Clasificación del Riesgo de Inversión en Acciones de la Bolsa Mexicana de Valores mediante Redes Kolmogorov-Arnold

Universidad Autónoma de Querétaro

Facultad de Ingeniería

Clasificación del Riesgo de Inversión en Acciones de la Bolsa Mexicana de Valores mediante Redes Kolmogorov-Arnold

Tesis

Que como parte de los requisitos para obtener el Grado de

Maestro en Ciencias (Ingeniería Matemática)

Presenta

Lic. Omar Nayib García Sabag

Dirigido por:
Dr. Victor Antonio Aguilar Arteaga

Co-Director:
Dr. Gilberto Reynoso Meza

Querétaro, Qro. a 3 de noviembre de 2026

REPOSITORIO INSTITUCIONAL DGBSDI-UAQ

La presente obra está bajo la licencia:
https://creativecommons.org/licenses/by-nc-nd/4.0/deed.es

CC BY-NC-ND 4.0 DEED

Atribución-NoComercial-SinDerivadas 4.0 Internacional

Usted es libre de:

Compartir — copiar y redistribuir el material en cualquier medio o formato

La licenciante no puede revocar estas libertades en tanto usted siga los términos de la licencia

Bajo los siguientes términos:

Atribución — Usted debe dar crédito de manera adecuada , brindar un enlace a la licencia, e indicar si se han realizado cambios . Puede hacerlo en cualquier forma razonable, pero no de forma tal que sugiera que usted o su uso tienen el apoyo de la licenciante.

NoComercial — Usted no puede hacer uso del material con propósitos comerciales .

SinDerivadas — Si remezcla, transforma o crea a partir del material, no podrá distribuir el material modificado.

No hay restricciones adicionales — No puede aplicar términos legales ni medidas tecnológicas que restrinjan legalmente a otras a hacer cualquier uso permitido por la licencia.

Avisos:

No tiene que cumplir con la licencia para elementos del material en el dominio público o cuando su uso esté permitido por una excepción o limitación aplicable.

No se dan garantías. La licencia podría no darle todos los permisos que necesita para el uso que tenga previsto. Por ejemplo, otros derechos como publicidad, privacidad, o derechos morales pueden limitar la forma en que utilice el material.

Universidad Autónoma de Querétaro

Facultad de Ingeniería

Maestría en Ciencias (Ingeniería Matemática)

Clasificación del Riesgo de Inversión en Acciones de la Bolsa Mexicana de Valores mediante Redes Kolmogorov-Arnold

Tesis

Que como parte de los requisitos para obtener el Grado de

Maestro en Ciencias (Ingeniería Matemática)

Presenta

Lic. Omar Nayib García Sabag

Dirigido por

Dr. Victor Antonio Aguilar Arteaga

Presidente - Dr. Victor Antonio Aguilar Arteaga
(Universidad Autónoma de Querétaro)

Secretario - Dr. Gilberto Reynoso Meza
(Pontificia Universidad Católica de Paraná)

Vocal - MC. Iván González García
(Universidad Autónoma de Querétaro)

Suplente - Dr. Francisco Gerardo Jiménez López
(Universidad Autónoma de Querétaro)

Suplente - Dr. José Alfonso Niembro Ceceña
(Universidad Autónoma de Querétaro)

Centro Universitario, Querétaro, Qro. México

A mis padres, Raúl y Nasime,
por su amor incondicional y por enseñarme
que todo esfuerzo vale la pena.

A mi hermano Rashid,
por caminar siempre a mi lado.

Este logro también es suyo.

Agradecimientos

A la Secretaría de Ciencia, Humanidades, Tecnología e Innovación (SECIHTI) por la beca otorgada para la realización de mis estudios de maestría (CVU: 1170274). Sin este apoyo, este trabajo no habría sido posible.

A la Universidad Autónoma de Querétaro y a la Facultad de Ingeniería, por abrirme sus puertas y brindarme la formación, los espacios y las oportunidades que hicieron posible esta etapa de mi vida.

A mi director, el Dr. Victor Antonio Aguilar Arteaga, y a mi co-director, el Dr. Gilberto Reynoso Meza, por su guía constante, su paciencia y su confianza. Gracias por proponer las Redes Kolmogorov-Arnold como eje de esta investigación y por acompañarme en cada paso, desde la primera idea hasta la última página. Su apoyo fue fundamental, tanto en lo académico como en lo humano.

A los miembros del sínodo, por el tiempo dedicado a revisar este trabajo y por sus valiosos comentarios, que contribuyeron a enriquecerlo y darle mayor solidez.

Y, sobre todo, a mi familia. A mis padres, Raúl y Nasime, por su amor, por sus sacrificios y por creer en mí incluso en los momentos en que yo dudaba. Todo lo que soy se lo debo a ustedes. A mi hermano Rashid, por su cariño, su compañía y por recordarme siempre lo que realmente importa. Gracias por ser mi hogar y mi impulso.

Abreviaturas y siglas

Sigla	Significado
Afore	Administradora de Fondos para el Retiro
API	Interfaz de Programación de Aplicaciones (*Application Programming Interface*)
ARCH	Heterocedasticidad Condicional Autorregresiva (*Autoregressive Conditional Heteroskedasticity*)
ARIMA	Modelo Autorregresivo Integrado de Media Móvil (*Autoregressive Integrated Moving Average*)
AUC	Área bajo la curva (*Area Under the Curve*)
Banxico	Banco de México
BIS	Banco de Pagos Internacionales (*Bank for International Settlements*)
BMV	Bolsa Mexicana de Valores
CAGR	Tasa de Crecimiento Anual Compuesta (*Compound Annual Growth Rate*)
CAPM	Modelo de Valoración de Activos de Capital (*Capital Asset Pricing Model*)
CESF	Consejo de Estabilidad del Sistema Financiero
CETES	Certificados de la Tesorería de la Federación
CNBV	Comisión Nacional Bancaria y de Valores
CNSF	Comisión Nacional de Seguros y Fianzas
CONSAR	Comisión Nacional del Sistema de Ahorro para el Retiro
CORAL	Logits de Rango Consistente (*COnsistent RAnk Logits*)
CORE-10	Conjunto de las diez variables predictoras empleadas en este trabajo
CPU	Unidad Central de Procesamiento (*Central Processing Unit*)
CVaR	Valor en Riesgo Condicional o *Expected Shortfall* (*Conditional Value at Risk*)
DCVaR	Diferencial del Valor en Riesgo Condicional
DL	Aprendizaje profundo (*Deep Learning*)
ECE	Error de Calibración Esperado (*Expected Calibration Error*)
FSB	Consejo de Estabilidad Financiera (*Financial Stability Board*)

Sigla	Significado
GARCH	Heterocedasticidad Condicional Autorregresiva Generalizada (*Generalized Autoregressive Conditional Heteroskedasticity*)
GPU	Unidad de Procesamiento Gráfico (*Graphics Processing Unit*)
IA	Inteligencia Artificial
IC	Intervalo de Confianza
IOSCO	Organización Internacional de Comisiones de Valores (*International Organization of Securities Commissions*)
IPC	Índice de Precios y Cotizaciones de la Bolsa Mexicana de Valores
IQR	Rango Intercuartílico (*Interquartile Range*)
KAN	Redes Kolmogorov-Arnold (*Kolmogorov-Arnold Networks*)
KART	Teorema de Representación de Kolmogorov-Arnold (*Kolmogorov-Arnold Representation Theorem*)
LIME	Explicaciones Locales Interpretables Agnósticas al Modelo (*Local Interpretable Model-agnostic Explanations*)
LSTM	Memoria de Corto y Largo Plazo (*Long Short-Term Memory*)
MaxDD	Máxima caída acumulada (*Maximum Drawdown*)
MCO	Mínimos Cuadrados Ordinarios
ML	Aprendizaje automático (*Machine Learning*)
MLP	Perceptrón Multicapa (*Multilayer Perceptron*)
MMK	Mezcla Multicapa de KAN (*Multi-layer Mixture-of-KAN*)
NSGA-III	Algoritmo Genético de Ordenamiento No Dominado III (*Non-dominated Sorting Genetic Algorithm III*)
OOS	Fuera de muestra (*Out-of-Sample*)
OvR	Uno contra el resto (*One-vs-Rest*)
PIB	Producto Interno Bruto
RAM	Memoria de Acceso Aleatorio (*Random Access Memory*)
REMEF	Revista Mexicana de Economía y Finanzas
REST	Transferencia de Estado Representacional (*Representational State Transfer*)
RF	Bosques Aleatorios (*Random Forest*)

Sigla	Significado
RiskIndex	Índice multivariado de riesgo construido en este trabajo a partir de la volatilidad futura, $\|\beta\|$, $\|\text{CVaR}\|$ y $1 - R^2$
RNA	Red Neuronal Artificial
ROC	Característica Operativa del Receptor (*Receiver Operating Characteristic*)
SAR	Sistema de Ahorro para el Retiro
SECIHTI	Secretaría de Ciencia, Humanidades, Tecnología e Innovación
SHAP	Explicaciones Aditivas de Shapley (*SHapley Additive exPlanations*)
SHCP	Secretaría de Hacienda y Crédito Público
SIE	Sistema de Información Económica del Banco de México
SIEFORE	Sociedad de Inversión Especializada de Fondos para el Retiro
SVM	Máquina de Vectores de Soporte (*Support Vector Machine*)
UAQ	Universidad Autónoma de Querétaro
UPAEP	Universidad Popular Autónoma del Estado de Puebla
VaR	Valor en Riesgo (*Value at Risk*)
Walk-Forward PRO	Protocolo de validación temporal con ventana expansiva, purga y embargo empleado en este trabajo
XAI	Inteligencia Artificial Explicable (*Explainable Artificial Intelligence*)

Resumen

Se desarrolla un sistema de **clasificación ordinal del riesgo accionario** (*Bajo / Medio / Alto*) para 30 emisoras de la BMV (2020–2025) basado en Redes Kolmogorov-Arnold (KAN). El riesgo se mide con un índice multivariado (`RiskIndex`: volatilidad futura, $|\beta|$, |CVaR|, $1-R^2$) y etiquetado anti *look-ahead* por pliegue. Bajo un protocolo **Walk-Forward PRO** con purga y embargo (13 pliegues, $n = 16{,}133$ obs. OOS), KAN alcanza F1 macro de 0.750 —equivalente a MLP— y supera a SVM en la clase transicional (recall Medio 65.3% vs. 46.6%, $p = 0.016$). Su interpretabilidad intrínseca —funciones $\varphi(x)$ inspeccionables— identifica volatilidad, CVaR y Beta como variables determinantes. El trabajo extiende el único antecedente directo en la literatura mexicana [34] en interpretabilidad estructural, validación temporal rigurosa y reproducibilidad con fuentes públicas.

Palabras clave: KAN, clasificación ordinal de riesgo, Walk-Forward, interpretabilidad intrínseca, BMV, RiskIndex.

Abstract

An **ordinal equity-risk classification system** (*Low / Medium / High*) is developed for 30 stocks on the Mexican Stock Exchange (2020–2025) using Kolmogorov-Arnold Networks (KAN). Risk is measured via a multivariate index (`RiskIndex`: future volatility, $|\beta|$, |CVaR|, $1-R^2$) with fold-specific anti-look-ahead labeling. Under a **Walk-Forward PRO** protocol with purging and embargo (13 folds, $n = 16{,}133$ OOS obs.), KAN achieves F1 macro of 0.750 —equivalent to MLP— and outperforms SVM in the transitional class (Medium recall 65.3% vs. 46.6%, $p = 0.016$). Its intrinsic interpretability —inspectable $\varphi(x)$ functions— identifies volatility, CVaR, and Beta as key drivers. This work extends the only direct antecedent in the Mexican literature [34] along three dimensions: structural interpretability, temporally honest validation, and full reproducibility from public data.

Keywords: KAN, ordinal risk classification, walk-forward, intrinsic interpretability, Mexican Stock Exchange, RiskIndex.

Índice

Índice de cuadros **V**

Índice de figuras **VI**

I Introducción **2**

Contribuciones principales del trabajo 7

II Antecedentes **10**

2.1 Modelos Tradicionales para Riesgo Financiero: Alcances y Limitaciones . 10

2.2 ML y DL en Finanzas: Poder Predictivo y Fragilidad Empírica 11

2.2.1 Panorama de aplicaciones globales 11

2.2.2 Aplicaciones en el contexto de la BMV 11

2.2.3 Limitaciones metodológicas recurrentes en ML financiero 13

2.3 Métodos de Ensamble: Desempeño Tabular Superior con Límites de Auditabilidad 14

2.4 Interpretabilidad en IA Financiera: *Post-hoc* vs. Intrínseca 15

2.4.1 XAI *post-hoc*: aportes y límites 15

2.4.2 Modelos interpretables por diseño: motivación en finanzas 15

2.5 Redes Kolmogorov-Arnold: Fundamento, Arquitectura y Evidencia . . . 16

2.5.1 Del Problema 13 de Hilbert al Teorema de Representación 16

2.5.2 Arquitectura KAN: interpretabilidad por construcción 16

2.5.3 Aplicaciones en finanzas y series temporales 17

2.6 Vacío en la Literatura y Posición del Estudio 17

2.6.1 El vacío específico: tres ausencias simultáneas 18

2.6.2 Posición de la investigación y originalidad 18

III Hipótesis y Objetivos **20**

3.1 Hipótesis de Investigación 20

3.2 Objetivos de Investigación 22

3.2.1 Objetivo General 22

3.2.2 Objetivos Específicos 22

IV Marco Teórico **25**

4.1 Fundamentos del Riesgo Financiero y su Medición 25

4.1.1 Concepto, Naturaleza y Tipologías del Riesgo Financiero 25
4.1.2 Rendimientos Financieros y Volatilidad 26
4.1.3 Valor en Riesgo (VaR) y Expected Shortfall (CVaR) 26
4.1.4 Riesgo Sistemático, No Sistemático y CAPM 27
4.1.5 Clasificación del Riesgo y Analogía con Calificadoras 27
4.2 Aprendizaje Automático para Clasificación Financiera: Síntesis 28
4.2.1 Paradigmas de Aprendizaje y Tarea de Clasificación 28
4.2.2 Modelos Base: SVM y MLP 28
4.2.3 Clasificación Ordinal: Fundamento y Relevancia para la Tarea . . 29
4.3 Redes Kolmogorov-Arnold: Fundamentos Matemáticos y Arquitectura . . 29
4.3.1 Del Problema 13 de Hilbert al Teorema de Representación 29
4.3.2 Arquitectura Matemática de las Redes KAN 31
4.3.3 Propiedades Matemáticas Clave de las KAN 32
4.4 Evaluación de Clasificadores: Métricas y Validación Temporal 33
4.4.1 Métricas de Desempeño para Clasificación Multiclase 33
4.4.2 Validación Temporal: Walk-Forward con Purga y Embargo . . . 34
4.4.3 Comparación Estadística entre Clasificadores 35
4.5 Interpretabilidad Matemática en Contextos de Alto Impacto 35
4.5.1 Taxonomía de la Interpretabilidad 35
4.5.2 *Post-hoc* vs. Interpretabilidad Intrínseca 35
4.5.3 Interpretabilidad de las KAN: Operacionalización 36
4.6 Distinciones Conceptuales Necesarias para Interpretar este Estudio 36

V Metodología **38**
5.1 Diseño General del Estudio 39
5.2 Datos: Universo, Período y Estructura 39
5.2.1 Fuentes de datos y benchmark 39
5.2.2 Universo de emisoras y control de sesgo de supervivencia 39
5.2.3 Estructura y resumen del dataset 41
5.3 Ingeniería y Selección de Variables Financieras (CORE-10) 41
5.3.1 Cálculo de Alpha, Beta y R^2 (CAPM rolling) 42
5.3.2 Riesgo asimétrico y de cola 43
5.3.3 Proxy de liquidez (Amihud) 43
5.4 Construcción de Etiquetas de Riesgo 43
5.4.1 Volatilidad futura realizada 43

5.4.2 Índice de riesgo multivariado (RiskIndex) 44
5.4.3 Discretización ordinal con cuantiles por pliegue (anti look-ahead) 45
5.5 Preprocesamiento, Depuración y Estandarización 45
5.5.1 Eliminación controlada de faltantes por construcción temporal . 45
5.5.2 Estandarización por pliegue 45
5.5.3 Control explícito de columnas con información futura 46
5.6 Protocolo Experimental: Walk-Forward PRO con Purga y Embargo 46
5.6.1 Motivación 46
5.6.2 Esquema Walk-Forward PRO con validación temporal interna . . 46
5.6.3 Configuración del protocolo 47
5.7 Modelos Evaluados y Criterios de Comparabilidad 48
5.7.1 Condiciones de comparabilidad *apples-to-apples* 48
5.7.2 Red Kolmogorov-Arnold (KAN) 48
5.7.3 Red Neuronal Multicapa (MLP) *arch-matched* 49
5.7.4 SVM lineal calibrado 50
5.8 Entrenamiento, Regularización e Hiperparámetros 51
5.8.1 Criterio de optimización y selección temporal 51
5.8.2 Tabla de hiperparámetros 51
5.8.3 Regularización específica en KAN 51
5.8.4 Filtrado de frontera de decisión (*fringe denoising*) 53
5.8.5 Actualización de malla de nudos 53
5.8.6 Ajuste de logits por priors de clase 53
5.9 Evaluación del Desempeño Predictivo 53
5.10 Comparación Estadística por Pliegue 54
5.11 Eficiencia Computacional 54
5.12 Metodología de Interpretabilidad 54
5.13 Verificación Económica: Backtesting con Predicciones OOS 55
5.13.1 Costos de transacción y benchmarks 55
5.13.2 Métricas del backtest 56
5.14 Reproducibilidad y Trazabilidad 56

VI Resultados y discusión **57**
6.1 Hallazgo 1 — Desempeño Global Competitivo bajo Validación Temporal Estricta 57

6.2 Hallazgo 2 — La Zona de Riesgo Transicional: Principal Reto y Ventaja Diferencial . . . 58
6.2.1 Patrones de error: matrices de confusión por modelo . . . 59
6.2.2 Separabilidad probabilística: curvas ROC y AUC por clase . . . 61
6.3 Hallazgo 3 — Estabilidad Temporal: Consistencia a través de Regímenes de Mercado . . . 61
6.3.1 El patrón de colapso de SVM: cuatro pliegues consecutivos . . . 62
6.3.2 Distribución de recall(Medio) por pliegue . . . 63
6.3.3 Dinámica de entrenamiento y rol del *early stopping* . . . 64
6.3.4 Respaldo estadístico por pliegue . . . 65
6.4 Eficiencia Computacional: Costo del Ciclo de Reentrenamiento . . . 65
6.5 Hallazgo 4 — Interpretabilidad Estructural como Ventaja Operativa Diferencial . . . 66
6.5.1 Ranking de importancia de variables: KAN (primera capa) . . . 66
6.5.2 Funciones $\varphi(x)$: trazabilidad funcional por variable . . . 67
6.6 Calibración Probabilística de las Predicciones OOS . . . 69
6.6.1 Error de calibración esperado (ECE) y Brier Score . . . 69
6.6.2 Diagramas de fiabilidad . . . 69
6.6.3 Lectura de los resultados de calibración . . . 70
6.7 Verificación Económica Ilustrativa: Coherencia del Pipeline OOS . . . 71
6.8 Lectura Ejecutiva de Resultados . . . 73
6.9 Perspectiva Histórica: Evolución desde el Antecedente Directo . . . 74
6.10 Síntesis de Hallazgos . . . 76

VII Implicaciones para el Sistema Financiero Mexicano 78
7.1 Adopción Operativa en Instituciones Financieras Mexicanas . . . 78
7.1.1 Fuentes de datos públicamente disponibles . . . 78
7.1.2 Integración como filtro de primer nivel en el proceso de análisis . 78
7.1.3 Protocolo de reentrenamiento operativo . . . 79
7.1.4 Alcance y límites del uso operativo: clasificación ordinal vs. probabilidades calibradas . . . 79
7.1.5 Beneficio incremental frente a la práctica actual . . . 80
7.2 Implicaciones para el Sector Gobierno: CONSAR, CNBV y Banxico . . . 81
7.2.1 CONSAR y las SIEFORES: señal granular complementaria para supervisar el riesgo accionario del ahorro nacional . . . 81

7.2.2 CNBV y la gobernanza algorítmica: IA explicable como requisito de supervisión . . . 82
7.2.3 Banxico y el CESF: el protocolo Walk-Forward como estándar de validación para modelos de riesgo . . . 83
7.3 Extensiones Naturales del Marco Propuesto . . . 84
7.3.1 Renta fija y deuda corporativa . . . 84
7.3.2 Análisis sectorial con variables macroeconómicas . . . 84
7.3.3 Integración con indicadores de riesgo sistémico del CESF . . . 84
7.4 Agenda de Investigación para México: el Vacío que Este Trabajo Identifica 84
7.4.1 La ausencia de un índice de riesgo accionario exógeno para la BMV 85
7.4.2 Una propuesta concreta para cerrar el vacío . . . 85
7.4.3 La contribución metodológica de este trabajo a esa agenda . . . 85
7.5 Evolución Financiera Sostenible: Fuentes Públicas como Democratización del Análisis de Riesgo . . . 86

VIII Conclusiones **87**
8.1 Síntesis General del Trabajo . . . 87
8.2 Cierre de Hipótesis . . . 87
8.3 Cumplimiento de Objetivos . . . 88
8.4 Aportaciones Principales . . . 89
8.5 Limitaciones . . . 90
8.6 Recomendaciones . . . 92
8.7 Líneas de Trabajo Futuro . . . 93
8.8 Conclusión Final . . . 94

Índice de cuadros

2 Interpretación de los coeficientes α y β del CAPM. . . . 27
3 Universo de 30 emisoras de la BMV utilizadas en el estudio. . . . 40
4 Resumen del dataset RAW y de la muestra efectiva utilizada para modelado. . 41
5 Variables de entrada al modelo (CORE-10) y su interpretación financiera. . . 42
6 Configuración del protocolo Walk-Forward PRO (purge + embargo). . . . 47
7 Hiperparámetros de los tres modelos (pipeline v3.4). . . . 52
8 Desempeño global OOS agregado en 13 pliegues ($n = 16{,}133$). AUC corresponde a ROC multiclase OvR con promedio macro [63, 93]. . . . 58

9 Reporte de clasificación OOS por clase, 13 pliegues ($n = 16{,}133$). 58
10 AUC OvR por clase en evaluación OOS [63, 93]. 61
11 Desempeño por pliegue: F1 macro y recall(Medio) en ventanas OOS (julio 2023 – septiembre 2025). Valores v3.4. 62
12 Resumen de estabilidad: media ± desviación estándar por pliegue (v3.4). . . 62
13 Colapso sostenido de SVM en recall(Medio): pliegues 5–8 (mar.–nov. 2024). 63
14 Estadística descriptiva del recall(Medio) por pliegue (n = 13). 64
15 Inferencia pareada Wilcoxon + IC bootstrap 95% para diferencias (KAN menos *baseline*) sobre 13 pliegues. 65
16 Eficiencia computacional promedio por pliegue (media ± std, v3.4). Columna "Calib." aplica exclusivamente a SVM. 66
17 KAN: ranking de importancia de variables por magnitud en la primera capa. . 66
18 SVM: importancia por norma ℓ_2 de coeficientes agregada entre clases. 67
19 Error de Calibración Esperado (ECE, 10 bins) y Brier Score OOS por modelo y clase. Menor es mejor en ambas métricas. Calculados sobre $n = 16{,}133$ predicciones estrictamente fuera de muestra. 69
20 Resumen del backtest OOS (KAN). La estrategia es *long-only* equiponderada en emisoras clasificadas Bajo riesgo, con costo de transacción del 0.15% sobre el *turnover* de cada rebalanceo semanal. El benchmark equiponderado (bench_eqw) es la media aritmética de los retornos de las 30 emisoras del universo en las fechas OOS. 71
21 Evolución del campo: antecedente directo vs. presente investigación. Las diferencias en período, frecuencia y protocolo hacen que la comparación de métricas no sea directa; se presentan como referencia contextual de la evolución metodológica del campo. 75

Índice de figuras

1 Esquema Walk-Forward PRO con partición interna: TRAIN_CORE, PURGE, VAL, GAP+EMB y OOS. La ventana de entrenamiento se expande en cada pliegue. 47
2 Arquitectura KAN usada en el estudio: $10 \to 32 \to 3$. 49
3 Arquitectura MLP *arch-matched*: $10 \to 32 \to 3$. 50
4 Esquema de la SVM lineal calibrada usada en el estudio. 51
5 SVM: matrices de confusión OOS (conteos y normalizada por fila). 59

6 MLP: matrices de confusión OOS (conteos y normalizada por fila). 60
7 KAN: matrices de confusión OOS (conteos y normalizada por fila). 60
8 Curvas ROC multiclase OvR en OOS para SVM (izq.), MLP (centro) y KAN (der.). 61
9 MLP (último pliegue): pérdida entrenamiento vs. validación (izq.) y F1 macro de validación con mejor época marcada (der.). 64
10 KAN (último pliegue): pérdida entrenamiento vs. validación (izq.) y F1 macro de validación con mejor época marcada (der.). 64
11 KAN: funciones univariadas $\varphi(x)$ de la primera capa para variables seleccionadas, con importancia asociada. 68
12 Diagramas de fiabilidad OvR para KAN, MLP y SVM (filas) por clase Alto, Bajo y Medio (columnas). Calculados sobre $n = 16{,}133$ predicciones OOS. La línea discontinua representa calibración perfecta ($y = x$). 70
13 Backtest OOS (KAN): capital acumulado de la estrategia (bruta y neta) frente a benchmarks (equal-weight OOS de 30 emisoras e IPC). 72

I. Introducción

La expansión del aprendizaje automático en los mercados financieros ha generado una tensión que los marcos regulatorios internacionales han formalizado como un problema de gobernanza: los modelos de mayor precisión predictiva tienden a ser los de menor explicabilidad, y en entornos de gestión de riesgo regulada esa opacidad no es una limitación académica menor, sino un obstáculo operativo concreto. La Organización Internacional de Comisiones de Valores (IOSCO), el Banco de Pagos Internacionales (BIS) y el Consejo de Estabilidad Financiera (FSB) han codificado esta tensión en marcos publicados entre 2021 y 2025 [65, 117, 118, 175], estableciendo que la interpretabilidad del mecanismo de decisión es un requisito de gobernanza, no una preferencia académica. En ese contexto, la pregunta relevante para la práctica institucional no es solo cuál modelo predice mejor, sino cuál puede adoptarse de forma responsable.

En México, esta tensión tiene consecuencias directas para el sistema de ahorro nacional. Las Administradoras de Fondos para el Retiro (Afores) gestionan recursos del Sistema de Ahorro para el Retiro (SAR) por 8.3 billones de pesos —equivalentes al 23.8% del PIB nacional— distribuidos en cerca de 70 millones de cuentas individuales [42]. Una fracción significativa de esos recursos se invierte en renta variable nacional a través de las Sociedades de Inversión Especializadas de Fondos para el Retiro (SIEFORES), cuyo régimen de inversión en la Bolsa Mexicana de Valores (BMV) [26] está sujeto a supervisión de la Comisión Nacional del Sistema de Ahorro para el Retiro (CONSAR). Sin un sistema robusto y auditable de clasificación del riesgo accionario por emisora, las instituciones que gestionan ese ahorro operan sin señales sistemáticas de cambio de régimen: toman decisiones de exposición basadas en modelos que, o bien subestiman el riesgo en mercados no estacionarios, o bien lo cuantifican con precisión pero sin poder explicar por qué. Esta tensión estructural define el problema central que la presente investigación aborda.

El contexto macroeconómico amplifica la urgencia. Los mercados accionarios de economías emergentes como México exhiben mayor volatilidad, sensibilidad a choques externos y episodios recurrentes de inestabilidad que sus contrapartes desarrolladas [94, 124]. Informes recientes del Fondo Monetario Internacional confirman que la incertidumbre macrofinanciera persiste en economías emergentes [116], y en el caso mexicano la interacción entre factores globales —tipo de cambio, tasas de la Reserva Federal, precios de materias primas— y dinámicas internas ha intensificado los ciclos de percepción del riesgo de inversión [16, 43, 82].

La ausencia de herramientas robustas, interpretables y temporalmente honestas para clasificar el riesgo accionario en la BMV representa, en ese entorno, una brecha con consecuencias directas tanto para el sistema de ahorro nacional como para la supervisión regulatoria del mercado de valores.

La evaluación cuantitativa del riesgo financiero tiene una historia metodológica consolidada. Desde la teoría de carteras de Markowitz [154] y el Capital Asset Pricing Model [140, 171, 198] hasta los modelos multifactoriales contemporáneos [61, 62], la literatura ha buscado formalizar la relación entre riesgo y rendimiento esperado con creciente precisión. Los enfoques econométricos —en particular las familias ARCH [58] y GARCH [24, 25]— han aportado herramientas ampliamente adoptadas para la modelización de la varianza condicional y el Valor en Riesgo [121].

Sin embargo, estos enfoques comparten un límite estructural: sus supuestos de linealidad, estacionariedad y distribuciones paramétricas son sistemáticamente violados en mercados emergentes, donde la distribución empírica de rendimientos exhibe colas pesadas, asimetría y cambios de régimen que los modelos clásicos no capturan [48, 158, 214]. La consecuencia práctica es crítica: precisamente en los períodos de estrés —cuando la clasificación de riesgo resulta más valiosa— estos modelos tienden a subestimar la exposición real. Este reconocimiento motivó la adopción de técnicas de aprendizaje automático capaces de aprender patrones no lineales directamente desde los datos.

El aprendizaje automático (*Machine Learning*, ML) y el aprendizaje profundo (*Deep Learning*, DL) han expandido sustancialmente el horizonte metodológico en finanzas [85, 145]. Estudios recientes documentan mejoras considerables en predicción de precios, clasificación de riesgo crediticio y modelización de series temporales financieras mediante redes neuronales profundas [66, 86, 163, 196]. En el mercado mexicano, investigaciones previas han explorado el uso de redes neuronales y técnicas de ML para la predicción del IPC, el pronóstico de rendimientos accionarios y la optimización de portafolios [77, 88, 133, 183].

No obstante, este incremento en poder predictivo viene acompañado de un problema estructural que lo hace inviable en entornos de gestión de riesgo regulada: la pérdida de interpretabilidad [89, 141]. La denominada problemática de la caja negra (*black box*) implica que, aunque un modelo produzca predicciones precisas, la lógica interna que conduce desde las variables de entrada hasta la salida permanece opaca [55]. En contextos donde las decisiones afectan estabilidad sistémica, asignación de capital y cumplimiento regulatorio —como la supervisión de CONSAR sobre las SIEFORES o la gobernanza de modelos algorítmicos que exige la

Comisión Nacional Bancaria y de Valores (CNBV) en línea con los estándares IOSCO [117]— esta opacidad no es una limitación académica menor: es un obstáculo operativo concreto [76, 175, 189].

La literatura en Inteligencia Artificial Explicable (XAI) ha intentado mitigar este problema mediante métodos *post-hoc* como SHAP [151] o LIME [184]. Sin embargo, estos enfoques no modifican la estructura del modelo: producen aproximaciones de la lógica de decisión, no la lógica misma. Un regulador que exige comprensión exacta del mecanismo de decisión —no una aproximación de él— no puede satisfacerse con herramientas cuya fidelidad depende de supuestos adicionales propios del método explicativo [118, 189]. Esta fragilidad se agrava en el contexto financiero específico: los métodos *post-hoc* generan explicaciones calibradas para la distribución de datos del período de entrenamiento. Cuando el régimen de mercado cambia —precisamente cuando la clasificación de riesgo tiene mayor valor— la explicación generada puede no corresponder a la lógica que el modelo aplica en los nuevos datos, sin que el usuario disponga de ningún mecanismo para detectarlo [55, 189].

De la revisión anterior se desprende que los enfoques disponibles no resuelven simultáneamente tres limitaciones críticas para el mercado accionario mexicano. **Primera**: los modelos tradicionales imponen supuestos de linealidad y estacionariedad que los mercados emergentes violan de forma recurrente, subestimando el riesgo precisamente cuando más importa. **Segunda**: los modelos de ML de alta precisión son cajas negras cuyo mecanismo de decisión no puede auditarse, lo que los hace inviables en entornos de gestión de riesgo regulada. **Tercera**: los estudios de ML financiero publicados en el contexto mexicano revisados en esta investigación presentan, con notable frecuencia, problemas de validación temporal que inflan artificialmente el desempeño aparente: al usar particiones aleatorias en series de tiempo, el modelo aprende, en la práctica, del futuro [5, 12, 145].

El antecedente más directo de esta investigación —Carmona Vega y González Garzón [35], cuya tesis doctoral fue reconocida con el Segundo Lugar del Premio Nacional Bolsa Mexicana de Valores 2011 (categoría Investigación)— demostró que una Red Neuronal Artificial Clasificadora (MLP con retropropagación de gradiente conjugado escalado) puede asignar categorías de riesgo de mercado (Bajo, Medio, Alto) a las 16 emisoras más activas del IPC, alcanzando una efectividad de clasificación del 74.02% en el conjunto de prueba sobre 384 observaciones trimestrales (2004–2009), con partición aleatoria 80/20 y etiquetado basado en umbrales fijos de beta. Ese trabajo establece la viabilidad del enfoque de clasificación; la presente investigación lo extiende en tres dimensiones simultáneas: arquitectura interpretable

por construcción (KAN vs. RNA clásica opaca), índice de riesgo multivariado con cuatro componentes (RiskIndex vs. indicador CAPM unidimensional) y validación temporal estricta (Walk-Forward PRO con purga y embargo vs. partición aleatoria estática).

Estos tres problemas actuando en conjunto generan una brecha verificable. Hasta donde alcanza la revisión sistemática realizada —que cubre el período 2014–2025 en REMEF, SCIELO México, Scopus y repositorios nacionales de posgrado— no se encuentra en la literatura financiera mexicana un sistema de clasificación de riesgo accionario que sea simultáneamente preciso en su evaluación fuera de muestra (*out-of-sample*, OOS), auditable en su mecanismo de decisión, y metodológicamente honesto en su protocolo de validación temporal. La pregunta que esta investigación responde no es cuál modelo predice mejor en términos globales, sino cuál arquitectura, entre las evaluadas bajo el protocolo adoptado, resuelve las tres condiciones al mismo tiempo —condición necesaria para que un sistema de clasificación de riesgo sea adoptable en entornos financieros regulados como las SIEFORES supervisadas por CONSAR o los intermediarios bajo escrutinio de la CNBV.

En este contexto emerge una línea de investigación con potencial para cerrar la brecha descrita: arquitecturas cuya interpretabilidad se deriva de su propia construcción matemática, sin requerir métodos de explicación añadidos *a posteriori*. Las Redes Kolmogorov-Arnold (KAN) [144] constituyen una propuesta reciente que materializa uno de los resultados más profundos del análisis matemático del siglo XX: el Teorema de Representación de Kolmogorov-Arnold (KART) [6, 130], cuyo origen intelectual se remonta al décimo tercer problema de Hilbert [100, 217]. El teorema establece que toda función continua multivariada puede representarse como superposición finita de funciones continuas univariadas, resultado que KAN traslada directamente a una arquitectura neuronal.

A diferencia de las redes neuronales tradicionales —cuyo fundamento es el Teorema de Aproximación Universal [47, 104, 105]— las KAN sustituyen las transformaciones lineales seguidas de activaciones fijas por funciones univariadas *aprendibles*, parametrizadas mediante B-splines [27]. La no linealidad se aprende en cada arista de la red, haciendo que la lógica de transformación por variable sea directamente inspeccionable sin herramientas externas. Análisis teóricos recientes demuestran que esta arquitectura exhibe menor sesgo espectral que las MLP y puede capturar con mayor eficiencia funciones con estructura de composición [221, 226].

Las aplicaciones financieras de KAN son recientes pero crecientes. Se han documentado resultados en predicción de series temporales bursátiles [92, 96, 224], predicción de volatilidad

implícita [39], valoración de instrumentos [142] y detección de fraude [134, 150]. Su aplicación sistemática a la clasificación ordinal del riesgo de inversión en mercados emergentes —bajo esquemas rigurosos de validación temporal y comparación estadística por pliegue— permanece, según la revisión efectuada, inexplorada. Esta investigación atiende exactamente ese vacío en el mercado accionario mexicano.

La comparación empírica se realiza frente a dos modelos de referencia: un clasificador de Máquinas de Soporte Vectorial lineal con calibración probabilística de Platt [176] y una Red Neuronal Multicapa (MLP) con dimensiones de capas emparejadas [10→32→3]. Los métodos de ensamble de árboles —Random Forest, *gradient boosting* y variantes como XGBoost— son altamente competitivos en clasificación de datos tabulares financieros y podrían superar a MLP en métricas globales. Sin embargo, comparten la limitación estructural que motiva esta investigación: su lógica de decisión requiere métodos *post-hoc* aproximados para ser explicada, y no es accesible directamente desde la arquitectura. La comparación relevante para la pregunta de investigación es entre arquitecturas que difieren en interpretabilidad estructural, no entre todas las variantes de ML disponibles. La extensión a ensambles queda propuesta como línea de trabajo futuro en el Capítulo VIII.

La presente investigación desarrolla y evalúa un sistema de clasificación del riesgo de inversión en acciones de la BMV basado en Redes Kolmogorov-Arnold. El sistema asigna a cada observación diaria una categoría ordinal —Bajo, Medio o Alto— y es evaluado mediante un protocolo Walk-Forward PRO con purga y embargo sobre 30 emisoras representativas del mercado mexicano en el período 2020–2025. Su desempeño predictivo OOS e interpretabilidad estructural se contrastan con los modelos de referencia descritos en §5, bajo condiciones estrictamente comparables en cuanto a información de entrada, protocolo de particiones y métricas de evaluación.

La pregunta de investigación que articula el trabajo es la siguiente: *¿Es posible construir, para la BMV, un sistema de clasificación de riesgo accionario que sea simultáneamente competitivo en desempeño predictivo fuera de muestra, auditable en su mecanismo de decisión, y metodológicamente honesto en su validación temporal?* Los resultados empíricos demuestran que la respuesta es afirmativa. Bajo el protocolo adoptado, KAN es la única arquitectura, entre las comparadas, que satisface las tres condiciones de forma simultánea —hallazgo que, hasta donde alcanza la revisión de la literatura mexicana realizada, no había sido documentado para la BMV.

