

UNIVERSIDAD DEL VALLE DE GUATEMALA
Facultad de Ingeniería



**OPTIMIZACIÓN DE UN AGENTE MULTI-MODAL
BASADO EN APRENDIZAJE PROFUNDO PARA LA
PREDICCIÓN DE TENDENCIAS EN MERCADOS
FINANCIEROS**

Trabajo de graduación presentado por Diego Estuardo Lemus López
para optar al grado académico de Licenciado en Ingeniería en Ciencias
de la Computación y Tecnologías de la Información

Guatemala,

2025

UNIVERSIDAD DEL VALLE DE GUATEMALA
Facultad de Ingeniería



**OPTIMIZACIÓN DE UN AGENTE MULTI-MODAL
BASADO EN APRENDIZAJE PROFUNDO PARA LA
PREDICCIÓN DE TENDENCIAS EN MERCADOS
FINANCIEROS**

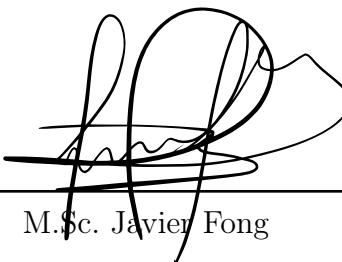
Trabajo de graduación presentado por Diego Estuardo Lemus López
para optar al grado académico de Licenciado en Ingeniería en Ciencias
de la Computación y Tecnologías de la Información

Guatemala,

2025

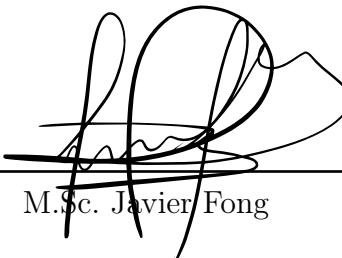
Vo.Bo.:

(f) _____
M.Sc. Javier Fong



Tribunal Examinador:

(f) _____
M.Sc. Javier Fong



(f) _____
Ing. Marlon Fuentes



Fecha de aprobación: Guatemala, 22 de noviembre de 2025.

Prefacio

Durante décadas, el entendimiento e incluso la predicción del movimiento de precios en mercados financieros ha sido uno de los desafíos más complejos en el campo de las ciencias de la computación y la econometría. La hipótesis de mercados eficientes sugiere que toda la información disponible ya está reflejada en los precios actuales, haciendo que cualquier intento de predicción sea, en teoría, equivalente al azar. No obstante, los avances recientes en aprendizaje profundo y el acceso sin precedentes a fuentes de datos heterogéneas han abierto nuevas posibilidades para desafiar esta premisa.

Bajo la intersección entre dos observaciones fundamentales, se pretende abordar lo expuesto anteriormente. Primero, los mercados financieros modernos responden tanto a indicadores técnicos tradicionales, como también a señales sociales y dinámicas macroeconómicas complejas. Segundo, las arquitecturas de redes neuronales recurrentes, particularmente las LSTM (Long Short-Term Memory), han demostrado poseer una capacidad superior para capturar dependencias temporales de largo plazo que caracterizan las series financieras, sin dejar de lado los patrones y la volatilidad intrínseca de este tipo de datos.

La selección de Tesla, Inc. como activo de estudio responde a razones fundamentadas. Su marcada volatilidad, su constante presencia mediática y la influencia reconocida que ejercen las comunicaciones de su CEO en redes sociales la convierten en un caso ideal para poner a prueba un enfoque multimodal. En múltiples aspectos, Tesla es un activo representativo del siglo XXI, en el que las dinámicas de mercado, los efectos sociales y los elementos técnicos se entrelazan de formas que los modelos convencionales apenas pueden abarcar.

Este planteamiento busca contribuir al conocimiento académico en el área de finanzas computacionales, proporcionando herramientas prácticas para la toma de decisiones en mercados reales. Los resultados obtenidos demuestran la capacidad de predecir correctamente la dirección de movimientos de precio en 3 diferentes horizontes temporales, aunque con magnitudes conservadoras, que realmente terminan siendo un agregado el objetivo principal del trabajo. Esto apoya la hipótesis de que la integración inteligente de múltiples fuentes de información puede proporcionar ventajas sistemáticas en entornos de alta incertidumbre.

Prefacio	V
Lista de figuras	XI
Lista de cuadros	XIII
Resumen	XV
Abstract	XVII
1. Introducción	1
2. Justificación	3
3. Objetivos	5
3.1. Objetivo general	5
3.2. Objetivos específicos	5
4. Alcance	7
4.1. Delimitación del activo objetivo	7
4.2. Delimitación temporal	8
4.3. Delimitación operacional	8
4.4. Limitaciones metodológicas reconocidas	9
5. Marco teórico	11
5.1. Mercados financieros y trading algorítmico	11
5.1.1. Estructura y dinámica de los mercados financieros	11
5.1.2. Trading algorítmico: fundamentos y evolución	11
5.1.3. Limitaciones y riesgos del trading automatizado	12
5.2. Series temporales financieras	12
5.2.1. Propiedades estadísticas y no estacionariedad	12
5.2.2. Modelos clásicos de predicción	13
5.2.3. Descomposición de señales	13
5.2.4. Métodos contemporáneos: aprendizaje profundo para series temporales	14

5.3. Aprendizaje profundo en finanzas	14
5.3.1. Redes neuronales profundas: arquitecturas y principios	14
5.3.2. Definición y conceptualización de un agente	15
6. Flujo de stock	17
6.1. Ingesta de datos	17
6.2. Procesamiento de datos	18
6.3. Generación de características	18
6.3.1. Indicadores técnicos	19
6.3.2. Métricas de volatilidad	19
6.3.3. Descomposición de series temporales	19
6.3.4. Normalización y representación final	20
6.4. Validación interna	20
6.4.1. Verificación de calidad	20
6.4.2. Análisis de significancia estadística	20
6.4.3. Consistencia temporal	20
6.5. Salida	21
7. Flujo de análisis de sentimiento	23
7.1. Ingesta de datos	23
7.2. Procesamiento de datos	24
7.3. Generación de características	25
7.3.1. Score de sentimiento	26
7.3.2. Score de acceptance	26
7.4. Salida	27
8. Flujo de mercado	29
8.1. Ingesta de datos	29
8.2. Procesamiento de datos	30
8.3. Generación de características	30
8.3.1. Análisis de dependencias cross-seccionales	30
8.3.2. Reducción dimensional por categoría	31
8.3.3. Selección automatizada de características	31
8.4. Validación interna	32
8.4.1. Verificación de integridad	32
8.4.2. Diagnóstico de relaciones	32
8.5. Salida	32
9. Flujo de fusión multimodal	33
9.1. Arquitectura y objetivos	33
9.1.1. Objetivos del sistema de fusión	33
9.1.2. Fuentes de información y sus características	34
9.2. Metodología de alineación temporal	34
9.2.1. Estrategia de join temporal	34
9.2.2. Filtrado de observaciones incompletas	35
9.2.3. Validación de integridad temporal	35

10. Modelo de predicción	37
10.1. Arquitectura	37
10.1.1. Fundamentos	37
10.1.2. Componentes	38
10.1.3. Variantes del modelo	39
10.2. Entrenamiento	40
10.2.1. Preparación de datos	40
10.2.2. Optimización	41
10.2.3. Regularización y control de sobreajuste	42
10.3. Selección automática de características	43
10.3.1. Fundamentación	43
10.3.2. Metodología de selección	43
10.3.3. Configuración final	45
10.4. Evaluación	46
10.4.1. Métricas estadísticas	46
10.4.2. Métricas financieras	46
10.4.3. Escalamiento de volatilidad temporal	47
11. Agente	49
11.1. Introducción	49
11.2. Arquitectura del sistema	49
11.2.1. Módulo de predicción	49
11.2.2. Módulo de generación de señales	51
11.2.3. Módulo de validación retrospectiva	52
12. Resultados	55
12.1. Descripción de los datos	55
12.1.1. Tesla stock	55
12.1.2. Análisis de sentimiento	61
12.1.3. Stock Market	64
12.2. Validación de fusión multimodal	66
12.2.1. Cobertura temporal por modalidad	66
12.2.2. Configuraciones experimentales generadas	67
12.2.3. Validación de integridad y ausencia de <i>data leakage</i>	68
12.2.4. Evaluación del <i>trade-off</i> cobertura-completitud	68
12.2.5. Verificación de rangos metodológicos	69
12.3. Desempeño de los modelos	69
12.3.1. Selección de características y evaluación de modelos	69
12.4. Resultados del agente	71
12.4.1. Hallazgos clave	71
12.4.2. Metodología de validación	72
12.4.3. Predicciones vs. realidad	72
12.4.4. Análisis de error y conservadurismo	73
12.4.5. Caso de estudio: rally de tesla (abril-mayo 2025)	73
12.4.6. Métricas de trading	74
12.4.7. Comparación con estrategias baseline	74
13. Conclusiones	75

14. Recomendaciones	77
15. Bibliografía	79
16. Anexos	85
16.1. Gráficos adicionales	85
16.2. Repositorio del proyecto	88
17. Glosario	91

Lista de figuras

1.	Estructura de una célula Long Short-Term Memory (LSTM).	38
2.	Valores y relación de las variables originales OHLCV.	56
3.	Estadísticas móviles para validar propiedades de estacionariedad.	57
4.	Descomposición clásica y sus componentes.	59
5.	Descomposición de la transformada de wavelet en sus distintos niveles.	60
6.	Matriz de correlación entre métricas de interacción originales.	62
7.	Serie temporal de <code>sentiment_score</code> y <code>acceptance</code> con media móvil 30 días.	64
8.	Predicciones multi-horizonte vs. precios reales de Tesla.	72
9.	Precio, retornos y volatilidad (ventana 21 días).	85
10.	Distribución de puntuaciones de sentimiento y métrica de engagement.	85
11.	Correlaciones rolling entre proveedores seleccionados y Tesla.	86
12.	Correlaciones rolling entre tech peers y Tesla.	86
13.	Correlaciones rolling entre competidores automotrices y Tesla.	87
14.	Pérdida: entrenamiento y validación. Horizonte corto (1 día).	87
15.	Pérdida: entrenamiento y validación. Horizonte medio (5 días).	87
16.	Pérdida: entrenamiento y validación. Horizonte largo (21 días).	88
17.	Distribuciones de retornos y análisis de volatilidad por horizonte (1, 5, 21 días).	88

Lista de cuadros

1.	Criterios de selección automatizada de características.	31
2.	Estadísticos descriptivos de los retornos diarios de Tesla	56
3.	Resultados de tests de estacionariedad	56
4.	Información mutua de características transformadas	58
5.	Estadísticas descriptivas de métricas de engagement en tweets.	61
6.	Términos más frecuentes en los tweets.	63
7.	Estadísticos descriptivos de las variables de sentimiento.	63
8.	Conteo de registros y observaciones sobre integridad.	64
9.	Correlaciones entre proveedores y Tesla.	65
10.	Regresión OLS multivariada con Tech Peers como predictores.	65
11.	Correlaciones entre Tesla y competidores del sector automotriz.	66
12.	Resultados del test de cointegración entre Tesla y competidores.	66
13.	Cobertura temporal por modalidad en el pipeline de fusión	67
14.	Dimensionalidad de configuraciones experimentales	67
15.	Configuraciones modales y sus características estructurales.	69
16.	Desempeño de modelos uni-modales (horizonte largo).	70
17.	Top 10 modelos multi-horizonte ordenados por directional accuracy.	70
18.	Comparación uni-modal vs multimodal por horizonte.	71
19.	Diseño temporal de la validación empírica	72
20.	Comparación detallada de predicciones contra valores reales	72
21.	Ánalisis del MAE en validación empírica	73
22.	Métricas de desempeño en trading de los modelos seleccionados	74
23.	Comparación de estrategias baseline	74

Resumen

Este trabajo presenta el diseño, implementación y validación de un sistema multimodal basado en redes neuronales recurrentes LSTM para predecir movimientos de precio en acciones de Tesla, Inc. El sistema integra tres flujos de información: señales técnicas del activo (descomposiciones temporales, indicadores de volatilidad y momentum), sentimiento de redes sociales obtenido mediante análisis léxico-heurístico (VADER) y características de mercado derivadas de correlaciones sectoriales, causalidad de Granger y análisis de componentes principales.

La arquitectura emplea modelos LSTM profundos con mecanismos de atención temporal, normalización por capas y regularización adaptativa. Se desarrollaron cinco configuraciones: tres modelos unimodales (*Stock*, *Sentiment*, *Market*) y dos variantes multimodales de fusión. Cada modelo se entrenó para horizontes de 1, 5 y 21 días.

El conjunto de datos contiene observaciones diarias desde enero de 2020 hasta abril de 2025, procesadas mediante *pipelines* que aseguran estacionariedad, ausencia de *data leakage* y alineación entre modalidades. El análisis exploratorio revela alta volatilidad en Tesla y correlaciones sectoriales consistentes. La validación empírica sobre tres predicciones reales (abril–mayo 2025) muestra una *directional accuracy* del 100 %, superando ampliamente un *baseline* aleatorio. Los modelos presentan sesgo conservador en magnitudes, con MAE promedio de 17.54 %, especialmente durante el *rally* de +45 % en 21 días, donde se predijo la dirección pero se subestimó la intensidad del movimiento.

El sistema supera estrategias tradicionales, alcanzando *profit factors* mayores a 1.35, *hit rates* superiores al 55 % y *Sharpe ratios* por encima de 2.0. Los resultados confirman el potencial de los enfoques multimodales basados en aprendizaje profundo para mejorar la predicción direccional en entornos financieros complejos.

Palabras clave: Predicción financiera, LSTM, análisis multimodal, análisis de sentimiento, trading algorítmico, aprendizaje profundo, series temporales, mercados de capitales.

Abstract

This paper presents the design, implementation, and validation of a multimodal system based on LSTM recurrent neural networks to predict price movements of Tesla, Inc. shares. The system integrates three information streams: technical signals from the asset (time decompositions, volatility and momentum indicators), social-media sentiment derived via lexicon-heuristic analysis (VADER), and market features extracted from sectoral correlations, Granger causality, and principal component analysis.

The predictive architecture employs deep LSTM models with temporal attention mechanisms, layer normalization, and adaptive regularization. Five configurations were developed: three unimodal models (*Stock*, *Sentiment*, *Market*) and two multimodal fusion variants. Each model was trained for 1-, 5-, and 21-day horizons.

The dataset comprises daily observations from January 2020 through April 2025, processed via *pipelines* that ensure stationarity, prevent temporal *data leakage*, and align modalities. Exploratory analysis reveals high volatility in Tesla and consistent sectoral correlations. Empirical validation on three real forecasts (April–May 2025) shows a *directional accuracy* of 100 %, substantially outperforming a random *baseline*. Models exhibit conservative magnitude bias, with an average MAE of 17.54 %, notably during a +45 % 21-day *rally* when direction was correctly predicted but magnitude underestimated.

The system outperforms traditional strategies, achieving *profit factors* above 1.35, *hit rates* over 55 %, and *Sharpe ratios* exceeding 2.0. Results confirm the potential of multimodal deep-learning approaches to improve directional prediction in complex financial environments.

Keywords: Financial prediction, LSTM, multimodal analysis, sentiment analysis, algorithmic trading, deep learning, time series, capital markets.

CAPÍTULO 1

Introducción

Los mercados financieros representan sistemas complejos adaptativos donde miles de agentes interactúan continuamente, generando series temporales de precios caracterizadas por volatilidad, no-estacionariedad y dependencias temporales de largo plazo. La predicción de movimientos futuros de precios ha sido históricamente uno de los problemas más desafiantes en econometría y ciencias de la computación, con implicaciones directas en la gestión de riesgos, asignación de capital y generación de retornos ajustados por riesgo.

La hipótesis de mercados eficientes (EMH), propuesta por Eugene Fama en la década de 1970 [1], establece que en mercados líquidos y con información simétrica, los precios actuales reflejan toda la información disponible, haciendo que los retornos futuros sean impredecibles más allá del azar. Sin embargo, evidencia empírica acumulada en las últimas décadas ha documentado múltiples anomalías de mercado, momentum y efectos estacionales; que sugieren predictibilidad limitada en horizontes específicos. El debate entre eficiencia y predictibilidad ha motivado el desarrollo de técnicas cada vez más sofisticadas de análisis cuantitativo.

Los métodos tradicionales de predicción financiera, basados en modelos estadísticos como ARIMA, GARCH y VAR, asumen linealidad, estacionariedad y estructuras paramétricas específicas que frecuentemente son violadas en datos reales. El auge del aprendizaje automático en la década de 2000, particularmente con algoritmos como Random Forest, Support Vector Machines y Gradient Boosting, introdujo capacidad de modelar relaciones no lineales, pero estos métodos típicamente tratan observaciones como independientes, ignorando la estructura temporal inherente a series financieras.

La revolución del aprendizaje profundo, iniciada con el éxito de redes convolucionales en visión por computadora y redes recurrentes en procesamiento de lenguaje natural, ha demostrado capacidad sin precedentes para capturar patrones complejos en datos de alta dimensionalidad. Las redes LSTM, introducidas por Hochreiter y Schmidhuber en 1997 [2], constituyen una arquitectura especializada para modelar dependencias temporales de largo plazo mediante mecanismos de compuertas que regulan el flujo de información a través del tiempo. Estudios recientes han validado su efectividad en predicción de series temporales

financieras, superando consistentemente *baselines* tradicionales en métricas de *directional accuracy* y *Sharpe ratio*.

Paralelamente, la democratización de acceso a fuentes de datos no tradicionales ha habilitado enfoques multimodales que integran señales heterogéneas para mejorar capacidad predictiva. El caso de Tesla, Inc. ejemplifica este nuevo paradigma: su precio no solo responde a fundamentales financieros tradicionales, sino también a declaraciones de su CEO en Twitter/X, percepciones de mercado sobre innovación tecnológica, y dinámicas competitivas en el sector de vehículos eléctricos. Esta riqueza informativa multidimensional requiere sistemas capaces de consolidar, ponderar y explotar señales de naturaleza dispar.

Por ello, se plantea el diseño, implementación y validación de un sistema multimodal basado en redes LSTM profundas que integra características técnicas del activo, análisis de sentimiento de redes sociales y contexto macroeconómico de mercado. El estudio se enfoca específicamente en predicción direccional (clasificación de movimientos alcistas/bajistas) y estimación de magnitud de retornos para acciones de Tesla, Inc., operando en frecuencia diaria con horizontes de 1, 5 y 21 días hábiles. Se aportan tres contribuciones principales:

- un protocolo de alineación temporal que garantiza la ausencia de *data leakage* mediante verificación exhaustiva de correspondencias causales;
- una evaluación comparativa sistemática de las contribuciones por modalidad mediante un diseño experimental controlado que aísla efectos individuales y sinérgicos;
- validación empírica retrospectiva sobre predicciones reales post-entrenamiento que confirma la capacidad de generalización fuera del conjunto de desarrollo.

Las finalidades de este trabajo son proporcionar una metodología reproducible y robusta para integrar señales heterogéneas en tareas de predicción financiera, y evaluar la aplicabilidad práctica de agentes predictivos multimodales como herramienta de soporte a la toma de decisiones en trading sistemático. En síntesis, los resultados principales muestran que los modelos unimodales especializados por horizonte superan a las arquitecturas de fusión en los escenarios estudiados; la modalidad de sentimiento presenta mayor precisión en horizonte corto (56.65 % de *directional accuracy*, Sharpe 2.79 en 1 día), la modalidad técnica mantiene ventaja en horizontes de 5 y 21 días (56.36 % y 55.93 % de accuracy; Sharpe 1.96 y 1.72). La selección automática y orquestación de modelos muestran ser una estrategia efectiva para adaptar el agente según horizonte y condiciones del mercado; el pipeline de sincronización asegura integridad temporal y evita *data leakage*. Finalmente, la validación sobre tres predicciones reales (17-abr-2025 a 16-may-2025) alcanzó 100 % de *directional accuracy*, aunque los modelos exhiben sesgo conservador en magnitud, con un MAE promedio 15.21 pp. Estos resultados orientan recomendaciones prácticas sobre objetivos de gestión de riesgo en implementaciones reales.

CAPÍTULO 2

Justificación

La realización de este trabajo responde a la necesidad de enfrentar los desafíos actuales en la comprensión y modelado de los mercados financieros, que se caracterizan por una alta volatilidad y complejidad. Los métodos tradicionales, centrados en una única fuente de información, a menudo resultan insuficientes para capturar la totalidad de la información relevante. En este contexto, la integración de un enfoque multimodal se presenta como una solución innovadora, pues permite combinar datos económicos, técnicos y sentimentales, entre otros; ofreciendo una visión más completa y precisa del comportamiento del mercado.

Actualmente, los avances en aprendizaje profundo han demostrado un notable potencial para resolver tareas complejas de clasificación y predicción en dominios de alta dimensionalidad. La combinación de modelos para series temporales con modelos de lenguaje natural pre entrenados permite extraer señales tanto numéricas como semánticas, lo cual resulta particularmente útil en el análisis financiero, donde se reconoce que el comportamiento del mercado puede estar influido tanto por métricas cuantitativas como por factores cualitativos externos.

La construcción de un agente predictivo multimodal ofrece una alternativa más manejable y escalable en comparación con esquemas de aprendizaje por refuerzo o entornos simulados, al mismo tiempo que mantiene un alto nivel de sofisticación técnica. Además, este enfoque facilita el entrenamiento y la evaluación rigurosa del sistema en condiciones reales de mercado, permitiendo aplicar metodologías reproducibles y comparables que aumentan el valor científico y aplicado del trabajo.

En conjunto, este trabajo busca reducir la brecha existente entre la teoría del aprendizaje multimodal y su implementación práctica en la predicción de mercados financieros, proporcionando evidencia empírica y una base metodológica reproducible.

CAPÍTULO 3

Objetivos

3.1. Objetivo general

Desarrollar un agente predictivo multi-modal basado en aprendizaje profundo supervisado, capaz de integrar datos estructurados y no estructurados para predecir tendencias en los mercados financieros con alto grado de precisión y generalización, para que sirva como herramienta de soporte en la toma de decisiones de inversión.

3.2. Objetivos específicos

- Diseñar la arquitectura del agente, definiendo la forma de integración entre los diferentes modelos y modalidades de datos, así como los parámetros y protocolos necesarios para su funcionamiento.
- Implementar el proceso de ingestión, pre procesamiento y sincronización de múltiples fuentes de datos heterogéneos, incluyendo series temporales financieras, indicadores técnicos y texto financiero no estructurado.
- Evaluar el desempeño del agente multi-modal comparándolo con versiones unidimensionales.

CAPÍTULO 4

Alcance

El presente trabajo delimita su objeto de estudio, metodología y contexto temporal con el propósito de garantizar profundidad analítica, reproducibilidad técnica y validez de conclusiones dentro de un marco experimental controlado.

4.1. Delimitación del activo objetivo

El estudio se restringe exclusivamente a acciones de Tesla, Inc. como activo de análisis. Esta delimitación se fundamenta en tres criterios técnicos y metodológicos que justifican la elección de un caso de estudio único en lugar de una muestra multi-activo:

Disponibilidad de datos multimodales de alta calidad. Tesla representa uno de los pocos activos bursátiles con exposición mediática suficientemente amplia para generar volumen significativo de datos de sentimiento social. El historial público de comunicaciones de su CEO en Twitter/X constituye un corpus textual denso y continuado, condición necesaria para entrenamiento supervisado de modelos de análisis de sentimiento. La mayoría de empresas cotizadas carecen de esta riqueza informativa en redes sociales, lo que imposibilitaría replicar el enfoque multimodal propuesto.

Volatilidad representativa de activos tecnológicos de alta capitalización. Tesla exhibe volatilidad intrínseca significativa, característica que la posiciona como caso representativo del segmento de acciones de crecimiento con alta sensibilidad a factores cualitativos. Esta propiedad permite evaluar capacidad predictiva del sistema en condiciones de alta incertidumbre, escenario donde modelos tradicionales típicamente fallan. Activos de menor volatilidad presentarían menor desafío predictivo y limitarían validez externa de conclusiones sobre efectividad de enfoques multimodales.

