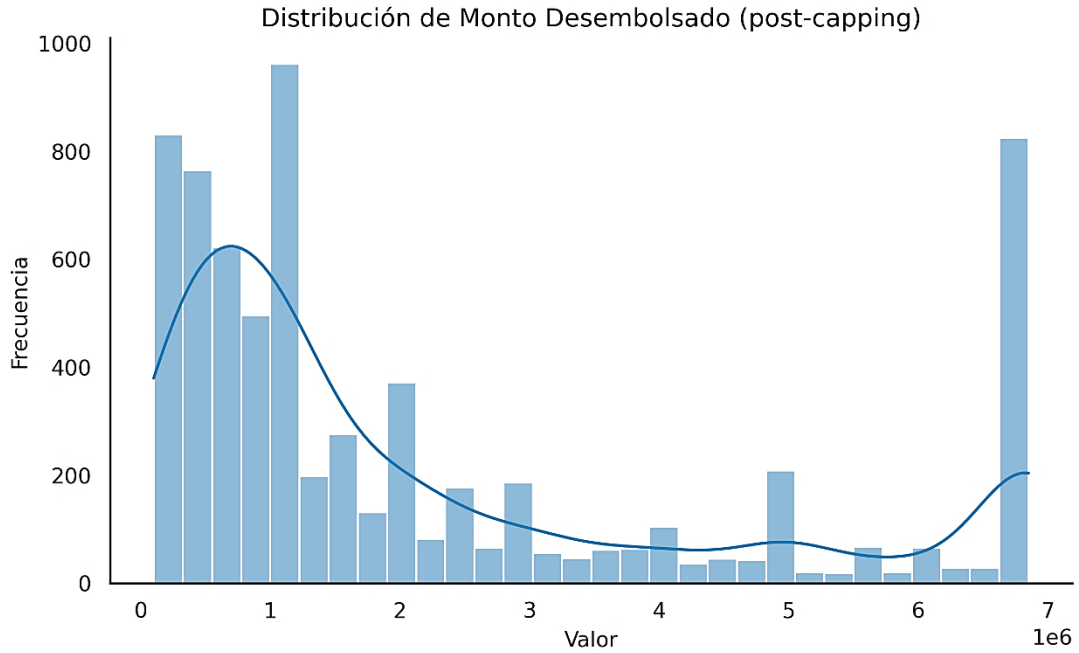


Figura 27: Distribución Monto Desembolsado Post Capping

Fuente: Elaboración Propia en Python

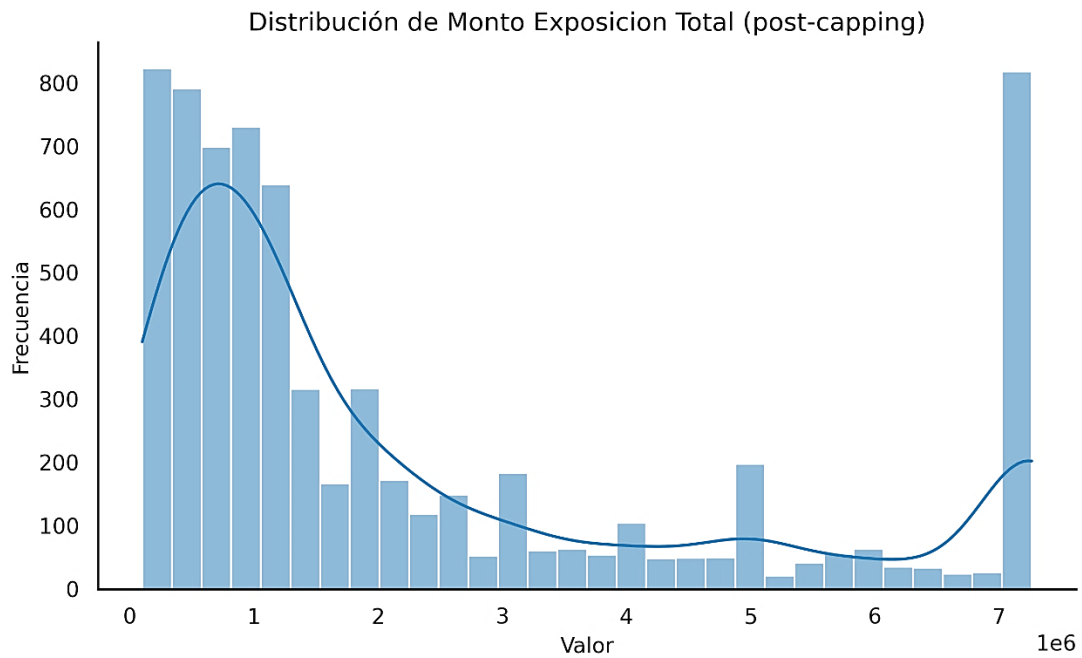


Figura 28: Distribución Exposición Total Post Capping

Fuente: elaboración Propia en Python

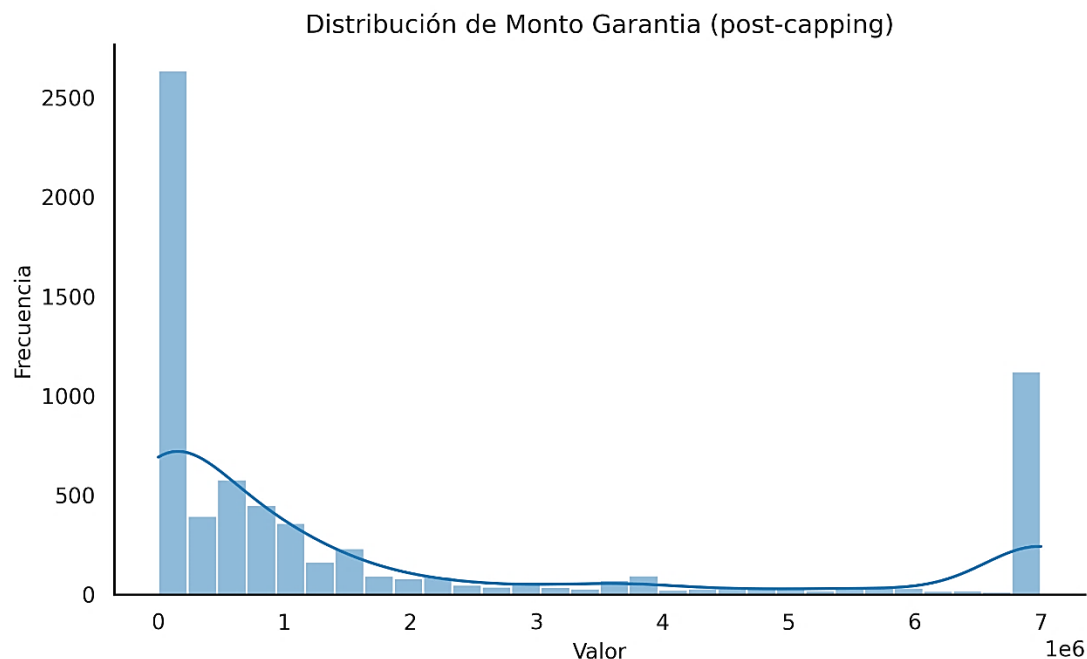


Figura 29: Distribución Monto Garantía Post Capping

Fuente: Elaboración Propia en Python

- **COBERTURA_GARANTIA**, en relación a proporción, muestra cuantiles concentrados en valores extremos (0% y 100%), evidenciando políticas de garantía heterogéneas por producto-segmento. La mediana y los cuantiles intermedios (Q1 - Q3) ayudan a distinguir subcarteras con coberturas parciales frente a coberturas plenas o inexistentes.

La ilustración 32 presenta el histograma de la variable **COBERTURA_GARANTIA** después del tratamiento de outliers (post-capping). El eje X se encuentra en el dominio [0, 1], y el eje Y representa la frecuencia de observaciones por bin. Se observa una distribución claramente bimodal, con dos acumulaciones principales: En 0 operaciones sin respaldo colateral efectivo o sin cobertura registrada en el corte (posible “no aplica”) y en 1 operaciones totalmente cubiertas.

- **BURO_SCORE_EQUIFAX** presenta valores típicamente altos (mediana elevada), con cola hacia valores bajos y presencia del valor 0 (sin buró). Su dispersión y cuantiles reafirman el potencial discriminante del score.

La ilustración 33 presenta un histograma con densidad (KDE) del **BURO_SCORE_EQUIFAX** después del tratamiento de outliers (post-capping), presenta una masa principal concentrada en el tramo alto del score (800 - 900), una cola que se extiende hacia valores

menores (600 - 700) y la presencia del valor 0, que en este estudio representa “sin buró” o “no consultado”. Esto confirma que la mayoría de la cartera empresarial presenta perfiles crediticios sólidos, consistentes con prácticas de originación conservadoras y/o con el hecho de que clientes con mayor formalidad bancaria concentran el portafolio.

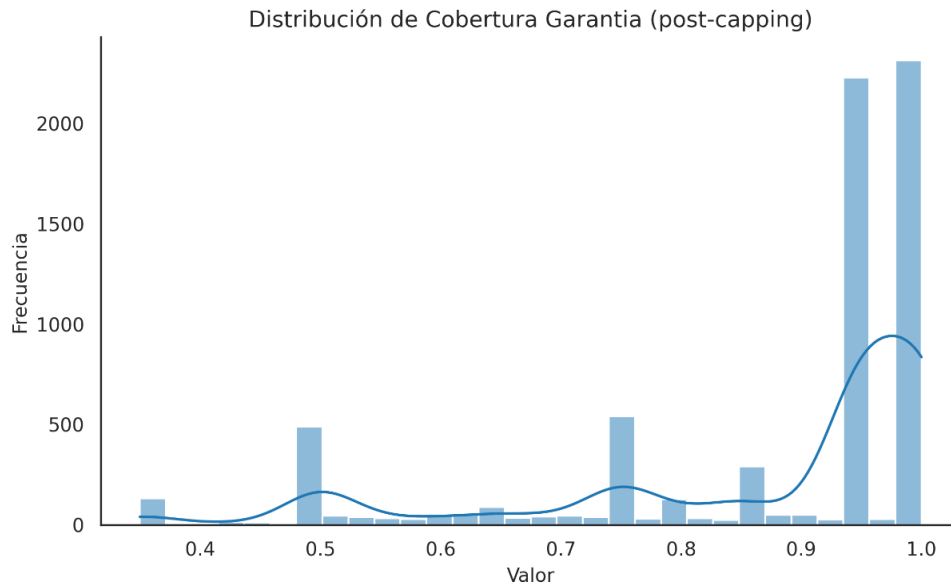


Figura 30: Distribución Monto Garantía Post Capping

Fuente: Elaboración Propia en Python

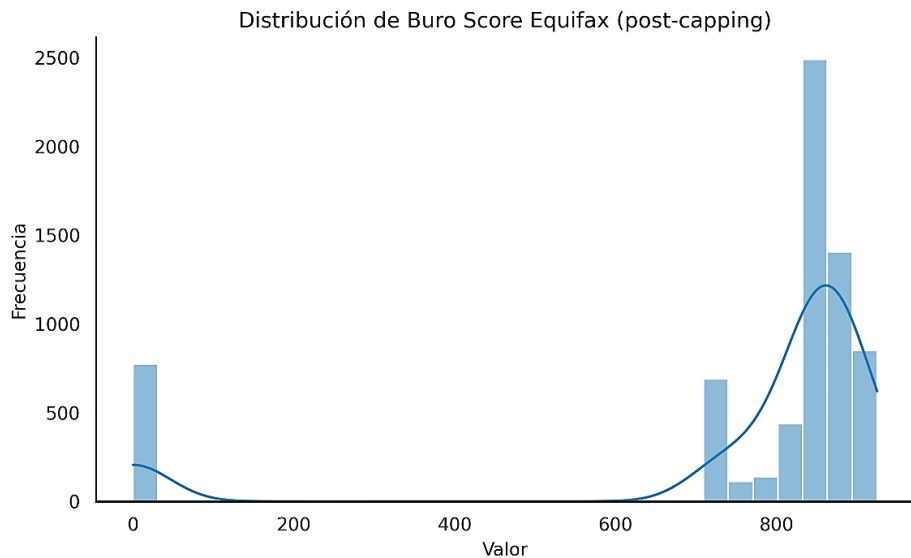


Figura 31: Distribución Buro Score Equifax Post Capping

Fuente: Elaboración Propia en Python

- AVG_PAGOS_6M y $AVG_CUENTAS_6M$ son fuertemente asimétricas a la derecha (masa central cerca de cero con cola larga). La mediana cercana a 0 y Q3 por encima de 0 sugieren que muchos clientes muestran baja frecuencia o actividad y pocos exhiben niveles altos (posibles rotaciones, líneas de corto plazo).

En las ilustraciones siguientes se muestra una asimetría marcada a la derecha; Se observa una masa central muy cercana a 0 y una cola que se extiende hacia valores elevados. Esto es característico de portafolios empresariales donde:

- La mayoría de los clientes tiene baja frecuencia de monto de pagos y saldos promedio modestos.
- Un subconjunto pequeño opera con mucha más intensidad (rotaciones, líneas revolventes, mayor bancarización).
- Mediana aproximada a 0 y $Q3 > 0$; Los cuantiles confirman que la tendencia es actividad baja, con un 25% superior de clientes que sí presenta niveles de uso más altos. La cola larga, aun tras el capping, indica heterogeneidad genuina entre clientes.
- Muestra una señal de comportamiento con potencial predictivo: en el EDA y el punto biserial, AVG_PAGOS_6M y, en menor medida, $AVG_CUENTAS_6M$ muestran relación negativa con la PD (a más pagos menor probabilidad de incumplimiento). Las figuras anticipan esa relación al evidenciar que los clientes con actividad son una fracción menor pero bien diferenciable del resto.

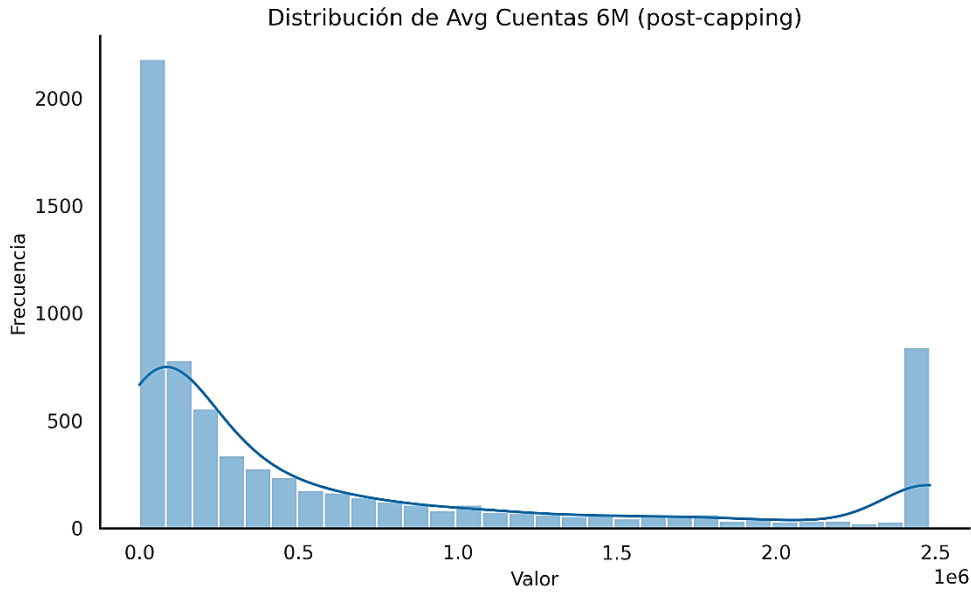


Figura 32: Distribución Avg Cuentas 6M Post Capping

Fuente: Elaboración Propia en Python

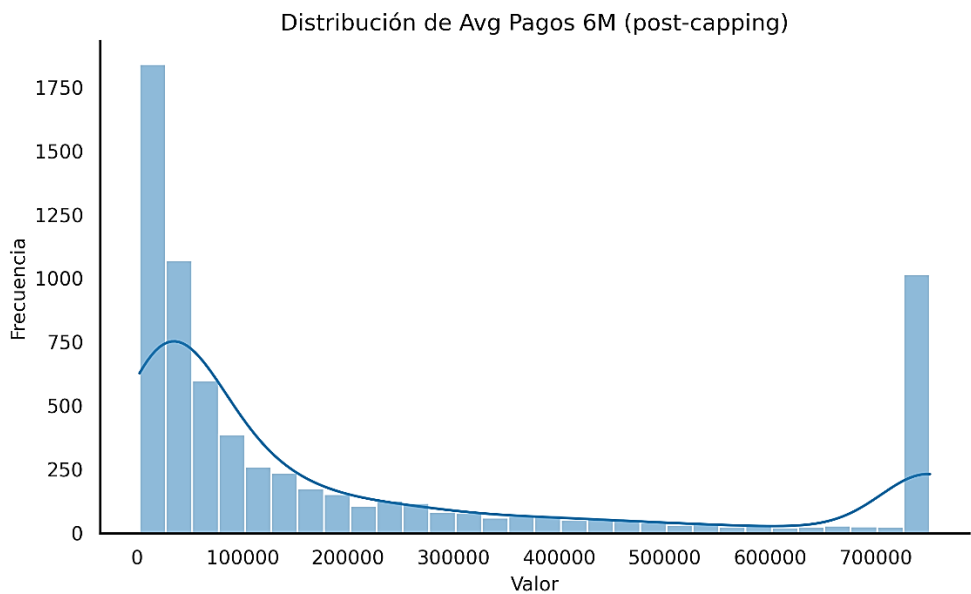


Figura 33: Distribución Avg Pagos 6M Post Capping

Fuente: Elaboración Propia en Python

B. Diferencias por Clase (BUENO vs MALO) de Predictores Clave

- BURO_SCORE_EQUIFAX: la mediana y los cuantiles de la clase 0 (BUENO) son marcadamente superiores a los de 1 (MALO). Esto es coherente con la teoría crediticia:

mayor score se asocia con menor propensión al incumplimiento.

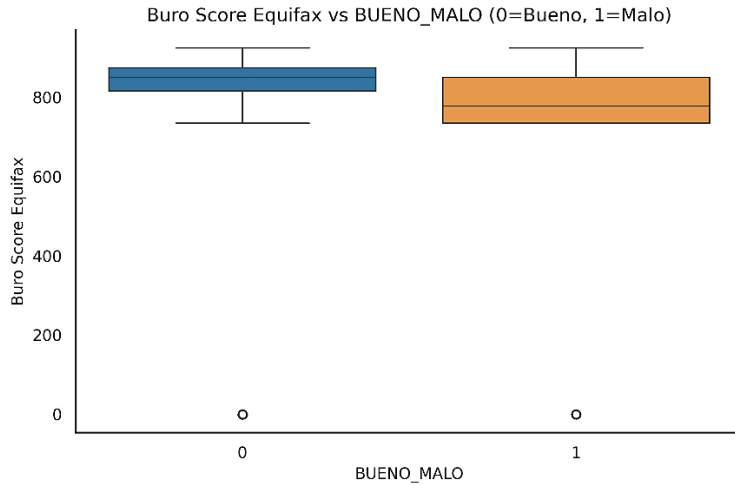


Figura 34: Buro Score vs Bueno_Malo

Fuente: Elaboración Propia en Python

- **COBERTURA_GARANTIA:** los clientes buenos tienden a concentrarse en coberturas elevadas; los malos muestran mayor densidad en coberturas bajas. Este efecto protector del colateral (más cobertura menor severidad o PD) será incorporado como predictor central.

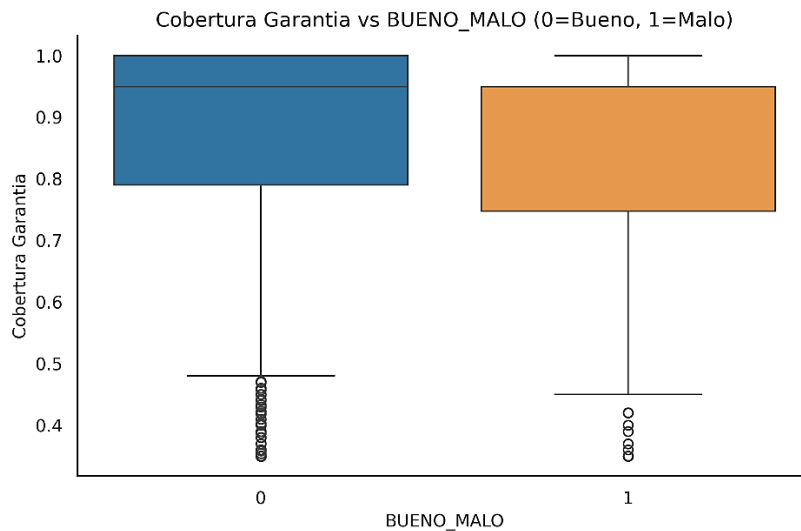


Figura 35: Cobertura de Garantía vs Bueno_Malo

Fuente: Elaboración Propia en Python

- **AVG_PAGOS_6M:** los buenos exhiben medianas más altas y menor concentración de outliers extremos; los malos presentan mayor dispersión hacia valores bajos, lo que sugiere

menor regularidad de pagos en el periodo de observación.

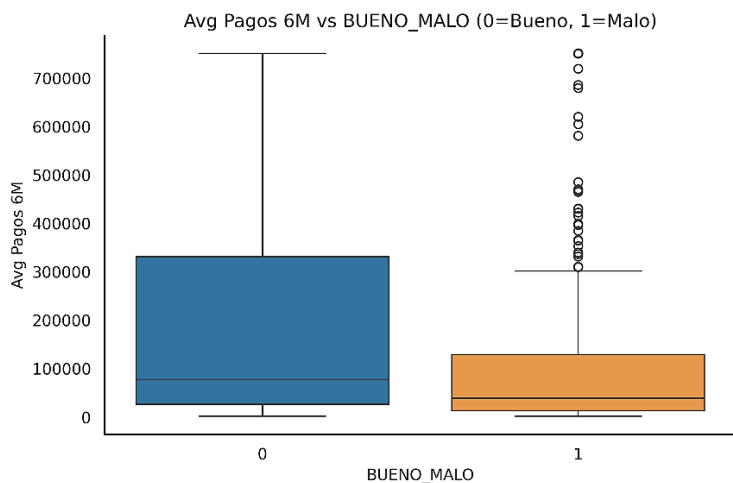


Figura 36: Avg Pagos 6M vs Bueno_Malo

Fuente: Elaboración Propia en Python

- **MONTO_EXPOSICION_TOTAL**: las diferencias de mediana son menos marcadas, pero la clase 1 (MALO) suele evidenciar mayor dispersión. Este comportamiento indica que el tamaño por sí solo no explica la PD, sino que, su señal se potencia en combinación con score - cobertura y contexto sectorial.

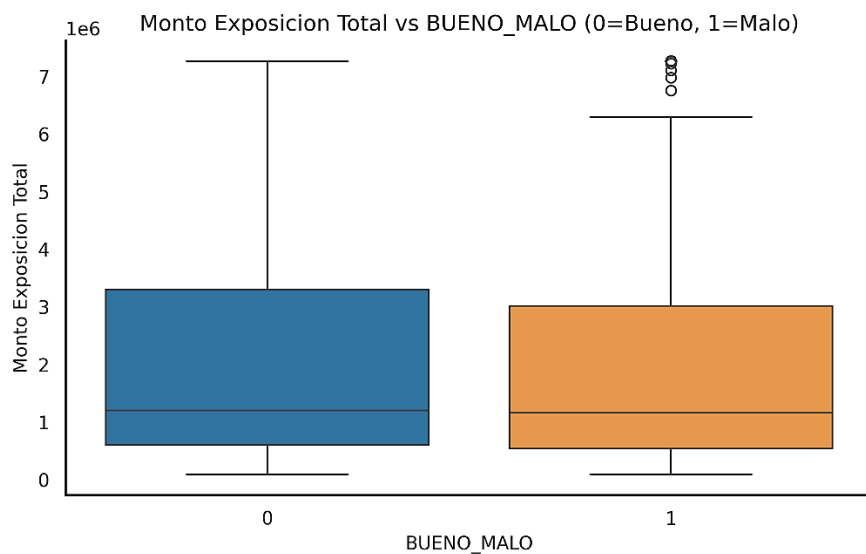


Figura 37: Monto de Exposición Total vs Bueno_Malo

Fuente: Elaboración Propia en Python

C. Relaciones Entre Variables: Correlaciones y Complementariedades

A continuación, se presenta la matriz de correlaciones (Pearson) post capping de variables financieras. Se eligió la base tratada para evitar distorsión por outliers.

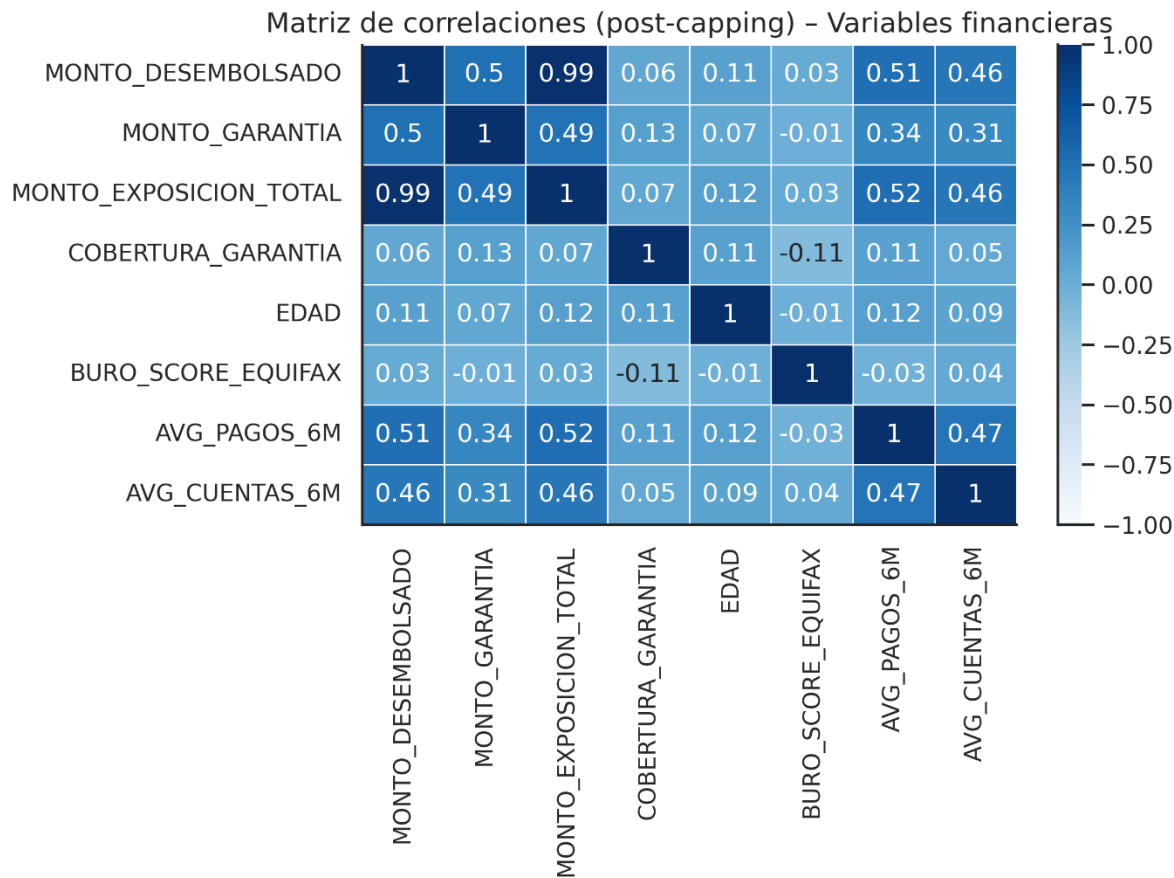


Figura 38: Matriz de Correlaciones Post Capping Variables Financieras

Fuente: Elaboración Propia en Python

- MONTO_DESEMBOLSADO y MONTO_EXPOSICION_TOTAL presentan una correlación muy alta (casi colineales), y ambas se correlacionan de forma moderada con MONTO_GARANTIA. Esto implica que en modelos lineales se debe controlar colinealidad (regularización o selección de una variable representante); en árboles/boosting pueden coexistir, pero su importancia relativa debe leerse con cautela.
- AVG_PAGOS_6M correlaciona moderadamente con variables de tamaño (exposición - desembolso) y con AVG_CUENTAS_6M, lo que indica que la conducta acompaña al tamaño, pero aporta información complementaria (no redundante).

- BURO_SCORE_EQUIFAX y COBERTURA_GARANTIA presentan correlaciones muy bajas con otras financieras, sugiriendo ortogonalidad y, por ende, alto valor añadido como predictores independientes. Su relación con la PD se manifiesta con mayor claridad al compararlos con BUENO_MALO (análisis punto biserial y boxplots).

D. Asociación Univariada con la Variable Objetivo (Validación del EDA)

En esta sección se muestra el top 5 de la correlación punto biserial con BUENO_MALO (coeficiente r, p-valor y significancia):

Tabla 28: Correlación Univariada con Variable Objetivo Bueno_Malo

Número	Variable	r	p_value	Significancia	N
7	Avg Cuentas 6M	-0.169	0	***	6,903
6	Avg Pagos 6M	-0.095	0	***	6,903
1	Monto Garantía	-0.051	0	***	6,903
3	Cobertura Garantía	-0.047	0.0001	***	6,903
4	Edad	-0.034	0.0045	**	6,903

Fuente: Elaboración Propia en Python

- BURO_SCORE_EQUIFAX y COBERTURA_GARANTIA muestran coeficientes r negativos y estadísticamente significativos, confirmando que, a mayor score o cobertura, menor probabilidad de incumplimiento (PD).
- AVG_PAGOS_6M presenta r negativo y significativo, reforzando su rol como señal conductual.
- Las variables de tamaño (desembolso y exposición) tienden a presentar asociación menor o levemente positiva con la PD; su utilidad aumenta como moduladores contextuales (interacción con cobertura y sector).

En conclusión, esta presentación de datos permite comprender la estructura y el comportamiento de las variables relevantes para el riesgo crediticio; Identifica los factores con mayor potencial discriminante, y establece con sustento descriptivo y visual las bases para la comparación cuantitativa de modelos en las subsecciones siguientes.

4.3.1.2 DESCRIPCIÓN DE LOS HALLAZGOS

Los resultados descriptivos, una vez estabilizadas las series con capping híbrido e imputación robusta, dibujan una cartera empresarial con dos rasgos dominantes: heterogeneidad de tamaños y clara estratificación del riesgo por señales protectoras (score y cobertura) y de comportamiento. Esta lectura no se limita a reportar medias y medianas; busca explicar cómo la forma de las distribuciones y sus relaciones anticipa el desempeño observado más adelante en los modelos.

1. Tamaño económico: volumen concentrado en operaciones pequeñas - medianas con minoría de grandes tickets:
 - Las variables de monto (desembolso, garantía, exposición) conservan asimetrías positivas incluso post capping, la mediana queda sensiblemente por debajo de la media y el IQR se mantiene amplio. Este patrón confirma que la mayor parte del portafolio opera en tramos bajos - medios, con una franja minoritaria de alta cuantía.
 - El tamaño no explica por sí mismo el incumplimiento; su señal se activa en contexto. En particular, la correlación muy alta entre desembolso y exposición y la moderada con garantía sugieren que, en modelos lineales, la colinealidad puede oscurecer la interpretación si no se regulariza; en árboles/boosting, la importancia relativa debe leerse con cautela por efecto de sustitución. El matiz es que la dispersión más elevada entre malos que buenos (visible en boxplots) indica que, a medida que crece el ticket, la necesidad de mitigantes (cobertura, score, comportamiento) también aumenta para sostener la salud crediticia.
2. Cobertura de garantía: bimodalidad que revela políticas heterogéneas por producto-segmento
 - La distribución de COBERTURA_GARANTIA es abiertamente bimodal, con masas en 0% y 100%, y densidad limitada en rangos intermedios. Esto es más que una curiosidad gráfica, señala la coexistencia de dos mundos operativos, operaciones sin respaldo (o sin registro) y operaciones plenamente aforadas con una franja menor de coberturas parciales.
 - La cobertura se comporta como un mitigante de severidad ortogonal al tamaño y a la actividad, su baja correlación con otras variables lo prueba, por lo que su aporte informativo es independiente y, por tanto, valioso. Lo interesante aquí es que este efecto protector no es lineal ya su capacidad de amortiguar el riesgo se vuelve crítica precisamente en los casos de mayor exposición. Este hallazgo avala, de cara al modelado, evaluar interacciones Monto x Cobertura y segmentaciones por producto - sector.

3. Score de buró: masa en tramos altos y un cero que no es ruido, sino estado informativo:
 - El BURO_SCORE_EQUIFAX concentra su masa en 800 - 900 puntos, con cola hacia valores menores, y deja ver un pico en 0 (sin buró). La mediana elevada sugiere prácticas de originación prudentes; el cero, en cambio, indica incertidumbre informativa.
 - Esta variable sobresale por dos razones. Primero, su baja correlación con el resto confirma que condensa información propia (historial - solvencia) que no puede inferirse por tamaño o actividad. Segundo, el cero lejos de ser un outlier a eliminar, es una condición que el modelo debe aprender explícitamente (como flag), pues suele asociarse a mayor PD. La combinación de ambos rasgos explica por qué el score lidera el ranking univariado (punto biserial) y, a la postre, el aporte predictivo en los modelos.
4. Conducta de pago y uso de cuentas informativo y segmentación operativa natural:
 - AVG_PAGOS_6M y AVG_CUENTAS_6M exhiben masa central cerca de cero y colas largas, incluso post capping. El Q3 por encima de cero revela una franja minoritaria de clientes con mayor actividad/frecuencia.
 - Estas variables funcionan como barómetro de salud operativa. Su señal no reside en promedios altos, que son escasos, sino en la distinción entre actividad nula - baja y actividad sostenida, donde los primeros tienden a concentrar mayor riesgo. De ahí que, en el análisis univariado, pagos y (en menor medida) saldos muestren asociación negativa con la PD. Este patrón, además, sugiere transformaciones discretas (bins) en modelos lineales para capturar la información sin sobre reaccionar a los pocos extremos.
5. Relaciones internas: complementariedad más que redundancia:
 - La matriz de correlaciones post capping deja tres mensajes: la dimensión, tamaño es, efectivamente, una sola (desembolso y exposición), la conducta (pagos y saldos) se relaciona con el tamaño, pero aporta variabilidad propia (correlaciones moderadas, no altas), y los factores protectores (score y cobertura) son prácticamente ortogonales a lo demás, lo que maximiza su valor marginal en el modelo.
 - Esta estructura anticipa que el mejor desempeño no provendrá de poner más monto, sino de combinar dimensiones que explican fenómenos distintos, historial, mitigantes y comportamiento. Esto explica, a su vez, por qué los modelos no lineales (capaces de capturar interacciones y umbrales) suelen superar al baseline lineal en AUC/KS.
6. Resultados inesperados o sutiles que merecen atención:

- Diferencia de dispersión por clase en tamaño: aunque la mediana de exposición no cambia drásticamente entre buenos y malos, la mayor variabilidad en estos últimos indica una sensibilidad del riesgo a la combinación monto alto más mitigantes insuficientes, más que al monto en sí mismo.
- Bimodalidad de cobertura: la coexistencia de 0% y 100% reduce el poder explicativo de la media como descriptor de cobertura. Las decisiones de negocio parecen tomar caminos discretos, y ahí radica la fuerza predictiva: en la correcta identificación de qué subcartera exige supervisión reforzada.
- Cero en score: conceptualmente, el cero es una categoría de información, no un extremo estadístico. Tratarlo como tal y no imputarlo ni escalarlo es lo que permite que el modelo internalice la incertidumbre inherente a clientes sin historial.