La hipótesis de trabajo es que KAN supera a SVM en la recuperación de la clase Medio

(H_a: $recall_{\text{KAN}}(\text{Medio}) > recall_{\text{SVM}}(\text{Medio})$) de manera estadísticamente consistente a través de los pliegues temporales, mientras mantiene un desempeño global OOS competitivo. La hipótesis nula correspondiente (H_0) establece que no existe diferencia sistemática entre KAN y SVM en esa dimensión a través de las 13 particiones independientes. Las hipótesis completas, desagregadas en tres componentes verificables con métricas específicas, se formalizan en el Capítulo III.

Contribuciones Principales del Trabajo

Este trabajo realiza cinco contribuciones originales al campo de las finanzas cuantitativas aplicadas al mercado mexicano:

1. **Primera arquitectura documentada para la BMV que, entre las comparadas bajo el protocolo adoptado, resuelve simultáneamente precisión predictiva OOS, auditabilidad estructural y validación temporal honesta.** Hasta donde alcanza la revisión realizada, los modelos disponibles resuelven a lo sumo dos de las tres condiciones: MLP logra métricas globales marginalmente superiores pero es una caja negra; SVM es interpretable en sus coeficientes pero colapsa sistemáticamente la región de riesgo transicional bajo cambios de régimen. KAN es la única, de las tres evaluadas, que combina desempeño OOS competitivo, interpretabilidad estructural directamente auditable y estabilidad temporal a través de 13 regímenes de mercado distintos. Esa triple condición —no la superioridad en una métrica aislada— es el criterio de adoptabilidad en entornos financieros regulados como las SIEFORES supervisadas por CONSAR o los intermediarios bajo escrutinio de la CNBV.

2. **Diseño y construcción del RiskIndex: un índice multivariado de riesgo accionario para la BMV.** Se propone un índice continuo que integra cuatro dimensiones del riesgo accionario: volatilidad futura realizada (`VolFut`), exposición sistemática ($|\beta|$ rodante), riesgo de cola (|CVaR| al 5%) y componente idiosincrático ($1 - R^2$). La razón de iliquidez de Amihud forma parte del vector de variables predictoras (CORE-10) pero no del objetivo, a fin de no agravar la circularidad conceptual del constructo. La discretización ordinal se realiza con cuantiles estimados estrictamente dentro de cada partición de entrenamiento, eliminando la filtración de información futura en el etiquetado.

3. **Protocolo Walk-Forward PRO con purga y embargo como estándar metodológico para ML financiero en mercados no estacionarios.** El esquema de 13 particiones temporales

secuenciales con purga de observaciones dependientes, período de embargo y ventana de validación interna responde a los riesgos metodológicos documentados por López de Prado [145], Bailey et al. [12] y Arian et al. [5], y establece un estándar más riguroso que el empleado en los estudios comparables revisados para la BMV. El diseño del *gap* entre entrenamiento y prueba OOS (igual al horizonte H de la variable objetivo) garantiza que ninguna observación OOS solapa con el período de etiquetado, cerrando la principal fuente de *look-ahead* de manera exacta.

4. **Evidencia empírica de ventaja de KAN en la clasificación de riesgo transicional, con respaldo estadístico por pliegue.** Se identifica que la clase Medio —la zona de mayor ambigüedad operativa en un problema ordinal de tres niveles— es donde KAN supera consistentemente al modelo lineal: *recall* Medio de 65.3% frente al 46.6% de SVM lineal, con diferencias estadísticamente significativas a través de 13 particiones temporales independientes (Wilcoxon pareado, $p = 0.016$, diferencia media $= +0.201$, intervalo *bootstrap* $[+0.059, +0.346]$). KAN supera a SVM en esta dimensión en 9 de los 13 pliegues, con un período de colapso sostenido del modelo lineal documentado en cuatro pliegues consecutivos de 2024.

5. **Pipeline reproducible de extremo a extremo con utilidad directa para la investigación financiera y la práctica institucional.** El flujo completo —adquisición de datos con fuentes públicas (Yahoo Finance, API REST de Banxico), ingeniería de variables CORE-10, etiquetado por pliegue, entrenamiento, validación OOS e interpretación estructural— es reproducible sin acceso a datos propietarios y documentado con suficiente detalle para su adopción por equipos de análisis cuantitativo, instituciones financieras y organismos reguladores mexicanos.

El resto del documento se estructura en siete capítulos. El Capítulo II revisa críticamente la literatura sobre clasificación de riesgo financiero mediante ML e IA, incluyendo antecedentes en el mercado mexicano, el límite de interpretabilidad de los métodos de ensamble, y las aplicaciones emergentes de KAN en finanzas. El Capítulo IV desarrolla los fundamentos matemáticos del KART, la arquitectura KAN y los conceptos financieros que sustentan el diseño experimental; incluye una subsección de distinciones conceptuales necesarias para interpretar correctamente los resultados. El Capítulo III enuncia la hipótesis general con sus componentes verificables y sus correspondientes hipótesis nulas, junto con los seis objetivos específicos. El Capítulo V describe el pipeline completo: datos, CORE-10, construcción del

RiskIndex, protocolo Walk-Forward PRO y configuración de los tres modelos. El Capítulo VI reporta y discute los resultados empíricos bajo validación estrictamente OOS. El Capítulo VII articula las implicaciones del sistema para CONSAR, la CNBV, Banxico y la agenda de investigación financiera en México. El Capítulo VIII cierra con las conclusiones, limitaciones, recomendaciones concretas y líneas de trabajo futuro.

El argumento central que recorre todo el documento es este: en el mercado accionario mexicano, bajo validación temporal estricta, es posible construir un sistema de clasificación de riesgo que sea simultáneamente competitivo en desempeño OOS, auditable en su mecanismo de decisión y metodológicamente honesto. La evidencia empírica del Capítulo VI demuestra que KAN, de las arquitecturas evaluadas bajo el protocolo adoptado, es la que mejor realiza esa combinación —y que la combinación en sí, no la superioridad en ninguna métrica individual, es el criterio de adoptabilidad en entornos financieros regulados como el que este trabajo atiende.

II. Antecedentes

Este capítulo construye el argumento empírico y bibliográfico que justifica la investigación. El objetivo no es catalogar exhaustivamente la literatura, sino demostrar que los enfoques disponibles —tanto globales como en el contexto mexicano— no resuelven de forma simultánea las tres limitaciones identificadas en la introducción: supuestos paramétricos insuficientes, opacidad del mecanismo de decisión y protocolos de validación temporalmente inadecuados. La revisión se organiza de lo general a lo específico: modelos clásicos y sus límites; adopción de ML y DL en finanzas con sus debilidades metodológicas; el problema de los métodos de ensamble y el límite de su auditabilidad; el problema de la interpretabilidad y la distinción entre explicación *post-hoc* e interpretabilidad intrínseca; las Redes Kolmogorov-Arnold como arquitectura emergente; y la delimitación del vacío específico que esta investigación atiende.

2.1 Modelos Tradicionales para Riesgo Financiero: Alcances y Limitaciones

La gestión cuantitativa del riesgo financiero se consolidó sobre herramientas estadísticas y econométricas que sostuvieron la práctica profesional y la regulación prudencial durante décadas. El Valor en Riesgo (VaR) se convirtió en el estándar de facto para estimar pérdidas máximas esperadas bajo niveles de confianza específicos [121], y las familias ARCH [58] y GARCH [24, 25] aportaron el marco para modelizar la varianza condicional y el agrupamiento de volatilidad. Para el análisis de dependencia temporal, los modelos ARIMA representaron durante décadas la herramienta estándar de caracterización y pronóstico [28].

Sin embargo, estos enfoques comparten una limitación estructural que la literatura ha documentado extensamente: sus supuestos —normalidad de rendimientos, estacionariedad, relaciones lineales— son violados de forma sistemática en los mercados financieros contemporáneos, y de manera especialmente pronunciada en economías emergentes [48, 158, 214]. En períodos de estrés, donde la gestión del riesgo es más crítica, estos modelos tienden a subestimar la exposición real precisamente porque las distribuciones empíricas exhiben colas pesadas, asimetría y cambios de régimen que los modelos paramétricos no capturan.

En el mercado mexicano, esta tensión entre modelos paramétricos y arquitecturas no lineales ha sido documentada empíricamente. Gómez-Ramos et al. [84] compararon directamente modelos GARCH y redes neuronales artificiales en la predicción del IPC y del Dow Jones durante 2010–2011, encontrando que la RNA captura mejor los patrones no lineales de la serie,

mientras que GARCH mantiene ventaja en métricas de error convencionales. Este resultado —coherente con la literatura internacional sobre el compromiso entre parsimonia paramétrica y expresividad no lineal— motivó los enfoques híbridos y, en última instancia, la búsqueda de arquitecturas que resuelvan ambas dimensiones simultáneamente [133, 144].

Esta brecha motivó la búsqueda de técnicas capaces de aprender patrones complejos directamente desde los datos, sin imponer una forma funcional rígida.

2.2 ML y DL en Finanzas: Poder Predictivo y Fragilidad Empírica

El aprendizaje automático y profundo ampliaron sustancialmente el horizonte metodológico en finanzas por su capacidad para modelar no linealidades, interacciones de alta dimensionalidad e interdependencias que los modelos econométricos no capturan [21, 85, 145]. En predicción, clasificación y gestión de riesgo, métodos como SVM, ensambles de árboles y redes neuronales profundas han mostrado desempeño superior frente a sus contrapartes lineales en múltiples contextos. No obstante, la evidencia empírica debe leerse con cautela: el desempeño reportado en muchos estudios está condicionado por la calidad del protocolo experimental, y ahí reside la fragilidad del campo.

2.2.1 Panorama de aplicaciones globales

Las arquitecturas recurrentes, y en particular LSTM, han mostrado capacidad para capturar dependencias temporales complejas en predicción de volatilidad y dirección de precios [66, 74, 103]. Los modelos de ensamble basados en árboles han sido estudiados por su robustez en clasificación financiera [2]. En riesgo de crédito, predicción de incumplimiento y detección de quiebra, la literatura reporta mejoras predictivas documentadas mediante múltiples arquitecturas de ML y DL [81, 181, 200]. Las revisiones sistemáticas de Sezer et al. [196] y Mienye et al. [163] sintetizan el estado del campo con énfasis en la amplitud de aplicaciones y la diversidad de métricas reportadas, poniendo de relieve que la comparabilidad entre estudios es con frecuencia limitada por diferencias en protocolo, ventana temporal y universo de activos.

2.2.2 Aplicaciones en el contexto de la BMV

En México, la investigación ha explorado de manera creciente el uso de técnicas de inteligencia artificial para evaluar el riesgo y apoyar las decisiones de inversión en la BMV. Los antecedentes se organizan en tres líneas temáticas.

En predicción de rendimientos y volatilidad, Landassuri-Moreno et al. [133] combinaron modelos econométricos y redes neuronales evolutivas para pronosticar el IPC mexicano, y

Reyna Miranda et al. [183] propusieron un modelo neuro-wavelet para datos de alta frecuencia sobre 25 emisoras de gran capitalización. Gallardo Del Ángel [77] documentó el desempeño de redes neuronales con retropropagación y propagación resiliente para series financieras individuales, y la predicción de volatilidad accionaria ha motivado trabajos de posgrado mediante modelos recurrentes [99]. La generalización de estos enfoques al mercado de renta variable mexicana ha sido explorada con distintas técnicas de ML [37, 182].

En optimización de portafolios, Guerrero Camargo [88] y Samaniego Alcántar [191] aplicaron LSTM y metaheurísticas para mejorar la construcción de carteras con activos de la BMV. Rodríguez Rodríguez [186] desarrollaron un sistema de dos etapas para la BMV: clasificación de acciones mediante SVM, Random Forest, redes neuronales y el clasificador `PBC4cip`, seguida de optimización multiobjetivo con el algoritmo genético NSGA-III para perfiles agresivo, conservador y moderado. Los portafolios resultantes superaron al CETES, al S&P 500 y al S&P/BMV IPC en el período evaluado (dic. 2019 – dic. 2020). Este trabajo establece que la clasificación de acciones con ML para construir portafolios en la BMV es factible; la presente investigación se diferencia en la arquitectura (KAN vs. SVM/NN/RF), en la interpretabilidad estructural del mecanismo de decisión y en el rigor del protocolo de validación temporal.

En clasificación de riesgo y análisis de emisoras, el antecedente más directo de esta investigación en la literatura mexicana es la tesis doctoral de Carmona Vega y González Garzón [35], presentada en la Universidad Popular Autónoma del Estado de Puebla (UPAEP) y reconocida con el Segundo Lugar del Premio Nacional Bolsa Mexicana de Valores 2011 (categoría Investigación). En ella, se implementó un **perceptrón multicapa clasificador** (MLP con retropropagación de gradiente conjugado escalado, 18 neuronas en la capa oculta, activación logística) para asignar categorías ordinales de riesgo de mercado —Bajo, Medio y Alto— a las 16 emisoras más activas del IPC. Sobre 384 observaciones trimestrales (2004–2009) con partición aleatoria 80/20, el clasificador alcanzó una efectividad del 74.02% en el conjunto de prueba (77 observaciones fuera de la muestra de entrenamiento), con la tasa de rendimiento requerida (K_j), los CETES a 91 días y los rendimientos accionarios como variables de entrada. El etiquetado se realizó mediante umbrales fijos de beta (< 0.70 / 0.70–1.14 / > 1.14), con clases balanceadas por diseño. Este trabajo establece la viabilidad del enfoque de clasificación de riesgo con IA en la BMV. La presente investigación lo extiende en tres dimensiones: (*i*) arquitectura de interpretabilidad intrínseca (KAN vs. MLP opaco), (*ii*) índice de riesgo multivariado con cuatro componentes de mercado (RiskIndex vs. indicador CAPM unidimensional) y (*iii*) validación temporal rigurosa sin filtración de información futura (Walk-Forward PRO con purga y embargo, $n = 16{,}133$ observaciones OOS, vs. partición aleatoria estática

sobre 384 obs. trimestrales).

La clasificación de emisoras mediante métodos de discriminante múltiple, Logit y redes neuronales fue explorada por García Salgado y Morales Castro [78], y en el ámbito del riesgo de mercado, Trejo Becerril y Gallegos David [212] aplicaron VaR y la Beta del CAPM a portafolios de acciones mexicanas. Fuera del mercado accionario pero en el contexto financiero mexicano, Melchor Pérez et al. [160] aplicaron aprendizaje computacional para predicción de riesgo crediticio con datos nacionales, y Golbayani et al. [83] compararon redes neuronales, SVM y árboles de decisión para la predicción de calificaciones corporativas, estudio relevante como referente metodológico de clasificación ordinal supervisada.

El panorama trazado revela un patrón: la literatura mexicana sobre ML/IA en la BMV ha producido avances valiosos en predicción de rendimientos [77, 133, 183], optimización de portafolios [88, 186, 191] y clasificación de emisoras por método discriminante [78]. El único trabajo que abordó explícitamente la misma tarea —clasificación ordinal del riesgo accionario en la BMV mediante IA— fue Carmona Vega y González Garzón [35], quien demostró su viabilidad con RNA clásica y validación estática. Lo que ninguno de estos trabajos integra de forma simultánea son las tres condiciones que definen la adoptabilidad operativa de un sistema de clasificación de riesgo en entornos regulados: (i) desempeño OOS con integridad temporal y sin filtración de información futura, (ii) interpretabilidad estructural del mecanismo de decisión directamente auditable por un regulador, y (iii) clasificación ordinal explícita del riesgo por emisora con protocolo de validación estadísticamente riguroso. Esta triple ausencia simultánea —no la ausencia de trabajos previos en general— es el vacío específico que la presente investigación atiende [5, 12, 117, 145, 189].

2.2.3 Limitaciones metodológicas recurrentes en ML financiero

La validez externa de resultados en ML financiero puede verse comprometida por tres prácticas experimentales problemáticas, documentadas de forma recurrente en la literatura especializada. **Primera**, la validación temporal incorrecta: las particiones aleatorias de series de tiempo introducen *look-ahead bias*, dado que el modelo puede haber sido entrenado con información cronológicamente posterior a los períodos que evalúa [12, 145]. **Segunda**, el sobreajuste por selección de hiperparámetros sin control de multiplicidad, que produce métricas en muestra artificialmente elevadas y generalización débil en uso real [5, 95]. **Tercera**, el *backtest overfitting*: cuando se prueban múltiples configuraciones sobre el mismo conjunto histórico, la probabilidad de encontrar una que parezca funcionar por azar crece de forma no despreciable, aun cuando el desempeño real fuera de muestra sea nulo [12, 95].

Un análisis comparativo reciente de métodos de evaluación fuera de muestra en un entorno controlado sintético concluye que el protocolo de validación cruzada purificada y con embargo es marcadamente superior a la validación cruzada estándar y al walk-forward simple para mitigar el sobreajuste en modelos financieros [5]. Esta evidencia refuerza la decisión metodológica central de este trabajo: adoptar un protocolo Walk-Forward PRO con purga y embargo que responde directamente a los tres vectores de fuga de información documentados. El debate no debe restringirse a cuál modelo predice mejor, sino bajo qué protocolo se establecen esas comparaciones y qué tan robustas son frente a cambios de régimen [19, 52, 54].

2.3 Métodos de Ensamble: Desempeño Tabular Superior con Límites de Auditabilidad

Los métodos de ensamble de árboles de decisión —particularmente las variantes de *gradient boosting* como XGBoost y las arquitecturas de bosque aleatorio (*Random Forest*)— han establecido el estado del arte en clasificación de datos tabulares en múltiples benchmarks financieros [2, 86]. Cuando los datos tienen estructura tabular con un número moderado de variables, estas arquitecturas igualan o superan con frecuencia a las redes neuronales profundas en métricas agregadas, manteniendo una eficiencia computacional considerablemente mayor. En el contexto de la predicción financiera, esta competitividad los hace opciones relevantes para muchas tareas de clasificación y regresión.

Sin embargo, para el problema específico que esta investigación aborda —clasificación de riesgo con interpretabilidad estructural en entornos regulados— los métodos de ensamble comparten una limitación de diseño que los hace insuficientes: la lógica de decisión de un ensamble de cientos o miles de árboles no es accesible por construcción, y para ser explicada requiere métodos *post-hoc* aproximados como SHAP o LIME [151, 184]. La diferencia entre KAN y un modelo de ensamble con SHAP no es de grado: es cualitativa. Las funciones $\varphi(x)$ de la primera capa de KAN *son* el mecanismo de decisión; la explicación SHAP de un ensamble es una aproximación del mecanismo de un modelo diferente, con fidelidad que depende de supuestos propios del método explicativo que raramente son transparentes para el regulador [189]. Esta distinción es precisa en el contexto que IOSCO y el BIS han formalizado: la interpretabilidad que los marcos regulatorios demandan es la lógica exacta del proceso de decisión, no una representación local aproximada de ella [117, 175].

Por estas razones, el diseño experimental de esta investigación utiliza SVM lineal y MLP como modelos de referencia, no métodos de ensamble. Los primeros representan, respectivamente, la

frontera de decisión lineal clásica y la red neuronal sin interpretabilidad estructural intrínseca: los dos polos del eje interpretabilidad-desempeño que KAN aspira a superar de forma conjunta. La comparación KAN vs. ensambles, si bien sería metodológicamente interesante para comparar desempeño tabular agregado, no abordaría la pregunta central de esta investigación. Dicha extensión queda propuesta como línea de trabajo futuro en el Capítulo VIII.

2.4 Interpretabilidad en IA Financiera: *Post-hoc* vs. Intrínseca

El incremento en poder predictivo de los modelos de ML y DL se logra frecuentemente sacrificando interpretabilidad, dando lugar al paradigma de la caja negra [89, 141]. En finanzas, esta opacidad tiene consecuencias concretas: dificulta la validación económica del modelo, obstaculiza la auditoría regulatoria, impide detectar sesgos indeseados y debilita la rendición de cuentas en decisiones de alto impacto [55, 76, 189]. La interpretabilidad no es, por tanto, un objetivo complementario al desempeño, sino una condición de viabilidad operativa en entornos financieros regulados —una distinción que organismos como IOSCO han formalizado en su agenda normativa sobre el uso de IA en los mercados de capitales [117, 118, 175].

2.4.1 XAI *post-hoc*: aportes y límites

Las técnicas de Inteligencia Artificial Explicable (XAI) *post-hoc* buscan explicar modelos complejos sin modificar su arquitectura. LIME aproxima localmente la frontera de decisión mediante modelos simples interpretables [184]; SHAP asigna contribuciones a variables mediante valores de Shapley derivados de teoría de juegos cooperativos [151]. En finanzas, estos métodos se han aplicado para explicar modelos de calificación crediticia y análisis de riesgo [161, 209]. Sin embargo, su naturaleza aproximada implica que la explicación no coincide necesariamente con la lógica interna exacta del modelo, puede ser sensible a perturbaciones en los datos de referencia y depende de supuestos del método explicativo que raramente son transparentes para el usuario final [189]. Explicar una caja negra con una aproximación externa no elimina la opacidad intrínseca: la desplaza.

Esta distinción —entre desplazar la opacidad y eliminarla— es el fundamento de la perspectiva que orienta el diseño de esta investigación, desarrollada en la sección siguiente.

2.4.2 Modelos interpretables por diseño: motivación en finanzas

Ante las limitaciones de los métodos *post-hoc*, Rudin [189] argumenta que en decisiones de alto impacto la alternativa correcta no es mejorar las explicaciones de modelos opacos, sino construir modelos con interpretabilidad intrínseca que sean igualmente competitivos

en desempeño. Esta perspectiva es especialmente pertinente en finanzas: la necesidad de justificar decisiones ante comités de inversión, auditores internos y reguladores crea un incentivo estructural para modelos cuya lógica de decisión sea directamente inspeccionable, no aproximada [55, 76]. En este marco, las arquitecturas cuya transparencia proviene de su construcción matemática —y no de herramientas adicionales— ofrecen una vía genuina para conciliar desempeño con explicabilidad sin comprometer la integridad del mecanismo de decisión. La distinción entre interpretabilidad intrínseca y explicabilidad *post-hoc* es, como se desarrollará en el Capítulo IV, uno de los ejes conceptuales fundamentales de este trabajo.

Esta perspectiva encuentra eco directo en el lenguaje de los marcos regulatorios más recientes: IOSCO FR06/2021 establece que los intermediarios que emplean modelos algorítmicos en decisiones de alto impacto deben poder explicar el mecanismo de decisión del modelo a supervisores y otras partes interesadas relevantes, distinguiendo explícitamente entre la disponibilidad de ese mecanismo exacto y sus aproximaciones *post-hoc* [117, 118, 175]. La interpretabilidad que KAN ofrece por construcción —no a través de un método adicional— es la única clase de interpretabilidad que satisface ese estándar de forma literal.

2.5 Redes Kolmogorov-Arnold: Fundamento, Arquitectura y Evidencia

2.5.1 Del Problema 13 de Hilbert al Teorema de Representación

La base conceptual de las KAN se conecta con el decimotercer problema de Hilbert [100, 101]: ¿puede toda función algebraica de varias variables expresarse mediante superposición de funciones de dos variables? Vitushkin [217] y trabajos posteriores contribuyeron a delimitar el alcance del problema [218]. Kolmogorov estableció el resultado fundamental en 1957: toda función continua multivariada puede expresarse mediante sumas finitas de funciones univariadas compuestas [129, 130]. Arnold refinó y expandió estas ideas [6, 7, 8], y Sprecher [204] y Sprecher [205] aportaron estructuras más simétricas para las representaciones. Trabajos posteriores desarrollaron pruebas constructivas y extensiones analíticas del resultado [29, 147, 148, 206].

2.5.2 Arquitectura KAN: interpretabilidad por construcción

Liu et al. [144] propusieron las Redes Kolmogorov-Arnold como arquitectura neuronal donde la no linealidad se desplaza desde los nodos hacia las aristas: en lugar de transformaciones lineales seguidas de activaciones fijas, cada conexión es una función univariada aprendible parametrizada mediante *splines* [27], mientras que los nodos agregan únicamente por suma. Esta característica estructural permite inspeccionar directamente las funciones aprendidas

por variable, ofreciendo interpretabilidad sin herramientas externas de explicación. Análisis teóricos recientes confirman que las KAN exhiben menor sesgo espectral que las MLP y pueden capturar con mayor eficiencia funciones con estructura de composición [221]. En comparaciones controladas por número de parámetros y operaciones de punto flotante, la ventaja de KAN frente a MLP se atribuye principalmente a la expresividad de sus funciones de activación *spline* [226]. El marco fue ampliado en KAN 2.0, integrando estructuras simbólicas para el descubrimiento científico [143].

2.5.3 Aplicaciones en finanzas y series temporales

La literatura de KAN aplicada a finanzas es reciente pero de crecimiento acelerado. En predicción de series bursátiles, Haryono et al. [96] propusieron redes KAN temporales permutadas combinadas con análisis de sentimiento generativo para el pronóstico de precios accionarios, documentando mejoras sobre modelos LSTM. Un estudio comparativo amplio [92] propuso la arquitectura Multi-layer Mixture-of-KAN (MMK) para pronóstico multivariado de series de tiempo y evidenció la efectividad de KAN en entornos de alta dimensionalidad con estructura de composición. Para series temporales genéricas con aplicación financiera, Xu et al. [224] analizaron el balance entre capacidad predictiva e interpretabilidad de las KAN. En valoración de instrumentos, Liu et al. [142] aplicaron KAN con estructura finance-informed para la fijación de precios de opciones. En predicción de volatilidad implícita, Cho et al. [39] documentaron resultados competitivos de KAN frente a GARCH y modelos basados en LSTM. En detección de fraude financiero, Le et al. [134] y Lu y Zhan [150] reportaron que las variantes eficientes de KAN alcanzan precisión comparable a las arquitecturas de referencia con ventajas en trazabilidad del mecanismo de decisión.

En conjunto, estos trabajos sugieren que la combinación de flexibilidad no lineal e interpretabilidad estructural hace de las KAN candidatas relevantes para tareas de riesgo financiero. Sin embargo, su aplicación a la *clasificación ordinal* del riesgo de inversión en acciones de mercados emergentes —bajo protocolos rigurosos de validación temporal y comparación estadística por pliegue— no ha sido documentada en la literatura existente según la revisión efectuada.

2.6 Vacío en la Literatura y Posición del Estudio

La evidencia revisada revela un patrón consistente en dos niveles. A nivel global, ML y DL han impulsado mejoras predictivas en tareas financieras, pero con frecuencia bajo tensiones entre desempeño e interpretabilidad, y con riesgos metodológicos que comprometen la validez

fuera de muestra cuando la evaluación temporal no se diseña con rigor [5, 12, 145]. A nivel local, la literatura mexicana sobre riesgo accionario en la BMV ha producido contribuciones valiosas —en predicción de rendimientos, volatilidad y clasificación de emisoras— pero sin integrar los tres requisitos que definen la viabilidad operativa de un sistema de clasificación de riesgo en entornos regulados: desempeño OOS con integridad temporal, interpretabilidad estructural del mecanismo de decisión, y clasificación ordinal explícita del riesgo por emisora.

2.6.1 El vacío específico: tres ausencias simultáneas

La revisión sistemática realizada —que cubre el período 2014–2025 en REMEF, SCIELO México, Scopus y repositorios nacionales de posgrado, con los términos de búsqueda: KAN, Kolmogorov-Arnold, BMV, riesgo accionario, clasificación de riesgo, Walk-Forward, ML finanzas México— no encuentra evidencia de estudios que integren de forma simultánea: *(i)* arquitecturas KAN orientadas a interpretabilidad estructural; *(ii)* una tarea explícita de clasificación ordinal del riesgo de inversión en tres niveles discretos —Bajo, Medio, Alto— con un índice de riesgo multivariado como variable objetivo; y *(iii)* protocolos rigurosos de evaluación temporal —validación Walk-Forward con purga y embargo, y comparación estadística entre clasificadores— aplicados al mercado accionario mexicano. Esta ausencia es especialmente relevante porque la utilidad práctica de un sistema de clasificación de riesgo requiere las tres condiciones en conjunto: desempeño competitivo fuera de muestra, lógica de decisión auditable, y evaluación metodológicamente honesta [19, 52, 144, 189].

El único trabajo en la literatura mexicana que abordó parcialmente este vacío —Carmona Vega y González Garzón [35]— resolvió la condición (*ii*) (clasificación ordinal de riesgo) con RNA clásica, pero no las condiciones (*i*) (interpretabilidad estructural) ni (*iii*) (validación temporal sin filtración de información): el sistema propuesto en la presente investigación es el primero, en la literatura revisada, en integrar las tres de forma simultánea para el mercado accionario mexicano.

2.6.2 Posición de la investigación y originalidad

Esta investigación se posiciona en ese vacío: aplica y evalúa KAN en la clasificación ordinal del riesgo de inversión para acciones de la BMV, bajo validación Walk-Forward PRO con purga y embargo, comparación estadística pareada por pliegue, e interpretabilidad estructural verificable sin herramientas *post-hoc*. Según la revisión efectuada, es el primer estudio en el contexto mexicano que integra simultáneamente estas cuatro condiciones.

La pertinencia científica se sustenta en tres planos complementarios. **Primero**, aportar eviden-

cia empírica sobre una arquitectura emergente aplicada a un dominio financiero concreto y relevante para México [143, 144]. **Segundo**, responder a la demanda de modelos auditables en decisiones de alto impacto, demanda que organismos internacionales como IOSCO y el FSB han formalizado en sus marcos regulatorios recientes [65, 117, 118, 189]. **Tercero**, establecer para la BMV un protocolo de evaluación más riguroso que el empleado en los estudios comparables revisados para el mercado mexicano, contribuyendo a elevar el estándar metodológico del campo. La investigación no solo propone un modelo: propone una forma más robusta de construir y evaluar sistemas de clasificación de riesgo en mercados emergentes, con independencia de la arquitectura específica que se adopte.

III. Hipótesis y Objetivos

3.1 Hipótesis de Investigación

Con base en la revisión del estado del arte, la identificación de la brecha en la literatura y las propiedades matemáticas de interpretabilidad y eficiencia de las Redes Kolmogorov-Arnold documentadas en el marco teórico [144, 145, 189], se plantea la siguiente hipótesis general y sus componentes verificables.

Hipótesis General

Un sistema de clasificación del riesgo de inversión en acciones de la Bolsa Mexicana de Valores basado en Redes Kolmogorov-Arnold (KAN) logra, **entre las arquitecturas comparadas bajo el protocolo adoptado**, desempeño predictivo fuera de muestra **competitivo** frente a los modelos de referencia —Máquinas de Soporte Vectorial (SVM) y Redes Neuronales Multicapa (MLP)— bajo validación temporal rigurosa, con una **ventaja diferencial en la zona de riesgo transicional** estadísticamente consistente a través de particiones temporales independientes, y proveyendo **interpretabilidad estructural intrínseca** que dichos modelos de referencia no ofrecen por construcción. La combinación de estas tres condiciones —competitividad predictiva, auditabilidad intrínseca y honestidad temporal— y no la superioridad en ninguna métrica individual, hace del sistema propuesto una arquitectura viable para la gestión de riesgo en entornos financieros regulados. En la literatura revisada, esta combinación no ha sido documentada para el mercado accionario mexicano.

Hipótesis Específicas

La hipótesis general se descompone en tres componentes verificables mediante los resultados del Capítulo VI, junto con sus correspondientes hipótesis nulas:

1. **H1 — Desempeño predictivo competitivo.** KAN alcanza métricas de clasificación fuera de muestra —F1 macro y AUC OvR macro— que no son estadísticamente inferiores a las de SVM lineal calibrado, y se encuentran dentro de un margen no mayor a 3 puntos porcentuales respecto a las de MLP, evaluados bajo el mismo protocolo Walk-Forward PRO con purga y embargo sobre 13 particiones temporales independientes [52, 145].

   Hipótesis nula correspondiente ($H_{0,1}$): No existe diferencia estadísticamente significativa en F1 macro OOS entre KAN y MLP, ni entre KAN y SVM, a través de los 13 pliegues temporales (Wilcoxon pareado, $\alpha = 0.05$, bilateral).

2. **H2 — Ventaja consistente en la región de riesgo transicional.** En la clasificación de la clase Medio —la zona de mayor ambigüedad operativa en un problema ordinal de tres niveles— KAN muestra un *recall* consistentemente superior al de SVM lineal, con diferencias estadísticamente significativas a través de los 13 pliegues temporales según prueba de Wilcoxon pareada [52, 54]. Esta hipótesis contrasta específicamente KAN contra SVM; la diferencia KAN–MLP en la clase Medio no forma parte del alcance inferencial formal de H2 dado que se espera que ambas arquitecturas no lineales sean estadísticamente equivalentes en esta dimensión. La clase Medio concentra los estados frontera donde la frontera lineal del SVM presenta la mayor degradación sistemática bajo cambios de régimen: en un problema ordinal, los errores en la clase intermedia —que comparte rasgos con el riesgo bajo y el alto— representan la mayor dificultad intrínseca y la mayor consecuencia operativa para decisiones de exposición [75, 91].

   Hipótesis nula correspondiente ($H_{0,2}$): No existe diferencia sistemática en *recall*(Medio) entre KAN y SVM a través de los 13 pliegues temporales (Wilcoxon pareado, $\alpha = 0.05$, unilateral: KAN $>$ SVM).

3. **H3 — Interpretabilidad estructural verificable.** Las funciones univariadas $\phi(x)$ aprendidas por KAN por variable son directamente inspeccionables sin herramientas *post-hoc*, y la importancia relativa de las variables financieras obtenida de la arquitectura muestra coherencia con la teoría financiera del riesgo: variables de volatilidad, riesgo de cola (CVaR) y exposición sistemática (Beta) deben aparecer entre las de mayor peso [55, 76, 144, 221].

   Hipótesis nula correspondiente ($H_{0,3}$): La jerarquía de importancias de la primera capa KAN no es coherente con la teoría financiera del riesgo (Vol_252, CVaR y Beta no emergen entre las tres variables de mayor peso relativo en el ranking de importancias).

Las tres hipótesis como sistema. H1, H2 y H3 no son independientes: constituyen el triple argumento de adoptabilidad que articula la tesis central del trabajo. H1 establece que el sistema es competitivo en desempeño OOS —condición necesaria para cualquier uso práctico—. H2 establece que esa competitividad se concentra donde más importa operativamente: la región transicional donde los modelos lineales fallan sistemáticamente bajo cambio de régimen —condición que diferencia a KAN de SVM en el eje de robustez—. H3 establece que el mecanismo de decisión es directamente auditable —condición que diferencia a KAN de MLP en el eje de gobernanza—. Un sistema que cumple H1 pero no H2 puede ser competitivo pero inestable bajo estrés; uno que cumple H1 y H2 pero no H3 puede ser robusto pero no

adoptable en entornos regulados. La tesis afirma que KAN, de las tres arquitecturas evaluadas bajo el protocolo adoptado, cumple las tres simultáneamente. La evidencia del Capítulo VI lo confirma.

3.2 Objetivos de Investigación

3.2.1 Objetivo General

Desarrollar, implementar y evaluar un sistema de clasificación del riesgo de inversión en acciones de la Bolsa Mexicana de Valores basado en Redes Kolmogorov-Arnold (KAN), que combine desempeño predictivo fuera de muestra competitivo, validación temporal metodológicamente rigurosa e interpretabilidad estructural intrínseca, con el fin de proveer a analistas financieros, gestores de riesgo e investigadores una herramienta auditable de apoyo a la toma de decisiones en el mercado accionario mexicano.

3.2.2 Objetivos Específicos

1. **Construir un conjunto de datos financiero público, robusto y reproducible para la BMV.** Diseñar y construir un *dataset* con 30 emisoras representativas del mercado mexicano, basado en fuentes públicamente disponibles (Yahoo Finance, API REST de Banxico), que incluya variables de exposición sistemática (Beta, Alpha, R^2 rodantes), volatilidad en múltiples horizontes (20, 60 y 252 días hábiles), riesgo de cola (CVaR al 5%), volatilidad a la baja y liquidez de Amihud (conjunto CORE-10). La reproducibilidad sin acceso a datos propietarios es un requisito de diseño [1, 145].

2. **Diseñar y construir el RiskIndex: un índice multivariado de riesgo accionario para la BMV.** Proponer un índice continuo que integre **cuatro dimensiones** del riesgo accionario —volatilidad futura realizada (VolFut), exposición sistemática ($|\beta|$ rodante), riesgo de cola ($|$CVaR$|$ al 5%) y componente idiosincrático ($1 - R^2$)— como variable objetivo de la clasificación. Establecer un protocolo de discretización en tres categorías ordinales —Bajo, Medio, Alto— con cuantiles estimados exclusivamente dentro de cada partición de entrenamiento, eliminando la filtración de información futura y garantizando que el orden entre clases refleje magnitudes de riesgo relativas reales [91, 145, 158].

3. **Implementar el protocolo Walk-Forward PRO con purga y embargo.** Diseñar e implementar un esquema de validación temporal con 13 particiones cronológicas secuenciales, purga de observaciones con solapamiento temporal y período de embargo igual al horizonte de la variable objetivo ($H = 60$ días), que elimine las tres fuentes principales de

fuga de información en ML financiero: *look-ahead* en etiquetado, escalamiento global y contaminación temporal entre particiones [5, 12, 13, 145, 208].

4. **Desarrollar y entrenar el sistema KAN para clasificación de riesgo accionario.** Implementar en Python una Red Kolmogorov-Arnold con parametrización B-spline cúbica ($p = 3$, $n = 10$ nudos internos), conexión residual SiLU, actualización adaptativa de nudos, ajuste de logits para manejo de desbalance de clases, y *early stopping* sobre F1 macro en validación; entrenado bajo el protocolo Walk-Forward PRO. Documentar la arquitectura completa, los hiperparámetros y los tiempos de entrenamiento para reproducibilidad [27, 144]. La elección de $p = 3$ (cúbico) y $n = 10$ nudos internos no es arbitraria: produce funciones $\phi(x)$ con suavidad C^2 y soporte local suficiente para distinguir regímenes de volatilidad alta, moderada y baja, mientras mantiene la inspeccionabilidad directa que el Objetivo 6 requiere [27, 144].