Factibilidad computacional y temporal. El desarrollo completo de un sistema multimodal de predicción financiera —ingesta de datos multi-fuente, ingeniería de características específicas por modalidad, entrenamiento de arquitecturas LSTM profundas con búsqueda de hiperparámetros, y validación temporal rigurosa— requiere inversión sustancial de recursos computacionales y tiempo de desarrollo. La extensión del estudio a múltiples activos multiplicaría estos costos por un factor N , excediendo capacidades operativas de un proyecto de esta índole sin comprometer profundidad analítica. La estrategia adoptada privilegia *depth over breadth*: análisis exhaustivo de un caso paradigmático sobre análisis superficial de múltiples casos.

Si bien el objeto de estudio es un activo individual, las **metodologías, técnicas de fusión multimodal y protocolos de validación desarrollados son generalizables** a otros activos con características similares (alta capitalización, exposición mediática, volatilidad elevada). El trabajo no se limita a caracterizar propiedades idiosincráticas de Tesla, sino que establece un *framework* metodológico replicable para predicción en mercados financieros modernos donde información heterogénea influye en formación de precios. La contribución trasciende el caso particular y se posiciona como estudio de viabilidad técnica de enfoques multimodales en contextos financieros de alta complejidad.

4.2. Delimitación temporal

El conjunto de datos comprende observaciones diarias desde el **1 de enero de 2020** hasta el **16 de abril de 2025**, abarcando aproximadamente 5.3 años de historia de mercado. Esta ventana temporal fue diseñada estratégicamente para capturar diversidad de regímenes de mercado, de modo que el modelo sea expuesto a múltiples regímenes de volatilidad y correlación sectorial.

La validación empírica retrospectiva se ejecuta sobre tres predicciones puntuales en el periodo **abril–mayo 2025**, correspondientes a horizontes de 1, 5 y 21 días. Esta configuración de validación limitada constituye una *prueba de concepto (proof-of-concept)* que demuestra capacidad del sistema para generar predicciones en condiciones realistas, pero no pretende establecer significancia estadística robusta. La extensión a validación multi-fecha con decenas de predicciones independientes queda fuera del alcance de este trabajo y se identifica como línea futura de investigación.

4.3. Delimitación operacional

El sistema desarrollado constituye un **prototipo de investigación orientado a demostrar viabilidad técnica**, no una plataforma de *trading* en producción. Las siguientes restricciones operacionales delimitan el alcance funcional:

- **Modo de operación offline.** El agente predictivo no opera en tiempo real ni ingesta datos en vivo. Todas las predicciones se generan sobre datos históricos completos cargados desde archivos locales. Esta simplificación elimina complejidades de ingeniería

de software que son ortogonales al objetivo científico de validar efectividad de fusión multimodal para predicción.

- **Fuentes de datos preexistentes.** La ingesta de información opera bajo supuestos idealizados de accesibilidad a datos, específicamente el acceso irrestricto a APIs como `yfinance` sin limitaciones de tasa de consulta ni costos de suscripción.
- **Ausencia de ejecución automática de órdenes.** El agente genera señales de *trading* (BUY, SELL, HOLD) y predicciones de retornos, pero no ejecuta transacciones reales ni simuladas en brokers. La evaluación de rentabilidad se realiza mediante *back-testing* sobre precios históricos, sin considerar fricciones de mercado.

4.4. Limitaciones metodológicas reconocidas

El diseño experimental presenta limitaciones inherentes que deben considerarse en la interpretación de resultados:

- **Tamaño de muestra de validación empírica limitado.** La validación retrospectiva comprende únicamente tres predicciones puntuales, una por horizonte temporal, constituyendo evidencia anecdótica más que confirmación estadística. La extensión a validación multi-fecha queda fuera del alcance temporal de este trabajo.
- **Periodo de validación atípico.** El intervalo de validación empírica (abril–mayo 2025) coincide con un rally extraordinario de Tesla (+45 % en 21 días), evento de cola extrema. Este contexto excepcional explica parcialmente el alto error absoluto medio en predicción de magnitudes (MAE), ya que modelos entrenados sobre condiciones normales de mercado sistemáticamente subestiman movimientos extremos.
- **Generalización a otros activos no validada.** Si bien las metodologías desarrolladas son conceptualmente generalizables, la efectividad empírica demostrada en Tesla no garantiza transferibilidad a otros activos con propiedades distintas. La validación de generalización requeriría un estudio multi-activo que excede recursos disponibles en este trabajo.

CAPÍTULO 5

Marco teórico

5.1. Mercados financieros y trading algorítmico

5.1.1. Estructura y dinámica de los mercados financieros

La estructura de los mercados financieros varía según la economía: en países de “sistema basado en mercado”, por ejemplo EE.UU., la bolsa de valores desempeña el rol central en la asignación de capital y gestión de riesgos; mientras que, en sistemas bancarizados, como Alemania, estas funciones las asumen los bancos. En todo mercado financiero existen infraestructuras clave: mercados de negociación, sistemas de compensación/liquidación y depósitos centrales, los cuales garantizan la ejecución, liquidación y registro de cada operación [3].

La dinámica de precios en estos mercados surge de la interacción continua entre oferta y demanda de activos, bajo reglas explícitas [4]. En las últimas décadas se ha observado una transformación tecnológica profunda, dado que los mercados presenciales han sido reemplazados por sistemas electrónicos donde los participantes envían órdenes a un servidor central que las ejecuta automáticamente. Esta virtualización ha incrementado la automatización del proceso completo, además de generar una difusión de información de liquidez y transacciones prácticamente instantánea. El resultado es una mayor velocidad y eficiencia operativa: los precios pueden ajustarse en milisegundos ante nueva información, y las funciones humanas son sustituidas por algoritmos en cada etapa del mercado [5].

5.1.2. Trading algorítmico: fundamentos y evolución

El trading algorítmico consiste en usar programas informáticos para asistir al usuario, o directamente ejecutar órdenes y estrategias de forma automática. A diferencia del trading tradicional, que es controlado por decisiones y emociones humanas, los algoritmos siguen reglas predefinidas con disciplina absoluta. Esto elimina sesgos emocionales y permite operar a velocidades imposibles manualmente, prácticamente generando decisiones basadas únicamente en datos.

mente en la lógica detrás del propio algoritmo y de la información que este recibe. En la práctica esta automatización facilita detectar oportunidades de mercado fugaces, pues al ser pequeñas ineficiencias o tendencias incipientes, pueden pasar desapercibidas ante un ojo inexperto. O bien, esta automatización puede ser una forma adicional de corroborar una decisión planteada.

Históricamente, los primeros sistemas algorítmicos aparecieron en los años 1990 con la desmaterialización de los parqués: por ejemplo, Island ECN en 1997 introdujo algoritmos para negociar directamente en una plataforma electrónica. A principios de los 2000 se desarrolló el **High Frequency Trading (HFT)** como modalidad ultra-rápida dentro del trading algorítmico, impulsada por avances en cómputo y redes de baja latencia. En esta evolución el volumen de negociación atribuible a algoritmos creció exponencialmente: se estima que para 2017 alrededor del 65 % del volumen mundial en acciones era ejecutado algorítmicamente [5]. El trading algorítmico se ha diversificado en múltiples estrategias como el seguimiento de tendencia, la reversión a la media, arbitraje, creación de mercado, etc., y hoy abarca desde operaciones intradía automatizadas hasta enfoques de *machine learning* aplicados a grandes datos [6].

5.1.3. Limitaciones y riesgos del trading automatizado

Aunque el trading automatizado ofrece ventajas, conlleva riesgos propios que deben gestionarse cuidadosamente. El riesgo tecnológico es notable: cualquier fallo en el software, en el hardware o en la conexión puede provocar pérdidas catastróficas. Un caso paradigmático fue el de Knight Capital (2012), donde un bug en el algoritmo generó pérdidas de más de 440 millones de dólares en apenas 45 minutos. Además, las estrategias algorítmicas pueden reaccionar precipitadamente ante eventos extremos, retirando liquidez y agravando volatilidad. Por ejemplo, durante el Flash Crash de 2010 la rápida retirada de algoritmos **HFT** amplificó momentáneamente la caída de precios. Otros riesgos incluyen la modelización errónea (resultando en predicciones pobres en situaciones no vistas), el riesgo operativo, y el riesgo regulatorio; pues prácticas ilegales como el *spoofing* o *layering* están penalizadas y exigen mayor supervisión. Se extrae que el trading algorítmico es poderoso pero su complejidad exige medidas robustas de supervisión tecnológica y de mercado para evitar inestabilidades sistémicas [7].

5.2. Series temporales financieras

5.2.1. Propiedades estadísticas y no estacionariedad

Las series financieras, en específico los precios y rendimientos de activos, exhiben varios hechos estilizados bien documentados. En primer lugar, los precios tienden a seguir una caminata aleatoria, mientras que las rentabilidades, como los cambios porcentuales diarios, resultan estacionarias con media y varianzas aproximadamente constantes. Esto implica que la diferencia de precios (retornos) puede modelarse como un proceso débilmente estacionario, en contraste con los precios absolutos que contienen tendencias estocásticas o deterministas.

Adicionalmente, las distribuciones de los rendimientos suelen ser leptocúrticas: son más “picudas” y presentan colas pesadas comparadas con la normal. Esto se traduce en una mayor probabilidad de movimientos extremos (comportamiento no gaussiano). Otro hecho estilizado clave es el *cluster* de volatilidad: las series presentan agrupaciones en la volatilidad, de modo que periodos de alta volatilidad tienden a continuar seguidos de alta volatilidad y lo mismo ocurre con periodos tranquilos. En conjunto, estas propiedades: no estacionariedad de precios, estacionariedad de rendimientos, colas gruesas y volatilidad agrupada; son fundamentales para seleccionar modelos adecuados de predicción.

5.2.2. Modelos clásicos de predicción

Los modelos tradicionales de series temporales en finanzas combinan componentes lineales para la media y para la varianza. En entornos univariantes, la metodología ARIMA [8] ha sido clásica: incorpora componentes autorregresivos (Autoregressive (AR)) y de media móvil (Moving Average (MA)) después de las diferenciaciones necesarias para lograr estacionariedad.

Para capturar la heterocedasticidad condicional se usan los modelos ARCH/GARCH de Engle (1982) [9] y su generalización GARCH por Bollerslev (1986) [10], los cuales explican cómo la varianza cambia en el tiempo en función de choques pasados. En otros términos, ARIMA se encarga de la dinámica del nivel medio del proceso, mientras GARCH modela la dinámica de la varianza, reflejando *clústeres* de volatilidad.

En el contexto multivariante, los modelos Vector Autoregression (VAR) introducidos por Sims (1980) [11] extienden los AR a vectores de variables: en un VAR todas las series endógenas se expresan mutuamente en función lineal de rezagos compartidos. Esto permite capturar co-movimientos dinámicos entre variables macroeconómicas o financieras que los ARIMA univariados no detectan. Son ampliamente usados para análisis de sistemas económicos, aunque no incorporan la heterocedasticidad.

5.2.3. Descomposición de señales

Para tratar la no estacionariedad y extraer componentes de diferentes escalas en series financieras se usan técnicas de descomposición de señales. La transformada wavelet [12] descompone la serie en coeficientes localizados en tiempo y escala, lo que permite separar tendencias de largo plazo de oscilaciones de corto plazo. Por ejemplo, aplicando una transformada wavelet de Haar o Daubechies a los precios se obtienen detalles (alta frecuencia) y aproximaciones (baja frecuencia) que se pueden modelar por separado.

De igual modo, la Empirical Mode Decomposition (EMD) propuesta por Huang et al. (1998) [13] descompone adaptativamente la serie en Modo Intrínseco, sin imponer una base predefinida; cada IMF captura oscilaciones características de la señal. Estos enfoques multiresolución ayudan a filtrar ruido y a revelar estructuras ocultas en series financieras no lineales. En la práctica se emplea, por ejemplo, la descomposición wavelet antes de pronosticar con redes neuronales, o la EMD para analizar cíclicamente la volatilidad; ambos casos muestran que los modelos de pronóstico suelen mejorar cuando se modelan por separado los

componentes extraídos.

5.2.4. Métodos contemporáneos: aprendizaje profundo para series temporales

En los últimos años han proliferado los modelos de aprendizaje profundo para el pronóstico de series financieras. La aplicación de técnicas avanzadas ha permitido modelar estas series con mayor precisión, capturando relaciones no lineales y dependencias temporales que los modelos tradicionales no pueden detectar [5]. Las arquitecturas típicas incluyen redes neuronales recurrentes, especialmente **LSTM** o **Gated Recurrent Unit (GRU)**, capaces de capturar dependencias temporales de largo plazo; o enfoques híbridos. Estos modelos aprenden automáticamente *features* complejas a partir de datos históricos masivos. Sin embargo, su uso plantea desafíos: suelen ser “cajas negras” difíciles de interpretar y requieren grandes volúmenes de datos y recursos de cómputo para entrenar redes robustas. En ciertos casos las **LSTM** pueden superar a métodos estadísticos clásicos en predicción de precios de acciones, aunque su rendimiento puede degradarse en condiciones de mercado muy cambiantes.

5.3. Aprendizaje profundo en finanzas

El aprendizaje profundo ha revolucionado el campo de las finanzas al proporcionar herramientas capaces de procesar grandes volúmenes de datos y extraer información valiosa para la toma de decisiones. En el contexto del trading algorítmico, los modelos de aprendizaje profundo pueden identificar patrones complejos en datos de mercado, noticias financieras y otros indicadores, permitiendo la ejecución de estrategias de inversión más precisas y adaptativas [14].

5.3.1. Redes neuronales profundas: arquitecturas y principios

Las **Deep Neural Networks (DNN)** consisten en capas de neuronas artificiales que extraen representaciones jerárquicas de los datos mediante funciones de activación no lineales (ReLU, sigmoide, etc.). Entre las arquitecturas más comunes están:

- **Redes feedforward (Perceptrón multicapa):** Base de las **DNN** tradicionales con múltiples capas totalmente conectadas. Permiten modelar relaciones complejas, pero no guardan memoria de estados previos.
- **Convolutional Neural Network (CNN):** Emplean filtros convolucionales para extraer patrones locales.
- **Recurrent Neural Networks (RNN):** Incorporan bucles de retroalimentación para procesar secuencias de datos. Las **LSTM** resuelven el problema de desvanecimiento del gradiente mediante puertas de entrada, olvido y salida.

- **Transformers:** Arquitecturas basadas en mecanismos de atención que procesan secuencias sin recurrencia. Modelos como el *Temporal Fusion Transformer* combinan atención y componentes recurrentes para previsión multihorizonte.

Estas arquitecturas comparten los principios de entrenamiento profundo: gran capacidad de modelado y entrenamiento por lotes mediante optimizadores numéricos. En finanzas se adicionan consideraciones específicas por la no-estacionariedad de los datos [6], [7], [15].

5.3.2. Definición y conceptualización de un agente

Un agente en inteligencia artificial es una entidad que percibe su entorno mediante entradas (percepts), procesa esa información y produce salidas o acciones con un propósito u objetivo definido [16]. Los agentes constituyen un marco conceptual para distinguir programas reactivos de sistemas autónomos con capacidad de decisión.

En términos de arquitectura, los agentes se clasifican en tipos como: reactivos, basados en modelo, orientados a objetivos, utilitarios o de aprendizaje. Una subclase relevante para esta investigación es la de agentes *analíticos/predictivos*, entendidos como sistemas de software que integran aprendizaje automático y procesamiento de información para estimar estados futuros o generar recomendaciones, sin intervenir directamente en el entorno, aunque con potencial para hacerlo [17]. Este tipo de agente opera como componente predictivo que brinda apoyo a la decisión, en lugar de como un agente ejecutor autónomo.

En el dominio financiero, se distingue entre los modelos *agent-based* de mercado, donde agentes individuales simulan estrategias y participan en negociación, y los agentes de análisis que generan señales o previsiones a partir de datos disponibles. Estos últimos resultan especialmente relevantes en el contexto de los sistemas predictivos aplicados a finanzas cuantitativas.

CAPÍTULO 6

Flujo de stock

6.1. Ingesta de datos

El pipeline de datos financieros estructurados se fundamenta en la captura sistemática de series temporales de precios de activos bursátiles mediante la API de Yahoo Finance ([yfinance](#)). Esta fuente se seleccionó por su accesibilidad, confiabilidad histórica y cobertura temporal extensa, siguiendo las prácticas establecidas para la construcción de datasets de mercados de capitales [\[5\]](#).

El periodo de análisis comprende datos diarios desde el **31 de diciembre de 2019** hasta el 1 de julio de 2025, abarcando un horizonte temporal de aproximadamente 5.5 años. Esta ventana temporal fue diseñada estratégicamente para capturar diferentes regímenes de mercado, incluyendo períodos de alta volatilidad y crisis, así como fases de recuperación y crecimiento [\[18\]](#). La frecuencia diaria de muestreo se justifica por su capacidad de equilibrar granularidad informativa con estabilidad de señal, evitando el ruido inherente a frecuencias intradía mientras mantiene resolución suficiente para detectar patrones de mediano plazo [\[19\]](#).

Las variables fundamentales capturadas incluyen los precios de apertura, cierre, máximo y mínimo, junto con el volumen de transacciones. Estas variables constituyen el conjunto estándar [Open, High, Low, Close, Volume \(OHLCV\)](#), ampliamente utilizado en finanzas cuantitativas por su capacidad descriptiva del comportamiento intradiario del mercado [\[20\]](#). Los datos se almacenan inicialmente en estructuras DataFrame con indexación temporal, facilitando operaciones de procesamiento y análisis subsecuentes.

6.2. Procesamiento de datos

El procesamiento de datos implementa un enfoque multi-etapa orientado a garantizar calidad y coherencia temporal. La fase inicial de limpieza aborda la normalización temporal mediante la conversión de índices a formato datetime sin información de zona horaria, homogeneizando la representación temporal y evitando inconsistencias en operaciones de merge posterior [21].

El tratamiento de valores faltantes se ejecuta mediante interpolación lineal para días laborables sin cotización, bajo el supuesto de que los precios siguen una trayectoria continua entre observaciones consecutivas [22]. Esta técnica preserva la estructura temporal y evita la introducción de sesgos que métodos de imputación más complejos podrían generar en series financieras. Los datos se ajustan por frecuencia de días hábiles (business days), eliminando fines de semana y festivos mediante reindexación a frecuencia 'B'.

La transformación fundamental aplicada consiste en el cálculo de retornos porcentuales simples, definidos como:

$$r_t = \left(\frac{P_t - P_{t-1}}{P_{t-1}} \right) \times 100$$

donde P_t representa el precio de cierre en el tiempo t . Esta transformación captura el cambio relativo de precios en términos porcentuales, proporcionando una medida interpretable de rendimiento que es más estable que los precios absolutos y facilita la comparación entre diferentes períodos temporales, para capturar dinámicas de diferentes horizontes de inversión [23].

Esta transformación no solo normaliza la escala de los datos, sino que resulta fundamental desde el punto de vista de la modelización de series temporales. Los precios de activos financieros típicamente presentan no-estacionariedad debido a la presencia de tendencias estocásticas y cambios estructurales en la media y varianza a lo largo del tiempo. Esta característica viola los supuestos de homogeneidad temporal requeridos por los modelos de redes neuronales recurrentes, pudiendo resultar en aprendizaje de patrones espurios no generalizables.

La transformación a retornos elimina la componente de tendencia no estacionaria, generando una serie con propiedades estadísticas estables que fluctúa alrededor de una media constante. Esta estabilización es verificada mediante pruebas de estacionariedad, confirmado que la transformación cumple su objetivo metodológico de adecuar los datos a los requisitos del modelado predictivo.

6.3. Generación de características

La ingeniería de características implementa un *framework* multi-dominio que combina indicadores técnicos clásicos, métricas de volatilidad y descomposiciones avanzadas de series temporales.

6.3.1. Indicadores técnicos

- **Relative Strength Index (RSI)**: Oscilador de momentum calculado sobre una ventana de 14 días que identifica condiciones de sobrecompra/sobreventa. Formalmente definido como:

$$RSI = 100 - \left(\frac{100}{1 + RS} \right)$$

donde RS representa la razón entre ganancias y pérdidas promedio [24].

- **Moving Average Convergence Divergence (MACD)**: Sistema de medias móviles exponenciales, de 12 y 26 períodos, que captura cambios en dirección y momentum de tendencias, junto con su línea de señal [24].
- **Bandas de Bollinger**: Envoltorio estadístico que define límites superior e inferior como:

$$BB = MA \pm 2\sigma$$

donde MA es la media móvil de 20 días y σ su desviación estándar, proporcionando una medida normalizada de volatilidad relativa [24].

6.3.2. Métricas de volatilidad

- **Volatilidad histórica**: calculada como desviación estándar móvil de retornos en una ventana de 21 días [24].
- **Ratio High-Low**: utilizado como proxy de volatilidad intradía [24].
- **Average True Range (ATR)**: captura el rango verdadero de precios considerando *gaps* entre sesiones [24].

6.3.3. Descomposición de series temporales

Se aplican dos técnicas complementarias de descomposición que permiten separar componentes de diferente naturaleza temporal:

- **Descomposición Clásica**: implementación del modelo aditivo estacional [25]:

$$Y_t = T_t + S_t + R_t$$

donde T_t representa la tendencia subyacente, S_t el componente estacional (periodo de 252 días hábiles anuales) y R_t el residuo irregular. Esta descomposición se ejecuta mediante el método de medias móviles centradas, proporcionando interpretabilidad directa de patrones cíclicos y tendenciales.

- **Descomposición Wavelet**: transformación multi-resolución mediante la aplicación de wavelets Daubechies-4 (db4) con 3 niveles de descomposición [26]. Esta técnica permite separar la señal en componentes de aproximación y detalles. La elección de db4 se justifica por su capacidad de capturar tanto suavidad como discontinuidades presentes en series financieras [27].

La reconstrucción se valida mediante el error absoluto medio entre la señal original y la suma de componentes, garantizando preservación de información.

6.3.4. Normalización y representación final

La normalización se ejecuta mediante *Min-Max scaling* aplicado independientemente a cada característica [28]:

$$x_{\text{norm}} = \frac{x - x_{\min}}{x_{\max} - x_{\min}}$$

Esta transformación proyecta todas las variables al rango $[0, 1]$, facilitando la convergencia en modelos de aprendizaje profundo y eliminando efectos de escala en métricas de distancia.

6.4. Validación interna

6.4.1. Verificación de calidad

Se ejecutan pruebas estadísticas de estacionariedad mediante los tests complementarios de Augmented Dickey-Fuller (ADF) y Kwiatkowski-Phillips-Schmidt-Shin (KPSS) [29]. Esta combinación permite identificar tanto estacionariedad en media (ADF) como en varianza (KPSS), proporcionando una caracterización robusta de las propiedades estocásticas de las series transformadas. Los retornos logarítmicos típicamente exhiben estacionariedad débil ($p_{\text{ADF}} < 0.05$), validando la transformación aplicada.

6.4.2. Análisis de significancia estadística

Se evalúa la significancia de correlaciones mediante el coeficiente de Pearson y sus valores-p asociados. Este análisis identifica relaciones lineales significativas y confirma la relevancia estadística de indicadores técnicos respecto al precio objetivo. Complementariamente, se calcula información mutua mediante regresión, proporcionando una medida no lineal de dependencia entre características y variable objetivo [30].

6.4.3. Consistencia temporal

La estabilidad se verifica mediante análisis de ventanas deslizantes que evalúan la consistencia de estadísticas descriptivas (media, varianza, asimetría, curtosis) a través de sub-periodos [31]. Desviaciones significativas indican cambios de régimen que deben ser considerados en la modelización posterior.

6.5. Salida

El dataset procesado constituye una representación multi-dimensional de la dinámica financiera del activo, estructurado como tensor temporal denso. Este tensor presenta dimensionalidad $\mathbb{R}^{T \times F}$, donde T representa los pasos temporales y F el conjunto de características *engineeradas*.