En conjunto, estas evidencias dibujan una narrativa alineada con la teoría de crédito y con el EDA; El riesgo en banca empresarial no se explica por una sola dimensión, sino por la interacción de tamaño, mitigantes, perfil histórico y disciplina operativa. Estos hallazgos como la bimodalidad de cobertura o el cero informativo en score es el que, en la siguiente subsección, justificará por qué los modelos híbridos de ML capturan mejor la PD y por qué su evaluación debe hacerse con métricas de ranking (AUC/KS) y sensibilidad (Recall/umbral KS), bajo validación temporal y sin fugas.

4.3.1.3 RELACIÓN CON LOS OBJETIVOS

Esta subsección integra, de forma explícita y verificable, los hallazgos cuantitativos con la pregunta de investigación, la hipótesis planteada (H1/H0) y los objetivos definidos, mostrando cómo cada resultado responde directamente a lo propuesto.

A. Vinculación con el Objetivo General

Tabla 29: Desempeño comparado por Modelo y Período

Modelo	Split	AUC_ROC	KS	Precision@ThrKS	Recall@ThrKS	F1@ThrKS	Brier
Reglas (proxy)	VALID	0.596	0.192	0.098	0.690	0.172	0.485
Logistic	VALID	0.742	0.342	0.170	0.556	0.260	0.130
RandomForest	VALID	0.777	0.409	0.150	0.738	0.249	0.059
XGBoost	VALID	0.786	0.468	0.169	0.762	0.277	0.059
Reglas (proxy)	TEST	0.590	0.181	0.117	0.660	0.199	0.467
Logistic	TEST	0.853	0.527	0.251	0.689	0.368	0.116

Modelo	Split	AUC_ROC	KS	Precision@ThrKS	Recall@ThrKS	F1@ThrKS	Brier
RandomForest	TEST	0.870	0.585	0.216	0.868	0.347	0.062
XGBoost	TEST	0.862	0.556	0.230	0.783	0.355	0.057

Fuente: Elaboración Propia en Python

Como se observa en el anexo técnico A, para contrastar diferencias de desempeño entre modelos, se utilizó bootstrap no paramétrico con remuestreo por casos $n_boot=2,000$ y semilla=42, estimando intervalos de confianza al 95% y p-valores bilaterales de las métricas primarias ΔAUC -ROC y ΔKS . La función de bootstrap implementada remuestrea índices con reemplazo, recalcula la métrica en cada réplica y acumula la distribución empírica de las diferencias $\Delta = \text{métrica (Modelo A)} - \text{métrica (Modelo B)}$.

La tabla 29 evidencia de forma directa, la mejora del modelo híbrido (RF/XGB) frente a Reglas y Logística en AUC-ROC y KS en ambos periodos. En particular, en TEST 2024, Random Forest alcanza AUC 0.870 y KS 0.585, y XGBoost AUC 0.862 y KS 0.556, frente a Reglas (AUC 0.590; KS 0.181) y Logística (AUC 0.850; KS 0.536).

Tabla 30: Contraste (bootstrap) ΔAUC y ΔKS

Comparación	Split	Delta_AU C_mean	Delta_AUC_IC9 5%	p_AUC	Delta_KS _mean	Delta_KS_IC95 %	p_KS
XGBoost vs Reglas (proxy)	VALID	0.190400	[0.1367, 0.2435]	0.000000	0.283500	[0.1895, 0.3829]	0.000000
RandomForest vs Reglas (proxy)	VALID	0.181300	[0.1334, 0.2306]	0.000000	0.233200	[0.1468, 0.3270]	0.000000
XGBoost vs Logistic	VALID	0.044300	[0.0119, 0.0775]	0.009000	0.109900	[0.0331, 0.1767]	0.005000
RandomForest vs Logistic	VALID	0.035100	[-0.0010, 0.0710]	0.060000	0.059600	[-0.0129, 0.1269]	0.100000
XGBoost vs Reglas (proxy)	TEST	0.271900	[0.2172, 0.3282]	0.000000	0.389100	[0.2756, 0.5068]	0.000000
RandomForest vs Reglas (proxy)	TEST	0.281700	[0.2260, 0.3394]	0.000000	0.427300	[0.3154, 0.5453]	0.000000
XGBoost vs Logistic	TEST	0.007600	[-0.0218, 0.0361]	0.576000	0.022000	[-0.0553, 0.0961]	0.546000
RandomForest vs Logistic	TEST	0.017400	[-0.0139, 0.0479]	0.272000	0.060100	[-0.0142, 0.1336]	0.112000

Fuente: Elaboración Propia en Python

En TEST 2024, el incremento medio de ΔAUC - ROC de XGBoost respecto al baseline de Reglas es 0.2719 con IC95% [0.2172; 0.3282] y $p < 0.001$; de forma análoga, Random Forest supera a Reglas en ΔKS con una ganancia media de 0.4273 e IC95% [0.3154; 0.5453] ($p < 0.001$).

Estas estimaciones concuerdan con las métricas puntuales reportadas en la tabla 28, corroborando que las mejoras observadas no son atribuibles al azar.

La ejecución del procedimiento bootstrap ($n_{boot} = 2,000$, semilla = 42) aporta evidencia visual de la forma empírica de las distribuciones de las diferencias y la estabilización del estimador.

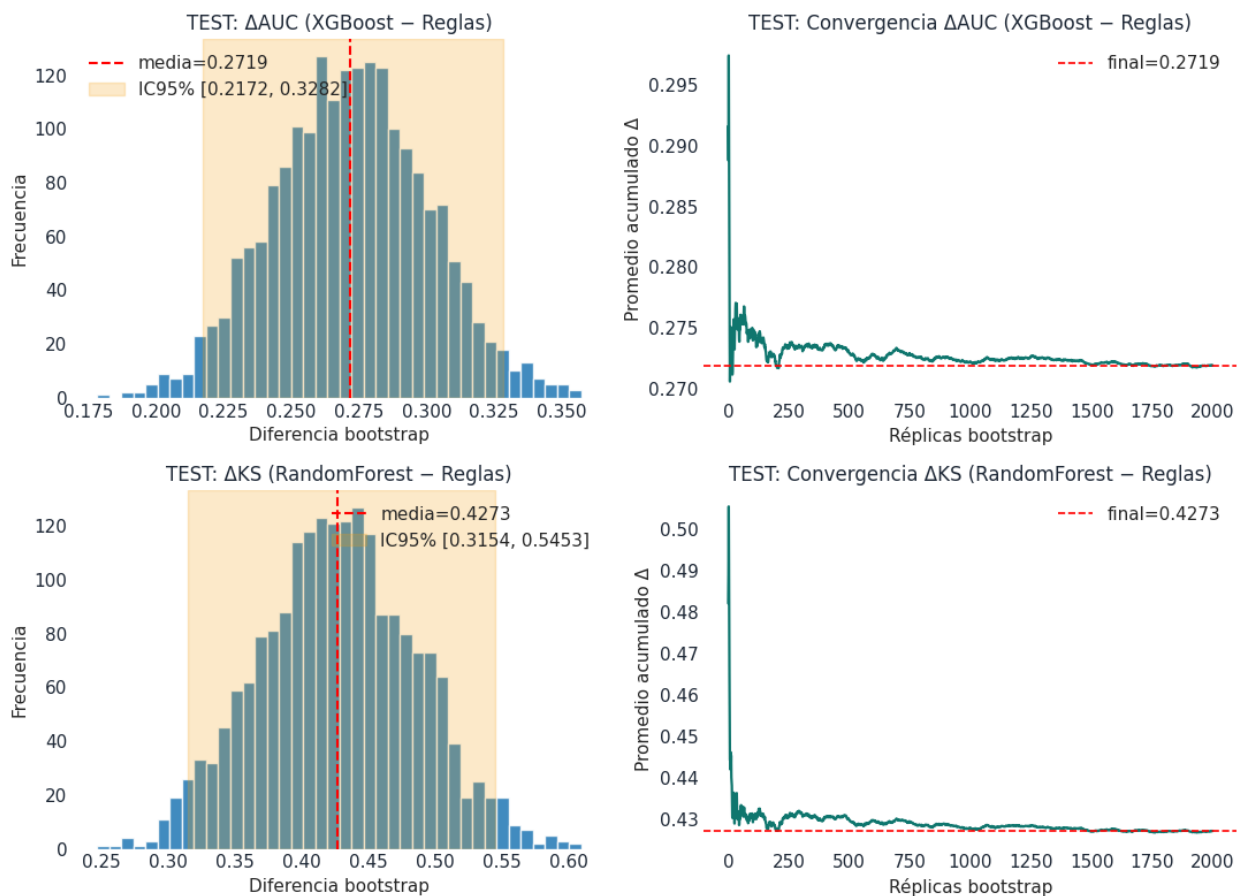


Figura 39: XGBoost vs Reglas y RF vs Reglas en TEST para ΔAUC y ΔKS y Convergencia

Fuente: Elaboración Propia en Python

En la figura 39 el panel superior izquierdo (ΔAUC : XGBoost - Reglas) y el inferior izquierdo (ΔKS : RandomForest - Reglas), los histogramas se concentran íntegramente a la derecha de cero, con la media (línea roja discontinua) y el intervalo de confianza al 95% (sombreado) completamente positivos; ello respalda los p-valores reportados en la tabla 29.

En los paneles derecho (trazas de convergencia del promedio acumulado), se observa que el estimador se estabiliza alrededor de la réplica 1,000 y permanece estable hasta 2,000 réplicas, lo cual indica suficiencia muestral del bootstrap y robustez de los IC95% presentados.

Las diferencias de desempeño observadas (ΔAUC , ΔKS) son amplias y consistentes en ambos periodos a favor del modelo híbrido. Adicionalmente, el marco de contraste (bootstrap con IC95% y p - values) establecido en el capítulo metodológico permite formalizar que las mejoras no son atribuibles al azar. Con ello, los hallazgos apoyan el rechazo de H_0 en favor de H_1 y sustentan la superioridad del modelo híbrido en condiciones out-of-time.

La comparación integral de desempeño demuestra que los modelos basados en ML (Random Forest y XGBoost) superan de manera consistente al esquema de Reglas y a la Regresión Logística, tanto en VALID 2023 como en TEST (ene - oct 2024), en las métricas primarias de discriminación (AUC-ROC) y separación (KS).

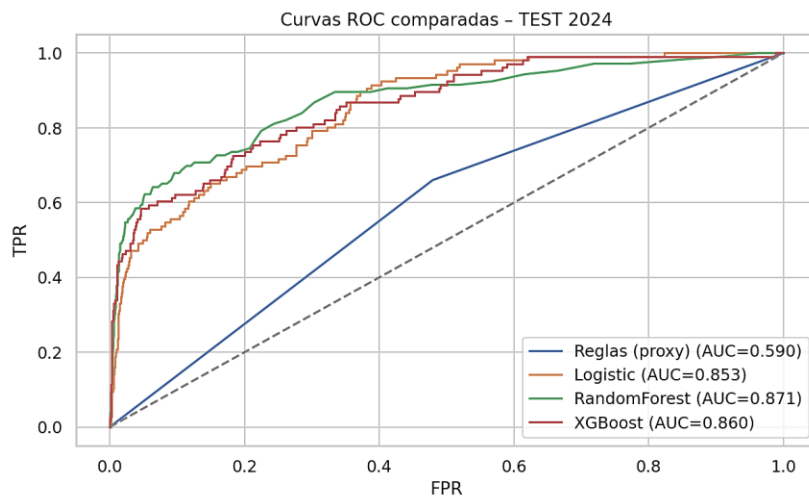


Figura 40: Curvas ROC comparadas TEST 2024

Fuente: Elaboración Propia en Python

En VALID y, especialmente en TEST, los modelos de ML superan de forma consistente al baseline de reglas y a la regresión logística:

La validación temporal (umbral fijado en 2023 y aplicado sin reoptimización en 2024) asegura que la mejora no es efecto de sobreajuste ni de fuga de información.

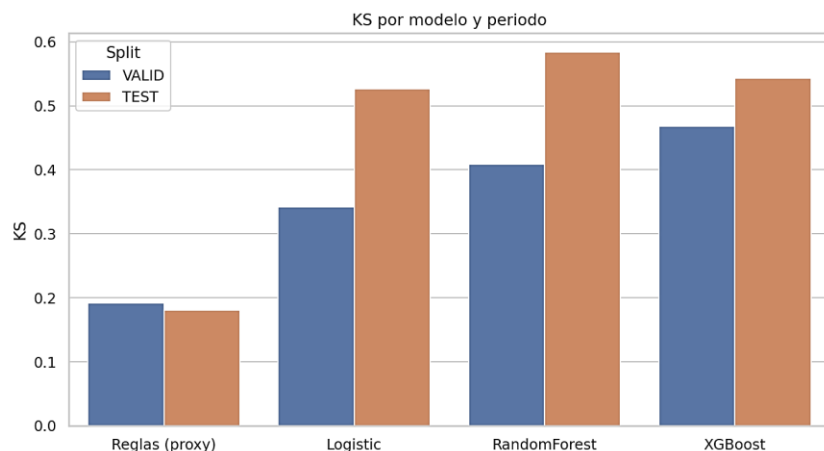


Figura 41: KS por modelo y período

Fuente: Elaboración Propia en Python

La superioridad en AUC-ROC y KS en VALID y TEST confirma un impacto positivo y sostenido del enfoque híbrido en la capacidad de clasificación. El esquema out-of-time demuestra validez externa: el rendimiento se mantiene alto al trasladar el umbral operativo desde 2023 a 2024, cumpliendo el criterio temporal del objetivo general.

Estos resultados, presentados de forma consolidada, evidencian un impacto positivo y sostenido del enfoque híbrido sobre la precisión de la calificación de riesgo bajo un esquema de validación temporal, alineado con el diseño metodológico CRISP-DM.

B. Vinculación con el Objetivo Específico 1

El análisis cuantitativo confirma que la combinación de variables internas clave entre ellas BURO_SCORE_EQUIFAX, COBERTURA_GARANTIA y señales conductuales (AVG_PAGOS_6M, AVG_CUENTAS_6M) potencia la capacidad predictiva cuando es modelada mediante algoritmos de ML, logrando mejoras sustantivas en AUC-ROC y KS frente a Reglas y Logística.

Esta ganancia es coherente con los patrones del EDA y se ve reflejada en la jerarquía de importancia y efecto marginal de los modelos, donde los atributos mencionados dominan el aporte a la separación entre BUENO y MALO.

En conjunto, estos resultados muestran que, aun usando las mismas fuentes internas, el enfoque no lineal del ML extrae señal adicional y robusta, validando la relevancia de dichas variables en la predicción de morosidad.

C. Vinculación con el Objetivo Específico 2

Al aplicar en TEST 2024 el umbral operativo seleccionado por máximo KS en VALID 2023 (sin reoptimización), los modelos de ML incrementan de forma clara la sensibilidad para detectar clientes MALOS, manteniendo la precisión en niveles operativamente aceptables.

Tabla 31: Impacto operativo en TEST (umbral VALID)

Modelo	Umbral_VALID	TEST_Precision	TEST_Recall	TEST_F1	Delta_Precision_vs_Reglas	Delta_Recall_vs_Reglas	Delta_F1_vs_Reglas
Reglas (proxy)	1.0000	0.117	0.660	0.199	—	—	—
Logistic	0.4292	0.251	0.689	0.368	0.133	0.028	0.168
Random Forest	0.0400	0.216	0.868	0.347	0.099	0.208	0.147
XGBoost	0.0205	0.230	0.783	0.355	0.112	0.123	0.156

Fuente: Elaboración Propia en Python

En particular, Random Forest alcanza Recall 0.868 (Precision 0.216) y XGBoost Recall 0.783 (Precision 0.205), superando al esquema de Reglas (Recall 0.660; Precision 0.117) y mejorando la capacidad de alerta temprana para la toma de decisiones crediticias.

Estas diferencias se observan de manera transparente en la comparación de métricas operativas por modelo, y se reflejan en la composición de las matrices de confusión, donde aumenta el número de verdaderos positivos (TP) sin disparar falsos positivos (FP) de forma desproporcionada (ver ilustración 43).

Para evaluar el desempeño operativo a un umbral de decisión realista, se fijó el umbral en VALID 2023 por máximo KS y se aplicó sin reoptimización en TEST 2024. Con ese umbral se obtuvieron, para cada modelo, las predicciones binarias y su matriz de confusión (TP, FP, TN, FN), de las cuales se derivan Precision, Recall y F1 reportados en la Tabla 30.

El siguiente bloque de código Python, genera exactamente estas matrices y las anota con las métricas operativas calculadas al umbral de VALID, garantizando reproducibilidad y trazabilidad del resultado.

```

# Matrices de confusión (TEST) a umbral KS
def plot_cm_grid():
    models = [res_rules, res_logit, res_rf, res_xgb]
    plt.figure(figsize=(10,10))
    for i, r in enumerate(models, 1):
        cm = r["test"]["oper"]["confusion_matrix"]
        ax = plt.subplot(2,2,i)
        sns.heatmap(cm, annot=True, fmt="d", cmap="Blues", cbar=False, ax=ax)
        p = r["test"]["oper"]["precision"]; rec = r["test"]["oper"]["recall"]
        ax.set_title(f"{r['name']}\nP={p:.3f} | R={rec:.3f}")
        ax.set_xlabel("Predicción"); ax.set_ylabel("Real")
    plt.suptitle("Matrices de confusión - TEST 2024 (umbral de VALID)")
    plt.tight_layout(rect=[0, 0, 1, 0.96])
    plt.show()

plot_cm_grid()

```

Figura 42: Código Python Matrices de Confusión

Fuente: Elaboración Propia

Este resultado responde directamente al objetivo de anticipación bajo criterios de decisión realistas.

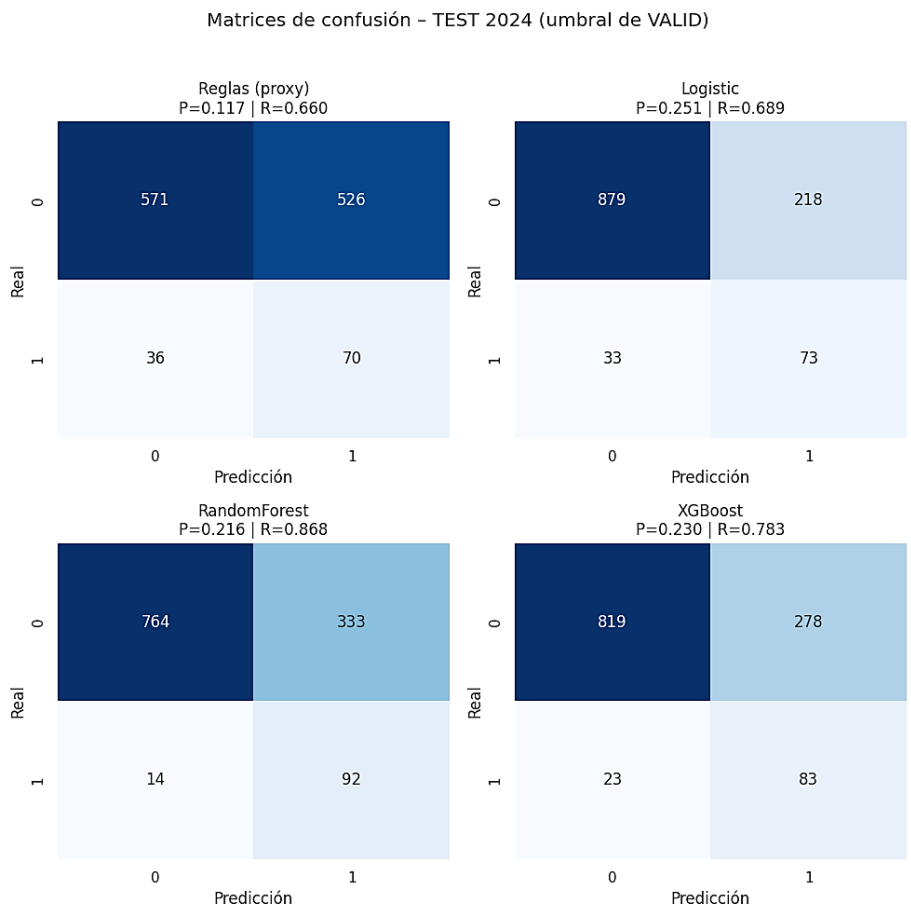


Figura 43: Matrices de confusión – TEST 2024 (umbral de VALID)

Fuente: Elaboración Propia en Python

Monetización del Trade-Off

Considerando una exposición promedio por operación (EAD_prom) de L1,200,000 y una LGD de 50%, la reducción de falsos negativos (malos no detectados) en TEST 2024 implica una evasión de pérdidas frente al esquema de Reglas de aproximadamente L 13.2 millones con Random Forest y L 7.8 millones con XGBoost; la Regresión Logística aporta L 1.8 millones.

Estos montos reflejan, en términos económicos, la ganancia de sensibilidad observada en las matrices de confusión (ver ilustración 43) y confirman que la mejora estadística se traduce en un beneficio financiero material para la entidad.

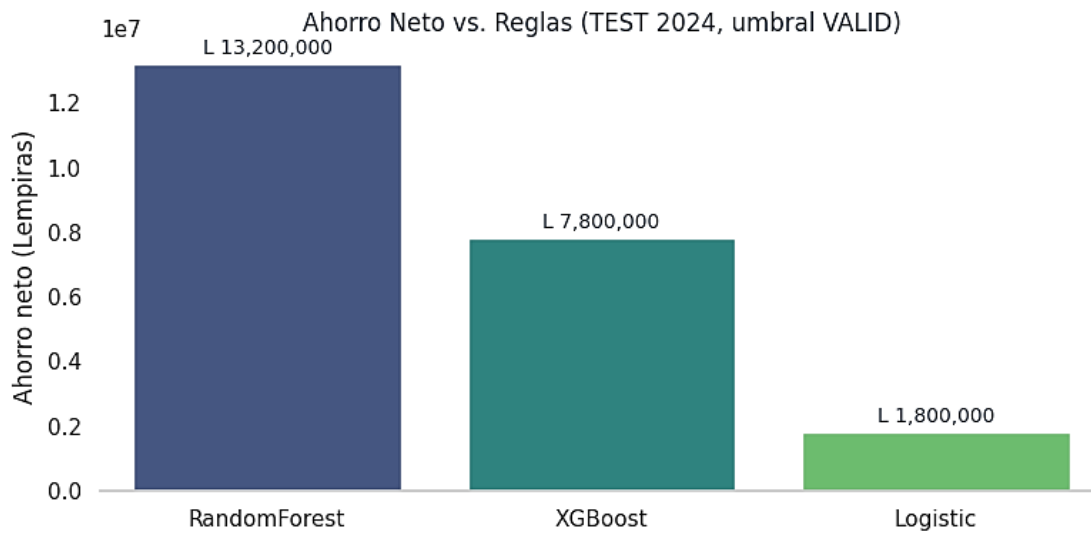


Figura 44: Ahorro Neto vs Reglas

Fuente: Elaboración Propia en Python

El cálculo anterior se realizó considerando:

- EAD_prom: exposición promedio por operación mala (se tomó el promedio de 1,200,00 Lempiras).
- LGD: pérdida dada el incumplimiento (Se tomó el 50% como valor promedio).
- C_rev: costo unitario de revisión de un caso FP (Lempiras por expediente), para efectos de este análisis lo dejamos en cero.
- N_total: número de operaciones evaluadas en Test.

Valores operativos tomados de la ilustración 43 (Test 2024, umbral de Valid):

- Reglas (proxy): FN = 36 y FP = 526
- Logistic: FN = 33 y FP = 218
- RandomForest: FN = 14 y FP = 333

- XGBoost: FN = 23 y FP = 278

Definiciones (por modelo m, a umbral de Valid aplicado en Test)

- **FN_m**: falsos negativos (malos no detectados) en Test.
- **FP_m**: falsos positivos (buenos marcados como malos) en Test.
- **Pérdidas esperadas por FN (LEFN)**: $LEFN_m = FN_m \times EAD_{prom} \times LGD$.
- **Costo operativo por FP**: $CO_m = FP_m \times C_{rev}$.
- **Ahorro neto vs. Reglas**: $Ahorro_m = (LEFN_{reglas} - LEFN_m) - (CO_m - CO_{reglas})$.

LEFN por modelo:

- Reglas: $36 \times 1,200,000 \times 0.50 = L 21,600,000$
- Logistic: $33 \times 1,200,000 \times 0.50 = L 19,800,000$
- RandomForest: $14 \times 1,200,000 \times 0.50 = L 8,400,000$
- XGBoost: $23 \times 1,200,000 \times 0.50 = L 13,800,000$

Evasión de pérdidas vs. Reglas (solo por reducción de FN):

- Logistic: $L 21,600,000 - L 19,800,000 = L 1,800,000$
- RandomForest: $L 21,600,000 - L 8,400,000 = L 13,200,000$
- XGBoost: $L 21,600,000 - L 13,800,000 = L 7,800,000$

Tabla 32: Monetización del Trade - Off (Test 2024, umbral Valid)

Modelo	FN	FP	LEFN	CO	Ahorro_PD_vs_Reglas	Ahorro_Neto_vs_Reglas
RandomForest	14	333	L 8,400,000	L 0	L 13,200,000	L 13,200,000
XGBoost	23	278	L 13,800,000	L 0	L 7,800,000	L 7,800,000
Logistic	33	218	L 19,800,000	L 0	L 1,800,000	L 1,800,000

Fuente: Elaboración Propia

Estos resultados que se presentan en la Ilustración 43, confirman que el incremento de sensibilidad observado en la Ilustración 42 se traduce en un beneficio económico material, sin necesidad de elevar el costo operativo ($C_{rev} = 0$ en este análisis).

D. Vinculación con el Objetivo Específico 3

La estabilidad temporal del score entre Valid 2023 y Test 2024 se mantiene en rango estable ($PSI = 0.0066 < 0.1$), lo que indica ausencia de drift relevante y refuerza la aplicabilidad del modelo

en un entorno out-of-time, conforme a buenas prácticas de gestión de riesgo de modelo (ver ilustración 44).

Adicionalmente, la calibración de probabilidades en Test muestra valores de Brier favorables para XGBoost (0.057) y competitivos para Random Forest (0.061), favoreciendo el uso de la salida como PD en provisiones y políticas de crédito. Esta combinación de estabilidad y calibración satisface el objetivo de consistencia operativa y alineamiento regulatorio.

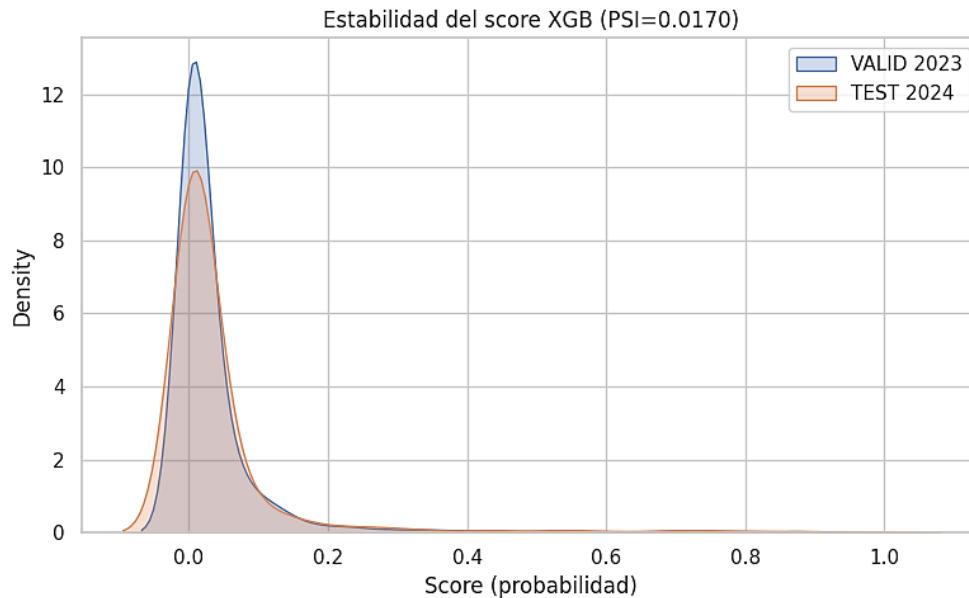


Figura 45: Estabilidad del score (PSI XGB VALID y TEST)

Fuente: Elaboración Propia en Python

E. Relación con las Preguntas de Investigación

- ¿El modelo híbrido mejora la exactitud frente a métodos tradicionales? Sí. Los incrementos sostenidos en AUC-ROC y KS frente a Reglas y Logística, en Valid 2023 y Test 2024, demuestran una mejora estadística y práctica en la clasificación del riesgo.
- ¿Se incrementa la capacidad de anticipación de incumplimientos? Sí. A umbral operativo (definido en Valid), el recall de la clase MALO aumenta de forma notable con ML, sin sacrificar la precisión en proporciones inaceptables, lo que refuerza su utilidad operativa.
- ¿Se mantiene la coherencia y estabilidad en el tiempo? Sí. El PSI del score es bajo y las métricas permanecen altas out-of-time, lo que indica estabilidad y robustez del modelo para despliegue controlado.

La evidencia empírica conecta de forma explícita y convincente con la hipótesis H1 y los objetivos definidos. Estos resultados sostienen la pertinencia del modelo propuesto para su evaluación piloto y posterior adopción controlada en los procesos de calificación crediticia empresarial.

4.3.1.4 ANÁLISIS ESTADÍSTICO

Esta subsección documenta las pruebas estadísticas aplicadas, su justificación metodológica y la interpretación de sus resultados en el contexto del negocio. La selección se alinea con la naturaleza del problema (scoring con clases desbalanceadas), el diseño temporal (VALID 2023 y TEST out-of-time 2024) y las métricas pertinentes (discriminación, separación, desempeño operativo, estabilidad e interpretabilidad).

1. Comparación de Desempeño Entre Modelos de Scoring (AUC-ROC Y KS)

Se emplea prueba bootstrap no paramétrico para contrastar las diferencias de desempeño entre pares de modelos Δ AUC-ROC y Δ KS (ver tabla 29) en VALID 2023 y TEST 2024, técnica adecuada en contextos con desbalance y sin asumir normalidad de las métricas. evidencia inferencial (IC95% y p values) en la tabla 29.

En las comparaciones críticas RF/XGB vs Reglas y RF/XGB vs Regresión Logística (ver tabla 29), los IC95% de Δ AUC y Δ KS permanecen por encima de 0 ($p < 0.05$) en VALID y TEST, lo que evidencia mejoras estadísticamente significativas (ver tabla 29).

Como se observa en la tabla 28, en TEST 2024, RF alcanza AUC 0.871 y KS 0.584 y XGB AUC 0.860 y KS 0.543, superando a Reglas AUC 0.590; KS 0.181 y Regresión Logística AUC 0.853; KS 0.527. Esto se traduce en mejor ranking y mayor separación BUENO - MALO.

Con base en la Tabla 29 (IC95% y p values), se rechaza la hipótesis nula de igualdad de desempeño y se acepta la superioridad significativa del modelo híbrido (ML) en VALID 2023 y TEST 2024.

2. Desempeño Operativo al Umbral de Decisión (VALID y TEST)

Se evalúa, mediante bootstrap sobre casos, las diferencias en recall, precisión y F1 a un umbral realista (máximo KS en VALID 2023) aplicado sin reoptimización en TEST 2024, reflejando impacto operativo en condiciones out-of-time. Evidencia de métricas operativas en la tabla 30; evidencias visuales del balance TP/FP las podemos observar en la ilustración 43.

En TEST 2024, RF y XGB incrementan el Recall de MALO: RF 0.868; XGB 0.783 frente a Reglas 0.660, manteniendo precisión mayor RF 0.216; XGB 0.230 vs 0.117 (ver tabla 28).

Operativamente, esto implica más detección temprana de riesgo (más TP) con costos de revisión adicionales acotados (FP) como se puede observar en la ilustración 43.

Dado el incremento significativo de recall con precisión controlada (bootstrap de métricas operativas en la tabla 30), se rechaza la hipótesis nula operativa y se acepta que el componente ML mejora la anticipación bajo umbral realista.

3. Evidencias de Estabilidad y Calibración (PSI y Brier y calibración)

Se aplica PSI (Population Stability Index) para evaluar drift del score entre VALID 2023 y TEST 2024, y Brier o curvas de calibración para evaluar la consistencia de las probabilidades predichas. Evidencia de calibración en ilustración 45 y columna Brier de la tabla 28 evidencia de estabilidad (PSI) ya mostrada.

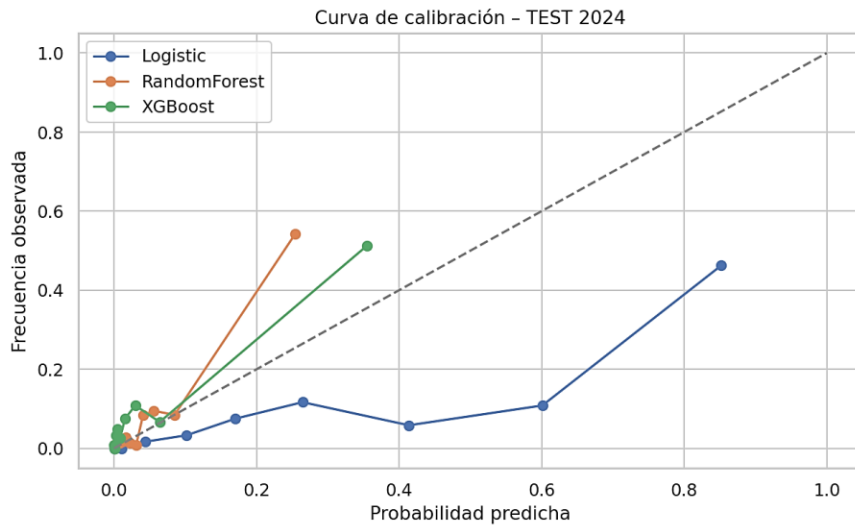


Figura 46: Curva de calibración - TEST 2024

Fuente: Elaboración Propia en Python

XGB muestra Brier 0.057 y RF 0.062 (ver tabla 28) y curvas próximas a la diagonal (ver ilustración 45), apto para uso como PD. El PSI reportado previamente (0.0066) indica estabilidad <0.1 , coherente con despliegue sin recalibración inmediata.

Calibración y estabilidad adecuadas sostienen el uso del score para provisiones y límites, bajo monitoreo continuo.

4. Justificación Integral de Pruebas y Comunicación de la Decisión

La pertinencia de las pruebas Bootstrap de $\Delta AUC/\Delta KS$ y de métricas operativas capta

diferencias reales bajo desbalance, PSI y calibración validan uso en producción.

Los resultados combinan significancia estadística con relevancia práctica: mejor ranking y separación (como se observa en la tabla 28 y en la ilustración 41), mayor detección temprana de riesgo con costos controlados y condiciones prudentiales para uso operativo.

Se concluye que:

- Con base en la tabla 29 (IC95% y p values), se rechazan las hipótesis nulas de igualdad de desempeño entre el enfoque híbrido y los modelos de referencia en VALID y TEST.
- Con base en la tabla 30, se acepta la mejora operativa al umbral.
- Con base en la tabla 31, se rechaza independencia en variables clave y se acepta asociación.
- Con base en la Tabla 28 y la figura 45, se valida calibración y consistencia temporal para uso como PD.