5. **Comparar el desempeño del sistema KAN frente a SVM y MLP bajo condiciones *apples-to-apples*.** Evaluar los tres clasificadores con el mismo conjunto CORE-10, el mismo protocolo de particiones temporales, el mismo escalamiento por pliegue y las mismas métricas (F1 macro, AUC OvR macro, precisión y recall por clase, matrices de confusión). Complementar con pruebas estadísticas pareadas de Wilcoxon e intervalos bootstrap por pliegue para distinguir mejoras consistentes de fluctuaciones temporales [19, 52]. En todos los casos, reportar tanto el estadístico p como el tamaño del efecto con intervalos de confianza al 95%, a fin de distinguir significancia estadística de relevancia práctica [56].

6. **Documentar y analizar la interpretabilidad estructural de KAN.** Extraer y visualizar las funciones univariadas $\phi(x)$ aprendidas por variable en la primera capa, construir el ranking de importancia global, y contrastar la coherencia de la jerarquía aprendida con la teoría financiera del riesgo accionario. Proveer evidencia de interpretabilidad auditable en el sentido de Doshi-Velez y Kim [55] y Rudin [189], sin recurrir a métodos *post-hoc* como SHAP o LIME, demostrando que la arquitectura misma provee la trazabilidad de decisión que los entornos financieros regulados requieren.

Correspondencia entre objetivos y contribuciones. Los seis objetivos específicos se corresponden directamente con las cinco contribuciones originales declaradas en la Introducción: el objetivo 1 sustenta la Contribución 5 (pipeline reproducible con datos públicos); los objetivos 2 y 3 sustentan las Contribuciones 2 y 3 (RiskIndex y protocolo Walk-Forward PRO); el

objetivo 4 sustenta la Contribución 1 (primera arquitectura documentada para la BMV, de las tres evaluadas bajo este protocolo, que resuelve simultáneamente los tres problemas identificados); el objetivo 5 sustenta la Contribución 4 (evidencia empírica en riesgo transicional con respaldo estadístico por pliegue); y el objetivo 6 sustenta el componente de interpretabilidad de la Contribución 1.

IV. Marco Teórico

Este capítulo establece el andamiaje teórico que sustenta la investigación. Se organiza en seis bloques interrelacionados: (1) fundamentos del riesgo financiero y su medición cuantitativa; (2) síntesis de los principios de aprendizaje automático relevantes para la tarea; (3) fundamentos matemáticos profundos de las Redes Kolmogorov-Arnold, núcleo teórico diferenciador del trabajo; (4) metodología de evaluación de clasificadores, con énfasis en validación temporal e interpretabilidad; (5) interpretabilidad matemática en contextos de alto impacto; y (6) distinciones conceptuales necesarias para interpretar correctamente los resultados y sus implicaciones.

4.1 Fundamentos del Riesgo Financiero y su Medición

4.1.1 Concepto, Naturaleza y Tipologías del Riesgo Financiero

El riesgo financiero es la posibilidad de que el rendimiento efectivo de una inversión difiera de su rendimiento esperado, especialmente en sentido adverso [111, 121]. Si R denota el rendimiento de un activo, el riesgo puede formalizarse como dispersión alrededor de $\mathbb{E}[R]$:

$$\mathrm{Var}(R) = \mathbb{E}\left[(R - \mathbb{E}[R])^2\right], \qquad \sigma_R = \sqrt{\mathrm{Var}(R)}. \tag{1}$$

La literatura distingue cuatro tipologías principales [111, 158]: riesgo de mercado (movimientos adversos en precios o tasas), riesgo de crédito (incumplimiento de contraparte), riesgo de liquidez (incapacidad de vender sin pérdida significativa) y riesgo operativo (fallos en procesos, personas o sistemas). Esta investigación se centra en el **riesgo de mercado** para acciones de la BMV.

Nota sobre la naturaleza no-*i.i.d.* del riesgo. En mercados reales, el riesgo financiero rara vez se comporta como un proceso independiente e idénticamente distribuido. Se observan dependencia temporal, cambios de régimen, asimetrías y colas pesadas [48, 158, 214]. Esta observación tiene dos consecuencias directas: las métricas de riesgo deben considerar el comportamiento de cola, y la validación de modelos predictivos debe preservar la causalidad temporal [12, 145].

4.1.2 Rendimientos Financieros y Volatilidad

El rendimiento logarítmico del activo con precio P_t en el período t se define como:

$$r_t = \ln\left(\frac{P_t}{P_{t-1}}\right), \tag{2}$$

preferido en análisis financiero por su aditividad temporal y simetría aproximada [214]. La volatilidad, medida por la desviación estándar σ de r_t, es la métrica primaria de riesgo de mercado. Para capturar el agrupamiento de volatilidad, Engle [58] propuso el modelo ARCH; Bollerslev [24] lo generalizó al GARCH:

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2, \tag{3}$$

con $\alpha_0 > 0$, $\alpha_i, \beta_j \geq 0$ y $\sum_i \alpha_i + \sum_j \beta_j < 1$ para garantizar estacionariedad. Aunque estos modelos son fundamentales, su capacidad para capturar relaciones no lineales de alta dimensionalidad es limitada —limitación documentada empíricamente para el IPC mexicano por Gómez-Ramos et al. [84]— motivando los enfoques de ML discutidos en el Capítulo II [25, 196].

4.1.3 Valor en Riesgo (VaR) y Expected Shortfall (CVaR)

El **Valor en Riesgo (VaR)** cuantifica la pérdida máxima que una cartera puede experimentar con probabilidad $1-c$ sobre un horizonte determinado [121]: $P(L > \mathrm{VaR}_c) = 1-c$. A pesar de su amplia adopción, el VaR no provee información sobre la severidad de pérdidas que lo exceden y no es siempre una medida coherente en el sentido de Artzner et al. [10].

El **Expected Shortfall (CVaR)** supera estas limitaciones cuantificando la pérdida esperada condicionada a que se exceda el VaR:

$$\mathrm{CVaR}_c(L) = \mathbb{E}[L \mid L > \mathrm{VaR}_c(L)]. \tag{4}$$

El CVaR satisface las propiedades de monotonicidad, invarianza a la traslación, homogeneidad positiva y subaditividad, siendo una medida coherente [10, 158, 185]. En el diseño experimental de este trabajo, el CVaR en ventana de 252 días es una de las cuatro dimensiones del RiskIndex y una de las diez variables del CORE-10.

4.1.4 Riesgo Sistemático, No Sistemático y CAPM

El riesgo total de un activo se descompone en **riesgo sistemático** (no diversificable) y **riesgo idiosincrático** (específico de la empresa, diversificable) [23, 57]. El CAPM formaliza la relación entre riesgo sistemático y rendimiento esperado en equilibrio [140, 171, 198]:

$$\mathbb{E}[R_i] = R_f + \beta_i\big(\mathbb{E}[R_m] - R_f\big), \tag{5}$$

donde $\beta_i = \mathrm{Cov}(R_i, R_m)/\mathrm{Var}(R_m)$ mide la sensibilidad del activo i al mercado y R_f es la tasa libre de riesgo. En este trabajo, β_i se estima de forma rodante (*rolling*) en ventanas de 252 días para capturar su evolución temporal.

Cuadro 2: Interpretación de los coeficientes α y β del CAPM.

Coeficiente	Interpretación	Implicación para el riesgo
$\beta > 1$	Activo más volátil que el mercado	Mayor exposición al riesgo sistemático
$\beta < 1$	Activo menos volátil que el mercado	Menor exposición sistémica
$\beta = 1$	Activo se mueve con el mercado	Exposición neutral
$\alpha > 0$	Rendimiento superior al esperado	Posible ventaja ajustada al riesgo
$\alpha < 0$	Rendimiento inferior al esperado	Destrucción de valor ajustada al riesgo

Aunque el CAPM es fundacional, su capacidad explicativa ha sido cuestionada y extendida mediante modelos multifactoriales [61, 62]. En el contexto predictivo de este trabajo, esta literatura motiva incluir variables que capturen dimensiones de riesgo no contenidas en β —cola, iliquidez, cambios de régimen— consistente con enfoques modernos de *asset pricing* mediante ML [1, 86].

4.1.5 Clasificación del Riesgo y Analogía con Calificadoras

Las agencias calificadoras de riesgo —S&P Global, Moody's, Fitch y HR Ratings en México— asignan categorías ordinales a emisores e instrumentos de deuda para reducir asimetría de información [67, 108, 166, 190]. Su lógica ordinal (grado de inversión $\rightarrow$ especulativo $\rightarrow$ alto riesgo) es conceptualmente análoga a la clasificación de este trabajo (Bajo $\rightarrow$ Medio $\rightarrow$ Alto): en ambos casos se asignan categorías interpretables que permiten decisiones de gestión bajo restricciones de trazabilidad y auditabilidad [76, 111]. La diferencia central es que el sistema propuesto opera sobre variables de mercado observables en tiempo real con fuentes públicas, no sobre análisis fundamental privado, y es evaluado con validación temporal estricta.

Esta analogía no es retórica: funda la elección de un esquema ordinal de tres niveles como variable objetivo, tema que el Capítulo V desarrolla en la construcción del RiskIndex.

4.2 Aprendizaje Automático para Clasificación Financiera: Síntesis

Esta sección sintetiza los principios de ML necesarios para contextualizar el diseño experimental. La revisión crítica de aplicaciones y sus limitaciones se presenta en el Capítulo II; aquí se establecen los conceptos que el Capítulo V operacionaliza.

4.2.1 Paradigmas de Aprendizaje y Tarea de Clasificación

Esta investigación se ubica íntegramente en **aprendizaje supervisado**: dado un conjunto etiquetado $\mathscr{D} = \{(\mathbf{x}_i, y_i)\}_{i=1}^{N}$ donde $\mathbf{x}_i \in \mathbb{R}^d$ son vectores de variables financieras e $y_i \in \{\text{Bajo}, \text{Medio}, \text{Alto}\}$, se busca aprender $f : \mathbb{R}^d \to \mathscr{Y}$ que generalice a observaciones futuras no vistas [21, 97].

Predicción vs. inferencia causal. Este trabajo adopta un objetivo de *predicción* (generalización OOS) en lugar de *inferencia causal*. Esta distinción justifica el protocolo walk-forward y el control estricto de fuga temporal: el modelo debe desempeñarse bien en el futuro, no explicar el pasado [86, 145, 197].

4.2.2 Modelos Base: SVM y MLP

Los dos modelos de referencia en este estudio son la SVM lineal calibrada y la MLP con dimensiones emparejadas. Se presentan en su rol de *baseline*, no como propuesta central.

Máquina de Soporte Vectorial (SVM). Una SVM busca el hiperplano de separación de máximo margen entre clases [45, 97]. En su versión lineal con calibración probabilística de Platt [176], la frontera de decisión es interpretable en el espacio original de variables, pero con capacidad limitada para capturar interacciones no lineales. La calibración convierte los márgenes en probabilidades de clase comparables con los demás modelos.

Red Neuronal Multicapa (MLP). Una MLP consiste en capas de neuronas conectadas en secuencia donde:

$$\mathbf{x}^{(l+1)} = \phi\left(W^{(l)}\mathbf{x}^{(l)} + \mathbf{b}^{(l)}\right), \tag{6}$$

con $W^{(l)}$ pesos, $\mathbf{b}^{(l)}$ sesgos y ϕ función de activación no lineal. El **Teorema de Aproximación Universal** [47, 104, 105] garantiza que una MLP con una capa oculta suficientemente ancha puede aproximar cualquier función continua; sin embargo, esta garantía de existencia no especifica la arquitectura óptima y, fundamentalmente, no asegura trazabilidad del mecanismo de decisión, lo que motiva la búsqueda de alternativas con interpretabilidad intrínseca [189].

Series temporales financieras: particularidades. A diferencia de dominios con datos aproximadamente *i.i.d.*, en finanzas se observan no-estacionariedad, dependencia temporal y cambios de régimen. Particiones aleatorias inducen fuga de información, y la normalización global contamina las particiones de prueba [13, 145]. Ambas prácticas se evitan en el diseño experimental del Capítulo V.

4.2.3 Clasificación Ordinal: Fundamento y Relevancia para la Tarea

La tarea de este trabajo es una **clasificación ordinal**: las etiquetas Bajo, Medio y Alto no son categorías nominales intercambiables, sino niveles con un orden intrínseco. Un error de clasificación Bajo $\rightarrow$ Alto es cualitativamente más grave que un error adyacente Bajo $\rightarrow$ Medio, y esta estructura debe estar presente en la evaluación [75, 91].

La literatura distingue dos estrategias principales para problemas ordinales con redes neuronales. La primera descompone el problema en $K-1$ clasificaciones binarias anidadas [75]. La segunda impone restricciones de consistencia de rango directamente en la función de salida —el marco CORAL y sus extensiones— garantizando que las probabilidades predichas respeten el orden de las clases [33]. Una revisión de 41 conjuntos de datos concluye que los enfoques que explotan explícitamente el orden entre clases superan de forma consistente a los clasificadores nominales aplicados al mismo problema [91].

En este trabajo, la ordinalidad se preserva de dos formas. *Primera*: la variable objetivo (RiskIndex) es continua y se discretiza en tres cuantiles ordenados, garantizando que el orden refleja magnitudes de riesgo reales. *Segunda*: las métricas de evaluación se analizan por clase, con atención explícita a la gravedad relativa de los errores y a la separabilidad de la clase Medio. Las extensiones futuras pueden incorporar funciones de pérdida ordinales explícitas como CORAL para mejorar la calibración en la región transicional.

4.3 Redes Kolmogorov-Arnold: Fundamentos Matemáticos y Arquitectura

Esta sección constituye el núcleo teórico diferenciador del trabajo.

4.3.1 Del Problema 13 de Hilbert al Teorema de Representación

En 1900, David Hilbert presentó su lista de 23 problemas para la matemática del siglo XX [100, 101]. El **decimotercer problema** cuestionaba si toda función algebraica de tres variables podía representarse como superposición de funciones de dos variables [218]. En 1957,

Kolmogorov demostró un resultado mucho más fuerte: toda función continua multivariada se puede representar mediante funciones de *una sola variable* [130].

Teorema IV.1 (Teorema de Superposición de Kolmogorov, 1957)**.** *Sea* $f : [0,1]^n \to \mathbb{R}$ *una función continua. Entonces existen funciones continuas univariadas* $\Phi_q : \mathbb{R} \to \mathbb{R}$ *y* $\psi_{p,q} : [0,1] \to \mathbb{R}$*, para* $q = 0, \ldots, 2n$ *y* $p = 1, \ldots, n$*, tales que:*

$$f(x_1, \ldots, x_n) = \sum_{q=0}^{2n} \Phi_q \left(\sum_{p=1}^{n} \psi_{p,q}(x_p) \right). \tag{7}$$

Arnold [6] y Arnold [7] refinaron el resultado; Sprecher [204] y Sprecher [205] aportaron estructuras más simétricas para las funciones internas; y Braun y Griebel [29] desarrollaron pruebas constructivas que ofrecen una vía explícita para encontrar las representaciones [147, 148, 206].

La implicación fundamental del KART para la teoría de la aproximación es profunda: cualquier función continua compleja en alta dimensión puede descomponerse y representarse fielmente mediante componentes univariadas y sumas, proveyendo el fundamento matemático de una arquitectura neuronal alternativa [147].

Del KART a la arquitectura computable: el rol de los B-splines. El KART original garantiza la *existencia* de funciones univariadas de representación, pero no que esas funciones sean suaves, diferenciables ni numéricamente estables para el aprendizaje por gradiente. Esta brecha entre el resultado de existencia matemático y la implementación práctica en una red neuronal es la que las B-splines resuelven: parametrizando cada función $\phi(x)$ como una combinación lineal de funciones base de B-spline, las KAN obtienen funciones univariadas con suavidad C^{p-1} garantizada, soporte local (cada base afecta solo una región del dominio) y coeficientes aprendibles de forma estable mediante descenso de gradiente [27]. El puente del KART a la arquitectura computable no es trivial: es precisamente este puente el que hace de KAN una arquitectura prácticamente implementable, no solo un resultado de existencia.

Nota epistemológica: KART en datos financieros no *i.i.d.*. El KART se formula para funciones continuas sobre dominios compactos. En el contexto financiero, esta hipótesis de continuidad no se adopta como descripción literal del proceso generador de datos, sino como *principio constructivo*: motiva arquitecturas donde la no linealidad se concentra en transformaciones univariadas inspeccionables. El rol del teorema en este trabajo es proveer

una estructura de modelización que favorece trazabilidad funcional, no afirmar que el mercado satisface condiciones analíticas exactas [12, 145, 214].

4.3.2 Arquitectura Matemática de las Redes KAN

Liu et al. [144] propusieron las Redes Kolmogorov-Arnold como arquitectura neuronal que materializa la estructura del KART. Su principio fundamental invierte el paradigma de las MLP: mientras que una MLP tiene activaciones no lineales *fijas* en los nodos y pesos lineales aprendibles en las aristas, una KAN tiene funciones de activación univariadas *aprendibles* en las aristas, y los nodos realizan únicamente sumas:

$$x_j^{(l+1)} = \sum_{i=1}^{n_l} \phi_{j,i}^{(l)} \left(x_i^{(l)} \right), \tag{8}$$

donde $\phi_{j,i}^{(l)}$ es la función de activación aprendible en la arista que conecta la neurona i de la capa l con la neurona j de la capa $l+1$.

Parametrización mediante B-splines. Cada función de activación $\phi(x)$ se parametriza como combinación lineal de funciones base de B-spline:

$$\phi(x) = \sum_{k=1}^{G} c_k B_k(x; \mathbf{t}), \tag{9}$$

donde c_k son los coeficientes aprendibles, G el número de funciones base y $\mathbf{t}$ el vector de nudos. Los B-splines se definen recursivamente mediante la fórmula de Cox–de Boor [27]:

Para grado $p = 0$:

$$B_{k,0}(x) = \begin{cases} 1, & t_k \leq x < t_{k+1}, \\ 0, & \text{en otro caso.} \end{cases} \tag{10}$$

Para grado $p \geq 1$:

$$B_{k,p}(x) = \frac{x - t_k}{t_{k+p} - t_k} B_{k,p-1}(x) + \frac{t_{k+p+1} - x}{t_{k+p+1} - t_{k+1}} B_{k+1,p-1}(x), \tag{11}$$

donde los cocientes con denominador cero se consideran nulos. Esta construcción induce dos propiedades críticas para la interpretabilidad:

- **Soporte local:** $B_{k,p}(x)$ es distinto de cero solo en $[t_k, t_{k+p+1})$, lo que permite inspeccionar cambios funcionales en regiones específicas del dominio. En el contexto financiero, el modelo puede aprender respuestas distintas para valores de volatilidad baja, moderada o extrema, sin que dichas regiones se contaminen mutuamente.

- **Suavidad controlada:** con nudos simples, la spline resultante es C^{p-1}; nudos con multiplicidad r reducen la continuidad en ese punto a C^{p-r}. La implementación de este trabajo emplea grado $p = 3$ (cúbico), con $n = 10$ nudos internos, produciendo funciones con suavidad C^2 y $G = 14$ funciones base por arista [27].

Contabilidad de parámetros y asimetría frente a MLP. Para una capa KAN que mapea $\mathbb{R}^{n_l} \to \mathbb{R}^{n_{l+1}}$ con G coeficientes por arista y con conexión residual (*base*), el número de parámetros de arista es:

$$\#\theta_{\text{KAN}} \approx n_l \cdot n_{l+1} \cdot (G+2) + n_{l+1}, \tag{12}$$

frente a $\#\theta_{\text{MLP}} = n_l \cdot n_{l+1} + n_{l+1}$ en una capa densa. Esta diferencia tiene consecuencias concretas: con la arquitectura [10→32→3] empleada en este trabajo y $G = 14$, el sistema KAN cuenta con aproximadamente 6,691 parámetros frente a los ≈451 del MLP con dimensiones emparejadas —una relación de ∼15:1. La comparación en este trabajo se denomina "*arch-matched*" porque iguala las *dimensiones* de capas, no la capacidad paramétrica; la justicia de la comparación descansa en la igualdad de información disponible (mismo CORE-10), la igualdad de particiones temporales y la evaluación exclusiva sobre datos OOS, donde el sobreajuste penaliza de forma independiente del número de parámetros [144, 221, 226].

Esta asimetría no compromete la justicia de la comparación: el criterio de igualdad relevante para la pregunta de investigación es la información disponible y el protocolo de validación, no el número de parámetros. Igualar parámetros entre KAN y MLP requeriría reducir los nudos spline a valores que eliminan la capacidad de inspeccionabilidad de las funciones $\phi(x)$ —exactamente la propiedad que justifica la investigación—. La comparación *arch-matched* es la única que pone a prueba la hipótesis central: ¿puede una arquitectura de interpretabilidad intrínseca igualar el desempeño de una sin esa propiedad, bajo el mismo volumen de información disponible? [52, 97].

4.3.3 Propiedades Matemáticas Clave de las KAN

Aproximación universal por construcción. Las KAN son aproximadores universales por consecuencia directa del KART, que garantiza representación *exacta*, no solo aproximada,

de toda función continua. Esto las diferencia de las MLP, cuya aproximación universal es de naturaleza existencial: garantiza que el aproximador existe, pero no cómo encontrarlo de forma eficiente [144, 221].

Eficiencia paramétrica. Si la función objetivo tiene estructura de composición dispersa, las KAN pueden capturarla haciendo que algunas $\phi(x)$ degeneren en funciones simples. Análisis teóricos sobre el sesgo espectral y las leyes de escalamiento de las KAN sugieren que pueden superar a las MLP en precisión con menos parámetros para funciones con estructura de composición [144, 221, 226].

Interpretabilidad intrínseca: soporte local y aditividad. La interpretabilidad descansa en dos pilares matemáticos: (i) el soporte local de los B-splines, que permite inspeccionar transformaciones en regiones específicas del dominio, y (ii) la estructura aditiva por nodo, que hace que cada neurona sea un modelo aditivo inspeccionable. Estas propiedades emergen de la arquitectura: no son una aproximación externa del modelo, sino su mecanismo de decisión [27, 144, 189].

KAN 2.0 y extensiones simbólicas. Liu et al. [143] extendieron el marco para integrar estructuras simbólicas y herramientas para descubrimiento científico. Si bien KAN 2.0 no es esencial para la implementación empírica de este trabajo, consolida la relevancia del paradigma en aplicaciones que demandan alta interpretabilidad y trazabilidad funcional.

4.4 Evaluación de Clasificadores: Métricas y Validación Temporal

4.4.1 Métricas de Desempeño para Clasificación Multiclase

La evaluación del clasificador se apoya en métricas derivadas de la matriz de confusión, que tabula predicciones contra etiquetas reales para cada clase [21, 97]:

$$\text{Accuracy} = \frac{\sum_k \text{TP}_k}{N}, \tag{13}$$

$$\text{Precision}_k = \frac{\text{TP}_k}{\text{TP}_k + \text{FP}_k}, \tag{14}$$

$$\text{Recall}_k = \frac{\text{TP}_k}{\text{TP}_k + \text{FN}_k}, \tag{15}$$

$$\text{F1}_k = 2 \cdot \frac{\text{Precision}_k \cdot \text{Recall}_k}{\text{Precision}_k + \text{Recall}_k}. \tag{16}$$

El **F1 macro** promedia el F1 por clase sin ponderar por frecuencia, siendo la métrica primaria por dos razones: (i) trata cada clase como igualmente importante, incluyendo la clase Medio, que es la más difícil y la de mayor consecuencia operativa; y (ii) penaliza de forma simétrica errores de precisión y recall [179, 201].

Curva ROC y AUC en multiclase. En clasificación multiclase se emplea la estrategia *one-vs-rest* (OvR) y se reporta el AUC macro, que promedia la separabilidad entre clases sin ponderar por frecuencia [63, 93]. Esta elección es coherente con la estructura ordinal del problema: la ambigüedad de la clase Medio —sistemáticamente la de menor AUC en los tres modelos evaluados— merece visibilidad explícita.

4.4.2 Validación Temporal: Walk-Forward con Purga y Embargo

Los datos financieros violan el supuesto de *i.i.d.* que subyace a la validación cruzada estándar. Particiones aleatorias inducen *look-ahead bias*: el modelo puede haber sido entrenado con información cronológicamente posterior a los períodos que evalúa [12, 145, 208]. El **Walk-Forward** preserva la causalidad temporal: el modelo se entrena sobre una ventana inicial, se evalúa en el período siguiente, y la ventana se desplaza iterativamente, simulando el uso operativo.

Las técnicas de **purga** y **embargo** refinan esta validación eliminando observaciones de entrenamiento que comparten información temporal con el conjunto de prueba, ya sea por solapamiento de ventanas o por etiquetas construidas con horizontes futuros [13, 145]. Un análisis comparativo en entornos controlados sintéticos concluye que el protocolo con purga y embargo es marcadamente superior a la validación cruzada estándar y al walk-forward simple para mitigar el sobreajuste en modelos financieros [5]. El protocolo Walk-Forward PRO con

purga y embargo implementado en este trabajo responde directamente a estos tres vectores de amenaza a la validez.

4.4.3 Comparación Estadística entre Clasificadores

Determinar si las diferencias de desempeño entre modelos son estadísticamente significativas requiere pruebas formales [52, 54]. En este trabajo se emplea la **prueba de los rangos con signo de Wilcoxon** entre pares de modelos, que compara las distribuciones de métricas por pliegue sin asumir normalidad de las diferencias. La evidencia inferencial se complementa con **intervalos bootstrap** para la estimación del tamaño del efecto [19, 56]. Esta combinación permite distinguir mejoras consistentes a través de los 13 pliegues temporales de fluctuaciones aisladas.

4.5 Interpretabilidad Matemática en Contextos de Alto Impacto

4.5.1 Taxonomía de la Interpretabilidad

La interpretabilidad en ML se refiere al grado en que un ser humano puede comprender la causa de una decisión del modelo [55, 165]. Se distinguen dos dimensiones:

- **Global:** comprensión del comportamiento completo del modelo. Permite auditoría holística y validación de coherencia económica.

- **Local:** explicación de una predicción individual. Facilita justificación ante comités y reguladores.

En finanzas, la interpretabilidad trasciende el interés académico: es un requisito funcional y regulatorio [76, 189]. Los marcos de IA responsable que IOSCO y la CNBV han incorporado en sus agendas hacen explícita la necesidad de sistemas auditables en decisiones de gestión de riesgo.

4.5.2 *Post-hoc* vs. Interpretabilidad Intrínseca

Las técnicas *post-hoc* como SHAP [151] y LIME [184] aproximan la lógica de un modelo opaco sin modificar su estructura. Su valor es indiscutible cuando el modelo de caja negra es irrenunciable; sin embargo, Rudin [189] argumenta que en decisiones de alto impacto la alternativa correcta es construir modelos con interpretabilidad intrínseca. La diferencia es cualitativa: una explicación *post-hoc* puede ser una buena aproximación del comportamiento del modelo sin reproducir su lógica exacta; una propiedad intrínseca del modelo *es* la lógica.

4.5.3 Interpretabilidad de las KAN: Operacionalización

En este trabajo, la interpretabilidad de KAN se operacionaliza en tres niveles verificables [55]:

1. **Importancia global por variable:** identificada mediante la norma de la función $\phi(x)$ aprendida en la primera capa, proveyendo un ranking directo de la contribución de cada variable sin requerir métodos externos.

2. **Análisis funcional por variable:** visualización directa de la curva $\phi_{j,i}(x)$ por arista, que permite identificar monotonías, umbrales, saturaciones y regiones de alta sensibilidad. El soporte local de los B-splines garantiza que estas propiedades son del modelo, no aproximaciones.

3. **Coherencia económica:** contraste cualitativo entre la forma de $\phi(x)$ y el sentido económico esperado de cada variable. La consistencia entre estructura aprendida y teoría financiera es evidencia de validez de constructo [46, 197].

Este enfoque prioriza interpretabilidad intrínseca, sin negar el valor complementario de SHAP o LIME en etapas de diagnóstico [89, 165].

4.6 Distinciones Conceptuales Necesarias para Interpretar este Estudio

Para interpretar correctamente los resultados reportados en los Capítulos VI e VII, se establecen cuatro distinciones que delimitan el alcance inferencial del trabajo y que, de no mantenerse explícitas, pueden generar expectativas incorrectas sobre lo que el sistema puede y no puede hacer.

1. Integridad temporal $\neq$ Validez de constructo. El protocolo Walk-Forward PRO con purga, embargo y $gap = H_{\text{futuro}}$ garantiza que el modelo no observa información futura durante el entrenamiento, el escalamiento ni el etiquetado (*integridad temporal*). Esto es independiente de si el RiskIndex captura el "riesgo verdadero" de las emisoras (*validez de constructo*). Este trabajo asegura la primera propiedad; sobre la segunda adopta una postura explícitamente modesta: el RiskIndex es una medida operativa de riesgo relativo de investigación, construida con variables de mercado observables, pero no equivale a un estándar exógeno certificado ni a un índice de riesgo oficial. El RiskIndex comparte componentes con algunos predictores del CORE-10 ($|\beta|$, $|\text{CVaR}|$, $1 - R^2$), lo que constituye una amenaza parcial a la validez de constructo documentada en el Capítulo VIII: el objetivo comparte información con el

espacio explicativo. Esta condición no introduce fuga temporal —los umbrales se estiman por pliegue y `VolFut` nunca entra como predictor— pero obliga a interpretar los resultados como clasificación predictiva de riesgo relativo, no como inferencia sobre el riesgo absoluto de las emisoras [46, 145, 197].

2. Interpretabilidad intrínseca $\neq$ Explicabilidad *post-hoc*. La interpretabilidad de KAN emerge de su arquitectura: las funciones $\varphi(x)$ aprendidas *son* el mecanismo exacto de decisión. La explicabilidad *post-hoc* (SHAP, LIME) produce aproximaciones del mecanismo de un modelo diferente, con fidelidad que depende de supuestos propios del método explicativo que raramente son transparentes para el regulador. Para efectos de gobernanza algorítmica, la distinción es cualitativa: una es la lógica exacta, la otra es una representación local aproximada de ella [55, 117, 189].

3. Clasificación de riesgo relativo $\neq$ Inferencia causal. Los resultados demuestran que KAN puede *predecir* la categoría ordinal del RiskIndex con precisión OOS competitiva. No demuestran que las variables del CORE-10 *causen* el nivel de riesgo, ni que las relaciones aprendidas por el modelo sean las causas estructurales de los episodios de estrés. El sistema es un clasificador predictivo operado bajo validación temporal rigurosa, no un modelo econométrico de identificación causal [86, 145, 197].

4. Señal de riesgo $\neq$ Motor de *alpha*. El sistema identifica perfiles de riesgo accionario, no estrategias de maximización de retorno. Un filtro que excluye emisoras de alto riesgo tiene, como consecuencia esperada, un retorno absoluto por debajo del benchmark equiponderado en períodos alcistas: exactamente lo que se observa en el backtest reportado en el Capítulo VI. El retorno nominal positivo del backtest OOS es evidencia de contenido informativo real de la señal de clasificación, no de ventaja ajustada por riesgo sobre el mercado. Cualquier uso en position sizing dinámico requiere una etapa adicional de calibración probabilística antes del despliegue [30, 90, 176].

V. Metodología

Este capítulo documenta el procedimiento metodológico con precisión y trazabilidad suficientes para garantizar replicabilidad y auditoría del diseño experimental. El enfoque adoptado es **cuantitativo**, basado en series numéricas diarias de mercado. El principio central es una comparación ***apples-to-apples***: mismo universo de activos, mismas variables predictoras, mismo esquema de validación temporal, mismo procedimiento de escalamiento, mismo mecanismo de etiquetado por pliegue y mismas métricas de evaluación para los tres modelos contrastados [12, 145].

1. **Construcción del dataset.** Selección de 30 emisoras representativas de la BMV; descarga de precios ajustados, benchmark (IPC) y tasa libre de riesgo (CETES); cálculo de rendimientos y generación de variables financieras en escala original.
2. **Depuración para modelado.** Eliminación controlada de observaciones con información insuficiente por ventanas móviles y horizonte futuro; construcción de la muestra efectiva de 40,649 observaciones utilizada por el protocolo temporal.
3. **Construcción de etiquetas de riesgo.** Definición del RiskIndex como índice multivariado continuo de cuatro dimensiones; discretización ordinal en tres clases (Bajo / Medio / Alto) mediante cuantiles estimados *estrictamente dentro de cada partición de entrenamiento*, sin contacto con información futura.
4. **Diseño experimental temporal.** Implementación del protocolo Walk-Forward PRO con purga y embargo: 13 particiones cronológicas con ventana de entrenamiento creciente, período de purga igual al horizonte de etiquetado, ventana de validación interna y período de embargo previo a la prueba fuera de muestra.
5. **Modelado *apples-to-apples*.** Entrenamiento de KAN, MLP y SVM lineal calibrado con el mismo conjunto CORE-10, el mismo escalamiento por pliegue y las mismas etiquetas construidas por pliegue; sin ningún acceso a datos fuera de muestra durante el entrenamiento o la selección de hiperparámetros.
6. **Evaluación metodológica.** Cálculo de métricas multiclase (F1 macro, AUC OvR, matrices de confusión); pruebas estadísticas pareadas de Wilcoxon e intervalos bootstrap por pliegue; registro de eficiencia computacional; y verificación económica ilustrativa mediante backtesting con predicciones OOS.

Sobre la comparabilidad y la asimetría paramétrica. La comparabilidad *apples-to-apples* se operacionaliza como igualdad de información disponible y particiones temporales, con control explícito sobre los pasos que suelen inducir fuga de información (etiquetado, escalamiento, calibración y selección). La comparación se denomina "*arch-matched*" porque iguala las dimensiones de capas de los modelos neuronales [10→32→3], no el número de parámetros: KAN aloja ≈6,691 parámetros frente a los ≈451 del MLP, dado que cada conexión KAN almacena $G = 14$ coeficientes spline en lugar de un solo peso escalar. La justicia de la comparación descansa en la igualdad de información disponible y en la evaluación exclusiva sobre datos OOS, donde el sobreajuste penaliza de forma independiente del número de parámetros [52, 97].

5.1 Diseño General del Estudio

La tarea principal es una **clasificación supervisada multiclase ordinal**: asignar niveles de riesgo de inversión (*Bajo*, *Medio*, *Alto*) a observaciones diarias de acciones listadas en la BMV. Las etiquetas no son categorías nominales intercambiables: poseen un orden intrínseco en el que un error Bajo → Alto es cualitativamente más grave que un error adyacente, lo que distingue al problema de una clasificación multiclase convencional [75, 91]. A diferencia de enfoques con etiquetas exógenas (calificaciones crediticias), el riesgo se define de forma **endógena** a partir de la dinámica futura del activo y de medidas de exposición sistemática, riesgo extremo y componente idiosincrático. Este enfoque mantiene coherencia econométrica y vincula la etiqueta con magnitudes financieras observables [145, 158, 214].

5.2 Datos: Universo, Período y Estructura

5.2.1 Fuentes de datos y benchmark

Los precios y volúmenes se obtuvieron mediante la librería `yfinance` [9], empleando precios **ajustados** por *splits* y dividendos para el cálculo de rendimientos logarítmicos. Como benchmark se utilizó el índice S&P/BMV IPC (`^MXX`) [225]. La tasa libre de riesgo diaria se aproximó con la tasa de CETES a 28 días, consultada mediante el API REST del Sistema de Información Económica (SIE) del Banco de México [17], transformada a tasa diaria por capitalización con *forward-fill* para días hábiles subsecuentes.

5.2.2 Universo de emisoras y control de sesgo de supervivencia

El universo comprende 30 emisoras representativas y líquidas de la BMV. La selección se realizó con base en tres criterios definidos antes del análisis: (i) cotización continua en la

BMV durante el período 2020–2025 sin interrupciones operativas superiores a 5 días hábiles consecutivos, (ii) cobertura sectorial diversificada que representara los principales sectores de la economía mexicana cotizados en bolsa, y (iii) suficiente liquidez para el cálculo estable de la razón de Amihud en ventanas de 60 días. La Tabla 3 reporta el *ticker*, empresa y giro operativo de cada emisora.

Cuadro 3: Universo de 30 emisoras de la BMV utilizadas en el estudio.

Ticker	**Empresa**	**Giro Operativo**
ASURB.MX	Grupo Aeroportuario del Sureste	Aeropuertos
GAPB.MX	Grupo Aeroportuario del Pacífico	Aeropuertos
OMAB.MX	Grupo Aeroportuario del Centro Norte	Aeropuertos
BIMBOA.MX	Grupo Bimbo	Consumo / Alimentos
GRUMAB.MX	Gruma	Consumo / Alimentos
KIMBERA.MX	Kimberly-Clark de México	Consumo / Hogar
LIVEPOLC-1.MX	El Puerto de Liverpool	Consumo / Retail
WALMEX.MX	Walmart de México y Centroamérica	Consumo / Retail
ALSEA.MX	Alsea	Consumo / Servicios
AC.MX	Arca Continental	Consumo / Bebidas
CUERVO.MX	Becle (José Cuervo)	Consumo / Bebidas
AGUA.MX	Grupo Rotoplas	Consumo / Hogar
TLEVISACPO.MX	Grupo Televisa	Telecomunicaciones / Medios
MEGACPO.MX	Megacable Holdings	Telecomunicaciones
AMXB.MX	América Móvil	Telecomunicaciones
CEMEXCPO.MX	CEMEX	Industrial / Materiales
GCC.MX	GCC	Industrial / Materiales
ORBIA.MX	Orbia Advance Corporation	Industrial / Químicos
ALPEKA.MX	Alpek	Industrial / Petroquímicos
NEMAKA.MX	Nemak	Industrial / Autopartes
GMEXICOB.MX	Grupo México	Minería / Industrial
PE&OLES.MX	Industrias Peñoles	Minería / Metales
GCARSOA1.MX	Grupo Carso	Conglomerado
FEMSAUBD.MX	FEMSA	Conglomerado
GFNORTEO.MX	Grupo Financiero Banorte	Financiero
Q.MX	Quálitas Controladora	Financiero (seguros)
VOLARA.MX	Volaris	Transporte
PINFRA.MX	Promotora y Operadora de Infraestructura	Infraestructura
LABB.MX	Genomma Lab Internacional	Consumo / Salud
ELEKTRA.MX	Grupo Elektra	Retail / Servicios financieros

Nota sobre sesgo de supervivencia. El universo fue construido con emisoras activas al momento del análisis (2026). Las empresas que cotizaban en 2020 pero que fueron descontinuadas, suspendidas o fusionadas con anterioridad no están representadas, lo que introduce un sesgo de supervivencia moderado. Este sesgo tiende a sobreestimar el desempeño histórico

de las categorías de menor riesgo, dado que las emisoras que sobrevivieron el período son, en promedio, las de menor probabilidad de evento adverso. Sin embargo, el sesgo afecta por igual a los tres modelos comparados y no compromete la validez de la comparación entre arquitecturas; sí debe considerarse al interpretar los valores de desempeño absoluto y al extrapolar las conclusiones a emisoras con perfiles históricos más adversos.