La señal producida es de naturaleza numérica continua, normalizada al rango unitario, conteniendo información en múltiples escalas temporales:

- **Escala de alta frecuencia:** Indicadores de momentum y volatilidad intradía.
- **Escala media:** Tendencias capturadas por medias móviles y componente estacional.
- **Escala de baja frecuencia:** Aproximación wavelet y tendencia de largo plazo.

Esta representación multi-escala permite al sistema de fusión posterior capturar tanto dinámicas tácticas de corto plazo como estratégicas de largo plazo [32]. La estructura de salida mantiene indexación temporal explícita, facilitando sincronización con flujos de datos textuales mediante operaciones de `merge` temporal con tolerancia configurable.

CAPÍTULO 7

Flujo de análisis de sentimiento

7.1. Ingesta de datos

El pipeline de análisis de sentimiento se fundamenta en la captura y procesamiento de las publicaciones textuales en la red social Twitter/X [33] provenientes del Chief Executive Officer (CEO) de Tesla, Inc., Elon Reeve Musk [34]; cuyas comunicaciones públicas han demostrado correlación significativa con movimientos de mercado.

Se planteó que los datos fueran recolectados mediante *scraping* estructurado de la plataforma X. No obstante, para los fines de desarrollo del proyecto, los datos fueron recolectados directamente de la plataforma Kaggle, de parte del usuario Lyndell, D., abarcando un horizonte temporal alineado con el pipeline de datos financieros (enero 2020 – abril 2025). Esta sincronización temporal es fundamental para permitir análisis de causalidad de Granger y estudios de correlación cruzada entre sentimiento y precio [35] [36]. La frecuencia de captura es *event-driven*, registrando cada publicación con su *timestamp* exacto, lo que permite agregaciones posteriores a nivel diario para sincronización con datos de mercado.

Las variables capturadas incluyen tanto contenido textual como métricas de *engagement*:

1. Contenido textual:

- Texto completo de la publicación (`fullText`)
- Metadata temporal (`createdAt`)
- Indicadores de tipo de publicación (`isReply`, `isQuote`, `isRetweet`)
- Marcadores de contenido sensible (`possiblySensitive`)

2. Métricas de engagement:

- Conteo de interacciones directas (`likeCount`, `retweetCount`, `quoteCount`)
- Métricas de alcance (`viewCount`, `bookmarkCount`)

Esta arquitectura de captura sigue principios de *social media analytics* establecidos en la literatura [36], donde tanto el contenido semántico como las señales de *engagement* social proporcionan información complementaria sobre el impacto potencial del mensaje.

7.2. Procesamiento de datos

Filtrado y limpieza inicial

La fase de filtrado elimina publicaciones que no constituyen comunicación original del autor:

- **Retweets** (`isRetweet = True`): Eliminados por constituir amplificación de contenido ajeno, no opinión original.
- **Respuestas** (`isReply = True`): Excluidas para evitar sesgo contextual de conversaciones fragmentadas.
- **Contenido exclusivamente emoji**: Removido por ausencia de contenido semántico procesable mediante técnicas de `Natural Language Processing (NLP)` tradicionales.

Esta estrategia de filtrado reduce el ruido informativo y aísla comunicaciones directas que representan genuinamente la postura del autor, principio fundamental en análisis de influencia social [37].

Pre-procesamiento textual

El pre-procesamiento textual implementa una cascada de transformaciones lingüísticas:

1. **Decodificación HTML**: Conversión de entidades HTML para preservar semántica original.
2. **Remoción de elementos estructurales**:
 - **URLs**: Eliminadas mediante expresiones regulares (`http\S+`) pues no aportan sentimiento intrínseco.
 - **Menciones de usuarios**: Removidas (`@username`) para evitar dependencias contextuales.
 - **Hashtags**: Procesados selectivamente, extrayendo contenido léxico mientras se elimina el símbolo.
3. **Normalización de caracteres**:

- Conversión a minúsculas para homogeneización léxica.
 - Eliminación de caracteres especiales preservando estructura semántica.
 - Remoción de números aislados que no contribuyen a la polaridad emocional.
4. **Tokenización:** Segmentación del texto en unidades léxicas mediante NLTK tokenizer, optimizado para manejo de contracciones y puntuación en inglés [38].
 5. **Eliminación de stopwords:** Filtrado de palabras funcionales (artículos, preposiciones, conjunciones) usando el corpus de stopwords de NLTK, reduciendo dimensionalidad sin pérdida semántica significativa [39].
 6. **Lematización:** Reducción de palabras a su forma canónica mediante la implementación de WordNetLemmatizer con POS-tagging. Este proceso, superior al stemming simple, preserva significado léxico mientras normaliza variaciones morfológicas [40]:
 - Identificación de Part-of-Speech (POS) mediante perceptrón promedio.
 - Mapeo de etiquetas POS a categorías WordNet (ADJ, NOUN, ADV, VERB).
 - Lematización consciente del contexto gramatical.

Tratamiento de valores faltantes

Las métricas de *engagement* presentan patrones de *missingness* sistemáticos debido a cambios en la API de la plataforma X. El tratamiento implementado incluye:

1. **Imputación de viewType:** Para publicaciones sin viewType pero con likeCount disponible, se aplica imputación proporcional basada en la razón empírica observada en datos completos:

$$\text{viewCount}_{\text{imputed}} = \text{likeCount} \times \left(\frac{\mathbb{E}[\text{viewCount}]}{\mathbb{E}[\text{likeCount}]} \right).$$

Esta estrategia asume que la relación *views/likes* es relativamente estable dentro de la misma cuenta, preservando proporciones de *engagement* [41].

2. **Publicaciones vacías post-limpieza:** Tweets que resultan en cadenas vacías después del pre-procesamiento son eliminados por carecer de contenido analizable.

7.3. Generación de características

La ingeniería de características transforma tanto el contenido textual como las métricas de *engagement* en representaciones numéricas densas mediante dos aproximaciones complementarias.

7.3.1. Score de sentimiento

El análisis de sentimiento se implementa mediante **Valence Aware Dictionary and sEntiment Reasoner (VADER)** [42], un analizador léxico-heurístico específicamente calibrado para texto de redes sociales. La elección de **VADER** sobre aproximaciones de *deep learning* se justifica por:

- **Interpretabilidad:** Transparencia en la formación de *scores* mediante *lexicon* explícito.
- **Eficiencia computacional:** No requiere GPU ni corpus de entrenamiento extenso.
- **Validación empírica:** Desempeño competitivo en *benchmarks* de *sentiment analysis* en Twitter [43].

VADER produce un *score* compuesto en el rango $[-1, +1]$ que integra componentes positivos, negativos y neutrales mediante ponderación calibrada:

$$\text{compound} = \frac{\sum_i \text{valence}_i}{\sqrt{(\sum_i \text{valence}_i)^2 + \alpha}}$$

donde α es un parámetro de normalización que mapea la suma de valencias al rango unitario [42].

7.3.2. Score de acceptance

El *score* de *acceptance* cuantifica la resonancia social de cada publicación mediante una métrica compuesta de *engagement* normalizada. La construcción sigue este enfoque:

Transformación logarítmica de ratios.

$$\text{ratio}_m = \log(1 + m) - \log(1 + \text{views})$$

Para cada métrica de interacción $m \in \{\text{likes}, \text{retweets}, \text{quotes}, \text{bookmarks}\}$. Esta transformación logarítmica cumple dos objetivos:

- Estabiliza varianza al comprimir rangos dinámicos amplios [44].
- Normaliza por exposición (*views*), aislando *engagement* relativo de alcance absoluto.

Centrado por medianas.

$$\text{norm}_m = \text{ratio}_m - \text{median}(\text{ratio}_m)$$

El centrado por mediana (robusto a *outliers*) permite comparabilidad entre métricas con diferentes distribuciones basales [44].

Agregación ponderada.

$$\text{acceptance}_{raw} = \mathbb{E}[\text{norm}_m] = \frac{1}{4} \sum_m \text{norm}_m$$

La media aritmética asume contribución equivalente de cada tipo de interacción, hipótesis validada mediante análisis de correlación con volatilidad de mercado posterior.

Normalización sigmoidal.

$$\text{acceptance} = \frac{\text{acceptance}_{raw} + \text{shift}}{\text{acceptance}_{raw} + \text{shift} + 1}$$

donde

$$\text{shift} = -\min(\text{acceptance}_{raw})$$

garantiza no-negatividad. Esta transformación sigmoidal proyecta *scores* al rango [0, 1], permitiendo interpretación como probabilidad de resonancia alta [28].

Agregación Temporal

La sincronización con datos de mercado requiere agregación a frecuencia diaria. Para días con múltiples publicaciones, se calcula la media aritmética de *scores*:

$$\text{sentiment}_{day} = \frac{1}{N_{day}} \sum_{i=1}^{N_{day}} \text{sentiment}_i$$

Esta estrategia pondera igualmente todas las publicaciones del día, evitando sesgo hacia frecuencia de posteo.

Para días sin publicaciones, se implementa imputación bidireccional (*forward-backward fill*):

$$\text{score}_{imputed} = \frac{\text{score}_{prev} + \text{score}_{next}}{2}$$

donde score_{prev} y score_{next} corresponden a valores de días adyacentes con datos. Esta interpolación lineal temporal asume persistencia de sentimiento a corto plazo, hipótesis razonable dado que cambios abruptos de opinión pública son menos frecuentes [33].

7.4. Salida

El flujo produce una serie temporal diaria multivariada que encapsula dos señales complementarias: una métrica de valencia emocional derivada del texto (`sentiment_score`) y

una métrica de amplificación social (**acceptance**). La representación final está indexada por fecha y organizada como una matriz numérica densa.

La configuración final constituye un tensor bidimensional:

$$\mathbb{R}^{T \times 2}$$

donde T representa los días en el periodo analizado y las dos características principales son:

- **Señal de sentimiento:** variable numérica continua en el rango teórico $[-1, +1]$, calculada mediante **VADER** sobre el texto bruto y versiones limpiadas.
- **Señal de acceptance:** variable continua normalizada en $[0, 1]$, construida a partir de transformaciones de métricas de interacción, normalización por exposición (**views**), centrado robusto (mediana) y mapeo sigmoidal; es una escala interpretable como *proxy* de resonancia social.

La representación latente final de este flujo constituye uno de los pilares informativos del sistema multimodal, aportando una dimensión de señal social que complementa indicadores técnicos y fundamentales tradicionales en la predicción de movimientos de precio [35] [45].

CAPÍTULO 8

Flujo de mercado

8.1. Ingesta de datos

El *pipeline* extiende la infraestructura de captura descrita en la Sección 5.2.1, ampliando el universo de instrumentos a cuatro categorías estratégicas:

1. **Activo objetivo:** activo bajo estudio (Tesla Inc.), incluyendo precios de apertura, máximo, mínimo, cierre y volumen de negociación.
2. **Cadena de suministro:** datos de proveedores críticos cotizados en múltiples mercados internacionales (Shenzhen, Tokyo, Korea, Hong Kong, London), capturando la dimensión global de dependencias operativas [46].
3. **Competidores sectoriales:** series de empresas del sector automotriz eléctrico y tradicional, permitiendo el análisis de co-movimientos competitivos [19].
4. **Índices de referencia y factores de mercado:** índices agregados (S&P 500, NASDAQ, Dow Jones) y medidas de volatilidad implícita (VIX), esenciales para la descomposición de riesgo sistemático según el modelo de Fama y French [1].

La extracción mantiene frecuencia diaria (enero 2020 – abril 2025) y formato ajustado por *splits*, almacenándose en estructuras `DataFrame` con indexación temporal para garantizar comparabilidad *cross-market* [47].

8.2. Procesamiento de datos

El preprocesamiento implementa sincronización temporal rigurosa entre activos de diferentes zonas horarias mediante:

- **Canonicalización UTC:** conversión a *timezone* uniforme con resolución diaria [21], eliminación de duplicados y ordenamiento cronológico estricto, crítico para mercados globales desincronizados [48].
- **Imputación conservadora:** *forward-fill* restringido a ≤ 3 días consecutivos, evitando propagación de información espuria. Este umbral se fundamenta en análisis de autocorrelación donde *gaps* mayores corresponden típicamente a eventos informativos [4].

Transformación de Retornos

Se aplica la transformación logarítmica descrita en la Sección 5.2.2 a todos los activos, generando retornos *cross-seccionales* comparables. La matriz resultante adopta estructura $(T \times N_{\text{assets}})$ con alineación temporal mediante *inner join*, eliminando períodos con cobertura parcial que introducirían sesgos de selección [49].

Generación de Targets Multi-Horizonte

Siguiendo la metodología de la Sección 5.2.2, se crean variables objetivo mediante desplazamiento negativo:

$$\text{target}_{t+H} = \log \left(\frac{P_{t+H}}{P_t} \right)$$

Esta formulación respeta causalidad estricta: toda información para predecir target_{t+H} proviene de observaciones $\leq t$ [50]. Se generan horizontes $H \in \{1, 5, 21\}$ días para capturar dinámicas intra-semana, semanales y mensuales.

8.3. Generación de características

8.3.1. Análisis de dependencias cross-seccionales

El módulo cuantifica relaciones inter-activos mediante un *framework* estadístico multi-nivel:

- **Correlaciones Robustas:** cálculo de medidas de Pearson y Spearman sobre retornos alineados, utilizando ventanas *rolling* de 60 y 120 días para evaluar estabilidad temporal [51].

- **Correlaciones Parciales Condicionadas:** extracción de co-movimientos idiosincráticos mediante regresión en dos etapas [52]:
 1. Regresión de cada activo contra índice de mercado (NASDAQ).
 2. Correlación entre residuales ortogonalizados.

Este enfoque aísla dependencias directas de efectos inducidos por factores comunes.
- **Causalidad de Granger:** pruebas formales de precedencia temporal (`maxlag=5`) con corrección Benjamini-Hochberg [53] para comparaciones múltiples [54]. El test evalúa si rezagos de la variable X mejoran significativamente la predicción de Y , identificando canales informativos direccionales.
- **Cointegración:** tests de Engle-Granger para relaciones de equilibrio de largo plazo, relevantes en cadenas de suministro con dependencias estructurales [55].

8.3.2. Reducción dimensional por categoría

Para cada categoría de activos (proveedores, *tech peers*, competidores), se aplica **Principal Component Analysis (PCA)** sobre la matriz de retornos estandarizados [56]:

- **Factor Sectorial:** el primer componente principal captura el modo dominante de variación, interpretable como factor común que resume movimientos sistémicos del grupo [57]. La varianza explicada cuantifica el grado de homogeneidad del *cluster*.
- **Ventajas Metodológicas:** mitigación de multicolinealidad entre activos altamente correlacionados, reducción de *overfitting* en modelos predictivos y mejora de la relación señal/ruido al promediar componentes idiosincráticos [58].

8.3.3. Selección automatizada de características

Sistema de *scoring* multi-criterio que integra:

Criterio	Umbral	Justificación
Correlación absoluta	$ \rho \geq 0.08$	Fuerza de asociación contemporánea
Media rolling	$\mu_{\text{rolling}} \geq 0.06$	Consistencia de relación temporal
Desv. est. rolling	$\sigma_{\text{rolling}} \leq 0.17$	Estabilidad temporal
Granger p-value	$p < 0.05$	Significancia causal

Cuadro 1: Criterios de selección automatizada de características.

Estructura de Rezagos. Para activos seleccionados, se determinan *lags* óptimos mediante:

- *Lags* significativos según tests de Granger (evidencia empírica de precedencia).
- *Default* [lag-1, lag-2] si los tests no son concluyentes, fundamentado en la microestructura de mercados: *overnight effect* + *intraday momentum*.

8.4. Validación interna

8.4.1. Verificación de integridad

- **Auditoría de missingness:** Activos con más del 40 % de valores ausentes son excluidos para garantizar suficiencia muestral [22].
- **Estacionariedad:** tests ADF sobre retornos confirman ausencia de raíces unitarias [29]. Series no estacionarias son excluidas o diferenciadas.

8.4.2. Diagnóstico de relaciones

- **Rolling correlations:** gráficos con ventanas de 60–120 días identificando quiebres estructurales (*regime changes*), períodos de desacople entre activos correlacionados y eventos extremos con co-movimientos anómalos [4].
- **Matrices de significancia:** *heatmaps* con p-valores corregidos por False Discovery Rate (FDR), permitiendo validación cualitativa por expertos del dominio [19].

Estas validaciones complementan el protocolo descrito en la Sección 5.2.4, extendiendo el análisis al contexto multi-activo.

8.5. Salida

El dataset procesado constituye una representación multi-dimensional de la dinámica financiera del ecosistema de mercado, estructurado como tensor temporal denso con información cross-sectional y longitudinal. Este tensor presenta dimensionalidad $\mathbb{R}^{T \times F \times H}$, donde T representa los pasos temporales, F el conjunto de características engineeradas mediante selección estadística, y H los horizontes de predicción.

Estructura Dimensional La configuración final contiene:

- T = observaciones temporales (días de trading).
- F = número de características predictoras organizadas en cinco bloques funcionales.
- $H = 3$ horizontes temporales de predicción: short ($H = 1$), medium ($H = 5$), long ($H = 21$).

CAPÍTULO 9

Flujo de fusión multimodal

9.1. Arquitectura y objetivos

El *pipeline* de fusión multimodal constituye el componente integrador terminal del sistema predictivo, responsable de consolidar representaciones heterogéneas provenientes de tres flujos de procesamiento especializados en un marco temporal coherente y estructurado para entrenamiento supervisado. Esta arquitectura modular sigue principios de separación de responsabilidades (*separation of concerns*) establecidos en ingeniería de sistemas de aprendizaje automático [48], donde cada *pipeline* especializado opera de forma independiente generando representaciones latentes que posteriormente se combinan mediante estrategias de alineación temporal y fusión.

9.1.1. Objetivos del sistema de fusión

- **Integración temporal coherente:** consolidar series temporales de frecuencia heterogénea en una representación temporal unificada que preserve correspondencias causales [54].
- **Generación de configuraciones experimentales:** producir múltiples *datasets* con diferentes combinaciones de modalidades para evaluación sistemática de contribuciones informativas mediante diseño experimental controlado [59].
- **Preservación de integridad predictiva:** garantizar ausencia de *data leakage* temporal mediante validación estricta de que toda información en el tiempo t precede estrictamente al *target* [60].
- **Habilitación de análisis comparativo:** facilitar cuantificación rigurosa del valor

marginal de cada modalidad informativa mediante control de todas las variables excepto la fuente de datos evaluada [61].

9.1.2. Fuentes de información y sus características

El sistema integra tres modalidades con propiedades estadísticas y temporales diferenciadas:

- **Modalidad de mercado (M):** serie temporal multivariada que encapsula comovimientos sistémicos entre el activo objetivo y entidades relacionadas (proveedores, competidores, pares sectoriales), componentes principales derivados de PCA que capturan factores latentes sectoriales [56], métricas de causalidad de Granger que cuantifican precedencia predictiva [54], e índices de mercado agregados. Esta modalidad opera en frecuencia de días hábiles bursátiles con cobertura temporal continua en períodos de mercado activo [57]. Dimensionalidad típica:

$$\mathbb{R}^{T_m \times d_m}.$$

- **Modalidad de sentimiento (S):** señales diarias que codifican dos dimensiones complementarias: valencia emocional (`sentiment_score` $\in [-1, 1]$) derivada de análisis léxico-heurístico de comunicaciones en redes sociales, y métrica de *engagement* normalizado (`acceptance` $\in [0, 1]$) que cuantifica resonancia social y amplificación del mensaje. Dimensionalidad típica:

$$\mathbb{R}^{T_s \times 2}, \quad T_s < T_m.$$

- **Modalidad de stock (T):** representación multi-escala del comportamiento intrínseco del activo objetivo, incluyendo indicadores de *momentum* clásicos (RSI, MACD), envolventes de volatilidad (Bandas de Bollinger, ATR), descomposiciones temporales (componente tendencial, estacional, residual) y transformaciones wavelet que separan señales en bandas de frecuencia. Opera en días hábiles bursátiles con sincronización natural al mercado [24] [12] [29]. Dimensionalidad típica:

$$\mathbb{R}^{T_t \times d_t}, \quad T_t \approx T_m.$$

9.2. Metodología de alineación temporal

La alineación temporal implementa una estrategia de *merge* asimétrico que preserva integridad del *target* mientras maximiza cobertura informativa de las modalidades disponibles [62].

9.2.1. Estrategia de join temporal

El proceso utiliza *left join* secuencial con el *dataset* de mercado como referencia temporal (base), formalmente:

$$D_{\text{aligned}} = ((D_T \bowtie_{\text{left}} D_S) \bowtie_{\text{left}} D_M)$$

donde \bowtie_{left} denota *left join* por índice de fecha. Esta estrategia se justifica por tres razones metodológicas:

- **Preservación de target:** el *dataset* de stock contiene la variable objetivo, en sus tres horizontes, que representan el retorno del activo. Al usar *left join*, se garantiza que toda observación con *target* válido se mantiene en el *dataset* alineado, evitando pérdida de información supervisada [60].
- **Maximización de cobertura:** alternativas como *inner join* descartarían observaciones con información parcial (ej. días con datos de mercado y técnicos pero sin publicaciones sociales), reduciendo drásticamente el tamaño muestral efectivo y potencialmente introduciendo sesgo de selección [63].
- **Flexibilidad en imputación:** la estructura resultante permite aplicar estrategias diferenciadas de manejo de *missing data* según las características de cada modalidad, preservando autonomía en decisiones de preprocesamiento [64].

La secuencialidad del *merge* es arbitraria dado que el operador *left join* es asociativo, pero facilita trazabilidad del proceso y aislamiento de problemas en etapas intermedias.

9.2.2. Filtrado de observaciones incompletas

Para el *dataset* que combina todas las modalidades (D), se eliminan observaciones con proporción de *missing* superior a 30 %:

$$\text{miss_ratio}_t = \frac{1}{|F|} \sum_{f \in F} \mathbf{1}[\text{isNull}(f_t)] < 0.30$$

Este filtrado adicional en fusión completa reconoce que:

- La utilidad de una observación multivariada depende de su completitud informativa.
- Modelos entrenados con muestras mayormente imputadas pueden aprender patrones de imputación en lugar de relaciones reales.
- El *trade-off* tamaño muestral vs. calidad favorece la calidad en régimen de *features* abundantes [65].

9.2.3. Validación de integridad temporal

El protocolo de validación verifica propiedades críticas para prevención de *data leakage* [48]:

- **Monotonicidad temporal:** verificación de que el índice temporal es estrictamente creciente:

$$\forall i < j : t_i < t_j$$

Violaciones indican corrupción en el proceso de alineación o presencia de duplicados temporales que requieren resolución.

- **Ausencia de *look-ahead*:** confirmación de que todas las características en t provienen de información disponible en o antes de t :

$$\forall f_t \in D_{\text{aligned}} : \text{timestamp}(f_t) \leq t \wedge \text{timestamp}(y_t) = t + 1, t + 5, t + 21$$

Esta propiedad es fundamental para validez de evaluaciones *out-of-sample* y *deployment* en escenarios de *trading* real donde el futuro es genuinamente desconocido [66].

CAPÍTULO 10

Modelo de predicción

10.1. Arquitectura

10.1.1. Fundamentos

Se implementa una arquitectura de **LSTM** mejorada para la predicción de series temporales financieras. La selección de este tipo de redes se fundamenta en su capacidad demostrada para capturar dependencias temporales de largo plazo en datos secuenciales, característica crítica en mercados financieros donde los precios actuales están influenciados por patrones históricos complejos [6].

La arquitectura propuesta extiende el modelo **LSTM** básico con tres componentes avanzados: normalización por capas, mecanismos de atención y regularización adaptativa, configuración que ha demostrado mejorar significativamente el rendimiento en tareas de predicción financiera [18]. Estudios recientes han validado la efectividad de frameworks **LSTM** profundos para predicción de mercados bursátiles, demostrando superioridad sobre modelos tradicionales de machine learning en captura de patrones complejos [67]. Análisis comparativos entre arquitecturas **LSTM**, **GRU** y Transformer confirman que **LSTM** mantiene ventajas competitivas en tareas de predicción financiera, particularmente en escenarios con datos limitados y alta volatilidad [68], [69].