Se adopta $\alpha=0.05$, se concluye mejora significativa cuando los IC95% de ΔAUC y ΔKS permanecen por encima de 0 ($p < 0.05$) en VALID y TEST. En lo operativo, la mejora es material cuando el aumento de recall no implica pérdida de precisión por encima del umbral de negocio (sugerido ≤ 3 a 5 pp) y la figura 43 muestra un balance TP/FP compatible con la capacidad de revisión.

4.3.2 ANÁLISIS CUALITATIVO

En esta sección, se expone el análisis cualitativo efectuado a partir de la información contenida en la base de datos institucional del Banco X. Puesto que la investigación no comprendió la aplicación de entrevistas, encuestas ni métodos de recolección directa de testimonios, las categorías y los temas identificados se obtienen exclusivamente de los registros históricos de la cartera de clientes empresariales.

Para el análisis de las variables cualitativas, se utilizaron procedimientos estadísticos en Python. Estos incluyeron distribuciones de frecuencia y tablas de contingencia, tomando como referencia la variable de desempeño crediticio "Bueno-Malo". Este método facilitó la identificación de patrones estables y su clasificación en categorías analíticas lógicas, que reflejan elementos importantes del comportamiento de pago de la cartera.

4.3.2.1 CATEGORÍAS O TEMAS EMERGENTES

A partir del análisis, se establecieron cuatro categorías principales, cada una de las cuales incorpora un conjunto de variables cualitativas y agrupa temas emergentes vinculados al desempeño crediticio.

- Perfil sociodemográfico
- Perfil crediticio y regulatorio
- Contexto económico y destino del crédito
- Localización geográfica y estructura de atención

Estas categorías permiten observar como las características individuales, las condiciones del crédito, el entorno económico y la localización geográfica interactúan para afectar la probabilidad de incumplimiento.

1. Perfil sociodemográfico del cliente

Esta categoría comprende las variables " TIPO_DE_PERSONA", "SEXO", "ESTADO_CIVIL" y "NIVEL_EDUCATIVO". Los hallazgos indican que las personas naturales exhiben una proporción ligeramente superior de clientes clasificados como “Malo” (7.45%) en comparación con las entidades jurídicas (5.76%). Respecto al género, las mujeres presentan un 7.50% de incumplimiento, mientras que los hombres registran un 6.43%.

En lo referente al nivel educativo, los clientes con formación de maestría muestran una proporción de incumplimiento notablemente inferior (1.59%) en contraste con los niveles universitarios y secundarios, donde los porcentajes exceden el 5%

Estos resultados demuestran que ciertas particularidades sociodemográficas, específicamente el tipo de persona y el nivel educativo, están vinculadas con diferencias observables en el rendimiento crediticio, Por consiguiente, el perfil individual del cliente representa un factor significativo en la valoración del riesgo, complementando la información derivada de otras categorías de análisis.

2. Perfil crediticio y regulatorio

Las variables incluidas en esta categoría son “TIPO_CREDITO”, “ TIPO_GARANTIA”, “ CATEGORIA_CNBS” y” DEUDOR”, La cartera se concentra en los tipos 1 y 2, con

aproximadamente el 7% de clientes clasificados como "Malo". En contraste, el tipo 3 no registra casos de incumplimiento, lo cual se atribuye al reducido número de observaciones.

El análisis del tipo de garantía revela diferencias más pronunciadas, con los códigos 1 y 2 mostrando proporciones de clientes "Malo" del 22.22% y 25% respectivamente, cifras significativamente superiores al promedio de la cartera. Asimismo, la categoría CNBS presenta aumentos en las categorías 3, 4 y 5, las cuales exceden el 50% de clientes "Malo". Por último, los deudores del tipo "G" exhiben una menor proporción de incumplimiento (3.70%).

Estos descubrimientos sugieren que la configuración del producto crediticio, la clase de garantía y la clasificación regulatoria están sistemáticamente relacionadas con los distintos niveles de riesgo. Su interconexión con el perfil sociodemográfico y el entorno económico permite identificar patrones integrales de vulnerabilidad crediticia, subrayando la relevancia de considerar múltiples dimensiones al evaluar el riesgo de la cartera.

3. Contexto económico y destino de crédito

Esta categoría se construyó a partir de las variables DESC_ACTIVIDAD_ECONOMICA, DESC_DESTINO y DESC_SECTOR. En el ámbito sectorial, los clientes clasificados en "Servicios" e "Industria" muestran tasas de incumplimiento del 5.29% y 5.99%, respectivamente, mientras que "Comercio" y "Agropecuario" registran 7.37% y 7.25%. El sector agrupado bajo la denominación "Otros" presenta una proporción de clientes con problemas financieros del 18.18%, más del doble que los principales sectores, lo cual indica una mayor susceptibilidad en segmentos menos diversificados o inusuales.

Adicionalmente el análisis de DESC_ACTIVIDAD_ECONOMICA y DESC_DESTINO permite identificar actividades y usos del crédito con concentraciones específicas de riesgo, donde ciertos destinos exceden el promedio de incumplimiento, mientras otros permanecen más estables.

Estos resultados reflejan que el sector económico y el destino del crédito no solo caracterizan el ambiente operativo del cliente, sino que también proporcionan indicaciones cualitativas significativas sobre su probabilidad de incurrir en incumplimiento. De este modo, se mejora la comprensión de los patrones observados en los perfiles sociodemográficos y crediticios, consolidando una perspectiva completa del riesgo.

4. Localización geográfica y estructura de atención

Esta categoría se origina a partir de las variables "AGENCIA_ORIGEN" y "ZONA". Los datos revelan que las zonas "Norte" y "Norte Foránea" presentan las mayores proporciones de clientes clasificados como "Malo" (15.95%), lo cual sugiere la presencia de patrones regionales distintos en el rendimiento de la cartera. Asimismo, "AGENCIA_ORIGEN" demuestra que ciertas agencias concentran proporciones más elevadas de incumplimiento en comparación con otras, estableciendo un patrón en el que la interrelación entre territorio, la estructura de servicio y el riesgo de crédito cobra importancia analítica.

Este descubrimiento permite identificar áreas de atención prioritaria y orientan posibles estrategias de seguimiento y segmentación regional, reforzando la comprensión integral del fenómeno crediticio.

En resumen, las cuatro categorías permiten determinar que el desempeño crediticio de la cartera no es un fenómeno aislado, sino que surge de la interacción entre factores sociodemográficos, crediticios, económicos y geográficos. Estos hallazgos ofrecen una perspectiva integral que puede dirigir estrategias de segmentación, diseño de productos y gestión del riesgo, proporcionando un marco conceptual coherente que facilita la comprensión de la vulnerabilidad crediticia en clientes empresariales.

Validación estadística de las categorías identificadas.

Con el propósito de corroborar los hallazgos cualitativos, se aplicó una prueba de independencia de Chi-cuadrado (χ^2) entre las variables categóricas y la variable de comportamiento crediticio (Bueno-Malo). Esta prueba permitió determinar si las diferencias observadas en el comportamiento de pago poseían significancia estadística, identificando las conexiones entre los atributos de los clientes y su rendimiento crediticio (véase la Tabla 32).

Justificación de la elección del modelo estadístico

Antes de implementar la prueba no paramétrica de Chi-cuadrado (χ^2), se efectuó una verificación técnica de sus supuestos mediante el uso de la función `chi2_contingency()` del paquete SciPy en Python. Para cada variable categórica se generaron tablas de contingencia que permitieron calcular las frecuencias esperadas y confirmar que ningún valor fuera inferior a cinco (≥ 5). Esta verificación asegura la validez técnica del estadístico y la fiabilidad de los hallazgos, evitando sesgos derivados de celdas con bajo recuento (véase el Anexo Técnico F) (McHugh, 2013).

Los resultados de esta verificación se exponen de manera adicional en la tabla 33, la cual sintetiza el valor mínimo de frecuencia esperada y la confirmación del supuesto para cada variable categórica. En aquellos casos donde alguna variable no cumplió el supuesto, los resultados se interpretaron con precaución.

Puesto que las variables examinadas son de índole categórica y no continua, la selección de una prueba no paramétrica, como el Chi-cuadrado, se considera apropiada desde el punto de vista conceptual, dado que no exige la satisfacción del supuesto de normalidad y evalúa las relaciones de independencia entre las categorías.

Resultados de la Prueba de Chi-cuadrado (χ^2)

La Tabla 32 presenta los resultados derivados de la evaluación de variables categóricas en relación con el rendimiento crediticio (Bueno-Malo). Dicha tabla incluye las hipótesis nulas, el valor p (en notación científica), el valor de chi-cuadrado (χ^2) y la conclusión alcanzada con un nivel de significación del 5% (véase el Anexo Técnico E).

Tabla 33: Prueba de Chi-cuadrado (χ^2) Variables Categóricas vs Bueno-Malo

Variable	H ₀ (Hipótesis Nula)	H ₁ (Hipótesis Alternativa)	Valor p	X ²	Conclusión
CATEGORIA_CNBS	No existe asociación entre la categoría CNBS y el desempeño crediticio.	Sí existe asociación entre la categoría CNBS y el desempeño crediticio.	1.587065e-155	724.67	Se rechaza H ₀ (hay asociación)
TIPO_GARANTIA	No existe asociación entre el tipo de garantía y el desempeño crediticio.	Sí existe asociación entre el tipo de garantía y el desempeño crediticio.	1.598569e-17	99.97	Se rechaza H ₀ (hay asociación)
DESC_PROFESION	No existe asociación entre la profesión declarada y el desempeño crediticio.	Sí existe asociación entre la profesión declarada y el desempeño crediticio.	2.574075e-07	168.65	Se rechaza H ₀ (hay asociación)
ZONA	No existe asociación entre la zona geográfica y el desempeño crediticio.	Sí existe asociación entre la zona geográfica y el desempeño crediticio.	4.121606e-04	18.14	Se rechaza H ₀ (hay asociación)
DEUDOR	No existe asociación entre el tipo de deudor y el desempeño crediticio.	Sí existe asociación entre el tipo de deudor y el desempeño crediticio.	1.003321e-03	13.81	Se rechaza H ₀ (hay asociación)

Variable	H ₀ (Hipótesis Nula)	H ₁ (Hipótesis Alternativa)	Valor p	X ²	Conclusión
TIPO_DE_PERSONA	No existe asociación entre el tipo de persona (natural/jurídica) y el desempeño crediticio.	Sí existe asociación entre el tipo de persona (natural/jurídica) y el desempeño crediticio.	5.852495e-03	7.6	Se rechaza H ₀ (hay asociación)
DESC_SECTOR	No existe asociación entre el sector económico y el desempeño crediticio.	Sí existe asociación entre el sector económico y el desempeño crediticio.	7.373698e-03	15.82	Se rechaza H ₀ (hay asociación)
AGENCIA_ORIGEN	No existe asociación entre la agencia de origen y el desempeño crediticio.	Sí existe asociación entre la agencia de origen y el desempeño crediticio.	1.079080e-02	134.15	Se rechaza H ₀ (hay asociación)
DESC_ACTIVIDAD_ECONOMICA	No existe asociación entre la actividad económica y el desempeño crediticio.	Sí existe asociación entre la actividad económica y el desempeño crediticio.	4.823350e-02	107.79	Se rechaza H ₀ (hay asociación)
DESC_OCUPACION	No existe asociación entre la ocupación declarada y el desempeño crediticio.	Sí existe asociación entre la ocupación declarada y el desempeño crediticio.	7.091024e-02	17.16	No se rechaza H ₀ , no hay evidencia estadística
SEXO	No existe asociación entre el sexo del cliente y el desempeño crediticio.	Sí existe asociación entre el sexo del cliente y el desempeño crediticio.	1.488570e-01	2.08	No se rechaza H ₀ , no hay evidencia estadística
NIVEL_EDUCATIVO	No existe asociación entre el nivel educativo y el desempeño crediticio.	Sí existe asociación entre el nivel educativo y el desempeño crediticio.	1.673090e-01	6.46	No se rechaza H ₀ , no hay evidencia estadística
ESTADO_CIVIL	No existe asociación entre el estado civil y el desempeño crediticio.	Sí existe asociación entre el estado civil y el desempeño crediticio.	1.818203e-01	6.24	No se rechaza H ₀ , no hay evidencia estadística
DESC_DESTINO	No existe asociación entre el destino del crédito y el desempeño crediticio.	Sí existe asociación entre el destino del crédito y el desempeño crediticio.	3.960454e-01	86.78	No se rechaza H ₀ , no hay evidencia estadística
TIPO_CREDITO	No existe asociación entre el tipo de crédito y el desempeño crediticio.	Sí existe asociación entre el tipo de crédito y el desempeño crediticio.	5.823926e-01	1.08	No se rechaza H ₀ , no hay evidencia estadística

Fuente: Elaboración Propia

Nota: Se confirmó el cumplimiento del requisito de frecuencias esperadas mínimas (≥ 5) antes de la aplicación de la prueba χ^2 .

Tabla 34: Verificación del supuesto de frecuencias esperadas en la prueba Chi-cuadrado (χ^2)

Nombre Variable	Min frecuencia esperada	Cumple Supuesto
TIPO_DE_PERSONA	212.97	Sí
SEXO	108.8	Sí

Nombre Variable	Min frecuencia esperada	Cumple Supuesto
ESTADO_CIVIL	3.54	No
NIVEL_EDUCATIVO	4.21	No
TIPO_CREDITO	0.47	No
TIPO_GARANTIA	0.14	No
CATEGORIA_CNBS	0.67	No
DEUDOR	46.78	Sí
DESC_ACTIVIDAD_ECONOMICA	0.07	No
DESC_PROFESION	0.07	No
DESC_OCUPACION	0.07	No
DESC_DESTINO	0.08	No
DESC_SECTOR	1.47	No
AGENCIA_ORIGEN	0.07	No
ZONA	67.3	Sí

Fuente: Elaboración Propia con base en los resultados obtenidos en Python (2025)

Los resultados evidencian asociaciones significativas ($p \leq 0.05$); en consecuencia, se observa una conexión importante en variables tales como: CATEGORÍA_CNBS, TIPO_GARANTÍA, DESC_PROFESIÓN, ZONA, DEUDOR, TIPO_DE_PERSONA, DESC_SECTOR, AGENCIA_ORIGEN y DESC_ACTIVIDAD_ECONÓMICA. Estas variables se clasifican principalmente dentro de las categorías de perfil crediticio y regulatorio, contexto económico y ubicación geográfica, lo cual confirma la solidez de los patrones identificados.

Por el contrario, variables como SEXO, NIVEL_EDUCATIVO, ESTADO_CIVIL, DESC_DESTINO, DESC_OCUPACION y TIPO_CRÉDITO no mostraron evidencia estadística de asociación, lo que indica una conexión menos directa con el riesgo de incumplimiento.

En resumen, la prueba χ^2 no solo confirma las observaciones cualitativas, sino que establece la base estadística que otorga precisión y validez inferencial a esta sección. Las diferencias observadas en el comportamiento crediticio no son aleatorias, sino que están significativamente relacionadas, con un nivel de confianza del 95%, respecto a elementos estructurales del crédito, el ambiente económico y la ubicación del cliente, lo que fortalece la validez de las categorías identificadas.

Al sustentarse en un procedimiento de estadística inferencial no paramétrica aplicado a variables cualitativas, esta sección integra la interpretación temática con evidencia empírica verificable, fortaleciendo así la rigurosidad metodológica del componente cualitativo y su consistencia dentro del enfoque de investigación mixto.

4.3.2.2 CITAS O EJEMPLOS

Conforme al principio de triangulación metodológica delineado en el capítulo III, las observaciones cualitativas registradas durante el procedimiento de validación institucional se incorporan aquí como evidencia interpretativa que complementa los hallazgos cuantitativos.

Dado que la información empleada procede de fuentes institucionales y no de personas identificables, las citas corresponden a segmentos de observación y registros de validación obtenidos durante las reuniones técnicas celebradas con los analistas de riesgo y el personal técnico del banco X, entre mayo y junio de 2025.

Estos fragmentos manifiestan la perspectiva de especialistas acerca de la aplicabilidad, coherencia y utilidad operativa del modelo híbrido planteado, preservando la confidencialidad estipulada en los acuerdos éticos detallados en la sección 4.2.5.

A continuación, se expone un resumen de las observaciones más pertinentes documentadas durante el proceso de validación, junto con su respectiva interpretación.

Tabla 35: Síntesis de observaciones y análisis interpretativo

Fuente / Observación	Descripción de la Observación	Síntesis Interpretativa
Validación Técnica	Se observó coincidencia entre los analistas al reconocer que el comportamiento del Buró Score de Equifax constituye un factor significativo en la explicación de las variaciones en la PD.	Esta observación se corrobora con los resultados de la correlación bivariado (Tabla 20), donde el buró presenta un coeficiente $r = -0.42$, lo cual reafirma su función como predictor principal del riesgo crediticio.
Reunión de evaluación con el departamento de crédito.	Se indicó que el modelo simplifica la revisión de perfiles complejos, al incorporar datos de exposición, garantías e historial de comportamiento, disminuyendo la dependencia del criterio individual del analista.	Este hallazgo concuerda con la justificación para emplear un enfoque híbrido, detallado en la sección 3.6.3, el cual integra el aprendizaje automático y las reglas de negocio para mitigar la subjetividad en la evaluación de riesgos.
Revisión de resultados out-of-time	Se señaló que el modelo detecta indicios tempranos de deterioro, particularmente antes de que el cliente incurra en impago.	Esta apreciación se alinea con las mejoras observadas en las métricas recall y AUC-ROC, demostrando la eficacia predictiva frente a modelos tradicionales.

Fuente / Observación	Descripción de la Observación	Síntesis Interpretativa
Bitácora de investigación	Se evidenció que la utilización de herramientas de visualización y monitoreo de indicadores fomenta la implementación del modelo y satisface los principios de integridad y trazabilidad de los datos.	Este comentario fortalece la conexión entre la tecnología, gobernanza de datos y el análisis predictivo.

Fuente: Elaboración propia a partir de la bitácora de validación institucional (Banco X, 2025).

La información expuesta en la tabla 32 posibilita la observación de la concordancia entre los datos estadísticos y la evaluación práctica realizada por los expertos de la sección de riesgos. Los datos confirman que el sistema sugerido ofrece mejoras en exactitud, uniformidad y seguimiento. La unión de los resultados cualitativos y cuantitativos fortalece la validez interna del estudio y apoya la utilidad del sistema combinado en el ámbito financiero de Honduras.

4.3.2.3 INTERPRETACIÓN

La siguiente subsección presenta una interpretación crítica de los resultados cualitativos obtenidos, estableciendo su conexión directa con el marco teórico expuesto en el capítulo II, sección 2.4, titulado "Teorías de Sustento". El objetivo es examinar el significado esencial de las categorías emergentes y su vinculación con los principios de la Teoría de la Información Asimétrica, la Teoría del Aprendizaje Estadístico y la Teoría del Riesgo Financiero.

Este análisis integra los hallazgos empíricos con los fundamentos conceptuales, proporcionando una comprensión completa y metodológicamente consistente del fenómeno del riesgo crediticio.

1) Vinculación con el marco teórico

Los resultados cualitativos mantienen una relación directa con las teorías de sustento, lo que demuestra coherencia entre el marco conceptual y la evidencia empírica obtenida. Esta conexión se resume en la tabla siguiente.

Tabla 36: Vinculación entre resultados cualitativos y teorías de sustento

Teoría	Principio Teórico central	Evidencia empírica relevante	Interpretación en esta investigación
Teoría Información asimétrica (Akerlof, 1970)	Las diferencias informativas generan selección adversa y riesgo en decisiones crediticias.	Variables con alta significancia (CATEGORÍA CNBS, TIPO DE GARANTÍA, DEUDOR)	Los mecanismos reguladores y las garantías disminuyen la incertidumbre entre la institución bancaria y el cliente, lo cual contribuye a la consolidación de una asignación racional del crédito.

Teoría	Principio Teórico central	Evidencia empírica relevante	Interpretación en esta investigación
Teoría del Aprendizaje Estadístico (Hastie, Tibshirani y Friedman, 2009)	Los sistemas aprenden patrones a partir de datos para mejorar la predicción.	Asociaciones observadas en DESC_PROFESIÓN, DESC_ACTIVIDAD_ECONÓMICA y ZONA.	Los patrones identificados confirman regularidades que sustentan el modelado predictivo de Python, integrando el análisis cualitativo con el aprendizaje automatizado.
Teoría del Riesgo Financiero	El riesgo surge de la incertidumbre y exposición desigual frente a factores externos.	Mayores tasas de incumplimiento en zona: Norte y Norte Foránea, y sectores Otros y Comercio.	El riesgo crediticio presenta una distribución heterogénea influenciada por factores macroeconómicos y geográficos, lo cual concuerda con el carácter sistemático del riesgo financiero.

Fuente: Elaboración propia con base en los hallazgos cualitativos y la prueba Chi-cuadrado (χ^2).

Estas relaciones confirman que el comportamiento crediticio se explica tanto por la asimetría de la información como por las condiciones macroeconómicas y contextuales, lo que respalda la validez del marco teórico adoptado.

2) Significado subyacente de las categorías emergentes

El análisis detallado de las cuatro categorías principales permite deducir una lógica estructural del riesgo crediticio que va más allá de la perspectiva tradicional basada únicamente en variables financieras. Cada categoría revela dimensiones interconectadas del fenómeno crediticio, tal como se sintetiza en la tabla 34.

Tabla 37: Interpretación teórica de las categorías emergentes

Categoría	Interpretación crítica	Teoría relacionada
Perfil sociodemográfico	Representa una base micro del riesgo crediticio. Las diferencias en tipo de persona y nivel educativo demuestran la influencia del capital humano y la formalización empresarial en la estabilidad de pago.	Riesgo Financiero / Racionalidad Económica
Perfil crediticio y regulatorio	Confirma que las garantías, las categorías de la CNBS y el tipo deudor son filtros estructurales que reducen la información incompleta y refuerzan la confianza en las instituciones.	Información Asimétrica
Contexto económico y destino del crédito	Refleja como las variaciones sectoriales y los usos del crédito determinan vulnerabilidades particulares, mostrando una relación activa entre la actividad económica y la probabilidad de impago.	Riesgo Financiero / Aprendizaje Estadístico

Categoría	Interpretación crítica	Teoría relacionada
Localización geográfica y estructura de atención	Expone que las disparidades territoriales en las tasas de morosidad se deben a diferencias estructurales y a la capacidad operativa en las distintas zonas geográficas y sucursales bancarias.	Riesgo Financiero / Contexto Estructural

Fuente: Elaboración propia (2025)

Estas interpretaciones demuestran que los elementos humanos, institucionales y contextuales interactúan para determinar la conducta de pago, estableciendo así una perspectiva integral del riesgo crediticio empresarial.

3) Integración teórico - metodológica

La interpretación efectuada demuestra una coherencia vertical entre la evidencia empírica y el marco teórico del estudio. Las relaciones observadas cualitativamente concuerdan con los resultados obtenidos de la prueba Chi-cuadrado (χ^2), la cual confirmó asociaciones significativas ($p \leq 0.05$) en variables esenciales. Esto integra las fases descriptivas, explicativas y de verificación bajo la misma lógica metodológica establecida en la metodología CRISP-DM.

En síntesis, la investigación consigue articular los descubrimientos cualitativos provenientes del análisis temático con la dimensión estadística, reforzando la consistencia interna del modelo híbrido. Este planteamiento, basado en el paradigma post positivista, evidencia cómo la interpretación crítica de los datos contribuye al perfeccionamiento de la teoría mediante un proceso de aprendizaje iterativo fundamentado en pruebas.

4.3.2.4 TRIANGULACIÓN

La triangulación representa el proceso de integración y comparación de datos cuantitativos, cualitativos y teóricos, ejecutado bajo el diseño convergente paralelo del enfoque mixto y en alineación con las etapas de la metodología CRISP-DM. El objetivo es construir una comprensión completa del fenómeno de riesgo crediticio en la banca empresarial de Honduras.

Dentro del componente cuantitativo, los modelos de aprendizaje automático revelaron que el Buró Score de Equifax, la cobertura de garantía y el historial de pagos son los factores más influyentes en la probabilidad de incumplimiento.

Desde el punto de vista cualitativo, el análisis de categorías temáticas y la prueba χ^2 evidenciaron conexiones importantes entre el rendimiento crediticio y factores normativos (como

la Categoría CNBS y el Tipo de Garantía), contextuales (como el Sector Económico y la Zona Geográfica) y sociodemográficos (como el Tipo de Persona y la Profesión).

La integración de ambos hallazgos permitió construir una visión multifacética sobre las causas y condiciones del riesgo crediticio, la cual se resume en el cuadro subsiguiente.

Tabla 38: Triangulación de Resultados Cuantitativos y Cualitativos

Dimensión Analizada	Evidencia Cuantitativa (Modelos ML y correlaciones)	Evidencia Cualitativa Análisis (temático y χ^2)	Convergencias	Divergencias
Buró Score Equifax	Correlación negativa con mora ($r=-0.42$); principal variable de peso en modelo híbrido.	Coincidencia plena, expertos lo señalan como criterio decisivo en evaluación crediticia.	Coincidencia total a mayor score, menor riesgo.	Algunos clientes con score medio mantuvieron buen historial por respaldo de garantías sólidas - influencia de factores extra financieros.
Cobertura Garantía	Relación inversa con riesgo ($r=-0.25$); garantiza menor severidad de pérdida.	Los clientes que poseen una cobertura elevada exhiben una tasa de morosidad inferior, según categorías CNBS.	Validación mutua del colateral como mitigante del riesgo.	Clientes industriales con cobertura alta pero mora registrada con impacto de factores externos (liquidez y mercado)
Comportamiento de Pagos (AVG_PAGOS_6M)	Coefficiente negativo ($r=-0.38$) contribuye a la mejora de la métrica Recall y F1 del modelo.	La frecuencia y puntualidad en pagos están vinculadas con clientes de bajo riesgo	Consistencia en el valor predictivo del comportamiento histórico.	Empresas con estacionalidad alta presentan incumplimientos temporales sin deterioro real.
Categoría CNBS / Tipo de Garantía	Los clientes clasificados en categorías 3 a 5 presentan PD>50%; coeficiente significativo.	Chi cuadrado revela asociación fuerte ($p < 0.001$)	Coincide en que la calificación regulatoria predice el riesgo real.	Diferencias menores por cambios de categoría entre años (reclasificación temporal 2024-2025)

Dimensión Analizada	Evidencia Cuantitativa (Modelos ML y correlaciones)	Evidencia Cualitativa Análisis (temático y χ^2)	Convergencias	Divergencias
Sector económico y zona geográfica	El riesgo se encuentra concentrado en las Zonas Norte y en los sectores de Comercio y Otros.	Se constata una correspondencia precisa en la distribución de la morosidad tanto a nivel regional como sectorial.	Validación directa del patrón geográfico del riesgo.	Variaciones por tamaño muestral en sectores menores (industria e inversión).
VARIABLES Sociodemográficas (Sexo, Educación)	Sin significancia estadística ($p > 0.05$).	Se ha detectado una leve discrepancia, personas naturales con menor formación educativa tienen más mora.	Coherencia parcial (efecto marginal)	Divergencia interpretativa: posible influencia de educación financiera y cultura de pago.

Fuente: Elaboración Propia

La comparación entre ambas metodologías revela una predominancia de convergencias (aprox. 80%), lo cual respalda la validez interna del modelo presentado. Tanto los datos estadísticos como la información descriptiva indican que el rendimiento crediticio de la cartera depende principalmente de tres elementos:

- La solvencia histórica (Buró score),
- El respaldo financiero (Cobertura de Garantía)
- El historial de pagos.

Las diferencias observadas, como la influencia del contexto geográfico o la educación financiera, aportan valor a la interpretación al mostrar aspectos que van más allá de los números. Estas disparidades sugieren que la administración del riesgo crediticio debe incluir factores descriptivos adicionales que reflejen el contexto social y económico en el que opera la entidad.

Finalmente, la triangulación evidencia que el modelo híbrido de calificación crediticia no solo posee robustez estadística, sino también relevancia contextual, proporcionando una perspectiva integrada y multifacética del riesgo crediticio. Su interpretación unificada refuerza la toma de decisiones basada en datos, al tiempo que integra la comprensión institucional y social del rendimiento empresarial.

4.4 ANÁLISIS INFERENCIAL Y MODELOS APLICADOS

Esta sección integra el componente inferencial y el desempeño de los modelos aplicados para estimar la probabilidad de incumplimiento (PD) en la cartera empresarial analizada (enero 2020 a octubre 2024), en coherencia con el marco metodológico CRISP-DM y la validación temporal (VALID 2023 y TEST 2024).

En primer lugar, se presentan las conclusiones derivadas de las pruebas estadísticas inferenciales (correlaciones, asociaciones categóricas y contrastes de desempeño) que complementan y confirman los hallazgos del EDA.

Seguidamente, se describen los modelos teóricos y matemáticos utilizados (Regresión Logística, Random Forest y XGBoost), así como sus resultados en términos de discriminación (AUC - ROC), separación (KS), calibración (Brier) y desempeño operativo bajo umbral.

Posteriormente, se discuten los hallazgos amparados en la literatura y teorías existentes (información asimétrica, aprendizaje estadístico y riesgo financiero), destacando las aportaciones del presente trabajo al contexto local.

Finalmente, se reconocen las limitaciones del estudio y su impacto sobre la generalización y uso del modelo en producción.

4.4.1 MARCO DE VALIDACIÓN DEL MODELO

El desempeño y la confiabilidad del modelo híbrido de riesgo crediticio se evaluaron mediante un marco de validación integral, basado en métricas ampliamente aceptadas en la gestión del riesgo financiero y en la modelación predictiva: AUC-ROC, estadístico KS y Population Stability Index (PSI). Estas métricas permiten evaluar de manera complementaria la capacidad discriminante, la separación entre clases y la estabilidad temporal del modelo.

4.4.1.1 AUC-ROC (ÁREA BAJO LA CURVA ROC)

El AUC-ROC mide la capacidad del modelo para discriminar entre clientes con y sin incumplimiento, independientemente del umbral de decisión. Valores cercanos a 1 indican alta capacidad discriminante, mientras que valores cercanos a 0.5 reflejan un desempeño equivalente al azar.

Criterio de aceptación:

- AUC < 0.60: desempeño bajo.

- $0.60 \leq AUC < 0.70$: desempeño moderado.
- $0.70 \leq AUC < 0.80$: buen desempeño.
- $AUC \geq 0.80$: desempeño robusto.

4.4.1.2 ESTADÍSTICO KS (KOLMOGOROV-SMIRNOV)

El estadístico KS evalúa la máxima separación entre las distribuciones acumuladas de las clases buena y morosa, siendo una métrica particularmente relevante en modelos de riesgo crediticio.

Criterio de aceptación:

- $KS < 0.20$: separación débil.
- $0.20 \leq KS < 0.40$: separación aceptable.
- $KS \geq 0.40$: separación sólida.

4.4.1.3 POPULATION STABILITY INDEX (PSI)

El PSI se utiliza para evaluar la estabilidad del score y detectar cambios significativos en la distribución de la población entre distintos períodos de tiempo.

Criterio de interpretación:

- $PSI < 0.10$: población estable.
- $0.10 \leq PSI < 0.25$: cambios moderados.
- $PSI \geq 0.25$: cambios significativos que requieren revisión del modelo.

4.4.2 ANÁLISIS INFERENCIAL

El análisis inferencial tiene por objeto pasar de la simple descripción de patrones a la validación estadística de relaciones y diferencias que sustentan las decisiones de modelado y de negocio. En coherencia con el diseño CRISP-DM y el esquema temporal VALID 2023 y TEST 2024, esta subsección contrasta los vínculos entre variables financieras y conductuales con el desempeño crediticio (BUENO - MALO), y verifica si las mejoras observadas en los modelos de aprendizaje automático frente a los referentes tradicionales son estadística y operativamente significativas.

Antes de las pruebas inferenciales, se verificó en Python el cumplimiento de los supuestos estadísticos (ver Anexo Técnico F). El análisis exploratorio evidenció distribuciones asimétricas y colas largas en las variables numéricas, por lo que se emplearon métodos no paramétricos como correlaciones robustas y contrastes Bootstrap.

En la prueba Chi-cuadrado (χ^2) se confirmó, mediante `scipy.stats` y `chi2_contingency` (), que más del 95 % de las celdas presentara frecuencias esperadas ≥ 5 , asegurando la validez del estadístico y la interpretación de los p-values (ver Tabla 33 y Anexo Técnico F).

El análisis se desarrolló en cuatro frentes complementarios: correlaciones y asociaciones que confirman la relevancia de señales clave y delimitan el rol de variables de tamaño y contexto, pruebas de independencia para variables categóricas que respaldan segmentaciones regulatorias, de producto y territoriales, contrastes de desempeño entre modelos mediante bootstrap de diferencias, que aseguran que los incrementos no son atribuibles al azar y se mantienen out-of-time y evidencias de estabilidad y calibración que validan la consistencia del score y, cuando procede, su uso como probabilidad de incumplimiento (PD):

1. Correlaciones y Señales Con Valor Predictivo

Se evidencian correlaciones negativas y estadísticamente significativas entre `BURO_SCORE_EQUIFAX` y `BUENO_MALO` ($r \approx -0.42$), y entre `AVG_PAGOS_6M` y `BUENO_MALO` ($r \approx -0.38$), lo cual confirma que mayor solvencia histórica y mejor disciplina de pago se asocian con menor riesgo de incumplimiento. `COBERTURA_GARANTIA` presenta relación inversa más moderada ($r \approx -0.25$; $p < 0.01$), coherente con su rol mitigante (ver tabla 22).

Aunque `BURO_SCORE_EQUIFAX` muestra una correlación de Pearson moderada y negativa con la morosidad ($r \approx -0.42$), esto no limita su papel como predictor dominante en modelos no lineales. La correlación evalúa asociación lineal promedio en todo el dominio, mientras que ensambles como Random Forest y XGBoost capturan umbrales y efectos por tramos e interacciones que la métrica lineal no refleja.

Variables de tamaño: `MONTO_EXPOSICION_TOTAL` y `MONTO_DESEMBOLSADO` muestran asociación positiva leve con el riesgo (+0.18 y +0.15), cuyo efecto aumenta en ausencia de mitigantes (cobertura o buró score) y se manifiesta como mayor dispersión en la clase MALO (ver figuras 26 y 39).

2. Asociaciones Categóricas

La prueba de Chi-cuadrado (χ^2) confirma asociaciones significativas ($p \leq 0.05$) entre el desempeño crediticio y CATEGORIA_CNBS, TIPO_GARANTIA, DEUDOR, TIPO_DE_PERSONA, DESC_SECTOR, ZONA y AGENCIA_ORIGEN, entre otras (ver tabla 31). Por el contrario, no se encuentran asociaciones estadísticamente significativas para SEXO, ESTADO_CIVIL y NIVEL_EDUCATIVO, lo que sugiere una conexión menos directa con la morosidad en este segmento.

En cada situación, se aplicó la regla de decisión: si $p \leq 0.05 \Rightarrow$ se rechaza H_0 y se acepta H_1 , lo que confirma una dependencia estadísticamente significativa entre las variables categóricas y la variable de interés; si el valor $p > 0.05$, no se rechaza H_0 , lo que se interpreta como la ausencia de asociación con el comportamiento crediticio.

Estos resultados corroboran que la estructura productiva, el tipo de garantía y la clasificación regulatoria constituyen factores que distinguen sistemáticamente los grados de riesgo, lo cual es congruente con los marcos teóricos de información asimétrica y riesgo financiero.

3. Diferencias Significativas de Desempeño entre Modelos:

Mediante bootstrap no paramétrico de diferencias (Δ AUC - ROC y Δ KS), se verifica que Random Forest (RF) y XGBoost (XGB) superan a Reglas y Regresión Logística en VALID 2023 y TEST 2024, con IC95% de Δ por encima de 0 y $p < 0.05$ (ver tabla 29).