5.2.3 Estructura y resumen del dataset

Para garantizar reproducibilidad y separación entre ingeniería de variables y experimentación, se generaron dos artefactos: `dataset_bmv_final_raw.csv` (escala original, sin estandarización) y `dataset_bmv_final_scaled.csv` (estandarizado globalmente, para inspección). El entrenamiento emplea **escalamiento por pliegue** (Sección 5.5) para evitar fuga de información [21, 97].

Cuadro 4: Resumen del dataset RAW y de la muestra efectiva utilizada para modelado.

Elemento	**Dataset RAW**	**Muestra efectiva**
Número de emisoras	30	30
Observaciones	43,470	40,649
Período	2020-01-03 a 2025-10-03	2020-03-27 a 2025-10-03
Fechas únicas	1,449	1,391
Pliegues Walk-Forward PRO	—	13
Observaciones OOS totales	—	16,133

Con esta base de datos y su estructura temporal verificada —43,470 observaciones brutas reducidas a 40,649 efectivas tras la eliminación controlada de filas con información insuficiente por ventanas móviles y horizonte futuro— la sección siguiente detalla las variables financieras que se calculan sobre ella y los criterios de selección del conjunto CORE-10.

5.3 Ingeniería y Selección de Variables Financieras (CORE-10)

Sea $P_{i,t}$ el precio de cierre ajustado del activo i en el día t. Los rendimientos logarítmicos diarios se definen como:

$$r_{i,t} = \log(P_{i,t}) - \log(P_{i,t-1}), \tag{17}$$

elección estándar por aditividad temporal y estabilidad de escala [214].

El vector de entrada utiliza $d = 10$ variables financieras (**CORE-10**) diseñadas para capturar riesgo sistemático, desempeño relativo, riesgo total, asimetría, riesgo de cola y liquidez. La

Tabla 5 resume las variables exactas, compartidas por **todos** los modelos evaluados.

Cuadro 5: Variables de entrada al modelo (CORE-10) y su interpretación financiera.

Variable	Ventana	Interpretación
`Beta`	252 d	Exposición al mercado (CAPM rolling), riesgo sistemático [198]
`Alpha`	252 d	Rendimiento anormal vs. benchmark [198]
`R2`	252 d	Proporción de varianza explicada por el mercado
`r_log`	1 d	Rendimiento logarítmico diario [214]
`Vol_20`	20 d	Volatilidad histórica de corto plazo
`Vol_60`	60 d	Volatilidad histórica de mediano plazo
`Vol_252`	252 d	Volatilidad histórica anual aproximada
`DownsideVol_60`	60 d	Volatilidad a la baja, riesgo asimétrico [203]
`CVaR_5`	252 d	Expected Shortfall al 5%, medida coherente de riesgo de cola [10, 185]
`Amihud_60`	60 d	Iliquidez de Amihud, fricción de mercado [1]

El CORE-10 como decisión de diseño deliberada. La restricción a diez variables endógenas de mercado no es una omisión de factores macroeconómicos externos —tipo de cambio, tasas de interés, indicadores de actividad económica—, sino una decisión de diseño con fundamento explícito. Los factores macro se manifiestan *endógenamente* a través del CORE-10: el Beta rodante captura la sensibilidad cambiante al mercado (que internaliza choques sistémicos); el CVaR absorbe episodios de estrés de cualquier origen que se materializan en colas extremas; y la volatilidad en múltiples horizontes refleja la respuesta de cada emisora a perturbaciones macro y de microestructura. Este enfoque mantiene el pipeline reproducible con fuentes públicas sin requerir series adicionales y garantiza disponibilidad en tiempo real para cualquier institución financiera mexicana [1, 86, 214].

5.3.1 Cálculo de Alpha, Beta y R^2 (CAPM rolling)

Para capturar la evolución temporal del riesgo sistemático, se estima un CAPM en ventanas móviles de 252 días sobre rendimientos en exceso:

$$r_{i,t} - r_{f,t} = \alpha_{i,t} + \beta_{i,t}(r_{m,t} - r_{f,t}) + \varepsilon_{i,t}, \tag{18}$$

donde $r_{m,t}$ es el rendimiento del IPC [225] y $r_{f,t}$ es la tasa diaria de CETES [17]. Los coeficientes α y β se obtienen mediante estimadores rolling de momentos equivalentes a MCO con intercepto. El R^2 se calcula como el cuadrado de la correlación rolling entre excesos

de rendimiento del activo y del mercado.

5.3.2 Riesgo asimétrico y de cola

La volatilidad a la baja (`DownsideVol_60`) restringe la desviación estándar a rendimientos negativos en ventana de 60 días, penalizando solo el riesgo de pérdida [203]. El riesgo de cola se aproxima con CVaR histórico al 5% en ventana de 252 días:

$$\mathrm{CVaR}_{0.05}(r) = -\mathbb{E}[\, r \mid r \leq Q_{0.05}(r)\,], \tag{19}$$

siguiendo las definiciones de medida coherente de riesgo de Artzner et al. [10] y Rockafellar y Uryasev [185].

5.3.3 Proxy de liquidez (Amihud)

La iliquidez de Amihud captura la fricción de mercado como cociente entre variación absoluta del retorno y volumen transado, agregado en ventana de 60 días:

$$\mathrm{ILLIQ}_{i,t} \propto \frac{|r_{i,t}|}{\mathrm{Volumen}_{i,t}}, \tag{20}$$

medida ampliamente validada en la literatura de liquidez y retornos [1].

5.4 Construcción de Etiquetas de Riesgo

5.4.1 Volatilidad futura realizada

Un componente esencial del índice de riesgo es la volatilidad futura realizada, definida como la desviación estándar de los rendimientos en los $H = 60$ días hábiles posteriores a la observación:

$$\mathrm{VolFut}_{i,t} = \mathrm{sd}(\{r_{i,t+1}, r_{i,t+2}, \ldots, r_{i,t+H}\})\,. \tag{21}$$

Esta variable ubica la señal en el tiempo t pero utiliza información de $t+1$ a $t+H$; por ello se usa **exclusivamente para el etiquetado** y nunca como variable predictora. El horizonte $H = 60$ días hábiles (equivalente a un trimestre calendario de mercado) balancea señal —suficiente varianza futura para discriminar clases— con aplicabilidad operativa —clasificaciones relevantes para decisiones trimestrales y ciclos de rebalanceo de SIEFORES.

5.4.2 Índice de riesgo multivariado (RiskIndex)

La variable objetivo se construye mediante un índice continuo que integra **cuatro dimensiones** de riesgo accionario. La estandarización se realiza **por fecha** (transversal), respecto a la distribución de emisoras en cada día t. Sea $\mathrm{z}_t(x_{i,t}) = (x_{i,t} - \mu_t)/\sigma_t$ la estandarización *cross-sectional*. El índice se define como:

$$\mathrm{RiskIndex}_{i,t} = \mathrm{z}_t(\mathrm{VolFut}_{i,t}) + \mathrm{z}_t(|\beta_{i,t}|) + \mathrm{z}_t(|\mathrm{CVaR}_{i,t}|) + \mathrm{z}_t(1 - R^2_{i,t}). \tag{22}$$

Este diseño produce un *score* de riesgo relativo transversal que integra: volatilidad futura realizada (señal prospectiva), exposición sistemática (riesgo de mercado), riesgo de cola (pérdida esperada en el percentil extremo) y componente idiosincrático (fracción de riesgo no explicada por el mercado) [145, 158].

La elección de estos cuatro componentes tiene lógica financiera explícita: son las cuatro dimensiones del riesgo que la literatura establece como fundamentales y no redundantes entre sí [23, 121, 158, 214]. VolFut captura el riesgo futuro **realizado** —lo que efectivamente ocurrirá en el horizonte H— como señal de la magnitud de exposición. $|\beta|$ rodante captura el riesgo **sistemático** vigente al momento de la clasificación, con sensibilidad temporal al cambio de régimen. |CVaR| captura el riesgo de **cola**, la dimensión que los modelos paramétricos típicamente subestiman. $1 - R^2$ captura el componente **idiosincrático** —lo que el mercado no explica— que concentra los eventos de riesgo específico de empresa. Un activo puede tener beta alta pero CVaR bajo (exposición sistemática sin estrés de cola), o beta baja pero $(1 - R^2)$ alto (riesgo idiosincrático elevado): ningún componente individual captura el perfil de riesgo integral, y su combinación reduce la circularidad conceptual respecto al CORE-10 al limitar la superposición directa.

Alcance inferencial del objetivo y posible circularidad conceptual. El RiskIndex incorpora magnitudes relacionadas con algunos predictores del CORE-10 ($|\beta|$, |CVaR|, $1 - R^2$), lo que constituye una **amenaza a la validez de constructo**: la definición operacional del objetivo comparte componentes con el espacio de variables explicativas [46, 197]. Esta situación **no introduce fuga temporal**: VolFut se usa exclusivamente en el etiquetado y nunca como predictor; los umbrales de discretización se estiman por pliegue sobre TRAIN_CORE. El modelo no observa información futura como predictor (*no look-ahead*). No obstante, el objetivo debe interpretarse como **clasificación predictiva de riesgo relativo**, no como inferencia causal sobre el riesgo absoluto. Esta limitación conceptual se desarrolla en el Capítulo VIII como la

principal extensión de alto impacto: un índice de riesgo exógeno, separado limpiamente del espacio de predictores, eliminaría la circularidad [145, 197].

5.4.3 Discretización ordinal con cuantiles por pliegue (anti look-ahead)

El score continuo se discretiza en tres clases mediante cuantiles $q_{1/3}$ y $q_{2/3}$ estimados **en cada pliegue** usando únicamente el subconjunto TRAIN_CORE:

$$q_{\text{low}} = Q_{1/3}(S \mid \text{TRAIN_CORE}), \qquad q_{\text{high}} = Q_{2/3}(S \mid \text{TRAIN_CORE}), \tag{23}$$

y las etiquetas se asignan como:

- **Bajo**: $S_{i,t} \leq q_{\text{low}}$,
- **Medio**: $q_{\text{low}} < S_{i,t} \leq q_{\text{high}}$,
- **Alto**: $S_{i,t} > q_{\text{high}}$.

Este procedimiento garantiza que la regla de etiquetado aplicada a datos fuera de muestra depende únicamente de información de entrenamiento del pliegue [145].

5.5 Preprocesamiento, Depuración y Estandarización

5.5.1 Eliminación controlada de faltantes por construcción temporal

Las variables con ventanas móviles largas (hasta 252 días) y el horizonte futuro ($H = 60$) generan observaciones sin información suficiente al inicio del período. Estas se eliminan de forma controlada, junto con cualquier fila con valores faltantes en CORE-10 o en el RiskIndex, para construir la muestra efectiva (Tabla 4).

5.5.2 Estandarización por pliegue

El entrenamiento emplea un `StandardScaler` ajustado **exclusivamente en TRAIN_CORE** de cada pliegue y aplicado en modo transformación a VAL y OOS. El escalamiento global —práctica común pero incorrecta en series temporales— contamina los conjuntos de prueba con estadísticos del período futuro, inflando artificialmente el desempeño aparente [21, 97, 120].

5.5.3 Control explícito de columnas con información futura

Las columnas asociadas al etiquetado o al futuro (`RiskIndex`, `VolFut`, retornos futuros) no forman parte del vector de entrada bajo ninguna condición. El pipeline verifica este control en tiempo de ejecución antes de iniciar el walk-forward, lanzando un error si alguna columna del conjunto de futuros se cuela en el vector de características [145].

5.6 Protocolo Experimental: Walk-Forward PRO con Purga y Embargo

5.6.1 Motivación

En series temporales financieras, particiones aleatorias o K-Fold estándar violan la causalidad temporal e inducen estimaciones optimistas del desempeño fuera de muestra [12, 145, 208]. El problema se amplifica cuando existen ventanas móviles y etiquetas construidas con información futura. Un análisis comparativo reciente en entornos controlados sintéticos demuestra que el protocolo con purga y embargo es marcadamente superior al walk-forward simple y a la validación cruzada estándar para mitigar el sobreajuste en modelos financieros [5]. Por ello se adopta un protocolo Walk-Forward que respeta el orden cronológico, con purga y embargo para reducir la dependencia residual entre particiones [13, 145].

5.6.2 Esquema Walk-Forward PRO con validación temporal interna

La Figura 1 ilustra el esquema. En cada pliegue se definen cinco zonas no solapadas:

- **TRAIN_CORE:** ventana de entrenamiento *expanding*; se usa para ajustar el escalador, estimar umbrales de etiquetado y entrenar los modelos.

- **PURGE:** bloque posterior a TRAIN_CORE eliminado para reducir traslape informacional inducido por ventanas móviles y etiquetas con horizonte futuro.

- **VAL:** ventana de validación temporal interna usada para *early stopping* y selección de hiperparámetros, sin contacto con OOS.

- **GAP + EMB:** separación adicional que combina *gap* cronológico y embargo, bloqueando cualquier filtración de información entre VAL y OOS.

- **OOS (TEST):** ventana de prueba fuera de muestra, utilizada **exclusivamente** para evaluación; nunca vista por el modelo durante entrenamiento o selección.

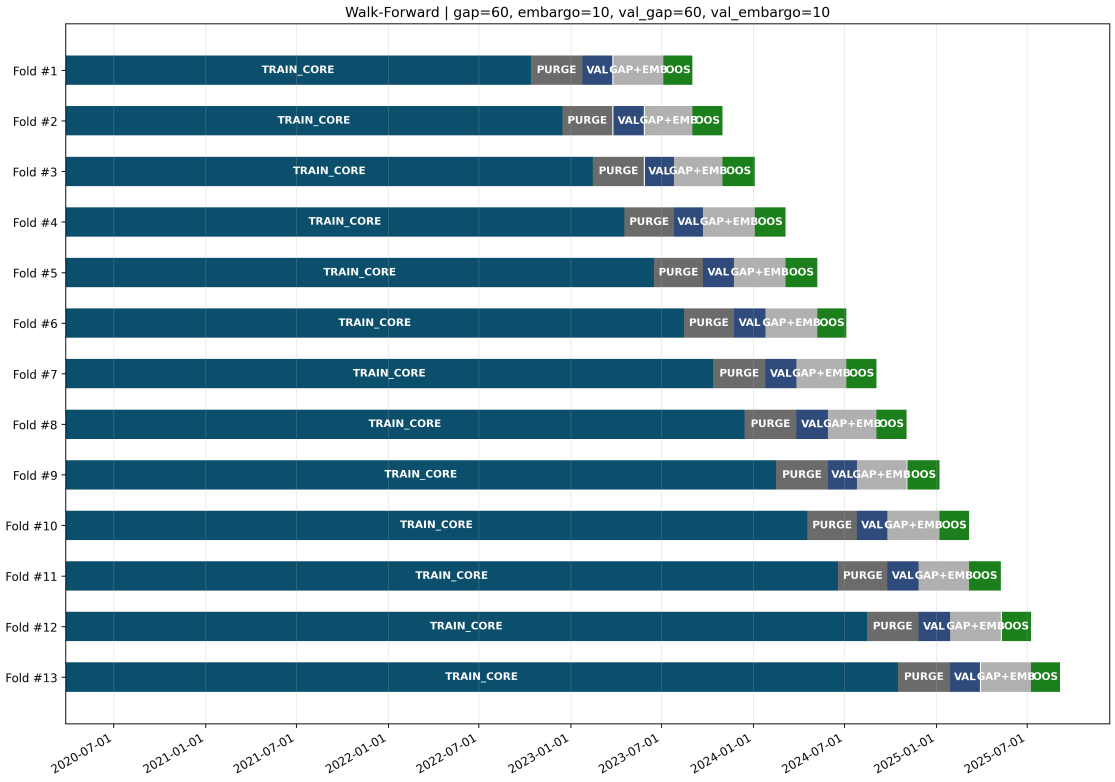


Figura 1: Esquema Walk-Forward PRO con partición interna: TRAIN_CORE, PURGE, VAL, GAP+EMB y OOS. La ventana de entrenamiento se expande en cada pliegue.

5.6.3 Configuración del protocolo

La Tabla 6 detalla los parámetros utilizados.

Cuadro 6: Configuración del protocolo Walk-Forward PRO (purge + embargo).

Parámetro	**Valor**	**Justificación**
`train_size` (inicial)	756 fechas (≈3 años)	Mínimo para estabilidad de CAPM rolling
`gap` (purge OOS)	60 fechas $= H$	Cierra look-ahead de VolFut exactamente
`embargo`	10 fechas	Buffer de seguridad adicional
`test_size` (OOS)	42 fechas (≈2 meses)	Ventana bimestral no solapada
`step_size`	42 fechas = test_size	OOS no solapados entre pliegues
`val_size`	42 fechas	Misma escala que OOS para simetría
`val_gap`	60 fechas $= H$	Misma lógica que gap principal
`val_embargo`	10 fechas	Consistente con embargo principal
`min_train_size`	252 fechas	Al menos 1 año de datos de mercado
`n_folds`	13	

Diseño del gap = H como decisión de integridad temporal exacta. La elección `gap = val_gap = 60 = H` no es arbitraria: garantiza que el primer día de cualquier ventana de evaluación (VAL u OOS) esté separado del último día de TRAIN_CORE por exactamente el horizonte de etiquetado H. Con gap $< H$, el etiquetado del último tramo de TRAIN_CORE usaría rendimientos de días que caen dentro de la ventana de evaluación, introduciendo *look-ahead* residual. Con gap $= H$, esta filtración queda cerrada de forma *exacta, no aproximada*. Este es el blindaje más fino del protocolo y el más difícil de replicar accidentalmente; su correcta implementación es lo que distingue el Walk-Forward PRO del walk-forward simple [13, 145].

El resultado es 13 particiones temporales independientes que cubren el período de evaluación julio 2023 a septiembre 2025, con un total de 16,133 observaciones OOS.

El número de 13 pliegues resulta de la combinación del período disponible (enero 2020 – octubre 2025), una ventana mínima de entrenamiento de aproximadamente 500 días hábiles para estabilidad de las estimaciones CAPM rolling de 252 días, y un período OOS de 2 meses por pliegue que provee suficiente cobertura temporal para el test de Wilcoxon pareado con potencia adecuada: $n = 13$ permite detectar diferencias medianas con $\beta \geq 0.80$ para efectos de tamaño mediano [52]. Menos pliegues reducirían la potencia estadística; más pliegues reducirían la longitud de cada período OOS, disminuyendo la representatividad de las métricas por pliegue.

5.7 Modelos Evaluados y Criterios de Comparabilidad

5.7.1 Condiciones de comparabilidad *apples-to-apples*

Los tres modelos comparten: (i) mismo universo y variables CORE-10, (ii) idénticas particiones Walk-Forward PRO, (iii) escalamiento ajustado exclusivamente en TRAIN_CORE, (iv) umbrales de etiquetado estimados en TRAIN_CORE, y (v) mismas métricas de evaluación [52, 97]. La comparabilidad iguala *información disponible*; no iguala el número de parámetros (KAN tiene $\approx 15\times$ más que MLP), lo que se reconoce y justifica por la naturaleza diferente de las dos arquitecturas [144, 226].

5.7.2 Red Kolmogorov-Arnold (KAN)

Las KAN reemplazan activaciones puntuales por **funciones univariadas aprendibles en las aristas**, parametrizadas como B-splines, lo que favorece aproximación flexible e interpretabilidad estructural [27, 144]. La implementación emplea una variante eficiente en PyTorch

[22].

La arquitectura es `width` $= [10, 32, 3]$ con dos capas `KANLayer`. Cada arista se modela con splines cúbicos (`degree` $= 3$) y malla de nudos adaptativa por cuantiles de los datos de activación. Las funciones $\phi(x)$ aprendidas en la primera capa son directamente inspeccionables, proveyendo el mecanismo de interpretabilidad intrínseca descrito en el Marco Teórico.

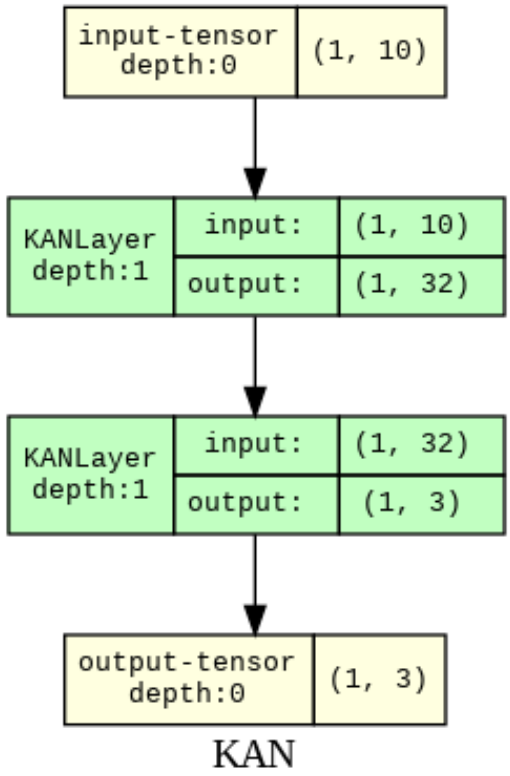


Figura 2: Arquitectura KAN usada en el estudio: $10 \to 32 \to 3$.

5.7.3 Red Neuronal Multicapa (MLP) *arch-matched*

Como *baseline* neuronal se emplea un MLP de dos capas totalmente conectadas con arquitectura emparejada con KAN: `width` $= [10, 32, 3]$. Esta igualdad de dimensiones de capas reduce el riesgo de atribuir diferencias de desempeño a capacidad estructural no comparable, aunque la capacidad paramétrica efectiva difiera ($\approx$ 451 parámetros en MLP vs. $\approx$ 6,691 en KAN) por la naturaleza aditiva de los splines [85, 97, 144].

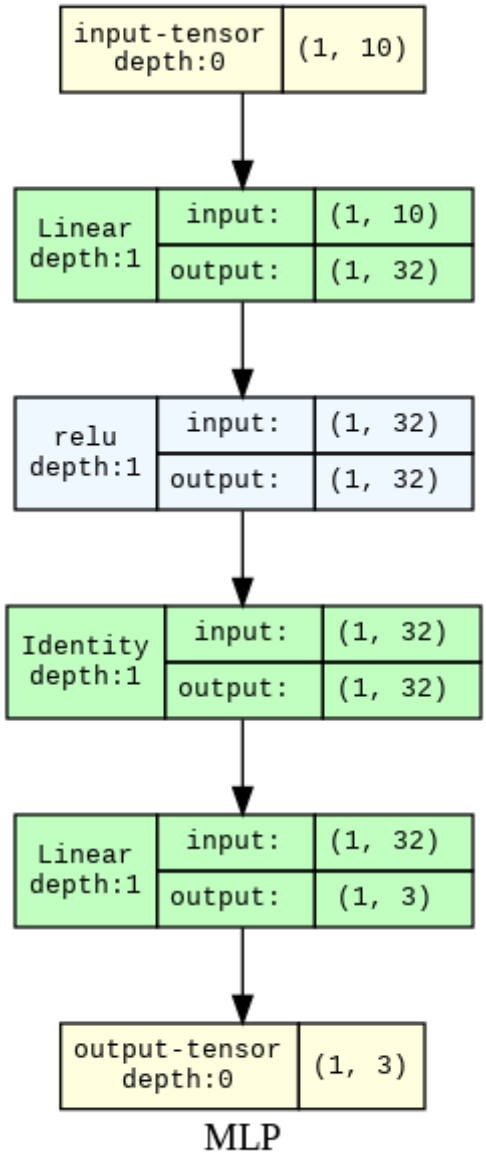


Figura 3: Arquitectura MLP *arch-matched*: $10 \to 32 \to 3$.

5.7.4 SVM lineal calibrado

Como *baseline* clásico se usa `LinearSVC` [45]. Dado que produce márgenes (no probabilidades), se aplica calibración probabilística sigmoidal (*Platt scaling*) [173, 176] ajustada **sobre el bloque VAL de cada pliegue** (modo *prefit*), obteniendo probabilidades comparables para métricas ROC/AUC sin acceso a datos OOS. Este esquema preserva la integridad temporal del protocolo y es consistente con la calibración por pliegue de los modelos neuronales.

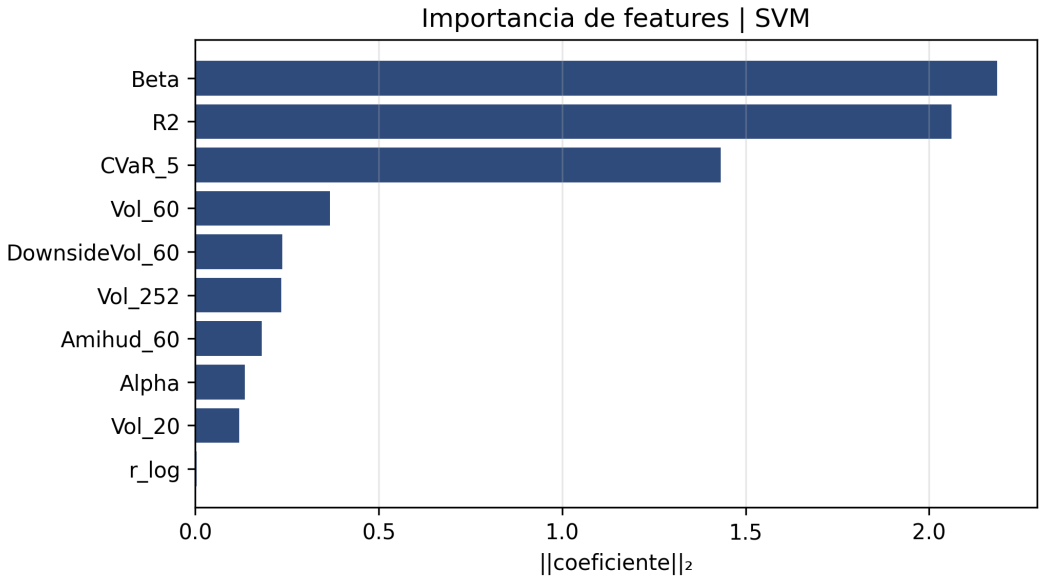


Figura 4: Esquema de la SVM lineal calibrada usada en el estudio.

5.8 Entrenamiento, Regularización e Hiperparámetros

5.8.1 Criterio de optimización y selección temporal

KAN y MLP se entrenan minimizando entropía cruzada multiclase con *early stopping* basado en F1 macro sobre VAL. El optimizador es AdamW [127, 149] con `ReduceLROnPlateau` que reduce la tasa de aprendizaje cuando el F1 en VAL deja de mejorar. Se aplica recorte de norma del gradiente (*gradient clipping*, max_norm $= 1.0$) para estabilidad numérica [180].

5.8.2 Tabla de hiperparámetros

La Tabla 7 documenta todos los hiperparámetros de los tres modelos tal como fueron configurados en el pipeline v3.4. Los valores de KAN y MLP resultan de una búsqueda en la ventana VAL de cada pliegue; los de SVM se fijaron con base en el desempeño en VAL. Estos parámetros fueron fijados antes del análisis de los resultados OOS.

5.8.3 Regularización específica en KAN

En KAN se incorpora regularización sobre parámetros spline: (i) ℓ_1 sobre coeficientes ($\lambda_1 = 7.5 \times 10^{-5}$, parsimonia), y (ii) penalización de segundas diferencias ($\lambda_{\text{smooth}} = 2.5 \times 10^{-5}$, suavidad, evita oscilaciones). Estas penalizaciones buscan controlar sobreajuste bajo ruido financiero y mantener funciones $\phi(x)$ interpretables [27, 97].

Cuadro 7: Hiperparámetros de los tres modelos (pipeline v3.4).

Hiperparámetro	**KAN**	**MLP**	**SVM**
Arquitectura			
Dimensiones de capas	[10, 32, 3]	[10, 32, 3]	Lineal
N° nudos internos (n)	10	—	—
Grado spline (p)	3 (cúbico)	—	—
Funciones base por arista (G)	14	—	—
Conexión residual (SiLU)	Sí	No	—
Parámetros totales ($\approx$)	6,691	451	30
Entrenamiento			
Optimizador	AdamW	AdamW	—
Tasa de aprendizaje (lr)	7.5×10^{-4}	1×10^{-3}	—
Weight decay	1×10^{-4}	1×10^{-4}	—
Épocas máximas	1,000	1,000	—
Paciencia (*early stopping*)	100 épocas	100 épocas	—
Monitor *early stopping*	F1 macro (VAL)	F1 macro (VAL)	—
Tamaño de batch	1,024	1,024	—
Clipping de gradiente (max_norm)	1.0	1.0	—
Scheduler	ReduceLROnPlateau	ReduceLROnPlateau	—
factor scheduler	0.5	0.5	—
paciencia scheduler	15 épocas	15 épocas	—
lr mínima	3×10^{-5}	3×10^{-5}	—
Semilla global	42	42	42
Regularización (KAN)			
λ_1 (L1 coef. spline)	7.5×10^{-5}	—	—
λ_{smooth} (curvatura)	2.5×10^{-5}	—	—
Dropout	—	0.0	—
SVM			
Tipo de kernel	—	—	Lineal
Parámetro de regularización (C)	—	—	1.0
Calibración probabilística	—	—	Platt/sigmoid
Modo calibración	—	—	*prefit* (VAL)
Máx. iteraciones	—	—	10,000
Procedimientos especiales (KAN)			
Actualización de nudos	Cada 100 épocas (primeras 500)	—	—
Recorte de colas para nudos	$q = 1\%$ bilateral	—	—
Manejo de desbalance de clases			
Ajuste de logits (logit adj.)	Sí ($\tau = 0.7$)	Sí ($\tau = 0.7$)	—

5.8.4 Filtrado de frontera de decisión (*fringe denoising*)

Para reducir el ruido de entrenamiento en las regiones de mayor ambigüedad del RiskIndex, se eliminan del conjunto TRAIN_CORE las observaciones cuyo score cae dentro de un margen del $\pm 5\%$ alrededor de los umbrales de cuantil ($q_{1/3}$ y $q_{2/3}$). Este filtro (`FRINGE_FRAC = 0.05`) se activa únicamente si el resultado conserva al menos el 20% de las observaciones de TRAIN_CORE, evitando reducción excesiva del conjunto de entrenamiento. La eliminación de ejemplos frontera con etiquetas ambiguas mejora la señal de gradiente sin afectar la composición del conjunto OOS. El valor `FRINGE_FRAC` $= 0.05$ corresponde al 5% estándar recomendado en la implementación de referencia de Liu et al. [144] para datos tabulares con distribuciones no uniformes, y fue mantenido sin ajuste para preservar la reproducibilidad directa de la arquitectura base.

5.8.5 Actualización de malla de nudos

De forma periódica durante la primera mitad del entrenamiento (cada 100 épocas, durante las primeras 500), la malla de nudos se actualiza usando cuantiles de las activaciones de la capa anterior. Este procedimiento adapta la discretización spline a la distribución observada de activaciones y se acompaña de limpieza del estado del optimizador para parámetros spline [27, 144].

5.8.6 Ajuste de logits por priors de clase

Para robustecer el entrenamiento ante desbalance de clases por pliegue, se aplica el ajuste de logits propuesto por Menon et al. [162]:

$$\tilde{\ell}_c = \ell_c - \tau \cdot \log(\pi_c), \tag{24}$$

donde ℓ_c es el logit de salida para la clase c, π_c es la frecuencia de esa clase en TRAIN_CORE del pliegue, y $\tau = 0.7$ es el hiperparámetro de temperatura. El ajuste se aplica durante el entrenamiento y la inferencia OOS, incentivando predicciones calibradas respecto a la distribución real de clases.

5.9 Evaluación del Desempeño Predictivo

El desempeño se reporta por pliegue y en forma agregada sobre todas las predicciones OOS ($n = 16{,}133$). Las métricas son:

- **Accuracy, Precisión, Recall y F1** por clase, con promedios macro y ponderado [179, 201].

- **F1 macro** como métrica primaria: pondera por igual las tres clases, sin que las más frecuentes oculten fallas en la clase Medio.

- **Matrices de confusión** en conteos y normalizadas por fila. Su inspección es informativa en este problema ordinal: los errores no adyacentes (Bajo $\rightarrow$ Alto) revelan fallos estructurales más graves que los errores entre clases contiguas [91, 179].

- **Curvas ROC OvR** y **AUC macro** para separabilidad probabilística [63, 93].

5.10 Comparación Estadística por Pliegue

La unidad de análisis es el desempeño OOS de cada pliegue bajo particiones idénticas. Este diseño pareado reduce variabilidad atribuible al período y condiciones de mercado [52, 54].

- **Prueba de Wilcoxon pareada:** contrasta diferencias de desempeño por pliegue sin asumir normalidad; se aplica de forma unilateral para verificar si un modelo supera consistentemente a otro [222].

- **Intervalos bootstrap:** sobre las diferencias pareadas de métricas, reportando media y límites al 95% para cuantificar magnitud y estabilidad [19, 56].

Con $n = 13$ pares, la prueba de Wilcoxon tiene potencia moderada para detectar diferencias pequeñas; los resultados estadísticamente significativos reportados corresponden a diferencias con dirección consistente a través de la mayoría de los pliegues, lo que refuerza la confiabilidad de la inferencia más allá del valor p aislado.

5.11 Eficiencia Computacional

Por pliegue se registran: tiempo de entrenamiento, tiempo de inferencia, tiempo de calibración (SVM), memoria RAM pico y memoria GPU pico cuando aplica. El objetivo es documentar el costo computacional relativo bajo un entorno consistente y evaluar la viabilidad práctica de cada modelo para uso operativo.

5.12 Metodología de Interpretabilidad

Se prioriza evidencia de interpretabilidad intrínseca sobre *post-hoc* [55, 89, 189]. El criterio operativo comprende tres niveles verificables:

- **Importancia global por variable:** magnitud media de coeficientes spline en la primera

capa de KAN, sin recurrir a SHAP o LIME.

- **Análisis funcional de $\phi(x)$:** inspección de formas aprendidas —monotonía, saturaciones, umbrales, regiones de alta sensibilidad— para auditar cómo cada variable modifica la señal interna.

- **Coherencia económica:** contraste cualitativo entre la forma de $\phi(x)$ y las relaciones esperadas por teoría financiera.

En SVM lineal, la interpretabilidad se apoya en la norma ℓ_2 de coeficientes por clase. En MLP, la interpretación directa de pesos no es estable ni transparente; el modelo sirve como *baseline* de desempeño predictivo sin pretensión de interpretabilidad estructural [165, 189].

5.13 Verificación Económica: Backtesting con Predicciones OOS

Como extensión metodológica, las predicciones fuera de muestra se utilizan como señal de selección de activos en un escenario de simulación histórica. El objetivo es verificar si las predicciones de riesgo tienen contenido económico real: una señal que funcione bien predictivamente *debería* mostrar valor informativo sobre la distribución futura de retornos.

La estrategia es *long-only*: en cada fecha de rebalanceo semanal, se invierte de forma equiponderada en acciones clasificadas como **Bajo** riesgo. Para eliminar *look-ahead bias*, el backtest utiliza **exclusivamente predicciones OOS** y aplica el retorno del día siguiente al de la predicción (*next return*) [12, 145].

5.13.1 Costos de transacción y benchmarks

En cada rebalanceo semanal se calcula el *turnover* respecto a la asignación previa y se descuenta un costo de 0.15% sobre el *turnover* del período, reflejando fricciones realistas de mercado. Se definen dos benchmarks de referencia:

- **Benchmark equiponderado (bench_eqw):** media aritmética del retorno del día siguiente para las 30 emisoras del universo, evaluada exclusivamente en las fechas OOS. Este es el benchmark primario de comparación, ya que usa el mismo universo y el mismo período que la estrategia KAN.

- **Benchmark de mercado (bench_mkt):** retorno del IPC durante el mismo período OOS, como referencia de desempeño del mercado amplio.

La distinción entre ambos benchmarks es importante: el bench_eqw refleja lo que habría rendido simplemente mantener todas las emisoras en partes iguales —la referencia de costo de oportunidad de *no* aplicar el filtro de riesgo— mientras que el bench_mkt refleja el mercado en su conjunto.

5.13.2 Métricas del backtest

Se reportan: CAGR bruto y neto de costos, ratio de Sharpe, *maximum drawdown* y tasa de aciertos, junto con curvas de capital acumulado. La sensibilidad a costos de transacción (diferencia bruto/neto) demuestra que el análisis no ignora fricciones reales del mercado [23, 145].

Alcance interpretativo del backtest. El backtest es una **verificación de coherencia económica del pipeline**, no evidencia concluyente de ventaja sobre el mercado. No incorpora optimización de portafolio, restricciones de inversión ni controles de múltiples hipótesis. Cualquier conclusión operativa sobre retornos requiere un programa de validación más amplio [5, 12, 95].

5.14 Reproducibilidad y Trazabilidad

Toda la implementación se realizó en Python con librerías estándar de cómputo científico y ML. Para reproducibilidad se fija una semilla pseudoaleatoria global (`SEED = 42`). Por pliegue se registran: configuración de etiquetado (q_{low}, q_{high}), distribución de clases en TRAIN/VAL/OOS, todos los hiperparámetros de entrenamiento y artefactos de evaluación (matrices de confusión serializadas, métricas por pliegue). Los índices temporales exactos de cada partición y los parámetros del escalador ajustado en TRAIN_CORE se almacenan para auditoría completa del diseño experimental [97, 157].