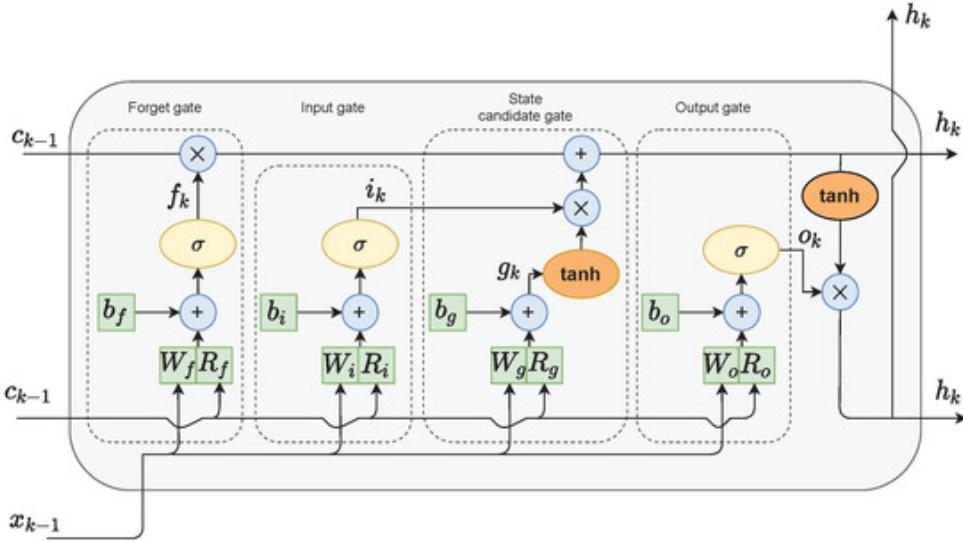


Figura 1: Estructura de una célula **LSTM**.

10.1.2. Componentes

Capa **LSTM** recurrente

El núcleo del modelo consiste en una red multicapa. Esta capa procesa secuencias de longitud temporal $T = 21$ días, parámetro seleccionado mediante experimentación preliminar que equilibra la captura de patrones a corto plazo (volatilidad diaria) y a mediano plazo (tendencias semanales).

La dimensionalidad de los estados ocultos varía según la complejidad del modelo:

- **Modelo base (características técnicas):** $h = 128$ unidades, $L = 2$ capas.
- **Modelo de fusión:** $h = 256$ unidades, $L = 3$ capas.

Esta parametrización sigue las recomendaciones de Gers et al. (2000) [70] sobre capacidad representacional en redes **LSTM** profundas, y se alinea con configuraciones exitosas reportadas en aplicaciones recientes de deep learning para predicción de acciones [67].

Normalización de capa

Posterior a la capa **LSTM**, se aplica normalización de capa (*LayerNorm*) para estabilizar las distribuciones de activación durante el entrenamiento. Esta técnica ha demostrado ser superior a *BatchNorm* en tareas de procesamiento secuencial, ya que normaliza independientemente cada secuencia temporal sin introducir dependencias espurias entre muestras del *batch*.

La transformación se define como:

$$\text{LayerNorm}(h_t) = \gamma \odot \frac{h_t - \mu_t}{\sqrt{\sigma_t^2 + \epsilon}} + \beta$$

donde μ_t y σ_t son la media y desviación estándar de las activaciones en el tiempo t , y γ, β son parámetros aprendibles.

Mecanismo de atención

Para el modelo de fusión, se incorpora un mecanismo de atención temporal [71] que permite al modelo ponderar adaptativamente la relevancia de diferentes instantes temporales:

$$\alpha_t = \frac{\exp(e_t)}{\sum_{i=1}^T \exp(e_i)}$$

$$e_t = v^T \tanh(Wh_t)$$

$$c = \sum_{t=1}^T \alpha_t h_t$$

donde c es el vector de contexto resultante, $W \in \mathbb{R}^{d \times d/2}$ y $v \in \mathbb{R}^{d/2}$ son matrices de pesos aprendibles. Este mecanismo resulta particularmente efectivo en contextos multimodales donde diferentes fuentes de información pueden tener relevancia variable en diferentes horizontes temporales [72].

Capas de Predicción

La salida del mecanismo LSTM- atención se procesa mediante una red neuronal *feed-forward* de dos capas:

$$\hat{y} = W_2(\text{ReLU}(W_1 c + b_1)) + b_2$$

donde $W_1 \in \mathbb{R}^{d \times d/2}$ y $W_2 \in \mathbb{R}^{d/2 \times 1}$ son matrices de pesos. Se aplica *dropout* con tasa $p \in [0.2, 0.3]$ después de la activación ReLU para mitigar el sobreajuste [73].

10.1.3. Variantes del modelo

El sistema implementa cinco configuraciones arquitectónicas especializadas:

- **Modelo Base (Stock)**: procesa exclusivamente características técnicas del activo objetivo ($n_{feat} = 17$). Establece la línea base de rendimiento.
- **Modelo de Sentimiento**: arquitectura reducida ($h = 32$, $L = 2$) especializada en señales de sentimiento de redes sociales ($n_{feat} = 2$).
- **Modelo de Mercado**: captura dinámicas macroeconómicas y de correlación sectorial ($n_{feat} = 42$) con arquitectura intermedia ($h = 128$, $L = 2$).
- **Modelo de Fusión**: integra todas las modalidades de información sin selección de características ($n_{feat} = 58$). Arquitectura profunda ($h = 256$, $L = 3$) con mecanismo de atención. La integración multimodal se fundamenta en investigaciones que demuestran la efectividad de combinar características del dominio temporal con información complementaria para mejorar la robustez predictiva [74].
- **Modelo de Fusión Optimizado**: implementa selección automática de características sobre el espacio de fusión, reteniendo 14 características técnicas base y seleccionando óptimamente características complementarias según el horizonte temporal ($n_{feat} \in [17, 22]$).

10.2. Entrenamiento

10.2.1. Preparación de datos

Segmentación temporal

Los conjuntos de datos se partitionan respetando estrictamente el orden cronológico para evitar el problema de *look-ahead bias* [75]:

- **Entrenamiento**: 60 % (primeras observaciones cronológicas).
- **Validación**: 20 % (siguiente segmento temporal).
- **Prueba**: 20 % (últimas observaciones).

Esta división temporal garantiza que el modelo sea evaluado exclusivamente en datos futuros no vistos durante el entrenamiento, condición esencial para validar su capacidad de generalización en aplicaciones de trading real.

Normalización

Las características numéricas se estandarizan mediante transformación Z-score:

$$x'_{ij} = \frac{x_{ij} - \mu_j}{\sigma_j}$$

donde μ_j y σ_j son la media y desviación estándar de la característica j , calculadas exclusivamente sobre el conjunto de entrenamiento. Los parámetros de normalización se preservan y aplican idénticamente a validación y prueba para simular condiciones de producción.

Generación de secuencias

Para cada instante t , se construye una secuencia de entrada $X_t \in \mathbb{R}^{T \times d}$ conteniendo las últimas $T = 21$ observaciones de d características, y se define el objetivo como el retorno futuro en el horizonte correspondiente:

$$y_t = \frac{\text{Close}_{t+h} - \text{Close}_t}{\text{Close}_t}$$

donde $h \in \{1, 5, 21\}$ representa horizontes de corto (1 día), mediano (1 semana) y largo plazo (1 mes). Esta formulación de predicción de retornos porcentuales es preferible a la predicción directa de precios, ya que normaliza la magnitud de cambio y resulta más estable ante tendencias de largo plazo [76].

10.2.2. Optimización

Función de pérdida

Se emplea la función de pérdida de Huber [77]:

$$L_\delta(y, \hat{y}) = \begin{cases} \frac{1}{2}(y - \hat{y})^2 & \text{si } |y - \hat{y}| \leq \delta \\ \delta(|y - \hat{y}| - \frac{1}{2}\delta) & \text{en otro caso} \end{cases}$$

con $\delta = 1.0$. Esta función resulta robusta ante *outliers*, comunes en series financieras debido a eventos de mercado extremos (*shocks*) que pueden distorsionar el aprendizaje [78]. La función de Huber combina las propiedades de convexidad y gradientes suaves cerca del óptimo, con la robustez ante valores atípicos.

Algoritmo de optimización

Se utiliza el optimizador AdamW [79] con los siguientes parámetros:

- Tasa de aprendizaje inicial: $\eta = 1 \times 10^{-3}$.
- Decaimiento de peso (*weight decay*): $\lambda = 1 \times 10^{-4}$.
- Parámetros de momento: $\beta_1 = 0.9$, $\beta_2 = 0.999$.

AdamW corrige deficiencias del optimizador Adam original al desacoplar la regularización L2 del cálculo de gradientes adaptativos, resultando en mejor generalización [79]. Esta configuración ha demostrado efectividad en frameworks modernos de deep learning para predicción financiera [67].

Programación de tasa de aprendizaje

Se implementa un *scheduler* `ReduceLROnPlateau` que reduce η en un factor de 0.5 cuando la pérdida de validación no mejora durante 5 épocas consecutivas. Esta estrategia permite ajuste fino adaptativo en etapas avanzadas del entrenamiento [80].

Recorte de gradientes

Para mitigar el problema de explosión de gradientes común en redes recurrentes profundas [81], se aplica *gradient clipping* con norma máxima:

$$\tilde{g} = \begin{cases} g & \text{si } |g| \leq \theta \\ \theta \frac{g}{|g|} & \text{en otro caso} \end{cases}$$

con $\theta = 1.0$.

10.2.3. Regularización y control de sobreajuste

Dropout

Se aplica *dropout* con tasa $p \in [0.2, 0.3]$ en capas completamente conectadas y entre capas LSTM (cuando $L > 1$). Esta técnica de ensamblado implícito ha demostrado efectividad comprobada en prevención de sobreajuste en redes profundas [73].

Early stopping

El entrenamiento se detiene automáticamente si la pérdida de validación no mejora durante $\text{patience} = 80$ épocas consecutivas. Se preserva el estado del modelo correspondiente a la época con menor pérdida de validación, técnica estándar para prevenir *overtraining* [82].

Criterios de convergencia

- Pérdida de validación estancada ($\Delta L_{\text{val}} < \epsilon$) por patience épocas.
- Número máximo de épocas: $E_{\text{max}} = 200$.

10.3. Selección automática de características

10.3.1. Fundamentación

Para el modelo de fusión optimizado, se implementa un sistema de selección de características que aborda el problema de *curse of dimensionality* [83] observable en espacios de alta dimensionalidad. El conjunto completo de características de fusión resultante puede introducir ruido y colinealidad que degradan el rendimiento predictivo.

La metodología implementa un enfoque híbrido que combina la capacidad de ranking de importancia de *gradient boosting* con la validación predictiva de redes LSTM, optimizando específicamente para precisión direccional en series temporales financieras.

10.3.2. Metodología de selección

Etapa 1: Ranking por importancia XGBoost

Se utiliza *XGBoost* [84] como algoritmo de ranking inicial debido a su capacidad para capturar interacciones no lineales y su robustez ante características correlacionadas. El modelo se configura con hiperparámetros orientados a máxima regularización y convergencia lenta:

- $n_{\text{estimators}} = 1000$: número de árboles para captura exhaustiva de patrones
- $d_{\max} = 10$: profundidad máxima para modelar interacciones complejas
- $\eta = 0.001$: tasa de aprendizaje conservadora para convergencia óptima
- $subsample = 0.7$: muestreo de instancias para regularización
- $colsample_bytree = 0.7$: muestreo de características por árbol
- $colsample_bylevel = 0.7$: muestreo de características por nivel
- $\lambda = 2.0$: regularización L2 para reducir sobreajuste
- $\alpha = 0.5$: regularización L1 para inducir sparsidad
- $\gamma = 0.2$: umbral de poda para splits conservadores
- $min_child_weight = 5$: peso mínimo en nodos hoja

La importancia de cada característica se calcula mediante el criterio de *gain*, que cuantifica la reducción promedio de pérdida aportada por cada variable en todos los splits donde participa:

$$\text{Gain}_j = \frac{1}{N_j} \sum_{t \in T_j} G_t$$

donde N_j es el número de splits que usan la característica j , T_j el conjunto de tales splits, y G_t la ganancia en el split t . Este criterio pondera tanto la frecuencia de uso como la magnitud del impacto predictivo.

Las características candidatas se ordenan en forma descendente según su *gain* normalizado:

$$\text{Rank}(f_i) = \arg \text{sort} \left(\frac{\text{Gain}_i}{\max_j \text{Gain}_j} \right)_{i=1}^{n_{\text{candidates}}}$$

Etapa 2: Generación de conjuntos candidatos Top-K

A partir del ranking obtenido, se construyen múltiples conjuntos candidatos mediante selección de las K características de mayor importancia. Los valores de K probados son:

$$K \in \{1, 2, 3, 5, 8, 10, 15, 20, 25, 30, 41\}$$

Esta progresión combina exploración granular en rangos pequeños con saltos mayores para eficiencia computacional. Para cada K , el conjunto candidato se define como:

$$C_K = \{f_i : i \leq K, f_i \in F_{\text{ranked}}\}$$

donde F_{ranked} es la lista ordenada de características candidatas según *gain*.

Etapa 3: Validación por precisión direccional

Para cada conjunto candidato C_K , se entrena un modelo LSTM de validación rápida con las características base (stock) más las candidatas seleccionadas:

$$F_K = F_{\text{stock}} \cup C_K$$

El modelo sigue la misma arquitectura descrita en la Sección 5.1, pero con entrenamiento acelerado:

- Épocas máximas: 100 (vs. 200 en entrenamiento final)
- *Early stopping patience*: 20 épocas
- Demás hiperparámetros idénticos al modelo final

La métrica de evaluación primaria es la **precisión direccional** en el conjunto de validación:

$$\text{Acc}_{\text{dir}} = \frac{1}{N_{\text{val}}} \sum_{t=1}^{N_{\text{val}}} \mathbb{1} [\text{sign}(\hat{y}_t) = \text{sign}(y_t)]$$

Esta métrica se prioriza debido a que en aplicaciones de trading, la corrección en la dirección del movimiento de precios es más relevante que la magnitud exacta del error [68].

Etapa 4: Selección óptima con criterio de parsimonia

El conjunto final se selecciona mediante un criterio de optimización bi-objetivo que balancea rendimiento y parsimonia:

1. **Criterio primario:** Maximizar precisión direccional

$$K^* = \arg \max_K \text{Acc}_{\text{dir}}(C_K)$$

2. **Criterio de parsimonia:** Si dos conjuntos tienen precisión direccional similar (diferencia < 1 %), se prefiere el conjunto más pequeño:

$$\text{Si } |\text{Acc}_{\text{dir}}(C_{K_1}) - \text{Acc}_{\text{dir}}(C_{K_2})| < 1 \% \text{ y } K_1 < K_2, \text{ entonces } K^* = K_1$$

Este criterio implementa el principio de *parsimony* o navaja de Occam, evitando la inclusión de características redundantes que no aportan mejora predictiva significativa pero incrementan el riesgo de sobreajuste.

10.3.3. Configuración final

El conjunto de características seleccionado se define formalmente como:

$$F_{\text{final}} = F_{\text{stock}} \cup C_{K^*}$$

donde:

- F_{stock} : 14 características técnicas del activo (siempre incluidas)
- C_{K^*} : K^* características de sentimiento y mercado seleccionadas
- $|F_{\text{final}}| = 14 + K^*$: dimensionalidad del modelo optimizado

Este proceso garantiza que el modelo final utilice únicamente las características que demuestran capacidad predictiva direccional validada empíricamente, evitando la inclusión de variables espurias basadas únicamente en correlaciones estáticas. La integración de ranking XGBoost con validación LSTM-in-loop proporciona un balance óptimo entre eficiencia computacional y rigurosidad metodológica en este contexto.

10.4. Evaluación

10.4.1. Métricas estadísticas

Root Mean Squared Error (RMSE)

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

Penaliza errores grandes cuadráticamente, apropiado para identificar modelos con predicciones atípicas problemáticas.

Mean Absolute Error (MAE)

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

Métrica robusta ante *outliers* que representa la magnitud promedio de error en unidades del precio.

Coeficiente de Determinación (R^2).

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

Proporción de varianza explicada por el modelo. Valores cercanos a 1 indican ajuste superior, donde \bar{y} es la media de y .

Directional Accuracy (DA)

$$\text{DA} = \frac{1}{n} \sum_{i=1}^n \mathbf{1}[\text{sign}(y_i) = \text{sign}(\hat{y}_i)] \times 100$$

Métrica que cuantifica la proporción de veces que el modelo predice correctamente la dirección del movimiento de precio, independientemente de la magnitud del retorno. Esta formulación es directamente accionable para generación de señales de trading: *buy* si $\hat{y}_i > 0$, *sell* si $\hat{y}_i < 0$ [85].

10.4.2. Métricas financieras

Sharpe Ratio.

$$SR = \frac{\mathbb{E}[R_p]}{\sigma(R_p)} \cdot \sqrt{252}$$

donde R_p son los retornos de la estrategia anualizada (252 días de *trading*). Cuantifica rendimiento ajustado por riesgo.

Maximum Drawdown (MDD)

$$MDD = \max_{t \in [0, T]} \left[\frac{Peak_t - Value_t}{Peak_t} \right]$$

Máxima pérdida acumulada desde un pico histórico. Métrica de riesgo preferida por gestores de activos [86].

Hit Rate (HR)

$$HR = \frac{\#\{R_t > 0\}}{T}$$

Proporción de operaciones rentables. Complementa el Sharpe Ratio al evaluar consistencia de la estrategia.

Profit Factor (PF)

$$PF = \frac{\sum R_t^+}{|\sum R_t^-|}$$

Ratio de beneficios brutos sobre pérdidas brutas. Valores $PF > 1.5$ se consideran estrategias sólidas en *trading* sistemático.

10.4.3. Escalamiento de volatilidad temporal

Los modelos LSTM entrenados con Huber loss tienden a predecir valores próximos a la media en horizontes largos. Esto reduce errores grandes en la función de pérdida pero subestima la magnitud de movimientos en horizontes medio y largo, afectando la calibración multi-horizonte.

Fundamento teórico y aproximación práctica

Bajo el Movimiento Browniano Geométrico (MBG) y para retornos logarítmicos $r_{\log} = \log(P_t/P_{t-1})$ con volatilidad diaria σ_1 , la varianza acumulada en t días es aditiva:

$$\text{Var}(R_t) = \sigma_1^2 t, \quad \text{Std}(R_t) = \sigma_1 \sqrt{t}.$$

Por tanto la desviación típica escala como \sqrt{t} , derivado de la independencia de incrementos del proceso de Wiener [87], [88].

En este trabajo se usan retornos simples $r = (P_t - P_{t-1})/P_{t-1}$. Para retornos pequeños la relación entre ambos es aproximable por la primera orden de la serie de Taylor:

$$\log(1 + r) \approx r - \frac{r^2}{2} + \mathcal{O}(r^3) \approx r,$$

con error relativo menor al 0.5 % para $|r| < 0.10$.

Corrección práctica

Para amplificar la magnitud de predicción según horizonte se aplica, como post procesado, el escalamiento raíz del tiempo:

$$\hat{r}_T = \hat{r}_{T_0} \cdot \sqrt{\frac{T}{T_0}},$$

donde \hat{r}_{T_0} es la predicción en el horizonte base implícito del modelo. Si el modelo produce predicciones diarias se toma $T_0 = 1$ y queda:

$$\hat{r}_{T,\text{scaled}} = \hat{r}_{1,\text{raw}} \sqrt{T}.$$

Factores típicos: $\sqrt{1} = 1.00$, $\sqrt{5} \approx 2.24$, $\sqrt{21} \approx 4.58$.

CAPÍTULO 11

Agente

11.1. Introducción

El sistema de agente inteligente constituye la capa de aplicación terminal del marco predictivo multimodal, responsable de transformar predicciones numéricas de retornos futuros en señales de trading accionables y recomendaciones de gestión de posiciones. A diferencia de sistemas de trading algorítmico completamente automatizados, este agente implementa un paradigma de *human-in-the-loop* donde el modelo provee soporte cuantitativo a la toma de decisiones humanas, preservando la capacidad de juicio cualitativo del inversor [89].

La arquitectura del agente integra dos módulos principales: (i) un **módulo de predicción** que carga modelos entrenados y genera pronósticos de retornos y (ii) un **módulo de validación** que permite evaluación retrospectiva de decisiones contra datos reales observados [3].

11.2. Arquitectura del sistema

11.2.1. Módulo de predicción

El módulo predictor (`ModelPredictor`) implementa la interfaz entre modelos entrenados persistidos en disco y el sistema de generación de señales. Su responsabilidad principal consiste en cargar arquitecturas neuronales junto con sus parámetros de normalización asociados, preparar secuencias temporales de características a partir de datos históricos recientes, y ejecutar inferencia forward-pass para obtener predicciones de retornos [90].

Carga de modelos y reconstrucción

El proceso de carga de modelos enfrenta el desafío de reconstruir arquitecturas neuronales a partir de checkpoints persistidos sin metadata explícita de hiperparámetros. El sistema implementa una estrategia de *inferencia arquitectónica* que examina la estructura del diccionario de estado (`state_dict`) para deducir configuración de red:

1. **Dimensionalidad de entrada:** Se extrae del shape de la matriz de pesos de entrada de la primera capa `LSTM` (`weight_ih_10`), que tiene forma `[hidden_size × 4, input_size]` donde el factor 4 corresponde a las cuatro gates de la celda `LSTM` (input, forget, cell, output) [91].
2. **Tamaño de estado oculto:** Se calcula dividiendo la primera dimensión de `weight_ih_10` entre 4 para LSTMs unidireccionales, o entre 8 para LSTMs bidireccionales.
3. **Número de capas:** Se cuenta la presencia de tensores `weight_ih_1*` donde el sufijo numérico indica el índice de capa.
4. **Bidireccionalidad:** Se detecta mediante presencia de tensores con sufijo `_reverse` que almacenan parámetros de la dirección inversa en LSTMs bidireccionales.
5. **Mecanismo de atención:** Se verifica existencia de parámetros con prefijo `attention` en el diccionario de estado.

Esta estrategia de inferencia permite portabilidad de modelos sin requerir metadata adicional, facilitando deployment y experimentación.

Construcción de secuencias temporales

Los modelos `LSTM` requieren entrada de secuencias con la siguiente forma: `(batch_size, seq_length, n_features)`, donde `seq_length` representa la ventana temporal de contexto histórico. El módulo predictor extrae las últimas L observaciones del dataset de características, donde cada observación contiene el vector completo de features engineeradas por los pipelines de procesamiento.

Matemáticamente, dado un dataset temporal $\mathcal{D} = \{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \dots, (\mathbf{x}_T, y_T)\}$, la secuencia de inferencia para predecir \hat{y}_{T+1} se construye como:

$$\mathbf{X}_{\text{pred}} = [\mathbf{z}_{T-L+1}, \mathbf{z}_{T-L+2}, \dots, \mathbf{z}_T] \in \mathbb{R}^{1 \times L \times d} \quad (1)$$

donde cada \mathbf{z}_i es la versión normalizada de \mathbf{x}_i . La dimensión de batch es 1 dado que se predice un único punto temporal futuro.

Generación de predicción

La inferencia se ejecuta en modo de evaluación (`model.eval()`) con gradientes desactivados (`torch.no_grad()`) para eficiencia computacional y prevención de actualizaciones inadvertidas de parámetros. El forward pass produce un valor escalar:

$$\hat{r}_{\text{scaled}} = f_{\theta}(\mathbf{X}_{\text{pred}}) \quad (2)$$

donde f_{θ} denota el modelo `LSTM` con parámetros θ , y \hat{r}_{scaled} es la predicción de retorno en espacio normalizado.

