



Desarrollo de un modelo predictivo basado en aprendizaje automático para trading algorítmico

Autor:

Ing. Gustavo Uñapillco

Director:

DSc. Camilo Argoty (FIUBA)

*Esta planificación fue realizada en el curso de Gestión de proyectos
entre el 29 de abril de 2025 y el 17 de junio de 2025.*

Índice

1. Descripción técnica-conceptual del proyecto a realizar	5
2. Identificación y análisis de los interesados	6
3. Propósito del proyecto	7
4. Alcance del proyecto	7
5. Supuestos del proyecto.	8
6. Requerimientos	9
7. Historias de usuarios (<i>Product backlog</i>).	10
8. Entregables principales del proyecto	11
9. Desglose del trabajo en tareas	11
10. Diagrama de Activity On Node.	12
11. Diagrama de Gantt	15
12. Presupuesto detallado del proyecto	17
13. Gestión de riesgos	17
14. Gestión de la calidad	20
15. Procesos de cierre	21

Registros de cambios

Revisión	Detalles de los cambios realizados	Fecha
0	Creación del documento	29 de abril de 2025
1	Se completa hasta el punto 5 inclusive	10 de mayo de 2025
2	Se completa hasta el punto 9 inclusive y se realizan las correcciones de la entrega anterior	16 de mayo de 2025
3	Se completa hasta el punto 12 inclusive y se realizan las correcciones de la entrega anterior	25 de mayo de 2025
4	Se completa el plan	02 de junio de 2025

Acta de constitución del proyecto

Buenos Aires, 29 de abril de 2025

Por medio de la presente se acuerda con el Ing. Gustavo Uñapillco que su Trabajo Final de la Carrera de Especialización en Inteligencia Artificial se titulará “Desarrollo de un modelo predictivo basado en aprendizaje automático para trading algorítmico” y consistirá en el desarrollo de un modelo predictivo basado en aprendizaje automático para trading algorítmico en el Futuro E-mini Nasdaq-100 (MNQ). El trabajo tendrá un presupuesto preliminar estimado de 600 horas y un costo estimado de US\$ 15.630, con fecha de inicio el 29 de abril de 2025 y fecha de presentación pública el 20 de diciembre de 2025.

Se adjunta a esta acta la planificación inicial.

Dr. Ing. Ariel Lutenberg
Director posgrado FIUBA

Ing. Gustavo Uñapillco
Alumno posgrado FIUBA

DSc. Camilo Argoty
Director del Trabajo Final

1. Descripción técnica-conceptual del proyecto a realizar

La transformación digital en el ámbito financiero ha dado lugar a nuevas formas de operar en los mercados, donde la velocidad, la precisión y la capacidad de procesar grandes volúmenes de datos se vuelven esenciales para la toma de decisiones. En este contexto, el *trading* algorítmico se ha consolidado como una herramienta clave para automatizar estrategias de inversión, reducir el sesgo humano y aumentar la eficiencia operativa. No obstante, los desafíos inherentes a la alta volatilidad y la complejidad de los mercados financieros siguen demandando enfoques más robustos e inteligentes.

Este proyecto se enmarca como un emprendimiento personal, impulsado por el interés del autor en la intersección entre inteligencia artificial, programación y mercados financieros. Su objetivo principal es desarrollar un modelo predictivo basado en modelos de aprendizaje automático, orientado a la anticipación de movimientos intradía del contrato de futuros E-mini Nasdaq-100 (MNQ), un instrumento financiero caracterizado por su alta liquidez y comportamiento dinámico minuto a minuto.

La propuesta consiste en construir un modelo que, a partir de datos históricos de alta frecuencia, permita predecir la evolución del precio en horizontes de muy corto plazo. Para ello, se emplearán series temporales que contienen los valores OHLCV (Open, High, Low, Close, Volume), extraídas desde la plataforma NinjaTrader, complementadas con indicadores técnicos tales como RSI, MACD y medias móviles exponenciales. Además, se explorará la generación de factores compuestos o transformados, a partir de la combinación de estos indicadores, conocidos como *alpha factors*, con el objetivo de capturar relaciones no lineales y patrones de mayor valor predictivo. Estas series serán sometidas a procesos de normalización, estructuración temporal y reducción de ruido, con el fin de capturar la dinámica subyacente del mercado.

