

Lic. Héctor Juliarena hector.juliarena@docentes.udemm.edu.ar

Lic. Rodolfo González rodolfo.gonzalez@docentes.udemm.edu.ar

Dra. Jesica Allende jesica.allende@docentes.udemm.edu.ar

Grupo de Investigación en Tecnología Financiera

Facultad de Administración y Economía

Universidad de la Marina Mercante

PREDICCIÓN DE PRECIOS DE INSTRUMENTOS FINANCIEROS MEDIANTE INTELIGENCIA ARTIFICIAL

Avances, desafíos y perspectivas de un proyecto de investigación aplicada

Resumen

El presente artículo describe los avances de un proyecto de investigación orientado al desarrollo de una aplicación capaz de predecir el precio futuro de instrumentos financieros mediante técnicas de machine learning (ML) y deep learning (DL). A lo largo de dos años de trabajo, logramos construir una infraestructura técnica robusta, recopilar y procesar datos históricos de mercado, e implementar un primer modelo predictivo basado en XGBoost con optimización de hiperparámetros. Se analizan los resultados obtenidos, las limitaciones identificadas —en

particular el sobreajuste del modelo y la ausencia de variables sentimentales— y se traza un plan de continuidad orientado a incorporar datos no estructurados y arquitecturas de DL más sofisticadas. El proyecto articula formación universitaria, investigación aplicada y desarrollo tecnológico en el campo de las finanzas computacionales.

Palabras clave: machine learning, deep learning, predicción financiera, XGBoost, series temporales, mercados financieros.

PREDICTION OF FINANCIAL INSTRUMENT PRICES USING ARTIFICIAL INTELLIGENCE

Advances, challenges, and perspectives of an applied research project

Abstract

This article describes the progress of a research project aimed at developing an application capable of predicting future prices of financial instruments using machine learning (ML) and deep learning (DL) techniques. Over two years of work, we built a robust technical infrastructure, collected

and processed historical market data, and implemented an initial predictive model based on XGBoost with hyperparameter optimization. The results obtained, identified limitations—particularly model overfitting and the lack of sentiment variables—are analyzed, and a roadmap is outlined to incorporate unstructured data and more advanced DL architectures. The project

integrates university training, applied research, and technological development in the field of computational finance.

1. Introducción

Los mercados financieros constituyen uno de los pilares del desarrollo económico moderno. Desde la fundación de la Bolsa de Valores de Ámsterdam en 1602¹, estos mercados han evolucionado de manera continua, facilitando la transferencia de capital, promoviendo la inversión y permitiendo la acumulación de riqueza a escala global. Su funcionamiento eficiente resulta clave para que los precios reflejen de manera fidedigna la información disponible.

La historia también revela la fragilidad de estos sistemas. La Gran Depresión de 1929, desencadenada por el colapso del mercado de valores, evidenció la capacidad de los mercados financieros para incidir profundamente en el bienestar social. Sin embargo, la progresiva sofisticación regulatoria y la resiliencia demostrada en sucesivas crisis han contribuido a mantener la estabilidad del sistema financiero global.

En este contexto, la capacidad de anticipar el comportamiento de los precios de los activos financieros reviste una importancia estratégica. La irrupción de la inteligencia artificial y, en particular, de las técnicas de aprendizaje automático (*machine learning*) y aprendizaje profundo (*deep learning*), ha abierto nuevas posibilidades en

Keywords: machine learning, deep learning, financial prediction, XGBoost, time series, financial markets.

este campo, superando las limitaciones de los modelos estadísticos tradicionales.

El presente artículo describe el estado de avance de un proyecto desarrollado en la Facultad de Administración y Economía de la UdeMM, cuyo objetivo central es diseñar y construir una aplicación que prediga el precio de instrumentos financieros —acciones y criptomonedas— haciendo uso de estas tecnologías. Se exponen los fundamentos teóricos, la metodología empleada, los principales logros alcanzados en dos años de trabajo y los desafíos que orientan la continuidad de la investigación.

2. Estado del conocimiento

Las aplicaciones orientadas a la predicción del valor de activos financieros han ganado relevancia creciente en las últimas décadas, impulsadas por los avances en capacidad de cómputo y el desarrollo de algoritmos de aprendizaje automático y aprendizaje profundo (López de Prado, 2018).