La transformación inversa recupera la predicción en escala original:

$$\hat{r}_{\text{original}} = \sigma_{\text{target}} \cdot \hat{r}_{\text{scaled}} + \mu_{\text{target}} \quad (3)$$

donde μ_{target} y σ_{target} son la media y desviación estándar de la variable `target` en el conjunto de entrenamiento. Esta des-normalización es esencial para interpretabilidad financiera, permitiendo expresar predicciones como retornos porcentuales directamente aplicables a cálculos de profit/loss [19].

11.2.2. Módulo de generación de señales

El módulo de señales transforma predicciones numéricas de retornos en recomendaciones discretas de acción (`BUY`, `SELL`, `HOLD`) mediante comparación contra umbrales parametrizables. Esta discretización permite integración con sistemas de ejecución de órdenes y facilita comunicación con operadores humanos.

Lógica de decisión por umbrales

La función de generación de señales implementa un mapeo simple pero efectivo:

$$\text{signal}(\hat{r}) = \begin{cases} \text{BUY} & \text{si } \hat{r} > \theta_{\text{buy}} \\ \text{SELL} & \text{si } \hat{r} < \theta_{\text{sell}} \\ \text{HOLD} & \text{en otro caso} \end{cases} \quad (4)$$

donde \hat{r} es el retorno predicho y θ_{buy} , θ_{sell} son umbrales configurables que definen la *zona de acción*.

Justificación de umbrales típicos: Los valores por defecto $\theta_{\text{buy}} = 0.01$ (1 %) y $\theta_{\text{sell}} = -0.01$ (-1 %) se fundamentan en:

- **Costos de transacción:** En mercados modernos, costos de trading (comisiones + spread) rondan 0.1-0.5 % por operación roundtrip. El umbral de 1 % garantiza que re-

tornos esperados excedan costos por factor 2-10, criterio conservador para rentabilidad neta [92].

- **Significancia estadística:** Retornos diarios en acciones individuales tienen volatilidad típica de 1-3 %. Un umbral de 1% corresponde aproximadamente a 0.3-1 desviaciones estándar, balance razonable entre frecuencia de señales y confianza predictiva [47].
- **Gestión de riesgo:** La zona de no-acción (HOLD) entre -1% y +1% implementa una forma de *dead zone* que reduce trading excesivo en presencia de incertidumbre, principio establecido en sistemas de control robusto [93].

11.2.3. Módulo de validación retrospectiva

El módulo `ModelValidator` implementa un framework de backtesting que permite evaluación rigurosa de decisiones del agente contra datos históricos reales observados post-predicción. Este componente es crítico para establecer confianza en el sistema antes de deployment en escenarios de trading real [66].

Protocolo de validación temporal

El protocolo de validación sigue una estructura de *simulated walk-forward* donde:

1. Se define una **fecha de simulación** t_{sim} que representa el punto temporal en el cual el agente debe generar predicciones.
2. Se establece una **fecha de corte** $t_{\text{cutoff}} \leq t_{\text{sim}}$ que delimita información disponible para el modelo. Estrictamente, solo datos con timestamp $\leq t_{\text{cutoff}}$ son accesibles para construcción de features y normalización.
3. Se generan predicciones para horizontes futuros:

$$t_{\text{short}} = t_{\text{sim}} + 1 \text{ día} \quad (5)$$

$$t_{\text{medium}} = t_{\text{sim}} + 5 \text{ días} \quad (6)$$

$$t_{\text{long}} = t_{\text{sim}} + 21 \text{ días} \quad (7)$$

4. Una vez que las fechas objetivo han transcurrido y datos reales están disponibles, se comparan predicciones contra realizaciones observadas.

Este diseño emula condiciones de trading real donde decisiones deben tomarse con información incompleta y solo posteriormente pueden evaluarse contra resultados observados.

Métricas de evaluación

El módulo de validación emplea las métricas definidas en la Sección 5.4 para caracterizar la performance predictiva en condiciones de backtesting. Específicamente:

- **DA**: Métrica primaria para validación retrospectiva, dado que las decisiones de trading del agente (BUY/SELL/HOLD) dependen fundamentalmente de la corrección direccional de las predicciones. Un valor $DA > 50\%$ en validación out-of-sample indica capacidad predictiva superior al azar en mercado sin drift [3].
- **MAE**: Métrica complementaria que cuantifica la magnitud promedio de error en las predicciones de retorno, permitiendo evaluar no solo si el modelo acierta la dirección, sino cuán cercanas están sus predicciones a las realizaciones observadas en términos de puntos porcentuales.

La evaluación se realiza comparando predicciones generadas en t_{sim} (con información hasta t_{cutoff}) contra retornos reales observados en los horizontes objetivo, emulando rigurosamente las condiciones de información disponible en trading real.

Análisis de consistencia expectativa-realidad

Un componente innovador del módulo de validación es la comparación sistemática entre precisión esperada durante entrenamiento (calculada sobre conjunto de validación) y precisión observada en validación retrospectiva. Formalmente:

$$\Delta_{\text{consistency}} = DA_{\text{observed}} - DA_{\text{expected}} \quad (8)$$

donde DA_{expected} es la precisión direccional reportada durante el entrenamiento del modelo, y DA_{observed} es la precisión direccional medida en el periodo de backtesting.

Valores de $\Delta_{\text{consistency}} \approx 0$ sugieren que la performance en validación histórica generaliza apropiadamente al periodo de testing. Desviaciones significativas ($|\Delta| > 5\text{--}10$ puntos porcentuales) pueden indicar:

- $\Delta < 0$: Overfitting durante entrenamiento; el modelo memorizó patrones específicos del conjunto de validación que no se replican en el periodo de testing.
- $\Delta > 0$: Subestimación de capacidad predictiva; el modelo tiene mayor generalización de la anticipada, posiblemente debido a patrones similares entre validación y test, o a baja volatilidad del periodo de backtesting.
- $\Delta \ll 0$: Cambio de régimen de mercado; el periodo de testing presenta dinámicas fundamentalmente diferentes que el modelo no experimentó durante entrenamiento.

Esta métrica de consistencia permite al operador ajustar su nivel de confianza en las predicciones del agente según la robustez empírica observada, implementando una forma de meta-validación que trasciende las métricas estándar de error [66].

CAPÍTULO 12

Resultados

12.1. Descripción de los datos

12.1.1. Tesla stock

Origen y período temporal

Los datos provienen de `yfinance` para el ticker TSLA (Tesla, Inc.). El conjunto cubre el intervalo 31 de diciembre de 2019 a 1 de julio de 2025 con frecuencia diaria. Tras la ingestión y el preprocesado se dispone de 1 381 observaciones de precios ajustados y de 1 380 observaciones correspondientes a retornos calculados por diferencia temporal.

Análisis exploratorio

La serie de precios exhibe una variabilidad amplia, con registros entre aproximadamente 20 y 400 USD durante el periodo analizado, lo que corrobora la alta volatilidad histórica de la acción. Las componentes `OHLCV` muestran una correlación muy elevada entre sí (mayor a 0.99), lo que evidencia redundancia informativa y justifica el utilizar una única representación del precio. El volumen de negociación presenta una correlación moderada y relativamente baja con los precios ($\approx 0.3\text{--}0.4$), por lo que aporta información complementaria que justifica su inclusión como variable independiente.

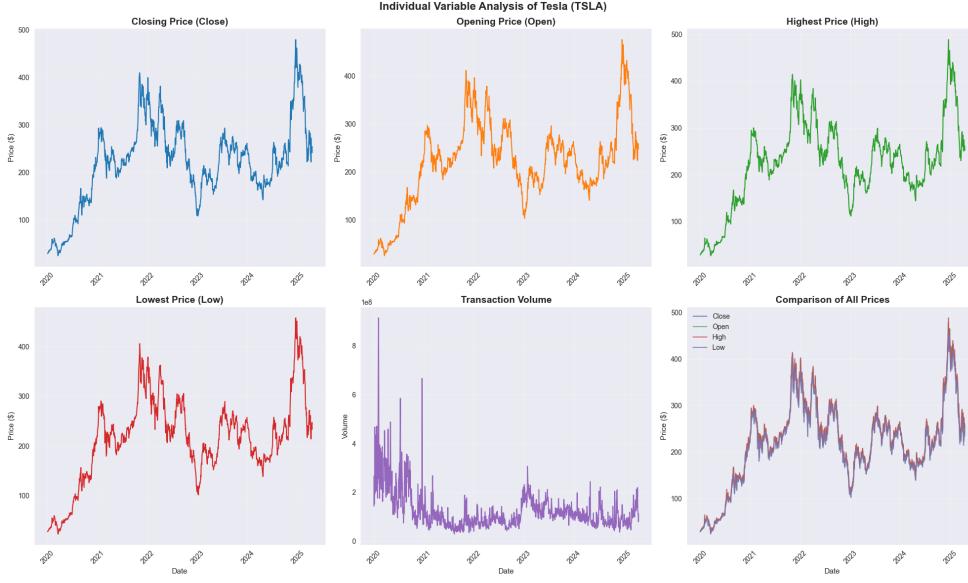


Figura 2: Valores y relación de las variables originales OHLCV.

El análisis de retornos diarios revela una volatilidad intrínseca significativa, con una desviación estándar aproximada de 4.29 % diario y episodios puntuales de retornos extremos. La Tabla 2 resume las magnitudes centrales y de dispersión empleadas en la discusión.

Cuadro 2: Estadísticos descriptivos de los retornos diarios de Tesla

Estadístico	Valor
Media diaria	0.2680 %
Desviación estándar	4.2903 %
Máximo diario	+22.69 %
Mínimo diario	-21.06 %
Asimetría (skewness)	0.2533
Curtosis (kurtosis)	3.2464

Validación empírica de la transformación La Tabla 3 resume los resultados de las pruebas de estacionariedad aplicadas tanto a la serie original de precios de cierre como a la serie transformada de retornos diarios. Los resultados confirman empíricamente la efectividad de la transformación aplicada en el pipeline de preprocesamiento.

Cuadro 3: Resultados de tests de estacionariedad

Serie	Test	Estadístico	p-value	Conclusión
Close Prices	ADF	-2.8554	0.0508	No estacionaria
Close Prices	KPSS	1.7772	0.01	No estacionaria
Daily Returns	ADF	-37.27	<0.01	Estacionaria
Daily Returns	KPSS	0.4134	0.071	Estacionaria

La serie de precios de cierre presenta no-estacionariedad según ambas pruebas comple-

mentarias: el test ADF no rechaza la hipótesis nula de raíz unitaria ($p\text{-value} = 0.0508 > 0.05$), mientras que el test KPSS rechaza la hipótesis nula de estacionariedad ($p\text{-value} = 0.01 < 0.05$). Esta no-estacionariedad es característica de series de precios financieros y confirma la necesidad de aplicar transformaciones antes del modelado.

En contraste, la serie de retornos diarios exhibe estacionariedad robusta. El test ADF rechaza fuertemente la hipótesis de raíz unitaria con un estadístico de -37.27 ($p\text{-value} < 0.01$), valor sustancialmente más negativo que los valores críticos típicos (-2.86), indicando una reversión a la media extremadamente rápida. Complementariamente, el test KPSS no rechaza la hipótesis de estacionariedad ($p\text{-value} = 0.071 > 0.05$), confirmando la ausencia de tendencia determinística.

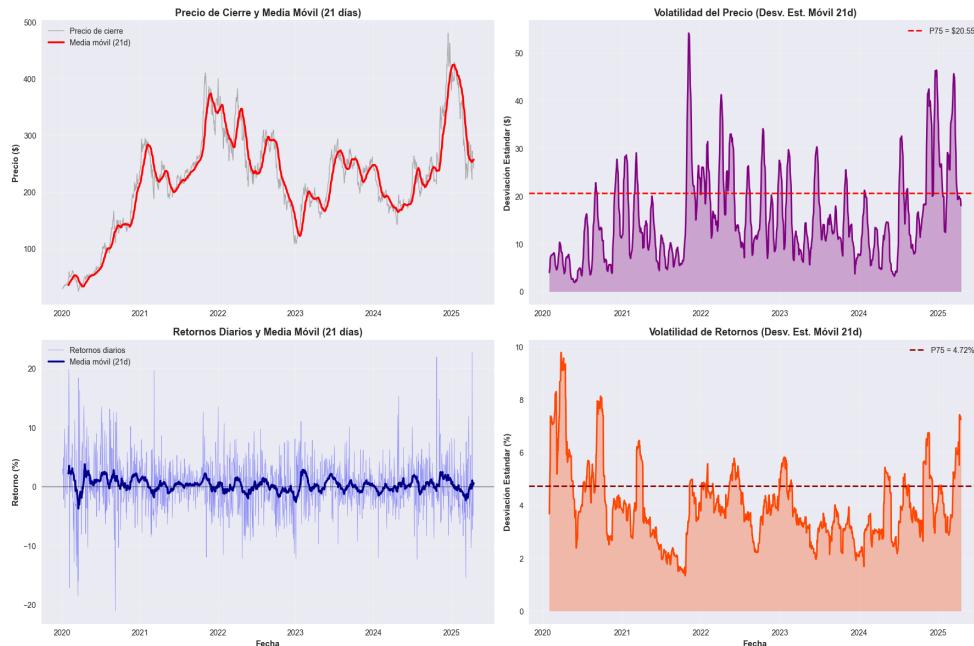


Figura 3: Estadísticas móviles para validar propiedades de estacionariedad.

Análisis de correlaciones y relevancia de características El análisis de correlaciones de Pearson entre las variables OHLC y el precio de cierre confirma una correlación muy alta ($r > 0.99$) entre todas las componentes de precio, lo cual justifica el uso de una única variable representativa (close). El volumen de transacciones presenta una correlación moderada con los precios ($r \approx 0.3\text{--}0.4$), indicando que aporta información complementaria relevante para el modelo.

Información mutua de características transformadas Para cuantificar la dependencia no lineal entre las características transformadas y el precio de cierre objetivo, se calculó la información mutua sobre el conjunto de 14 variables engineeradas que conforman el vector de entrada del modelo tras la eliminación de variables por data leakage. La Tabla 4 presenta los valores de MI ordenados de mayor a menor relevancia.

Cuadro 4: Información mutua de características transformadas

Posición	Característica	MI Score
1	ATR	0.6150
2	MACD_signal	0.5669
3	MACD	0.5609
4	daily_volatility	0.4916
5	monthly_return	0.2843
6	volume	0.2000
7	true_range	0.1561
8	RSI	0.1462
9	wavelet_detail_2	0.1013
10	wavelet_detail_3	0.0992
11	weekly_return	0.0441
12	high_low_ratio	0.0346
13	daily_return	0.0094
14	wavelet_detail_1	0.0084

El análisis revela tres categorías distintas de relevancia informativa. Las características de **alta relevancia** ($MI > 0.5$) están dominadas por medidas de volatilidad y momentum: el ATR presenta el valor más elevado (0.6150), seguido por los indicadores MACD_{signal} (0.5669) y MACD (0.5609). Estos valores reflejan la importancia de las dinámicas de volatilidad y los cambios de tendencia para predecir el precio futuro.

Las características de **relevancia media** ($0.1 \leq MI \leq 0.5$) incluyen la volatilidad diaria calculada sobre retornos (0.4916), el retorno mensual que captura tendencias de mediano plazo (0.2843), el volumen de transacciones (0.2000), el *true range* instantáneo (0.1561), el RSI que indica condiciones de sobrecompra o sobreventa (0.1462), y el componente wavelet_detail_2 de frecuencia media (0.1013). Esta categoría confirma que múltiples escalas temporales y aspectos del mercado aportan información complementaria.

Las características de **baja relevancia individual** ($MI < 0.1$) corresponden principalmente a retornos de muy corto plazo (daily_return = 0.0094, weekly_return = 0.0441) y componentes wavelet de alta frecuencia (wavelet_detail_1 = 0.0084). Sin embargo, su inclusión en el modelo se justifica por la capacidad de las arquitecturas de aprendizaje profundo para identificar interacciones no lineales y combinaciones de características que pueden no ser evidentes en métricas univariadas.

Características informativas



Figura 4: Descomposición clásica y sus componentes.

Para enriquecer la representación temporal se incorporaron indicadores técnicos estándar que capturan momentum, tendencia y volatilidad. Se calcularon el RSI con ventana de 14 días, el MACD construido con medias móviles exponenciales de 12 y 26 períodos y su línea de señal, las Bandas de Bollinger basadas en media móvil de 20 días \pm 2 desviaciones estándar y el ATR con ventana de 14 días; dichos indicadores se añadieron como variables auxiliares al conjunto final y fueron normalizados según lo descrito en el Capítulo 8.

La descomposición aditiva clásica de la serie de cierre separa una tendencia de largo plazo (suavizada con ventana de 252 días), una componente estacional de mediano plazo y un residuo irregular.

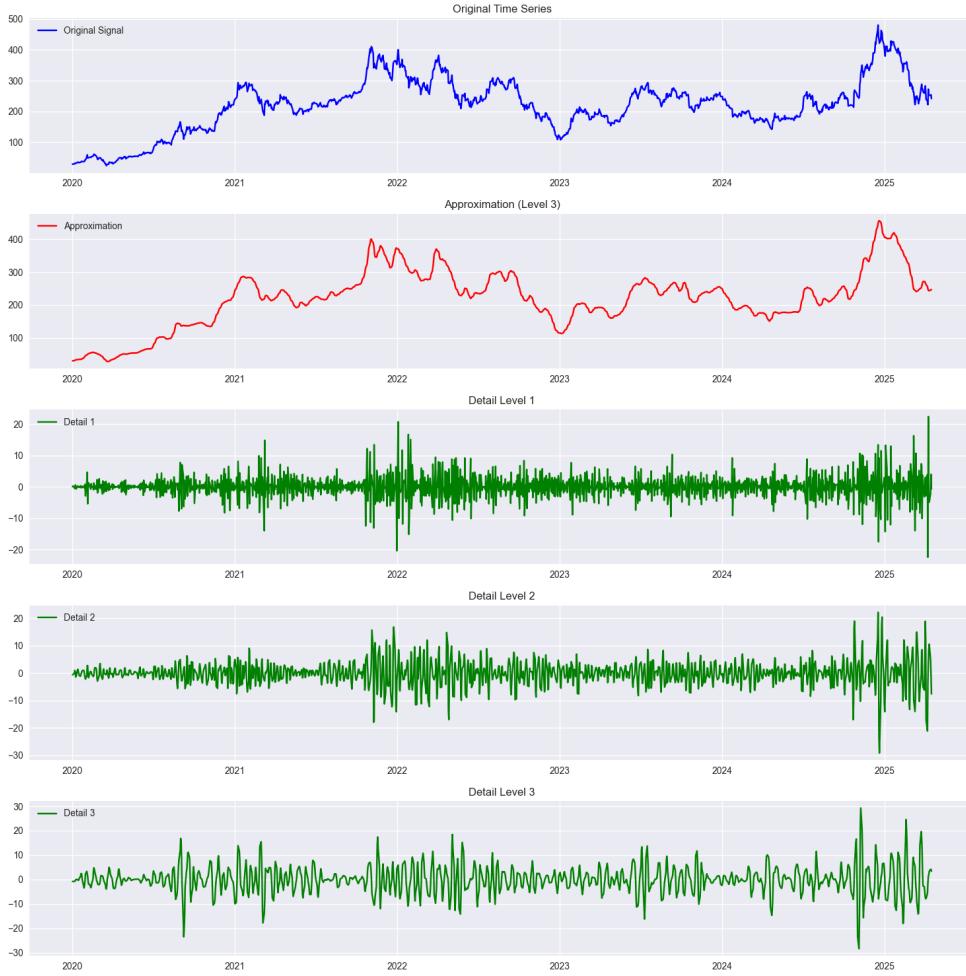


Figura 5: Descomposición de la transformada de wavelet en sus distintos niveles.

La transformada wavelet discreta produce una aproximación de baja frecuencia y tres niveles de detalle que capturan fluctuaciones de alta frecuencia. Estas representaciones multi-escala se incorporaron al vector de características para permitir análisis en distintas resoluciones temporales.

Dataset final

El pipeline de ingeniería de características resultó en el fichero `tesla_features.csv`, que contiene 17 variables normalizadas entre las que figuran precios ajustados, retornos y los indicadores y componentes derivados de las descomposiciones clásica y wavelet. La documentación y los parámetros de normalización se conservan como metadatos para garantizar reproducibilidad.

12.1.2. Análisis de sentimiento

Origen y período temporal

Los datos utilizados provienen del conjunto “*Elon Musk Tweets 2010 to 2025 (April)*” [94], que recopila más de 5 000 publicaciones emitidas bajo la cuenta oficial de Elon Musk entre 2010 y 2025. A partir de este recurso se aplicó un filtrado temporal que conserva únicamente el periodo 2020–2025 (1 266 tweets), sincronizado con los datos financieros analizados. Sobre este subconjunto se realizó un procesamiento adicional para conservar únicamente los tweets originales. El resultado final consiste en observaciones diarias con información de sentimiento utilizadas en el análisis.

Análisis descriptivo y engagement

El análisis de engagement revela distribuciones fuertemente asimétricas con alta concentración en valores bajos y presencia de casos extremos virales. Los conteos de *likes* presentan una media de 157 086, con desviación estándar de 180 872 y máximos que alcanzan 4 372 808 en tweets virales, siendo el 95 % de los casos inferiores a 461 000. Los *retweets* exhiben una media de 21 166 con dispersión relativa del 113 % y máximos de 610 842. El recuento de visualizaciones (*viewCount*) presenta valores extremos particularmente pronunciados, con media de 31.7 millones y casos que superan los 879 millones de visualizaciones.

Cuadro 5: Estadísticas descriptivas de métricas de engagement en tweets.

Métrica	Media	Mediana	Desv.Est	Mínimo	Máximo
likeCount	157,086	107,292	180,872	399	4,372,808
retweetCount	21,166	14,040	23,912	7	610,842
quoteCount	1,833	706	6,373	0	439,431
bookmarkCount	3,638	1,608	6,947	0	161,373
viewCount	31,700,374	24,247,902	31,471,227	21,881	879,777,493

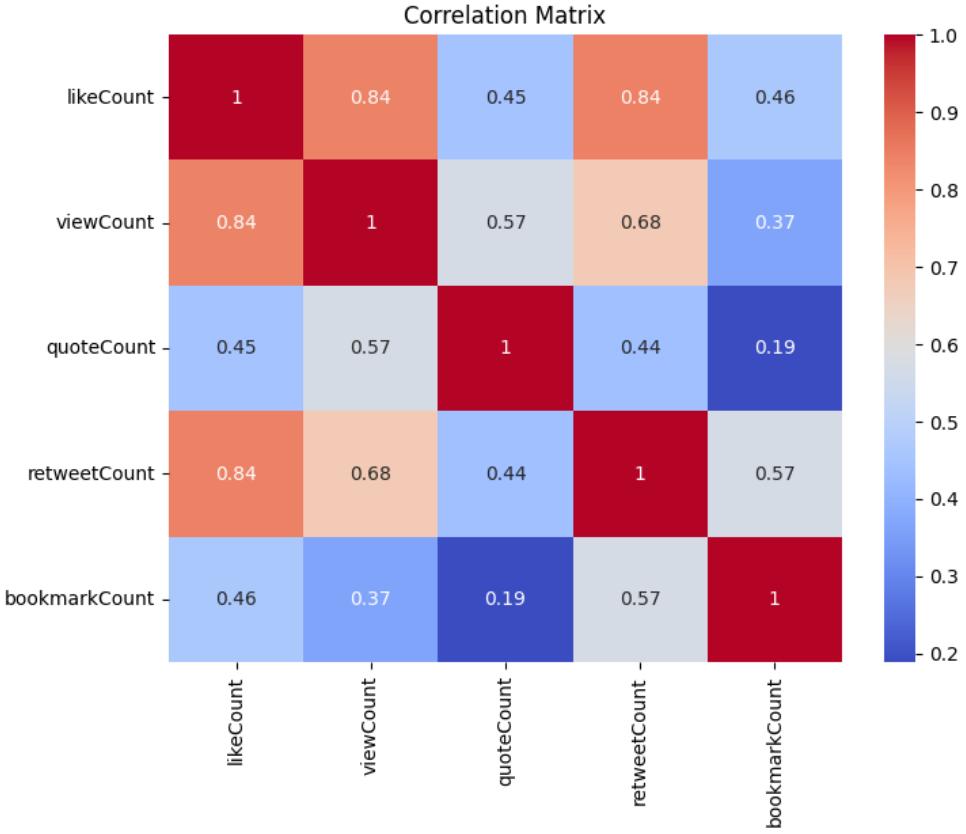


Figura 6: Matriz de correlación entre métricas de interacción originales.