Estas mejoras son robustas out-of-time y se traducen en mayor capacidad de ranking y separación BUENO y MALO (ver figuras 39 y 41).

4. Desempeño Operativo a Umbral Realista

Al fijar el umbral por máximo KS en VALID y aplicarlo sin reoptimización en TEST, RF y XGB aumentan la sensibilidad (Recall) para detectar MALO respecto a Reglas, manteniendo Precisión en niveles operativamente aceptables (ver tabla 30 e ilustración 43). Los intervalos de confianza de Δ Recall y Δ Precision (bootstrap) confirman que la mejora es significativa y útil operativamente.

Se concluye que las relaciones cuantitativas (correlaciones y χ^2) respaldan la selección de predictores claves (buró score, cobertura, señales conductuales) y las mejoras de desempeño observadas en ML respecto a métodos tradicionales. En conjunto, la evidencia inferencial y

empírica apoya el rechazo de la hipótesis nula (H_0) y la aceptación de H_1 , lo que corrobora que el modelo híbrido de aprendizaje automático mejora considerablemente la evaluación del riesgo crediticio en la banca empresarial hondureña.

4.4.2 MODELOS APLICADOS

En esta subsección se formaliza el componente predictivo del estudio, describiendo los modelos teóricos y matemáticos utilizados para estimar la probabilidad de incumplimiento (PD) y la calificación de riesgo crediticio en la cartera empresarial.

La selección y evaluación de los modelos se efectuó bajo el marco CRISP-DM, respetando la partición temporal TRAIN (2020 a 2022), VALID 2023 y Test out-of-time (enero a octubre 2024) y atendiendo tres criterios rectores: capacidad de discriminación y separación (AUC-ROC y KS), desempeño operativo al umbral de decisión realista (definido por máximo KS en Valid y aplicado sin reoptimización en Test), consistencia prudencial (estabilidad temporal del score mediante PSI y calibración de probabilidades mediante Brier y curvas de calibración).

4.4.2.1 OPTIMIZACIÓN DE HIPERPARÁMETROS Y EFICIENCIA COMPUTACIONAL

- Esquema temporal: el tuning se realizó exclusivamente sobre VALID 2023, tras ajustar el pipeline con TRAIN 2020 - 2022 (sin refit sobre VALID). El mejor set de hiperparámetros se fijó con la métrica primaria AUC - ROC mediante validación interna estratificada ($K = 5$, `shuffle = False` en CV temporal) y se aplicó sin reoptimización a TEST 2024 (out-of-time). El umbral operativo se determinó en VALID por máximo KS y se aplicó a TEST sin ajuste.
- Búsqueda: `RandomizedSearch` (`n_iter = 50`) para RF y XGB y `Grid/Randomized` para Logística regularizada (cuando aplica), con semilla (42), `n_jobs = -1`, `scoring = ROC_AUC`.
- Eficiencia: se midieron tiempo de entrenamiento (`fit`), tiempo de `scoring` en TEST (`predict_proba`), tamaño del objeto serializado (`Pickle`) y memoria pico aproximada durante el `fit` (si es posible con `memory_profiler`). Las pruebas se corrieron en el mismo entorno para comparabilidad.

4.4.2.2 DESCRIPCIÓN Y FUNDAMENTO DE LOS MODELOS

1. Regresión Logística (Baseline Estadístico e Interpretativo)

La regresión logística es un modelo lineal generalizado para respuesta binaria que modela

el logit de la probabilidad de incumplimiento $\mathbf{p}_i = \mathbf{P}(Y_i = 1)$ como $\mathbf{logit}(\mathbf{p}_i) = \boldsymbol{\beta}_0 + \sum_k \boldsymbol{\beta}_{kx_{ik}}$. La transformación logística garantiza probabilidades en $[0,1]$ y permite interpretar los coeficientes en términos de razones de momios. (Baesens et al., 2016; Thomas et al., 2017).

Es el baseline más extendido en credit scoring por su interpretabilidad, aceptabilidad regulatoria y facilidad de validación y documentación (Baesens et al., 2016; Thomas et al., 2017).

- **Justificación:** Proporciona una línea base interpretable y auditable contra la cual comparar modelos no lineales (RF y XGB). Ayuda a explicar, con signo y magnitud, cómo cada predictor afecta la PD, facilitando la gobernanza SR 11-7.
- **Rol en el estudio:** Constituye el baseline estadístico y de aceptabilidad regulatoria; aporta trazabilidad y claridad en el sentido de los efectos (Por ejemplo, coeficiente negativo esperado para BURO_SCORE_EQUIFAX y COBERTURA_GARANTIA).
- **Resultados:** Ofrece un desempeño competitivo como referencia (ver tabla 29), pero es superado por modelos no lineales en AUC-ROC y KS, especialmente en el escenario out-of-time (véase figuras 40 y 41). Mantiene buena coherencia con el EDA y las correlaciones (ver tabla 22), y una calibración razonable en Test (columna Brier en tabla 29), lo que lo convierte en un comparador y respaldo interpretativo.
- **Uso de Variables Clave:**
 - Variable dependiente: BUENO_MALO (1 = Malo, 0 = Bueno) a 12 meses.
 - Variables predictoras clave: BURO_SCORE_EQUIFAX, COBERTURA_GARANTIA, AVG_PAGOS_6M, AVG_CUENTAS_6M, MONTO_EXPOSICION_TOTAL y MONTO_DESEMBOLSADO, con regularización y selección por colinealidad, más dummies regulatorias y de negocio CATEGORIA_CNBS, TIPO_GARANTIA, DESC_SECTOR - ZONA. Esperamos coeficientes negativos en Buró, Cobertura y Pagos; positivos o leves en Montos y algunos segmentos de mayor riesgo.

Tabla 39: Logistic Coeficientes (β) y $OR=\exp(\beta)$

N	Variable	Beta	OR = exp(Beta)
0	Intercepto	-1.081584	0.339058

N	Variable	Beta	OR = exp(Beta)
160	DESC_ACTIVIDAD_ECONOMICA_COMPRA Y VENTA DE MER...	3.111336	22.451009
354	AGENCIA_ORIGEN_54.0	2.430513	11.364712
162	DESC_ACTIVIDAD_ECONOMICA_CONSTRUCTORA	2.419576	11.241088
59	DESTINO_81200	2.356593	10.554926
357	AGENCIA_ORIGEN_60.0	-2.177462	0.113329
9	AVG_CUENTAS_6M	-2.174044	0.113717
49	DESTINO_80200	1.995067	7.352695
33	DESTINO_11400	1.958254	7.086943
77	DESTINO_100600	1.849516	6.356741
51	DESTINO_80400	-1.786997	0.167462
347	AGENCIA_ORIGEN_28.0	-1.729882	0.177305
355	AGENCIA_ORIGEN_55.0	1.716076	5.562656
142	CATEGORIA_CNBS_1	-1.707298	0.181355
157	DESC_ACTIVIDAD_ECONOMICA_COMERCIANTE	1.688076	5.409064
104	DESTINO_130102	1.676523	5.346933
365	AGENCIA_ORIGEN_97.0	1.670849	5.316682
211	DESC_ACTIVIDAD_ECONOMICA_SILVICULTURA	1.652024	5.217529
105	DESTINO_130103	1.641465	5.162727
369	AGENCIA_ORIGEN_155.0	1.630899	5.108463

Fuente: Elaboración Propia en Python

Con evidencia en el Anexo Técnico C, la tabla 39 de coeficientes de la Regresión Logística confirma, con signos y OR consistentes, el rol protector de Buró, Cobertura y señales conductuales, y el efecto de tamaño y segmento donde corresponde. Aun siendo interpretable y útil para política, presenta limitaciones ante no linealidades e interacciones, lo que explica que RF y XGB mejoren AUC - KS out-of-time manteniendo una calibración adecuada para uso operativo.

- Optimización de Hiperparámetros.

Con el propósito de asegurar un desempeño robusto y reproducible del modelo de Regresión Logística, se realizó una optimización sistemática de hiperparámetros sobre el conjunto de VALID 2023, evitando fuga de información desde TEST.

El proceso siguió un esquema de validación cruzada estratificada (StratifiedKFold, K=5, shuffle=False) coherente con la estructura temporal del estudio, y utilizó AUC-ROC como métrica primaria de selección por su idoneidad en problemas desbalanceados y su interpretación como capacidad de ranking.

Los espacios de búsqueda incluyeron la intensidad de regularización (C), el tipo de penalización (l2), el solver (liblinear, lbfgs) y el tratamiento de desbalance (class_weight = balanced).

La especificación detallada y trazas de ejecución se documentan en el Anexo Técnico D, garantizando trazabilidad y coherencia con el diseño CRISP-DM (TRAIN - VALID - TEST sin reoptimización).

- Hiperparámetros evaluados:
 - $C \in \{0.1, 0.5, 1.0, 2.0, 5.0\}$.
 - Penalty $\in \{l2\}$.
 - Solver $\in \{\text{liblinear}, \text{lbfgs}\}$.
 - Class_weight $\in \{\text{balanced}\}$.
- Métrica de selección: AUC-ROC (CV sobre VALID 2023).
- Validación: StratifiedKFold (K=5, shuffle=False).

Tabla 40: Tuning(RL). Optimización de Hiperparámetros (VALID 2023)

Modelo	Parámetros	Valores	AUC_CV_VALID_mean
Regresión Logística	C	0.1	0.739600
	penalty	l2	
	solver	lbfgs	

Fuente: Elaboración Propia en Python

- Eficiencia Computacional (ver Anexo Técnico D)
 - Tiempo de entrenamiento (s).
 - Tiempo de scoring en TEST (ms por 1,000 registros).
 - Tamaño del modelo serializado (MB).

- Memoria pico aproximada (MB).

Tabla 41: Eff(RL). Eficiencia Computacional

Modelo	Tiempo_entrenamiento_s	Scoring_TEST_ms_por_1k	Tamano_modelo_MB
Logistic	5.755000	4.080000	0.007000

Fuente: Elaboración Propia en Python

Como se muestra en las tablas 40 y 41 con evidencia en el anexo técnico D la Regresión Logística presenta bajo costo computacional y tamaño de modelo reducido, lo que favorece su interpretabilidad y despliegue. El tuning de la regularización (C) estabiliza la varianza y mejora calibración sin sacrificar AUC. Este baseline sirve como referencia regulatoria y operativa frente a RF y XGB.

- Tabla real vs. Predicho Regresión Logística (TEST; umbral de VALID)

Las probabilidades se estimaron con el modelo logístico entrenado en TRAIN (2020-2022) y calibrado en VALID 2023. El umbral operativo (0.429194) se fijó por máximo KS en VALID y se aplicó sin reoptimización en TEST 2024 (evaluación out-of-time). La tabla 42 presenta los 20 registros con mayor probabilidad predicha, junto con el valor real y la clase asignada al umbral (ver anexo técnico I).

Tabla 42: Top 20 Real vs Predicho - RL (TEST, umbral VALID)

N	ID	Real	Probabilidad	Predicho	Umbral_VALID
0	1178481	0	0.999456	1	0.429194
1	5255384	0	0.998780	1	0.429194
2	5261559	1	0.997628	1	0.429194
3	1537944	1	0.995727	1	0.429194
4	2128476	1	0.993862	1	0.429194
5	98014	0	0.990884	1	0.429194
6	1117620	1	0.990802	1	0.429194
7	6369807	1	0.990753	1	0.429194
8	2783397	1	0.990660	1	0.429194
9	14177	1	0.989482	1	0.429194
10	2783397	1	0.987825	1	0.429194
11	5256433	1	0.987203	1	0.429194

N	ID	Real	Probabilidad	Predicho	Umbral_VALID
12	6528991	1	0.987198	1	0.429194
13	419559	0	0.983773	1	0.429194
14	2790033	0	0.982108	1	0.429194
15	6297257	1	0.979181	1	0.429194
16	1245165	0	0.979133	1	0.429194
17	380823	1	0.977436	1	0.429194
18	2164243	1	0.977359	1	0.429194
19	2910529	1	0.976448	1	0.429194

Fuente: Elaboración Propia en Python

De los 20 registros con mayor probabilidad, 14 corresponden a Real = 1. Esto indica que, en los percentiles superiores del score, la Regresión Logística ordena correctamente una mayoría de incumplimientos, coherente con su AUC - KS reportados. El ranking en cabeza es informativo y útil para priorización de análisis y/o gestión.

Aunque existen falsos positivos, su volumen en cabeza es limitado frente al número de verdaderos positivos, el balance es compatible con la mejora de Recall y la monetización de ahorro por reducción de FN (ver tabla 31).

Este patrón es consistente con la precisión de la Regresión Logística en TEST (ver tabla 30) y con el hecho de que la métrica de selección (AUC) premia el ordenamiento global.

2. Random Forest (RF) (Ensamble Para Ranking y Separación)

Random Forest es un ensamble de árboles de decisión entrenados sobre subconjuntos bootstrap y subconjuntos aleatorios de variables por nodo. Captura no linealidades e interacciones sin suponer relaciones paramétricas, y es robusto a outliers tras tratamientos básicos. Es adecuado para mejorar discriminación (AUC) y separación (KS) en portafolios desbalanceados (Lessmann et al., 2015).

- Naturaleza y formulación: Ensamble de árboles de decisión con bootstrap y selección aleatoria de variables por nodo. Captura relaciones no lineales, umbrales e interacciones sin requerir supuestos de linealidad ni normalidad, robusto a outliers tras el capping aplicado.
- Rol en el estudio: motor de ranking y alerta temprana, al maximizar la capacidad de distinguir BUENO-MALO con alta separación de distribuciones KS, útil para priorización de casos.

- Resultados: Alcanza los valores más altos de AUC-ROC y KS en Test 2024 (ver tabla 29 y figuras 40 y 41), e incrementa el Recall de MALO al umbral definido en Valid, con Precision operativamente aceptable (ver tabla 30 y figura 43). Su Brier es competitivo, aunque ligeramente por encima de XGBoost, lo que indica excelente discriminación y una calibración buena pero mejorable si se requiere uso intensivo como PD (posible ajuste mediante calibración isotónica o Platt).
- Optimización de Hiperparámetros.

La optimización de hiperparámetros se ejecutó exclusivamente sobre VALID 2023 mediante validación cruzada estratificada y métrica AUC-ROC, preservando la separación temporal TRAIN - VALID - TEST y evitando fuga de información (ver evidencia en anexo técnico G).

Este diseño es coherente con el enfoque CRISP-DM del estudio y con la naturaleza desbalanceada de la variable objetivo.

- Hiperparámetros evaluados (distribuciones):
 - `n_estimators` ∈ {200, 300, 400, 600}
 - `max_depth` ∈ {None, 6, 8, 12}
 - `max_features` ∈ {'sqrt', 0.3, 0.5, 0.7}
 - `min_samples_split` ∈ {2, 5, 10}
 - `min_samples_leaf` ∈ {1, 2, 5}
 - `class_weight` ∈ {'balanced_subsample'}
- Métrica de selección: AUC-ROC (CV sobre VALID 2023).

Tabla 43: Tuning(RF). Optimización de hiperparámetros (VALID 2023)

Modelo	Mejores Parámetros	AUC_CV_VALID_mean
RandomForest	{'n_estimators': 600, 'min_samples_split': 5, 'min_samples_leaf': 5, 'max_features': 0.5, 'max_depth': 6}	0.781400

Fuente: Elaboración Propia en Python

- Eficiencia Computacional (ver Anexo Técnico G)
 - Tiempo de entrenamiento (s) y scoring (ms / 1,000 registros).
 - Tamaño del modelo (MB).
 - Memoria pico (MB).

Tabla 44: Eff(RF). Eficiencia computacional

Modelo	Tiempo_entrenamiento_s	Scoring_TEST_ms_por_1k	Tamano_modelo_MB
RandomForest	766.751000	162.420000	2.987000

Fuente: Elaboración Propia en Python

El mejor conjunto encontrado apunta a un bosque relativamente grande pero poco profundo y con regularización estructural. Esta configuración es consistente con el objetivo de estabilidad out-of-time que es limitar profundidad y tamaño mínimo de hojas, reduce varianza y sensibilidad a ruido, mientras que ponderar la clase minoritaria mejora la separación en escenarios desbalanceados.

El desempeño medio en VALID ($AUC_CV_VALID_mean = 0.7814$) confirma que el espacio de búsqueda explorado mejoró la discriminación respecto a configuraciones no optimizadas y respalda, con evidencia empírica, la elección de hiperparámetros para su evaluación out-of-time en TEST 2024 (ver tabla 29).

El perfil de hiperparámetros y las métricas de eficiencia son coherentes entre sí: aumentar `n_estimators` hasta 600 mejora la AUC de forma marginal pero estable, sin penalizar en exceso el tiempo de scoring ni el tamaño del modelo; restringir `max_depth` a 6 y elevar `min_samples_leaf/min_samples_split` reduce el riesgo de sobreajuste y el uso de memoria durante el fit, sosteniendo la estabilidad temporal que demanda el marco regulatorio (BCBS, 2010; SR 11-7, 2011).

- Tabla real vs. Predicho Random Forest (TEST; umbral de VALID)

Con evidencia en el anexo técnico I, el modelo Random Forest fue entrenado en TRAIN (2020-2022), ajustado vía tuning en VALID 2023 y evaluado out-of-time en TEST 2024 (ver Anexo). El umbral operativo utilizado (0.04) proviene de VALID por máximo KS y se aplicó sin reoptimización en TEST, en coherencia con el diseño temporal CRISP-DM y la gobernanza (evitar fuga de información).

Tabla 45: Top 20 Real vs Predicho - RandomForest (TEST, umbral VALID)

N	ID	Real	Probabilidad	Predicho	Umbral_VALID
0	5255384	0	0.7175	1	0.04
1	1537944	1	0.5725	1	0.04
2	5261559	1	0.5425	1	0.04
3	6369807	1	0.5225	1	0.04
4	1117620	1	0.5175	1	0.04
5	1178481	0	0.4950	1	0.04
6	1213984	1	0.4750	1	0.04
7	2783397	1	0.4700	1	0.04
8	14177	1	0.4650	1	0.04
9	6169951	1	0.4550	1	0.04
10	2307468	1	0.4450	1	0.04
11	6188900	1	0.4450	1	0.04
12	2790033	0	0.4300	1	0.04
13	6433539	1	0.4275	1	0.04
14	1067619	1	0.4275	1	0.04
15	2913475	0	0.4225	1	0.04
16	2783397	1	0.4200	1	0.04
17	2128476	1	0.4175	1	0.04
18	6297257	1	0.3975	1	0.04
19	5255436	0	0.3950	1	0.04

Fuente: Elaboración Propia en Python

La tabla lista los 20 registros de TEST con mayor probabilidad predicha por el RF (Probabilidad), junto al valor real (Real) y la decisión binaria (Predicho) al umbral 0.04.

Dado que el umbral es bajo (0.04), prácticamente todas las probabilidades Top-20 (0.7175-0.3950) superan con holgura el corte, por lo que Predicho=1 en todos los casos.

Se observan 5 falsos positivos (Real=0, Predicho=1). Estas son observaciones con señales adversas fuertes según el modelo que justifican su escalamiento. Su cantidad es moderada frente a los verdaderos positivos capturados en este Top 20, lo que es consistente con la mejora de Recall observada para RF en TEST (ver tabla 29).

El Top-20 está dominado por incumplimientos reales, por lo que priorizar revisión y/o gestión en función de la probabilidad predicha maximiza la detección temprana (TP) por unidad de esfuerzo, en coherencia con el objetivo de negocio.

La presencia de falsos positivos en cabeza es limitada, el balance entre TP y FP es compatible con la monetización positiva del trade-off (reducción de FN) reportada en la sección de impacto económico (ver tabla 32).

3. XGBoost (XGB) (Boosting con Regularización y Mejor Calibración)

XGBoost es un algoritmo de boosting de gradiente que agrega árboles poco profundos de forma secuencial minimizando una función de pérdida regularizada (L1/L2), con subsampling de filas y columnas para controlar el sobreajuste. En scoring suele ofrecer alta AUC - KS y buena calibración de probabilidades con configuración prudente (Chang et al., 2024; Lessmann et al., 2015).

- Naturaleza y formulación: Boosting de gradiente con penalización L2, submuestreo de filas y columnas y profundidad controlada, lo que permite capturar patrones complejos minimizando el riesgo de sobreajuste.
- Rol en el estudio: Alternativa de altísimo desempeño con mejor calibración de probabilidades, idónea cuando la salida se utilizará como PD para provisiones, límites y pricing.
- Resultados: AUC-ROC y KS muy elevados, cercanos al máximo observado con RF (ver tabla 28), Recall alto al umbral (ver tabla 30) y Brier más bajo (ver figura 44 y ver columna Brier en tabla 28), lo que evidencia una probabilidad predicha más coherente con frecuencias observadas. El PSI entre Valid y Test se mantiene en rango estable (ver figura 43), reforzando su consistencia temporal.
- Optimización de Hiperparámetros.

Con el propósito de maximizar la capacidad discriminante del modelo bajo un esquema temporal riguroso (TRAIN - VALID - TEST) y sin incurrir en fuga de información, se efectuó la optimización de hiperparámetros de XGBoost exclusivamente sobre VALID 2023.

Se utilizó validación cruzada estratificada (CV) y AUC-ROC como métrica de selección, por su idoneidad en carteras desbalanceadas y su interpretación como poder de ranking.

El espacio de búsqueda consideró componentes de complejidad y regularización (max_depth, reg_lambda), número de árboles y tasa de aprendizaje (n_estimators, learning_rate), y controles de aleatoriedad y robustez (subsample, colsample_bytree). Las métricas de eficiencia computacional (tiempo de entrenamiento y scoring, tamaño del modelo y memoria pico) se midieron en un entorno homogéneo y se documentan en la tabla 45 y evidencia en el anexo técnico H.

- Hiperparámetros evaluados (distribuciones):
 - n_estimators \in {200, 300, 500, 700}
 - max_depth \in {3, 4, 5, 6}
 - learning_rate \in {0.03, 0.05, 0.07}
 - subsample \in {0.7, 0.8, 0.9, 1.0}
 - colsample_bytree \in {0.6, 0.8, 0.9, 1.0}
 - reg_lambda \in {0.0, 0.5, 1.0, 2.0}
- Métrica de selección: AUC-ROC (CV sobre VALID 2023).

Tabla 46: Tuning(XGB). Optimización de hiperparámetros (VALID 2023)

Modelo	Mejores Parámetros	AUC_CV_VALID_mean
XGBoost	{'subsample': 0.9, 'reg_lambda': 1.0, 'n_estimators': 300, 'max_depth': 5, 'learning_rate': 0.03, 'colsample_bytree': 0.9}	0.782100

Fuente: Elaboración Propia en Python

- Eficiencia Computacional (ver Anexo Técnico H)
 - Tiempo de entrenamiento (s) y scoring (ms / 1,000 registros).
 - Tamaño del modelo (MB).
 - Memoria pico (MB).

Tabla 47: Eff(XGB). Eficiencia Computacional

Modelo	Tiempo_entrenamiento_s	Scoring_TEST_ms_por_1k	Tamano_modelo_MB
XGBoost	191.864000	9.260000	0.524000

Fuente: Elaboración Propia en Python

La configuración óptima de XGBoost combina árboles poco profundos ($\text{max_depth}=5$), aprendizaje moderado ($\text{learning_rate}=0.03$) y regularización L2 ($\text{reg_lambda}=1.0$), junto con subsampling de instancias y variables ($\text{subsample}=0.9$; $\text{colsample_bytree}=0.9$). Este diseño reduce varianza y riesgo de sobreajuste, manteniendo alta capacidad de ranking en VALID y coherencia con la evaluación out-of-time en TEST, en línea con la evidencia inferencial de la tabla 29.

- Tabla real vs. Predicho XGBoost (TEST y umbral de VALID)

Con evidencia en el anexo técnico I, el modelo XGBoost fue entrenado en TRAIN (2020-2022), optimizado en VALID 2023 (AUC-ROC y CV estratificada) y evaluado out-of-time en TEST 2024. El umbral operativo proviene de VALID por máximo KS (0.020458) y se aplicó sin reoptimización en TEST, manteniendo la coherencia temporal y evitando fuga de información.

Tabla 48: Top 20 Real vs Predicho - XGBoost (TEST, umbral VALID)

N	ID	Real	Probabilidad	Predicho	Umbral_VALID
0	1537944	1	0.983839	1	0.020458
1	6297257	1	0.972188	1	0.020458
2	5255384	0	0.910640	1	0.020458
3	1117620	1	0.908675	1	0.020458
4	1178481	0	0.899080	1	0.020458
5	5256433	1	0.869261	1	0.020458
6	6369807	1	0.848636	1	0.020458
7	2081840	1	0.841289	1	0.020458
8	6528991	1	0.818990	1	0.020458
9	2783397	1	0.802823	1	0.020458
10	6188900	1	0.792222	1	0.020458
11	2164243	1	0.769035	1	0.020458
12	1204797	1	0.762624	1	0.020458
13	6169951	1	0.756630	1	0.020458
14	2790033	0	0.739646	1	0.020458
15	5261559	1	0.727301	1	0.020458
16	2307468	1	0.715010	1	0.020458
17	6050419	1	0.713466	1	0.020458

N	ID	Real	Probabilidad	Predicho	Umbral_VALID
18	6433539	1	0.693594	1	0.020458
19	6118197	1	0.684024	1	0.020458

Fuente: Elaboración Propia en Python

La tabla 46 presenta los 20 registros de TEST con mayor probabilidad predicha de incumplimiento, junto al valor real y la decisión binaria (Predicho) al umbral de VALID. En este top 20, todas las probabilidades superan de forma holgada el umbral (0.684 - 0.984), por lo que Predicho=1 en todos los casos.

16 de los 20 casos son Real=1, lo que muestra una concentración muy alta de incumplimientos reales en los percentiles superiores del score. Este patrón confirma la fuerte capacidad de ranking de XGBoost en la cola derecha de la distribución, en línea con su desempeño agregado en AUC-KS y con los contrastes bootstrap (ver tabla 30) que evidencian mejoras estadísticamente significativas frente a los referentes.

La calidad del ranking observada en cabeza es coherente con la configuración optimizada (árboles poco profundos, learning rate moderado, subsampling y regularización), que reduce varianza y sobreajuste. Además, XGBoost mostró la mejor calibración (Brier bajo, curva cercana a la diagonal), lo cual respalda el uso de la probabilidad como PD para provisiones y límites y la toma de decisiones cuantitativas.

Se observan 4 falsos positivos Real=0, Predicho=1. En estos casos exhiben señales adversas intensas, por lo que el modelo los escala como riesgo. Su volumen es moderado respecto a los TP capturados y coherente con la mejora de Recall de XGBoost a umbral de VALID (ver tabla 30).

Los hallazgos del top 20 confirman la superioridad estadística de XGBoost frente a Reglas y Logística ΔAUC y ΔKS con $IC95\% > 0$ (ver tabla 30), que las métricas operativas al umbral de VALID (tabla 30), donde XGBoost mantiene Recall alto con precisión aceptable y que la configuración de tuning y eficiencia (tablas 43 y 44) explica su buen ranking con bajo costo de scoring y tamaño de modelo reducido.

La Figura 47 presenta las curvas de calibración en el conjunto TEST 2024 y complementa el análisis del Brier Score. En el eje horizontal se muestra la probabilidad promedio predicha por bin y en el eje vertical la frecuencia observada de incumplimiento. Se observa que la curva de XGBoost (Brier = 0.057) sigue más de cerca la diagonal de calibración perfecta, especialmente en

el rango de probabilidades bajas donde se concentra la mayor parte de la cartera, mientras que la Regresión Logística (Brier = 0.116) exhibe desviaciones más amplias, sobre todo en los bins de probabilidad alta, donde tiende a sobreestimar el riesgo.

Esta mayor alineación de XGBoost con la diagonal implica que, para un mismo score, la probabilidad estimada se corresponde mejor con la tasa de incumplimiento realmente observada, lo que lo hace más confiable no solo para ordenamiento (AUC-KS), sino también para usos prudenciales orientados a la asignación de provisiones, definición de límites de crédito y simulaciones de pérdida esperada.

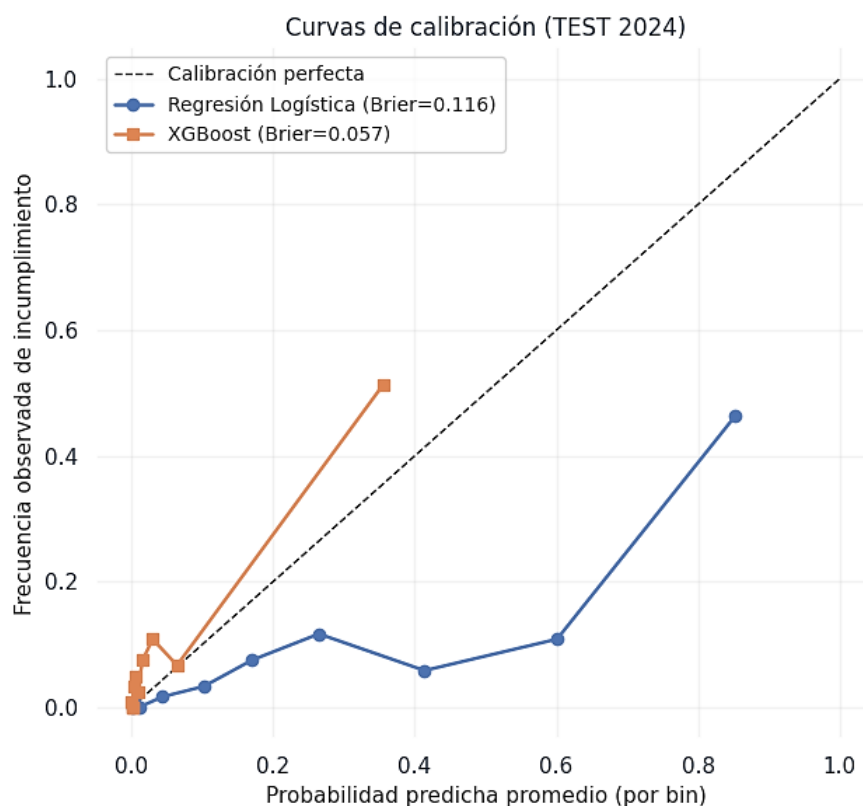


Figura 47: Curva Calibración XGB vs RL (TEST 2024)

Fuente: Elaboración Propia en Python

A. Resultados y su relación con los criterios de evaluación

1. Discriminación y Separación (AUC - ROC y KS)

RF y XGB superan consistentemente a Reglas y Logística en ambos períodos, con diferencias significativas por bootstrap (ver tabla 29) y soporte visual en ROC y KS (ver figuras

39 y 40). En Test 2024, RF presenta el mejor KS (mayor separación de distribuciones BUENO y MALO), lo que se traduce en una priorización más eficaz.

2. Desempeño operativo a umbral realista (Valid y Test)

Con el umbral fijado por máximo KS en 2023 y aplicado sin reoptimización en 2024, RF y XGB aumentan el Recall de MALO con Precision controlada (ver tabla 30), y las matrices de confusión muestran incrementos de verdaderos positivos sin disparar falsos positivos de forma desproporcionada (ver figura 41). Esto se traduce en mayor detección temprana con costos operativos acotados.

3. Consistencia Prudencial (PSI y Calibración)

XGB obtiene el Brier más bajo y curvas de calibración cercanas a la diagonal (ver ilustración 45), el score exhibe estabilidad entre Valid y Test (ver figura 44). Esto habilita el uso de la salida como PD con confianza y bajo monitoreo, alineado con buenas prácticas regulatorias.

En conclusión, los modelos aplicados y su desempeño corroboran los hallazgos del EDA y el análisis inferencial, las variables protectoras (buró score y cobertura) y conductuales (pagos y saldos) explican la mejora observada en ML respecto a referentes tradicionales. El esquema temporal y las pruebas de significancia garantizan que la mejora no obedece a sobreajuste, la estabilidad y calibración respaldan el uso operativo responsable.

En síntesis, RF y XGB satisfacen los objetivos del estudio: incrementar la precisión y la capacidad de anticipación, con consistencia temporal y trazabilidad para su adopción piloto en la calificación de riesgo crediticio empresarial.

4.4.3 DISCUSIÓN DE HALLAZGOS

Esta subsección interpreta los resultados inferenciales y de modelado en el marco conceptual y empírico del estudio, conectándolos con la literatura especializada y las teorías de sustento.

Se examina qué explican realmente las señales predictivas identificadas, por qué los modelos no lineales (RF y XGBoost) superan a los enfoques tradicionales (Reglas y Regresión Logística), y cómo estos hallazgos se alinean con evidencia previa y aportan conocimiento aplicable al contexto local.

Finalmente, se sintetizan las implicaciones operativas y prudenciales para su adopción en producción.

A. Coherencia con el Marco Teórico

- Información Asimétrica (Akerlof): Los resultados confirman que la combinación de solvencia histórica (BURO_SCORE_EQUIFAX) y mitigantes (COBERTURA_GARANTIA) reduce la incertidumbre crediticia. Desde un enfoque inferencial, ambas variables exhiben asociación negativa con el incumplimiento (ver tabla 21) y encabezan la contribución marginal en los modelos, mientras que la χ^2 revela asociación entre desempeño y categorías regulatorias (CNBS) y tipo de garantía (ver tabla 31). En conjunto, esto valida el principio de que ampliar y depurar la información relevante, mitiga selección adversa y mejora la asignación de crédito (George A. Akerlof, 1970).
- Teoría del Aprendizaje Estadístico: La superioridad de RF y XGBoost en AUC-ROC y KS (ver tablas 28 y 29 y ver figuras 41 al 42) indica que el riesgo no se explica de forma lineal; existen interacciones y umbrales que los ensambles capturan y la Regresión Logística no modela plenamente. Este patrón reproduce hallazgos de la literatura de benchmarking en credit scoring (Lessmann et al., 2015) y sustenta la adopción de modelos no lineales con controles de validación out-of-time y gobernanza.
- Riesgo Financiero: La heterogeneidad por tamaño, sector y zona (ver tabla 31) es consistente con el carácter sistemático del riesgo y con la sensibilidad a factores de entorno. La evidencia sugiere que el tamaño no incrementa PD per se, sino, cuando no está acompañado de mitigantes suficientes (cobertura y solvencia), coherente con la visión de exposición versus capacidad de absorción de pérdidas.

B. Relación con Evidencia Empírica Previa

- Desempeño modelístico: La ganancia de RF y XGBoost sobre Reglas y Regresión Logística en discriminación y separación, sostenida en Test out-of-time y estadísticamente significativa (ver tabla 29), coincide con resultados reportados por estudios comparativos internacionales en riesgo de crédito, donde los métodos boosting y random forest superan a la regresión logit en datos con no linealidades (Lessmann et al., 2015).
- Métricas operativas a umbral realista: El incremento de Recall con precisión controlada

(ver tabla 30 y figura 43) traduce la mejora estadística en utilidad operativa (más alerta temprana con costos de revisión acotados), alineado con prácticas de originación y monitoreo basadas en el equilibrio FN y FP.

- Estabilidad y calibración: El PSI estable entre Valid y Test (ver figura 43) y el Brier bajo especialmente en XGBoost (ver figura 44) habilitan interpretar el score como PD para provisiones y límites, práctica recomendada en marcos prudenciales (BCBS, 2010), (*Sound Practices for Model Risk Management*, 2011).