2.1. Inicios y evolución

Los primeros intentos sistemáticos de predicción de precios financieros se remontan a las décadas de 1970 y 1980, con la aplicación de modelos estadísticos clásicos como la regresión lineal y el análisis de series temporales —en particular, el modelo ARIMA (modelo autorregresivo integrado de media móvil)²—. Estos enfoques descansaban en la premisa de que los precios seguían patrones

¹La Bolsa de Valores de Ámsterdam (Amsterdamsche Beurs) fue fundada en 1602 por la Compañía Neerlandesa de las Indias Orientales (VOC) y es considerada el primer mercado bursátil moderno del mundo.

²El modelo ARIMA fue sistematizado en: Box, G. E. P., & Jenkins, G. M. (1970). *Time Series Analysis: Forecasting and Control*. Holden-Day.

susceptibles de ser capturados y proyectados hacia el futuro (Chan, 2013).

Con la expansión de la informática y el incremento en la capacidad de procesamiento, emergieron algoritmos más sofisticados capaces de gestionar grandes volúmenes de datos y detectar patrones complejos. A partir de la década de 1990, las redes neuronales artificiales marcaron un hito al introducir mejoras sustanciales en la capacidad predictiva, al capturar relaciones no lineales en los datos financieros que los modelos anteriores no podían representar adecuadamente (Bishop, 2006).

2.2. Técnicas de aprendizaje profundo en finanzas

Los modelos de predicción más avanzados de la actualidad recurren a arquitecturas de aprendizaje profundo como las redes neuronales convolucionales y las redes neuronales recurrentes, entre las cuales destacan las redes de memoria de largo y corto plazo (LSTM, por sus siglas en inglés)³ por su capacidad para retener patrones a largo plazo en series temporales. Estas técnicas integran no solo precios históricos, sino también fuentes no estructuradas —noticias financieras, publicaciones en redes sociales— con el propósito de capturar el estado de ánimo del mercado.

Una tendencia emergente es el uso de modelos de aprendizaje por refuerzo (*reinforcement learning*), en los cuales un agente toma decisiones simulando millones de escenarios posibles. En condiciones normales, algunos algoritmos alcanzan tasas

de acierto de entre el 60 % y el 80 % en predicciones de corto plazo, aunque esto no siempre se traduce en rentabilidad efectiva, dado el peso de los costos de transacción.

2.3. Herramientas tecnológicas y costos de implementación

Python se consolida como el lenguaje de programación de referencia para este tipo de desarrollos, gracias a su ecosistema de bibliotecas especializadas —TensorFlow, Keras, PyTorch, scikit-learn—. R, C++ y Java complementan este panorama en contextos de análisis estadístico avanzado, aplicaciones de alto rendimiento y sistemas empresariales, respectivamente. El costo de implementación varía de algunos miles de dólares para modelos básicos a varios millones para aplicaciones comerciales de alta complejidad (López de Prado, 2020).

3. Fundamentación teórica

3.1. La Hipótesis de los Mercados Eficientes y sus límites

La base conceptual del proyecto se inscribe en la tensión entre la Hipótesis de los Mercados Eficientes (HME)⁴ y las evidencias empíricas que la cuestionan. La HME sostiene que en un mercado eficiente los precios incorporan toda la información disponible, lo que haría imposible superar al mercado de manera consistente sin información privilegiada.

Sin embargo, la existencia documentada de anomalías —efectos calendario, volatilidad excesiva, comportamiento en manada— ha erosionado

³Las redes LSTM fueron introducidas en: Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, 9(8), 1735-1780.

⁴La Hipótesis de los Mercados Eficientes fue formulada en: Fama, E. F. (1970). Efficient Capital Markets: A Review of Empirical Work. *The Journal of Finance*, 25(2), 383-417.

la validez universal de esta hipótesis. En este marco, el uso de inteligencia artificial y aprendizaje profundo reabre la posibilidad de identificar oportunidades predictivas mediante el análisis de patrones complejos en datos históricos y fuentes externas, inaccesibles a los métodos estadísticos tradicionales (López de Prado, 2018).

3.2. Aprendizaje automático y aprendizaje profundo en la predicción financiera

El recorrido del aprendizaje automático en finanzas siguió una trayectoria ascendente: de las técnicas lineales iniciales hacia modelos cada vez más complejos a medida que la disponibilidad de datos y la capacidad de cómputo lo permitieron. La incorporación del aprendizaje profundo representó un salto cualitativo: arquitecturas como las redes neuronales recurrentes y LSTM permiten capturar dependencias temporales de largo alcance en series de precios, superando las limitaciones de los modelos que asumen independencia entre observaciones (De La Torre, 2021).