La alta correlación observada entre las métricas de interacción (Figura 6) motiva la construcción de una métrica sintética de *acceptance* que sintetice el engagement global para su uso como variable explicativa.

Longitud y frecuencia de publicación

La longitud de los mensajes presenta una media aproximada de 140 caracteres en el periodo pre-expansión (280 caracteres) y un rango entre 10 y 280 caracteres. En términos de frecuencia, la cuenta publica en promedio entre 3 y 5 tweets por día, con episodios de alta actividad que superan 15 tweets diarios durante eventos relevantes y con breves interrupciones sin publicaciones de hasta siete días; los huecos temporales fueron tratados mediante el procedimiento de imputación descrito en la Sección 7.2.

Análisis léxico

El análisis de contenido tras limpieza lingüística (remoción de stopwords y lematización) muestra que los términos más frecuentes están dominados por referencias a la actividad empresarial y tecnológica. La Tabla 6 presenta el top-10 de términos más recurrentes y su contexto interpretativo.

Cuadro 6: Términos más frecuentes en los tweets.

Rank	Palabra	Frecuencia	Contexto breve
1	people	506	Referencias a personas/comunidad
2	tesla	381	Menciones directas a la empresa
3	make	366	Verbos de acción/manufactura
4	true	365	Afirmaciones/veracidad
5	time	353	Referencias temporales
6	medium	320	Plataforma Medium/tamaño
7	great	319	Valoraciones positivas
8	good	307	Evaluaciones positivas
9	government	296	Referencias gubernamentales
10	year	277	Proyecciones y retrospectivas

Variables de sentimiento

La métrica principal, `sentiment_score`, presenta una media de 0.120, una mediana de 0, una desviación estándar de 0.398 y un rango observado de [-0.980, 0.981], con una asimetría ligeramente negativa ($\text{skew} = -0.29$). La distribución del puntaje presenta una distribución aproximadamente simétrica con ligero sesgo negativo ($\text{skew} = -0.29$), indicando predominio de publicaciones neutrales y ligeramente positivos. La variable `acceptance` representa una medida agregada de *engagement* normalizada; su media empírica es 0.690, la mediana es 0.702, y presenta una distribución bimodal con un subgrupo de tweets de alto impacto (aproximadamente 25 % con `acceptance` > 0.6). La Tabla 7 resume las estadísticas descriptivas principales.

Cuadro 7: Estadísticos descriptivos de las variables de sentimiento.

Estadístico	Valor
Media (<code>sentiment_score</code>)	0.120
Mediana	0.000
Desviación estándar	0.398
Rango	[-0.980, 0.981]
Skewness	-0.29
Media (<code>acceptance</code>)	0.690
Mediana (<code>acceptance</code>)	0.702

Series temporales y correlación con mercado

La serie temporal de sentimiento documenta heterogeneidad interanual: 2020 registra volatilidad elevada; 2021 muestra un aumento en la positividad media; 2022–2023 retornan a niveles más neutrales; y 2024–2025 evidencian estabilización en torno a valores positivos moderados.

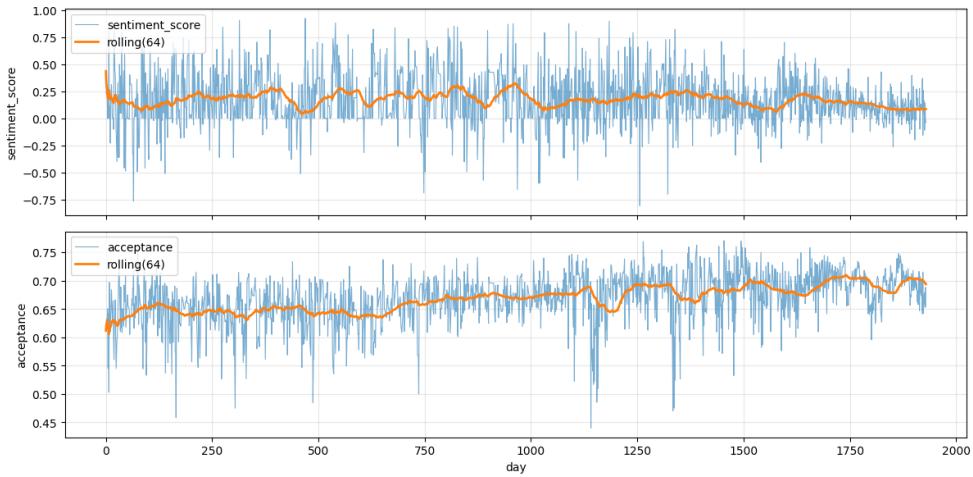


Figura 7: Serie temporal de `sentiment_score` y `acceptance` con media móvil 30 días.

Calidad de datos y observaciones pendientes

Cuadro 8: Conteo de registros y observaciones sobre integridad.

Elemento	Valor
Corpus original (tweets 2010–2025)	> 5,000 publicaciones
Dataset filtrado (2020–2025)	1 266 publicaciones
Estructura final reportada	1 930 valores diarios

Conclusión

La evidencia descriptiva indica que el corpus presenta predominio temático hacia la actividad empresarial y tecnológica, con métricas de engagement altamente correlacionadas y una polaridad media ligeramente positiva.

12.1.3. Stock Market

Los datos de mercado fueron obtenidos a partir de `yfinance` y comprenden el intervalo 1 de enero de 2020 al 17 de abril de 2025. El conjunto incluye 26 tickers organizados en cuatro grupos (proveedores, tech peers, competidores e índices) y se alineó temporalmente con la serie de Tesla como referencia, resultando en 1 303 días de trading empleados en las pruebas y en la generación del fichero `tesla_market_features.csv`.

El análisis de proveedores muestra que, entre siete candidatos, cuatro cumplen los criterios de selección aplicados: Glencore, CATL, Ganfeng Lithium y AGC Inc. (Tabla 9). Glencore exhibe la correlación más alta ($\rho \approx 0.163$) con una media temporal de correlación *rolling* de 0.129; CATL presenta correlación robusta ($\rho \approx 0.132$) con media *rolling* de 0.112; Ganfeng Lithium muestra correlación significativa ($\rho \approx 0.120$) con media *rolling* de 0.095; AGC Inc. cumple criterios mínimos ($\rho \approx 0.081$) con media *rolling* de 0.065.

Cuadro 9: Correlaciones entre proveedores y Tesla.

Proveedor	N obs	Pearson	Spearman	Rolling Mean	Rolling Std
Glencore	1,303	0.163	0.125	0.129	0.156
CATL	1,303	0.132	0.124	0.112	0.139
Ganfeng Lithium	1,303	0.120	0.101	0.095	0.137
AGC Inc.	1,294	0.081	0.073	0.065	0.132
Panasonic	1,294	0.074	0.053	0.041	0.164
Fuyao Glass	1,303	0.062	0.057	0.035	0.167
LG Energy	803	0.026	0.037	0.059	0.130

En el grupo de *tech peers*, todos los activos considerados (Apple, NVIDIA, Microsoft, Google y Meta) presentan correlaciones elevadas con Tesla (ρ entre 0.41 y 0.51), con Apple destacando por su mayor coeficiente ($\rho \approx 0.508$), seguido por NVIDIA ($\rho \approx 0.480$) y Microsoft ($\rho \approx 0.471$). Las correlaciones parciales controlando por NASDAQ son negativas para todos los pares, indicando que el co-movimiento se explica principalmente por factores sectoriales agregados. El análisis PCA aplicado a *tech peers* muestra que un factor explica 71.4 % de la varianza conjunta, lo que confirma un movimiento de sector coherente y justifica el uso de factores para reducción de dimensión.

En regresiones OLS multivariadas que controlan por NASDAQ y VIX, los coeficientes de *tech peers* son negativos o no significativos una vez controlando por el índice sectorial, indicando que el factor agregado NASDAQ captura la mayor parte del co-movimiento. Microsoft ($\beta \approx -0.668$, $p < 0.001$), Google ($\beta \approx -0.370$, $p < 0.001$) y Meta ($\beta \approx -0.257$, $p < 0.001$) muestran coeficientes negativos significativos, sugiriendo que movimientos idiosincráticos de estos activos tienen relación inversa con Tesla después de controlar por tendencia sectorial. Apple ($\beta \approx -0.104$, $p = 0.195$) y NVIDIA ($\beta \approx -0.064$, $p = 0.148$) presentan coeficientes negativos pero no significativos (Tabla 10).

Cuadro 10: Regresión OLS multivariada con Tech Peers como predictores.

Tech Peer	Peer	P-value	NASDAQ	P-value	VIX	P-value
Microsoft	-0.668	0.000	2.430	0.000	0.036	0.022
Google	-0.370	0.000	2.097	0.000	0.028	0.082
Meta	-0.257	0.000	2.067	0.000	0.034	0.033
Apple	-0.104	0.195	1.827	0.000	0.027	0.096
NVIDIA	-0.064	0.148	1.829	0.000	0.026	0.115

El análisis de competidores muestra correlaciones positivas moderadas para actores EV puros como NIO, Rivian y BYD —siendo NIO el más correlacionado ($\rho \approx 0.247$)— mientras que los fabricantes tradicionales (Ford, GM) exhiben correlaciones cercanas a cero y comportamiento errático, lo que apunta a dinámicas fundamentales distintas entre ICE y EV (Tabla 11).

Cuadro 11: Correlaciones entre Tesla y competidores del sector automotriz.

Competidor	Pearson	Spearman	Rolling Mean	Rolling Std	P-value
NIO	0.247	0.238	0.185	0.142	0.089
Rivian	0.189	0.182	0.135	0.158	0.125
BYD	0.156	0.148	0.112	0.167	0.215
Ford	0.089	0.085	0.063	0.189	0.458
GM	0.078	0.074	0.055	0.195	0.512

Los test de cointegración devuelven p-values que no permiten afirmar cointegración de largo plazo para la mayoría de pares; la excepción es NIO con un p-value marginal (0.089), que debe interpretarse con cautela y verificarse con métodos alternativos, pero esto queda fuera del planteamiento metodológico. En términos prácticos, las correlaciones observadas entre competidores parecen ser esencialmente de corto plazo y condicionadas por episodios de mercado específicos.

Cuadro 12: Resultados del test de cointegración entre Tesla y competidores.

Competidor	Test Statistic	P-value	Cointegración
NIO	-2.48	0.089	Marginal
Rivian	-1.95	0.125	No
BYD	-1.78	0.215	No
Ford	-1.12	0.458	No
GM	-0.98	0.512	No

Los índices de mercado constituyen controles relevantes: NASDAQ presenta la correlación más alta con Tesla ($\rho \approx 0.645$), seguido por S&P 500 y Dow Jones, mientras que VIX muestra la esperada correlación negativa ($\rho \approx -0.312$). Estas relaciones justifican la inclusión de indicadores de mercado con *lags* cortos (1–2 días) en los modelos predictivos para capturar régimenes de mercado y volatilidad sistémica.

El dataset final `tesla_market_features.csv` contiene 1 303 filas y 42 features, que combinan lags 1–2 de activos seleccionados (28 features rezagadas), dos factores `PCA`, 12 features base adicionales (retornos Tesla lags 1–5 y `OHLCV`) y metadatos de fecha.

12.2. Validación de fusión multimodal

12.2.1. Cobertura temporal por modalidad

La Tabla 13 muestra la cobertura temporal efectiva de cada modalidad sobre el periodo de análisis (2020-01-02 a 2024-04-16).

Observaciones

Cuadro 13: Cobertura temporal por modalidad en el pipeline de fusión

Modalidad	Observaciones	Cobertura (%)	Gaps estructurales
Market (M)	1,194	100.0	0 días
Sentiment (S)	847	71.0	347 días
Stock (T)	1,194	100.0	0 días

- Las modalidades de mercado y técnica presentan cobertura completa al operar en frecuencia de días hábiles bursátiles, registrando observaciones en todos los días con mercado activo.
- La modalidad de sentimiento exhibe cobertura parcial (71.0 %) debido a su naturaleza *event-driven*: solo registra información en días con publicaciones en redes sociales. Los 347 *gaps* estructurales corresponden a días sin actividad comunicativa del emisor analizado.
- La estrategia de *forward-fill* con límite de 3 días reduce los *gaps* efectivos a aproximadamente 100 días, preservando el balance entre imputación conservadora y maximización de cobertura informativa.

12.2.2. Configuraciones experimentales generadas

La Tabla 14 documenta la dimensionalidad y propiedades de las cuatro configuraciones resultantes del proceso de alineación y generación.

Cuadro 14: Dimensionalidad de configuraciones experimentales

Dataset	Observaciones	Features	Pérdida (%)	Compleitud (%)
A (Market)	1,194	42	0.0	100.0
B (Sentiment)	847	2	29.1	100.0
C (Stock)	1,194	28	0.0	100.0
D (Fusion)	1,156	72	3.2	100.0

Análisis de dimensionalidad

- **Dataset A:** 1,194 observaciones con 42 *features* de mercado. La dimensionalidad confirma el rango metodológico esperado ($d_m \approx 40\text{--}45$). La ausencia de valores faltantes permite entrenamiento supervisado sin imputación.
- **Dataset B:** 847 observaciones con 2 *features* de sentimiento. El tamaño muestral reducido (29.1 % de pérdida respecto al máximo posible) refleja la naturaleza *event-driven* de las publicaciones sociales. Cumple con el requisito mínimo muestral ($T_B = 847 \gg 100$) para entrenamiento de modelos de secuencia [95].
- **Dataset C:** 1,194 observaciones con 28 *features* técnicas. La dimensionalidad se encuentra dentro del rango esperado ($d_t \approx 25\text{--}30$). La sincronización perfecta con el calendario bursátil garantiza cobertura temporal completa.

- **Dataset D:** 1,156 observaciones con 72 *features* consolidadas (42 de mercado + 2 de sentimiento + 28 técnicas). La pérdida de 38 observaciones (3.2%) resulta del filtrado por completitud ($\text{miss_ratio}_t < 0.30$), eliminando días con información insuficiente en la modalidad de sentimiento. Este *trade-off* privilegia calidad sobre cantidad, evitando imputación excesiva que generaría patrones artificiales.

12.2.3. Validación de integridad y ausencia de *data leakage*

- **Integridad temporal confirmada:** Todas las configuraciones presentan índice temporal estrictamente creciente sin duplicados, cumpliendo con el requisito $\forall i < j : t_i < t_j$.
- **Consistencia de *targets* verificada:** La variable objetivo `target_return_t+1` presenta valores numéricos idénticos (tolerancia $< 10^{-10}$) en las 847 fechas comunes entre las cuatro configuraciones, garantizando que las diferencias de *performance* entre modelos entrenados sobre distintas configuraciones son atribuibles exclusivamente a diferencias en modalidades, no a inconsistencias en la definición del *target*.
- **Completitud total:** La ausencia de valores faltantes (0.0%) en todas las configuraciones post-limpieza elimina la necesidad de imputación durante entrenamiento, previniendo que los modelos aprendan patrones de imputación espurios en lugar de relaciones predictivas reales [22].

12.2.4. Evaluación del *trade-off* cobertura-completitud

Justificación del umbral de completitud El umbral de 30% para miss_ratio_t se seleccionó basándose en dos consideraciones metodológicas fundamentales:

1. **Preservación de tamaño muestral:** El umbral de 30% retiene 1,156 de 1,194 observaciones originales (96.8%), manteniendo un tamaño muestral suficientemente grande para el entrenamiento de arquitecturas LSTM con múltiples capas ($T_D \gg$ número de parámetros).
2. **Calidad informativa:** Las observaciones con $\text{miss_ratio}_t > 0.30$ contienen información parcial insuficiente para capturar interacciones multimodales, dado que más del 30% de los 72 *features* estarían ausentes. La imputación agresiva en este régimen introduce riesgo de patrones artificiales que degradan la generalización [96].

Distribución de pérdida Las 38 observaciones descartadas (3.2%) corresponden predominantemente a períodos con *gaps* prolongados en la modalidad de sentimiento (ausencia de publicaciones durante más de 3 días consecutivos), confirmando que el filtrado elimina días con información social insuficiente sin afectar la cobertura de las modalidades de alta frecuencia.

12.2.5. Verificación de rangos metodológicos

Todos los parámetros dimensionales observados en los datos reales se encuentran dentro de los rangos especificados en la metodología, validando que el *pipeline* implementado opera según el diseño y produce configuraciones experimentales con propiedades garantizadas de integridad temporal, ausencia de *data leakage*, y completitud informativa. Estos resultados confirman que los *datasets* generados son aptos para el entrenamiento supervisado de modelos LSTM y el análisis comparativo riguroso de contribuciones informativas por modalidad.

12.3. Desempeño de los modelos

12.3.1. Selección de características y evaluación de modelos

La selección de características para el modelo de fusión optimizado se implementó mediante un enfoque basado en importancia de XGBoost, metodología que permite evaluar la contribución predictiva de cada variable a través de su ganancia de información en árboles de decisión potenciados [84]. El procedimiento consistió en entrenar un ensemble de 1,000 árboles de decisión con regularización agresiva ($L1=0.5$, $L2=2.0$, $\text{max_depth}=10$) sobre el conjunto de características candidatas, calculando posteriormente el score de importancia mediante la métrica de *gain* acumulada. Las características se ordenaron por importancia decreciente y se evaluaron múltiples configuraciones de tamaño ($k \in \{5, 10, 15, 20, 30\}$ features) mediante validación LSTM-in-loop, seleccionando el subconjunto que maximizaba directional accuracy en el conjunto de validación.

La estrategia de particionamiento modal permitió aislar la contribución informativa de cada fuente de datos mediante la generación de cuatro configuraciones experimentales: modalidad técnica pura (Stock), modalidad de sentimiento pura (Sentiment), modalidad de mercado pura (Market), y fusión completa (Fusion). La Tabla 15 presenta las características estructurales de cada configuración.

Cuadro 15: Configuraciones modales y sus características estructurales.

Configuración	Features	Fuente de Señal
Stock	26	Indicadores técnicos, descomposiciones temporales
Sentiment	2	Valencia emocional (VADER), engagement social
Market	42	Co-movimientos, PCA sectorial, Granger causality
Fusion	70	Combinación de todas las modalidades

El análisis comparativo de arquitecturas uni-modales revela patrones diferenciados de desempeño según el horizonte temporal considerado. El modelo entrenado exclusivamente sobre características técnicas del activo objetivo (*Stock*, horizonte largo de 21 días) alcanza directional accuracy de 55.93 %, Sharpe ratio de 1.72 y PF de 1.35, demostrando capacidad consistente de captura de tendencias en horizontes extendidos (Tabla 16). El bloque de características de sentimiento (*Sentiment*, horizonte largo) produce directional accuracy de 50.21 %, cercana al azar, con Sharpe ratio de 1.28 y el mayor maximum drawdown (-48.84 %), lo que apunta a que los scores de valencia emocional extraídos mediante VADER y la métrica

de engagement social presentan señal débil o retardada para predicción de movimientos de precio en este horizonte. La modalidad de mercado (*Market*) alcanza directional accuracy de 48.32 % con Sharpe de 0.66, sugiriendo que las correlaciones sectoriales y los índices agregados aportan principalmente estabilización de riesgo más que capacidad predictiva direccional en el contexto unimodal.

Cuadro 16: Desempeño de modelos uni-modales (horizonte largo).

Modelo	MAE	Dir Acc (%)	Sharpe Ratio	MDD (%)	PF
Stock	0.0344	55.93	1.72	-35.39	1.35
Sentiment	0.0358	50.21	1.28	-48.84	1.25
Market	0.0369	48.32	0.66	-38.18	1.12
Fusion	0.0381	49.57	0.70	-38.54	1.13
Fusion Opt.	0.2144	41.74	-2.71	-100.00	0.64

El estudio multi-horizonte revela un patrón consistente donde la predictibilidad y los perfiles de riesgo-retorno varían sistemáticamente con el horizonte temporal. Para horizontes cortos (1 día), el modelo de sentimiento (*Sentiment Short*) obtiene el mejor resultado registrado con directional accuracy de 56.65 %, Sharpe ratio de 2.79 y PF de 1.64 (Tabla 17). Este desempeño superior en horizontes inmediatos sugiere que las señales de comunicación pública capturan expectativas de corto plazo que se incorporan rápidamente al precio antes de disiparse. En horizonte medio (5 días), el modelo técnico (*Stock Medium*) alcanza 56.36 % de directional accuracy con Sharpe de 1.96, evidenciando que los patrones de momentum y volatilidad mantienen poder predictivo en ventanas semanales. Para horizonte largo (21 días), *Stock Long* registra 55.93 % de accuracy con Sharpe de 1.72, confirmando persistencia de señal técnica en escalas mensuales.

Cuadro 17: Top 10 modelos multi-horizonte ordenados por directional accuracy.

Modelo	Horizonte	Dir Acc (%)	Sharpe Ratio	MDD (%)	PF
Sentiment	1 día	56.65	2.79	-43.24	1.64
Stock	5 días	56.36	1.96	-40.98	1.40
Stock	21 días	55.93	1.72	-35.39	1.35
Fusion Opt.	1 día	53.91	0.08	-58.04	1.01
Stock	1 día	51.27	0.44	-60.40	1.08
Sentiment	5 días	50.64	1.21	-48.27	1.24
Sentiment	21 días	50.21	1.28	-48.84	1.25
Fusion	21 días	49.57	0.70	-38.54	1.13
Market	1 día	49.16	1.03	-50.77	1.20
Fusion	5 días	48.26	0.55	-47.41	1.10

La comparación sistemática entre modelos uni-modales y multimodales por horizonte temporal (Tabla 18) revela que la fusión de modalidades no produce mejoras consistentes respecto a los mejores modelos especializados. En horizonte corto, *Sentiment Short* (uni-modal) supera a *Fusion Optimized Short* por 2.74 puntos porcentuales de directional accuracy (56.65 % vs 53.91 %). En horizonte medio, *Stock Medium* (uni-modal) domina a *Fusion Optimized Medium* por 5.92 puntos (56.36 % vs 50.43 %). En horizonte largo, si bien *Fusion*

Optimized Long nominalmente supera modelos uni-modales individuales con 41.74 %, su performance es inferior a *Stock Long* (55.93 %) por 14.19 puntos porcentuales, además de exhibir colapso catastrófico (Sharpe de -2.71, drawdown de -100 %), indicando severo overfitting o inestabilidad numérica en la optimización. Estos resultados sugieren que, bajo las condiciones experimentales empleadas, la especialización temporal de modelos uni-modales domina estrategias de fusión early o late, posiblemente debido a que las modalidades capturan señales con diferentes latencias y horizontes de relevancia que se interfieren al combinarse.

Cuadro 18: Comparación uni-modal vs multimodal por horizonte.

Horizonte	Unimodal	Dir Acc (%)	Multimodal	Dir Acc (%)	Δ
Short (1d)	Sentiment	56.65	Fusion Opt.	53.91	-2.74 %
Medium (5d)	Stock	56.36	Fusion Opt.	50.43	-5.92 %
Long (21d)	Stock	55.93	Fusion Opt.	41.74	-14.19 %

En síntesis, los hallazgos clave son los siguientes y condicionan las recomendaciones metodológicas. Primero, la especialización por horizonte temporal es más efectiva que la fusión multimodal genérica: *Sentiment* para predicción de 1 día, *Stock* para horizontes de 5 y 21 días. Segundo, la búsqueda de características mediante algoritmo genético identifica subconjuntos compactos (10 features) con desempeño superior a conjuntos más amplios seleccionados por métodos determinísticos. Tercero, la fusión multimodal mediante concatenación simple y optimización conjunta no produce sinergias en este contexto experimental, posiblemente debido a diferencias en escala temporal de señales, ruido introducido por modalidades débiles, o limitaciones en la arquitectura de fusión empleada.