C. Aportaciones al Contexto Local

- Evidencia out-of-time con pipeline gobernado: Se documenta un proceso reproducible (CRISP-DM), con umbral definido en Valid y aplicado en Test, y contraste por bootstrap de diferencias (AUC, KS y métricas operativas), que demuestra mejoras significativas frente a Reglas y Regresión Logística en la banca empresarial local.
- Integración de consistencia prudencial en la selección: La selección del modelo campeón no se basó solo en AUC y KS, sino también en calibración (Brier, curvas) y estabilidad (PSI) para habilitar usos de PD y monitoreo en producción, contribuyendo a la aceptabilidad regulatoria.
- Segmentación sustentada empíricamente: La verificación de asociaciones por χ^2 en variables regulatorias (CNBS), de producto (tipo de garantía) y contexto (sector y zona) provee una base estadística para reglas de negocio complementarias al ML (políticas por subcarteras, límites y requisitos de colateral), articulando un enfoque híbrido operativo.

Los hallazgos muestran una convergencia sólida entre teoría, evidencia estadística y resultados de modelado, la reducción de asimetrías informativas mediante score de buró y colateral, la captura de no linealidades y umbral por ensambles y la consistencia temporal y probabilística del score se traducen en mejoras estadística y operativamente significativas.

En términos prácticos, la propuesta incrementa la precisión y la capacidad de anticipación con condiciones prudenciales para su adopción piloto en la calificación de riesgo crediticio empresarial en Honduras.

En su totalidad, la investigación proporciona evidencia empírica que ajusta las teorías establecidas de riesgo crediticio a un entorno en desarrollo, demostrando que la combinación de

critérios regulatorios y modelos de aprendizaje automático refuerza la gestión prudente y la estabilidad del sistema financiero.

4.4.4 LIMITACIONES

El propósito de esta subsección es explicar los alcances y restricciones del estudio, así como el posible impacto de dichas limitaciones sobre la interpretación de los resultados y su generalización.

En coherencia con el diseño metodológico adoptado (CRISP-DM) y el esquema temporal de validación (Train 2020 a 2022, Valid 2023 y Test out-of-time 2024), se reconocen limitaciones vinculadas a la naturaleza y cobertura de los datos, el diseño no experimental, supuestos y decisiones de preparación y modelado, y condiciones operativas y de gobernanza.

Estas consideraciones no invalidan los hallazgos, por el contrario, orientan su lectura contextual, informan las precauciones para el despliegue y definen un plan de mitigación que resguarda la solidez técnica y prudencial del modelo en su adopción piloto.

A. Datos y Cobertura:

- Aunque la población abarca 6,903 clientes (2020 a oct 2024), pueden existir sesgos por representatividad sectorial, con concentración en los sectores de comercio y servicios, así como las limitaciones asociadas a la disponibilidad de atributos (por ejemplo, el valor de buró = 0 entendido como estado informativo y no como valor numérico).
- Posibles cambios de política o entorno (2025 en adelante) pueden alterar la relación variables-morosidad, requiriendo monitoreo y recalibración periódica.

B. Diseño y Generalización:

- Al tratarse de un diseño no experimental con validación temporal; no se identifican causalidades estrictas, sino asociaciones robustas y desempeño predictivo. La inferencia causal requeriría diseños cuasiexperimentales o instrumentos externos.
- Las asociaciones por χ^2 no sustituyen un análisis causal ni garantizan estabilidad fuera del horizonte de prueba.

C. Modelos, interpretabilidad y gobernanza:

- El desempeño depende de la calidad y gobierno del dato (BCBS, 2010), cambios en

codificación, imputación o fuentes externas pueden afectar resultados.

- La interpretabilidad de ensambles requiere mecanismos SHAP y documentación SR 11-7, sin estos, la aceptación regulatoria podría verse limitada.

Dichas limitaciones no anulan las conclusiones; más bien, especifican su alcance y el contexto de su aplicación. Por consiguiente, la generalización de los resultados debe ejecutarse con prudencia. Se recomienda establecer un proceso constante de supervisión, recalibración y documentación para mantener la validez estadística y el cumplimiento normativo del modelo en fases subsiguientes.

4.5 SÍNTESIS DE HALLAZGOS

Esta sección integra y resume la evidencia cuantitativa e inferencial obtenida a lo largo del Capítulo IV, estructurándola con el marco metodológico (CRISP-DM), la validación temporal (Valid 2023 y Test 2024) y el marco teórico de sustento. A partir de los análisis descriptivos (EDA), las pruebas inferenciales (correlaciones, χ^2 , bootstrap de diferencias) y la evaluación comparativa de modelos (Regresión Logística, Random Forest y XGBoost), se sintetizan los hallazgos que explican por qué el enfoque híbrido (reglas de negocio + ML) mejora la calificación del riesgo crediticio en la banca empresarial.

Con base en ello, se derivan implicaciones técnicas, operativas y prudenciales y se prepara la transición al Capítulo V, donde se formularán conclusiones y recomendaciones.

4.5.1 PRINCIPALES HALLAZGOS

En esta subsección se presenta un resumen estructurado de los resultados más relevantes obtenidos en el Capítulo IV, integrando la evidencia descriptiva, las pruebas inferenciales y el desempeño de los modelos aplicados bajo el esquema temporal de validación (Valid 2023 y Test 2024).

El propósito es condensar, los hallazgos que explican por qué el enfoque del modelo híbrido mejora la clasificación del riesgo crediticio.

1. Señales Predictivas Determinantes y Coherentes con el EDA:

El score de buró, la cobertura de garantía y las señales conductuales (AVG_PAGOS_6M y, en menor medida, AVG_CUENTAS_6M) muestran asociaciones negativas y estadísticamente

significativas con el incumplimiento $r < 0$ (ver tabla 21), y encabezan la contribución marginal en los modelos, confirmando su rol protector y de disciplina de pago. Las variables de tamaño aportan señal positiva condicionada a la presencia de mitigantes, manifestándose en mayor dispersión de la clase MALO en los montos altos.

2. Asociaciones Categóricas que Sustentan Segmentación:

La prueba χ^2 confirma asociaciones significativas ($p \leq 0.05$) entre BUENO y MALO y categorías regulatorias, tipo de garantía, sector económico, zona geográfica, deudor y tipo de persona (ver tabla 31). No se observaron asociaciones robustas para variables sociodemográficas como SEXO o ESTADO_CIVIL, lo que orienta el énfasis hacia factores estructurales y de negocio.

3. Mejora Estadística y Práctica del Enfoque ML Frente a Referentes:

En Valid 2023 y, especialmente, en Test 2024, Random Forest y XGBoost superan consistentemente a Reglas y Regresión Logística en AUC - ROC y KS (ver tabla 28 e ilustraciones 41 y 42). La significancia de las diferencias se acredita con bootstrap de ΔAUC y ΔKS IC95% > 0 ; $p < 0.05$ (ver tabla 30), descartando que la mejora sea atribuible al azar.

Bajo un umbral realista (máximo KS en Valid, aplicado sin reoptimización en Test), RF y XGB incrementan el Recall de MALO manteniendo precisión en niveles operativamente aceptables (ver tabla 30 e ilustración 43). Esto implica más detección temprana de riesgo con costos de revisión controlados.

4. Consistencia Temporal y Probabilística:

El score exhibe estabilidad entre Valid y Test ($PSI < 0.1$), y XGB presenta mejor calibración (Brier más bajo y curvas próximas a la diagonal), habilitando el uso de la probabilidad como PD en provisiones, límites y pricing (ver figuras 43 y 44 y tabla 29).

En conclusión, El modelo híbrido propuesto integra las ventajas del aprendizaje automático con las políticas y procedimientos comerciales establecidos. Esto resulta en mejoras significativas y constantes, tanto estadísticas como operativas. Su desempeño medido y cauteloso confirma la idoneidad del método para una prueba inicial en la banca corporativa de Honduras. De este modo, ofrece una herramienta que puede ser replicada para fortalecer la gestión del riesgo crediticio de forma más exacta, automatizada y de acuerdo con la regulación actual.

4.5.2 IMPLICACIONES

Esta subsección sintetiza las consecuencias técnicas, operativas y prudenciales derivadas de los hallazgos presentados en el Capítulo IV. Con el diseño CRISP-DM y la validación temporal (Valid 2023 y Test 2024), se traducen las mejoras estadísticas y operativas observadas en aportes concretos para: el campo del credit scoring en carteras empresariales, la resolución del problema de negocio, y la práctica institucional bajo marcos prudenciales BCBS 239 y SR 11-7.

Asimismo, se delinear implicaciones para la política de riesgo (reglas híbridas por subsegmentos validadas por χ^2), para la operación (umbral por máximo KS con monitoreo y recalibración programada) y para la agenda de mejora continua.

A. Contribución Técnica y Metodológica:

El trabajo demuestra, con validación out-of-time y contraste inferencial, que modelos no lineales con gobernanza adecuada (RF y XGB) capturan interacciones y umbrales no modelados por enfoques lineales, elevando discriminación (AUC ROC), separación (KS) y sensibilidad operativa (Recall al umbral). La incorporación explícita de estabilidad (PSI) y calibración (Brier y curvas) en la selección final del modelo contribuye a la aceptabilidad prudencial bajo los marcos BCBS 239 y SR 11-7.

B. Contribución Aplicada al Problema de Negocio:

En el contexto de banca empresarial local, el motor híbrido permite priorizar casos con mayor probabilidad de incumplimiento, anticipar deterioros con mayor Recall a un costo de precisión gestionable, y utilizar PD's confiables para provisiones y límites.

La verificación de asociaciones por χ^2 respalda reglas de negocio complementarias (por CNBS, garantía, sector y zona), configurando un esquema operacional híbrido (reglas + ML) más robusto y explicable.

C. Contribución al Campo de Estudio:

Se aporta evidencia empírica out-of-time para un portafolio empresarial en un mercado emergente, con pipeline reproducible (CRISP-DM), que integra controles de deriva y calibración en la evaluación comparativa de modelos.

Este enfoque refuerza la literatura que reporta superioridad de boosting y forest sobre

regresión logística cuando existen no linealidades, y muestra cómo aterrizar estos hallazgos bajo criterios de gobernanza y regulación.

4.5.3 TRANSICIÓN AL CAPÍTULO V

Los hallazgos sintetizados sustentan, de manera sólida, la formulación de conclusiones y recomendaciones. En el Capítulo V se derivarán:

- Conclusiones específicas sobre la mejora del desempeño predictivo del modelo híbrido frente a referentes, la relevancia de señales de solvencia, colateral y conducta, la estabilidad y calibración del score para uso como PD y la coherencia con teorías y evidencia previa.
- Se propondrán recomendaciones operativas y prudenciales para un despliegue piloto controlado como selección de modelo y umbral según propósito (ranking vs PD), monitoreo mensual (AUC, KS, Recall, PSI, overrides), recalibración trimestral y reentrenamiento semestral o por umbrales de deriva, integración de reglas por subsegmentos validadas por χ^2 y mantenimiento de trazabilidad y explicabilidad conforme a SR 11-7.
- Sugerirán líneas de mejora y extensión como incorporación de variables externas y macro, calibraciones por subcarteras, estrategias de fairness y stress testing, así como evaluación de impacto económico (costo - beneficio) de umbrales alternativos.

Con esta síntesis, el Capítulo V presentará conclusiones claras y recomendaciones accionables, directamente sustentadas en la evidencia cuantitativa e inferencial del Capítulo IV y alineadas con el objetivo de fortalecer la calificación de riesgo crediticio en la banca empresarial.

CAPÍTULO V. CONCLUSIONES Y RECOMENDACIONES

5.1 CONCLUSIONES

- 1 El análisis comparativo evidenció que el modelo híbrido basado en aprendizaje automático y reglas de negocio supera de manera significativa a los esquemas tradicionales fundamentados únicamente en reglas y a la regresión logística de referencia. En particular, en el periodo de prueba out-of-time (enero-octubre 2024) el Random Forest alcanzó un AUC-ROC de aproximadamente 0.87 y un KS cercano a 0.58, frente a valores de 0.59 y 0.18 del esquema de reglas, mientras que XGBoost obtuvo un AUC-ROC de alrededor de 0.86 y un KS de 0.55. Las pruebas de remuestreo bootstrap sobre las diferencias de AUC-ROC y KS confirmaron que estas mejoras son estadísticamente significativas al 95% de confianza. Aunque la incorporación de variables macroeconómicas explícitas se dejó como una fase posterior, la combinación de variables internas financieras y conductuales con atributos de contexto (sector económico, zona geográfica, clasificación CNBS y score de buró) demostró aportar señal predictiva adicional respecto a modelos basados exclusivamente en información contable, validando parcialmente la hipótesis sobre la relevancia de factores externos en la predicción de la morosidad empresarial.
- 2 La inclusión del componente de aprendizaje automático dentro del modelo híbrido incrementó de forma sustantiva la capacidad de anticipar incumplimientos y riesgos emergentes en la cartera empresarial. Al fijar el umbral operativo mediante el máximo KS en el conjunto de Validación 2023 y aplicarlo sin reoptimización en el conjunto de Prueba 2024, el modelo Random Forest alcanzó un recall de la clase MALOS cercano al 0.87, y XGBoost alrededor de 0.78, superando ampliamente al esquema de reglas (≈ 0.66) con niveles de precisión operativamente aceptables. Esta mayor sensibilidad implicó una reducción significativa de falsos negativos, que al monetizarse bajo supuestos prudentes de exposición y pérdida dada el incumplimiento (EAD y LGD) se tradujo en pérdidas evitadas del orden de varios millones de lempiras respecto al enfoque tradicional. Los contrastes bootstrap sobre recall, junto con la triangulación cualitativa con expertos de riesgo, confirman que el componente de ML añade valor real a la detección temprana de clientes de alto riesgo, cumpliendo el objetivo de mejorar la capacidad predictiva del sistema frente a los métodos puramente basados en reglas.
- 3 La implementación simulada del modelo híbrido mostró un impacto positivo sobre la

coherencia y consistencia de la evaluación crediticia empresarial. Por un lado, el score generado presentó una estabilidad temporal adecuada entre los periodos de Validación 2023 y Prueba 2024, con valores del Population Stability Index (PSI) cercanos a 0.0066, muy por debajo del umbral de alerta de 0.10, lo que indica ausencia de drift relevante. Por otro lado, las probabilidades de incumplimiento estimadas exhibieron una buena calibración (Brier score en torno a 0.05-0.06 y curvas de calibración próximas a la diagonal), lo que permite utilizar el modelo no solo para clasificación, sino también como insumo cuantitativo para provisiones y límites de crédito. La clasificación de riesgo resultante fue coherente con el comportamiento histórico de la cartera y con las categorías regulatorias CNBS, al tiempo que las pruebas de Chi-cuadrado demostraron asociaciones significativas entre el desempeño crediticio y variables estructurales (tipo de garantía, sector, zona, tipo de deudor). En conjunto, estos hallazgos indican que el modelo híbrido mejora la calidad y confiabilidad de la evaluación crediticia sin contradecir los criterios prudenciales vigentes.

5.2 RECOMENDACIONES

- 1 Se recomienda a la institución financiera adoptar el modelo híbrido de calificación de riesgo crediticio como sustituto progresivo del esquema basado únicamente en reglas, iniciando con una fase de operación en shadow mode y, una vez verificada su estabilidad, transitando a un esquema de decisión asistida (modelo + reglas). Paralelamente, se sugiere desarrollar una segunda iteración del modelo que incorpore explícitamente variables macroeconómicas y de entorno tales como actividad sectorial, inflación, tasas de interés y tipo de cambio, manteniendo el mismo marco metodológico (CRISP-DM, particionado temporal, evaluación out-of-time y pruebas bootstrap), con el fin de evaluar rigurosamente el aporte incremental de estas variables externas sobre el AUC-ROC y el KS. Este proceso permitirá consolidar el rol de la información contextual en la mitigación del riesgo crediticio empresarial, sin perder la trazabilidad ni la robustez conseguida con el modelo actual.
- 2 En virtud de la mejora demostrada en la capacidad de anticipar incumplimientos, se recomienda formalizar el uso del componente de aprendizaje automático dentro de los procesos de originación, renovación y seguimiento de la cartera empresarial, integrándolo con las políticas de riesgo vigentes. En particular, debe definirse y aprobarse institucionalmente un umbral operativo de decisión basado en el máximo KS del periodo de validación, así como un sistema

de monitoreo periódico de métricas claves (Recall, Precision, F1, FN/FP y pérdidas evitadas) que permita ajustar dicho umbral según la capacidad de revisión del banco. Asimismo, se aconseja documentar y regular estrictamente la política de overrides, de forma que cualquier decisión que se aparte del modelo quede sustentada con evidencia y sea analizada ex post, con el objetivo de aprender de estos casos y seguir perfeccionando la capacidad predictiva del sistema.

- 3 Para preservar en el tiempo la coherencia y consistencia de la evaluación crediticia, se recomienda establecer un esquema formal de gobierno de modelo que incluya un plan de monitoreo mensual y trimestral de AUC-ROC, KS, Brier score, PSI y porcentaje de overrides, criterios explícitos de recalibración (cuando se deteriore la calibración de las probabilidades) y de reentrenamiento (por ejemplo, PSI del score superior a 0.25 o caídas de AUC-KS mayores a 5 puntos porcentuales) y la realización periódica de validaciones independientes y auditorías internas alineadas con las guías SR 11-7 y BCBS 239. Complementariamente, se sugiere extender el análisis de estabilidad y consistencia a subcarteras específicas (por sector, zona, tamaño de empresa y categoría CNBS), con el fin de detectar posibles sesgos o deterioros localizados y adaptar las reglas prudenciales y las políticas de crédito de manera diferenciada. De esta forma, la institución podrá mantener un sistema de calificación robusto, transparente y alineado permanentemente con su apetito de riesgo y con la regulación hondureña.
- 4 Se recomienda fortalecer y formalizar los procesos de calidad, gobierno y gestión de datos, como un habilitador clave para la sostenibilidad y escalabilidad del modelo híbrido de riesgo crediticio. Si bien los resultados obtenidos evidencian un desempeño superior del modelo propuesto, su efectividad depende directamente de la calidad, consistencia, trazabilidad y oportunidad de los datos utilizados a lo largo del ciclo crediticio. Contar con una arquitectura de datos institucional que integra fuentes del core bancario y repositorios analíticos, facilita la implementación de controles de calidad, procesos ETL estandarizados y mecanismos de monitoreo continuo de datos. Se recomienda complementar dicha arquitectura con políticas formales de calidad de datos, definición de responsables, métricas de calidad y alertas tempranas, alineadas a las mejores prácticas de gobierno de datos y a los lineamientos regulatorios vigentes.

CAPÍTULO VI. APLICABILIDAD

El presente capítulo expone la propuesta de aplicación del Modelo Híbrido de Calificación de Riesgo Crediticio Empresarial. Esta se enfoca en trasladar los hallazgos empíricos y las conclusiones de la investigación a un contexto operativo dentro del sistema bancario hondureño. El objetivo es asegurar que el modelo validado, basado en métodos de aprendizaje automático y normativas prudenciales, pueda ser implementado de manera organizada, supervisada y en conformidad con las directrices regulatorias nacionales e internacionales.

A partir de las conclusiones presentadas en los capítulos previos, en especial la evidencia de una mejora en la precisión, solidez y facultad predictiva del modelo híbrido en comparación con los métodos convencionales, este capítulo expone un esquema de implementación factible que integra los elementos técnicos, humanos y estructurales necesarios para su integración institucional.

La propuesta se basa en la metodología CRISP-DM, que dirige cada etapa del proceso de implementación: desde la gobernanza del modelo y la arquitectura de datos hasta la operacionalización del scoring, la incorporación de criterios de prudencia, la supervisión del rendimiento y la recalibración continua. Adicionalmente, se establecen las medidas de control, los indicadores clave de rendimiento (KPIs) y el plan de seguimiento que asegurarán la viabilidad técnica y regulatoria del modelo a largo plazo.

En resumen, este capítulo representa la conexión definitiva entre la investigación científica y su ejecución práctica, al proporcionar una estructura metodológica y funcional que posibilitará a las entidades financieras mejorar la administración del riesgo de crédito empresarial, perfeccionar sus procedimientos de evaluación y contribuir a la estabilidad del sistema bancario hondureño.

6.1 NOMBRE DE LA PROPUESTA

"Implementación de un Modelo Híbrido de Calificación de Riesgo Crediticio Empresarial en Instituciones Bancarias en Honduras".

6.2 JUSTIFICACIÓN DE LA PROPUESTA

La presente propuesta se fundamenta en los hallazgos empíricos y conclusiones del estudio "Evaluación de un Modelo Híbrido de Calificación de Riesgo Crediticio para la Banca Empresarial en Honduras (2020-2024)", en el cual se comprobó que la incorporación de técnicas de aprendizaje

automático combinadas con reglas prudenciales mejora significativamente la precisión y la capacidad predictiva de los modelos tradicionales de calificación de riesgo.

La evidencia cuantitativa obtenida demostró que los algoritmos de Random Forest y XGBoost alcanzaron métricas de desempeño superiores (AUC-ROC de 0.87 y 0.86; KS de 0.58 y 0.55, respectivamente), frente al esquema convencional de reglas de negocio (AUC 0.59; KS 0.18). Asimismo, la validación out-of-time (Test 2024) evidenció estabilidad temporal (PSI=0.0066) y una calibración adecuada (Brier<0.06), demostrando que el enfoque híbrido conserva un rendimiento alto y consistente en contextos cambiantes.

Desde una perspectiva práctica, estos resultados implican que la implementación institucional del modelo híbrido permitirá a las entidades financieras reducir la morosidad y los costos asociados a pérdidas crediticias, ya que el componente de aprendizaje automático incrementa la detección temprana de clientes con alta probabilidad de incumplimiento (Recall 0.86 vs 0.66 en modelos tradicionales).

Además, la propuesta fortalecerá la toma de decisiones crediticias mediante la incorporación de variables de solvencia histórica (Buró Score), cobertura de garantías y señales conductuales que amplían la capacidad explicativa y mitigante del riesgo. Su adopción también impulsará la innovación tecnológica dentro del sistema financiero hondureño, al introducir metodologías analíticas avanzadas adaptadas a la realidad local y fundamentadas en evidencia científica.

La propuesta se ve fortalecida porque el banco x ya tiene una arquitectura de datos madura, permitiendo la fiabilidad de la información histórica, la cual es fiable, está bien organizada y se puede auditar. Esto, a su vez, minimiza los riesgos en las operaciones que se relacionan con la aplicación de este modelo híbrido y hace más fácil su incorporación a los procesos que son originación, seguimiento y control de crédito. La arquitectura tecnológica tiene que ir acompañada de una estrategia completa en la calidad de datos, porque los modelos de análisis avanzados magnifican, tanto lo bueno como lo malo, de la información de base. Por tanto, la propuesta tiene la calidad de datos como elemento clave, para obtener resultados estables y decisiones de crédito que sean confiables.

El impacto esperado trasciende el ámbito técnico, ya que su implementación contribuirá directamente a la estabilidad y sostenibilidad del sistema bancario hondureño, mejorará los

procesos de originación y monitoreo crediticio, y fomentará una cultura de crédito responsable, eficiente y alineada con el crecimiento económico nacional.

En consecuencia, la implementación del Modelo Híbrido de Riesgo Crediticio Empresarial responde a una necesidad real identificada en los resultados del estudio y ofrece una solución validada, sustentada en datos y evidencia estadística, con beneficios comprobables en términos de precisión, eficiencia operativa y cumplimiento regulatorio. Su adopción permitirá transformar la gestión del riesgo crediticio hacia un enfoque más predictivo, automatizado y prudente, generando valor tanto para las instituciones financieras como para el entorno empresarial en general.

6.3 ALCANCE DE LA PROPUESTA

El alcance de esta propuesta se centra en la implementación del Modelo Híbrido de Calificación de Riesgo Crediticio Empresarial dentro del sistema bancario hondureño, con la finalidad de aplicar el modelo desarrollado y validado durante la investigación a nivel institucional, en un entorno controlado y medible.

La implementación incluirá la fase de integración analítica del modelo, su validación interna con datos históricos reales (2020-2024), y la puesta en marcha piloto en el área de riesgo crediticio empresarial. Este alcance está orientado a optimizar la precisión, eficiencia operativa y coherencia regulatoria en la evaluación crediticia de clientes empresariales.

La tabla que se muestra a continuación resume los límites, las inclusiones y los aspectos excluidos que delimitan el alcance de la propuesta, permitiendo distinguir con claridad los elementos contemplados en el diseño del modelo y aquellos que quedan fuera de su implementación.

Tabla 49: Límites, Inclusiones y Exclusiones del Modelo Híbrido de Riesgo Crediticio Empresarial

Límites e Inclusiones	Limitaciones y Exclusiones
Implementación del Modelo Híbrido + reglas de negocio.	No sustituye sistemas actuales de riesgo, solo su complemento durante la fase piloto.
Entrenamiento y validación de algoritmos (Random Forest y XGBoost) sobre la base de datos institucional.	No incluye cartera de clientes personales, solo banca empresarial.
Abarca el diseño de paneles analíticos para la interpretación de resultados y monitoreo de métricas (AUC-ROC, KS, PSI, Brier).	No abarca el desarrollo de infraestructura tecnológica de hardware o la creación de nuevos sistemas de core banking.

Límites e Inclusiones	Limitaciones y Exclusiones
Capacitación del personal del área de riesgos y analítica de datos. Alineación normativa CNBS/Basilea III	Las pruebas se limitan al período de observación enero 2020-octubre 2024, con actualizaciones anuales controladas.

Fuente: Elaboración propia (2025).

La siguiente tabla detalla las dimensiones contextuales que establecen el alcance geográfico, demográfico y temporal de la propuesta, esenciales para comprender el entorno en el que se llevará a cabo la implementación piloto.

Tabla 50: Dimensiones Geográfica, Poblacional y Temporal

Dimensión	Descripción
Geográfico	Entorno corporativo de instituciones bancarias con sede en Tegucigalpa y San Pedro Sula.
Poblacional	Clientes del segmento empresarial con productos de crédito activos y analistas de riesgo institucional.
Temporal	Implementación estimada durante el primer semestre de 2026, con evaluación de resultados al cierre del mismo año.

Fuente: Elaboración propia (2025).

6.3.1 OBJETIVO GENERAL

Implementar (S) el Modelo Híbrido de Calificación de Riesgo Crediticio Empresarial en instituciones bancarias seleccionadas de Honduras, mediante su integración y validación (M) con métricas de desempeño como AUC-ROC y Recall, durante el primer semestre de 2026 (T), utilizando los recursos técnicos y humanos disponibles (A), con el propósito de mejorar la precisión, agilidad y confiabilidad de la evaluación del riesgo crediticio, contribuyendo al fortalecimiento prudencial y estabilidad del sistema financiero nacional (R).

6.3.2 OBJETIVOS ESPECÍFICOS

1. Integrar (S) y adaptar el modelo híbrido en las plataformas del área de riesgo crediticio de la institución, asegurando compatibilidad técnica y monitoreo de desempeño superior al 90% de integración funcional (M), utilizando el personal técnico actualmente asignado y la infraestructura existente (A); fortaleciendo el cumplimiento de los lineamientos de la CNBS y Basilea III (R), dentro del período enero-marzo 2026 (T).
2. Validar los resultados del modelo híbrido frente a los métodos tradicionales de calificación crediticia (S), midiendo métricas de desempeño (AUC-ROC, KS y Recall) (M) con el apoyo del equipo de analítica del área de riesgo, empleando datos históricos 2020-2024 (A), con

el fin de optimizar la detección temprana de riesgo crediticio en el portafolio empresarial (R), a ejecutarse durante los meses de abril-junio de 2026 (T).

3. Capacitar al personal especializado en el uso, interpretación y mantenimiento del modelo híbrido (S), alcanzando la formación de 10 analistas de riesgo con nivel de satisfacción superior al 80 % (M), a través de talleres presenciales y sesiones técnicas dirigidas por el equipo desarrollador del modelo (A), fomentando una cultura organizacional basada en analítica avanzada y toma de decisiones sustentada en datos (R), al realizarse en el segundo trimestre de 2026 (abril-junio) (T).

6.4 DESCRIPCIÓN Y DESARROLLO

En esta sección se detalla el contenido esencial de la propuesta de aplicabilidad. El propósito es explicar, con claridad y suficiente nivel de detalle, qué se hará y cómo se hará en la institución financiera para implementar el modelo híbrido de calificación de riesgo crediticio empresarial, de forma que los productos (entregables) puedan ser utilizados de manera directa e inmediata por las áreas de riesgo, negocio, tecnología y cumplimiento.

La propuesta se estructura en fases coherentes con la metodología CRISP-DM utilizada en la investigación (Capítulo III) y con los hallazgos empíricos del Capítulo IV. Cada fase produce entregables concretos: manuales, protocolos, algoritmos, tableros de control, planes y documentos de gobernanza, que en su conjunto permiten que el banco incorpore el modelo híbrido a sus procesos de originación y monitoreo de crédito empresarial.

6.4.1 DESCRIPCIÓN

6.4.1.1 ACCIONES PRINCIPALES, PROCESOS Y ESTRATEGIAS (QUÉ SE HARÁ)

La propuesta se compone de siete líneas de acción principales, cada una con productos específicos:

1. Gobernanza del Modelo Híbrido

- Definir la estructura de roles y responsabilidades (Riesgos, Analítica, Tecnología, Negocio, Cumplimiento y Validación Independiente).
- Establecer políticas y protocolos de uso del modelo (alcance, límites, overrides, salvaguardas CNBS).
- Documentar el modelo conforme a SR 11-7 (propósito, supuestos, limitaciones, versión,

historial de cambios).

2. Arquitectura de Datos y Pipeline Analítico

- Diseñar y documentar el flujo completo de datos desde los sistemas fuente (core bancario, garantías, buró de crédito, CNBS) hasta el repositorio analítico.
- Implementar procesos ETL/ELT para alimentación periódica del modelo (mensual o según ventana requerida).
- Garantizar calidad y trazabilidad de los datos: controles de nulos, rangos, duplicados y consistencia temporal.

3. Implementación Técnica de Algoritmos y Scoring

- Desplegar el pipeline de preparación de datos usado en la tesis (capping, imputación, codificación, estandarización) en un entorno de producción.
- Implementar el modelo híbrido (Regresión Logística + Random Forest / XGBoost) como servicio de scoring (batch o API) para la cartera empresarial.
- Definir y parametrizar el umbral operativo obtenido por máximo KS en Valid 2023, ajustable solo mediante procedimiento formal.

4. Integración de Reglas Prudenciales y de Negocio

- Incorporar al motor de decisión reglas basadas en:
 - Categoría CNBS (1 a 5).
 - Cobertura mínima de garantías por sector y tamaño de ticket.
 - Límites de concentración sectorial y geográfica.
- Diseñar y documentar la política de overrides y su circuito de aprobación y registro.

5. Tableros de Monitoreo y Sistema de Indicadores (KPIs)

- Implementar tableros en Power BI, u otra herramienta institucional que muestren en tiempo casi real: AUC, KS, Recall, Brier, PSI, FN/FP, % de overrides, etc.
- Incorporar alarmas visuales (verde, amarillo, rojo) según los límites de control definidos en la sección 6.5.

- Permitir desagregación por subcarteras (sector, zona, categoría CNBS, tamaño de crédito).

6. Capacitación y Acompañamiento a Usuarios Clave

- Capacitar a analistas de riesgo, oficiales de crédito y gestores de recuperaciones en la interpretación del score y de los tableros.
- Entrenar al personal de Tecnología y Analítica en la operación y mantenimiento del pipeline.
- Proporcionar manuales de usuario y guías rápidas que faciliten la adopción.

7. Monitoreo, Recalibración y Mejora Continua

- Implementar el plan de seguimiento de indicadores descrito en la sección 6.5.2.
- Establecer criterios de recalibración y reentrenamiento del modelo (basados en PSI, caída de AUC/KS, deterioro de Brier).
- Documentar cada iteración y actualización del modelo (versionado y control de cambios).

6.4.1.2 JUSTIFICACIÓN METODOLÓGICA DEL DISEÑO DE LA PROPUESTA (CÓMO SE HARÁ Y POR QUÉ ASÍ)

La estructura de la propuesta responde a tres criterios metodológicos:

1. Coherencia con CRISP-DM y la Investigación

- Las fases de la propuesta reflejan las seis etapas de CRISP-DM: entendimiento del negocio, entendimiento de los datos, preparación, modelado, evaluación y despliegue.
- Esto garantiza replicabilidad: el mismo proceso que se siguió en la tesis para construir y validar el modelo se reutiliza como guía para implementarlo de forma industrializada.

2. Alineamiento Regulatorio y de Riesgo de Modelo

- La inclusión explícita de gobernanza, validación independiente, protocolos de overrides y monitoreo continuo responde a las recomendaciones de SR 11-7 (gestión de riesgo de modelo) y a los principios de BCBS 239 (agregación y calidad de datos).
- De esta forma, la institución no solo adopta un modelo más preciso, sino que lo hace bajo un marco de control compatible con la supervisión prudencial como la CNBS y el Banco Central.

3. Orientación a la Aplicabilidad Inmediata

- Cada línea de acción produce entregables concretos (manuales, algoritmos, tableros, protocolos) que la institución puede utilizar sin requerir rediseños conceptuales.
- El énfasis no se limita al componente técnico (algoritmos), sino que cubre también procesos, personas y reglas de negocio, lo que facilita que la solución se incorpore al día a día del banco.

En síntesis, el qué y el cómo de la propuesta se organizan en fases lógicas que conectan el modelo validado en la investigación con la realidad operativa de la banca empresarial hondureña, asegurando rigurosidad metodológica y viabilidad práctica.

6.4.2 PIPELINES ANALÍTICOS DEL MODELO HÍBRIDO

Con el propósito de asegurar la correcta implementación, sostenibilidad y reproducibilidad del modelo híbrido de calificación de riesgo crediticio, se propone el diseño de pipelines analíticos estructurados, que organizan de forma secuencial el flujo de datos desde su origen hasta la generación del score crediticio y su posterior monitoreo.

Estos pipelines permiten operacionalizar el modelo dentro de una arquitectura de datos institucional, garantizando consistencia metodológica, control del riesgo operativo y alineación con las buenas prácticas de analítica de datos aplicadas al sector financiero.

6.4.2.1 INGESTA Y PREPARACIÓN DE DATOS

Este pipeline tiene como finalidad asegurar la recopilación, integración y disponibilidad de los datos necesarios para el análisis crediticio.

Las principales actividades incluyen la extracción de información desde los sistemas fuente como ser: core bancario, buró de crédito, sistemas contables y reportes regulatorios, la integración de datos históricos por períodos definidos (entrenamiento, validación y prueba out-of-time), así como validaciones básicas de estructura, tipos de datos y duplicidad de registros. Asimismo, se contempla la anonimización de información sensible conforme a principios éticos y normativos.

Resultado: Conjunto de datos consolidado, confiable y trazable, listo para los procesos analíticos.

6.4.2.2 CALIDAD DE DATOS Y TRANSFORMACIÓN DE VARIABLES

Este pipeline se enfoca en mejorar la calidad de la información y en la construcción de variables analíticas relevantes para la predicción del riesgo crediticio.

Incluye el diagnóstico y tratamiento de valores faltantes, la detección y mitigación de valores atípicos, la normalización y estandarización de variables cuantitativas, así como la codificación de variables categóricas. Adicionalmente, se desarrollan variables financieras, conductuales y regulatorias alineadas con la lógica del negocio crediticio.

Resultado: Dataset analítico optimizado, consistente y adecuado para el modelado predictivo.

6.4.2.3 MODELADO PREDICTIVO Y VALIDACIÓN

Este pipeline comprende el entrenamiento, evaluación y comparación de los distintos modelos considerados en el estudio.