La ventaja comparativa del aprendizaje profundo radica, adicionalmente, en su capacidad para integrar datos no estructurados —noticias, publicaciones en redes sociales, comunicados corporativos— como insumos del proceso predictivo. Investigaciones recientes demuestran que la incorporación de este tipo de información mejora significativamente la precisión de los modelos respecto del uso exclusivo de datos históricos estructurados.

4. Objetivos e hipótesis de la investigación

El objetivo general del proyecto consiste en desarrollar una aplicación basada en aprendizaje automático y aprendizaje profundo capaz de predecir con precisión los

precios futuros de instrumentos financieros, especialmente acciones y criptomonedas.

Los objetivos específicos son: (a) implementar y evaluar modelos como redes neuronales recurrentes y LSTM para el análisis de series temporales financieras; (b) medir el desempeño de los modelos mediante el error cuadrático medio (ECM) y el error absoluto porcentual medio (EAPM); y (c) integrar variables de carácter económico y sentimental junto a los datos históricos de precios.

Las hipótesis que orientan la investigación sostienen que los modelos de aprendizaje profundo mejorarán la precisión predictiva respecto de los modelos tradicionales, y que la incorporación de datos no estructurados incrementará la exactitud de las predicciones. Las preguntas de investigación indagan acerca de qué factores externos inciden más en la precisión y cómo varía dicha precisión entre los distintos modelos considerados.

5. Metodología

El proyecto se estructuró en seis grandes etapas: (1) definición del problema y requisitos, (2) recopilación y preparación de datos, (3) desarrollo del modelo, (4) construcción de la aplicación, (5) pruebas y validación, y (6) despliegue y mantenimiento.

5.1. Definición del problema y recopilación de datos

La primera etapa estableció con precisión los instrumentos financieros objeto de análisis, los requisitos funcionales —frecuencia de predicción, interfaz de usuario— y los no funcionales —velocidad, escalabilidad—, considerando el marco regulatorio aplicable. La recopilación de

datos se realizó mediante interfaces de programación de aplicaciones (API) de servicios reconocidos: Yahoo Finance, Alpha Vantage, CoinMarketCap y CoinGecko, obteniendo series históricas de precios, volúmenes de transacción e indicadores financieros.

Particular énfasis se puso en la preparación de los datos: limpieza, normalización, escalado y generación de nuevas variables —promedios móviles, índice de fuerza relativa (IFR) y convergencia/divergencia de medias móviles (MACD)— que potencialmente enriquecen la capacidad predictiva del modelo.

5.2. Desarrollo del modelo

Se exploraron tanto enfoques de aprendizaje automático clásico —regresión lineal, árboles de decisión, bosques aleatorios— como arquitecturas de aprendizaje profundo orientadas a series temporales, incluyendo redes neuronales recurrentes, LSTM y unidades recurrentes con compuerta (GRU, por sus siglas en inglés). Como primer modelo implementado se optó por un regresor XGBoost (potenciación extrema del gradiente)⁵ con optimización de hiperparámetros mediante búsqueda exhaustiva en cuadrícula (*GridSearchCV*). El entrenamiento se organizó en conjuntos de entrenamiento, validación y prueba, con técnicas de regularización y validación cruzada. La evaluación se realizó mediante la raíz del error cuadrático medio (RECM), el error absoluto medio (EAM) y el error absoluto porcentual medio (EAPM).

5.3. Arquitectura de la aplicación

⁵XGBoost fue presentado en: Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. Proceedings of the 22nd ACM SIGKDD, 785-794.

En la etapa actual, la arquitectura del sistema se orienta principalmente al backend, priorizando la ejecución y validación de modelos predictivos en entornos controlados. Las pruebas se realizaron mediante interfaces como notebooks (por ejemplo, Jupyter) y línea de comandos (CLI), sin una capa frontend dedicada. La infraestructura se basa en esquemas de virtualización completa (KVM) y virtualización a nivel de software (Docker), lo que permitió evaluar distintos algoritmos en entornos reproducibles y aislados. En este contexto, el backend centraliza la gestión de datos y la ejecución de modelos. La incorporación de una arquitectura cliente-servidor con interfaz de usuario será considerada en etapas futuras, en función de los requerimientos específicos de interacción, visualización y carga de datos.