VI. Resultados y discusión

Cómo leer este capítulo. Los resultados se presentan como cuatro hallazgos encadenados. Una aclaración de lectura es importante antes de proceder: el objetivo de este capítulo no es demostrar supremacía universal de KAN en toda métrica. Es evaluar qué arquitectura, bajo validación temporal estricta, ofrece el mejor equilibrio entre desempeño global competitivo, ventaja en la zona de mayor dificultad operativa, estabilidad temporal e interpretabilidad auditable. MLP obtiene métricas globales marginalmente superiores; eso no invalida el argumento central. SVM colapsa la región transicional de forma sostenida durante múltiples pliegues consecutivos; ese es el hallazgo más relevante para la gestión de riesgo. La argumentación no se apoya en una métrica aislada, sino en la combinación de cuatro dimensiones evaluadas bajo el mismo protocolo.

La evaluación comprende 16,133 observaciones OOS a lo largo de 13 particiones cronológicas Walk-Forward PRO (julio 2023 – septiembre 2025). Los tres modelos evaluados —**KAN** (propuesta), **MLP** (*baseline* neuronal) y **SVM lineal calibrado** (*baseline* clásico)— comparten el mismo conjunto de variables (CORE-10), el mismo protocolo de particiones y las mismas métricas, garantizando comparabilidad estricta [12, 13, 145].

En conjunto, la evidencia de este capítulo sustenta la tesis central del trabajo: en el mercado accionario mexicano, bajo validación temporal estricta, es posible combinar desempeño OOS competitivo, ventaja diferencial en la región de mayor dificultad operativa, estabilidad temporal e interpretabilidad estructural auditable en un mismo sistema. KAN, entre las arquitecturas evaluadas bajo el protocolo adoptado, es la que mejor realiza esa combinación.

6.1 Hallazgo 1 — Desempeño Global Competitivo bajo Validación Temporal Estricta

El primer resultado establece que KAN es viable como clasificador de riesgo en la BMV: alcanza desempeño predictivo OOS **competitivo** bajo las condiciones de validación más exigentes documentadas para este tipo de estudio en el mercado mexicano.

La Tabla 8 resume el desempeño agregado sobre las 16,133 observaciones OOS, distribuidas por clase como: Alto = 5,683, Bajo = 5,962, Medio = 4,488.

MLP obtiene el mejor desempeño global; KAN se mantiene competitivo y supera a SVM en todas las métricas agregadas. La diferencia KAN–MLP en F1 macro (0.0082) no alcanza

Cuadro 8: Desempeño global OOS agregado en 13 pliegues ($n = 16{,}133$). AUC corresponde a ROC multiclase OvR con promedio macro [63, 93].

Modelo	Accuracy	Precisión macro	F1 macro	AUC OvR macro
KAN	0.7570	0.7571	0.7496	0.8799
MLP (arch-matched)	0.7668	0.7639	0.7578	0.8926
SVM (lineal calibrado)	0.7539	0.7298	0.7275	0.8761

significancia estadística bajo comparación Wilcoxon pareada por pliegue ($p = 0.812$, bilateral): KAN y MLP son **estadísticamente equivalentes en métricas globales** bajo este protocolo. En el contexto de series financieras no estacionarias, diferencias del orden de centésimas son habituales por variabilidad temporal y cambios de régimen [52, 145]. Lo sustantivo de este hallazgo no es quién lidera por centésimas: es que KAN alcanza este nivel de desempeño bajo un protocolo de validación más riguroso que el empleado en los estudios comparables revisados para la BMV. Esto confirma **H1**: el sistema es predictivamente competitivo OOS.

El análisis de desempeño por clase, estabilidad y estadística que sigue revela que el cuadro global oculta el hallazgo más relevante para la gestión de riesgo, que emerge precisamente en la zona de mayor dificultad.

6.2 Hallazgo 2 — La Zona de Riesgo Transicional: Principal Reto y Ventaja Diferencial

El segundo hallazgo —el más relevante desde la perspectiva operativa— es que la clase *Medio*, donde los perfiles de activos comparten rasgos con ambos extremos, concentra la dificultad de clasificación en los tres modelos, y es precisamente en esta región donde KAN muestra su ventaja consistente frente al modelo lineal.

Cuadro 9: Reporte de clasificación OOS por clase, 13 pliegues ($n = 16{,}133$).

	Alto			Bajo			Medio		
Modelo	Prec.	Rec.	F1	Prec.	Rec.	F1	Prec.	Rec.	F1
KAN	0.8093	0.8564	0.8322	0.8996	0.7409	0.8125	0.5623	**0.6526**	0.6041
MLP	0.8142	0.8761	0.8440	0.9044	0.7553	0.8231	0.5731	0.6435	0.6063
SVM	0.7789	0.8494	0.8126	0.8207	0.8797	0.8492	0.5898	0.4659	0.5206

La asimetría es sistemática: las clases extremas son sustancialmente más fáciles de separar que Medio. Esta asimetría tiene interpretación económica directa: Medio representa estados donde

la volatilidad es moderada pero la exposición sistemática puede ser significativa, generando superposición en el espacio de características [145, 214].

El dato central es el recall de la clase Medio: **KAN recupera el 65.3% de los estados de riesgo transicional**, frente al 64.4% de MLP y el 46.6% de SVM. Es necesario distinguir dos comparaciones con consecuencias diferentes. La **ventaja KAN–SVM** es de 18.7 puntos porcentuales, estadísticamente significativa (Wilcoxon $p = 0.016$) y de magnitud operativa concreta: el SVM colapsa sistemáticamente los casos intermedios hacia los extremos bajo cambios de régimen, produciendo decisiones de asignación más volátiles. La **diferencia KAN–MLP** es de 0.9 puntos porcentuales ($p = 0.420$, no significativa); ambos modelos son estadísticamente equivalentes en esta dimensión. La afirmación de ventaja diferencial de KAN se sostiene fundamentalmente contra la frontera lineal, no contra el *baseline* neuronal.

6.2.1 Patrones de error: matrices de confusión por modelo

Las matrices de confusión revelan *cómo* se equivoca cada modelo, aspecto crítico en clasificación de riesgo ordinal: un error Bajo → Alto es cualitativamente más grave que un error adyacente, y su presencia indica fallos estructurales que ninguna métrica global resume adecuadamente [91, 179, 201].

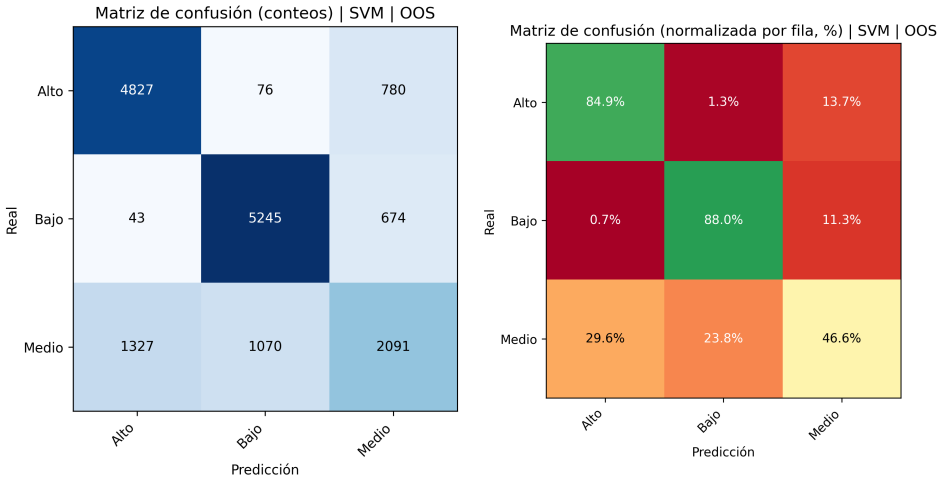


Figura 5: SVM: matrices de confusión OOS (conteos y normalizada por fila).

SVM lineal calibrado. La diagonal de SVM es alta en Bajo y Alto, pero se observa degradación sustantiva en Medio: una fracción importante de verdaderos Medio se asigna hacia los extremos (especialmente hacia Alto). Este patrón es consecuencia directa de la frontera lineal rígida en regiones de superposición [21, 97].

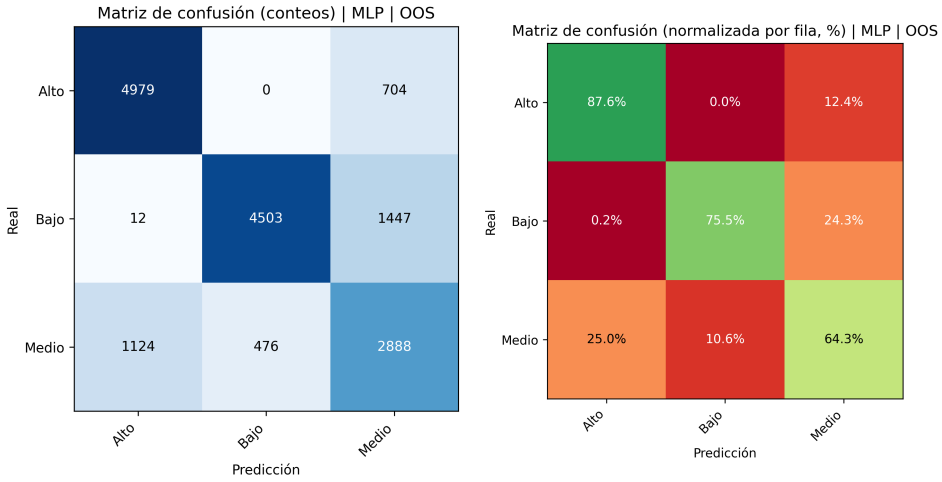


Figura 6: MLP: matrices de confusión OOS (conteos y normalizada por fila).

MLP (*arch-matched*). El MLP mejora la retención en la diagonal para Medio respecto a SVM, consistente con su capacidad de modelar fronteras no lineales. La confusión residual entre Medio y los extremos refleja la dificultad intrínseca de la región transicional [85, 145].

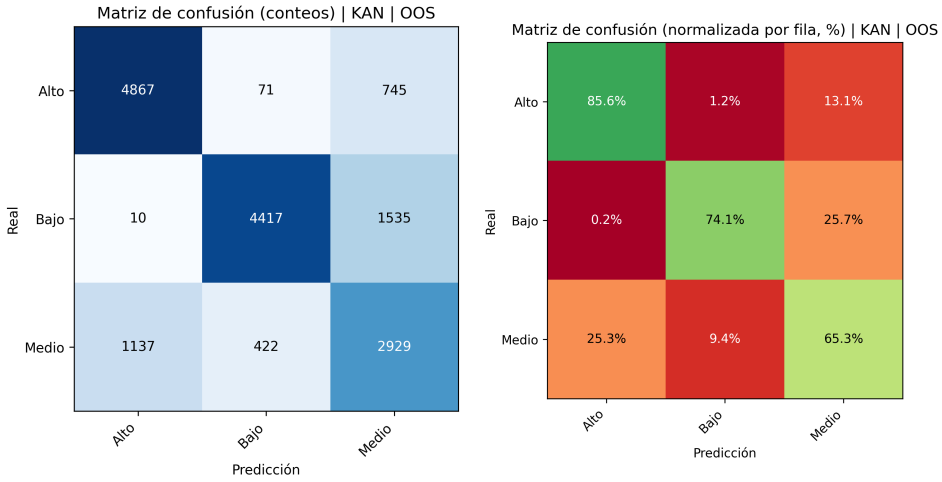


Figura 7: KAN: matrices de confusión OOS (conteos y normalizada por fila).

KAN. KAN presenta un patrón cercano a MLP en las clases extremas y retención competitiva en Medio. La parametrización spline favorece transiciones suaves en la frontera de decisión, consistente con el mayor recall en la región intermedia [27, 144].

6.2.2 Separabilidad probabilística: curvas ROC y AUC por clase

En los tres modelos, Medio es sistemáticamente la clase de menor AUC, confirmando que la dificultad principal radica en la región intermedia, no en discriminar los extremos. KAN obtiene el mayor AUC en la clase Medio (0.7951), superando a SVM (0.7421) en 5.3 puntos porcentuales —diferencia coherente con la ventaja en recall y F1(Medio)—. La diferencia KAN–MLP en AUC(Medio) es de 0.62 puntos porcentuales, muy pequeña para ser interpretada de forma aislada; la evidencia robusta de ventaja en esta clase descansa en el recall y F1 con respaldo estadístico por pliegue, no en el AUC por clase.

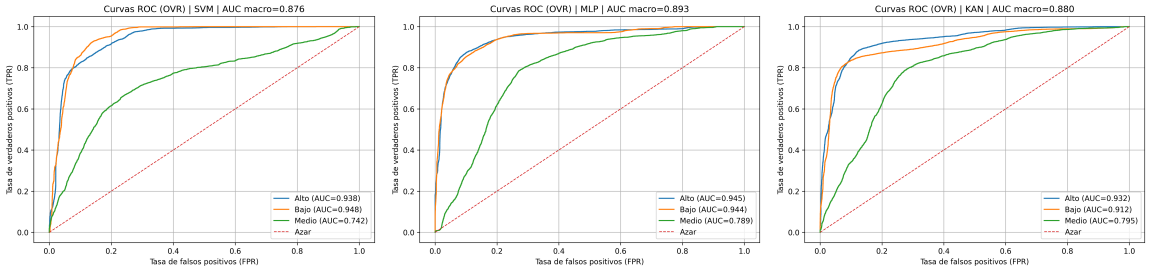


Figura 8: Curvas ROC multiclase OvR en OOS para SVM (izq.), MLP (centro) y KAN (der.).

Cuadro 10: AUC OvR por clase en evaluación OOS [63, 93].

Clase	AUC (SVM)	AUC (MLP)	AUC (KAN)
Alto	0.9382	0.9450	0.9324
Bajo	0.9482	0.9440	0.9122
Medio	0.7421	0.7889	**0.7951**

6.3 Hallazgo 3 — Estabilidad Temporal: Consistencia a través de Regímenes de Mercado

Un buen clasificador de riesgo no solo debe funcionar en promedio: debe mantener su desempeño a través de diferentes condiciones de mercado. El tercer hallazgo es que KAN exhibe la menor variabilidad de F1 macro entre pliegues, confirmando su robustez a cambios de régimen en el período evaluado.

Cuadro 11: Desempeño por pliegue: F1 macro y recall(Medio) en ventanas OOS (julio 2023 – septiembre 2025). Valores v3.4.

Pliegue	Período OOS	F1 KAN	F1 MLP	F1 SVM	Rec.Med. KAN	Rec.Med. MLP	Rec.Med. SVM	Δ(KAN−SVM)
1	2023-07-05 – 2023-08-31	0.721	0.724	0.720	0.628	0.732	0.728	−0.100
2	2023-09-01 – 2023-10-30	0.756	0.704	0.650	0.635	0.690	0.646	−0.012
3	2023-10-31 – 2024-01-03	0.695	0.727	0.703	0.528	0.708	0.696	−0.168
4	2024-01-04 – 2024-03-04	0.775	0.663	0.672	0.762	0.605	0.326	+0.437
5	2024-03-05 – 2024-05-07	0.764	0.714	0.665	0.790	0.650	**0.168**	+0.622
6	2024-05-08 – 2024-07-04	0.697	0.775	0.685	0.655	0.662	**0.178**	+0.476
7	2024-07-05 – 2024-09-02	0.686	0.748	0.631	0.550	0.616	**0.100**	+0.450
8	2024-09-03 – 2024-11-01	0.748	0.820	0.601	0.568	0.716	**0.080**	+0.488
9	2024-11-04 – 2025-01-06	0.768	0.808	0.812	0.914	0.770	0.700	+0.214
10	2025-01-07 – 2025-03-06	0.783	0.769	0.802	0.754	0.629	0.739	+0.015
11	2025-03-07 – 2025-05-09	0.770	0.797	0.812	0.627	0.617	0.750	−0.123
12	2025-05-12 – 2025-07-08	0.707	0.739	0.663	0.444	0.389	0.249	+0.196
13	2025-07-09 – 2025-09-04	0.803	0.805	0.744	0.580	0.550	0.464	+0.116

Cuadro 12: Resumen de estabilidad: media ± desviación estándar por pliegue (v3.4).

Modelo	F1 macro (media ± std)	Recall(Medio) (media ± std)
KAN	$0.744 \pm \mathbf{0.038}$	0.649 ± 0.127
MLP	0.753 ± 0.047	$0.641 \pm \mathbf{0.097}$
SVM	0.705 ± 0.070	0.448 ± 0.271

KAN exhibe la **menor desviación estándar de F1 macro entre pliegues** ($\sigma = 0.038$), inferior a MLP ($\sigma = 0.047$) y a SVM ($\sigma = 0.070$). Esto confirma que KAN es la arquitectura más estable en desempeño global. Un matiz relevante: en recall(Medio) específicamente, MLP muestra la mayor estabilidad ($\sigma = 0.097 < \sigma_{\text{KAN}} = 0.127$); ambos son, sin embargo, marcadamente más estables que SVM ($\sigma = 0.271$, más de $2.1\times$ la variabilidad de KAN). La inestabilidad de SVM en esta métrica es el hallazgo operativamente más significativo.

6.3.1 El patrón de colapso de SVM: cuatro pliegues consecutivos

La Tabla 11 revela un patrón que va más allá de la media: en los primeros tres pliegues (julio 2023 – enero 2024), SVM y KAN son comparables en recall(Medio), con SVM incluso ligeramente superior. A partir del pliegue 4 (enero 2024), el comportamiento de SVM diverge de forma dramática y sostenida. Los pliegues 5 a 8 (marzo a noviembre de 2024) documentan un **colapso sostenido** de SVM en la región transicional, mientras KAN mantiene capacidad clasificatoria estable:

Cuadro 13: Colapso sostenido de SVM en recall(Medio): pliegues 5–8 (mar.–nov. 2024).

Pliegue	Período	SVM	KAN	Ratio KAN/SVM
5	2024-03-05 – 2024-05-07	0.168	0.790	4.7×
6	2024-05-08 – 2024-07-04	0.178	0.655	3.7×
7	2024-07-05 – 2024-09-02	0.100	0.550	5.5×
8	2024-09-03 – 2024-11-01	0.080	0.568	7.1×

Este período coincide con episodios de incertidumbre política y financiera en México que alteraron la percepción de riesgo de forma pronunciada y diferenciada por emisora. La frontera de decisión lineal de SVM, calibrada con régimen anterior, pierde su validez en la zona transicional cuando el patrón de riesgo relativo entre emisoras cambia sustancialmente. KAN, con sus funciones $\phi(x)$ localmente adaptativas por región del dominio, mantiene mayor coherencia clasificatoria bajo ese cambio de régimen. Esta es la dimensión de robustez operativa más relevante para sistemas de gestión de riesgo: la confiabilidad precisamente cuando el entorno es más adverso.

Concretamente, el pliegue 5 (marzo–mayo 2024) precede al proceso electoral federal, el pliegue 6 (mayo–julio 2024) captura el período inmediatamente posterior a los comicios del 2 de junio y la depreciación del peso de aproximadamente 15% frente al dólar registrada en esas semanas, y los pliegues 7 y 8 (julio–noviembre 2024) cubren la transición del gobierno y el ajuste de expectativas del mercado respecto a la política energética y regulatoria. En este entorno, la percepción de riesgo relativo entre emisoras cambió de forma pronunciada: empresas del sector infraestructura, energía y retail cambiaron de categoría de forma sincronizada, un patrón que la frontera lineal del SVM —calibrada sobre régimen anterior— no tenía capacidad de capturar en la zona transicional.

La ventaja de KAN en este escenario no es una propiedad estadística abstracta: es exactamente la clase de robustez que un supervisor de SIEFORES necesita verificar antes de adoptar cualquier sistema clasificador, y que la interpretabilidad estructural permite documentar de forma directa.

6.3.2 Distribución de recall(Medio) por pliegue

Para completar el análisis de estabilidad, la Tabla 14 reporta estadística descriptiva del recall(Medio) por pliegue para los tres modelos.

El IQR de SVM (0.178 a 0.700) abarca casi todo el espacio posible, reflejando su extrema

Cuadro 14: Estadística descriptiva del recall(Medio) por pliegue (n = 13).

Modelo	Media	Mediana	IQR	Mínimo (pliegue)	Máximo (pliegue)
KAN	0.649	0.628	[0.568, 0.754]	0.444 (12)	0.914 (9)
MLP	0.641	0.650	[0.616, 0.708]	0.389 (12)	0.770 (9)
SVM	0.448	0.464	[0.178, 0.700]	0.080 (8)	0.750 (11)

sensibilidad al régimen. El de MLP (0.616 a 0.708) es el más estrecho, coherente con su mayor estabilidad en recall(Medio). El de KAN (0.568 a 0.754) es intermedio, con alta variabilidad explicada por los valores extremos de pliegue 9 (0.914, máximo global) y pliegue 12 (0.444, segundo mínimo tras SVM).

6.3.3 Dinámica de entrenamiento y rol del *early stopping*

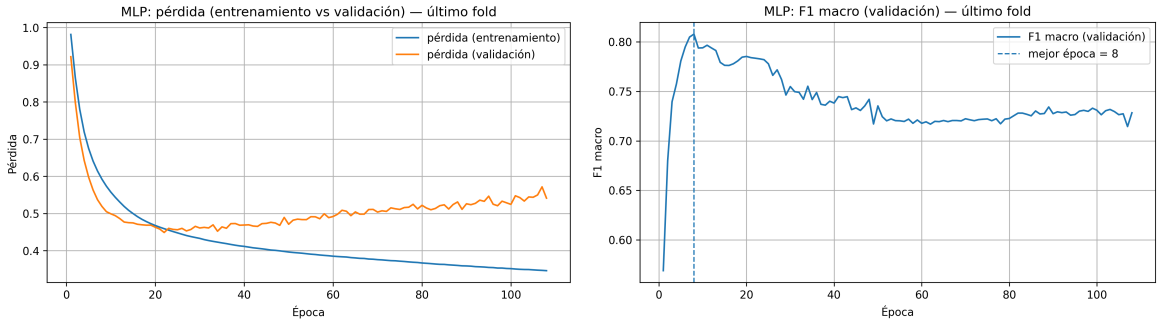


Figura 9: MLP (último pliegue): pérdida entrenamiento vs. validación (izq.) y F1 macro de validación con mejor época marcada (der.).

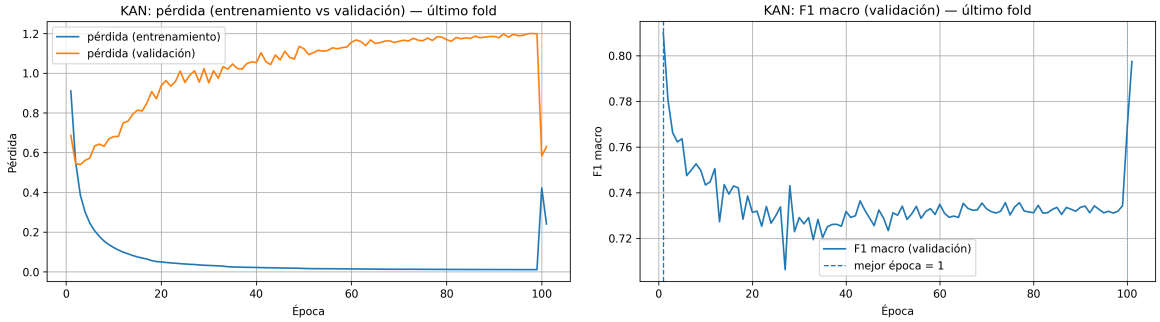


Figura 10: KAN (último pliegue): pérdida entrenamiento vs. validación (izq.) y F1 macro de validación con mejor época marcada (der.).

MLP muestra una reducción rápida de pérdida y un máximo temprano de F1 de validación, seguido por degradación gradual [21, 85]. KAN exhibe mayor expresividad inicial pero también sobreajuste si no se controla temporalmente; el *early stopping* no es una conveniencia

computacional sino un componente metodológico central para preservar la generalización [144, 180].

6.3.4 Respaldo estadístico por pliegue

La Tabla 15 reporta la inferencia pareada (Wilcoxon) e intervalos bootstrap al 95% sobre diferencias de métricas. Los valores son los calculados sobre los 13 pliegues del protocolo Walk-Forward PRO.

Cuadro 15: Inferencia pareada Wilcoxon + IC bootstrap 95% para diferencias (KAN menos *baseline*) sobre 13 pliegues.

Métrica	**Comparación**	n	W	p-valor	**Dif. media**	**IC95% low**	**IC95% high**
F1 macro	KAN > SVM	13	73.0	**0.0287**	+0.040	+0.007	+0.072
F1 macro	KAN > MLP	13	33.0	0.8121	−0.009	−0.037	+0.021
Recall macro	KAN > SVM	13	63.0	0.1219	+0.024	−0.006	+0.054
Recall macro	KAN > MLP	13	34.0	0.7929	−0.009	−0.037	+0.021
Recall(Medio)	KAN > SVM	13	76.0	**0.0164**	+0.201	+0.059	+0.346
Recall(Medio)	KAN > MLP	13	49.0	0.4197	+0.008	−0.053	+0.067
F1(Medio)	KAN > SVM	13	77.0	**0.0133**	+0.134	+0.048	+0.225
F1(Medio)	KAN > MLP	13	42.0	0.6066	−0.006	−0.046	+0.038

Tres lecturas emergen de estos datos. **Primera**: KAN supera a SVM en F1 macro ($p = 0.029$) con intervalo bootstrap completamente positivo, evidencia de mejora consistente en desempeño global. **Segunda**: la ventaja de KAN sobre SVM en recall(Medio) y F1(Medio) es estadísticamente significativa ($p < 0.02$) y de magnitud sustantiva (+20 pp en recall, +13 pp en F1), con KAN ganando en 9 de los 13 pliegues. **Tercera**: KAN no supera a MLP en ninguna métrica global (todos los intervalos incluyen cero), lo que es consistente con H1 —competitividad estadística, no superioridad—. KAN ofrece desempeño comparable a MLP añadiendo interpretabilidad estructural que MLP no provee. Esto confirma **H2**: ventaja consistente sobre el modelo lineal en la región transicional, con respaldo estadístico por pliegue.

6.4 Eficiencia Computacional: Costo del Ciclo de Reentrenamiento

La viabilidad operativa de un clasificador de riesgo depende también de su costo computacional de reentrenamiento. La Tabla 16 reporta los tiempos y consumo de memoria por pliegue.

SVM es el modelo más eficiente, coherente con su optimización convexa lineal. MLP requiere del orden de 3 minutos por pliegue. KAN tiene el mayor costo temporal —aproximadamente

Cuadro 16: Eficiencia computacional promedio por pliegue (media ± std, v3.4). Columna "Calib." aplica exclusivamente a SVM.

Modelo	Entrenamiento (s)	Inferencia (s)	RAM pico (MB)	Calib. (s)
SVM (lineal)	0.18 ± 0.04	0.006 ± 0.000	3.01 ± 0.55	0.031 ± 0.002
MLP	180.1 ± 65.9	0.002 ± 0.003	1.51 ± 0.16	—
KAN	656.8 ± 204.0	0.144 ± 0.055	1.45 ± 0.17	—

3.6× el de MLP—, coherente con la evaluación de transformaciones spline por arista. Significativamente, el consumo de memoria RAM de KAN (1.45 MB) es comparable al de MLP e inferior al de SVM, confirmando que la mayor carga de KAN es temporal, no espacial. Un ciclo semanal de reentrenamiento del sistema KAN sobre el universo de 30 emisoras requeriría del orden de 10–15 minutos en CPU convencional sin GPU dedicada: compatible con un flujo operativo de actualización semanal de señales de riesgo [76, 117, 189].

6.5 Hallazgo 4 — Interpretabilidad Estructural como Ventaja Operativa Diferencial

El cuarto hallazgo documenta la trazabilidad directa del mecanismo de decisión de KAN, cerrando el argumento central del trabajo: KAN no solo clasifica razonablemente bien, sino que expone directamente *por qué* clasifica como lo hace —propiedad que hace al sistema viable en entornos financieros regulados donde la auditabilidad del modelo es un requisito de gobernanza [55, 76, 189].

6.5.1 Ranking de importancia de variables: KAN (primera capa)

Cuadro 17: KAN: ranking de importancia de variables por magnitud en la primera capa.

Variable	Importancia
`Vol_252`	0.011429
`CVaR_5`	0.010458
`Beta`	0.009856
`R2`	0.009494
`Vol_60`	0.009470
`Amihud_60`	0.008358
`Alpha`	0.007901
`DownsideVol_60`	0.007837
`Vol_20`	0.007348
`r_log`	0.006698

Las variables de mayor peso son `Vol_252` (volatilidad anualizada), `CVaR_5` (riesgo de cola) y `Beta` (exposición sistemática) — las tres dimensiones canónicas del riesgo accionario según la teoría financiera [158, 214]. El rango de importancias es relativamente estrecho (0.0114 a 0.0067), indicando que KAN distribuye la contribución de forma equilibrada entre las diez variables en lugar de concentrarla en pocas: un clasificador de riesgo robusto no debe depender críticamente de una sola variable. La coherencia entre la jerarquía aprendida y el sentido económico esperado constituye evidencia de validez de constructo y confirma **H3** [46, 55].

En contraste, la Tabla 18 muestra que SVM concentra su peso en Beta y R^2: esto explica su alto desempeño en clases extremas (activos claramente sistemáticos vs. claramente idiosincrásicos), pero su rigidez en la zona transicional, donde los casos de riesgo intermedio con alta exposición sistemática se colapsan hacia la clase Alto.

Cuadro 18: SVM: importancia por norma ℓ_2 de coeficientes agregada entre clases.

Variable	**Importancia**
`Beta`	2.416779
`R2`	2.133818
`CVaR_5`	1.045585
`Vol_252`	0.392044
`Vol_60`	0.178534
`Alpha`	0.165874
`Amihud_60`	0.145324
`Vol_20`	0.128040
`DownsideVol_60`	0.096746
`r_log`	0.003576

6.5.2 Funciones $\varphi(x)$: trazabilidad funcional por variable

Las funciones $\varphi(x)$ aprendidas revelan propiedades de interés económico más allá del ranking de importancia. `Vol_252` exhibe una función monótonamente creciente con pendiente pronunciada en la región de volatilidad alta ($\sigma > 0.35$ anualizada), consistente con la evidencia de que el riesgo es no lineal en la cola de la distribución [158, 214]. `CVaR_5` muestra una respuesta asimétrica: suave en valores moderados y con incremento abrupto en la región de pérdidas extremas, capturando la discontinuidad empírica que separa episodios de estrés de mercado ordinario. `Beta` presenta una función en forma de U asimétrica, con mayor peso tanto para betas negativas (activos descorrelacionados con el mercado) como para betas superiores a 1.5 (activos de alta exposición sistemática), que la escala lineal del SVM no puede representar. La inspeccionabilidad directa de estas formas funcionales —sin necesidad de herramientas

externas— es la dimensión que diferencia cualitativamente a KAN de cualquier arquitectura de caja negra [55, 144, 189].

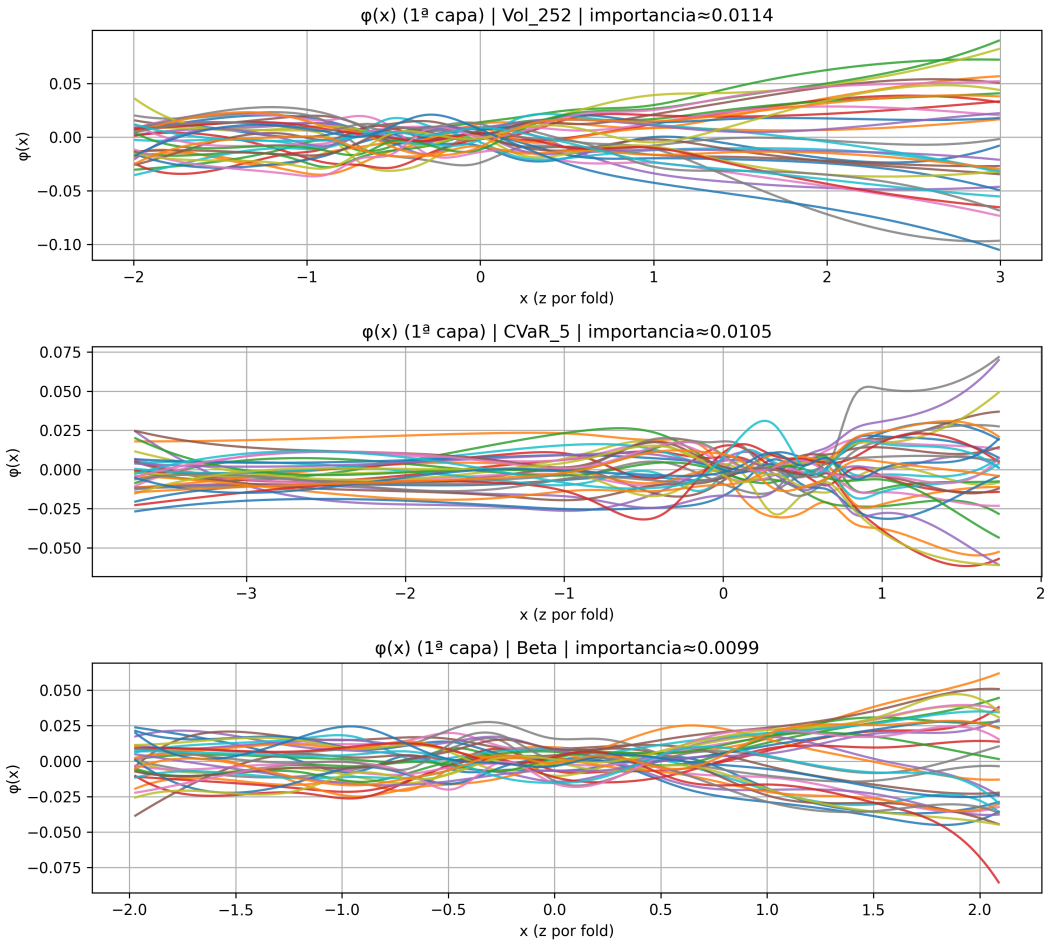


Figura 11: KAN: funciones univariadas $\varphi(x)$ de la primera capa para variables seleccionadas, con importancia asociada.

Esta visualización provee lo que ningún método *post-hoc* puede garantizar: el mecanismo exacto de transformación por variable. Un analista de riesgo puede identificar directamente en estas curvas qué valores de volatilidad disparan una reclasificación hacia Alto, dónde están los umbrales implícitos del modelo, y en qué regiones el sistema es más sensible a variaciones del mercado [27, 144]. Esto es precisamente lo que se requiere para defender una decisión de clasificación ante un comité de riesgo o un auditor externo [76, 189].

6.6 Calibración Probabilística de las Predicciones OOS

Los hallazgos anteriores caracterizan el desempeño *discriminativo* del sistema: qué tan bien separa las clases. Este hallazgo complementario caracteriza su **calibración probabilística**: qué tan confiables son las probabilidades emitidas como estimaciones de la frecuencia real de cada clase. La distinción es operativamente crítica para cualquier uso más allá de la clasificación ordinal binaria —especialmente para *position sizing* o estrategias que requieran probabilidades como insumo directo.

6.6.1 Error de calibración esperado (ECE) y Brier Score

La Tabla 19 reporta el Error de Calibración Esperado (ECE) por clase y modelo, calculado sobre las 16,133 predicciones OOS con 10 intervalos de confianza de amplitud uniforme. El ECE cuantifica la discrepancia promedio —ponderada por el número de observaciones en cada intervalo— entre la confianza predicha y la precisión real observada [90, 173].

Cuadro 19: Error de Calibración Esperado (ECE, 10 bins) y Brier Score OOS por modelo y clase. Menor es mejor en ambas métricas. Calculados sobre $n = 16{,}133$ predicciones estrictamente fuera de muestra.

Modelo	ECE Alto	ECE Bajo	ECE Medio	Brier global	Posición
KAN	0.0434	0.0791	0.0840	0.1248	3ro
MLP	**0.0367**	0.0723	0.0800	**0.1172**	1ro
SVM	0.0571	**0.0686**	**0.0761**	0.1198	2do

6.6.2 Diagramas de fiabilidad

La Figura 12 muestra los diagramas de fiabilidad (*reliability diagrams*) para los tres modelos y las tres clases en esquema OvR. Cada diagrama confronta la confianza predicha (eje x) con la precisión real observada (barra, eje y) y la línea de calibración perfecta ($y = x$, línea discontinua). Las barras por encima de la diagonal indican subestimación de confianza (*underconfidence*); por debajo, sobreestimación (*overconfidence*).