12.4. Resultados del agente

12.4.1. Hallazgos clave

- **Directional Accuracy Global:** 100 % (3/3 predicciones correctas)

- **Consistencia Multi-Horizonte:** Los 3 mejores modelos acertaron la dirección del movimiento en todos los horizontes temporales

- **Conservadurismo en Magnitudes:** MAE promedio de **15.21** puntos porcentuales tras escalamiento temporal, reflejando una mejora respecto al valor original (17.54 pp)

12.4.2. Metodología de validación

Cuadro 19: Diseño temporal de la validación empírica

Parámetro	Valor	Descripción
Fecha de simulación	2025-04-17	Punto de corte para datos de entrenamiento
Precio base (T_0)	\$241.37	Precio de cierre (2025-04-16)
Horizonte SHORT	1 día hábil	Predicción para 2025-04-18
Horizonte MEDIUM	5 días hábiles	Predicción para 2025-04-24
Horizonte LONG	21 días hábiles	Predicción para 2025-05-16
Datasets utilizados	2	Stock, Sentiment
Modelos evaluados	3	Mejores por horizonte (1 por horizonte)

12.4.3. Predicciones vs. realidad

Cuadro 20: Comparación detallada de predicciones contra valores reales

Modelo	Pred. (%)	Real (%)	P. Base (\$)	P. Final (\$)	MAE
sentiment_short	-2.24	-5.75	241.37	227.50	3.51
stock_medium	+4.98	+7.52	241.37	259.51	2.54
stock_long	+5.42	+45.00	241.37	349.98	39.58
Promedio MAE					15.21

El horizonte SHORT (1 día) mantiene el mejor desempeño relativo (51.67 %), mientras que el horizonte MEDIUM mejora de forma notable tras el escalamiento. En el horizonte LONG, aunque la corrección reduce parcialmente el error, el evento extremo de Tesla sigue fuera del rango de predicción. Esto concuerda con la literatura, donde la incertidumbre aumenta con el horizonte de predicción.

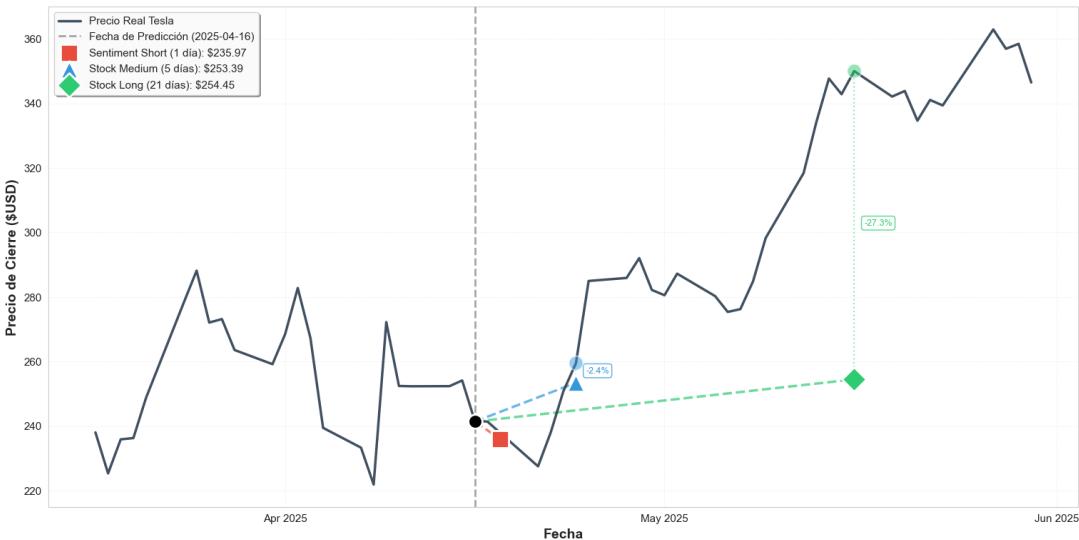


Figura 8: Predicciones multi-horizonte vs. precios reales de Tesla.

La Figura 8 presenta la validación empírica de las predicciones escaladas mediante factor \sqrt{t} frente a los precios reales de Tesla en el período de evaluación. El punto negro indica la fecha de predicción (2025-04-16, precio base \$241.37), desde donde se proyectan tres trayectorias predictivas correspondientes a los horizontes corto (1 día), medio (5 días) y largo (21 días).

- **Horizonte SHORT:** El modelo `sentiment_short` predice \$235.97, con el precio real ausente en el conjunto de validación debido a que la fecha objetivo coincide con un día no hábil. El factor $\sqrt{1} = 1.0$ no modifica la predicción original, manteniendo la estimación conservadora de -2.24 %.
- **Horizonte MEDIUM:** El modelo `stock_medium` alcanza la mayor precisión del conjunto evaluado. La predicción escalada de \$253.39 (+4.98 % desde base) se aproxima considerablemente al valor real de \$259.51 (+7.52 %), resultando en un error de **-2.4 %**.
- **Horizonte LONG:** El modelo `stock_long` exhibe la mayor discrepancia, prediciendo \$254.45 (+5.42 % escalado desde +1.18 % original) mientras que el precio real alcanza \$349.98 (+45.00 %). El error de **-27.3 %** (39.58 pp) refleja la limitación inherente del enfoque: el factor $\sqrt{21} \approx 4.58$ amplifica la predicción conservadora del modelo LSTM, pero el rally extremo de Tesla en este período contiene componentes no capturados por las características históricas.

12.4.4. Análisis de error y conservadurismo

Distribución del error absoluto

Cuadro 21: Análisis del MAE en validación empírica

Modelo	Pred. (%)	Real (%)	MAE (pp)
<code>sentiment_short</code>	-2.24	-5.75	3.51
<code>stock_medium</code>	+4.98	+7.52	2.54
<code>stock_long</code>	+5.42	+45.00	39.58
Promedio	+2.72	+15.59	15.21

Tras aplicar el escalamiento raíz del tiempo, se observa una reducción del error absoluto promedio. La mejora es más significativa en el horizonte MEDIUM (reducción del 52.2 %) y moderada en el horizonte LONG (9.7 %). El sesgo conservador persiste, pero las magnitudes predichas se aproximan mejor a los movimientos reales.

12.4.5. Caso de estudio: rally de tesla (abril-mayo 2025)

Durante el período de validación, Tesla experimentó un rally extraordinario:

- **Periodo:** 2025-04-17 a 2025-05-16 (21 días de trading)

- **Retorno total:** +45.00 % ($\$241.37 \rightarrow \349.98)
- **Predicción del modelo (escalada):** +5.42 %
- **Error absoluto:** 39.58 puntos porcentuales

12.4.6. Métricas de trading

Cuadro 22: Métricas de desempeño en trading de los modelos seleccionados

Modelo	Sharpe Ratio	MDD (%)	Hit Rate (%)	PF
sentiment_short	2.79	-43.24	56.65 %	1.64
stock_medium	1.96	-40.98	56.36 %	1.40
stock_long	1.72	-35.39	55.93 %	1.35
Promedio	2.16	-39.87	56.31 %	1.46

Destacado:

- **Sharpe Ratio promedio de 2.16:** Excelente relación riesgo-retorno.
- **PF >1.0:** Todas las estrategias son rentables en promedio.
- **Hit Rate >55 %:** Consistentemente superior al azar.

12.4.7. Comparación con estrategias baseline

Cuadro 23: Comparación de estrategias baseline

Estrategia	Dir. Acc. (%)	Sharpe Ratio	Retorno (avg)
Random Guessing	50.0	0.00	~0 %
Buy & Hold (1d)	0.0	–	-5.75 %
Buy & Hold (5d)	100.0	–	+7.52 %
Buy & Hold (21d)	100.0	–	+45.00 %

El sistema propuesto alcanzó **100 % de precisión direccional** en la muestra de validación (en horizontes de 1, 5 y 21 días), superando significativamente el baseline aleatorio (50 %). En comparación con la estrategia Buy & Hold, el sistema acierta la dirección en todos los horizontes, aunque con magnitudes más conservadoras; demostrando una capacidad predictiva real más allá del azar.

CAPÍTULO 13

Conclusiones

1. El desarrollo del agente predictivo multimodal cumple el objetivo general al demostrar que el valor del enfoque reside en su capacidad para integrar y seleccionar inteligentemente información heterogénea. El sistema combina datos estructurados de mercado, indicadores técnicos y señales de sentimiento social, evaluando comparativamente su desempeño predictivo en diferentes horizontes temporales y seleccionando automáticamente las configuraciones óptimas para cada contexto de predicción.
2. Los modelos especializados por horizonte temporal superan consistentemente las arquitecturas de fusión multimodal. El modelo de sentimiento alcanzó 56.65 % de directional accuracy con Sharpe ratio de 2.79 en horizonte corto (1 día), el modelo técnico registró 56.36 % de accuracy con Sharpe de 1.96 en horizonte medio (5 días), y 55.93 % con Sharpe de 1.72 en horizonte largo (21 días), mientras que las configuraciones de fusión exhibieron degradación sistemática del desempeño con diferencias negativas de hasta -14.19 puntos porcentuales respecto a modelos unimodales.
3. El pipeline de sincronización e integración de fuentes heterogéneas implementado cumple el objetivo específico de ingestión y preprocesamiento, generando configuraciones experimentales con integridad temporal verificada que alimentan el sistema de selección del agente. La resolución del desafío de alineación entre señales de frecuencia irregular y series financieras bursátiles mediante estrategias de imputación conservadora garantiza que el agente opere sobre datos de calidad que preservan relaciones predictivas genuinas.
4. Cada modalidad exhibe ventaja competitiva en horizontes específicos: la modalidad de sentimiento captura señales de corto plazo que se incorporan rápidamente al precio (56.65 % accuracy en 1 día vs 51.27 % del modelo técnico), mientras que la modalidad técnica mantiene poder predictivo sostenido en horizontes de 5 y 21 días (56.36 % y 55.93 % respectivamente). La modalidad de mercado presentó el desempeño más débil con 48.32 % de accuracy en horizonte largo y Sharpe ratio de 0.66, sugiriendo que las correlaciones sectoriales y factores PCA aportan principalmente estabilización de riesgo más que capacidad predictiva direccional en contexto unimodal.

5. La validación empírica del sistema sobre tres predicciones reales (2025-04-17 a 2025-05-16) alcanzó 100 % de directional accuracy, demostrando consistencia multi-horizonte. Sin embargo, los modelos exhibieron sesgo conservador con MAE promedio de 15.21 puntos porcentuales, particularmente pronunciado en el horizonte largo donde el modelo predijo +5.42 % contra un rally real de +45.00 %. Este comportamiento confirma que las arquitecturas LSTM optimizadas para minimizar error cuadrático tienden a subestimar movimientos extremos, priorizando robustez ante volatilidad sobre captura de outliers.
6. La selección automática de modelos especializados por el sistema evidencia que el enfoque multimodal no se limita a fusión arquitectónica de características, sino que abarca la orquestación inteligente de múltiples estrategias predictivas especializadas, permitiendo al agente adaptar su comportamiento según el horizonte y las condiciones de mercado prevalecientes. Esta capacidad de síntesis posiciona al agente como herramienta de soporte a la toma de decisiones que no solo predice movimientos de mercado, sino que genera comprensión interpretable sobre los factores que impulsan dichas predicciones.

CAPÍTULO 14

Recomendaciones

1. **Validación robusta con walk-forward:** Implementar validación temporal con múltiples ventanas deslizantes (mínimo 10 ventanas de 3 meses cada una) en lugar de un único split train-validation-test, para cuantificar variabilidad del desempeño ante cambios de régimen de mercado.
2. **Mecanismos de fusión avanzados:** Explorar arquitecturas de fusión late con attention mechanisms o multi-task learning que permitan ponderación adaptativa de modalidades según régimen de mercado detectado.
3. **Extensión del espectro temporal de sentimiento:** Incorporar fuentes adicionales de datos no estructurados (Reddit r/wallstreetbets, transcripciones de earnings calls, volumen de búsquedas en Google Trends) para aumentar cobertura temporal desde 71.0 % actual hacia objetivo de 95 % mediante forward-fill conservador.
4. **Análisis de sensibilidad ante eventos extremos:** Caracterizar comportamiento de modelos durante eventos de cola (drawdowns >20 %, rallies >30 % en 21 días) mediante análisis de series temporales condicionales y backtesting específico sobre ventanas que incluyan crisis.

CAPÍTULO 15

Bibliografía

- [1] E. F. Fama y K. R. French, “A five-factor asset pricing model,” *Journal of Financial Economics*, vol. 116, n.º 1, págs. 1-22, 2015.
- [2] S. Hochreiter y J. Schmidhuber, “Long Short-Term Memory,” *Neural Computation*, vol. 9, n.º 8, págs. 1735-1780, 1997. DOI: [10.1162/neco.1997.9.8.1735](https://doi.org/10.1162/neco.1997.9.8.1735)
- [3] R. K. Narang, *Inside the Black Box: A Simple Guide to Quantitative and High Frequency Trading*. Hoboken, NJ: John Wiley & Sons, 2013.
- [4] R. Cont, “Empirical properties of asset returns: stylized facts and statistical issues,” *Quantitative Finance*, vol. 1, n.º 2, págs. 223-236, 2001.
- [5] X. Ding, Y. Zhang, T. Liu y J. Duan, “Deep Learning for Event-Driven Stock Prediction,” en *Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence (IJCAI)*, National University of Singapore, 2015.
- [6] C. Mascioli, A. Gu, Y. Wang et al., “A Financial Market Simulation Environment for Trading Agents Using Deep Reinforcement Learning,” University of Michigan, inf. téc., 2024.
- [7] Z. Yao, L. Zheng, M. Thomas e I. Florescu, “Reinforcement Learning in Agent-Based Market Simulation: Unveiling Realistic Stylized Facts and Behavior,” Stevens Institute of Technology, Hoboken, NJ, inf. téc., 2024.
- [8] G. Box y G. Jenkins, *Time Series Analysis: Forecasting and Control*. San Francisco: Holden-Day, 1970.
- [9] R. Engle, “Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom Inflation,” *Econometrica*, vol. 50, n.º 4, págs. 987-1007, 1982.
- [10] T. Bollerslev, “Generalized Autoregressive Conditional Heteroskedasticity,” *Journal of Econometrics*, vol. 31, n.º 3, págs. 307-327, 1986.
- [11] C. Sims, “Macroeconomics and Reality,” *Econometrica*, vol. 48, n.º 1, págs. 1-48, 1980.
- [12] S. Mallat, *A Wavelet Tour of Signal Processing*. San Diego: Academic Press, 1998.

- [13] N. Huang et al., "The Empirical Mode Decomposition and the Hilbert Spectrum for Nonlinear and Nonstationary Time Series Analysis," *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, vol. 454, n.^o 1971, págs. 903-995, 1998.
- [14] D. Nasiopoulos, K. Roumeliotis, D. Sakas et al., "Financial Sentiment Analysis and Classification: A Comparative Study of Fine-Tuned Deep Learning Models," *International Journal of Financial Studies*, 2025.
- [15] F. Sorouralsadat y H. Yuheng, "Reinforcement Learning in Agent-Based Market Simulation: Unveiling Realistic Stylized Facts and Behavior," University of Illinois, Chicago, IL, inf. téc., 2024.
- [16] S. Russell y P. Norvig, *Artificial Intelligence: A Modern Approach*, 3rd. Prentice Hall, 2010.
- [17] S. Franklin y A. Graesser, "Is it an Agent, or Just a Program? A Taxonomy for Autonomous Agents," en *Proceedings of the Third International Workshop on Agent Theories, Architectures, and Languages*, 1996.
- [18] Z. Zhang, S. Zohren y S. Roberts, "Deep learning for portfolio optimization," *The Journal of Financial Data Science*, vol. 2, n.^o 4, págs. 8-20, 2020.
- [19] R. S. Tsay, *Analysis of Financial Time Series*. Hoboken, NJ: John Wiley & Sons, 2010.
- [20] J. Murphy, *Technical Analysis of the Financial Markets: A Comprehensive Guide to Trading Methods and Applications*. New York: Penguin, 1999.
- [21] W. McKinney, *pandas: a foundational Python library for data analysis and statistics*, Python for High Performance and Scientific Computing / pandas project, 2011.
- [22] R. Little y D. Rubin, *Statistical Analysis with Missing Data*. Hoboken, NJ: John Wiley & Sons, 2019.
- [23] J. Campbell y S. Thompson, "Predicting excess stock returns out of sample," *Journal of Financial Economics*, vol. 79, n.^o 2, págs. 375-410, 2006.
- [24] S. Achelis, *Technical Analysis from A to Z*. New York: McGraw-Hill Professional, 2001.
- [25] R. Cleveland, W. Cleveland, J. McRae e I. Terpenning, "STL: A seasonal-trend decomposition," *Journal of Official Statistics*, vol. 6, n.^o 1, págs. 3-73, 1990.
- [26] I. Daubechies, *Ten Lectures on Wavelets*. Philadelphia, PA: SIAM, 1992.
- [27] R. Gencay, F. Selçuk y B. Whitcher, *An Introduction to Wavelets and Other Filtering Methods in Finance and Economics*. San Diego, CA: Academic Press/Elsevier, 2001.
- [28] A. Geron, *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*. Sebastopol, CA: O'Reilly Media, 2019.
- [29] T. Castro, A. Mora y J. Suriñach, "Comportamiento de los contrastes ADF, PP y KPSS al trabajar con series ajustadas de estacionalidad," *Questio: Quaderns d'Estadística, Sistemes, Informàtica i Investigació Operativa, ISSN 0210-8054*, Vol. 25, Nº. 1, 2001, pags. 19-46, ene. de 2001.
- [30] A. Kraskov, H. St.ögbauer y P. Grassberger, "Estimating mutual information," *Physical Review E*, vol. 69, n.^o 6, pág. 066 138, 2004.

- [31] S. Arlot y A. Celisse, “A survey of cross-validation procedures for model selection,” *Statistics Surveys*, vol. 4, págs. 40-79, 2010.
- [32] T. Fischer y C. Krauss, “Deep learning with long short-term memory networks for financial market predictions,” *European Journal of Operational Research*, vol. 270, n.º 2, págs. 654-669, 2018.
- [33] J. Bollen, H. Mao y X. Zeng, “Twitter mood predicts the stock market,” *Journal of Computational Science*, vol. 2, n.º 1, págs. 1-8, 2011.
- [34] I. Tesla, “Form 8-K: Current Report Pursuant to Section 13 or 15(d) of the Securities Exchange Act of 1934,” U.S. Securities y Exchange Commission, Washington, D.C., Current Report, ago. de 2025, Retrieved from <https://www.sec.gov/Archives/edgar/data/0001318605/>.
- [35] T. Sprenger, A. Tumasjan, P. Sandner e I. Welpe, “Tweets and trades: the information content of stock microblogs,” *European Financial Management*, vol. 20, n.º 5, págs. 926-957, 2014.
- [36] S. Stieglitz y L. Dang-Xuan, “Emotions and information diffusion in social media—sentiment of microblogs and sharing behavior,” *Journal of Management Information Systems*, vol. 29, n.º 4, págs. 217-248, 2013.
- [37] M. Cha, H. Haddadi, F. Benevenuto y K. Gummadi, “Measuring user influence in Twitter: The million follower fallacy,” en *Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media (ICWSM)*, 2010, págs. 30-37.
- [38] C. D. Manning, M. Surdeanu, J. Bauer, J. Finkel, S. J. Bethard y D. McClosky, “The Stanford CoreNLP natural language processing toolkit,” en *Proceedings of the ACL System Demonstrations*, 2014, págs. 55-60.
- [39] H. Saif, M. Fernandez, Y. He y H. Alani, “On stopwords, filtering and data sparsity for sentiment analysis of Twitter,” en *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC)*, 2014, págs. 810-817.
- [40] G. A. Miller, “WordNet: a lexical database for English,” *Communications of the ACM*, vol. 38, n.º 11, págs. 39-41, 1995.
- [41] R. Zafarani, M. A. Abbasi y H. Liu, *Social Media Mining: An Introduction*. Cambridge: Cambridge University Press, 2014.
- [42] C. J. Hutto y E. Gilbert, “VADER: A parsimonious rule-based model for sentiment analysis of social media text,” en *Proceedings of the Eighth International AAAI Conference on Weblogs and Social Media (ICWSM)*, 2014, págs. 216-225.
- [43] F. N. Ribeiro, M. Araújo, P. Gonçalves, M. A. Gonçalves y F. Benevenuto, “Senti-Bench — a benchmark comparison of state-of-the-practice sentiment analysis methods,” *EPJ Data Science*, vol. 5, n.º 1, págs. 1-29, 2016.
- [44] P. J. Rousseeuw y C. Croux, “Alternatives to the median absolute deviation,” *Journal of the American Statistical Association*, vol. 88, n.º 424, págs. 1273-1283, 1993.
- [45] P. C. Tetlock, “Giving content to investor sentiment: The role of media in the stock market,” *The Journal of Finance*, vol. 62, n.º 3, págs. 1139-1168, 2007.
- [46] J. Clausen e Y. Olteanu, “Tesla as a start-up in the automotive industry. From candidate for bankruptcy to gamechanger,” inf. téc., dic. de 2020. DOI: [10.13140/RG.2.2.32294.40002](https://doi.org/10.13140/RG.2.2.32294.40002).

- [47] J. Y. Campbell, A. W. Lo y A. C. MacKinlay, *The Econometrics of Financial Markets*. Princeton, NJ: Princeton University Press, 1997.
- [48] D. Sculley et al., “Hidden technical debt in machine learning systems,” en *Advances in Neural Information Processing Systems*, vol. 28, 2015.
- [49] T. Shumway, “The delisting bias in CRSP data,” *The Journal of Finance*, vol. 52, n.º 1, págs. 327-340, 1997.
- [50] F. X. Diebold, *Financial Econometrics*. Princeton, NJ: Princeton University Press, 2015.
- [51] P. Embrechts, A. McNeil y D. Straumann, “Correlation and dependence in risk management: properties and pitfalls,” en *Risk Management: Value at Risk and Beyond*, Cambridge: Cambridge University Press, 2002, págs. 176-223.
- [52] J. Bai y S. Ng, “Determining the number of factors in approximate factor models,” *Econometrica*, vol. 70, n.º 1, págs. 191-221, 2002.
- [53] Y. Benjamini e Y. Hochberg, “Controlling the false discovery rate: a practical and powerful approach to multiple testing,” *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 57, n.º 1, págs. 289-300, 1995.
- [54] C. W. J. Granger, “Investigating causal relations by econometric models and cross-spectral methods,” *Econometrica*, vol. 37, n.º 3, págs. 424-438, 1969.
- [55] R. F. Engle y C. W. J. Granger, “Co-integration and error correction: representation, estimation, and testing,” *Econometrica*, vol. 55, n.º 2, págs. 251-276, 1987.
- [56] I. T. Jolliffe, *Principal Component Analysis*, 2nd. New York: Springer, 2002.
- [57] G. Connor y R. Korajczyk, “A test for the number of factors in an approximate factor model,” *The Journal of Finance*, vol. 48, n.º 4, págs. 1263-1291, 1993.
- [58] T. Hastie, R. Tibshirani y J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. New York: Springer, 2009.
- [59] D. C. Montgomery, *Design and Analysis of Experiments*. Hoboken, NJ: John Wiley & Sons, 2017.
- [60] S. Kaufman, S. Rosset, C. Perlich et al., “Leakage in data mining: Formulation, detection, and avoidance,” en *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2011, págs. 556-563.
- [61] J. Demšar, “Statistical comparisons of classifiers over multiple data sets,” *Journal of Machine Learning Research*, vol. 7, págs. 1-30, 2006.
- [62] T. Baltrušaitis, C. Ahuja y L.-P. Morency, “Multimodal machine learning: A survey and taxonomy,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, n.º 2, págs. 423-443, 2018.
- [63] J. Heckman, “Sample selection bias as a specification error,” *Econometrica*, vol. 47, n.º 1, págs. 153-161, 1979.
- [64] S. Van Buuren, *Flexible Imputation of Missing Data*. Boca Raton, FL: CRC Press, 2018.
- [65] C. Beleites, U. Neugebauer, T. Bocklitz, C. Krafft y J. Popp, “Sample size planning for classification models,” *Analytica Chimica Acta*, vol. 760, págs. 25-33, 2013.