6. Principales logros del proyecto

A lo largo de los dos años de trabajo, el proyecto alcanzó resultados concretos en distintos planos.

En el plano institucional y formativo, se conformó un equipo interdisciplinario que integra investigadores del área de Ciencias de la Computación y estudiantes de carreras como Licenciatura en Sistemas, Administración de Sistemas, Contador Público, Marketing y Administración de Empresas. Esta articulación favoreció el intercambio entre disciplinas técnicas y económico-financieras, enriqueciendo tanto el enfoque investigativo como la formación de los participantes.

En el plano técnico, se construyó un entorno de trabajo en Python que incluye la gestión de entornos virtuales, el análisis del ciclo de vida de los datos y la implementación de escenarios en infraestructuras de

cómputo en la nube y en servidores propios. La recopilación y procesamiento de datos históricos —entre ellos, la serie completa de precios de las acciones de Microsoft desde su cotización inicial en la Bolsa de Nueva York hasta la actualidad— permitió contar con un conjunto de datos amplio y representativo.

Desde el punto de vista modelístico, se implementó el regresor XGBoost con optimización de hiperparámetros, incorporando indicadores técnicos clásicos: promedios móviles, índice de fuerza relativa (IFR) y convergencia/divergencia de medias móviles (MACD). El modelo fue entrenado, validado y evaluado bajo distintos escenarios, incluyendo el conjunto completo de datos y una submuestra del 60%, identificándose problemas de sobreajuste que deben abordarse en la etapa siguiente.

En el ámbito formativo, el proyecto impulsó instancias de capacitación en cómputo en la nube —en Amazon Web Services (AWS)—, logrando que estudiantes completaran módulos del programa AWS Cloud Foundations y, en al menos un caso, obtuvieran la certificación correspondiente. Asimismo, se avanzó en formación sobre análisis de grandes volúmenes de datos (*Big Data*), limpieza y normalización de datos, y técnicas avanzadas como la generación aumentada por recuperación (RAG, por sus siglas en inglés).

6.1. Resultados del modelo XGBoost

El modelo XGBoost entrenado sobre la serie histórica completa de precios de Microsoft mostró un desempeño satisfactorio en el conjunto de entrenamiento, con valores de raíz del error cuadrático medio (RECM) considerablemente bajos en ese subconjunto. Sin embargo, al evaluar el modelo sobre los datos de prueba, se

observó una brecha significativa entre el error en entrenamiento y el error en prueba, lo que evidencia un fenómeno de sobreajuste (*overfitting*). Esta limitación, característica de modelos de alta capacidad entrenados con conjuntos de datos de tamaño acotado, constituye uno de los principales desafíos técnicos a resolver en la etapa siguiente.

Los experimentos realizados con la submuestra del 60% de los datos permitieron confirmar que el comportamiento del modelo es sensible al volumen de información disponible: con menor cantidad de datos, el sobreajuste se acentúa y la capacidad de generalización disminuye. Estos resultados orientan las decisiones de diseño para la próxima etapa, donde se explorarán técnicas de regularización más robustas —como la penalización L1 y L2, el ajuste de la profundidad máxima de los árboles y la tasa de aprendizaje— con el objetivo de mejorar la estabilidad del modelo frente a datos no vistos.

La incorporación de indicadores técnicos como los promedios móviles simples (PMS) de 5, 10 y 20 días, el índice de fuerza relativa (IFR) y la convergencia/divergencia de medias móviles (MACD) aportó señales relevantes al modelo, aunque su poder predictivo resultó insuficiente de manera aislada. El análisis de importancia de variables (*feature importance*) indica que las variables de tendencia a corto plazo presentan mayor peso predictivo que las de largo plazo, lo que es consistente con la literatura sobre predicción de precios de acciones mediante aprendizaje automático (López de Prado, 2018).

6.2. Infraestructura de datos y entorno de desarrollo

El proceso de recopilación y preparación de datos constituyó una etapa de alta complejidad técnica. Las APIs utilizadas —Yahoo Finance, Alpha Vantage, CoinMarketCap y CoinGecko— presentaron diferencias en los formatos de respuesta, frecuencias de actualización y políticas de acceso, lo que requirió el desarrollo de rutinas de normalización y reconciliación de datos ad hoc. La gestión de valores faltantes, outliers y discontinuidades en la serie temporal de Microsoft demandó un tratamiento cuidadoso para preservar la integridad de los datos sin introducir sesgos artificiales.