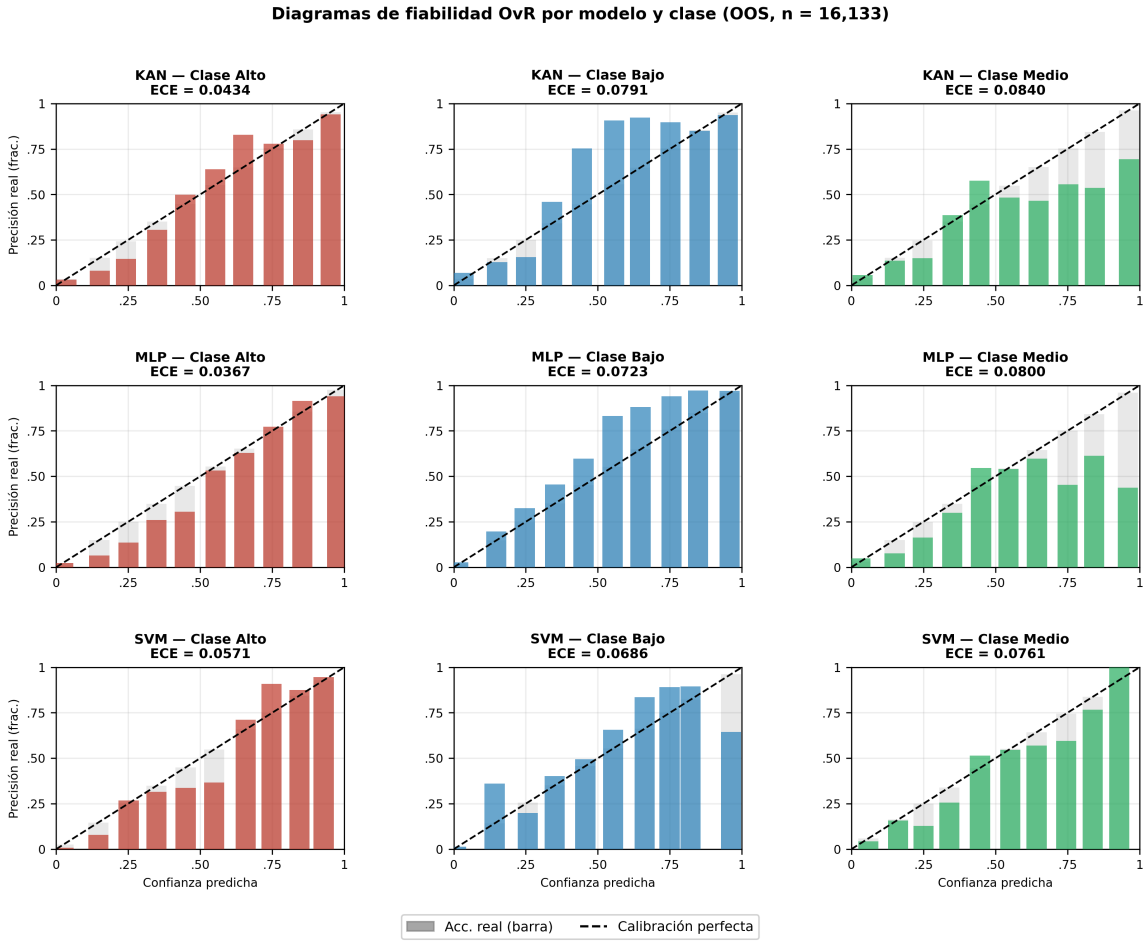


Figura 12: Diagramas de fiabilidad OvR para KAN, MLP y SVM (filas) por clase Alto, Bajo y Medio (columnas). Calculados sobre $n = 16{,}133$ predicciones OOS. La línea discontinua representa calibración perfecta ($y = x$).

6.6.3 Lectura de los resultados de calibración

Tres conclusiones emergen de los datos y la figura.

Primera: MLP tiene la mejor calibración global (Brier $= 0.1172$, ECE Alto $= 0.037$), coherente con que sea el modelo con mayor desempeño discriminativo agregado. KAN tiene la mayor ECE en las clases Bajo y Medio (0.079 y 0.084 respectivamente), sumando otro matiz a la relación entre los dos modelos neuronales: MLP es marginalmente superior en desempeño global, en estabilidad de recall(Medio) y en calibración probabilística. KAN diferencia por interpretabilidad intrínseca, no por ventaja en todo.

Segunda: la clase Medio tiene la peor calibración en los tres modelos —los ECE más altos

(0.076–0.084)— coherente con ser la región de mayor ambigüedad y menor separabilidad. Los diagramas muestran que en rangos de confianza media (0.4–0.7), todos los modelos sobreestiman su confianza para la clase Medio: predicen estar seguros con más frecuencia de la que aciertan. Este es el rango donde las probabilidades son más peligrosas si se usan directamente para *position sizing*.

Tercera: los ECE reportados (0.04–0.08) no son catastróficos —no estamos ante un modelo que emite probabilidades completamente aleatorias— pero sí son suficientes para descartar el uso directo de las probabilidades OvR como insumo de dimensionamiento de posiciones sin una etapa explícita de calibración. Un ECE de 0.084 en la clase Medio de KAN significa que, en promedio a través de los bins de confianza, el modelo sobreestima su certeza en esa clase en ≈ 8.4 puntos porcentuales. Para una institución que dimensiona exposición basada en esas probabilidades, esa discrepancia se traduciría en posiciones sistemáticamente mayores de lo que la evidencia OOS justifica en la región de riesgo transicional.

La corrección de esta limitación mediante *temperature scaling* sobre la ventana VAL de cada pliegue es directamente implementable con la infraestructura del pipeline existente y constituye la extensión más inmediata para habilitar el uso del sistema más allá de la clasificación ordinal [30, 90, 173].

6.7 Verificación Económica Ilustrativa: Coherencia del Pipeline OOS

La verificación económica del pipeline responde una pregunta específica: ¿tiene la señal de riesgo contenido informativo suficiente para que una regla de selección ingenua produzca retorno positivo fuera de muestra, sin ninguna optimización? La respuesta es afirmativa, con matices importantes.

Cuadro 20: Resumen del backtest OOS (KAN). La estrategia es *long-only* equiponderada en emisoras clasificadas Bajo riesgo, con costo de transacción del 0.15% sobre el *turnover* de cada rebalanceo semanal. El benchmark equiponderado (bench_eqw) es la media aritmética de los retornos de las 30 emisoras del universo en las fechas OOS.

Serie	CAGR	Sharpe	MaxDD	HitRate
Estrategia (bruta)	1.51%	0.178	−18.30%	50.4%
Estrategia (neta de costos)	0.77%	0.125	−18.74%	50.4%
Benchmark equal-weight OOS (30 em.)	6.79%	0.504	−15.40%	52.6%
Benchmark mercado (IPC)	4.94%	0.390	−16.82%	52.2%

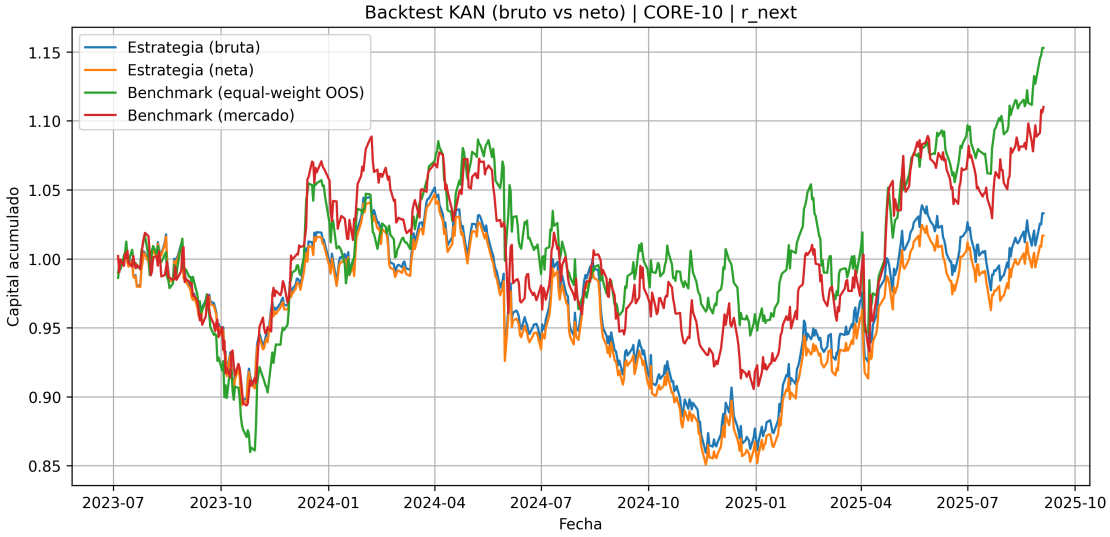


Figura 13: Backtest OOS (KAN): capital acumulado de la estrategia (bruta y neta) frente a benchmarks (equal-weight OOS de 30 emisoras e IPC).

Lectura honesta de los números. La estrategia genera un CAGR neto del 0.77%, que es un retorno nominal positivo pero por debajo de la inflación del período y marcadamente inferior al benchmark equiponderado (6.79%) y al IPC (4.94%). Esta brecha debe leerse con honestidad: la estrategia *long-only* de Bajo riesgo excluye las emisoras de mayor riesgo —y mayor retorno potencial en períodos alcistas—. Eso no es un defecto del clasificador: es la consecuencia esperada de su función como filtro de riesgo, no como motor de *alpha* [23].

La pregunta correcta: contenido económico, no superación del benchmark. El argumento relevante es si la señal de clasificación distingue activos con retornos futuros mejores de los peores, no si supera al IPC en períodos alcistas. La evidencia sugiere que sí: una regla sin ninguna optimización —sin gestión de pesos, sin restricciones sectoriales, sin control de riesgo de portafolio— genera CAGR bruto del 1.51% sobre 13 particiones cronológicas estrictamente OOS, lo que implica contenido informativo real en la señal [145]. La diferencia bruto/neto (1.51% vs. 0.77%) cuantifica el impacto de fricciones de mercado: la mayoría de estudios de ML financiero omite este cálculo, inflando el desempeño aparente; este análisis no lo hace [12].

El patrón de estabilidad bajo estrés. El valor más tangible de un filtro de riesgo no está en su CAGR promedio, sino en su comportamiento durante los períodos de mayor perturbación. La Tabla 13 documenta que en los pliegues 5 a 8 (marzo–noviembre de 2024), SVM colapsa

prácticamente su señal transicional: con recalls(Medio) de 0.168/0.178/0.100/0.080, el modelo lineal clasifica de forma prácticamente aleatoria la región de mayor ambigüedad, generando señales de selección volátiles e impredecibles. KAN, en esos mismos pliegues, mantiene recalls(Medio) de 0.790/0.655/0.550/0.568 —entre 3.7 y 7.1 veces superiores—. La inestabilidad de SVM ($\sigma_{\text{recall(Medio)}} = 0.271$) amplifica la incertidumbre de la señal precisamente cuando el mercado está bajo estrés, produciendo reclasificaciones volátiles que reducen la confiabilidad del filtro en el momento más crítico para un gestor de riesgo o un supervisor de SIEFORES [40, 41, 145].

Alcance y extensiones necesarias. El backtest es una **verificación de coherencia económica del pipeline**, no evidencia concluyente de ventaja de inversión. Las extensiones que convertirían esta verificación en evaluación concluyente requieren: (i) optimización formal de portafolio con restricciones de inversión, (ii) análisis de sensibilidad a costos y turnover, (iii) controles estadísticos de múltiples hipótesis, y (iv) evaluación del comportamiento durante episodios de Alto riesgo identificados correctamente por el modelo [12, 95, 145].

6.8 Lectura Ejecutiva de Resultados

Qué puede afirmarse con seguridad: KAN es estadísticamente equivalente a MLP en métricas globales OOS bajo un protocolo más riguroso que el empleado en los estudios comparables revisados para la BMV. Supera a SVM de forma consistente y estadísticamente significativa en la clasificación de la zona transicional (+18.7 pp recall Medio, $p < 0.02$, $n = 13$ pliegues). Tiene la menor variabilidad de F1 macro entre los tres modelos ($\sigma = 0.038$). Su jerarquía de importancias es coherente con la teoría financiera del riesgo. Su backtest OOS genera retorno nominal positivo sin ninguna optimización.

Qué no puede afirmarse: Que KAN supera a MLP en toda métrica (MLP es marginalmente superior en F1 global, AUC global y calibración probabilística: Brier $= 0.1172$ vs. 0.1248, ECE Medio $= 0.080$ vs. 0.084). Que KAN supera a MLP en recall(Medio) de forma estadísticamente significativa (la diferencia es 0.9 pp, no significativa). Que las probabilidades de clase son directamente usables para *position sizing* (los ECE de 0.076–0.084 en la clase Medio indican sobreestimación de confianza en esa región). Que la señal garantiza retorno ajustado por riesgo positivo (el backtest no incorpora optimización de portafolio). Que el RiskIndex equivale a un estándar exógeno de riesgo accionario.

Por qué ese alcance es suficiente: La pregunta de investigación no era cuál modelo tiene el

mejor F1 aislado. Era cuál arquitectura, de las evaluadas bajo el protocolo adoptado, cumple simultáneamente las tres condiciones de adoptabilidad en entornos financieros regulados. KAN es la única, de las tres evaluadas, que las cumple: equivalente a MLP en desempeño global, superior a SVM en la zona de mayor consecuencia operativa, más estable en F1 macro, e inspeccionable sin herramientas externas. Esa demostración, bajo un protocolo más riguroso que el empleado en los estudios comparables revisados para la BMV, constituye la contribución práctica del trabajo.

La contextualización histórica de estos resultados frente al único antecedente directo identificado en la literatura mexicana de clasificación de riesgo con IA en la BMV se desarrolla en la sección siguiente, que convierte las cuatro métricas en una narrativa de evolución del campo.

6.9 Perspectiva Histórica: Evolución desde el Antecedente Directo

Los resultados de este trabajo adquieren mayor dimensión cuando se contextualizan frente al único antecedente directo identificado en la literatura mexicana de clasificación de riesgo accionario con IA. Carmona Vega y González Garzón [35] resolvió este mismo problema en la BMV mediante una Red Neuronal Artificial Clasificadora sobre las 16 emisoras más activas del IPC, obteniendo una precisión clasificatoria superior al 70% sobre 384 observaciones trimestrales (2004–2009). Ese trabajo —reconocido con el Segundo Lugar del Premio Nacional BMV 2011 (Investigación)— estableció la viabilidad del enfoque y sentó un referente de facto para la clasificación de riesgo accionario con IA en México.

Dimensiones de evolución, no comparación directa. Una comparación directa de métricas entre ambos estudios no sería metodológicamente válida: los períodos, universos, frecuencias de observación y constructos de riesgo difieren de forma sustancial. Lo que sí es posible —y más iluminador— es identificar las dimensiones en que el campo ha avanzado en quince años:

Cuadro 21: Evolución del campo: antecedente directo vs. presente investigación. Las diferencias en período, frecuencia y protocolo hacen que la comparación de métricas no sea directa; se presentan como referencia contextual de la evolución metodológica del campo.

Dimensión	**Carmona Vega y González Garzón (2011)**	**Presente investigación**
Arquitectura	MLP (perceptrón multicapa, opaco)	KAN (interpretabilidad intrínseca)
Interpretabilidad	Ninguna (reconocida como "caja negra" por las autoras)	Funciones $\varphi(x)$ exactas por variable
Variable objetivo	Umbrales fijos de beta (< 0.70 / 0.70–1.14 / > 1.14)	RiskIndex: VolFut + $\|\beta\|$ + |CVaR| + ((cuantiles por pliegue)
Clases de riesgo	3 (Bajo, Medio, Alto)	3 (Bajo, Medio, Alto)
Variables de entrada	3 (K_j, CETES 91d, rendimientos accionarios)	10 (CORE-10: Beta, Alpha, R^2, Vol$\times$3, CVaR, Amihud, r_log, DsVol)
Frecuencia	Trimestral	Diaria
Emisoras evaluadas	16 más activas del IPC (2004–2009)	30 representativas del mercado (2020–2025)
Observaciones totales	384 (307 entrenamiento + 77 prueba, partición aleatoria)	16,133 (estrictamente fuera de muestra, Walk-Forward)
Precisión clasificatoria	74.02% accuracy (conjunto de prueba, $n = 77$)	75.7% accuracy, 75.0% F1 macro (OOS pooled, $n = 16{,}133$)
Protocolo validación	Partición aleatoria 80/20 (sin preservar causalidad temporal)	Walk-Forward PRO: 13 pliegues + purga + embargo
Comparación baselines	Sin comparación con baselines	MLP (arch-matched) + SVM lineal calibrado
Validación estadística	Métricas de clasificación estándar (MSE, matrices)	Wilcoxon pareado + IC bootstrap 95%, $n = 13$
Eficiencia computacional	No reportada	$\approx$657 s/pliegue, 1.45 MB RAM; ciclo semanal $\approx$10–15 min CPU
Reproducibilidad	Matlab 7.0	Python + fuentes públicas (Yahoo Finance, API Banxico) + código docu

Continuidad y superación. La diferencia más significativa entre ambos estudios no es la métrica de precisión —que no puede compararse directamente dado el cambio de protocolo— sino el estándar de evidencia al que se someten los resultados. Carmona Vega y González Garzón [35] demostró que el problema es soluble con RNA; la presente investigación demuestra que puede resolverse de forma simultáneamente auditable, temporalmente honesta y compu-

tacionalmente reproducible con datos públicos. Esta evolución responde a un cambio en las exigencias del campo: en 2011, la pregunta era si la IA podía clasificar riesgo; en 2025, la pregunta es si puede hacerlo de forma que un regulador, un auditor o un comité de inversión pueda verificar exactamente por qué [117, 118, 175, 189].

Impacto del protocolo en las métricas observadas. Es relevante notar que el Walk-Forward PRO con purga y embargo produce métricas conservadoras: el F1 macro de KAN sobre 16,133 observaciones OOS (0.750) se obtiene bajo condiciones donde el modelo nunca vio los datos que evalúa, los umbrales de clasificación se reestiman por pliegue sin contaminación y el período de prueba siempre es cronológicamente posterior al entrenamiento. Las métricas obtenidas bajo un protocolo estático —como el de Carmona Vega y González Garzón [35], cuya elección fue metodológicamente razonable para el estándar de la época— tienden a ser más elevadas en muestra, pero no replicables en condiciones reales de mercado [12, 145]. La comparación honesta no es entre el 70% de Carmona Vega y el 75% de KAN: es entre un resultado obtenido con validación estática y uno obtenido con un protocolo más riguroso que el empleado en los estudios comparables revisados para la BMV.

Una línea de continuidad verificable. El trabajo de Carmona Vega y González Garzón [35] y la presente investigación constituyen, en la literatura revisada, los únicos dos estudios de la literatura mexicana que abordan explícitamente la tarea de *clasificación* (y no meramente predicción) del riesgo accionario en la BMV mediante IA. La distancia entre ellos —quince años, una transición de MLP opaco a arquitecturas de interpretabilidad intrínseca y un salto de 384 observaciones con partición aleatoria a 16,133 con Walk-Forward PRO— refleja la maduración del campo a nivel global y, al mismo tiempo, el rezago relativo de la literatura financiera mexicana respecto a este problema específico: un rezago que este trabajo contribuye a cerrar.

6.10 Síntesis de Hallazgos

Los cuatro hallazgos responden directamente a las hipótesis planteadas:

1. **H1 confirmada**: KAN alcanza F1 macro de 0.750 y AUC de 0.880 sobre 16,133 observaciones OOS bajo validación temporal estricta. El desempeño es estadísticamente equivalente a MLP y superior a SVM en métricas agregadas — viable como herramienta de clasificación de riesgo en la BMV [86, 145].

2. **H2 confirmada**: KAN supera consistentemente a SVM en la clase Medio (recall +20 pp, F1 +13 pp), con diferencias estadísticamente significativas por Wilcoxon ($p < 0.02$) en 9 de 13 pliegues, incluyendo un colapso sostenido de SVM en 4 pliegues consecutivos. La menor variabilidad de KAN en recall(Medio) ($\sigma = 0.127$ vs. $\sigma_{\text{SVM}} = 0.271$) confirma que la ventaja es estable bajo diferentes regímenes de mercado.

3. **H3 confirmada**: el ranking de importancia estructural ubica a Vol_252, CVaR_5 y Beta como variables más influyentes —exactamente las que la teoría financiera predice como determinantes del riesgo accionario—. Las funciones $\varphi(x)$ son directamente inspeccionables sin herramientas *post-hoc* [55, 144].

4. **Verificación económica**: la señal *aparenta* capturar información con contenido económico real en el período evaluado (CAGR bruto OOS positivo sobre 13 pliegues independientes con costos explícitos de transacción) —resultado consistente con la hipótesis de viabilidad pero insuficiente, por sí solo, para establecer ventaja en portafolios de forma concluyente. La traducción a ventaja en portafolios requiere extensiones metodológicas adicionales [12, 95].

En conjunto, la evidencia respalda que KAN, entre las arquitecturas comparadas y bajo el protocolo adoptado, es la **única que cumple simultáneamente las tres condiciones de adoptabilidad en entornos financieros regulados**: MLP logra métricas globales marginalmente superiores pero no puede explicar sus decisiones sin herramientas de aproximación externa; SVM es interpretable en sus coeficientes pero colapsa la región transicional con variabilidad $\sigma = 0.271$ en recall(Medio). KAN combina desempeño OOS competitivo, interpretabilidad estructural directamente auditable, menor variabilidad en F1 macro entre pliegues —y esa combinación, en la literatura revisada, no había sido documentada para el mercado accionario mexicano. Esa triple condición, no la superioridad en una métrica aislada, es el argumento central de este trabajo [76, 144, 189].

VII. Implicaciones para el Sistema Financiero Mexicano

Los resultados empíricos del capítulo anterior establecen que el sistema propuesto es técnicamente viable y que KAN, entre las arquitecturas comparadas bajo el protocolo adoptado, es la única que simultáneamente alcanza desempeño OOS competitivo, provee interpretabilidad estructural directamente auditable y mantiene estabilidad temporal a través de múltiples regímenes de mercado —condición que, en la literatura revisada, no había sido documentada para el mercado accionario mexicano. Este capítulo aborda la aplicabilidad práctica del sistema y sus usuarios potenciales. Las respuestas se articulan en cinco planos: adopción operativa directa, implicaciones para el Sector Gobierno y la regulación financiera mexicana, extensiones naturales del marco, agenda de investigación que este trabajo habilita para México, y el alineamiento del sistema con la visión de evolución financiera sostenible del sistema financiero mexicano.

7.1 Adopción Operativa en Instituciones Financieras Mexicanas

El sistema de clasificación de riesgo propuesto es adoptable por instituciones financieras mexicanas sin requerir datos propietarios ni infraestructura especializada. Las tres características que habilitan esta adopción son: fuentes públicas, señal interpretable y protocolo de reentrenamiento definido.

7.1.1 Fuentes de datos públicamente disponibles

Las diez variables del CORE-10 se calculan íntegramente a partir de datos accesibles en tiempo real con costo cero: precios ajustados y volúmenes de Yahoo Finance, la tasa de CETES mediante el API REST del SIE de Banxico, y el benchmark IPC bajo el mismo protocolo. En ningún paso del pipeline se requiere acceso a terminales de datos propietarios, bases de datos de Bloomberg o Refinitiv, ni información de flujo de órdenes. Cualquier área de análisis cuantitativo de una casa de bolsa, fondo de inversión o institución regulatoria mexicana puede replicar el sistema en su entorno de trabajo con las mismas herramientas de uso libre que se emplearon en esta investigación [9, 17].

7.1.2 Integración como filtro de primer nivel en el proceso de análisis

El caso de uso más directo es como filtro de primer nivel antes del análisis fundamental. En lugar de que un equipo evalúe manualmente el perfil de riesgo de las 30 o más emisoras de un universo de inversión, el clasificador provee una preclasificación diaria automática (Bajo / Medio / Alto) basada en variables de mercado observables. El analista concentra su atención

en las emisoras cuyo perfil cambia de categoría —señal de cambio de régimen de riesgo— o en aquellas que permanecen en la zona Medio, donde la ambigüedad requiere criterio humano que ningún modelo reemplaza. La interpretabilidad estructural de KAN hace posible que, cuando un comité de inversión pregunta por qué una emisora fue clasificada como Alto riesgo, la respuesta sean las funciones $\varphi(x)$ directamente inspeccionables.

Límite operativo. El sistema provee **clasificaciones ordinales** (Bajo / Medio / Alto) como capa de filtrado de primer nivel. No reemplaza el análisis fundamental ni los modelos cuantitativos de gestión de portafolios; actúa como una señal preliminar estructurada que aumenta la eficiencia del análisis humano, no que lo sustituye.

7.1.3 Protocolo de reentrenamiento operativo

El esquema Walk-Forward PRO con ventana de entrenamiento expandida es directamente transferible a un ciclo operativo semanal: el modelo incorpora nuevos datos de mercado, actualiza los umbrales de discretización por cuantiles sobre el período acumulado y genera nuevas clasificaciones para la semana siguiente. El costo computacional documentado en la Tabla 16 confirma que este ciclo es viable en hardware convencional sin GPU dedicada para el universo de 30 emisoras [145].

7.1.4 Alcance y límites del uso operativo: clasificación ordinal vs. probabilidades calibradas

Una distinción operativa central que debe quedar explícita: el sistema provee **clasificaciones ordinales** (Bajo / Medio / Alto) diseñadas para uso como filtro de primer nivel. No provee probabilidades calibradas aptas para el dimensionamiento dinámico de posiciones.

La evidencia empírica cuantifica esta limitación. El Error de Calibración Esperado (ECE) de KAN en la clase Medio es 0.084, lo que indica que las probabilidades predichas sobreestiman la confianza del modelo en esa región en ≈ 8.4 puntos porcentuales en promedio a través de los intervalos de confianza OOS. El Brier Score global de KAN (0.1248) es superior al de MLP (0.1172), quien muestra mejor calibración general. Estos valores no implican que las probabilidades sean inútiles —la confianza de KAN es informativa: con predicciones de confianza superior al 80%, la *accuracy* real es del 85.1%— pero sí confirman que no son aptas para uso directo como insumo de dimensionamiento de posiciones sin ajuste previo [90, 173]. Para un gestor que dimensiona exposición basado en esas probabilidades, la discrepancia ECE de 0.084 en Medio se traduciría en posiciones sistemáticamente mayores de lo que la evidencia OOS justifica en la región transicional.

Por ello, cualquier uso como *overlay* de riesgo con position sizing debe incorporar una etapa de calibración probabilística —mediante *temperature scaling* sobre la ventana VAL de cada pliegue— y verificación de confiabilidad mediante diagramas de fiabilidad, ECE y Brier Score antes del despliegue [30, 90].

Esta limitación es intencional: el sistema fue diseñado para proveer la capa de clasificación auditable que *precede* a los modelos de gestión cuantitativa de portafolios, no para reemplazarlos. La extensión hacia probabilidades calibradas es la línea de trabajo futuro más directa para habilitar el uso del sistema en position sizing [90, 173].

7.1.5 Beneficio incremental frente a la práctica actual

El valor del sistema se articula con precisión al contrastarlo con lo que hoy existe:

- **Frente al *screening* discrecional de riesgo:** el sistema provee señales sistemáticas, reproducibles y auditables, eliminando el sesgo de confirmación del analista y aumentando la consistencia de la clasificación a lo largo del tiempo. Dos analistas aplicando el mismo modelo al mismo universo obtendrán la misma clasificación; eso no ocurre con el análisis discrecional.

- **Frente a modelos sin validación temporal rigurosa:** el protocolo Walk-Forward PRO garantiza que el desempeño reportado es el real OOS, no un artefacto de particiones aleatorias que inflan artificialmente la precisión aparente. Un modelo que parece funcionar con el 90% de accuracy en validación cruzada estándar puede degradarse a 60% en uso real bajo series temporales con cambios de régimen; el protocolo adoptado aquí cierra ese riesgo de forma exacta [5, 145].

- **Frente a modelos de caja negra con explicación post-hoc:** el sistema provee la lógica exacta de clasificación —las funciones $\varphi(x)$— que permite auditoría regulatoria sin herramientas cuya fidelidad no puede verificarse independientemente. Una explicación SHAP de un modelo de ensamble es una aproximación del mecanismo real; las funciones de KAN son el mecanismo real.

- **Frente a la ausencia de un índice exógeno de riesgo accionario:** el sistema cubre parcialmente ese vacío con una metodología reproducible y pública, mientras señala con precisión qué extensión lo cerraría definitivamente. No es el índice exógeno que México necesita, pero es la infraestructura sobre la cual puede construirse [108, 166].

- **Frente al único sistema de clasificación con IA previo en la BMV:** el sistema de Carmona Vega y González Garzón [35] demostró que la IA puede clasificar el riesgo de mercado de emisoras del IPC con más del 70% de precisión bajo un protocolo de validación estática. El sistema propuesto en esta investigación mantiene un nivel comparable de desempeño (F1 macro de 0.750 sobre 16,133 observaciones OOS) bajo un protocolo de validación sustancialmente más exigente, con el beneficio adicional de interpretabilidad estructural directamente inspeccionable —la dimensión que los marcos regulatorios actuales de IOSCO y el BIS identifican como el requisito pendiente de la primera generación de clasificadores de riesgo con IA— [117, 118, 175].

7.2 Implicaciones para el Sector Gobierno: CONSAR, CNBV y Banxico

El enfoque macrofinanciero de esta investigación —orientado al Sector Gobierno y al Mercado de Valores— exige articular el argumento regulatorio con sustancia concreta. Esta sección desarrolla tres casos de uso institucional gubernamental directos, verificados a partir de la normativa vigente.

7.2.1 CONSAR y las SIEFORES: señal granular complementaria para supervisar el riesgo accionario del ahorro nacional

La Comisión Nacional del Sistema de Ahorro para el Retiro (CONSAR) supervisa el régimen de inversión de las Sociedades de Inversión Especializadas de Fondos para el Retiro (SIEFORES), que gestionan 8.3 billones de pesos —equivalentes al 23.8% del PIB nacional— en cerca de 70 millones de cuentas individuales [42]. El régimen vigente, actualizado mediante modificaciones publicadas en el Diario Oficial de la Federación en octubre de 2024, establece límites de exposición a renta variable nacional diferenciados por cohorte generacional y utiliza el Diferencial del Valor en Riesgo Condicional (DCVaR) como métrica de riesgo regulatoria [40].

El DCVaR mide el riesgo *agregado* de la cartera de cada SIEFORE, pero no provee a CONSAR ni a las Afores una señal sistemática y auditable del **perfil de riesgo individual por emisora** que les permita anticipar cambios de régimen antes de que se reflejen en el DCVaR de la cartera completa. El sistema KAN propuesto llena exactamente ese vacío dimensional: una preclasificación diaria por emisora en tres niveles de riesgo, con interpretabilidad estructural verificable y validación temporal rigurosa. El caso de uso concreto: CONSAR podría integrar señales de este tipo en su sistema de monitoreo para identificar, antes de cada período de rebalanceo, si alguna SIEFORE está incrementando su exposición a emisoras clasificadas

como Alto riesgo —especialmente en las Siefores de cohortes más jóvenes, donde el régimen permite mayor exposición a renta variable pero donde un error de clasificación tiene el mayor impacto acumulado sobre el ahorro de largo plazo.

A modo ilustrativo, un flujo operativo compatible con los recursos técnicos disponibles en cualquier área de análisis cuantitativo mexicana podría diseñarse de la siguiente forma: cada lunes antes del inicio de operaciones, el sistema *procesaría* los precios de cierre del viernes y actualizaría el CORE-10 para las 30 emisoras. Las emisoras que cambiaran de categoría (p.ej., de Bajo a Medio o de Medio a Alto) generarían una alerta que CONSAR podría revisar antes del período de rebalanceo semanal de las SIEFORES. Si en un período de seis semanas consecutivas tres o más emisoras de un sector migraran de Bajo a Alto riesgo —patrón que captura cambios de régimen sectoriales antes de que se reflejen en el DCVaR agregado de cartera— la señal operativa precedería al indicador regulatorio vigente por semanas, habilitando intervención preventiva en lugar de reactiva. Esta no es una capacidad marginal: es la diferencia entre monitoreo prospectivo y monitoreo retrospectivo en la gestión del riesgo del ahorro nacional.

Límite operativo. El sistema provee una señal ordinal diaria por emisora como **capa auxiliar de monitoreo granular**. No reemplaza el DCVaR ni las métricas prudenciales de cartera existentes; no es una medida de riesgo regulatoria certificada ni equivale a un estándar exógeno de riesgo accionario. Actúa como señal complementaria que opera en la dimensión individual (por emisora) donde el DCVaR agregado de cartera no llega. La adopción regulatoria requeriría validación adicional por parte de CONSAR, incluyendo pruebas de calibración y backtesting sobre un universo de emisoras más amplio.

7.2.2 CNBV y la gobernanza algorítmica: IA explicable como requisito de supervisión

La Comisión Nacional Bancaria y de Valores (CNBV), como miembro activo de IOSCO, ha adoptado los marcos regulatorios internacionales sobre el uso responsable de IA en los mercados financieros. El reporte IOSCO FR06/2021 establece seis medidas de conducta esperada para intermediarios y gestores de activos que utilizan IA y ML, entre las que destacan la gobernanza y supervisión adecuada, el monitoreo continuo y la **transparencia y explicabilidad del mecanismo de decisión** [117]. En 2025, IOSCO publicó el reporte de consulta CR/01/2025, que amplía este análisis hacia los riesgos y desafíos de la IA en mercados de capitales [118]. El BIS y el FSB han publicado directrices adicionales sobre explicabilidad de IA en el sector financiero [65, 175].

La interpretabilidad estructural de KAN responde directamente a este marco. Cuando la CNBV solicita a un intermediario evidencia de que su sistema algorítmico de clasificación de riesgo es comprensible y auditable, la diferencia entre KAN y un modelo de caja negra no es técnica de grado: es cualitativa:

- **Modelo de caja negra con explicación post-hoc**: el intermediario presenta una aproximación SHAP o LIME de la lógica del modelo. El supervisor acepta una representación *aproximada* cuya fidelidad depende de supuestos del método explicativo que son, a su vez, opacos.

- **Sistema KAN**: el intermediario presenta directamente las funciones $\varphi(x)$ aprendidas por variable, que *son* el mecanismo exacto de decisión. No hay brecha entre el modelo y su explicación porque la arquitectura misma es la explicación.

Límite operativo. La interpretabilidad de KAN documenta la lógica exacta del proceso de clasificación, pero no garantiza que el RiskIndex sea un estándar exógeno y certificado de riesgo accionario. Las decisiones supervisoras deben considerar el alcance inferencial de la medida descrito en el Capítulo IV (§3.6, Distinción 1): el RiskIndex es una medida operativa de riesgo relativo construida con variables de mercado, no un estándar prudencial independiente. Su uso en supervisión regulatoria requeriría certificación adicional por parte del organismo regulador correspondiente.

7.2.3 Banxico y el CESF: el protocolo Walk-Forward como estándar de validación para modelos de riesgo

El Consejo de Estabilidad del Sistema Financiero (CESF), integrado por la SHCP, Banxico, CNBV, CONSAR y CNSF, publica evaluaciones periódicas del riesgo sistémico que combinan indicadores macroeconómicos, financieros y de mercado [43]. El clasificador puede funcionar como insumo de granularidad micro —por emisora y por fecha— que complemente los indicadores del CESF, proveyendo señales de alerta temprana para el mercado de renta variable antes de que el riesgo se materialice a nivel sistémico.

Más allá de este caso de uso, el protocolo Walk-Forward PRO tiene una implicación metodológica para la supervisión: establece un estándar de validación más riguroso que el utilizado en los modelos de riesgo de mercado revisados que los intermediarios presentan ante supervisores. La práctica de validación cruzada estándar con particiones aleatorias produce estimaciones de desempeño que no se replican en uso real [5, 12, 145]. Un regulador que adopte el protocolo

Walk-Forward con purga y embargo como criterio de evaluación de modelos internos tendría mayor certeza de que el desempeño reportado es el desempeño real fuera de muestra.

7.3 Extensiones Naturales del Marco Propuesto

El marco metodológico desarrollado es transferible a dominios relacionados dentro del sistema financiero mexicano.

7.3.1 Renta fija y deuda corporativa

La lógica de clasificación ordinal del riesgo es conceptualmente análoga a la de las calificadoras de riesgo crediticio. El marco puede adaptarse a la clasificación de riesgo de instrumentos de deuda corporativa mexicana, sustituyendo las variables del CORE-10 por variables de estructura de capital, cobertura de deuda, márgenes operativos y liquidez, con un índice de riesgo basado en spreads crediticios observados o probabilidades de incumplimiento históricas [108, 166].

7.3.2 Análisis sectorial con variables macroeconómicas

El RiskIndex puede adaptarse para análisis sectorial incorporando variables macroeconómicas específicas: tipo de cambio y tasa de interés para el sector financiero, precios de materias primas para minería e industria, variables de consumo para el sector retail. Esta extensión permitiría construir señales de riesgo con mayor especificidad económica, manteniendo el mismo protocolo de validación temporal y la misma arquitectura de clasificación interpretable.

7.3.3 Integración con indicadores de riesgo sistémico del CESF

El clasificador puede integrarse como insumo de granularidad micro para los modelos de riesgo sistémico del CESF, en un tablero de monitoreo que combine la clasificación diaria por emisora con los indicadores macro del CESF, proveyendo una vista de dos niveles del riesgo: el sistémico agregado y el específico por activo [43].

7.4 Agenda de Investigación para México: el Vacío que Este Trabajo Identifica

Esta investigación no solo propone un sistema: identifica un vacío de investigación con impacto directo sobre el sistema financiero mexicano y establece la infraestructura metodológica para cerrarlo.

7.4.1 La ausencia de un índice de riesgo accionario exógeno para la BMV

La limitación conceptual más relevante del RiskIndex apunta hacia una brecha más profunda: México no cuenta con un índice de riesgo accionario exógeno, certificado y de referencia pública para el mercado de renta variable, análogo al rol que las calificadoras desempeñan en el mercado de deuda. En el mercado de deuda corporativa, las calificaciones de HR Ratings, S&P y Fitch proveen una señal de riesgo independiente de las variables de mercado utilizadas para valorar los instrumentos. En el mercado accionario mexicano, no existe un equivalente estructurado, lo que implica que la supervisión del riesgo accionario descansa en métricas agregadas (DCVaR de cartera) sin una referencia externa por emisora que permita validar si esas métricas agregan correctamente riesgos individuales.

7.4.2 Una propuesta concreta para cerrar el vacío

El desarrollo de un índice de riesgo accionario exógeno para la BMV requeriría tres ingredientes que este trabajo contribuye a establecer: (i) un marco de variables observables de mercado que capturen múltiples dimensiones del riesgo —el CORE-10 es un punto de partida validado—; (ii) una variable objetivo construida exclusivamente con información de mercado futura totalmente separada del espacio de predictores; y (iii) un protocolo de validación temporal que garantice que el índice tiene poder predictivo real y no es un artefacto de ajuste histórico [12, 145].

Así como las calificadoras elevaron la transparencia y comparabilidad del riesgo crediticio, un índice de riesgo accionario certificado elevaría el estándar de gestión de riesgo en el mercado de renta variable, proveyendo a CONSAR y a la CNBV una referencia auditable del perfil de riesgo individual de las emisoras de la BMV, complementaria a las métricas de riesgo agregado existentes.

7.4.3 La contribución metodológica de este trabajo a esa agenda

El protocolo Walk-Forward PRO con purga y embargo, la arquitectura KAN con interpretabilidad estructural verificable, y el pipeline reproducible con fuentes públicas que este trabajo establece son la infraestructura metodológica sobre la cual puede construirse esa investigación. La contribución va más allá de los resultados empíricos inmediatos: establece un estándar de evaluación más riguroso para el ML financiero aplicado a la BMV y provee las herramientas con las que México puede cerrar de forma metódica el vacío identificado [12, 144, 145].