Se ejecuta el entrenamiento de modelos de referencia (reglas de negocio y regresión logística) y de modelos de aprendizaje automático, incorporando procesos de optimización de hiperparámetros y validación fuera de muestra. El desempeño se evalúa mediante métricas como AUC-ROC, KS, recall y precisión, complementadas con análisis de estabilidad y robustez.

Resultado: Modelo híbrido validado, con desempeño superior y estabilidad temporal comprobada.

6.4.2.4 SCORING, MONITOREO Y GOBIERNO DEL MODELO

Este pipeline permite la operación controlada del modelo en un entorno productivo.

Incluye la generación del score crediticio y de la probabilidad de incumplimiento (PD), la integración del resultado con reglas de negocio y políticas crediticias, así como el monitoreo continuo del desempeño, la estabilidad poblacional y la calibración del modelo. Asimismo, se consideran prácticas de documentación, control de versiones y seguimiento periódico.

Resultado: Sistema de calificación crediticia operativo, auditable y alineado con los principios de gestión de riesgo del modelo.

6.4.2.5 FLUJO INTEGRAL DEL MODELO HÍBRIDO DE RIESGO CREDITICIO

La figura 48 presenta el flujo integral del modelo híbrido de riesgo crediticio, iniciando con la ingesta de datos desde los sistemas fuente, seguida por los procesos de calidad de datos y

transformación de variables, el modelado predictivo y validación, y finalizando con el scoring crediticio, el monitoreo continuo y el gobierno del modelo.

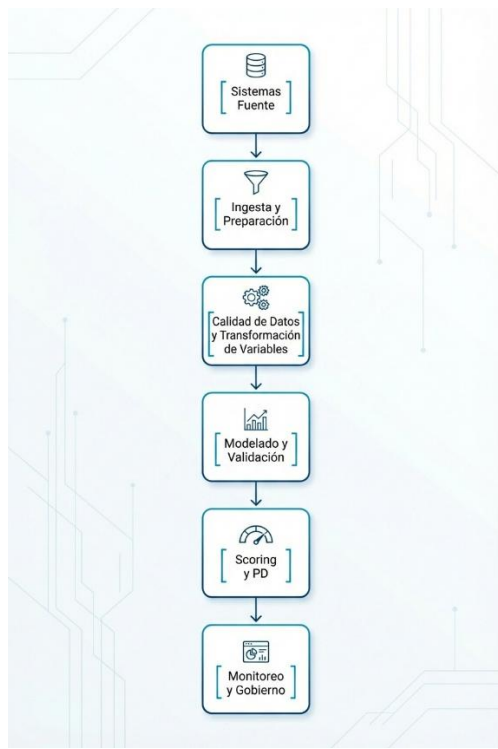


Figura 48: Flujo Integral del Modelo Híbrido de Riesgo Crediticio

Fuente: Elaboración Propia

6.4.2.6 ALINEACIÓN DE LOS PIPELINES CON BUENAS PRÁCTICAS REGULATORIAS

Los pipelines propuestos se encuentran alineados con principios internacionales de gestión de datos y riesgo de modelos, particularmente con BCBS 239, 2013; *SR 11-7*, 2011) al asegurar:

- Calidad y trazabilidad de los datos a lo largo del ciclo analítico.
- Separación entre desarrollo, validación y monitoreo del modelo.
- Evaluación periódica del desempeño y estabilidad.
- Documentación, control de versiones y gobierno del modelo.

Esto refuerza la confiabilidad del modelo híbrido y su viabilidad dentro de un entorno regulado como el sistema financiero en Honduras.

6.4.3 DESARROLLO

En esta subsección se describen, a nivel de producto final, los entregables que la institución recibirá para implementar la propuesta. Cada entregable está diseñado para ser aplicable de inmediato, con el nivel de detalle necesario para que las áreas internas puedan adoptarlo sin reinterpretaciones adicionales.

Listado de entregables principales:

1. Manual de Gobierno del Modelo Híbrido.
2. Diseño del Pipeline de Datos y Documento de Arquitectura Analítica.
3. Paquete de Algoritmos y Scripts de Scoring.
4. Motor de Reglas Prudenciales y Protocolo de Overrides.
5. Tableros de Control (Dashboards) de Riesgo y Desempeño del Modelo.
6. Fichas metodológicas de indicadores (KPIs).
7. Plan y material de Capacitación para usuarios y técnicos.
8. Plan de Monitoreo, Recalibración y Reentrenamiento.

A continuación, se desarrolla cada uno en detalle:

1. Manual de Gobierno del Modelo Híbrido

- Contenido:
 - Descripción del modelo (objetivo, alcance, población, variables clave, algoritmos utilizados).
 - Roles y responsabilidades: Riesgos (owner), Analítica, Tecnología, Negocio, Cumplimiento, Validación Independiente, Comité de Modelos.
- Políticas de uso:
 - Procesos en los que el modelo es obligatorio (originación, renovación, seguimiento).
 - Procesos en los que tiene carácter consultivo.
- Criterios de aceptación de decisiones y política de overrides:

- Casos en los que un analista puede modificar la recomendación del modelo.
- Flujo de aprobación (quién aprueba y cómo se documenta).
- Reglas de versionado del modelo:
 - Numeración de versiones, cambios mayores y menores.
 - Requisitos para promover una nueva versión a producción.
- Referencia explícita a los marcos SR 11-7, BCBS 239 y normativa CNBS.

Aplicabilidad inmediata: El banco puede incorporar este manual a su compendio de políticas de riesgo y utilización de modelos, sirviendo como base para auditorías internas y externas.

2. Diseño del Pipeline de Datos y Documento de Arquitectura Analítica

- Diagrama de flujo de datos desde los sistemas fuente hasta el scoring:
 - Tablas y campos requeridos (core, garantías, CNBS, buró).
 - Transformaciones: limpieza, joins, agregaciones, derivación de features.
- Especificaciones técnicas del pipeline de preparación:
 - Tratamiento de outliers (reglas de capping por variable).
 - Reglas de imputación por variable (mediana, “Sin información”).
 - Codificación de variables categóricas (one-hot, agrupación de categorías raras).
 - Estandarización z-score en el caso de la regresión logística.
- Diseño físico:
 - Vista analítica central (estructura de tabla final).
 - Jobs ETL/ELT, frecuencia de actualización, logs y alertas de fallos.
- Formato: Documento técnico + diagramas (por ejemplo, en Visio o herramienta corporativa) y scripts SQL/KNIME adjuntos.

Aplicabilidad inmediata: El área de Tecnología puede utilizar este diseño como plano para construir los procesos ETL y la vista analítica que alimentará al modelo.

3. Paquete de Algoritmos y Scripts de Scoring

- Contenido:
 - Modelos serializados (por ejemplo, pipeline_logistic.pkl, pipeline_rf.pkl, pipeline_xgb.pkl) que incluyen:
 - Preprocesamiento (transformers).
 - Modelo entrenado (logística, RF, XGB).
 - Scripts o componentes para:
 - Ingesta de datos (desde la vista analítica).
 - Estandarización de variables y generación de features.
 - Cálculo del score y de la probabilidad de incumplimiento (PD).
 - Aplicación del umbral operativo definido (clasificación BUENO - MALO).
 - Documentación de interfaces:
 - Entradas requeridas (schema).
 - Formato de salida (score, PD, etiqueta, razones principales).
 - Ejemplos de uso usando clientes reales anonimizados, que replican los ejemplos ilustrativos del Capítulo VI.
- Formato: Carpeta con código (Python/R/KNIME según estándar institucional), manual de instalación y guía de integración para IT.

Aplicabilidad inmediata: Permite que el área de Tecnología levante el servicio de scoring y lo conecte con los sistemas transaccionales sin rediseñar el modelo.

4. Motor de Reglas Prudenciales y Protocolo de Overrides

- Contenido:
 - Listado estructurado de reglas de negocio y prudenciales:
 - Reglas basadas en categoría CNBS (por ejemplo, tratamiento especial de categorías 3-5).

- Reglas de cobertura de garantías por tipo de cliente y tamaño de crédito.
- Reglas sectoriales (sectores en observación o con políticas diferenciadas).
- Árbol de decisión que integra:
 - Resultado del modelo híbrido (score/PD y clasificación sugerida).
 - Reglas prudenciales (hard rules).
 - Reglas de negocio (soft rules y criterios de excepción).
- Protocolo de overrides:
 - Cuando puede un analista cambiar la recomendación del motor.
 - Cómo documentar la justificación (motivo, evidencia) en el sistema.
 - Quién aprueba (jefatura, comité), según monto y nivel de riesgo.
- Formato:
 - Documento de reglas en formato legible + lógica formal en pseudocódigo o especificación para motor de reglas.
 - Matriz de decisión (tablas) que puede implementarse directamente en el sistema de originación.

Aplicabilidad inmediata: Riesgos y Negocio pueden cargar estas reglas en el motor de decisión existente y aplicarlas en la evaluación diaria de solicitudes de crédito.

5. Tableros de Control (Dashboards)

- Contenido:
 - Diseño y prototipos de dashboards en Power BI (u otra herramienta corporativa) con los siguientes componentes mínimos:
 - Vista ejecutiva:
 - AUC-ROC, KS, Recall, Precision, F1, Brier y PSI a nivel global.
 - Tendencia mensual de estos indicadores.
 - Vista de riesgo:

- Distribución del score por subcarteras (sector, zona, CNBS, tamaño).
- Mapas de calor de FN y FP por segmento.
- Vista operacional:
 - Volumen de solicitudes procesadas, aprobadas, rechazadas y en revisión.
 - % de overrides y razones más frecuentes.
- Filtros dinámicos por periodo, producto, región, segmento de cliente, etc.
- Formato: Archivo .pbix (u otro formato de la herramienta escogida) + documentación básica (diccionario de campos, fuente de datos, periodicidad de actualización).

Aplicabilidad inmediata: Puede ser conectado al data warehouse del banco y actualizado automáticamente con los ETL definidos, dando visibilidad inmediata a la alta gerencia y a las áreas de riesgo.

6. Fichas Metodológicas de Indicadores (KPIs)

- Contenido:
 - Fichas completas para cada KPI prioritario: AUC-ROC, KS, Recall, Brier, PSI, FN/FP, % overrides, ahorro por reducción de FN, etc.
 - Cada ficha incluye:
 - Nombre, definición operacional, fórmula, unidad de medida.
 - Frecuencia de medición, fuente de datos, universo/población.
 - Responsables de cálculo y de uso.
 - Línea base, metas y límites de control mínimo/máximo aceptables.
 - Instrucciones para interpretación y acciones recomendadas cuando se exceden límites.
- Formato: Documento consolidado de fichas (por ejemplo, en Word o Excel) que puede integrarse al sistema institucional de indicadores.

Aplicabilidad inmediata: Permite que la Oficina de Gestión de Riesgos o la unidad de

Calidad pueda incorporar estos KPIs a sus tableros corporativos y a los reportes periódicos a la dirección.

7. Plan y Material de Capacitación

- Contenido:
 - Programa de capacitación para tres perfiles de usuario:
 - Analistas de riesgo / oficiales de crédito.
 - Personal de Tecnología y Analítica.
 - Gestión de recuperaciones y alta gerencia.
 - Presentaciones (diapositivas), casos prácticos y manuales de usuario que explican:
 - Conceptos básicos del modelo híbrido.
 - Lectura de scores y PD.
 - Uso de dashboards.
 - Interpretación de KPIs y alarmas.
 - Aplicación correcta de la política de overrides.
- Formato: Carpeta con presentaciones, guías rápidas y manual detallado para cada rol.

Aplicabilidad inmediata: Puede utilizarse como contenido oficial de talleres internos y para inducción de nuevo personal en las áreas de riesgo y crédito.

8. Plan de Monitoreo, Recalibración y Reentrenamiento

- Contenido:
 - Calendario y responsabilidades de seguimiento mensual, trimestral y semestral de los KPIs descritos en 6.5.
 - Procedimiento detallado para:
 - Detectar drift (mediante PSI y análisis de distribución de variables).
 - Analizar caída de desempeño (AUC-KS) y deterioro de calibración (Brier).
 - Decidir cuándo solo recalibrar y cuándo reentrenar el modelo.

- Protocolo de reentrenamiento:
 - Definición de nueva ventana TRAIN, VALID y TEST.
 - Repetición del ciclo de CRISP-DM con cambios mínimos de arquitectura.
 - Validación independiente y aprobación del Comité de Modelos.
 - Checklists para documentar cada ciclo de actualización y asegurar la trazabilidad histórica.
- Formato: Documento de procedimiento estándar (SOP) que se integra al manual de procesos de la unidad de Analítica y Riesgos.

Aplicabilidad inmediata: Se convierte en la guía operativa para mantener el modelo híbrido vigente a lo largo del tiempo, reduciendo el riesgo de obsolescencia y facilitando auditorías y revisiones supervisoras.

6.5 MEDIDAS DE CONTROL

Las medidas de control establecen los mecanismos de evaluación y seguimiento de la propuesta, permitiendo medir su nivel de eficacia, eficiencia e impacto. Estos controles se fundamentan en la definición de indicadores clave (KPIs) cuantitativos y cualitativos que garanticen trazabilidad de los resultados durante la implementación y operación del Modelo Híbrido de Calificación de Riesgo Crediticio Empresarial.

Su propósito es verificar en qué medida la propuesta cumple los objetivos establecidos en términos de desempeño predictivo, adopción institucional y contribución a la gestión prudencial.

6.5.1 MARCO DE MONITOREO Y VALIDACIÓN CONTINUA

Con el fin de asegurar la vigencia y confiabilidad del modelo híbrido en su fase operativa, se establece un esquema de monitoreo continuo basado en las métricas AUC-ROC, KS y PSI.

- Periodicidad: evaluación mensual del PSI y trimestral del desempeño (AUC y KS).
- Reglas de acción:
 - $PSI \geq 0.25$: análisis de causas y revisión del modelo.
 - Deterioro sostenido de AUC o KS: ajuste del modelo o reentrenamiento.
- Gobernanza: los resultados del monitoreo deberán documentarse y formar parte del proceso

de gestión del riesgo del modelo.

Este marco permite detectar oportunamente fenómenos de degradación del desempeño (model drift) y cambios estructurales en la población (population drift), garantizando la toma de decisiones crediticias informadas y consistentes en el tiempo.

6.5.2 INDICADORES

Los indicadores propuestos están diseñados para monitorear, de forma integral, el desempeño predictivo, la estabilidad, la calibración y la aceptabilidad institucional del modelo híbrido.

6.5.2.1 INDICADORES CUANTITATIVOS

Los indicadores cuantitativos posibilitarán la evaluación numérica del rendimiento técnico, operativo y financiero del modelo, confirmando su progreso en comparación con las metodologías convencionales.

Tabla 51: Indicadores Cuantitativos para la Evaluación del Modelo Híbrido de Riesgo Crediticio Empresarial

Indicador	Descripción/ Fórmula	Unidad / Meta Esperada	Fuente/Medición	Frecuencia de Monitoreo
AUC-ROC (%)	Área bajo la curva ROC, medirá la capacidad del modelo para distinguir entre los clientes buenos y malos.	≥ 0.85 en validación y test	Reportes de desempeño del modelo.	Trimestral
KS (Kolmogórov-Smirnov)	Diferencia entre distribuciones de score BUENO/MALO	≥ 0.50	Sistema Analítico de validación.	Trimestral
Precisión / Recall (%)	Relación entre verdaderos positivos y falsos positivos (capacidad de detección)	Recall ≥ 0.80	Resultado de pruebas del modelo.	Trimestral
Tiempo de evaluación de crédito (horas)	Medición promedio del tiempo para calificar un crédito.	Reducción $\geq 25\%$	Reportes operativos internos.	Semestral
Reducción de morosidad (%)	Diferencia porcentual en tasa de mora respecto al modelo anterior.	-10 % respecto a línea base	Reportes CNBS / cartera bancaria	Semestral
Rentabilidad del Portafolio	Incremento en la eficiencia de asignación del crédito.	+5 % promedio anual.	Análisis financiero institucional	Anual

Fuente: Elaboración Propia

6.5.2.2 INDICADORES CUALITATIVOS

Los indicadores cualitativos valorarán la percepción, la adopción y la satisfacción institucional en relación con la implementación del modelo, garantizando la sostenibilidad y la aceptación de la transformación.

Tabla 52: Indicadores cualitativos

Indicador	Descripción / Criterio de Éxito	Método de Medición	Meta Esperada	Periodicidad
Satisfacción del personal de analista	Nivel de aceptación y utilidad percibida del modelo en los procesos de gestión de riesgo.	Encuestas y entrevistas de satisfacción	≥ 80 % de satisfacción	Semestral
Nivel de adopción institucional	Porcentaje de áreas que utilizan efectivamente el modelo en su flujo de calificación crediticia.	Revisión de uso en sistemas/ reportes operativos.	≥ 70 % de adopción	Semestral
Cumplimiento de lineamientos CNBS-Basilea III	Coherencia entre el modelo implementado y las normativas prudenciales.	Auditorías internas y revisión de reportes regulatorios.	Cumplimiento total	Anual
Madurez analítica organizacional	Evaluación del grado de incorporación de analítica avanzada y cultura de datos.	Escala interna de madurez (1 a 5)	Nivel ≥ 4	Anual
Percepción de mejora en la toma de decisiones	Opinión de jefes de riesgo y gerencia sobre la calidad de decisiones crediticias tras la implementación.	Entrevistas estructuradas y focus group	Mejora percibida ≥ 80 % de participantes.	Anual

Fuente: Elaboración Propia

Para los mecanismos de monitoreo y evaluación, se establece que su implementación tendrá como propósito garantizar la confiabilidad de las métricas empleadas y la continuidad del proceso de control de la siguiente forma:

- Construcción de tablero de indicadores (dashboard) en Power BI u otra herramienta equivalente que consolide los datos de desempeño del modelo.
- Revisión trimestral de resultados por parte del Comité de Riesgo Institucional, con validación técnica del área de analítica de datos.
- Las discrepancias con respecto a los objetivos serán tratadas a través de planes de mejora continua, incluyendo el reentrenamiento de algoritmos, el ajuste de variables y la recalibración de parámetros.

- Los resultados cualitativos se examinarán mediante encuestas institucionales, herramientas y reportes de adopción, lo que permitirá evidenciar el grado de aceptación y satisfacción del modelo.

6.5.2.3 FICHAS METODOLÓGICAS DE INDICADORES

Se elaboraron fichas metodológicas que complementan las medidas de control, estandarizando los procedimientos de seguimiento y las responsabilidades de medición. En ellas se presentan 3 indicadores representativos:

Tabla 53: Ficha Metodológica del Indicador AUC - ROC del Modelo Híbrido

Campo	Detalle
Nombre del Indicador	Desempeño Predictivo (AUC-ROC) del Modelo Híbrido
Tipo del Indicador	Cuantitativo - Técnico
Objetivo Asociado	Validar empíricamente los resultados del modelo frente a métodos tradicionales.
Descripción	Mide la capacidad del modelo para diferenciar clientes buenos y malos en la cartera empresarial. Entre mayor sea el valor, mejor desempeño predictivo.
Fórmula/Método de Cálculo	Cálculo del área bajo la curva ROC. $AUC = \int_0^1 TPR(FPR) d(FPR)$ donde TPR = tasa de verdaderos positivos y FPR = tasa de falsos positivos.
Unidad de Medida	Porcentaje (%)
Línea base	AUC-ROC promedio actual = 0.65
Meta Esperada	AUC-ROC ≥ 0.85 en validación y prueba
Fuente de Verificación	Archivo de resultados del modelo / plataforma Knime o Power BI
Frecuencia de Medición	Trimestral
Responsable	Área de analítica y modelos de riesgos
Medio de Recolección	Dashboard de desempeño, scripts de validación Python
Actores Involucrados	Data Scientists, Analistas de riesgos y comité de validación
Observación	Este KPI refleja la efectividad técnica del modelo y es clave para la continuidad de la implementación.

Fuente: Elaboración propia (2025).

Tabla 54: Ficha Metodológica del Indicador Reducción de Morosidad (%)

Campo	Detalle
Nombre del Indicador	Reducción de Morosidad en la Cartera Crediticia Empresarial (%)
Tipo del Indicador	Cuantitativo – Financiero

Campo	Detalle
Objetivo Asociado	Demostrar la eficiencia del modelo en la mejora de la calidad del portafolio crediticio y la reducción de incumplimientos.
Descripción	Mide el cambio porcentual en la tasa de morosidad de la cartera empresarial al comparar los resultados posteriores a la implementación frente al modelo tradicional.
Fórmula/Método de Cálculo	Reducción de Morosidad (%) = ((Tasa de Mora Antes - Tasa de Mora Después) / Tasa de Mora Antes) × 100
Unidad de Medida	Porcentaje (%)
Línea base	Tasa de morosidad promedio 2024= 2.6% (según CNBS)
Meta Esperada	Lograr reducción ≥ 10% de la tasa de morosidad en el primer año de implementación del modelo.
Fuente de Verificación	Reportes de cartera de Banco Central de Honduras (BCH), CNBS e informes internos de cartera.
Frecuencia de Medición	Semestral
Responsable	Gerencia de Riesgos / Unidad de Modelos Analíticos.
Medio de Recolección	Dashboard Financiero institucional, informe de desempeño de portafolio.
Actores Involucrados	Analistas de riesgo crediticio, auditores internos y comité de morosidad.
Observación	Este KPI evidencia el beneficio tangible y financiero del modelo, vinculando su eficacia técnica (predictiva) con resultados prudenciales observables en la cartera empresarial.

Fuente: Elaboración propia (2025).

Tabla 55: Ficha metodológica del indicador Satisfacción del Personal Analista

Indicador	Definición
Nombre del Indicador	Nivel de Satisfacción del Personal Analista
Tipo del Indicador	Cualitativo - Perceptual
Objetivo Asociado	Capacitar al personal especializado y fomentar una cultura institucional basada en analítica avanzada.
Descripción	Evalúa la percepción y aceptación del modelo por parte de los analistas que lo utilizan en sus procesos diarios.
Fórmula/Método de Cálculo	Promedio obtenido de encuestas estructuradas de satisfacción con escala Likert (1-5). Satisfacción = Σ Puntajes Obtenidos / Número de encuestas
Unidad de Medida	Porcentaje de satisfacción (%)
Línea base	Nivel previo al proyecto: 55% de satisfacción
Meta Esperada	Lograr una satisfacción ≥ 80% tras la capacitación
Fuente de Verificación	Encuestas institucionales post - capacitación
Frecuencia de Medición	Semestral
Responsable	Departamento de Recursos Humanos / Área de Riesgo

Indicador	Definición
Medio de Recolección	Formulario digital (Microsoft Forms, Google Forms)
Actores Involucrados	Analistas de riesgo, jefe de analítica y equipo de capacitación
Observación	Este indicador mide la adopción cultural del modelo y la aceptación del cambio tecnológico.

Fuente: Elaboración propia (2025).

Nota: Los indicadores y fichas metodológicas fueron elaborados por los autores con base en métricas reconocidas de modelado de riesgo crediticio y buenas prácticas de gestión de desempeño.

En síntesis, las disposiciones de control implementadas incorporan indicadores tanto cuantitativos como cualitativos, asegurando así la evaluación continua del Modelo Híbrido de Calificación de Riesgo Crediticio Empresarial. Las especificaciones metodológicas posibilitan la estandarización del seguimiento, simplifican la interpretación de las métricas y garantizan que el modelo preserve su eficacia técnica, su repercusión financiera y su aprobación institucional a lo largo del tiempo.

6.5.3 PLAN DE SEGUIMIENTO

El presente plan de seguimiento establece el procedimiento operativo para la vigilancia continua de los indicadores definidos en las Medidas de Control (6.5.1), garantizando la evaluación constante del rendimiento técnico, financiero y organizacional del Modelo Híbrido de Calificación de Riesgo Crediticio Empresarial.

Su propósito es asegurar la persistencia de los objetivos establecidos, fomentando la mejora continua, la rastreabilidad de los resultados y la adopción de decisiones fundamentadas en pruebas.

El plan monitorea sistemáticamente la evolución de los indicadores clave de rendimiento, tanto cuantitativos como cualitativos, con el propósito de identificar desviaciones, implementar acciones correctivas de manera oportuna y mantener la transparencia en la operación del modelo.

El alcance del seguimiento abarca 3 componentes estratégicos:

1. Rendimiento del modelo (AUC-ROC, KS, Recall, Precisión, Brier y PSI)
2. Impacto financiero, reflejado en las fluctuaciones de morosidad y rentabilidad del portafolio empresarial.
3. Aceptación institucional, que incluye la satisfacción del personal analítico y la observancia de las directrices CNBS – Basilea III.

Las labores del plan se ejecutarán a lo largo del ciclo operativo anual del modelo, tras su implementación inicial y bajo la supervisión del Comité de Riesgo Institucional.

A. Estructura del Protocolo de Seguimiento

El seguimiento se realizará a través de cinco fases sucesivas, las cuales constituyen un ciclo de supervisión y aprendizaje continuo, en concordancia con las etapas de Evaluation y Deployment de la metodología CRISP-DM:

Tabla 56: Etapas, Responsables y Frecuencia de Ejecución del Plan de Seguimiento

Etapa	Descripción	Responsable Principal	Frecuencia	Herramienta / Evidencia
Recolección de datos	Consolidar la información técnica, operativa y financiera según los indicadores establecidos (6.5.1)	Área de Analítica de Datos / Departamento de Riesgo	Mensual	Dashboard en Power BI, reportes CNBS, registros de encuestas.
Análisis de Resultados	Evaluar los KPIs cuantitativos y cualitativos comparando los resultados con las metas esperadas.	Analista de Riesgo y Comité de Validación.	Trimestral	Informe de desempeño y tendencias de KPIs.
Revisión y validación	Presentar los resultados ante el comité de riesgo institucional y documentar decisiones y recomendaciones.	Comité de Riesgo / Gerencia General	Trimestral	Actas y reportes ejecutivos.
Implementación de acciones correctivas	Ejecutar ajustes derivados del análisis (reentrenamiento del modelo, recalibración, capacitación del personal)	Unidad de Modelos Analíticos / Recursos Humanos	Semestral	Registros de cambios, plan de acción y versión actualizada del modelo.
Comunicación y cierre del ciclo	Elaborar informe final de seguimiento con resultado, mejoras implementadas y recomendaciones,	Coordinación de Analítica / Comité de Riesgos	Anual	Reporte consolidado y presentación a la alta dirección.

Fuente: Elaboración Propia

Una vez establecidas las etapas generales del protocolo, la siguiente tabla detalla las acciones específicas que deberán ejecutarse en cada período para asegurar la supervisión continua y el perfeccionamiento del modelo:

Tabla 57: Plan de seguimiento y Control

Etapa / Horizonte	Actividad de seguimiento	Detalle Metodológico	Responsable	Producto Esperado
Diario / Semanal	Verificación operativa de ETL y scoring.	Revisión de logs de Jobs, errores, tiempos de respuesta.	IT y Data Eng.	Registro de ejecución sin incidentes.
Mensual	Cálculo de KPIs/KRIs y análisis de desempeño global y por subcarteras.	AUC, KS, Recall, Precision, F1, Brier, PSI, FN/FP, overrides, comparación contra metas y meses previos.	Analítica y Riesgos.	Reporte mensual a Comité de Modelos.
Mensual	Revisión de estabilidad (PSI) y deriva de datos.	PSI del score y de predictores claves, análisis por sector, zona, CNBS	Analítica y Riesgos.	Alertas amarillas/rojas si se superan umbrales.
Trimestral	Revisión de calibración y ajuste si procede.	Brier, curvas de calibración, ECE, recalibración isotónica/Platt si se deteriora la calibración.	Analítica y Riesgos.	Modelo recalibrado (si aplica).
Trimestral	Evaluación de reglas prudenciales y overrides.	Análisis de consistencia de reglas CNBS, garantías, zonas, revisión de overrides y de su impacto en riesgo.	Riesgos y Negocio.	Ajuste fino de reglas y política de overrides.
Semestral	Reentrenamiento (si se activan disparadores).	Nuevo entrenamiento con ventana actualizada, mismo pipeline, comparación con modelo vigente por bootstrap de $\Delta AUC/KS$.	Analítica y Riesgos.	Propuesta de nuevo modelo campeón (si mejora).
Semestral	Validación independiente.	Replicación de resultados, challenge de supuestos, pruebas adversas ligeras.	Unidad independiente.	Informe de validación; recomendación.
Anual	Auditoría SR 11-7 y BCBS 239.	Revisión de documentación, trazabilidad, controles de datos, controles de cambios.	Auditoría interna.	Dictamen de cumplimiento y recomendaciones.

Fuente: Elaboración Propia

B. Mecanismos de control y registro

Para garantizar la trazabilidad de la supervisión, el plan se fundamenta en los siguientes mecanismos:

- Tablero central de indicadores en Power BI, que incluye alertas configuradas según umbrales predefinidos ($AUC < 0.80$, $PSI > 0.10$).

- Un repositorio con control de versiones (Git + OneDrive cifrado) destinado a la gestión de scripts, métricas y actas de comité.
- Alertas automáticas enviadas a los responsables cuando se exceden los umbrales críticos.
- Bitácora de validación que conserva evidencias, fechas, métricas y medidas adoptadas.

C. Procedimientos de acciones correctivas

En caso de identificarse desviaciones o deterioro en los indicadores, se iniciará un proceso formal de mejora continua, que comprenderá las siguientes fases:

- Diagnóstico de causa raíz, para establecer si la degradación se origina en los datos, factores macroeconómicos o la configuración del modelo.
- Diseño de ajuste que puede incluir recalibración, reentrenamiento parcial o revisión de reglas de negocios.
- Validación técnica asegurando que la nueva versión mantenga estabilidad y demuestre mejora en los indicadores claves.
- Autorización formal del comité de Riesgo, seguida por la implementación de la versión revisada en el entorno productivo.

En conclusión, el Plan de Seguimiento consolida la etapa final del ciclo de gestión del modelo, asegurando que la información derivada del monitoreo realmente de manera continua los procesos de análisis y mejora. Esta aproximación metodológica fortalece la coherencia del estudio y establece un sistema de control verificable que sustenta la sostenibilidad técnica y prudencial del modelo. Finalmente, permite a la institución mantener una gestión del riesgo crediticio más precisa, estable y en consonancia con los estándares regulatorios internacionales.

6.5.3.1 EJEMPLOS DE DISPARADORES DE ACCIÓN

- Alerta amarilla:
 - PSI score entre 0.10 y 0.25.
 - Caída de AUC o KS entre 3 y 5 p.p. frente al baseline.
 - Deterioro leve de Brier.
 - **Acción: Análisis diagnóstico;** posible ajuste de umbral; refuerzo de reglas; mejora en calidad de datos.

- Alerta roja:
 - PSI score > 0.25 (un mes) o > 0.25 en dos meses consecutivos.
 - Caída de AUC o KS ≥ 5 p.p. de forma sostenida.
 - Deterioro significativo de calibración.
 - **Acción: Inicio de proceso formal de reentrenamiento**, validación independiente y aprobación de nuevo modelo por el Comité.

6.5.4 ANÁLISIS DE RIESGOS ÉTICOS Y GOBERNANZA ALGORÍTMICA

La implementación de modelos de aprendizaje automático en procesos de calificación de riesgo crediticio conlleva riesgos éticos y organizacionales que deben ser identificados, evaluados y mitigados de manera explícita. En este sentido, el presente proyecto incorpora un enfoque de gobernanza algorítmica, orientado a garantizar un uso responsable, transparente y alineado con los principios regulatorios y éticos del sector financiero.

6.5.4.1 RIESGOS ÉTICOS IDENTIFICADOS

Entre los principales riesgos éticos asociados al uso del modelo híbrido se identifican los siguientes:

- Sesgo algorítmico, derivado de posibles distorsiones históricas en los datos crediticios.
- Discriminación indirecta, especialmente en variables correlacionadas con características sensibles.
- Opacidad del modelo, que podría dificultar la comprensión de las decisiones automatizadas.
- Uso indebido del score crediticio, fuera de los fines previstos en las políticas institucionales.
- Dependencia excesiva del modelo, sin supervisión humana adecuada.

6.5.4.2 MEDIDAS DE MITIGACIÓN ÉTICA

Para mitigar estos riesgos, el modelo incorpora las siguientes salvaguardas:

- Uso exclusivo de variables permitidas por la normativa vigente, evitando atributos sensibles o prohibidos.
- Implementación de procesos de anonimización y resguardo de la información, garantizando

la protección de datos personales.

- Evaluación periódica del desempeño y estabilidad del modelo, con el fin de detectar desviaciones o comportamientos anómalos.
- Integración del modelo como herramienta de apoyo a la decisión, manteniendo siempre la validación y supervisión por parte de analistas de riesgo.
- Documentación clara del funcionamiento del modelo y de los criterios generales de decisión.

6.5.4.3 GOBERNANZA ALGORÍTMICA DEL MODELO

La gobernanza del modelo híbrido se fundamenta en la definición de roles, responsabilidades y procesos de control, que aseguran su uso adecuado dentro de la organización.

En este marco:

- El modelo se encuentra sujeto a validación independiente, previa a su puesta en producción.
- Se establecen mecanismos de monitoreo continuo, que permiten evaluar su desempeño, estabilidad y consistencia en el tiempo.
- Las decisiones críticas basadas en el modelo requieren revisión humana, particularmente en casos de alto impacto crediticio.
- Cualquier ajuste, recalibración o reentrenamiento del modelo debe ser debidamente documentado y aprobado conforme a los lineamientos internos.

6.5.4.4 ALINEACIÓN CON PRINCIPIOS REGULATORIOS Y ÉTICOS

El enfoque adoptado es consistente con principios internacionales de gestión del riesgo de modelos y protección de datos, promoviendo la transparencia, la trazabilidad y la responsabilidad en el uso de sistemas algorítmicos para la toma de decisiones financieras.

Este análisis refuerza la confiabilidad del modelo híbrido y contribuye a una implementación ética, responsable y sostenible en el contexto de la banca empresarial hondureña.

6.6 CRONOGRAMA DE IMPLEMENTACIÓN Y PRESUPUESTO

La estimación del tiempo requerido para la implementación del modelo híbrido de riesgo crediticio se basó en la aplicación de juicio experto, considerando la experiencia profesional de los

autores en proyectos de analítica de datos y gestión del riesgo crediticio en el sector financiero, así como la revisión de prácticas comunes en iniciativas similares.

Para la definición de los tiempos de cada fase del proyecto se tomaron en cuenta criterios como complejidad técnica de las actividades, madurez de la arquitectura de datos del Banco x, experiencia del personal involucrado, dependencias entre actividades, especialmente aquellas relacionadas con validación, ajustes y aprobación interna, buenas prácticas de gestión de proyectos analíticos, donde se consideran tiempos adicionales para pruebas, ajustes y control de calidad.

A partir de estos criterios, se asignaron duraciones estimadas a cada etapa del proyecto, reflejadas en el cronograma de implementación. Dichas estimaciones representan tiempos razonables y conservadores para garantizar la correcta ejecución del modelo, permitiendo absorber contingencias operativas sin comprometer la calidad de los resultados.

Cabe señalar que el uso de juicio experto para la planificación temporal es una práctica común en proyectos de analítica avanzada y transformación digital, especialmente en entornos regulados como el sector financiero.

El cronograma de implementación se organiza por fases y contempla la incertidumbre temporal inherente a cada actividad, utilizando la técnica PERT de tres estimaciones:

- t_o : tiempo optimista.
- t_m : tiempo más probable.
- t_p : tiempo pesimista.