El entorno de desarrollo se desplegó sobre instancias virtualizadas mediante KVM y contenedores Docker, lo que permitió garantizar la reproducibilidad de los experimentos en distintos equipos del equipo de investigación. La adopción de entornos virtuales de Python (venv) para la gestión de dependencias y la documentación sistemática de los experimentos mediante notebooks de Jupyter contribuyeron a establecer una cultura de trabajo reproducible y colaborativa, hábitos fundamentales en el campo de la ciencia de datos aplicada.

7. Balance entre objetivos propuestos y alcanzados

Entre los logros consolidados se cuentan la implementación efectiva del modelo de predicción en Python, la evaluación de su desempeño en distintos escenarios y el diagnóstico preciso de sus limitaciones —en especial el sobreajuste y la reducida capacidad de generalización—.

Queda pendiente, asimismo, la experimentación sistemática con nuevas arquitecturas de aprendizaje profundo y la

evaluación de estrategias eficaces para reducir el sobreajuste.

Este diagnóstico de limitaciones no debe interpretarse como un resultado negativo, sino como una contribución metodológica en sí misma. La caracterización precisa del sobreajuste, la identificación de las variables con mayor poder predictivo y la delimitación de los escenarios donde el modelo opera de manera más confiable constituyen insumos valiosos para el diseño de la etapa siguiente. En línea con las mejores prácticas del campo, el equipo adoptó un enfoque iterativo que contempla el fracaso como parte del proceso de mejora continua (Bishop, 2006).

El equipo estima que un año adicional de investigación permitirá incorporar variables de sentimiento de mercado, probar arquitecturas de aprendizaje profundo como LSTM y GRU, y realizar una validación cruzada más exhaustiva sobre distintos activos financieros. Esta extensión del proyecto posibilitará también contrastar las hipótesis planteadas originalmente con evidencia empírica más robusta, avanzando hacia el cumplimiento integral de los objetivos e incrementando el valor científico y aplicado de los resultados.

8. Alcance social y ético de la investigación

El desarrollo de herramientas predictivas en el ámbito financiero conlleva implicancias sociales y éticas que el equipo ha asumido como parte integral del proceso de investigación.

En términos sociales, estas aplicaciones tienen el potencial de democratizar el acceso a información financiera avanzada, poniéndola al alcance de inversores minoristas y reduciendo las barreras de

entrada a los mercados. Al mismo tiempo, su uso masivo podría incrementar la volatilidad si muchos actores actúan sobre la base de las mismas predicciones, generando ciclos de retroalimentación no deseados.

En el plano ético, el proyecto se rige por principios de transparencia, privacidad y responsabilidad. La claridad acerca de las limitaciones predictivas —ningún modelo garantiza certeza absoluta— es un imperativo irrenunciable. El manejo de los datos de los usuarios debe ajustarse a los marcos normativos vigentes en materia de protección de datos personales⁶ y sustentarse en el consentimiento informado. La aplicación debe diseñarse también para prevenir usos que puedan derivar en manipulación del mercado o en comportamientos impulsivos, incorporando elementos educativos que orienten al usuario hacia una toma de decisiones responsable.

8.1. Impacto en la inclusión financiera

La democratización del acceso a herramientas de análisis financiero avanzado representa uno de los impactos sociales más relevantes del proyecto. En el contexto argentino, donde una proporción significativa de la población no tiene acceso a servicios de asesoría financiera especializada, una aplicación de predicción de precios de bajo costo o libre acceso podría contribuir a reducir las asimetrías de información entre inversores institucionales e individuales. Esta reducción de barreras resulta especialmente significativa para pequeños ahorristas que buscan proteger su patrimonio frente a la volatilidad macroeconómica.

⁶En Argentina rige la Ley 25.326 de Protección de los Datos Personales. En el ámbito europeo, el Reglamento General de Protección de Datos (RGPD) establece el marco de referencia internacional más extendido.

No obstante, la democratización tecnológica conlleva riesgos que deben gestionarse con responsabilidad. La sobre-dependencia en herramientas automatizadas de predicción, sin la formación financiera adecuada, puede generar decisiones de inversión erróneas con consecuencias patrimoniales graves para los usuarios. Por este motivo, la aplicación contempla desde su concepción la incorporación de módulos informativos que contextualicen las predicciones en el marco del riesgo inherente a los mercados financieros, promoviendo una cultura de inversión consciente e informada.