7.5 Evolución Financiera Sostenible: Fuentes Públicas como Democratización del Análisis de Riesgo

Un atributo del sistema que adquiere especial relevancia en el contexto de la agenda contemporánea de evolución financiera sostenible es su dependencia **exclusiva en fuentes de datos públicas y gratuitas** (Yahoo Finance, API REST de Banxico) y su implementación con herramientas de uso libre.

La democratización del análisis de riesgo riguroso no es un detalle técnico de implementación: es una condición de equidad del sistema financiero. Una herramienta de clasificación de riesgo que requiere datos propietarios de alto costo solo es accesible para los participantes mejor equipados —las casas de bolsa de mayor tamaño, los fondos institucionales con presupuestos significativos de infraestructura de datos, o las instituciones extranjeras con presencia en México—. Una herramienta construida con fuentes abiertas puede ser replicada, auditada y mejorada por cualquier participante del sistema financiero mexicano, incluyendo los organismos reguladores, las Afores de menor tamaño, las instituciones académicas y los investigadores independientes.

En ese sentido, la reproducibilidad sin costo del pipeline es, por sí misma, una contribución a la *evolución financiera sostenible*: las herramientas que elevan el estándar metodológico del análisis de riesgo en el mercado mexicano deben ser accesibles para todos sus participantes para que ese estándar tenga impacto real, no solo para quienes ya cuentan con los recursos para construirlas. El sistema propuesto demuestra que esa combinación —rigor metodológico, interpretabilidad auditable y reproducibilidad abierta— es alcanzable sin depender de datos propietarios ni de infraestructura de cómputo especializada. Esa es la dimensión de su contribución que más directamente se alinea con el objetivo de un sistema financiero mexicano más transparente, equitativo y metodológicamente robusto.

VIII. Conclusiones

8.1 Síntesis General del Trabajo

El argumento central que este trabajo ha construido capítulo a capítulo se puede enunciar de forma directa: en el mercado accionario mexicano, bajo validación temporal estricta, es posible construir un sistema de clasificación de riesgo que sea simultáneamente competitivo en desempeño OOS, auditable en su mecanismo de decisión y metodológicamente honesto en su protocolo de evaluación. KAN, entre las arquitecturas comparadas bajo el protocolo adoptado, es la que mejor demuestra esa posibilidad; y esa demostración, junto con la infraestructura metodológica que la sostiene, es la contribución durable de este trabajo.

La investigación integró un pipeline reproducible de adquisición de datos, ingeniería de variables, etiquetado y validación temporal. Se adoptó un esquema Walk-Forward PRO con purga y embargo para mitigar dependencia inducida por ventanas móviles y etiquetas con horizonte futuro, alineado con las buenas prácticas de ML financiero [12, 13, 145]. El desempeño se evaluó con métricas estándar de clasificación multiclase y separabilidad ROC/AUC, complementadas con evidencia inferencial pareada por pliegue (Wilcoxon) e intervalos bootstrap [52, 56, 63, 93, 179, 201].

8.2 Cierre de Hipótesis

Las tres hipótesis planteadas en el Capítulo III se cierran con la evidencia empírica del Capítulo VI:

1. **H1 confirmada — Desempeño predictivo competitivo.** Bajo evaluación OOS estricta, KAN alcanzó F1 macro de 0.750 y AUC OvR macro de 0.880 sobre 16,133 observaciones en 13 pliegues cronológicos. El desempeño es estadísticamente equivalente al de MLP (Wilcoxon $p = 0.812$, bilateral) y superior al de SVM en métricas agregadas ($p = 0.029$). En un entorno financiero no estacionario, alcanzar este nivel bajo validación temporal robusta constituye evidencia relevante de la viabilidad del enfoque [86, 145].

2. **H2 confirmada — Ventaja consistente en la región transicional.** La clase Medio concentra la dificultad de clasificación. KAN presentó recall de Medio de 65.3% frente al 64.4% de MLP y al 46.6% de SVM. La ventaja KAN–SVM de 18.7 pp está respaldada por evidencia inferencial pareada (Wilcoxon, $p = 0.016$) con intervalos bootstrap completamente positivos [+0.059, +0.346], confirmando que la mejora fue consistente a través de las

13 particiones temporales y no un fenómeno aislado. KAN supera a SVM en recall(Medio) en 9 de los 13 pliegues, con un colapso sostenido de SVM de 4 pliegues consecutivos en 2024. La diferencia KAN–MLP en recall(Medio) (0.9 pp, $p = 0.420$) no es estadísticamente significativa; ambos modelos son equivalentes en esta dimensión [91, 145, 214].

3. **H3 confirmada — Interpretabilidad estructural verificable.** KAN ofreció interpretabilidad estructural mediante: (i) ranking de importancia en la primera capa, donde `Vol_252`, `CVaR_5` y `Beta` emergen como las variables más influyentes —exactamente las que la teoría financiera predice como determinantes del riesgo accionario—, y (ii) visualización directa de funciones univariadas $\varphi(x)$ aprendidas por variable. Esta capacidad no depende de métodos *post-hoc*: emerge de la arquitectura misma. La coherencia entre jerarquía aprendida y sentido económico constituye evidencia de validez de constructo del sistema [46, 55, 144, 189].

4. **Verificación económica: evidencia exploratoria positiva.** El backtest OOS construido exclusivamente con predicciones fuera de muestra generó un CAGR bruto positivo (1.51%, neto 0.77%) sobre las 13 particiones temporales independientes, con costos de transacción explícitos. Esto sugiere que la señal *aparenta* capturar información con contenido económico real en el período evaluado —resultado consistente con la hipótesis de viabilidad pero insuficiente, por sí solo, para establecer ventaja en portafolios de forma concluyente. No obstante, este componente es una verificación de coherencia del pipeline, no evidencia concluyente de ventaja en portafolios; cualquier conclusión operativa requiere extensiones metodológicas con controles de múltiples hipótesis y optimización formal [12, 95, 145].

8.3 Cumplimiento de Objetivos

Los seis objetivos específicos del Capítulo III se consideran cumplidos en los siguientes términos:

1. **Dataset financiero público y reproducible.** Se construyó un conjunto de datos con 30 emisoras de la BMV, 40,649 observaciones efectivas en el período 2020–2025, basado íntegramente en fuentes públicas. El pipeline es reproducible sin acceso a datos propietarios [1, 214].

2. **RiskIndex: índice multivariado con etiquetado sin look-ahead.** Se diseñó el RiskIndex integrando cuatro dimensiones de riesgo con estandarización transversal, y se implementó

discretización ordinal con cuantiles estimados estrictamente dentro de cada partición de entrenamiento [145, 158].

3. **Protocolo Walk-Forward PRO con purga y embargo.** Se implementó el esquema con 13 particiones cronológicas, con gap $= H = 60$ días que cierra el look-ahead de forma exacta, estableciendo un estándar metodológico más riguroso que el empleado en los estudios comparables revisados para la BMV [12, 13, 145].

4. **Sistema KAN para clasificación de riesgo.** Se implementó KAN en PyTorch con B-splines cúbicos ($p = 3$, $n = 10$ nudos internos), ajuste de logits, fringe denoising y early stopping. El pipeline es reproducible para entrenamiento, predicción OOS e interpretación estructural [27, 144].

5. **Comparación *apples-to-apples* con SVM y MLP.** Se realizó la comparación bajo condiciones estrictamente equivalentes con contrastes Wilcoxon e intervalos bootstrap por pliegue [52, 56].

6. **Interpretabilidad estructural documentada.** Se extrajeron y visualizaron las funciones $\varphi(x)$ de la primera capa, se construyó el ranking de importancia global y se contrastó con la teoría financiera del riesgo, sin recurrir a métodos *post-hoc* [55, 144, 189].

8.4 Aportaciones Principales

Las aportaciones se organizan en tres planos que reflejan los distintos niveles de impacto del trabajo. La tesis de fondo es que KAN, entre las arquitecturas evaluadas bajo el protocolo adoptado, es la única que cumple simultáneamente las tres condiciones de adoptabilidad en entornos financieros regulados —condición que, hasta donde alcanza la revisión sistemática realizada, no había sido documentada para el mercado accionario mexicano:

Plano metodológico. El protocolo Walk-Forward PRO con purga, embargo y gap $= H$ establece un estándar de evaluación que elimina las tres fuentes de look-ahead documentadas en ML financiero. Su reproducibilidad con fuentes públicas lo hace transferible directamente a cualquier estudio de ML financiero en México que trabaje con series de tiempo. La adopción de este estándar como criterio de evaluación elevará la calidad inferencial del campo y facilitará la comparabilidad entre estudios.

Plano empírico. Sobre 16,133 observaciones OOS en 13 pliegues cronológicos, se documenta que KAN es estadísticamente equivalente a MLP en métricas globales, supera de forma consistente y significativa a SVM en recall(Medio) (+18.7 pp, $p < 0.02$, Wilcoxon), mantiene la menor variabilidad en F1 macro entre los tres modelos ($\sigma = 0.038$), y provee una jerarquía de importancias coherente con la teoría financiera del riesgo. Un matiz que la honestidad científica obliga a consignar: MLP es marginalmente superior a KAN en F1 macro global y en estabilidad de recall(Medio) ($\sigma_{\text{MLP}} = 0.097 < \sigma_{\text{KAN}} = 0.127$); la ventaja de KAN es relacional —la combinación de desempeño, estabilidad e interpretabilidad en un mismo sistema— no la superioridad en ninguna métrica individual.

Plano institucional. El pipeline reproducible con fuentes públicas, el diseño del CORE-10, la infraestructura Walk-Forward PRO y la arquitectura KAN son la base sobre la cual puede construirse el paso de mayor impacto institucional: un índice exógeno de riesgo accionario por emisora para la BMV, que provea a CONSAR y la CNBV una referencia granular análoga al rol de las calificadoras en el mercado de deuda. El sistema propuesto no es ese índice; es la demostración de que puede construirse de forma rigurosa y auditable.

8.5 Limitaciones

Delimitar con precisión el alcance de un estudio es una señal de madurez científica. Las siguientes limitaciones no invalidan las conclusiones del trabajo: las acotan con exactitud, y en varios casos esa acotación señala la agenda de investigación de mayor valor para el sistema financiero mexicano.

1. **Dependencia del esquema de etiquetado y horizonte temporal.** El desempeño depende de cómo se define el riesgo y del horizonte utilizado ($H = 60$ días). Cambios en el horizonte pueden alterar la separabilidad efectiva de clases y el balance entre ellas [214].

2. **No estacionariedad y sensibilidad a régimen.** Con validación temporal estricta, los modelos siguen siendo susceptibles a cambios estructurales no anticipados. La variabilidad por pliegue documentada es esperada en finanzas y debe considerarse al extrapolar a períodos futuros [86, 145].

3. **CORE-10 representativo pero no exhaustivo.** Las diez variables capturan las dimensiones fundamentales del riesgo, pero no incorporan variables fundamentales, macroeconómicas o de microestructura avanzada que podrían enriquecer la señal bajo regímenes específicos

[86, 214].

4. **Sesgo de supervivencia en el universo de emisoras.** Las 30 emisoras fueron seleccionadas por su cotización activa al momento del análisis (2026). Las empresas que cotizaban en 2020 pero que fueron descontinuadas, suspendidas o fusionadas no están representadas, lo que introduce un sesgo de supervivencia moderado que tiende a sobreestimar el desempeño de las categorías de menor riesgo. Este sesgo afecta por igual a los tres modelos y no compromete la comparación entre arquitecturas, pero sí debe considerarse al extrapolar los resultados de desempeño absoluto a universos más amplios.

5. **Backtest como evaluación exploratoria.** El backtest no incorpora optimización formal de portafolios y está sujeto a los riesgos de sobreajuste propios de la experimentación financiera [5, 12, 95].

6. **Asimetría paramétrica KAN–MLP.** La comparación "arch-matched" iguala dimensiones de capas pero no parámetros (≈6,691 en KAN vs. ≈451 en MLP). La mayor capacidad paramétrica de KAN contribuye a su potencial expresivo pero también a su costo computacional; este trade-off es consecuencia de la arquitectura y debe considerarse en entornos con recursos computacionales limitados.

Validez de constructo del RiskIndex y propuesta de extensión. El RiskIndex integra componentes relacionados con algunos predictores del CORE-10 ($|\beta|$, $|\text{CVaR}|$, $1 - R^2$), lo que constituye una amenaza a la validez de constructo: el objetivo comparte componentes con el espacio de variables explicativas [46, 197]. Esta limitación no introduce fuga temporal —los umbrales se estiman por pliegue y VolFut nunca entra como predictor— pero condiciona la interpretación económica del objetivo a **clasificación predictiva de riesgo relativo**, no inferencia causal. El hallazgo de esta circularidad es en sí mismo una contribución: señala con precisión el vacío que el sistema financiero mexicano necesita cerrar mediante un índice de riesgo accionario exógeno para la BMV, basado exclusivamente en variables de mercado futuras separadas limpiamente del espacio de predictores [145, 166].

Probabilidades no calibradas para uso operativo. La evidencia empírica cuantifica esta limitación con precisión: el ECE de KAN en la clase Medio es 0.084 y su Brier Score global es 0.1248, frente a 0.080 y 0.1172 de MLP respectivamente. Estos valores confirman que las probabilidades de clase sobreestiman la confianza del modelo en la región transicional en ≈ 8.4 puntos porcentuales en promedio. La señal sigue siendo informativa —con predicciones

de confianza superior al 80%, la *accuracy* real de KAN es 85.1%— pero no apta para dimensionamiento directo de posiciones. Cualquier uso operativo con position sizing debe incorporar calibración por pliegue (temperature scaling sobre la ventana VAL de cada pliegue) y verificación mediante diagramas de fiabilidad, ECE y Brier Score antes del despliegue [30, 90, 173].

8.6 Recomendaciones

La evidencia empírica y la experiencia metodológica acumulada en este trabajo permiten formular tres recomendaciones concretas dirigidas a audiencias distintas dentro del sistema financiero mexicano.

Recomendación 1 — Metodológica (comunidad investigadora). Adoptar el protocolo Walk-Forward con purga, embargo y gap $= H_{\text{futuro}}$ como **estándar mínimo de evaluación** en estudios de ML aplicado a series financieras en México, especialmente cuando el modelo emplee ventanas móviles o etiquetas con horizonte futuro. La validación cruzada estándar con particiones aleatorias sobre series de tiempo produce estimaciones de desempeño que no se replican en uso real y dificulta la comparabilidad entre estudios. La adopción de este estándar como criterio editorial en revistas de investigación financiera aplicada en México —como REMEF— elevaría la calidad inferencial del campo y reduciría el riesgo de publicar modelos con desempeño aparente que no se sostiene en condiciones operativas [5, 12, 145]. La implementación práctica de esta recomendación no requiere cambios editoriales onerosos: bastaría con añadir, en las instrucciones para autores de REMEF, un criterio de reporte que exija declarar explícitamente el protocolo de validación temporal empleado y justificar cualquier desviación del Walk-Forward con purga y embargo cuando el modelo utilice ventanas móviles o etiquetas con horizonte futuro. Esta adición transformaría la comparabilidad inferencial del campo de forma permanente.

Recomendación 2 — Institucional (CONSAR, CNBV e intermediarios supervisados). Al evaluar modelos algorítmicos de riesgo para uso en gestión de SIEFORES, gobernanza de modelos internos o supervisión del mercado de valores, distinguir explícitamente entre **tres propiedades** que los marcos regulatorios de IOSCO y el BIS reconocen como distintas y complementarias: (a) *poder predictivo*, medido como desempeño fuera de muestra bajo protocolo temporal riguroso; (b) *auditabilidad*, entendida como la disponibilidad del mecanismo exacto de decisión para inspección por supervisores y comités de riesgo; y (c) *calibración operativa*, que determina si las probabilidades del modelo son confiables para su uso en

dimensionamiento de posiciones. Evaluar los modelos únicamente por precisión predictiva omite las dos propiedades que determinan si el modelo puede adoptarse *responsablemente* en entornos regulados [90, 117, 118, 175, 189].

Recomendación 3 — Agenda pública (reguladores e investigadores de política financiera). Impulsar el desarrollo de un **índice exógeno de riesgo accionario por emisora para la BMV**, construido con variables de mercado futuras completamente separadas del espacio de predictores, que provea una referencia pública de acceso abierto análoga al rol que las calificadoras (HR Ratings, Fitch, S&P) desempeñan en el mercado de deuda. Este índice proveería a CONSAR y a la CNBV una referencia granular y auditable del perfil de riesgo individual de las emisoras, complementaria a las métricas de riesgo agregado (DCVaR) existentes. El protocolo Walk-Forward PRO, el CORE-10 y la arquitectura KAN establecidos en este trabajo son la infraestructura metodológica sobre la cual puede construirse ese índice con las garantías necesarias de integridad temporal, reproducibilidad y auditabilidad [40, 41, 108, 145, 166].

8.7 Líneas de Trabajo Futuro

1. **Búsqueda sistemática de hiperparámetros en KAN.** Explorar de forma exhaustiva número de nudos, grado spline, regularización y criterios de parada, buscando mejorar estabilidad OOS en la clase Medio y reducir el costo computacional [85, 144].

2. **Extensión a comparación con métodos de ensamble (XGBoost, RF).** Comparar KAN contra Random Forest y XGBoost en el eje de desempeño tabular puro, documentando la diferencia entre auditabilidad intrínseca (KAN) y explicabilidad post-hoc (SHAP sobre ensambles), con métricas de fidelidad de la explicación [151, 189].

3. **Pérdidas ordinales y funciones asimétricas.** Evaluar el marco CORAL y extensiones que penalicen errores no adyacentes (Bajo→Alto) más que adyacentes, incorporando costos coherentes con la estructura ordinal del problema [33, 91].

4. **Calibración probabilística por pliegue.** Incorporar temperature scaling u otros métodos sobre la ventana VAL de cada pliegue, con evaluación mediante diagramas de confiabilidad, ECE y Brier Score, habilitando uso del sistema con dimensionamiento de posiciones basado en probabilidades confiables [30, 90, 173].

5. **Ampliación del conjunto de variables.** Integrar variables fundamentales, macroeconómi-

cas y factores multifactoriales para enriquecer la señal bajo distintos regímenes de mercado [61, 86, 214].

6. **Extensión formal del módulo de portafolios.** Diseñar un estudio con reglas explícitas de asignación, restricciones de inversión, control de riesgo (minimización de CVaR/drawdown) y controles estadísticos de múltiples hipótesis. Un referente natural para esa extensión es el esquema de dos etapas de Rodríguez Rodríguez [186], que combina clasificación de acciones BMV con optimización multiobjetivo mediante NSGA-III: comparar esa arquitectura con KAN bajo un protocolo Walk-Forward PRO equivalente permitiría cuantificar el beneficio diferencial de la interpretabilidad estructural en el eje de construcción de portafolios [12, 23, 145].

7. **Desarrollo del índice de riesgo accionario exógeno para la BMV.** Construir una etiqueta de riesgo basada exclusivamente en variables de mercado futuras, separada del espacio de predictores, que sirva como referencia pública para la gestión del riesgo accionario —la extensión de mayor impacto institucional que este trabajo habilita [145, 158, 197].

8.8 Conclusión Final

Esta investigación demuestra que es factible incorporar una arquitectura basada en el Teorema de Kolmogorov-Arnold en un sistema de clasificación de riesgo accionario con evaluación temporalmente consistente y desempeño OOS competitivo. Los resultados confirman que KAN puede operar con desempeño comparable a *baselines* tradicionales y, de manera particularmente relevante, aporta **interpretabilidad estructural intrínseca** mediante funciones univariadas inspeccionables, característica valiosa en un contexto financiero donde la trazabilidad del modelo es un requisito de gobernanza [76, 144, 189].

La evidencia muestra que KAN ofrece ventajas consistentes frente a SVM en la clase de riesgo Medio, reforzando la idea de que las transiciones funcionales suaves —propiedad estructural de las B-splines— pueden modelar de forma más coherente las regiones frontera del riesgo que las fronteras rígidas de los modelos lineales. La verificación económica positiva (CAGR bruto OOS > 0, con costos explícitos) sugiere que la señal *aparenta* capturar contenido informativo real en el período evaluado, aunque la traducción a ventaja económica en portafolios requiere un programa de validación más amplio y controles estadísticos rigurosos [5, 12, 145].

Referencias bibliográficas

[1] Y. Amihud, «Illiquidity and stock returns: cross-section and time-series effects,» *Journal of Financial Markets*, vol. 5, nº 1, págs. 31-56, 2002. DOI: 10.1016/S1386-4181(01)00024-6.

[2] E. K. Ampomah, Z. Qin y G. Nyame, «Evaluation of Tree-Based Ensemble Machine Learning Models in Predicting Stock Price Direction of Movement,» *Information*, vol. 11, nº 6, pág. 332, 2020. DOI: 10.3390/info11060332.

[3] A. Ang, *Asset Management: A Systematic Approach to Factor Investing*. Oxford University Press, 2014, ISBN: 9780199959327.

[4] C. S. Ang, *Analysing Financial Data and Implementing Financial Models Using R*. Springer, 2015, ISBN: 9783319140756.

[5] H. R. Arian, D. Norouzi Mobarekeh y L. A. Seco, «Backtest overfitting in the machine learning era: A comparison of out-of-sample testing methods in a synthetic controlled environment,» *Knowledge-Based Systems*, vol. 305, pág. 112 477, 2024. DOI: 10.1016/j.knosys.2024.112477.

[6] V. I. Arnold, «On functions of three variables,» *Doklady Akademii Nauk SSSR*, vol. 114, nº 4, págs. 679-681, 1957.

[7] V. I. Arnold, «On the representation of continuous functions of three variables by superpositions of continuous functions of two variables,» *Matematicheskii Sbornik*, vol. 48(90), nº 1, págs. 3-74, 1959, English translation in: American Mathematical Society Translations, Series 2, 28, pp. 55–101, 1963.

[8] V. I. Arnold, *On the representability of functions of several variables as a superposition of functions of a smaller number of variables*. Moscow: Izdatel'stvo Akademii Nauk SSSR, 1962.

[9] R. Aroussi, *yfinance: Download market data from Yahoo! Finance's API*, https://github.com/ranaroussi/yfinance, Accessed: 2026-02-10, 2019.

[10] P. Artzner, F. Delbaen, J.-M. Eber y D. Heath, «Coherent Measures of Risk,» *Mathematical Finance*, vol. 9, nº 3, págs. 203-228, 1999. DOI: 10.1111/1467-9965.00068.

[11] arXiv. «A Practitioner's Guide to Kolmogorov–Arnold Networks.» Fecha de acceso: 10 febrero 2026. dirección: https://arxiv.org/html/2510.25781v1.

[12] D. H. Bailey, J. M. Borwein, M. López de Prado y Q. J. Zhu, «The Probability of Backtest Overfitting,» *Journal of Computational Finance*, vol. 20, nº 4, págs. 39-69, 2017. DOI: 10.21314/JCF.2017.323.

[13] D. H. Bailey y M. López de Prado, «The Deflated Sharpe Ratio: Correcting for Selection Bias, Backtest Overfitting and Non-Normality,» *Journal of Portfolio Management*, vol. 40, nº 5, págs. 94-107, 2014. DOI: 10.3905/jpm.2014.40.5.094.

[14] Banco de México, «Indicadores alpha y beta para medir riesgo en la Bolsa Mexicana de Valores (BMV),» Banco de México, inf. téc., 2024, Consultado en 2025. dirección: https://www.banxico.org.mx.

[15] Banco de México, «Reporte sobre Estabilidad Financiera: Primer semestre 2024,» Banco de México, inf. téc., 2024. dirección: https://www.banxico.org.mx/publicaciones-y-prensa/reportes-sobre-el-sistema-financiero/%7B9DCA595A-9844-00A1-CBFE-2B0CE096EE0B%7D.pdf.

[16] Banco de México, «Reporte sobre Estabilidad Financiera: Segundo semestre 2024,» Banco de México, inf. téc., 2024. dirección: https://www.banxico.org.mx/publicaciones-y-prensa/reportes-sobre-el-sistema-financiero/%7B0E8A0F80-B3CA-CB1C-5314-9781E0A8A70A%7D.pdf.

[17] Banco de México. «SIE API (REST): Documentación y ejemplos.» Consultado: febrero 2026. dirección: https://www.banxico.org.mx/SieAPIRest/.

[18] S. M. Bartram, J. Branke y M. Motahari, «Artificial Intelligence in Asset Management,» *Financial Analysts Journal*, vol. 76, nº 4, págs. 1-20, 2020. DOI: 10.2139/ssrn.3510343.

[19] A. Benavoli, G. Corani, J. Demšar y M. Zaffalon, «Time for a Change: A Tutorial for Comparing Multiple Classifiers through Bayesian Analysis,» *Journal of Machine Learning Research*, vol. 18, nº 77, págs. 1-36, 2017.

[20] Y. Bengio, A. Courville y P. Vincent, «Representation Learning: A Review and New Perspectives,» *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, nº 8, págs. 1798-1828, 2013. DOI: 10.1109/TPAMI.2013.50.

[21] C. M. Bishop, *Pattern Recognition and Machine Learning*. New York: Springer, 2006, ISBN: 978-0-387-31073-2.

[22] Blealtan. «efficient-kan: An efficient pure-PyTorch implementation of Kolmogorov-Arnold Network (KAN).» Fecha de acceso: 10 febrero 2026. dirección: https://github.com/Blealtan/efficient-kan.

[23] Z. Bodie, A. Kane y A. J. Marcus, *Investments*, 12th. McGraw-Hill Education, 2021, ISBN: 9781260013832.

[24] T. Bollerslev, «Generalized Autoregressive Conditional Heteroskedasticity,» *Journal of Econometrics*, vol. 31, nº 3, págs. 307-327, 1986. DOI: 10.1016/0304-4076(86)90063-1.

[25] T. Bollerslev, R. F. Engle y D. B. Nelson, «ARCH Models,» en *Handbook of Econometrics*, R. F. Engle y D. L. McFadden, eds., vol. 4, Elsevier, 1994, págs. 2959-3038. DOI: 10.1016/S1573-4412(05)80017-2.

[26] Bolsa Mexicana de Valores, *Índice S&P/BMV IPC: Metodología y composición*, https://www.bmv.com.mx/es/indices/principales/, Consultado en 2025, 2025.

[27] C. de Boor, *A Practical Guide to Splines* (Applied Mathematical Sciences), Revised. Springer, 2001, vol. 27, ISBN: 978-0-387-95366-3. DOI: 10.1007/978-0-387-95366-3.

[28] G. E. P. Box, G. M. Jenkins, G. C. Reinsel y G. M. Ljung, *Time Series Analysis: Forecasting and Control*, 5th. Wiley, 2015, ISBN: 9781118675021.

[29] J. Braun y M. Griebel, «On a constructive proof of Kolmogorov's superposition theorem,» *Constructive Approximation*, vol. 30, nº 3, págs. 653-675, 2009, Referencia clave para pruebas constructivas del KART.

[30] G. W. Brier, «Verification of Forecasts Expressed in Terms of Probability,» *Monthly Weather Review*, vol. 78, nº 1, págs. 1-3, 1950. DOI: 10.1175/1520-0493(1950)078<0001:VOFEIT>2.0.CO;2.

[31] O. Calin, *Deep Learning Architectures: A Mathematical Approach*. Springer, 2020, ISBN: 9783030347883.

[32] H. F. Calvo-Pardo, T. Mancini y J. Olmo, «Machine Learning and Model-Based Portfolio Allocation,» *Journal of Financial Econometrics*, vol. 18, nº 4, págs. 656-693, 2020. DOI: 10.1093/jjfinec/nbaa014.

[33] W. Cao, V. Mirjalili y S. Raschka, «Rank Consistent Ordinal Regression for Neural Networks with Application to Age Estimation,» *Pattern Recognition Letters*, vol. 140, págs. 325-331, 2020. DOI: 10.1016/j.patrec.2020.11.008. dirección: https://arxiv.org/abs/1901.07884.

[34] E. G. Carmona Vega, «Ajuste a la Calificación del Riesgo del Mercado de las Acciones más Volátiles que Conforman el Índice de Precios y Cotizaciones de la Bolsa Mexicana de Valores, con la Implementación de una Red Neuronal Artificial,» *Revista Mexicana de Economía y Finanzas*, vol. 8, nº 1, págs. 25-51, 2013, Segundo Lugar del Premio Nacional Bolsa Mexicana de Valores 2011 (categoría Investigación). DOI: 10.21919/remef.v8i1.40. dirección: https://remef.org.mx/index.php/remef/article/view/40.

[35] E. G. Carmona Vega y A. M. González Garzón, «Ajuste de la calificación del riesgo de mercado de las emisoras más activas que cotizan en la Bolsa Mexicana de Valores, con

la implementación de una Red Neuronal Artificial clasificadora,» Segundo Lugar del Premio Nacional Bolsa Mexicana de Valores 2011 (categoría Investigación). Artículo derivado publicado en REMEF, vol. 8, núm. 1, pp. 25–51, 2013, DOI: 10.21919/remef.v8i1.40, Tesis de Doctorado en Planeación Estratégica y Dirección de Tecnología, Universidad Popular Autónoma del Estado de Puebla, Puebla, México, 2011.

[36] E. G. Carmona Vega y A. M. González Garzón, «Ajuste de la Calificación del Riesgo de Mercado de las Empresas más Activas que Cotizan en la Bolsa Mexicana de Valores, con la Implementación de una Red Neuronal Artificial Clasificadora,» Tesis Doctoral. Segundo Lugar del Premio Nacional Bolsa Mexicana de Valores 2011 (categoría Investigación), Tesis doct., Universidad Popular Autónoma del Estado de Puebla, 2011.

[37] C. O. Chávez, A. Morales Castro y O. García Salgado, «Análisis de anomalías de mercado usando Machine Learning: un caso de la Bolsa Mexicana de Valores, BMV, 2000–2020,» *Un Espacio para la Ciencia*, vol. 2, nº 3, págs. 217-241, 2022. DOI: 10.5281/zenodo.7415936. dirección: https://dialnet.unirioja.es/servlet/articulo?codigo=9662528.

[38] F. Cheung. «Dissecting Kolmogorov-Arnold Network.» Fecha de acceso: 10 febrero 2026. dirección: https://feicheung2016.medium.com/dissecting-kolmogorov-arnold-network-f1bee719d949.

[39] S.-Y. Cho, S. Lee y H.-G. Kim, «Forecasting VIX Using Interpretable Kolmogorov-Arnold Networks,» *arXiv preprint arXiv:2502.00980*, 2025. dirección: https://arxiv.org/abs/2502.00980.

[40] Comisión Nacional del Sistema de Ahorro para el Retiro, *Se actualiza el régimen de inversión para potenciar las inversiones de las Afores en beneficio de los trabajadores*, Boletín de prensa No. 10/2024 (PDF), Publicado 2024-10-25; relacionado con modificaciones en DOF, oct. de 2024. dirección: https://www.gob.mx/cms/uploads/attachment/file/953159/BP__10_2024_Modificaciones_RI_2024.pdf.

[41] Comisión Nacional del Sistema de Ahorro para el Retiro, «Informe trimestral al H. Congreso de la Unión: octubre–diciembre 2024,» CONSAR, inf. téc., 2025, Alojado en el Sistema de Información Legislativa (Segob). dirección: https://sil.gobernacion.gob.mx/Archivos/Documentos/2025/03/asun_4847518_20250304_1741107229.pdf.

[42] Comisión Nacional del Sistema de Ahorro para el Retiro, «Informe trimestral al H. Congreso de la Unión sobre la situación del SAR: octubre–diciembre 2025,» CONSAR, inf. téc., 2026, Al cierre de diciembre de 2025, los activos netos del SAR alcanzaron

8.3 billones de pesos (23.8% del PIB). 69.7 millones de cuentas administradas por las Afores (77.8 millones incluyendo inhabilitadas y Fondo de Pensiones para el Bienestar). dirección: https://www.gob.mx/consar.

[43] Consejo de Estabilidad del Sistema Financiero (CESF), *Minutas y Comunicados*, https://www.gob.mx/shcp/acciones-y-programas/consejo-de-estabilidad-del-sistema-financiero-cesf, 2024.

[44] L. Corry, «Los 23 problemas de Hilbert y su trasfondo histórico,» *Boletín de la Asociación Matemática Venezolana*, vol. V, nº 2, págs. 119-125, 1998.

[45] C. Cortes y V. Vapnik, «Support-Vector Networks,» *Machine Learning*, vol. 20, nº 3, págs. 273-297, 1995. DOI: 10.1007/BF00994018.

[46] L. J. Cronbach y P. E. Meehl, «Construct Validity in Psychological Tests,» *Psychological Bulletin*, vol. 52, nº 4, págs. 281-302, 1955. DOI: 10.1037/h0040957.

[47] G. Cybenko, «Approximation by Superpositions of a Sigmoidal Function,» *Mathematics of Control, Signals and Systems*, vol. 2, nº 4, págs. 303-314, 1989. DOI: 10.1007/BF02551274.

[48] J. Danielsson, *Financial Risk Forecasting: The Theory and Practice of Forecasting Market Risk, Credit Risk and Operational Risk*. Wiley, 2011, ISBN: 9780470669433.

[49] S. Das y A. Das, «Fusion with Sentiment Scores for Market Research,» en *19th International Conference on Information Fusion (FUSION)*, IEEE, 2016, págs. 1003-1010. DOI: 10.1109/FUSION.2016.7528052.

[50] R. De Jesús-Gutiérrez, «El uso de la volatilidad implícita en el modelado de la varianza condicional puede mejorar la predicción de la volatilidad y la estimación del VaR y CVaR,» *Economía: Teoría y Práctica*, nº 58, págs. 173-198, 2023. DOI: 10.24275/etypuam/ne/582023/jesus. dirección: https://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S0188-33802023000100173.

[51] F. M. Dekking, C. Kraaikamp, H. Lopuhaä y L. Meester, *A Modern Introduction to Probability and Statistics*. Springer, 2005, ISBN: 9781852338961.

[52] J. Demšar, «Statistical Comparisons of Classifiers over Multiple Data Sets,» *Journal of Machine Learning Research*, vol. 7, págs. 1-30, 2006.

[53] P. Diaconis y M. Shahshahani, «On linear functions of linear combinations,» *Journal of Mathematical Analysis and Applications*, vol. 100, nº 2, págs. 388-398, 1984.

[54] T. G. Dietterich, «Approximate Statistical Tests for Comparing Supervised Classification Learning Algorithms,» *Neural Computation*, vol. 10, nº 7, págs. 1895-1923, 1998. DOI: 10.1162/089976698300017197.

[55] F. Doshi-Velez y B. Kim, «Towards a Rigorous Science of Interpretable Machine Learning,» *arXiv preprint arXiv:1702.08608*, 2017. dirección: https://arxiv.org/abs/1702.08608.

[56] B. Efron y R. J. Tibshirani, *An Introduction to the Bootstrap*. CRC Press, 1994.

[57] E. J. Elton, M. J. Gruber, S. J. Brown y W. N. Goetzmann, *Modern Portfolio Theory and Investment Analysis*, 9th. Wiley, 2014, ISBN: 9781118469941.

[58] R. F. Engle, «Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation,» *Econometrica*, vol. 50, nº 4, págs. 987-1007, 1982. DOI: 10.2307/1912773.

[59] D. Enke y S. Thawornwong, «The Use of Data Mining and Neural Networks for Forecasting Stock Market Returns,» *Expert Systems with Applications*, vol. 29, nº 4, págs. 927-940, 2005. DOI: 10.1016/j.eswa.2005.06.024.

[60] I. A. Fahad. «Understanding Walk Forward Validation in Time Series Analysis: A Practical Guide.» Fecha de acceso: 10 febrero 2026. dirección: https://medium.com/@ahmedfahad04/understanding-walk-forward-validation-in-time-series-analysis-a-practical-guide-ea3814015abf.

[61] E. F. Fama y K. R. French, «The Cross-Section of Expected Stock Returns,» *The Journal of Finance*, vol. 47, nº 2, págs. 427-465, 1992. DOI: 10.1111/j.1540-6261.1992.tb04398.x.

[62] E. F. Fama y K. R. French, «The Capital Asset Pricing Model: Theory and Evidence,» *Journal of Economic Perspectives*, vol. 18, nº 3, págs. 25-46, 2004. DOI: 10.1257/0895330042162430.

[63] T. Fawcett, «An Introduction to ROC Analysis,» *Pattern Recognition Letters*, vol. 27, nº 8, págs. 861-874, 2006. DOI: 10.1016/j.patrec.2005.10.010.

[64] R. Fernández, B. Palma Guizar y C. Rho, «A Sentiment-based Risk Indicator for the Mexican Financial Sector,» Banco de México, Documento de Investigación N° 2021-04, 2021. dirección: https://www.banxico.org.mx/publicaciones-y-prensa/documentos-de-investigacion-del-banco-de-mexico/%7B1A22D656-CAD6-B11A-6471-53160723B254%7D.pdf.

[65] Financial Stability Board, «The Financial Stability Implications of Artificial Intelligence,» Financial Stability Board, inf. téc., nov. de 2024. dirección: https://www.fsb.org/2024/11/the-financial-stability-implications-of-artificial-intelligence/.

[66] T. Fischer y C. Krauss, «Deep Learning with Long Short-Term Memory Networks for Financial Market Predictions,» *European Journal of Operational Research*, vol. 270, nº 2, págs. 654-669, 2018. DOI: 10.1016/j.ejor.2017.11.054.

[67] Fitch Ratings, «Metodología de Calificación de Empresas,» Fitch Ratings, inf. téc., 2020.

[68] Fitch Ratings, «Corporates Recovery Ratings and Instrument Ratings Criteria,» Fitch Ratings, inf. téc., 2021.

[69] Fitch Ratings, «Corporate Rating Criteria,» Fitch Ratings, inf. téc., 2022.

[70] Fitch Ratings, «Metodología de Calificaciones en Escala Nacional,» Fitch Ratings, inf. téc., 2022.

[71] Fitch Ratings, «Government-Related Entities Rating Criteria,» Fitch Ratings, inf. téc., 2024.