Desde el punto de vista metodológico, se desarrollarán y compararán modelos de aprendizaje supervisado. Se priorizará el uso de redes neuronales recurrentes, como LSTM y GRU, por su capacidad de modelar secuencias temporales y capturar dependencias de corto y largo plazo. También se evaluarán modelos como XGBoost, ampliamente utilizados por su capacidad de generalización y eficiencia. El enfoque no se limitará exclusivamente a estos algoritmos, sino que también se dejará abierta la posibilidad de explorar arquitecturas más complejas, como modelos basados en *Transformers* para series temporales, siempre que los recursos disponibles y el tiempo del proyecto lo permitan.

El desempeño de los modelos será evaluado mediante un *pipeline* automatizado de entrenamiento, validación y *backtesting*, utilizando métricas estadísticas (MAE, RMSE, R^2) y métricas específicas del dominio financiero, tales como el retorno promedio, ratio de Sharpe y tasa de aciertos operativos. El sistema será validado en un entorno *offline* con datos históricos no utilizados durante el entrenamiento, y se documentará todo el proceso de desarrollo técnico.

El proyecto aún no cuenta con un prototipo funcional. Hasta el momento, solo se ha dado un primer paso durante el trabajo práctico final de la materia *Análisis de Series Temporales 1*, correspondiente al curso de especialización. En esa instancia preliminar se exploraron modelos sencillos, como regresiones ARIMA, aplicados sobre una fracción de los datos, pero los resultados obtenidos fueron poco satisfactorios debido a la limitada capacidad de dichos modelos para capturar la complejidad del comportamiento del índice MNQ. El presente proyecto constituye así el primer intento formal y sistemático de abordar el problema desde un enfoque más robusto, con técnicas de aprendizaje automático, una arquitectura escalable y un proceso de validación más riguroso.

Es importante señalar que el objetivo inmediato de este proyecto es validar la capacidad predictiva del modelo. La generación automatizada de señales de compra y venta no será implementada en esta etapa, pero se contempla como una línea de desarrollo futuro, una vez que se haya verificado el rendimiento del modelo en condiciones controladas.

El proyecto se encuentra actualmente en una etapa de planificación y diseño. No existen acuerdos de confidencialidad ni restricciones de propiedad intelectual, dado que se trata de un desarrollo individual con datos de acceso público o comercial. La solución ha sido concebida con una arquitectura modular y replicable, lo que permitirá su futura adaptación a entornos reales o incluso su integración con fuentes de datos en tiempo real mediante APIs especializadas.

En la figura 1 se presenta el diagrama en bloques de la propuesta. Se observa que el flujo comienza con la recopilación de datos históricos de alta frecuencia del índice MNQ, los cuales son preprocesados y estructurados en secuencias temporales para su posterior uso en el entrenamiento del modelo de predicción. Una vez entrenado, el modelo es evaluado tanto desde un enfoque estadístico como financiero. Finalmente, se contempla una etapa futura donde las predicciones obtenidas podrían ser traducidas en señales operativas automatizadas de compra y venta, sujeto a la validación del rendimiento predictivo.



Figura 1. Diagrama en bloques del sistema.

2. Identificación y análisis de los interesados

- Responsable: Gustavo Uñapillco, alumno de la Carrera de Especialización en Inteligencia Artificial, será el encargado del desarrollo, implementación y documentación completa del sistema.

- Cliente: el mismo autor en su rol de solicitante del sistema, interesado en desarrollar una herramienta predictiva aplicable al análisis de mercados financieros. Será quien defina los requerimientos y valide el funcionamiento del sistema durante el desarrollo.
- Orientador: brindará orientación metodológica y técnica en las distintas etapas del proyecto, colaborando con la definición del enfoque de modelado y los criterios de validación.
- Usuario final: personas con formación técnica o económica interesadas en aplicar modelos predictivos al *trading* algorítmico. Este grupo podrá utilizar el sistema como apoyo para la toma de decisiones, o bien como base para investigación o desarrollo de estrategias cuantitativas.