8.2. Consideraciones regulatorias y de privacidad

El desarrollo de la aplicación requiere considerar el marco normativo vigente en materia de datos personales y actividades financieras. En Argentina, la Ley 25.326 de Protección de los Datos Personales regula el tratamiento de información personal, mientras que la Comisión Nacional de Valores (CNV) supervisa las actividades relacionadas con el mercado de capitales. El cumplimiento de estas normativas no solo representa una obligación legal, sino también un componente esencial para generar confianza en los usuarios y legitimar el uso de la herramienta en el ecosistema financiero local.

En cuanto a la transparencia algorítmica, el equipo adhiere al principio de explicabilidad de los modelos (explainability), según el cual los usuarios deben poder comprender, al menos en términos generales, por qué el modelo genera una determinada predicción. La adopción de técnicas de interpretabilidad como SHAP (Shapley Additive Explanations) está prevista para

etapas futuras del proyecto, con el objetivo de facilitar la auditoría de los modelos y fortalecer la confianza de los usuarios en las predicciones generadas.

9. Conclusiones

Este proyecto de investigación representa un aporte concreto al campo emergente de las finanzas computacionales en el contexto universitario argentino. En el transcurso de dos años se construyó una infraestructura técnica sólida, se implementó un primer modelo predictivo funcional y se generó un proceso formativo valioso para los estudiantes involucrados.

Los resultados obtenidos permiten identificar con claridad las limitaciones del modelo actual —particularmente el sobreajuste y la ausencia de variables

En cuanto a las proyecciones metodológicas, la siguiente etapa priorizará tres líneas de trabajo. En primer lugar, la incorporación de datos de sentimiento de mercado mediante técnicas de procesamiento del lenguaje natural (PLN), analizando noticias financieras, publicaciones en redes sociales y comunicados corporativos relacionados con los activos bajo análisis. En segundo lugar, la implementación y evaluación comparativa de arquitecturas de aprendizaje profundo —en particular LSTM y GRU— frente al modelo XGBoost ya desarrollado, utilizando métricas estandarizadas que permitan una comparación objetiva entre enfoques. En tercer lugar, la expansión del universo de activos analizados, incorporando criptomonedas y acciones de empresas de

sentimentales— y trazan una hoja de ruta precisa para la continuidad del proyecto. La incorporación de datos no estructurados mediante técnicas de procesamiento del lenguaje natural, junto a la exploración de arquitecturas LSTM y otros enfoques de aprendizaje profundo, constituye el eje central de la siguiente etapa.

Más allá de sus resultados técnicos, el proyecto contribuye a la formación de recursos humanos con competencias en inteligencia artificial, ciencia de datos y cómputo en la nube. El equipo aspira a que esta investigación sienta las bases para el desarrollo de herramientas financieras accesibles, transparentes y éticamente responsables, al servicio de la comunidad inversora y de la sociedad en su conjunto.

distintos sectores e índices bursátiles, con el fin de evaluar la generalización del modelo más allá del caso de Microsoft.

Desde una perspectiva institucional, el proyecto ha demostrado la viabilidad de desarrollar investigación aplicada en tecnología financiera dentro de una universidad argentina de tamaño mediano, articulando recursos humanos interdisciplinarios, infraestructura de cómputo disponible en el ecosistema académico y marcos teóricos sólidamente documentados. Esta experiencia puede servir de referencia para iniciativas similares en otras unidades académicas de la UdeMM o de universidades nacionales con perfiles análogos.

Referencias bibliográficas

- Bishop, C. M. (2006). *Pattern recognition and machine learning*. Springer.
- Chan, E. P. (2013). *Algorithmic trading: Winning strategies and their rationale*. Wiley.
- De La Torre, J. I. (2021). *Deep learning for finance*. Springer.
- Fama, E. F. (1970). Efficient capital markets: A review of empirical work. *The Journal of Finance*, 25(2), 383-417. <https://doi.org/10.2307/2325486>
- López de Prado, M. (2018). *Advances in financial machine learning*. Wiley.
- López de Prado, M. (2020). *Machine learning for asset managers*. Cambridge University Press.
- Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, 30. <https://doi.org/10.48550/arXiv.1705.07874>