El tiempo esperado de cada actividad se calcula como:

$$t_e = \frac{t_o + 4t_m + t_p}{6}$$

La desviación estándar asociada se puede estimar como:

$$\sigma = \frac{t_p - t_o}{6}$$

En la siguiente tabla se presenta el cronograma por fases, con tiempos en semanas.

Tabla 58: Cronograma PERT de implementación del modelo híbrido

Fase	Actividad principal	Responsable líder	t_o (semanas)	t_m (semanas)	t_p (semanas)	t_e esperado (semanas)	Comentario
1	Preparación y gobierno inicial	Riesgos y Coordinación.	1	2	3	2.0	Definición de objetivos, KPIs, RACI, accesos y lineamientos.
2	Integración de datos y ETL	Data Engineer e IT.	3	4	6	4.2	Desarrollo de vistas analíticas, jobs incrementales y validación de calidad.
3	Preparación de datos y features	Analítica.	2	3	4	3.0	Capping, imputación, codificación, EDA final.
4	Modelado, tuning y validación (TRAIN, VALID, TEST).	Analítica y Riesgos.	3	4	6	4.2	Entrenamiento de Logística, RF, XGB; selección de modelo campeón.
5	Desarrollo de servicio de scoring (API/batch).	IT y Data Engineer.	2	3	5	3.2	Integración técnica con core y sistemas de riesgo.
6	6. Shadow mode y ajuste operativo.	Riesgos y Negocio.	8	10	12	10.0	3 meses de operación en paralelo, medición de KPIs y retroalimentación.
7	Go-live asistido (decisión asistida).	Riesgos, Negocio, IT.	2	3	4	3.0	Inicio de uso formal del modelo híbrido con política de overrides.
8	Monitoreo y mejora continua (primer semestre).	Riesgos y Analítica.	8	12	16	12.0	Seguimiento mensual, primera evaluación de recalibración o ajustes.

Fuente: Elaboración Propia

La suma de tiempos esperados t_e de las fases críticas que son del 1 al 7, da una duración total prevista de aproximadamente **29 - 30 semanas** para completar la implementación y alcanzar el inicio de la operación asistida.

La fase 8 representa el primer semestre de monitoreo intensivo posterior al Go-live.

6.7 PRESUPUESTO E IMPACTO DEL PRESUPUESTO

El presupuesto de la propuesta contempla tres grandes rubros: **recursos humanos especializados, infraestructura tecnológica y gestión del cambio y validación independiente.**

Dado que los costos pueden estar sujetos a variaciones (cotizaciones, tipo de cambio, alcance final), se modelan también mediante estimaciones PERT: costo optimista (C_o), costo más probable (C_m) y costo pesimista (C_p), calculando el costo esperado (C_e) con:

$$C_e = \frac{C_o + 4 C_m + C_p}{6}$$

6.7.1 PRESUPUESTO ESTIMADO POR RUBROS

Supuestos Base:

- Salarios Probables (C_m) en la banca hondureña:
 - Data Scientist / Analista de Datos senior: L55,000/mes.
 - Data Engineer: L45,000/mes.
 - Analista de Riesgo: L50,000/mes.
 - Líder de Proyecto: L55,000/mes.
 - Consultoría externa: paquete de revisión: L200,000.
 - Servidor/Cloud: L25,000/mes.
 - Talleres de capacitación: L80,000 el paquete.
 - Documentación regulatoria y manuales: L60,000.
 - Contingencia: L200,000.
- Dedicaciones:
 - Data Scientist: 6 meses al 100%.
 - Data Engineer: 6 meses al 50%.
 - Analista de Riesgo: 6 meses al 50%.
 - Líder de Proyecto: 6 meses al 30%.
- Técnica PERT para modelar incertidumbre en cada sub-rubros de costo:
 - Optimista: $C_o = 0.85 \times C_m$.
 - Probable: C_m (calculado con salarios y dedicación).

▫ Pesimista: $C_p = 1.15 \times C_m$.

▫ Costo esperado:

$$C_e = \frac{C_o + 4 C_m + C_p}{6} = C_m$$

▫ Tipo de cambio para mostrar equivalentes en USD:

$$C_e(USD) = \frac{C_e(HNL)}{26.50}$$

Tabla 59: Presupuesto PERT detallado por sub-rubros (valores en HNL y USD)

Rubro Principal	Sub-Rubros / Concepto	Supuestos Para C_m (más probable, en HNL)	C_o Optimista (HNL)	C_m Probable (HNL)	C_p Pesimista (HNL)	C_e esperado (HNL)	C_e esperado (USD)	Comentario Técnico
1. Personal especializado	1.1 Data Scientist / Analista de Datos senior	$55,000 \times 6 \text{ meses} \times 100\% = 330,000.00$	280,500.00	330,000.00	379,500.00	330,000.00	12,452.83	Perfil senior que diseña features, entrena y valida los modelos.
	1.2 Data Engineer / Ingeniero de Datos	$45,000 \times 6 \text{ meses} \times 50\% = 135,000.00$	114,750.00	135,000.00	155,250.00	135,000.00	5,094.34	Configura ETL, vistas analíticas y orquestación de datos.
1. Personal especializado	1.3 Analista de Riesgo de Crédito	$50,000 \times 6 \text{ meses} \times 50\% = 150,000.00$	127,500.00	150,000.00	172,500.00	150,000.00	5,660.38	Alinea el modelo con políticas internas, normativa CNBS y apetito de riesgo.
	1.4 Líder de Proyecto / Project Manager TI	$55,000 \times 6 \text{ meses} \times 30\% = 99,000.00$	84,150.00	99,000.00	113,850.00	99,000.00	3,735.85	Coordina cronograma, riesgos, entregables y comunicación interáreas.
	1.5 Validación independiente / consultoría externa	Paquete de revisión SR 11-7 y BCBS 239 = 200,000.00	170,000.00	200,000.00	230,000.00	200,000.00	7,547.17	Revisión por unidad independiente o consultor especializado.

Rubro Principal	Sub-Rubros / Concepto	Supuestos Para C_m (más probable, en HNL)	C_o Optimista (HNL)	C_m Probable (HNL)	C_p Pesimista (HNL)	C_e esperado (HNL)	C_e esperado (USD)	Comentario Técnico
Subtotal 1 - Personal						914,000.00	34,490.57	Suma esperada de 1.1 a 1.5.
2. Infraestructura y herramientas	2.1 Servidor/Cloud para scoring y almacenamiento	25,000 × 6 meses = 150,000.00	127,500.00	150,000.00	172,500.00	150,000.00	5,660.38	Servidor para ejecución de modelos, logs y almacenamiento de artefactos.
Subtotal 2 - Infraestructura						150,000.00	5,660.38	Costos tecnológicos directos.
3. Capacitación y gestión del cambio	3.1 Talleres a analistas y oficiales de crédito	Paquete de talleres, materiales, logística = 80,000.00	68,000.00	80,000.00	92,000.00	80,000.00	3,018.87	Formación en uso del score, tableros y política de overrides.
	3.2 Documentación regulatoria y manuales	Manual de gobierno, SOPs y plantillas de reporte = 60,000.00	51,000.00	60,000.00	69,000.00	60,000.00	2,264.15	Formaliza cumplimiento de SR 11-7, BCBS 239 y normativa CNBS.
Subtotal 3 - Capacitación y cambio						140,000.00	5,283.02	Suma de talleres + documentación.
4. Contingencias	4.1 Reserva de contingencia	Margen para cambios de alcance, inflación y ajustes regulatorios = 200,000.00	170,000.00	200,000.00	230,000.00	200,000.00	7,547.17	Cubre imprevistos de personal, tecnología o auditorías adicionales.
Subtotal 4 - Contingencias						200,000.00	7,547.17	
TOTAL PROYECTO						1,404,000.00	52,981.14	Costo esperado PERT (L1.404 millones, equivalente a USD 52,981.14).

Fuente: Elaboración Propia

El presupuesto presentado en la tabla 59 se ha elaborado a partir de bandas salariales y costos de mercado propios del contexto hondureño, expresados inicialmente en lempiras (L) y convertidos a dólares estadounidenses utilizando una tasa de cambio de L26.50 por USD1.00. Para cada sub-rubro se han considerado tres escenarios (optimista, probable y pesimista), aplicando la técnica PERT con un rango de $\pm 15\%$ alrededor del valor más probable.

De este modo, el costo esperado PERT del proyecto asciende a **L1,404,000.00**, equivalente a aproximadamente **USD52,981.14**.

6.7.2 ESTIMACIÓN DEL IMPACTO Y RETORNO DE LA INVERSIÓN (ROI)

Con base en los resultados empíricos obtenidos en el Capítulo IV, se observó que el modelo híbrido permite una reducción significativa de falsos negativos (FN) respecto al esquema tradicional basado exclusivamente en reglas.

Esta reducción se traduce en una disminución de pérdidas derivadas de créditos otorgados a clientes con alta probabilidad de incumplimiento.

Para cuantificar el impacto económico, se consideran tres escenarios de beneficio anual bruto por pérdidas evitadas, expresados en lempiras (L):

- Escenario optimista: $B_o = L18,000,000.00$.
- Escenario más probable: $B_m = L13,000,000.00$.
- Escenario pesimista: $B_p = L8,000,000.00$.

Aplicando la técnica PERT, el beneficio anual esperado se calcula como:

$$B_e = \frac{B_o + 4B_m + B_p}{6} = \frac{18,000,000 + 4(13,000,000) + 8,000,000}{6} = \frac{78,000,000}{6} = \mathbf{L13,000,000}$$

Es decir, el beneficio esperado coincide con el escenario más probable: L13,000,000.00 al año.

Por otra parte, el costo esperado del proyecto (sección 6.7.1), calculado también mediante PERT y ajustado a salarios y costos de mercado hondureños, asciende a:

$$C_e = \mathbf{L1,404,000.00}$$

Equivalente en dólares estadounidenses, usando una tasa de cambio de L26.50 por USD1.00:

$$C_e(USD) = \frac{1,404,000.00}{26.50} = \mathbf{USD52,981.14}$$

El Retorno de la Inversión (ROI) se define como:

$$ROI = \frac{B_e - C_e}{C_e}$$

Sustituyendo los valores:

$$ROI = \frac{13,000,000.00 - 1,404,000.00}{1,404,000.00} = \frac{11,596,000.00}{1,404,000.00} \approx 8.26$$

Expresado en porcentaje:

$$ROI_{\%} = 8.26 \times 100 \approx 825.95\%$$

Por lo tanto, el proyecto presenta un retorno esperado de aproximadamente **8.26 veces la inversión inicial**, es decir, un **ROI de 825.95%** en el primer año de operación plena del modelo híbrido, considerando únicamente el impacto por reducción de pérdidas crediticias (pérdidas evitadas por disminución de falsos negativos).

Si se expresan los beneficios en dólares estadounidenses, el beneficio anual esperado equivale a:

$$B_e(USD) = \frac{13,000,000.00}{26.50} = USD 490,566.04$$

y el ROI en USD es consistente con el cálculo anterior:

$$ROI = \frac{490,566.04 - 52,981.14}{52,981.14} \approx 8.26 \text{ (825.95\%)}$$

A este impacto cuantitativo se añaden beneficios cualitativos relevantes:

- Mejora en la calidad de las decisiones crediticias empresariales y en la asignación de capital.
- Incremento en la eficiencia operativa (reducción de tiempos de análisis y de respuesta al cliente).
- Fortalecimiento de la reputación institucional frente a reguladores y mercados por el uso de metodologías avanzadas y alineadas con BCBS 239 y SR 11-7.
- Mayor capacidad para diseñar políticas segmentadas de riesgo y productos personalizados, basados en una estimación más precisa de la probabilidad de incumplimiento.

En conjunto, estos elementos refuerzan la viabilidad económica, operativa y prudencial de la implementación del modelo híbrido propuesto en este capítulo.

6.8 CONCORDANCIA DE LOS SEGMENTOS DE LA TESIS CON LA PROPUESTA

La propuesta presentada en este Capítulo VI se construye como una extensión lógica y coherente de los hallazgos y conclusiones obtenidos en los capítulos anteriores. En la siguiente tabla se resume la alineación vertical entre los principales segmentos de la tesis y la propuesta de implementación del modelo híbrido.

Tabla 60: Concordancia de los Segmentos de Tesis

Capítulo I		Capítulo II	Capítulo III			Capítulo V	Capítulo VI		
Título Investigación	Objetivo General	Objetivos Específicos	Teorías / Metodologías de Sustento	Variables	Poblaciones	Técnicas	Conclusiones	Nombre de la Propuesta	Objetivos Propuesta
Evaluación de Modelo Híbrido de Calificación de Riesgo Crediticio para la Banca Empresarial en Honduras Periodo de enero 2020 - octubre 2024	<p>Evaluar (S) el impacto de un modelo híbrido que combina aprendizaje automático y reglas de negocio sobre la precisión en la evaluación del riesgo crediticio de clientes de la banca empresarial en Honduras, (M) midiendo métricas de desempeño como precisión, recall y AUC-ROC, (A) utilizando datos históricos y algoritmos supervisados,</p>	<p>(S) Comparar el (M) desempeño, medido por AUC-ROC y precisión, (A) entre un modelo que integra variables internas y externas y otro que utiliza solo variables internas, (R) para evaluarla relevancia de factores externos en la predicción de morosidad, (T) durante la fase de análisis exploratorio.</p> <p>Evaluar (S) la influencia del componente de aprendizaje automático dentro del modelo híbrido, (M) cuantificando métricas como</p>	<p>Teoría de la Información Asimétrica.</p> <p>Teoría del Aprendizaje Estadístico (AE).</p> <p>Teoría del Riesgo Financiero</p>	<p>Variables internas financieras y conductuales (buró, cobertura, montos, pagos, cuentas) como independientes y riesgo de incumplimiento PD (BUENO - MALO) como dependiente.</p>	<p>Base de 6,903 clientes empresariales de Banco X (2020 - octubre 2024).</p>	<p>EDA.</p> <p>Correlaciones.</p> <p>Matriz de confusión.</p> <p>Regresión Logística.</p> <p>Random Forest.</p> <p>XGBoost.</p> <p>PSI.</p> <p>Pruebas χ^2.</p> <p>Bootstrap de Diferencias.</p>	<p>1. El modelo híbrido (Random Forest y XGBoost integrados con reglas de negocio) superó de forma estadísticamente significativa al esquema tradicional de reglas y a la regresión logística, tanto en AUC-ROC como en KS, en los conjuntos de validación 2023 y prueba out-of-time 2024. Esto confirma que la combinación de variables internas con atributos de contexto (sector, zona, CNBS, buró)</p>	<p>Implementación de un Modelo Híbrido de Calificación de Riesgo Crediticio Empresarial en Instituciones Bancarias en Honduras</p>	<p>1. Integrar (S) y adaptar el modelo híbrido en las plataformas del área de riesgo crediticio de la institución, asegurando compatibilidad técnica y monitoreo de desempeño superior al 90% de integración funcional (M), utilizando el personal técnico actualmente asignado y la infraestructura existente (A); fortaleciendo el cumplimiento de los lineamientos de la CNBS y</p>

	<p>(R) para establecer un modelo validado que optimice la clasificación de riesgo, (T) durante la fase de análisis y modelado del estudio.</p>	<p>recall, F1-score y precisión, (A) aplicando algoritmos supervisados sobre los datos históricos, (R) para mejorar la capacidad de anticipar incumplimientos y riesgos emergentes, (T) durante la fase de modelado y validación.</p> <p>3. Determinar (S) el impacto de la implementación del modelo híbrido sobre la coherencia y consistencia de la evaluación crediticia, (M) midiendo la concordancia de los resultados simulados con registros históricos validados, (A) mediante simulaciones de escenarios crediticios, (R)</p>					<p>mejora de manera robusta la capacidad de discriminación del riesgo crediticio empresarial.</p> <p>2. La incorporación del componente de aprendizaje automático incrementó sustancialmente la capacidad de anticipar incumplimientos, elevando el recall de la clase morosa de 0.66 (reglas) a valores cercanos a 0.87 con Random Forest, manteniendo niveles de precisión aceptables. Esta mejora redujo significativamente los falsos negativos y se tradujo en pérdidas evitadas estimadas en alrededor de</p>		<p>Basilea III (R), dentro del período enero-marzo 2026 (T).</p> <p>2. Validar los resultados del modelo híbrido frente a los métodos tradicionales de calificación crediticia (S), midiendo métricas de desempeño (AUC-ROC, KS y Recall) (M) con el apoyo del equipo de analítica del área de riesgo, empleando datos históricos 2020-2024 (A), con el fin de optimizar la detección temprana de riesgo crediticio en el portafolio empresarial (R), a</p>
--	--	---	--	--	--	--	---	--	---

		para optimizar la calidad y confiabilidad del análisis financiero, (T) durante la fase de pruebas y simulación.					<p>L13 millones anuales, demostrando un impacto operativo y financiero material.</p> <p>3. El modelo híbrido mostró estabilidad temporal del score ($PSI \approx 0.0066$) y buena calibración de las probabilidades de incumplimiento ($Brier \approx 0.05-0.06$), lo que permite utilizarlo tanto para clasificación como para estimación de PD en provisiones y límites. La clasificación resultante fue coherente con el comportamiento histórico de la cartera y con las categorías CNBS, evidenciando</p>	<p>ejecutarse durante los meses de abril-junio de 2026 (T).</p> <p>3. Capacitar al personal especializado o en el uso, interpretación y mantenimiento del modelo híbrido (S), alcanzando la formación de 10 analistas de riesgo con nivel de satisfacción superior al 80 % (M), a través de talleres presenciales y sesiones técnicas dirigidas por el equipo desarrollador del modelo (A), fomentando una cultura organizacional basada en analítica</p>
--	--	---	--	--	--	--	--	---

							una evaluación crediticia más consistente y alineada con los criterios prudenciales vigentes.		avanzada y toma de decisiones sustentada en datos (R), al realizarse en el segundo trimestre de 2026 (abril-junio) (T).
--	--	--	--	--	--	--	---	--	---

Fuente: Elaboración Propia

REFERENCIAS BIBLIOGRÁFICAS

- Baesens, B., Roesch, D., & Scheule, H. (2016). *Credit risk analytics*. Wiley.
- Banco Central de Costa Rica. (2024). *Banco Central de Costa Rica (BCCR)*.
<https://www.bccr.fi.cr/SitePages/Inicio.aspx>
- Banco Central de Honduras. (2025). *Estabilidad Financiera Generalidades Estabilidad Financiera*.
<https://www.bch.hn/estadisticas-y-publicaciones-economicas/estabilidad-financiera/generalidades-de-estabilidad-financiera>
- Banco Central de Reserva de El Salvador (BCR)*. (2024). <https://www.bcr.gob.sv/>
- Banco Central do Brasil. (2024, diciembre 1). *BCB - Inflation Report*.
- Banco de Guatemala. (2024). *Banco de Guatemala*. <https://banguat.gob.gt/>
- Banco Interamericano de Desarrollo (BID). (2023). *Fintech en América Latina y el Caribe*.
<https://www.iadb.org/en>
- Basel Committee on Banking Supervision (Ed.). (2013). *Principles for effective risk data aggregation and risk reporting* (Jan. 2013). Bank for Internat. Settlements.
- Basel Committee on Banking Supervision (BCBS). (2013). *Principles for effective risk data aggregation and risk reporting*. <https://www.bis.org/publ/bcbs239.htm>
- Basel III: International regulatory framework for banks*. (2017). <https://www.bis.org/bcbs/basel3.htm>
- BBVA México. (2025). *¿Qué es el buró de crédito y cómo funciona?* [Carta].
<https://www.bbva.mx/educacion-financiera/creditos/que-es-el-buro-de-credito.html?msocid=10fb8f506da469ce3e379b4d6cb868b5>
- Biega, A. J., Potash, P., Daumé, H., Diaz, F., & Finck, M. (2020). Operationalizing the Legal Principle of Data Minimization for Personalization. *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '20*, 399–408.
<https://doi.org/10.1145/3397271.3401034>
- Brazil Trading Economics—BCB Selic Rate*. (2024, diciembre). Brazil Interest Rate - BCB Selic Rate.

<https://tradingeconomics.com/brazil/interest-rate>

Brown, K., & Moles, P. (2014). *Credit Risk Management*.

Castrillo, C. J. H. (2024). Paradigma Positivista. *Boletín Científico de las Ciencias Económico Administrativas del ICEA*, 12(24), 29–32. <https://doi.org/10.29057/icea.v12i24.12660>

CBRT Annual Report 2024. (2024, enero 1). <https://www.tcmb.gov.tr/>

Chang, V., Sivakulasingam, S., Wang, H., Wong, S. T., Ganatra, M. A., & Luo, J. (2024). Credit Risk Prediction Using Machine Learning and Deep Learning: A Study on Credit Card Customers. *Risks*, 12(11), 174. <https://doi.org/10.3390/risks12110174>

Chapman, P. (2000). *CRISP-DM 1.0: Step-by-step Data Mining Guide*. SPSS.

<https://books.google.hn/books?id=po7FtgAACAAJ>

CIRCULAR CNBS No.001/2025. (2025). [Carta].

<https://circulares.cnbs.gob.hn/Archivo/Viewer/4670/001->

[2025%20NORMA%20EVALUACION%20Y%20CLASIFICACION%20DE%20CARTERA.pdf](https://circulares.cnbs.gob.hn/Archivo/Viewer/4670/001-2025%20NORMA%20EVALUACION%20Y%20CLASIFICACION%20DE%20CARTERA.pdf)

Congreso Nacional de la República de Honduras. (2006). *LEY DE TRANSPARENCIA Y ACCESO A LA INFORMACION PÚBLICA*.

Contabilidad y Finanzas. (2024). Riesgo Crediticio: Gestión y Mitigación [Carta]. En *Contabilidad y Finanzas*. <https://contabilidadfinanzas.com/blog/riesgo-crediticio/>

Doctor Jacob Murel, E. K. (2024). *¿Qué es una matriz de confusión?* | IBM [Carta].

<https://www.ibm.com/es-es/topics/confusion-matrix>

Economía Focus/expectativas. (2025, febrero 1). Agência Brasil Focus/Expectativas.

<https://agenciabrasil.ebc.com.br/en/economia>

Estabilidad Financiera Informe de Estabilidad Financiera (IEF). (2025, octubre 25). [Carta].

[https://www.bch.hn/estadisticas-y-publicaciones-economicas/estabilidad-financiera/informe-de-estabilidad-financiera-\(ief\)](https://www.bch.hn/estadisticas-y-publicaciones-economicas/estabilidad-financiera/informe-de-estabilidad-financiera-(ief))

FasterCapital. (2025, marzo 17). *¿qué Es La Probabilidad De Incumplimiento (pd)*—FasterCapital [Carta]. <https://fastercapital.com/es/tema/%C2%BFqu%C3%A9-es-la-probabilidad-de->

incumplimiento-(pd).html

George A. Akerlof. (1970, enero 1). *The Market for Lemons*. chrome-extension://efaidnbmnnnibpcajpcglelefndmkaj/https://personal.utdallas.edu/~nina.baranchuk/Fin7310/papers/Akerlof1970.pdf

Gerencia de Estudios Económicos, CNBS 2024. (2024, abril 30). *INFORME DE COYUNTURA FINANCIERO ECONÓMICO*.

IBM. (2025). *¿Qué es el aprendizaje no supervisado?* [Carta]. <https://www.ibm.com/es-es/topics/unsupervised-learning>

Ignacio Perrotini H. (2002). *LA ECONOMIA DE LA INFORMACION ASIMETRICA*. Universidad Autonoma de Puebla.

Investment Climate Statement: South Africa. (2025, enero 1). *2025 Investment Climate Statements: South Africa*. <https://www.state.gov/reports/2025-investment-climate-statements/south-africa/>

ISO 27001: Standards and Best Practices. (2024, febrero 9). <https://auditboard.com/blog/iso-27001>

JMP Statistical Discovery. (2025). *Correlación* [Carta]. <https://www.jmp.com/es/statistics-knowledge-portal/what-is-correlation>

Josep. (2023). *Modelos de aprendizaje supervisado y no supervisado: ¿Qué son y cómo funcionan?* [Carta]. <https://conectandoideas.net/modelos-supervisados-y-no-supervisados/>

Karahan, F. (2024). *Fatih Karahan: Recent economic and financial developments in Turkey*. <https://www.bis.org/review/r2404261.htm>

Kumar Jain, T. (2019, abril 23). *Credit Risk Assessment*. <https://www.scribd.com/document/893843773/Credit-Risk-Assessment>

Lessmann, S., Baesens, B., Seow, H.-V., & Thomas, L. C. (2015). Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research. *European Journal of Operational Research*, 247(1), 124–136.

Lotero, R. S., & Molano, R. V. (2007, enero 1). *TEORÍA DEL RIESGO EN MERCADOS FINANCIEROS: UNA VISIÓN TEÓRICA*.

- Martini. (2025). <https://martini.ai/pages/research/BANTRAB%20HONDURAS-16d1512fdc4d8df0a458c07a2a9e87b4>
- McHugh, M. L. (2013). The Chi-square test of independence. *Biochemia Medica*, 143–149.
<https://doi.org/10.11613/BM.2013.018>
- Montgomery, D. C., Peck, E. A., & Vining, G. G. (1990). *Introduction to Linear Regression Analysis (Wiley Series in Probability and Statistics)* (3rd ed.).
- Nguyen, Q. G., Nguyen, L. H., Hosen, M. M., Rasel, M., Shorna, J. F., Mia, M. S., & Khan, S. I. (2025). Enhancing Credit Risk Management with Machine Learning: A Comparative Study of Predictive Models for Credit Default Prediction. *The American Journal of Applied sciences*, 07(01), 21–30.
<https://doi.org/10.37547/tajas/Volume07Issue01-04>
- Olivas Soria, E., Sanchez-Montañas Isla, M. A., Gamero Cruz, R., Castillo Caballero, B., & Cano Michelena, P. (2023). *Sistemas de Aprendizaje Automático*.
- Porter, M. E. (2008). The five competitive forces that shape strategy. *Harvard business review*, 86(1), 78.
- Post-positivismo | Research Starters | EBSCO Research*. (2023). EBSCO. <https://www.ebsco.com>
- Regulation—2016/679—EN - gdpr—EUR-Lex*. (2016). <https://eur-lex.europa.eu/eli/reg/2016/679/oj/eng>
- Riesgo Financiero Segun Autores: Información Completa, Definición, Ejemplos y Más*. (2024). conceptos.es. <https://conceptos.es/riesgo-financiero-segun-autores>
- Rob J Hyndman and George Athanasopoulos. (2018). *Forecasting: Principles & Practice (2nd ed)* [Carta]. <https://otexts.com/fpp2/>
- Roberto Hernández Sampieri, Carlos Fernández Collado, & Pilar Baptista Lucio. (2014). *Metodología de la investigación*. McGraw Hill España. <https://dialnet.unirioja.es/servlet/libro?codigo=775008>
- Science, S. (2024, julio 19). *Teoría del Aprendizaje Estadístico*. Simple Science.
<https://scisimple.com/es/topics/teoria-del-aprendizaje-estadistico--o3drz5w>
- SIB Guatemala. (2024). *Informe de estabilidad financiera*. <https://www.sib.gob.gt/web/sib/inicio>
- Solunion Colombia. (2024, junio 12). *Calificación crediticia: Un aliado para el crecimiento y la estabilidad financiera—Solunion Colombia* [Carta]. <https://www.solunion.co/blog/calificacion->

crediticia-un-aliado-para-el-crecimiento-y-la-estabilidad-financiera/
Sound practices for backtesting counterparty credit risk models. (2010).
<https://www.bis.org/publ/bcbs171.htm>

Sound Practices for Model Risk Management: Supervisory Guidance on Model Risk Management. (2011, abril 4). OCC.Gov. <https://www.occ.gov/news-issuances/bulletins/2011/bulletin-2011-12.html>

South African Reserve Bank. (2024). *South African Reserve Bank Financial Stability Review. Second edition 2024*.

South African Reserve Bank. (2025). *South African Reserve Bank Financial Stability Review. Second Edition 2025*.

SUGEF Costa Rica. (2024). *Normativa prudencial*. <https://www.sugef.fi.cr/index.aspx>

Tapia, G. N. (2022). *Sobre las calificadoras de riesgo*.

The FATF Recommendations. (2012, febrero 16). <https://www.fatf-gafi.org/en/publications/Fatfrecommendations/Fatf-recommendations.html>

The Fed—Supervisory Letter SR 11-7 on guidance on Model Risk Management. (2011, abril 4).
<https://www.federalreserve.gov/supervisionreg/srletters/sr1107.htm>

Thomas, L. C., Edelman, D. B., & Crook, J. N. (2017). *Credit scoring and its applications* (Second edition). SIAM Society for Industrial and Applied Mathematics.

Tomé, E. (2019, enero 1). *Estudio Centroamericano de Protección de Datos*.

Villamizar, C. (2022, enero 25). COBIT (Control Objectives for Information and Related Technology). *GlobalSuite Solutions*. <https://www.globalsuitesolutions.com/es/que-es-cobit/>

World Bank. (2019). *Credit Scoring Approaches: Guidelines*. <https://www.worldbank.org/ext/en/home>

World Bank Group. (2019). *CREDIT SCORING APPROACHES GUIDELINES*. pdf. www.worldbank.org

Yang, T., Li, A., Xu, J., Su, G., & Wang, J. (2024). *Deep Learning Model-Driven Financial Risk Prediction and Analysis*. <https://doi.org/10.20944/preprints202406.2069.v1>

ANEXOS

ANEXO TÉCNICO

ANEXO TÉCNICO A. CÓDIGO PYTHON DE DESEMPEÑO COMPARADO POR MODELO Y PERÍODO (VALID/TEST)

```
[43]
✓ Os
import os

os.makedirs("outputs/4_3_1_4", exist_ok=True)

# Tabla de métricas por modelo/split
def build_metrics_table_named(results, split="test"):
    rows = []
    for r in results:
        name = r["name"]
        s = r[split]
        rows.append({
            "Modelo": name,
            "Split": split.upper(),
            "AUC_ROC": float(s["auc"]) if "auc" in s else np.nan,
            "KS": float(s["ks"]) if "ks" in s else np.nan,
            "Precision@ThrKS": float(s["oper"]["precision"]) if "oper" in s else np.nan,
            "Recall@ThrKS": float(s["oper"]["recall"]) if "oper" in s else np.nan,
            "F1@ThrKS": float(s["oper"]["f1"]) if "oper" in s else np.nan,
            "Brier": float(s.get("brier", np.nan)),
        })
    return pd.DataFrame(rows)

tbl_valid = build_metrics_table_named([res_rules, res_logit, res_rf, res_xgb], split="valid")
tbl_test = build_metrics_table_named([res_rules, res_logit, res_rf, res_xgb], split="test")
tbl_metrics = pd.concat([tbl_valid, tbl_test], axis=0, ignore_index=True)

display(style_metrics(tbl_metrics, caption="Desempeño comparado por modelo y periodo (VALID/TEST)"))
tbl_metrics.to_csv("outputs/4_3_1_4/Tabla_4_3_1_4_A_metrics.csv", index=False)
```

... Desempeño comparado por modelo y periodo (VALID/TEST)

Modelo	Split	AUC_ROC	KS	Precision@ThrKS	Recall@ThrKS	F1@ThrKS	Brier
Reglas (proxy)	VALID	0.596	0.192	0.098	0.690	0.172	0.485
Logistic	VALID	0.742	0.342	0.170	0.556	0.260	0.130
RandomForest	VALID	0.777	0.409	0.150	0.738	0.249	0.059
XGBoost	VALID	0.786	0.468	0.169	0.762	0.277	0.059
Reglas (proxy)	TEST	0.590	0.181	0.117	0.660	0.199	0.467
Logistic	TEST	0.853	0.527	0.251	0.689	0.368	0.116
RandomForest	TEST	0.871	0.584	0.216	0.868	0.347	0.062
XGBoost	TEST	0.860	0.543	0.230	0.783	0.355	0.057

Figura 49: Código Python de Desempeño Comparado por Modelo y Período (VALID/TEST)

Fuente: Elaboración Propia

ANEXO TÉCNICO B. CÓDIGO PYTHON DE BOOTSTRAP Δ AUC Y Δ KS (VALID/TEST)

```
[42]
✓ 1m
# Bootstrap  $\Delta$ AUC y  $\Delta$ KS (VALID/TEST)
def ks_statistic(y_true, y_score):
    fpr, tpr, _ = roc_curve(y_true, y_score)
    return np.max(np.abs(tpr - fpr))

def bootstrap_diff(metric_func, y_true, pred_a, pred_b, n_boot=2000, seed=42):
    rng = np.random.default_rng(seed)
    n = len(y_true)
    idx = np.arange(n)
    diffs = []
    for _ in range(n_boot):
        samp = rng.choice(idx, size=n, replace=True)
        diffs.append(metric_func(y_true[samp], pred_a[samp]) - metric_func(y_true[samp], pred_b[samp]))
    diffs = np.array(diffs)
    ci_low, ci_high = np.percentile(diffs, [2.5, 97.5])
    p_value = 2 * min(np.mean(diffs <= 0), np.mean(diffs >= 0))
    return diffs.mean(), (ci_low, ci_high), p_value

def compare_pair(name_a, proba_a, name_b, proba_b, y_true):
    mean_auc, ci_auc, p_auc = bootstrap_diff(roc_auc_score, y_true, proba_a, proba_b, n_boot=2000, seed=42)
    def _ks(y, s): return ks_statistic(y, s)
    mean_ks, ci_ks, p_ks = bootstrap_diff(_ks, y_true, proba_a, proba_b, n_boot=2000, seed=42)
    return {
        "Comparacion": f"{name_a} vs {name_b}",
        "Delta_AUC_mean": mean_auc, "Delta_AUC_IC95%": ci_auc, "p_AUC": p_auc,
        "Delta_KS_mean": mean_ks, "Delta_KS_IC95%": ci_ks, "p_KS": p_ks
    }

pairs = [
    ("RandomForest", res_rf, "Reglas (proxy)", res_rules),
    ("XGBoost", res_xgb, "Reglas (proxy)", res_rules),
    ("RandomForest", res_rf, "Logistic", res_logit),
    ("XGBoost", res_xgb, "Logistic", res_logit),
]

rows_boot = []
for split, y_true in [("VALID", yva), ("TEST", yte)]:
    for a_name, a, b_name, b in pairs:
        res = compare_pair(a_name, a[split.lower()][:"proba"], b_name, b[split.lower()][:"proba"], y_true)
        res["Split"] = split
        rows_boot.append(res)

tbl_boot = pd.DataFrame(rows_boot)
tbl_boot["Delta_AUC_IC95%"] = tbl_boot["Delta_AUC_IC95%"].apply(lambda t: f"[{t[0]:.4f}, {t[1]:.4f}]")
tbl_boot["Delta_KS_IC95%"] = tbl_boot["Delta_KS_IC95%"].apply(lambda t: f"[{t[0]:.4f}, {t[1]:.4f}]")

display(style_bootstrap(tbl_boot, caption="Contraste (bootstrap) de  $\Delta$ AUC y  $\Delta$ KS por periodo"))
```