Rol	Nombre y Apellido	Organización	Puesto
Responsable	Ing. Gustavo Uñapillco	FIUBA	Alumno
Cliente	Ing. Gustavo Uñapillco	Emprendimiento Personal	Solicitante del proyecto
Orientador	DSc. Camilo Argoty	FIUBA	Director del Trabajo Final
Usuario final	–	–	Profesionales interesados en <i>trading</i> algorítmico

3. Propósito del proyecto

El propósito de este proyecto es desarrollar un modelo predictivo basado en técnicas de aprendizaje automático que permita anticipar los movimientos de muy corto plazo del contrato de futuros E-mini Nasdaq-100 (MNQ), utilizando datos históricos de alta frecuencia. La intención es explorar el potencial de la inteligencia artificial como herramienta para el análisis de series temporales financieras, enfocándose en su aplicabilidad operativa y en la comprensión de patrones complejos del mercado. El proyecto busca validar si es posible construir una solución precisa y robusta que sirva como base para estrategias automatizadas en entornos financieros reales.

4. Alcance del proyecto

Este proyecto incluye la recolección y análisis de datos históricos de alta frecuencia del contrato de futuros E-mini Nasdaq-100 (MNQ), específicamente con resolución de un minuto, provenientes de la plataforma NinjaTrader. Los datos considerados abarcan los precios de apertura, cierre, máximo, mínimo y volumen negociado (OHLCV).

Dentro del alcance se encuentra el preprocesamiento de dichos datos, que implica tareas de limpieza, normalización y estructuración en ventanas temporales. También se incorporarán indicadores técnicos complementarios, tales como el RSI, medias móviles y MACD, junto con la generación de factores combinados o transformados conocidos como *alpha factors*, diseñados a partir de la combinación o transformación de variables existentes, con el objetivo de mejorar la capacidad predictiva del modelo.

Se realizará el diseño, entrenamiento y evaluación de modelos de predicción basados en técnicas de aprendizaje automático. En particular, se utilizarán redes neuronales recurrentes

(LSTM, GRU) y algoritmos de *boosting* como XGBoost. No obstante, el enfoque no se limitará exclusivamente a estos modelos, sino que se explorará también la viabilidad de arquitecturas más complejas y avanzadas, tales como *Transformers* adaptados a series temporales o modelos híbridos, en la medida en que los tiempos del proyecto y los recursos disponibles lo permitan.

La validación del desempeño se hará mediante un *pipeline* automatizado, utilizando métricas estadísticas (MAE, RMSE, R^2) y métricas financieras relevantes (retorno medio, ratio de Sharpe, tasa de aciertos).

El sistema será evaluado de forma *offline*, utilizando datos históricos no vistos, y se documentarán todos los resultados, decisiones técnicas y hallazgos. Se elaborará además la memoria técnica del proyecto y la presentación correspondiente para la defensa.

Quedan fuera del alcance del proyecto la integración en tiempo real con plataformas de *trading* o APIs financieras, así como la ejecución de operaciones reales en el mercado. Tampoco se contempla la generación de señales automatizadas de compra y venta como resultado final del trabajo. Esto se debe a que el proyecto contempla un máximo estimado de 600 horas de dedicación, lo cual delimita el alcance a la construcción, entrenamiento y validación de modelos predictivos en entorno *offline*. La exploración de funcionalidades operativas en tiempo real se plantea como una posible línea de trabajo futuro, sujeta a la validación del rendimiento del sistema en condiciones controladas y a la disponibilidad de recursos técnicos adicionales.

5. Supuestos del proyecto

Para el desarrollo del presente proyecto se supone que:

- El responsable del proyecto contará con una disponibilidad horaria estimada de 600 horas, distribuidas de forma regular a lo largo del cronograma previsto por la carrera.
- Se podrá acceder sin restricciones a los datos históricos minuto a minuto del contrato E-mini Nasdaq-100 (MNQ) mediante la plataforma NinjaTrader, o por otros medios alternativos equivalentes.
- Se contará con los recursos computacionales personales necesarios (una notebook equipada con entorno de desarrollo Python y bibliotecas compatibles) para realizar el procesamiento, entrenamiento y evaluación de los modelos propuestos.
- No habrá impedimentos legales, comerciales o de derechos de autor que limiten el uso de los datos descargados ni la publicación de los resultados obtenidos.
- El entorno técnico y académico (orientador, repositorio, bibliografía, plataforma de desarrollo) estará disponible para consultas, revisiones periódicas y documentación durante toda la duración del proyecto.
- Los modelos de aprendizaje automático seleccionados podrán entrenarse de forma local o en la nube sin requerir infraestructura especializada adicional.
- No se prevén cambios drásticos en las condiciones regulatorias del acceso a datos de mercado, ni restricciones en el uso de herramientas *open source* empleadas.