[72] Fitch Ratings, «Metodología de Calificación de Seguros,» Fitch Ratings, inf. téc., 2024.

[73] Fitch Ratings, «Sector Navigators — Addendum to the Corporate Rating Criteria,» Fitch Ratings, inf. téc., 2024.

[74] C. Fjellström, «Long Short-Term Memory Neural Network for Financial Time Series,» en *2022 IEEE International Conference on Big Data*, 2022, págs. 3496-3504. DOI: 10.1109/BigData55660.2022.10020784.

[75] E. Frank y M. Hall, «A Simple Approach to Ordinal Classification,» en *Proceedings of the 12th European Conference on Machine Learning (ECML 2001)*, ép. Lecture Notes in Computer Science, vol. 2167, Springer, 2001, págs. 145-156. DOI: 10.1007/3-540-44795-4_13.

[76] S. Fritz-Morgenthal, B. Hein y J. Papenbrock, «Financial Risk Management and Explainable, Trustworthy, Responsible AI,» *Frontiers in Artificial Intelligence*, vol. 5, pág. 779 799, 2022. DOI: 10.3389/frai.2022.779799.

[77] R. Gallardo Del Ángel, «Financial Time Series Forecasting Using Artificial Neural Networks,» *Revista Mexicana de Economía y Finanzas*, vol. 15, nº 1, 2020. DOI: 10.21919/remef.v15i1.376. dirección: https://www.remef.org.mx/index.php/remef/article/view/376.

[78] O. García Salgado y A. Morales Castro, «Empresas exitosas y no exitosas que cotizan en la BMV del Sector Comercial: Una clasificación con Análisis Discriminante Múltiple, Modelos Logit y Redes Neuronales Artificiales,» *Estocástica: Finanzas y Riesgo*, vol. 4, nº 1, págs. 33-62, 2014. dirección: https://ideas.repec.org/a/sfr/efruam/v4y2014i1p33-62.html.

[79] R. Genet y H. Inzirillo, *TKAN: Temporal Kolmogorov-Arnold Networks*, arXiv, 2024. DOI: 10.48550/arXiv.2405.07344. dirección: https://arxiv.org/abs/2405.07344.

[80] A. Géron, *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*, 2nd. O'Reilly Media, 2019, ISBN: 9781492032649.

[81] H. Gholampoor y M. Asadi, «Risk Analysis of Bankruptcy in the U.S. Healthcare Industries Based on Financial Ratios: A Machine Learning Analysis,» *Journal of Theoretical and Applied Electronic Commerce Research*, vol. 19, nº 2, pág. 66, 2024. DOI: 10.3390/jtaer19020066.

[82] Gobierno de México, Secretaría de Economía, «Indicadores financieros e inversión extranjera directa 2024,» Gobierno de México, inf. téc., 2024.

[83] P. Golbayani, I. Florescu y R. Chatterjee, «A comparative study of forecasting Corporate Credit Ratings using Neural Networks, Support Vector Machines, and Decision Trees,» *North American Journal of Economics and Finance*, vol. 54, pág. 101 251, 2020. DOI: 10.1016/j.najef.2020.101251.

[84] E. L. Gómez-Ramos, F. Venegas-Martínez y H. Allier-Campuzano, «Análisis comparativo entre modelos GARCH y redes neuronales en el pronóstico de los índices bursátiles IPC y Dow Jones,» *eseconomía*, vol. VI, nº 32, págs. 3-22, 2011, Escuela Superior de Economía, Instituto Politécnico Nacional. dirección: https://ideas.repec.org/a/ipn/esecon/vviy2011i32p3-22.html.

[85] I. Goodfellow, Y. Bengio y A. Courville, *Deep Learning*. MIT Press, 2016, ISBN: 9780262035613. dirección: https://www.deeplearningbook.org/.

[86] S. Gu, B. Kelly y D. Xiu, «Empirical Asset Pricing via Machine Learning,» *The Review of Financial Studies*, vol. 33, nº 5, págs. 2223-2273, 2020. DOI: 10.1093/rfs/hhaa009.

[87] Y. Gu, T. Shibukawa, Y. Kondo, S. Nagao y S. Kamijo, «Prediction of Stock Performance Using Deep Neural Networks,» *Applied Sciences*, vol. 10, nº 22, pág. 8142, 2020. DOI: 10.3390/app10228142.

[88] G. C. Guerrero Camargo, «Optimización de portafolios de inversión en la bolsa mexicana de valores utilizando redes neuronales artificiales,» Tesis de Maestría, Tesis de mtría., Universidad Autónoma de Querétaro, 2022. dirección: https://ri-ng.uaq.mx/handle/123456789/8672.

[89] R. Guidotti, A. Monreale, S. Ruggieri, F. Turini, F. Giannotti y D. Pedreschi, «A Survey of Methods for Explaining Black Box Models,» *ACM Computing Surveys*, vol. 51, nº 5, pág. 93, 2018. DOI: 10.1145/3236009.

[90] C. Guo, G. Pleiss, Y. Sun y K. Q. Weinberger, «On Calibration of Modern Neural Networks,» en *Proceedings of the 34th International Conference on Machine Learning*, ép. PMLR, vol. 70, 2017, págs. 1321-1330. dirección: https://proceedings.mlr.press/v70/guo17a.html.

[91] P. A. Gutiérrez, M. Pérez-Ortiz, J. Sánchez-Monedero, F. Fernández-Navarro y C. Hervás-Martínez, «Ordinal Regression Methods: Survey and Experimental Study,» *IEEE Transactions on Knowledge and Data Engineering*, vol. 28, nº 1, págs. 127-146, 2016. DOI: 10.1109/TKDE.2015.2457911.

[92] X. Han, X. Zhang, Y. Wu, Z. Zhang y Z. Wu, «Are KANs Effective for Multivariate Time Series Forecasting?» En *Proceedings of the 34th ACM International Conference on Information and Knowledge Management (CIKM 2025)*, 2025. DOI: 10.1145/3746252.3760836. dirección: https://arxiv.org/abs/2408.11306.

[93] D. J. Hand y R. J. Till, «A Simple Generalisation of the Area under the ROC Curve for Multiple Class Classification Problems,» *Machine Learning*, vol. 45, nº 2, págs. 171-186, 2001. DOI: 10.1023/A:1010920819831.

[94] C. R. Harvey, «Predictable Risk and Returns in Emerging Markets,» *The Review of Financial Studies*, vol. 8, nº 3, págs. 773-816, 1995. DOI: 10.1093/rfs/8.3.773.

[95] C. R. Harvey, Y. Liu y H. Zhu, «. . . and the Cross-Section of Expected Returns,» *The Review of Financial Studies*, vol. 29, nº 1, págs. 5-68, 2016. DOI: 10.1093/rfs/hhv059.

[96] A. T. Haryono, R. Sarno, R. N. E. Anggraini y K. R. Sungkono, «Permuted Temporal Kolmogorov-Arnold Networks for Stock Price Forecasting Using Generative Aspect-Based Sentiment Analysis,» *IEEE Access*, vol. 12, págs. 178 672-178 689, 2024. DOI: 10.1109/ACCESS.2024.3506658.

[97] T. Hastie, R. Tibshirani y J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (Springer Series in Statistics), 2nd. Springer, 2009, ISBN: 978-0-387-84858-7. DOI: 10.1007/978-0-387-84858-7.

[98] R. Hecht-Nielsen, «Kolmogorov's Mapping Neural Network Existence Theorem,» en *Proceedings of the First IEEE International Conference on Neural Networks*, vol. 3, San Diego, CA: IEEE, 1987, págs. 11-14.

[99] A. Hernández García, «Predicción de la volatilidad del mercado accionario mexicano con redes neuronales recurrentes,» Tesis de Maestría, Tesis de mtría., Universidad Anáhuac, 2023.

[100] D. Hilbert, «Mathematische Probleme,» *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse*, págs. 253-297, 1900, Vortrag, gehalten auf dem internationalen Mathematiker-Kongreß zu Paris 1900.

[101] D. Hilbert, «Mathematical problems,» *Bulletin of the American Mathematical Society*, vol. 8, nº 10, págs. 437-479, 1902, English translation of Hilbert (1900).

[102] Y. Hilpisch, *Python for Finance: Mastering Data-Driven Finance*. O'Reilly Media, 2018, ISBN: 9781492024330.

[103] S. Hochreiter y J. Schmidhuber, «Long Short-Term Memory,» *Neural Computation*, vol. 9, nº 8, págs. 1735-1780, 1997. DOI: 10.1162/neco.1997.9.8.1735.

[104] K. Hornik, «Approximation Capabilities of Multilayer Feedforward Networks,» *Neural Networks*, vol. 4, nº 2, págs. 251-257, 1991. DOI: 10.1016/0893-6080(91)90009-T.

[105] K. Hornik, M. Stinchcombe y H. White, «Multilayer Feedforward Networks are Universal Approximators,» *Neural Networks*, vol. 2, nº 5, págs. 359-366, 1989. DOI: 10.1016/0893-6080(89)90020-8.

[106] Y. Hou, T. Ji, D. Zhang y A. Stefanidis, *Kolmogorov-Arnold Networks: A Critical Assessment of Claims, Performance, and Practical Viability*, arXiv, Last revised 2025-09-14 (v8), 2024. DOI: 10.48550/arXiv.2407.11075. dirección: https://arxiv.org/abs/2407.11075.

[107] HR Ratings de México, S.A. de C.V., «Metodología para Calificar Sociedades de Inversión,» HR Ratings, inf. téc., 2008.

[108] HR Ratings de México, S.A. de C.V., «Metodología de Calificación de Empresas,» HR Ratings, inf. téc., 2021.

[109] HR Ratings de México, S.A. de C.V., «Metodología de Calificación para Infraestructura y Financiamiento de Proyectos,» HR Ratings, inf. téc., 2022.

[110] J. C. Hull, *Options, Futures, and Other Derivatives*, 10th. Pearson Education, 2018, ISBN: 9780134472089.

[111] J. C. Hull, *Risk Management and Financial Institutions*, 5th. Wiley, 2018, ISBN: 9781119448112.

[112] INEGI, «Sistema de Cuentas Nacionales de México. Fuentes y metodologías. Año base 2013,» INEGI, inf. téc., 2018. dirección: https://www.inegi.org.mx/contenidos/programas/pibmed/2018/doc/met_cab2018.pdf.

[113] INEGI, *Banco de Información Económica (BIE)*, https://www.inegi.org.mx/sistemas/bie/, 2025.

[114] INEGI, *Indicadores de Confianza Empresarial (ICE). Boletines Mensuales*, https://www.inegi.org.mx/temas/opinion/, 2025.

[115] INEGI, *Producto Interno Bruto por Sector Institucional. Informes Trimestrales*, https://www.inegi.org.mx/temas/sit/, 2025.

[116] International Monetary Fund, *Global Financial Stability Report*. Washington, DC: International Monetary Fund, 2024. dirección: https://www.imf.org/en/Publications/GFSR.

[117] International Organization of Securities Commissions, «The Use of Artificial Intelligence and Machine Learning by Market Intermediaries and Asset Managers,» IOSCO, Madrid, España, inf. téc. FR06/2021, 2021. dirección: https://www.iosco.org/library/pubdocs/pdf/IOSCOPD684.pdf.

[118] International Organization of Securities Commissions, «Artificial Intelligence in Capital Markets: Use Cases, Risks, and Challenges,» IOSCO, Madrid, España, inf. téc. CR/01/2025, 2025. dirección: https://www.iosco.org/library/pubdocs/pdf/IOSCOPD788.pdf.

[119] J.P. Morgan Private Bank, «Incertidumbre y volatilidad: Actualización del mercado mexicano,» J.P. Morgan Private Bank, inf. téc., 2024. dirección: https://privatebank.jpmorgan.com/latam/es/insights/markets-and-investing/el-pulso-de-america-latina/incertidumbre-y-volatilidad-actualizacion-del-mercado-mexicano.

[120] G. James, D. Witten, T. Hastie y R. Tibshirani, *An Introduction to Statistical Learning*, 2nd. Springer, 2021, ISBN: 9781071614174.

[121] P. Jorion, *Value at Risk: The New Benchmark for Managing Financial Risk*, 3rd. McGraw-Hill, 2007, ISBN: 9780071464956.

[122] W. Jurado Pedroza, «Around-the-Clock USD/MXN Volatility: Macroeconomic Announcement Spillovers and FX Market Intervention Mechanisms,» Banco de México, Documento de Investigación / Working Paper 2021-05, jun. de 2021. dirección: https://www.banxico.org.mx/publicaciones-y-prensa/documentos-de-investigacion-del-banco-de-mexico/%7BFD46369C-C33A-04A3-445A-6ACD66D19D71%7D.pdf.

[123] Kaggle. «Time Series Using Walk-Forward Validation.» Fecha de acceso: 10 febrero 2026. dirección: https://www.kaggle.com/code/justozner/time-series-using-walk-forward-validation.

[124] G. L. Kaminsky y C. M. Reinhart, «The Twin Crises: The Causes of Banking and Balance-of-Payments Problems,» *American Economic Review*, vol. 89, nº 3, págs. 473-500, 1999. DOI: 10.1257/aer.89.3.473.

[125] H. Kharabe, S. Jain, S. Bagdare, A. Jain, R. Malpani y P. Churi, «DeepPredict: A Deep Learning Based Investment Platform for Stocks and Mutual Funds,» *Journal of Big Data*, vol. 10, nº 1, págs. 1-23, 2023. DOI: 10.1186/s40537-023-00718-1.

[126] D. Khavinson, «A note on the thirteenth Hilbert problem,» *Proceedings of the American Mathematical Society*, vol. 124, nº 7, págs. 2275-2277, 1996.

[127] D. P. Kingma y J. Ba, «Adam: A Method for Stochastic Optimization,» *arXiv preprint arXiv:1412.6980*, 2014. dirección: https://arxiv.org/abs/1412.6980.

[128] W. Knottenbelt et al., «CoxKAN: Kolmogorov-Arnold networks for interpretable, high-performance survival analysis,» *Bioinformatics*, vol. 41, nº 8, btaf413, 2025. DOI: 10.1093/bioinformatics/btaf413. dirección: https://academic.oup.com/bioinformatics/article/41/8/btaf413/8209483.

[129] A. N. Kolmogorov, «On the representation of continuous functions of several variables by superpositions of continuous functions of a smaller number of variables,» *Doklady Akademii Nauk SSSR*, vol. 108, págs. 179-182, 1956, English translation in: American Mathematical Society Translations, Series 2, 17, pp. 369–373, 1961.

[130] A. N. Kolmogorov, «On the representation of continuous functions of many variables by superposition of continuous functions of one variable and addition,» *Doklady Akademii Nauk SSSR*, vol. 114, nº 5, págs. 953-956, 1957, English translation in: American Mathematical Society Translations, Series 2, 28, pp. 55–59, 1963.

[131] A. Kundu, A. Sarkar y A. Sadhu, «KANQAS: Kolmogorov-Arnold Network for Quantum Architecture Search,» *EPJ Quantum Technology*, vol. 11, nº 76, 2024. DOI: 10.1140/epjqt/s40507-024-00289-z. dirección: https://epjquantumtechnology.springeropen.com/articles/10.1140/epjqt/s40507-024-00289-z.

[132] V. Kůrková, «Kolmogorov's Theorem and Multilayer Neural Networks,» *Neural Networks*, vol. 5, nº 3, págs. 501-506, 1992. DOI: 10.1016/0893-6080(92)90012-8.

[133] V. M. Landassuri-Moreno, C. E. Bustillo-Hernández, J. J. Carbajal-Hernández y L. P. Sánchez Fernández, «Predicción del IPC mexicano combinando modelos econométricos e inteligencia artificial,» *Revista Mexicana de Economía y Finanzas*, vol. 13, nº 4, págs. 603-629, 2018. dirección: https://www.researchgate.net/publication/328000236.

[134] T.-T.-H. Le, Y. Hwang, H. Kang y H. Kim, «Robust Credit Card Fraud Detection Based on Efficient Kolmogorov-Arnold Network Models,» *IEEE Access*, vol. 12, págs. 157 006-157 020, 2024. DOI: 10.1109/ACCESS.2024.3485200.

[135] Y. LeCun, Y. Bengio y G. Hinton, «Deep Learning,» *Nature*, vol. 521, nº 7553, págs. 436-444, 2015. DOI: 10.1038/nature14539.

[136] Y. LeCun, J. S. Denker y S. A. Solla, «Optimal Brain Damage,» en *Advances in Neural Information Processing Systems*, D. S. Touretzky, ed., vol. 2, Denver, CO: Morgan Kaufmann, 1990, págs. 598-605.

[137] M. Leo, S. Sharma y K. Maddulety, «Machine learning in banking risk management: A literature review,» *Risks*, vol. 7, nº 1, pág. 29, 2019. DOI: 10.3390/risks7010029.

[138] Y. Li y Z. Tan, «Stock Portfolio Selection with Deep RankNet,» *IEEE Transactions on Neural Networks and Learning Systems*, vol. 34, nº 5, págs. 2478-2490, 2023. DOI: 10.1109/TNNLS.2022.3169011.

[139] Z. Li, «Kolmogorov-Arnold Networks are Radial Basis Function Networks,» *arXiv preprint arXiv:2405.06721*, 2024. dirección: https://arxiv.org/abs/2405.06721.

[140] J. Lintner, «The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets,» *The Review of Economics and Statistics*, vol. 47, nº 1, págs. 13-37, 1965. DOI: 10.2307/1924119.

[141] Z. C. Lipton, «The Mythos of Model Interpretability,» *Queue*, vol. 16, nº 3, págs. 31-57, 2018. DOI: 10.1145/3236386.3241340.

[142] C. Z. Liu, Y. Zhang, L. Qin e Y. Liu, «Kolmogorov–Arnold Finance-Informed Neural Network in Option Pricing,» *Applied Sciences*, vol. 14, nº 24, pág. 11 618, 2024. DOI: 10.3390/app142411618. dirección: https://www.mdpi.com/2076-3417/14/24/11618.

[143] Z. Liu, P. Ma, Y. Wang, W. Matusik y M. Tegmark, *KAN 2.0: Kolmogorov-Arnold Networks Meet Science*, arXiv, 2024. DOI: 10.48550/arXiv.2408.10205. dirección: https://arxiv.org/abs/2408.10205.

[144] Z. Liu et al., «KAN: Kolmogorov–Arnold Networks,» en *International Conference on Learning Representations (ICLR)*, Oral, 2025. dirección: https://openreview.net/forum?id=Ozo7qJ5vZi.

[145] M. López de Prado, *Advances in Financial Machine Learning*. Hoboken, NJ: Wiley, 2018, ISBN: 978-1-119-48208-6.

[146] M. López Pérez, «Análisis de sentimientos en redes sociales para la predicción de rendimientos en el mercado bursátil mexicano,» Tesis de Maestría, Tesis de mtría., Instituto Tecnológico Autónomo de México, 2021.

[147] G. G. Lorentz, «Metric entropy, widths, and superpositions of functions,» *American Mathematical Monthly*, vol. 69, nº 6, págs. 469-485, 1962.

[148] G. G. Lorentz, *Approximation of Functions*. New York: Holt, Rinehart y Winston, 1966.

[149] I. Loshchilov y F. Hutter, «Decoupled Weight Decay Regularization,» en *International Conference on Learning Representations (ICLR)*, 2019. dirección: https://openreview.net/forum?id=Bkg6RiCqY7.

[150] Y. Lu y F. Zhan, «Kolmogorov Arnold Networks in Fraud Detection: Bridging the Gap Between Theory and Practice,» *arXiv preprint arXiv:2408.10263*, 2024. dirección: https://arxiv.org/abs/2408.10263.

[151] S. M. Lundberg y S.-I. Lee, «A Unified Approach to Interpreting Model Predictions,» en *Advances in Neural Information Processing Systems*, vol. 30, 2017, págs. 4765-4774. dirección: https://proceedings.neurips.cc/paper/2017/file/8a20a8621978632d76c43dfd28b67
Paper.pdf.

[152] V. Maiorov y A. Pinkus, «Approximation with Superpositions of a Sigmoidal Function,» *Advances in Computational Mathematics*, vol. 25, nº 1–3, págs. 337-363, 2006. DOI: 10.1007/s10444-005-9002-3.

[153] R. Mamgai, H. Jain, Y. Aithani y A. Sharma, «Stock Prediction & Recommendation System Using KNN and Linear Regression,» *International Journal of Computer Applications*, vol. 184, nº 45, págs. 1-5, 2023. DOI: 10.5120/ijca2023922645.

[154] H. Markowitz, «Portfolio Selection,» *The Journal of Finance*, vol. 7, nº 1, págs. 77-91, 1952. DOI: 10.2307/2975974.

[155] J. Márquez Diez-Canedo, «Suficiencia de Capital y Riesgo de Crédito en Carteras de Préstamos Bancarios,» Banco de México, Documento de Investigación No. 2002-04, 2002. dirección: https://www.banxico.org.mx/publicaciones-y-prensa/documentos-de-investigacion-del-banco-de-mexico/%7BD7247E60-4FCD-F8AD-A6C1-C8A645812C80%7D.pdf.

[156] W. S. McCulloch y W. Pitts, «A logical calculus of the ideas immanent in nervous activity,» *Bulletin of Mathematical Biophysics*, vol. 5, nº 4, págs. 115-133, 1943. DOI: 10.1007/BF02478259.

[157] W. McKinney, *Python for Data Analysis*. O'Reilly Media, 2012, ISBN: 9781449319793.

[158] A. J. McNeil, R. Frey y P. Embrechts, *Quantitative Risk Management: Concepts, Techniques and Tools*, Revised. Princeton, NJ: Princeton University Press, 2015, ISBN: 9780691166278.

[159] P. D. McNelis, *Neural Networks in Finance: Gaining Predictive Edge in the Market*. Academic Press, 2005, ISBN: 9780124859678.

[160] E. Melchor Pérez, M. E. Ramírez Guzmán, A. Hernández Jiménez y A. Santiago Alvarado, «Predicción del riesgo crediticio a microfinanciera usando aprendizaje computacional,» *Revista Mexicana de Economía y Finanzas*, vol. 19, nº 4, 2024. DOI: 10.21919/remef.v19i4.868. dirección: https://www.remef.org.mx/index.php/remef/article/view/868.

[161] V. C. Melveetil y S. Mohanty, «A Review of Explainable Artificial Intelligence (XAI): Towards Finance,» *Nanotechnology Perceptions*, vol. 20, nº S16, págs. 411-435, 2024. dirección: https://nano-ntp.com/index.php/nano/article/view/3866.

[162] A. K. Menon, S. Jayasumana, A. S. Rawat, H. Jain, A. Veit y S. Kumar, «Long-tail Learning via Logit Adjustment,» en *International Conference on Learning Representations (ICLR)*, 2021. dirección: https://openreview.net/forum?id=37nvvqkCo5.

[163] E. Mienye, N. Jere, G. Obaido, I. D. Mienye y K. Aruleba, «Deep Learning in Finance: A Survey of Applications and Techniques,» *AI*, vol. 5, nº 4, pág. 101, 2024. DOI: 10.3390/ai5040101.

[164] M. Minsky y S. Papert, *Perceptrons: An Introduction to Computational Geometry*. Cambridge, MA: MIT Press, 1969, ISBN: 978-0-262-63022-1.

[165] C. Molnar, *Interpretable Machine Learning: A Guide for Making Black Box Models Explainable*. Lulu.com, 2020. dirección: https://christophm.github.io/interpretable-ml-book/.

[166] Moody's Investors Service, «The Evolving Meaning of Moody's Bond Ratings,» Moody's Investors Service, inf. téc., 2000.

[167] Moody's Investors Service, «Finance Companies Rating Methodology,» Moody's Investors Service, inf. téc., 2023.

[168] Moody's Investors Service, «Insurance Brokers and Service Companies Rating Methodology,» Moody's Investors Service, inf. téc., 2024, Rating methodology report, Moody's Investors Service, Inc.

[169] A. Morales Castro, E. Ramírez Reyes y G. Rodríguez Albor, «Pronóstico de ventas de las empresas del sector alimentos: una aplicación de redes neuronales,» *Semestre Económico*, vol. 22, nº 52, págs. 161-177, 2019. DOI: 10.22395/seec.v22n52a7.

[170] S. A. Morris, «Hilbert's 13th Problem and the Superposition of Continuous Functions,» *arXiv preprint math/0503211*, 2005.

[171] J. Mossin, «Equilibrium in a Capital Asset Market,» *Econometrica*, vol. 34, nº 4, págs. 768-783, 1966. DOI: 10.2307/1910098.

[172] M. Murillo Tsijli, «El Teorema de Aproximación Universal para Redes Neuronales,» *Matemática, Educación e Internet*, vol. 22, nº 1, 2021. dirección: https://revistas.tec.ac.cr/index.php/matematica/article/view/2282.

[173] A. Niculescu-Mizil y R. Caruana, «Predicting Good Probabilities with Supervised Learning,» en *Proceedings of the 22nd International Conference on Machine Learning (ICML)*, 2005, págs. 625-632. DOI: 10.1145/1102351.1102430.

[174] A. M. Ozbayoglu, M. U. Gudelek y O. B. Sezer, «Deep Learning for Financial Applications: A Survey,» *Applied Soft Computing*, vol. 93, pág. 106 384, 2020. DOI: 10.1016/j.asoc.2020.106384.

[175] F. Pérez-Cruz, J. Prenio, F. Restoy y J. Yong, «Managing explanations: how regulators can address AI explainability,» Bank for International Settlements, Financial Stability Institute, FSI Occasional Papers 24, sep. de 2025. dirección: https://www.bis.org/fsi/fsipapers24.htm.

[176] J. C. Platt, «Probabilistic Outputs for Support Vector Machines and Comparisons to Regularized Likelihood Methods,» en *Advances in Large Margin Classifiers*, A. J. Smola, P. L. Bartlett, B. Schölkopf y D. Schuurmans, eds., MIT Press, 1999, págs. 61-74.

[177] T. Poggio y F. Girosi, «A Theory of Networks for Approximation and Learning,» Artificial Intelligence Laboratory, Massachusetts Institute of Technology, inf. téc. A.I. Memo No. 1140, 1989.

[178] A. Polar y M. Poluektov, «A Deep Machine Learning Algorithm for Construction of the Kolmogorov–Arnold Representation,» *Engineering Applications of Artificial Intelligence*, vol. 99, pág. 104 137, 2021. DOI: 10.1016/j.engappai.2020.104137.

[179] D. M. W. Powers, «Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness & Correlation,» *Journal of Machine Learning Technologies*, vol. 2, nº 1, págs. 37-63, 2011.

[180] L. Prechelt, «Early Stopping — But When?» En *Neural Networks: Tricks of the Trade*, Springer, 1998, págs. 55-69. DOI: 10.1007/3-540-49430-8_3.

[181] Y. Qu, P. Quan, M. Lei e Y. Shi, «Review of Bankruptcy Prediction Using Machine Learning and Deep Learning Techniques,» en *Procedia Computer Science*, vol. 162, 2019, págs. 895-899. DOI: 10.1016/j.procs.2019.12.065.

[182] A. Rafiuddin, J. C. Tellez Gaytan, G. Ahmed y M. Alshurideh, «Machine Learning Techniques for Stock Market Predictions: A Case of Mexican Stocks,» en *The Effect of Information Technology on Business and Marketing Intelligence Systems*, ép. Studies in Computational Intelligence, vol. 1056, Springer, 2023, págs. 1833-1843. DOI: 10.1007/978-3-031-12382-5_100.

[183] M. Reyna Miranda, R. Massa Roldán y V. Gómez Salcido, «Neuro-wavelet Model for Price Prediction in High-Frequency Data in the Mexican Stock Market,» *Revista Mexicana de Economía y Finanzas*, vol. 17, nº 1, 2022. DOI: 10.21919/remef.v17i1.570. dirección: https://remef.org.mx/index.php/remef/article/view/570.

[184] M. T. Ribeiro, S. Singh y C. Guestrin, «"Why Should I Trust You?": Explaining the Predictions of Any Classifier,» en *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016, págs. 1135-1144. DOI: 10.1145/2939672.2939778.

[185] R. T. Rockafellar y S. Uryasev, «Optimization of Conditional Value-at-Risk,» *Journal of Risk*, vol. 2, págs. 21-42, 2000.

[186] S. G. Rodríguez Rodríguez, «Portfolio Selection Using Artificial Intelligence Techniques in the Mexican Stock Exchange,» Tesis de Maestría en Ciencias de la Computación. Campus Monterrey, Tesis de mtría., Instituto Tecnológico y de Estudios Superiores de Monterrey, 2022. dirección: https://hdl.handle.net/11285/650804.

[187] F. Rosenblatt, «The perceptron: a probabilistic model for information storage and organization in the brain,» *Psychological Review*, vol. 65, nº 6, págs. 386-408, 1958. DOI: 10.1037/h0042519.

[188] S. A. Ross, R. W. Westerfield y J. Jaffe, *Corporate Finance*, 12th. McGraw-Hill Education, 2019, ISBN: 9781260565553.

[189] C. Rudin, «Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead,» *Nature Machine Intelligence*, vol. 1, nº 5, págs. 206-215, 2019. DOI: 10.1038/s42256-019-0048-x.

[190] S&P Global Ratings, «Metodología de Calificación de Empresas,» S&P Global Ratings, inf. téc., 2019.

[191] Á. Samaniego Alcántar, «Portfolio Optimization with Long-Short Term Memory Deep Learning (LSTM),» *Revista Mexicana de Economía y Finanzas*, vol. 20, nº 2, 2025. DOI: 10.21919/remef.v20i2.862. dirección: https://remef.org.mx/index.php/remef/article/view/862.

[192] W. Samek, T. Wiegand y K.-R. Müller, «Explainable Artificial Intelligence: Understanding, Visualizing and Interpreting Deep Learning Models,» *ITU Journal: ICT Discoveries*, vol. 1, nº 1, págs. 39-48, 2017. dirección: https://www.itu.int/en/journal/001/Pages/default.aspx.

[193] P. A. Samuelson y W. D. Nordhaus, *Economía con aplicaciones a Latinoamérica*, Spanish, 19ª ed. McGraw-Hill Interamericana, 2019, ISBN: 9781456272197.

[194] J. Schmidt-Hieber, «Nonparametric Regression Using Deep Neural Networks with ReLU Activation Function,» *The Annals of Statistics*, vol. 48, nº 4, págs. 1875-1897, 2020. DOI: 10.1214/19-AOS1875.

[195] J. Schmidt-Hieber, «The Kolmogorov–Arnold representation theorem revisited,» *Neural Networks*, vol. 137, págs. 119-126, 2021. DOI: 10.1016/j.neunet.2021.01.020.

[196] O. B. Sezer, M. U. Gudelek y A. M. Ozbayoglu, «Financial Time Series Forecasting with Deep Learning: A Systematic Literature Review: 2005–2019,» *Applied Soft Computing*, vol. 90, pág. 106 181, 2020. DOI: 10.1016/j.asoc.2020.106181.

[197] W. R. Shadish, T. D. Cook y D. T. Campbell, *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston, MA: Houghton Mifflin, 2002, ISBN: 978-0395615560.

[198] W. F. Sharpe, «Capital Asset Prices: A Theory of Market Equilibrium Under Conditions of Risk,» *The Journal of Finance*, vol. 19, nº 3, págs. 425-442, 1964. DOI: 10.2307/2975924.

[199] Y. Shin, J. Yoon e I. Lee, «Improving the Trainability of Deep Neural Networks Using Kolmogorov Superposition Theorem,» *arXiv preprint arXiv:2106.07869*, 2021. dirección: https://arxiv.org/abs/2106.07869.

[200] J. Sirignano, A. Sadhwani y K. Giesecke, «Deep Learning for Mortgage Risk,» *Journal of Financial Econometrics*, vol. 19, nº 2, págs. 313-368, 2021. DOI: 10.1093/jjfinec/nbaa025.

[201] M. Sokolova y G. Lapalme, «A Systematic Analysis of Performance Measures for Classification Tasks,» *Information Processing & Management*, vol. 45, nº 4, págs. 427-437, 2009. DOI: 10.1016/j.ipm.2009.03.002.

[202] S. Somvanshi, S. A. Javed, M. M. Islam, D. Pandit y S. Das, «A Survey on Kolmogorov-Arnold Network,» *arXiv preprint arXiv:2411.06078*, 2024. dirección: https://arxiv.org/abs/2411.06078.

[203] F. A. Sortino y R. van der Meer, «Downside risk,» *Journal of Portfolio Management*, vol. 17, nº 4, págs. 27-31, 1991. DOI: 10.3905/jpm.1991.409343.

[204] D. A. Sprecher, «On the structure of representations of continuous functions of several variables as finite sums of continuous functions of one variable,» *Proceedings of the American Mathematical Society*, vol. 15, nº 6, págs. 945-952, 1964.

[205] D. A. Sprecher, «On the structure of continuous functions of several variables,» *Transactions of the American Mathematical Society*, vol. 115, nº 3, págs. 340-355, 1965.

[206] D. A. Sprecher, «A survey of solved and unsolved problems on the superposition of functions,» *Journal of Approximation Theory*, vol. 86, nº 2, págs. 233-254, 1996.

[207] Standard & Poor's, «Corporate Methodology,» Standard & Poor's, inf. téc., 2008.

[208] L. J. Tashman, «Out-of-sample tests of forecasting accuracy: an analysis and review,» *International Journal of Forecasting*, vol. 16, nº 4, págs. 437-450, 2000. DOI: 10.1016/S0169-2070(00)00065-0.

[209] C. A. Torres, R. B. Pereira, M. A. Gatti, E. A. Cerchiari y F. Gomide, «Explaining Deep Learning Models for Credit Scoring with SHAP: A Case Study Using Open Banking Data,» *Expert Systems with Applications*, vol. 215, pág. 119 255, 2023. DOI: 10.1016/j.eswa.2022.119255.

[210] Y. Touzani, K. Douzi y F. Khoukhi, «Stock Price Forecasting: New Model for Uptrend Detecting and Downtrend Anticipating,» *Journal of Forecasting*, vol. 42, nº 7, págs. 1641-1661, 2023. DOI: 10.1002/for.2884.

[211] Towards Data Science (Archive). «Understanding Kolmogorov Arnold Networks (KAN).» Fecha de acceso: 10 febrero 2026. dirección: https://medium.com/data-science/understanding-kolmogorov-arnold-networks-kan-e317b1b4d075.

[212] B. R. Trejo Becerril y A. Gallegos David, «Estimación del Riesgo de Mercado utilizando el VaR y la Beta del CAPM,» *Revista Mexicana de Economía y Finanzas*, vol. 16, nº 2, 2021. DOI: 10.21919/remef.v16i2.589. dirección: https://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S1665-53462021000200009.

[213] J. C. Trejo García, J. M. Nyssen y E. I. Téllez León, «Redes neuronales artificiales para la predicción del tipo de cambio peso-dólar,» *Contaduría y Administración*, vol. 61, nº 2, págs. 264-286, 2016. DOI: 10.1016/j.cya.2015.03.003.

[214] R. S. Tsay, *Analysis of Financial Time Series*, 3rd. Hoboken, NJ: Wiley, 2010, ISBN: 9780470414354.

[215] S. Valdivia y A. Morales, «Determinants of the Index of Prices and Quotations on the Mexican Stock Exchange: Sensitivity Analysis Based on Artificial Neural Networks,» *Global Journal of Business Research*, vol. 10, nº 2, págs. 27-32, 2016. dirección:

https://www.theibfr2.com/RePEc/ibf/gjbres/gjbr-v10n2-2016/GJBR-V10N2-2016-3.pdf.

[216] F. Venegas-Martínez, «Modelos econométricos de series de tiempo y redes neuronales artificiales para la predicción de la inflación en México,» *Comercio Exterior*, vol. 56, nº 5, págs. 393-403, 2006. dirección: https://almanaque.colmex.mx/wp-content/uploads/2019/11/redes.pdf.

[217] A. G. Vitushkin, «On Hilbert's thirteenth problem,» *Doklady Akademii Nauk SSSR*, vol. 95, págs. 701-704, 1954.

[218] A. G. Vitushkin, «On Hilbert's thirteenth problem and related questions,» *Russian Mathematical Surveys*, vol. 59, nº 1, págs. 11-25, 2004.

[219] T. Wang y S. Singh, *KAN based Autoencoders for Factor Models*, arXiv, 2024. DOI: 10.48550/arXiv.2408.02694. dirección: https://arxiv.org/abs/2408.02694.

[220] Y. Wang et al., «Kolmogorov-Arnold Informed Neural Network: A Physics-Informed Deep Learning Framework for Solving PDEs Based on Kolmogorov-Arnold Networks,» *arXiv preprint arXiv:2406.11045*, 2024. dirección: https://arxiv.org/abs/2406.11045.

[221] Y. Wang, J. W. Siegel, Z. Liu y T. Y. Hou, «On the Expressiveness and Spectral Bias of KANs,» en *International Conference on Learning Representations (ICLR 2025)*, 2025. dirección: https://arxiv.org/abs/2410.01803.

[222] F. Wilcoxon, «Individual Comparisons by Ranking Methods,» *Biometrics Bulletin*, vol. 1, nº 6, págs. 80-83, 1945. DOI: 10.2307/3001968.

[223] P. Wilmott, *Machine Learning: An Applied Mathematics Introduction*. Panda Ohana Publishing, 2019, ISBN: 978-1916081604.

[224] K. Xu, L. Chen y S. Wang, «Kolmogorov-Arnold Networks for Time Series: Bridging Predictive Power and Interpretability,» *arXiv preprint arXiv:2406.02496*, 2024. dirección: https://arxiv.org/abs/2406.02496.

[225] Yahoo Finance. «IPC MEXICO (^MXX): datos históricos y cotización.» Consultado: febrero 2026. dirección: https://finance.yahoo.com/quote/%5C%5EMXX/.

[226] R. Yu, W. Yu y X. Wang, «KAN or MLP: A Fairer Comparison,» *arXiv preprint arXiv:2407.16674*, 2024. dirección: https://arxiv.org/abs/2407.16674.

[227] A. L. Zhang. «The Annotated Kolmogorov-Arnold Network (KAN).» Fecha de acceso: 10 febrero 2026. dirección: https://alexzhang13.github.io/blog/2024/annotated-kan/.