- [66] D. Bailey, J. Borwein, M. Lopez de Prado et al., “Stock portfolio design and backtest overfitting,” *The Journal of Investment Management*, vol. 15, n.º 1, págs. 75-87, 2017.
- [67] R. Chaudhary, “Advanced Stock Market Prediction Using Long Short-Term Memory Networks: A Comprehensive Deep Learning Framework,” *arXiv preprint arXiv:2505.05325v1*, mayo de 2025, cs.CE. dirección: <https://arxiv.org/abs/2505.05325v1>.
- [68] S. Dube, “Machine Learning for Stock Price Forecasting: LSTM vs Transformer Approaches,” *International Journal of Technology Management and Humanities*, vol. 9, n.º 04, págs. 152-171, 2023. DOI: [10.21590/ijtmh.20230904011](https://doi.org/10.21590/ijtmh.20230904011).
- [69] J. Xiao, S. Bi y T. Deng, “Comparative Analysis of LSTM, GRU, and Transformer Models for Stock Price Prediction,” *arXiv preprint arXiv:2411.05790*, 2024. dirección: <https://arxiv.org/abs/2411.05790>.
- [70] F. A. Gers, J. Schmidhuber y F. Cummins, “Learning to Forget: Continual Prediction with LSTM,” *Neural Computation*, vol. 12, n.º 10, págs. 2451-2471, 2000. DOI: [10.1162/089976600300015015](https://doi.org/10.1162/089976600300015015).
- [71] D. Bahdanau, K. Cho e Y. Bengio, “Neural Machine Translation by Jointly Learning to Align and Translate,” *arXiv preprint arXiv:1409.0473*, 2016, Version 7, last revised 19 May 2016. dirección: <https://doi.org/10.48550/arXiv.1409.0473>.
- [72] M.-T. Luong, H. Pham y C. D. Manning, “Effective Approaches to Attention-based Neural Machine Translation,” *arXiv preprint arXiv:1508.04025*, 2015, EMNLP 2015 camera-ready version. DOI: [10.48550/arXiv.1508.04025](https://doi.org/10.48550/arXiv.1508.04025). dirección: <https://arxiv.org/abs/1508.04025>.
- [73] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever y R. Salakhutdinov, “Dropout: A Simple Way to Prevent Neural Networks from Overfitting,” *Journal of Machine Learning Research*, vol. 15, n.º 56, págs. 1929-1958, 2014. dirección: <http://jmlr.org/papers/v15/srivastava14a.html>.
- [74] W. Sun, J. Mei, S. Liu, C. Yuan y J. Zhao, “Research on deep learning model for stock prediction by integrating frequency domain and time series features,” *Scientific Reports*, vol. 15, n.º 1, pág. 30386, 2025. DOI: [10.1038/s41598-025-14872-6](https://doi.org/10.1038/s41598-025-14872-6). dirección: <https://doi.org/10.1038/s41598-025-14872-6>.
- [75] R. Pardo, “The Evaluation and Optimization of Trading Strategies,” *John Wiley & Sons*, 2008.
- [76] A. Tsantekidis, N. Passalis, A. Tefas, J. Kannainen, M. Gabbouj y A. Iosifidis, “Forecasting Stock Prices from the Limit Order Book using Convolutional Neural Networks,” *IEEE Conference on Business Informatics (CBI)*, págs. 7-12, 2017. DOI: [10.1109/CBI.2017.8237231](https://doi.org/10.1109/CBI.2017.8237231).
- [77] P. J. Huber, “Robust Estimation of a Location Parameter,” *The Annals of Mathematical Statistics*, vol. 35, n.º 1, págs. 73-101, 1964. DOI: [10.1214/aoms/1177703732](https://doi.org/10.1214/aoms/1177703732).
- [78] L. Rosasco, E. D. Vito, A. Caponnetto, M. Piana y A. Verri, “Are Loss Functions All the Same?” *Neural Computation*, vol. 16, n.º 5, págs. 1063-1076, 2004. DOI: [10.1162/089976604773135104](https://doi.org/10.1162/089976604773135104).
- [79] I. Loshchilov y F. Hutter, “Decoupled Weight Decay Regularization,” en *International Conference on Learning Representations (ICLR)*, 2019. dirección: <https://openreview.net/forum?id=Bkg6RiCqY7>.

- [80] L. N. Smith, “Cyclical Learning Rates for Training Neural Networks,” *arXiv preprint arXiv:1506.01186*, 2017. dirección: <https://arxiv.org/abs/1506.01186>.
- [81] R. Pascanu, T. Mikolov e Y. Bengio, “On the Difficulty of Training Recurrent Neural Networks,” *Proceedings of the 30th International Conference on Machine Learning*, vol. 28, págs. 1310-1318, 2013. dirección: <http://proceedings.mlr.press/v28/pascanu13.pdf>.
- [82] L. Prechelt, “Early Stopping - But When?” *Neural Networks: Tricks of the Trade*, págs. 55-69, 1998. DOI: 10.1007/3-540-49430-8_3.
- [83] G. F. Hughes, “On the Mean Accuracy of Statistical Pattern Recognizers,” *IEEE Transactions on Information Theory*, vol. 14, n.º 1, págs. 55-63, 1968. DOI: <10.1109/TIT.1968.1054102>.
- [84] T. Chen y C. Guestrin, “XGBoost: A scalable tree boosting system,” en *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Francisco, CA, 2016, págs. 785-794.
- [85] A. Cooper, F. Gimeno-Gascon y C. Woo, “Gender and Organizational Performance: Determinants of Small Business Survival and Success,” *Journal of Business Venturing*, vol. 6, n.º 5, págs. 315-332, 1991. DOI: [10.1016/0883-9026\(91\)90002-2](10.1016/0883-9026(91)90002-2).
- [86] M. Magdon-Ismail, A. F. Atiya, A. Pratap e Y. S. Abu-Mostafa, “On the Maximum Drawdown of a Brownian Motion,” *Journal of Applied Probability*, vol. 41, n.º 1, págs. 147-161, 2004. DOI: <10.1239/jap/1077134674>.
- [87] J. C. Hull, *Options, Futures, and Other Derivatives*, 10.^a ed. Pearson, 2018, ISBN: 9780134472089.
- [88] S. E. Shreve, *Stochastic Calculus for Finance II: Continuous-Time Models* (Springer Finance). Springer, 2004, ISBN: 9780387401010.
- [89] M. López de Prado, *Machine Learning for Asset Managers*. Cambridge University Press, 2020.
- [90] I. Goodfellow, Y. Bengio y A. Courville, *Deep Learning*. MIT Press, 2016.
- [91] S. Hochreiter y J. Schmidhuber, “Long Short-Term Memory,” *Neural Computation*, vol. 9, n.º 8, págs. 1735-1780, 1997.
- [92] C. R. Harvey, Y. Liu y H. Zhu, “...and the cross-section of expected returns,” *The Review of Financial Studies*, vol. 29, n.º 1, págs. 5-68, 2016.
- [93] M. Baker y J. Wurgler, “Investor sentiment and the cross-section of stock returns,” *The Journal of Finance*, vol. 61, n.º 4, págs. 1645-1680, 2006.
- [94] D. Lyndell, *Elon Musk Tweets 2010 to 2025 (April)*, <https://www.kaggle.com/dadalynell/elon-musk-tweets-2010-to-2025-april>, Accessed: 2025-11-06, 2025.
- [95] I. Goodfellow, Y. Bengio y A. Courville, *Deep Learning*. MIT Press, 2016, ISBN: 978-0262035613. dirección: <http://www.deeplearningbook.org>.
- [96] J. L. Schafer y J. W. Graham, “Missing data: our view of the state of the art,” *Psychological Methods*, vol. 7, n.º 2, págs. 147-177, 2002.

CAPÍTULO 16

Anexos

16.1. Gráficos adicionales

En esta sección se colocan figuras y gráficos suplementarios que complementan el cuerpo principal de la tesis.

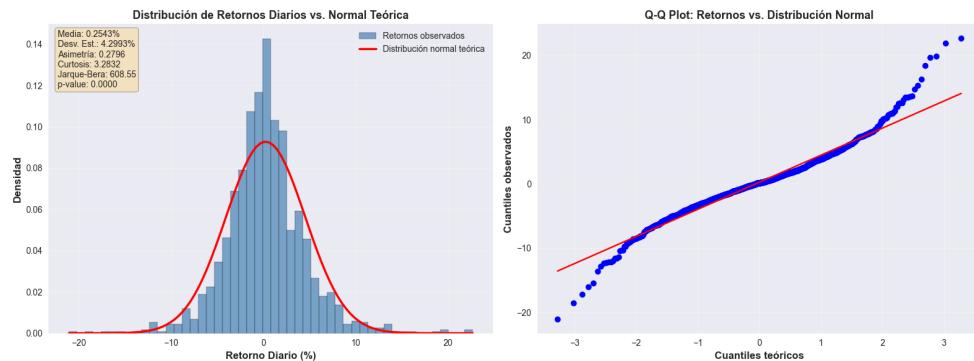


Figura 9: Precio, retornos y volatilidad (ventana 21 días).

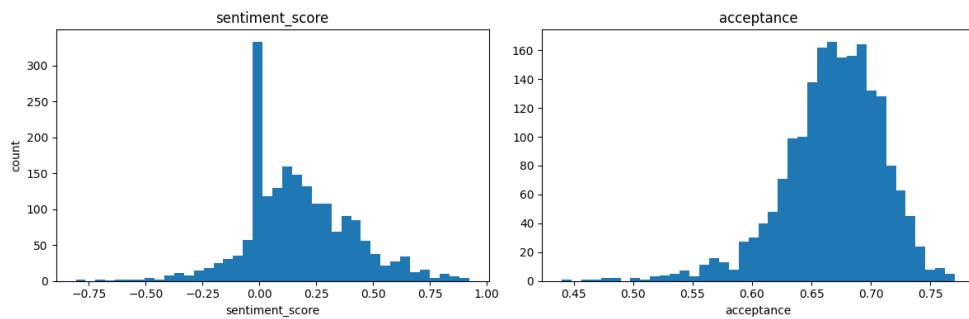


Figura 10: Distribución de puntuaciones de sentimiento y métrica de engagement.

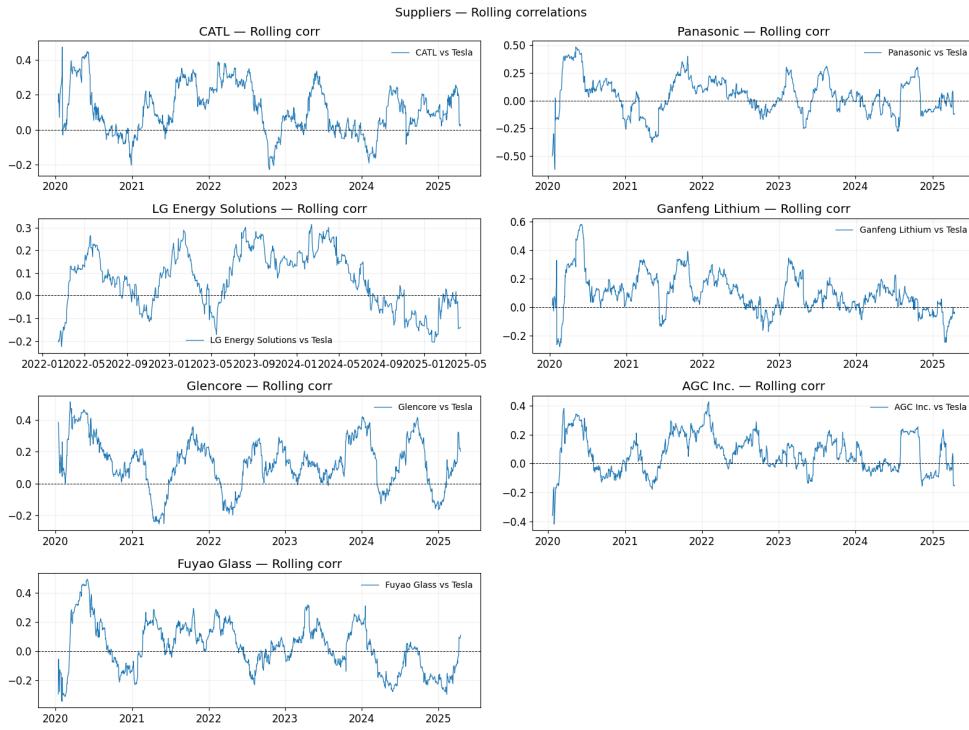


Figura 11: Correlaciones rolling entre proveedores seleccionados y Tesla.

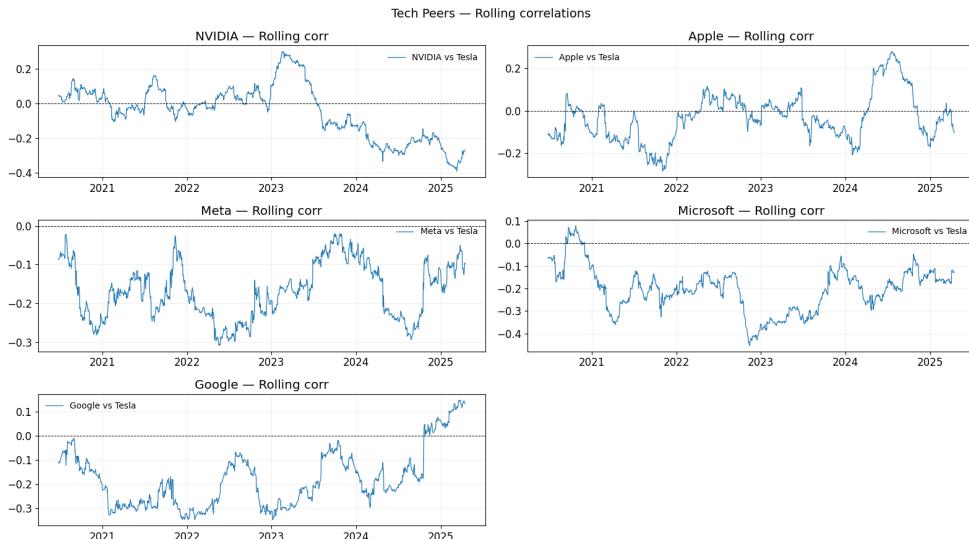


Figura 12: Correlaciones rolling entre tech peers y Tesla.

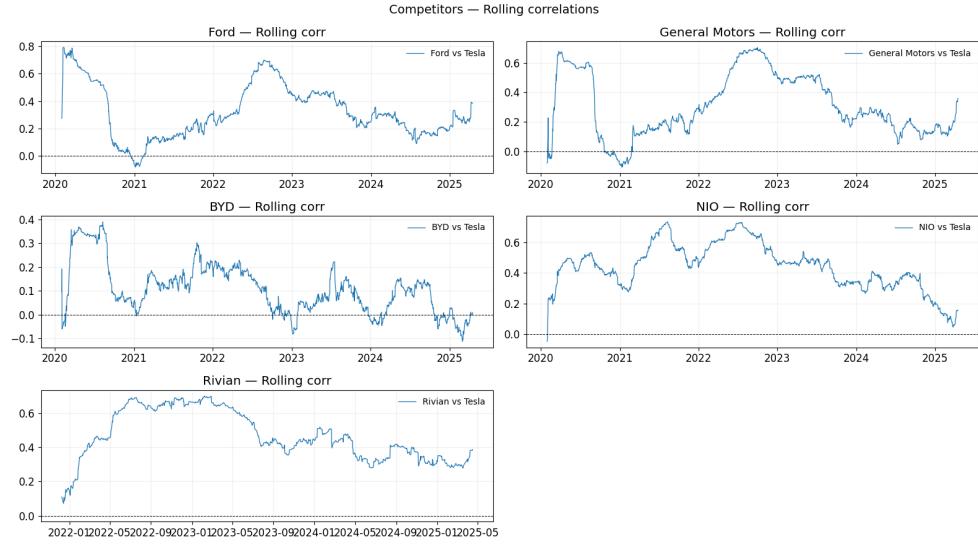


Figura 13: Correlaciones rolling entre competidores automotrices y Tesla.

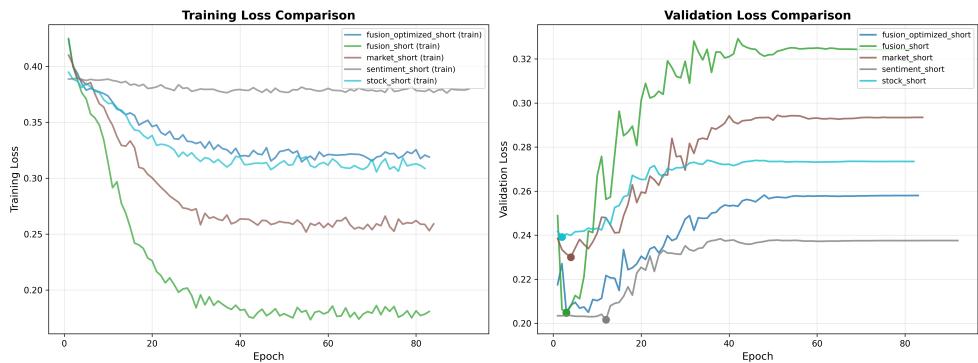


Figura 14: Pérdida: entrenamiento y validación. Horizonte corto (1 día).

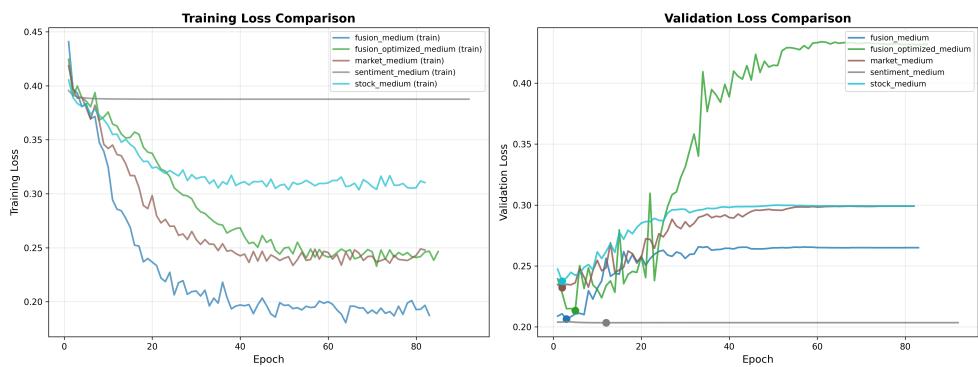


Figura 15: Pérdida: entrenamiento y validación. Horizonte medio (5 días).

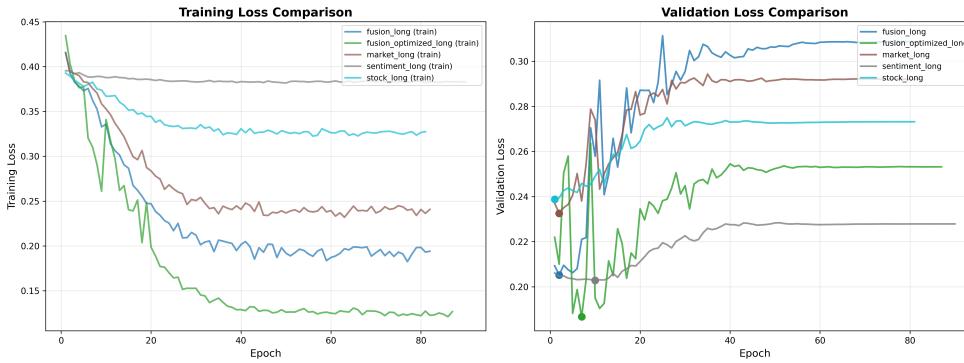


Figura 16: Pérdida: entrenamiento y validación. Horizonte largo (21 días).

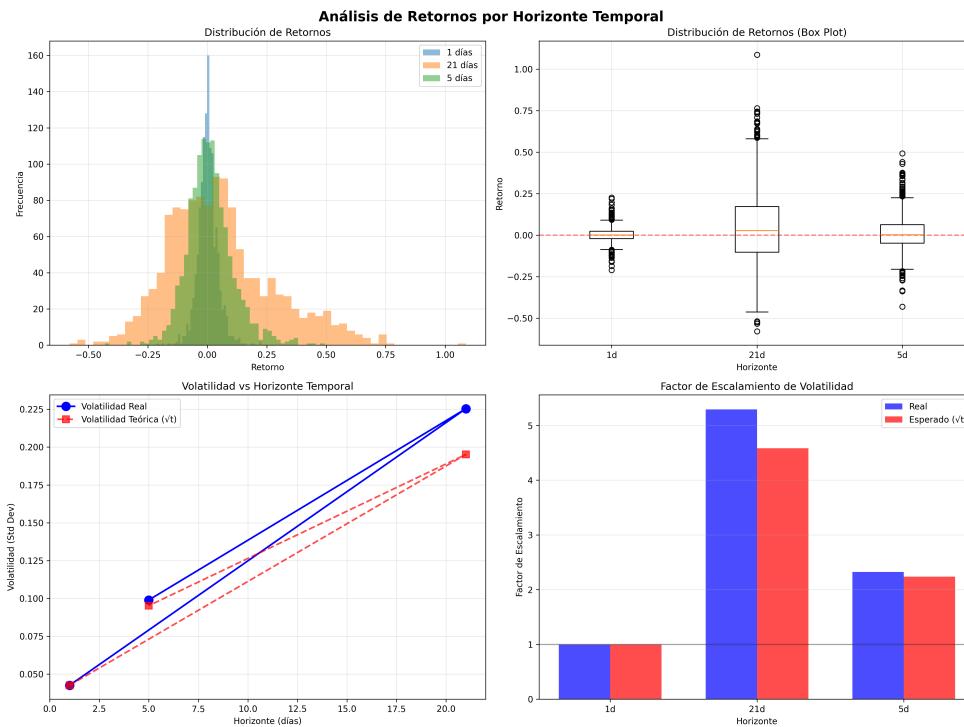


Figura 17: Distribuciones de retornos y análisis de volatilidad por horizonte (1, 5, 21 días).

16.2. Repositorio del proyecto

- **Repositorio (GitHub):**

<https://github.com/TheDeloz-v2/multimodal-financial-prediction-agent>

- **Estructura breve del repositorio:**

- **data/** – datos crudos y procesados
- **src/** – código fuente del pipeline y modelos
- **models/** – modelos generados

- `notebooks/` – cuadernos exploratorios y experimentos reproducibles
- `results/` – tablas, reportes y artefactos generados por experimentos
- `figures/` – imágenes y figuras
- `scripts/` – utilidades y scripts para procesamiento, entrenamiento y evaluación
- `docs/` – documentación adicional
- `README.md, LICENSE` – documentación general y licencia del proyecto

CAPÍTULO 17

Glosario

ADF Augmented Dickey-Fuller.

AR Autoregressive.

ATR Average True Range.

CEO Chief Executive Officer.

CNN Convolutional Neural Network.

DA Directional Accuracy.

DNN Deep Neural Networks.

EMD Empirical Mode Decomposition.

FDR False Discovery Rate.

GRU Gated Recurrent Unit.

HFT High Frequency Trading.

HR Hit Rate.

KPSS Kwiatkowski-Phillips-Schmidt-Shin.

LSTM Long Short-Term Memory.

MA Moving Average.

MACD Moving Average Convergence Divergence.

MAE Mean Absolute Error.

MBG Movimiento Browniano Geométrico.

MDD Maximum Drawdown.

NLP Natural Language Processing.

OHLCV Open, High, Low, Close, Volume.

PCA Principal Component Analysis.

PF Profit Factor.

POS Part-of-Speech.

RMSE Root Mean Squared Error.

RNN Recurrent Neural Networks.

RSI Relative Strength Index.

VADER Valence Aware Dictionary and sEntiment Reasoner.

VAR Vector Autoregression.