6. Requerimientos

1. Requerimientos funcionales (Alta prioridad):

- 1.1. El sistema experimental debe permitir la carga y el preprocesamiento de datos históricos en formato OHLCV con resolución de un minuto, correspondientes al contrato futuro E-mini Nasdaq-100 (MNQ).
- 1.2. El sistema debe permitir la estructuración de dichos datos en ventanas temporales adecuadas para entrenar modelos de predicción de series temporales.
- 1.3. El entorno debe facilitar la experimentación con distintos modelos de predicción, incluyendo al menos: ARIMA, LSTM y XGBoost, así como permitir la evaluación de arquitecturas más complejas como Transformers en etapas exploratorias.
- 1.4. Cada modelo desarrollado debe ser evaluado con métricas estadísticas (MAE, RMSE, R^2) y métricas del dominio financiero (retorno medio, tasa de aciertos, ratio de Sharpe, etc.), permitiendo comparaciones objetivas entre enfoques.
- 1.5. El objetivo de predicción será el valor futuro de la variable *retorno futuro* de la variable *close* en horizontes temporales de 15, 20 y 30 minutos, calculado a partir de bloques históricos de 60 minutos.
- 1.6. El sistema debe incluir herramientas para visualizar las predicciones vs. los valores reales, incluyendo gráficos de líneas y simulaciones de *backtesting* sobre datos históricos.
- 1.7. En esta etapa, el sistema no requiere funcionamiento en tiempo real ni integración con plataformas externas. Sin embargo, se deberá documentar su diseño de forma modular y extensible para una futura implementación operativa.

2. Requerimientos de documentación (Media prioridad):

- 2.1. El código debe estar completamente documentado siguiendo convenciones de buenas prácticas (PEP8, *docstrings*, etc.).
- 2.2. El proyecto debe incluir una memoria técnica que describa los datos utilizados, los modelos probados, los resultados obtenidos y las decisiones de diseño adoptadas.

3. Requerimiento de testing (Alta prioridad):

- 3.1. El sistema debe incluir pruebas de validación cruzada y evaluación fuera de muestra, para verificar que el modelo generaliza correctamente y no se ajusta únicamente a los datos de entrenamiento.
- 3.2. El desempeño del modelo debe medirse mediante métricas estadísticas como MAE, RMSE y R^2 . El valor mínimo aceptable para R^2 en la validación fuera de muestra será de 0,65, asegurando un nivel razonable de explicación de la variabilidad de los precios.
- 3.3. El sistema también debe calcular métricas financieras operativas, tales como el retorno medio por operación, ratio de Sharpe y tasa de aciertos, con el fin de evaluar su aplicabilidad en contextos de trading algorítmico mediante simulaciones de *backtesting*.

4. Requerimientos de la interfaz (opcional):

- 4.1. En caso de contar con tiempo suficiente, se desarrollará una interfaz básica en Jupyter Notebook o Streamlit para mostrar predicciones e indicadores visualmente.

5. Requerimientos interoperabilidad (Media prioridad):

- 5.1. El sistema debe ser capaz de integrarse en el futuro con plataformas externas de trading mediante exportación estructurada de señales.
- 5.2. Los datos de entrada deben poder importarse desde fuentes en formatos .TXT, CSV, Parquet o HDF5.

6. Requerimientos normativos (Alta prioridad):

- 6.1. Todo el desarrollo debe cumplir con licencias de uso libre (MIT o equivalente) y respetar los términos de uso de las plataformas de datos utilizadas (por ejemplo, NinjaTrader).

7. Historias de usuarios (*Product backlog*)

A continuación se presentan las historias de usuario del proyecto, redactadas en el formato ágil “Como [rol], quiero [objetivo] para [beneficio]”. Cada historia ha sido estimada en términos de dificultad (D), complejidad (C) e incertidumbre (R), sumadas y redondeadas al número más cercano de la serie de Fibonacci para asignar los *Story Points*.