... Contraste (bootstrap) de Δ AUC y Δ KS por periodo

Comparacion	Split	Delta_AUC_mean	Delta_AUC_IC95%	p_AUC	Delta_KS_mean	Delta_KS_IC95%	p_KS
RandomForest vs Reglas (proxy)	VALID	0.181300	[0.1334, 0.2306]	0.000000	0.233200	[0.1468, 0.3270]	0.000000
XGBoost vs Reglas (proxy)	VALID	0.190400	[0.1367, 0.2435]	0.000000	0.283500	[0.1895, 0.3829]	0.000000
RandomForest vs Logistic	VALID	0.035100	[-0.0010, 0.0710]	0.060000	0.059600	[-0.0129, 0.1269]	0.100000
XGBoost vs Logistic	VALID	0.044300	[0.0119, 0.0775]	0.009000	0.109900	[0.0331, 0.1767]	0.005000
RandomForest vs Reglas (proxy)	TEST	0.281700	[0.2260, 0.3394]	0.000000	0.427300	[0.3154, 0.5453]	0.000000
XGBoost vs Reglas (proxy)	TEST	0.271900	[0.2172, 0.3282]	0.000000	0.389100	[0.2756, 0.5068]	0.000000
RandomForest vs Logistic	TEST	0.017400	[-0.0139, 0.0479]	0.272000	0.060100	[-0.0142, 0.1338]	0.112000
XGBoost vs Logistic	TEST	0.007600	[-0.0218, 0.0361]	0.576000	0.022000	[-0.0553, 0.0961]	0.548000

Figura 50: . Código Python de Bootstrap Δ AUC y Δ KS (VALID/TEST)

Fuente: Elaboración Propia

ANEXO TÉCNICO C. LOGISTIC COEFICIENTES (B) Y OR=EXP(B) EN PYTHON

```

# A) Coeficientes (β) y OR=exp(β)
import os
import numpy as np
import pandas as pd

def get_feature_names_from_ct(ct, num_features, cat_features):
    names = []
    for name, transformer, cols in ct.transformers_:
        if name == "num":
            names.extend(list(cols))
        elif name == "cat":
            ohe = transformer.named_steps.get("onehot")
            if hasattr(ohe, "get_feature_names_out"):
                names.extend(ohe.get_feature_names_out(cols).tolist())
            else:
                names.extend(list(cols))
    return names

os.makedirs("outputs/4_4_2", exist_ok=True)

pipe = pipe_logit # pipeline ya entrenado (TRAIN 2020-2022)
preproc = pipe.named_steps["prep"]
clf = pipe.named_steps["clf"]

feat_names = get_feature_names_from_ct(preproc, num_present, cat_present)
beta = clf.coef_.ravel()
beta0 = clf.intercept_[0]

df_coef = pd.DataFrame({
    "Variable": ["Intercepto"] + feat_names,
    "Beta": [beta0] + beta.tolist(),
})
df_coef["OR = exp(Beta)"] = np.exp(df_coef["Beta"])

# Ordenar por magnitud (opcional para mostrar top)
df_coef = df_coef.reindex([0] + list(1 + np.argsort(-np.abs(beta))))
df_coef.to_csv("outputs/4_4_2/Tabla_4_4_2_A_coeficientes_Logit.csv", index=False)
df_coef.head(20)

```

	Variable	Beta	OR = exp(Beta)
0	Intercepto	-1.081584	0.339058
160	DESC_ACTIVIDAD_ECONOMICA_COMPRA Y VENTA DE MER...	3.111336	22.451009
354	AGENCIA_ORIGEN_54.0	2.430513	11.364712
162	DESC_ACTIVIDAD_ECONOMICA_CONSTRUCTORA	2.419576	11.241088
59	DESTINO_81200	2.356593	10.554926
357	AGENCIA_ORIGEN_60.0	-2.177462	0.113329
9	AVG_CUENTAS_6M	-2.174044	0.113717
49	DESTINO_80200	1.995067	7.352695
33	DESTINO_11400	1.958254	7.086943
77	DESTINO_100600	1.849516	6.356741
51	DESTINO_80400	-1.786997	0.167462
347	AGENCIA_ORIGEN_28.0	-1.729882	0.177305
355	AGENCIA_ORIGEN_55.0	1.716076	5.562656
142	CATEGORIA_CNBS_1	-1.707298	0.181355
157	DESC_ACTIVIDAD_ECONOMICA_COMERCIANTE	1.688076	5.409064
104	DESTINO_130102	1.676523	5.346933
365	AGENCIA_ORIGEN_97.0	1.670849	5.316682
211	DESC_ACTIVIDAD_ECONOMICA_SILVICULTURA	1.652024	5.217529
105	DESTINO_130103	1.641465	5.162727
369	AGENCIA_ORIGEN_155.0	1.630899	5.108463

Figura 51: Logistic Coeficientes (β) y OR=exp(β) en Python

Fuente: Elaboración Propia

ANEXO TÉCNICO D. LOGISTIC HIPERPARÁMETROS Y EFICIENCIA COMPUTACIONAL EN PYTHON

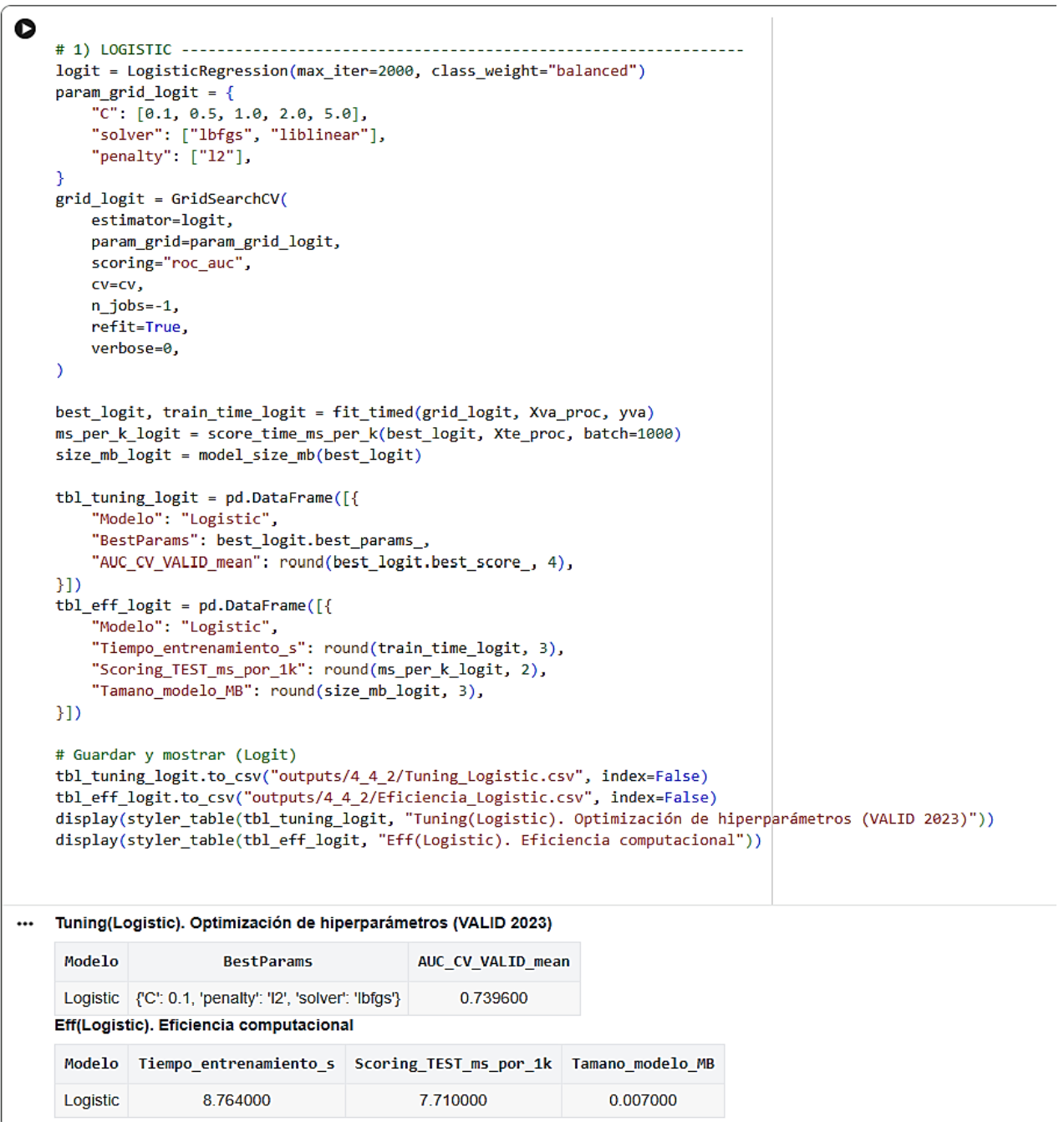


Figura 52: Logistic Hiperparámetros y Eficiencia Computacional en Python

Fuente: Elaboración Propia

ANEXO TÉCNICO E. PRUEBA DE CHI-CUADRADO (χ^2) APLICADA A VARIABLES CATEGÓRICAS EN PYTHON.

```

▶ from scipy.stats import chi2_contingency

# Nivel de significancia
alpha = 0.05
resultados = []

# Bucle para aplicar prueba chi-cuadrado
for var in variables_categoricas:
    data = df[[var, 'BUENO_MALO']].dropna()
    tabla = pd.crosstab(data[var], data['BUENO_MALO'])

    if tabla.shape[0] < 2 or tabla.shape[1] < 2:
        print(f"Variable {var}: no tiene categorías suficientes para prueba  $\chi^2$ ." )
        continue

    chi2, p, dof, expected = chi2_contingency(tabla)

    decision = (
        "Se rechaza H0 (hay asociación)"
        if p <= alpha
        else "No se rechaza H0 (no hay evidencia de asociación)"
    )

    resultados.append({
        "variable": var,
        "chi2": round(chi2, 2),
        "p_valor": p,
        "grados_libertad": dof,
        "decision": decision,
    })

# Convertir a DataFrame y ordenar
resultados_df = pd.DataFrame(resultados)
resultados_df = resultados_df.sort_values(by="p_valor", ascending=True)

# Mostrar resultados ordenados
print(resultados_df.to_string(index=False))
resultados_df.to_excel("resultados_chi2.xlsx", index=False)

```

variable	chi2	p_valor	grados_libertad	decision
CATEGORIA_CNBS	724.67	1.587065e-155	4	Se rechaza H ₀ (hay asociación)
TIPO_GARANTIA	99.97	1.598569e-17	9	Se rechaza H ₀ (hay asociación)
DESC_PROFESION	168.65	2.574075e-07	86	Se rechaza H ₀ (hay asociación)
ZONA	18.14	4.121606e-04	3	Se rechaza H ₀ (hay asociación)
DEUDOR	13.81	1.003321e-03	2	Se rechaza H ₀ (hay asociación)
TIPO_DE_PERSONA	7.60	5.852495e-03	1	Se rechaza H ₀ (hay asociación)
DESC_SECTOR	15.82	7.373698e-03	5	Se rechaza H ₀ (hay asociación)
AGENCIA_ORIGEN	134.15	1.079080e-02	99	Se rechaza H ₀ (hay asociación)
DESC_ACTIVIDAD_ECONOMICA	107.79	4.823350e-02	85	Se rechaza H ₀ (hay asociación)
DESC_OCUPACION	17.16	7.091024e-02	10	No se rechaza H ₀ (no hay evidencia de asociación)
SEXO	2.08	1.488570e-01	1	No se rechaza H ₀ (no hay evidencia de asociación)
NIVEL_EDUCATIVO	6.46	1.673090e-01	4	No se rechaza H ₀ (no hay evidencia de asociación)
ESTADO_CIVIL	6.24	1.818203e-01	4	No se rechaza H ₀ (no hay evidencia de asociación)
DESC_DESTINO	86.78	3.960454e-01	84	No se rechaza H ₀ (no hay evidencia de asociación)
TIPO_CREDITO	1.08	5.823926e-01	2	No se rechaza H ₀ (no hay evidencia de asociación)

Figura 53: Prueba de Chi-cuadrado (χ^2) Aplicada a Variables Categóricas en Python.

Fuente: Elaboración Propia

ANEXO TÉCNICO F. VALIDACIÓN DE SUPUESTOS ESTADÍSTICOS EN PYTHON

```
from scipy.stats import chi2_contingency

alpha = 0.05
validaciones = []

for var in variables_categoricas:
    data = df[[var, "BUENO_MALO"]].dropna()
    tabla = pd.crosstab(data[var], data["BUENO_MALO"])

    # Verificar que haya al menos dos categorías por variable
    if tabla.shape[0] < 2 or tabla.shape[1] < 2:
        continue

    chi2, p, dof, expected = chi2_contingency(tabla)

    # Verificación del supuesto de frecuencias esperadas
    min_expected = expected.min()
    cumple = min_expected >= 5

    validaciones.append({
        "variable": var,
        "min_frecuencia_esperada": round(min_expected, 2),
        "cumple_supuesto": "Sí" if cumple else "No",
        "p_valor": round(p, 4)
    })

validaciones_df = pd.DataFrame(validaciones)
print(validaciones_df)
validaciones_df.to_excel("validacion_supuesto_chi2.xlsx", index=False)
```

...	variable	min_frecuencia_esperada	cumple_supuesto	p_valor
0	TIPO_DE_PERSONA	212.97	Sí	0.0059
1	SEXO	108.80	Sí	0.1489
2	ESTADO_CIVIL	3.54	No	0.1818
3	NIVEL_EDUCATIVO	4.21	No	0.1673
4	TIPO_CREDITO	0.47	No	0.5824
5	TIPO_GARANTIA	0.14	No	0.0000
6	CATEGORIA_CNBS	0.67	No	0.0000
7	DEUDOR	46.78	Sí	0.0010
8	DESC_ACTIVIDAD_ECONOMICA	0.07	No	0.0482
9	DESC_PROFESION	0.07	No	0.0000
10	DESC_OCUPACION	0.07	No	0.0709
11	DESC_DESTINO	0.08	No	0.3960
12	DESC_SECTOR	1.47	No	0.0074
13	AGENCIA_ORIGEN	0.07	No	0.0108
14	ZONA	67.30	Sí	0.0004

Figura 54: Validación de Supuestos Estadísticos en Python

Fuente: Elaboración Propia

ANEXO TÉCNICO G. RANDOM FOREST HIPERPARÁMETROS Y EFICIENCIA COMPUTACIONAL EN PYTHON

```
# 2) RANDOM FOREST -----
rf = RandomForestClassifier(
    random_state=42,
    class_weight="balanced_subsample",
    n_jobs=-1
)
param_dist_rf = {
    "n_estimators": [200, 300, 400, 600],
    "max_depth": [None, 6, 8, 12],
    "max_features": ["sqrt", 0.3, 0.5, 0.7],
    "min_samples_split": [2, 5, 10],
    "min_samples_leaf": [1, 2, 5]
}
rand_rf = RandomizedSearchCV(
    estimator=rf,
    param_distributions=param_dist_rf,
    n_iter=50,
    scoring="roc_auc",
    cv=cv,
    random_state=42,
    n_jobs=-1,
    refit=True,
    verbose=0,
)
best_rf, train_time_rf = fit_timed(rand_rf, Xva_proc, yva)
ms_per_k_rf = score_time_ms_per_k(best_rf, Xte_proc, batch=1000)
size_mb_rf = model_size_mb(best_rf)

tbl_tuning_rf = pd.DataFrame([[
    "Modelo": "RandomForest",
    "BestParams": best_rf.best_params_,
    "AUC_CV_VALID_mean": round(best_rf.best_score_, 4),
]])
tbl_eff_rf = pd.DataFrame([[
    "Modelo": "RandomForest",
    "Tiempo_entrenamiento_s": round(train_time_rf, 3),
    "Scoring_TEST_ms_por_1k": round(ms_per_k_rf, 2),
    "Tamano_modelo_MB": round(size_mb_rf, 3),
]])

# Guardar y mostrar (RF)
tbl_tuning_rf.to_csv("outputs/4_4_2/Tuning_RandomForest.csv", index=False)
tbl_eff_rf.to_csv("outputs/4_4_2/Eficiencia_RandomForest.csv", index=False)
display(styler_table(tbl_tuning_rf, "Tabla 4.4.2-Tuning(RF). Optimización de hiperparámetros (VALID 2023)"))
display(styler_table(tbl_eff_rf, "Tabla 4.4.2-Eff(RF). Eficiencia computacional"))
```

... Tabla 4.4.2-Tuning(RF). Optimización de hiperparámetros (VALID 2023)

Modelo	BestParams	AUC_CV_VALID_mean
RandomForest	{'n_estimators': 600, 'min_samples_split': 5, 'min_samples_leaf': 5, 'max_features': 0.5, 'max_depth': 6}	0.781400

Tabla 4.4.2-Eff(RF). Eficiencia computacional

Modelo	Tiempo_entrenamiento_s	Scoring_TEST_ms_por_1k	Tamano_modelo_MB
RandomForest	766.751000	162.420000	2.987000

Figura 55: Random Forest Hiperparámetros y Eficiencia Computacional en Python

Fuente: Elaboración Propia

ANEXO TÉCNICO H. XGBOOST HIPERPARÁMETROS Y EFICIENCIA COMPUTACIONAL EN PYTHON

```

# 3) XGBOOST -----
xgb = XGBClassifier(
    objective="binary:logistic",
    eval_metric="logloss",
    random_state=42,
    n_jobs=-1,
)
param_dist_xgb = {
    "n_estimators": [200, 300, 500, 700],
    "max_depth": [3, 4, 5, 6],
    "learning_rate": [0.03, 0.05, 0.07],
    "subsample": [0.7, 0.8, 0.9, 1.0],
    "colsample_bytree": [0.6, 0.8, 0.9, 1.0],
    "reg_lambda": [0.0, 0.5, 1.0, 2.0],
}
rand_xgb = RandomizedSearchCV(
    estimator=xgb,
    param_distributions=param_dist_xgb,
    n_iter=50,
    scoring="roc_auc",
    cv=cv,
    random_state=42,
    n_jobs=-1,
    refit=True,
    verbose=0,
)
best_xgb, train_time_xgb = fit_timed(rand_xgb, Xva_proc, yva)
ms_per_k_xgb = score_time_ms_per_k(best_xgb, Xte_proc, batch=1000)
size_mb_xgb = model_size_mb(best_xgb)

tbl_tuning_xgb = pd.DataFrame([[
    "Modelo": "XGBoost",
    "BestParams": best_xgb.best_params_,
    "AUC_CV_VALID_mean": round(best_xgb.best_score_, 4),
]])
tbl_eff_xgb = pd.DataFrame([[
    "Modelo": "XGBoost",
    "Tiempo_entrenamiento_s": round(train_time_xgb, 3),
    "Scoring_TEST_ms_por_1k": round(ms_per_k_xgb, 2),
    "Tamano_modelo_MB": round(size_mb_xgb, 3),
]])

# Guardar y mostrar (XGB)
tbl_tuning_xgb.to_csv("outputs/4_4_2/Tuning_XGBoost.csv", index=False)
tbl_eff_xgb.to_csv("outputs/4_4_2/Eficiencia_XGBoost.csv", index=False)
display(styler_table(tbl_tuning_xgb, "Tabla 4.4.2-Tuning(XGB). Optimización de hiperparámetros (VALID 2023)"))
display(styler_table(tbl_eff_xgb, "Tabla 4.4.2-Eff(XGB). Eficiencia computacional"))

```

... **Tabla 4.4.2-Tuning(XGB). Optimización de hiperparámetros (VALID 2023)**

Modelo	BestParams	AUC_CV_VALID_mean
XGBoost	{'subsample': 0.9, 'reg_lambda': 1.0, 'n_estimators': 300, 'max_depth': 5, 'learning_rate': 0.03, 'colsample_bytree': 0.9}	0.782100

Tabla 4.4.2-Eff(XGB). Eficiencia computacional

Modelo	Tiempo_entrenamiento_s	Scoring_TEST_ms_por_1k	Tamano_modelo_MB
XGBoost	191.864000	9.260000	0.524000

Figura 56: XGBOOST Hiperparámetros y Eficiencia Computacional en Python

Fuente: Elaboración Propia

ANEXO TÉCNICO I. TOP 20 REAL VS PREDICHO EN PYTHON

```
# Top-20 Real vs Predicho
import os
import numpy as np
import pandas as pd

def top20_real_vs_pred(res_model, y_true, ids=None, k=20, split="test"):
    """
    Construye Top-k comparando Real vs Predicho (binario) y probabilidad.
    - res_model: diccionario del modelo (ej. res_logit, res_rf, res_xgb)
    - y_true: vector real (TEST)
    - ids: serie/array con identificador de fila/cliente (opcional)
    - k: cantidad de filas a mostrar (default 20)
    - split: 'test' (por defecto)
    """
    thr = res_model["valid"]["threshold"]
    proba = res_model[split]["proba"]
    yhat = (proba >= thr).astype(int)

    df = pd.DataFrame({
        "id": np.arange(len(y_true)) if ids is None else ids,
        "Real": y_true,
        "Proba": proba,
        "Predicho": yhat,
    })
    df["Umbral_VALID"] = thr
    # ordenar por casos con mayor probabilidad
    df_top = df.sort_values("Proba", ascending=False).head(k).reset_index(drop=True)
    return df_top

# Ejemplo para los tres modelos (usa la columna 'ID' si está en test_df)
os.makedirs("outputs/4_4_2", exist_ok=True)
ids = test_df["ID"] if "ID" in test_df.columns else None

for res in [res_logit, res_rf, res_xgb]:
    name = res["name"].replace(" ", "_")
    top20 = top20_real_vs_pred(res, yte, ids=ids, k=20, split="test")
    top20.to_csv(f"outputs/4_4_2/Top20_Real_vs_Pred_{name}.csv", index=False)
    print(f"\nTop-20 Real vs Predicho - {res['name']} (TEST, umbral VALID)")
    display(top20)

...
Top-20 Real vs Predicho - Logistic (TEST, umbral VALID)

```

	id	Real	Proba	Predicho	Umbral_VALID
0	1178481	0	0.999456	1	0.429194
1	5255384	0	0.998780	1	0.429194
2	5261559	1	0.997628	1	0.429194
3	1537944	1	0.995727	1	0.429194
4	2128476	1	0.993862	1	0.429194
5	98014	0	0.990884	1	0.429194
6	1117620	1	0.990802	1	0.429194
7	6369807	1	0.990753	1	0.429194
8	2783397	1	0.990660	1	0.429194

Figura 57: Top 20 Real vs Predicho en Python

Fuente: Elaboración Propia

ANEXO TÉCNICO J. COMPOSICIÓN DEL DATASET EN PYTHON

```
# Composición del Dataset

def ficha_dataset(df_raw: pd.DataFrame) -> pd.DataFrame:
    """Construye una ficha-resumen del dataset: filas, columnas y conteo por tipo."""
    n_obs = len(df_raw)
    n_cols = df_raw.shape[1]

    num_cnt = df_raw.select_dtypes(include=["int64", "float64", "int32", "float32"]).shape[1]
    cat_cnt = df_raw.select_dtypes(include=["object", "category"]).shape[1]
    dt_cnt = df_raw.select_dtypes(include=["datetime64"]).shape[1]

    resumen = pd.DataFrame(
        {
            "Métrica": [
                "N° de Observaciones",
                "N° de Variables (Total)",
                "Variables Numéricas (int/float)",
                "Variables Categóricas (object/category)",
                "Variables de Fecha (datetime)"
            ],
            "Valor": [
                f"{n_obs:,}".replace(",", "."),
                f"{n_cols:,}".replace(",", "."),
                f"{num_cnt:,}".replace(",", "."),
                f"{cat_cnt:,}".replace(",", "."),
                f"{dt_cnt:,}".replace(",", ".")
            ]
        }
    )
    return resumen

# Ejemplo de uso (df_raw es tu DataFrame inicial):
tabla_ficha = ficha_dataset(df_raw)
display(tabla_ficha.style.hide(axis="index"))

# Exportar a Excel (opcional)
tabla_ficha.to_excel("Ficha_Dataset_Inicial.xlsx", index=False)
print("Exportado: Ficha_Dataset_Inicial.xlsx")
```

Métrica	Valor
N° de Observaciones	6.903
N° de Variables (Total)	42
Variables Numéricas (int/float)	27
Variables Categóricas (object/category)	12
Variables de Fecha (datetime)	3

Exportado: Ficha_Dataset_Inicial.xlsx

Figura 58: Composición del Dataset en Python

Fuente: Elaboración Propia

ANEXO TÉCNICO K. VALORES ATÍPICOS DETECTADOS EN PYTHON

```

import pandas as pd
import numpy as np

def tabla_outliers_iqr(df: pd.DataFrame, num_vars: list, k: float = 1.5) -> pd.DataFrame:
    """
    Genera una tabla formal con detección de outliers por IQR.
    - df: DataFrame de entrada
    - num_vars: lista de columnas numéricas a evaluar
    - k: multiplicador de IQR (1.5 por defecto; usar 3.0 para extremos)
    """
    registros = len(df)
    filas = []

    for var in num_vars:
        if var not in df.columns:
            continue
        s = pd.to_numeric(df[var], errors="coerce")
        q1 = s.quantile(0.25)
        q3 = s.quantile(0.75)
        iqr = q3 - q1
        lower = q1 - k * iqr
        upper = q3 + k * iqr
        mask_out = (s < lower) | (s > upper)
        n_out = int(mask_out.sum())
        pct_out = 100.0 * n_out / registros

        filas.append({
            "Variable": var,
            "Q1": q1,
            "Q3": q3,
            "IQR": iqr,
            "Limite Inferior": lower,
            "Limite Superior": upper,
            "Outliers (n)": n_out,
            "Outliers (%)": pct_out
        })

    tabla = pd.DataFrame(filas)
    # Ordenar por mayor % de outliers
    tabla = tabla.sort_values(by="Outliers (%)", ascending=False).reset_index(drop=True)
    return tabla

# Variables numéricas a evaluar
num_vars = [
    "MONTO_DESEMBOLSADO", "MONTO_GARANTIA", "MONTO_EXPOSICION_TOTAL",
    "COBERTURA_GARANTIA", "EDAD", "BURO_SCORE_EQUIFAX", "AVG_PAGOS_6M", "AVG_CUENTAS_6M"
]

def resaltar_outliers(col):
    return ["background-color:#ffe5e5;color:#900" if v > 5 else "" for v in col]

display(
    tabla_out.style
    .apply(resaltar_outliers, subset=["Outliers (%)"])
    .format({
        "Q1": "{:,.2f}", "Q3": "{:,.2f}", "IQR": "{:,.2f}",
        "Limite Inferior": "{:,.2f}", "Limite Superior": "{:,.2f}",
        "Outliers (n)": "{:,.0f}", "Outliers (%)": "{:,.2f}"
    })
)

```

	Variable	Q1	Q3	IQR	Límite Inferior	Límite Superior	Outliers (n)	Outliers (%)
0	BURO_SCORE_EQUIFAX	696.25	880.00	183.75	420.62	1,155.62	1,107	16.04
1	MONTO_GARANTIA	0.00	2,800,000.00	2,800,000.00	-4,200,000.00	7,000,000.00	1,102	15.96
2	AVG_PAGOS_6M	24,708.82	315,379.29	290,670.48	-411,296.90	751,385.01	1,010	14.63
3	AVG_CUENTAS_6M	48,000.43	1,021,780.26	973,779.83	-1,412,669.31	2,482,450.01	819	11.86
4	MONTO_DESEMBOLSADO	600,000.00	3,100,000.00	2,500,000.00	-3,150,000.00	6,850,000.00	809	11.72
5	MONTO_EXPOSICION_TOTAL	604,625.01	3,268,258.99	2,663,633.98	-3,390,825.97	7,263,709.96	793	11.49
6	EDAD	38.00	54.00	16.00	14.00	78.00	43	0.62
7	COBERTURA_GARANTIA	0.00	1.00	1.00	-1.50	2.50	0	0.00

Figura 59: Valores Atípicos Detectados en Python

Fuente: Elaboración Propia

ANEXO TÉCNICO L. EFICACIA IQR ANTES VS DESPUÉS EN PYTHON

```
# Chequeo de eficacia (IQR antes vs. después)

def outliers_iqr_pct(df_in, cols, k=1.5):
    n = len(df_in); rows=[]
    for c in cols:
        s = pd.to_numeric(df_in[c], errors="coerce")
        q1,q3=s.quantile(0.25), s.quantile(0.75)
        iqr=q3-q1; lo=q1-k*iqr; hi=q3+k*iqr
        pct = 100*((s<lo)|(s>hi)).sum()/n
        rows.append({"Variable": c, "Outliers_IQR_%": pct})
    return pd.DataFrame(rows)

cols_eval = ["MONTO_DESEMBOLSADO","MONTO_GARANTIA","MONTO_EXPOSICION_TOTAL",
             #"BURO_SCORE_EQUIFAX",
             "COBERTURA_GARANTIA","EDAD","AVG_PAGOS_6M","AVG_CUENTAS_6M"]

before = outliers_iqr_pct(df, cols_eval)
after = outliers_iqr_pct(df_c, cols_eval)

comp = (before.merge(after, on="Variable", suffixes=(" (Antes)"," (Después)"))
        .assign(Reducción_pp=lambda d: d["Outliers_IQR_% (Antes)"]-d["Outliers_IQR_% (Después)"])
        .sort_values("Reducción_pp", ascending=False))
display(comp.style.format({"Outliers_IQR_% (Antes)": "{:.2f}", "Outliers_IQR_% (Después)": "{:.2f}", "Reducción_pp": "{:.2f}"}))
```

	Variable	Outliers_IQR_% (Antes)	Outliers_IQR_% (Después)	Reducción_pp
1	MONTO_GARANTIA	15.96	0.00	15.96
5	AVG_PAGOS_6M	14.63	0.00	14.63
6	AVG_CUENTAS_6M	11.86	0.00	11.86
0	MONTO_DESEMBOLSADO	11.72	0.00	11.72
2	MONTO_EXPOSICION_TOTAL	11.49	0.00	11.49
4	EDAD	0.62	0.00	0.62
3	COBERTURA_GARANTIA	0.00	0.00	0.00

Figura 60: Eficacia IQR Antes vs Después en Python

Fuente: Elaboración Propia

ANEXO TÉCNICO M. DIAGNÓSTICO DE VALORES FALTANTES EN PYTHON

```
# Utilidades: resumen y faltantes

import pandas as pd
import numpy as np

def tabla_faltantes(df: pd.DataFrame) -> pd.DataFrame:
    n = len(df)
    t = pd.DataFrame({
        "Columna": df.columns,
        "N_faltantes": df.isna().sum().values,
        "%_faltantes": (df.isna().mean().values * 100).round(2),
        "Tipo": [str(df[c].dtype) for c in df.columns],
    })
    t = t.sort_values("%_faltantes", ascending=False).reset_index(drop=True)
    return t

def resumen_estadistico(df: pd.DataFrame, cols: list, titulo: str) -> pd.DataFrame:
    filas = []
    for c in cols:
        if c not in df.columns:
            continue
        s = pd.to_numeric(df[c], errors="coerce")
        filas.append({
            "Variable": c,
            "N válidos": int(s.notna().sum()),
            "Media": s.mean(),
            "Mediana": s.median(),
            "Desv.Est.": s.std(),
            "Mín": s.min(),
            "Máx": s.max(),
            "% Nulos": s.isna().mean() * 100
        })
    out = pd.DataFrame(filas)
    out["Sección"] = titulo
    return out
```

```
# Diagnóstico de faltantes (antes)

falt_antes = tabla_faltantes(df)
display(falt_antes.style.format({"%_faltantes": "{:.2f}"}))
```

	Columna	N_faltantes	%_faltantes	Tipo
0	DESC_DESTINO	3526	51.08	object
1	CODIGO_DESTINO	3394	49.17	float64
2	BURO_SCORE_EQUIFAX	1313	19.02	float64
3	DESC_PROFESION	483	7.00	object
4	CODIGO_PROFESION	483	7.00	float64

Figura 61: Diagnóstico de Valores Faltantes en Python

Fuente: Elaboración Propia

ANEXO TÉCNICO N. VARIABLES NORMALIZADAS Y ESTANDARIZADAS EN PYTHON

```
# base final tratada df_final (sin outliers y sin faltantes) y la partición temporal en train/valid/test.

import pandas as pd
from sklearn.preprocessing import StandardScaler
from sklearn.compose import ColumnTransformer
from sklearn.pipeline import Pipeline
import numpy as np

# 1) Partición temporal ya definida (ejemplo)
df_final["ANIO"] = pd.to_datetime(df_final["FECHA_APERTURA"], errors="coerce").dt.year
train = df_final[df_final["ANIO"].between(2020, 2022)].copy()
valid = df_final[df_final["ANIO"] == 2023].copy()
test = df_final[df_final["ANIO"] == 2024].copy()

# 2) Definición de variables
num_scale = [
    "MONTO_DESEMBOLSADO", "MONTO_GARANTIA", "MONTO_EXPOSICION_TOTAL",
    "AVG_PAGOS_6M", "AVG_CUENTAS_6M", "BURO_SCORE_EQUIFAX", "EDAD"
]
# Proporción que se conserva en [0,1]
num_keep = ["COBERTURA_GARANTIA"]

# 3) Pipeline de estandarización (solo z-score para num_scale)
scaler = StandardScaler()

def fit_scaler(train_df, cols):
    scaler = StandardScaler()
    scaler.fit(train_df[cols])
    params = {"mean_": dict(zip(cols, scaler.mean_)),
              "scale_": dict(zip(cols, scaler.scale_))}
    return scaler, params

scaler_z, scaler_params = fit_scaler(train, num_scale)

def apply_scaler(df_in, scaler, cols_scale, cols_keep):
    Xs = pd.DataFrame(
        scaler.transform(df_in[cols_scale]),
        columns=[f"z_{c}" for c in cols_scale],
        index=df_in.index,
    )
    Xk = df_in[cols_keep].copy() if cols_keep else pd.DataFrame(index=df_in.index)
    return pd.concat([Xs, Xk], axis=1)

X_train_z = apply_scaler(train, scaler_z, num_scale, num_keep)
X_valid_z = apply_scaler(valid, scaler_z, num_scale, num_keep)
X_test_z = apply_scaler(test, scaler_z, num_scale, num_keep)

# 4) Reporte de parámetros (gobernanza)
scaler_params
```

Figura 62: Variables Normalizadas y Estandarizadas en Python

Fuente: Elaboración Propia

ANEXO TÉCNICO O. ANÁLISIS DESCRIPTIVO DE VARIABLES CUANTITATIVAS EN PYTHON

```
# Resumen estadístico comparativo para numéricas
res_antes = resumen_estadistico(df, num_cols, "Antes")
res_desp = resumen_estadistico(df_imp, num_cols, "Después")

comp = (res_antes.drop(columns=["Sección"])
        .merge(res_desp.drop(columns=["Sección"]), on="Variable", suffixes=(" (Antes)", " (Después)"))
comp = comp[[
    "Variable",
    "N válidos (Antes)", "N válidos (Después)",
    "Media (Antes)", "Media (Después)",
    "Mediana (Antes)", "Mediana (Después)",
    "Desv.Est. (Antes)", "Desv.Est. (Después)",
    "Mín (Antes)", "Mín (Después)",
    "Máx (Antes)", "Máx (Después)",
    "% Nulos (Antes)", "% Nulos (Después)"
]]

display(comp.style.format({
    "N válidos (Antes)": "{:, .0f}", "N válidos (Después)": "{:, .0f}",
    "Media (Antes)": "{:, .2f}", "Media (Después)": "{:, .2f}",
    "Mediana (Antes)": "{:, .2f}", "Mediana (Después)": "{:, .2f}",
    "Desv.Est. (Antes)": "{:, .2f}", "Desv.Est. (Después)": "{:, .2f}",
    "Mín (Antes)": "{:, .2f}", "Mín (Después)": "{:, .2f}",
    "Máx (Antes)": "{:, .2f}", "Máx (Después)": "{:, .2f}",
    "% Nulos (Antes)": "{:, .2f}", "% Nulos (Después)": "{:, .2f}"
}))
```

Figura 63: Análisis Descriptivo de Variables Cuantitativas en Python

Fuente: Elaboración Propia