1. “Como desarrollador, quiero entrenar distintos modelos de predicción (ARIMA, LSTM, XGBoost, etc.) para comparar su desempeño y seleccionar el más adecuado.”
Story points: 13 (complejidad: 5, dificultad: 5, incertidumbre: 3)
2. “Como desarrollador, quiero preprocesar y normalizar los datos OHLCV para garantizar que el modelo trabaje con información limpia y estructurada.”
Story points: 13 (complejidad: 5, dificultad: 5, incertidumbre: 3)
3. “Como investigador, quiero evaluar el modelo en datos no vistos durante el entrenamiento para estimar su capacidad de generalización.”
Story points: 13 (complejidad: 3, dificultad: 3, incertidumbre: 5)
4. “Como usuario, quiero visualizar el rendimiento histórico del modelo mediante gráficos de valores reales vs. predichos y marcar señales correctas y fallidas para evaluar su confiabilidad.”
Story points: 8 (complejidad: 3, dificultad: 3, incertidumbre: 2)
5. “Como usuario, quiero recibir predicciones claras del retorno futuro en horizontes de 15, 20 y 30 minutos para tomar decisiones de trading de corto plazo más informadas.”
Story points: 13 (complejidad: 5, dificultad: 4, incertidumbre: 3)
6. “Como usuario, quiero conocer con qué confianza o probabilidad el modelo realiza cada predicción para ponderar el riesgo asociado a cada señal.”
Story points: 13 (complejidad: 5, dificultad: 3, incertidumbre: 5)
7. “Como usuario, quiero poder ajustar los parámetros de entrada (como el horizonte de predicción) desde una interfaz simple para adaptar el modelo a diferentes estilos operativos.”
Story points: 8 (complejidad: 3, dificultad: 2, incertidumbre: 3)
8. “Como usuario, quiero obtener señales de compra o venta generadas por el modelo con su justificación gráfica y métrica para actuar rápidamente.”
Story points: 13 (complejidad: 5, dificultad: 4, incertidumbre: 4)

9. “Como usuario, quiero que la herramienta indique cuándo evitar operar (por baja confianza, alta volatilidad o datos inconsistentes) para reducir errores por falsas señales.”
Story points: 8 (complejidad: 3, dificultad: 2, incertidumbre: 3)
10. “Como usuario, quiero exportar los resultados del modelo, incluyendo métricas, predicciones y señales, en formatos como CSV o Excel para analizarlos o documentarlos externamente.”
Story points: 5 (complejidad: 2, dificultad: 2, incertidumbre: 1)
11. “Como investigador, quiero calcular métricas financieras como retorno promedio, Sharpe y tasa de aciertos para analizar la utilidad práctica del modelo en contexto de trading.”
Story points: 13 (complejidad: 5, dificultad: 3, incertidumbre: 5)

8. Entregables principales del proyecto

Los entregables del presente proyecto incluyen tanto artefactos técnicos como documentación de soporte, y se detallan a continuación:

- Plan de proyecto.
- Dataset limpio y estructurado, con variables OHLCV e indicadores técnicos derivados.
- Análisis exploratorio y descriptivo de los datos.
- Resultados de entrenamiento y validación de los distintos modelos (ARIMA, LSTM, XGBoost, etc.).
- Métricas estadísticas y financieras obtenidas (MAE, RMSE, R^2 , retorno medio, Sharpe, tasa de aciertos).
- Código fuente documentado para preprocesamiento, entrenamiento, evaluación y generación de predicciones.
- Visualizaciones de desempeño predictivo y simulaciones de backtesting.
- Archivo de predicciones exportado en formato CSV para posible integración futura con plataformas de trading.
- Memoria técnica del proyecto.
- Presentación final para la defensa oral.

9. Desglose del trabajo en tareas

1. Planificación y definición del proyecto (50 h)
 - 1.1. Revisión de objetivos y requerimientos funcionales (10 h).
 - 1.2. Estudio preliminar del estado del arte en predicción financiera e inteligencia artificial (20 h).
 - 1.3. Definición del enfoque metodológico y selección de herramientas (20 h).
2. Obtención, exploración y preparación de datos (140 h)

- 2.1. Descarga y organización de datos históricos desde NinjaTrader (20 h).
- 2.2. Limpieza, normalización y estructuración de series temporales (30 h).
- 2.3. Cálculo de indicadores técnicos (RSI, MACD, etc.) (20 h).
- 2.4. Generación de alpha factors y selección de variables (30 h).
- 2.5. Generación de variables derivadas: lags, medias móviles, z-scores, volatilidad GARCH (20 h).
- 2.6. División temporal del dataset en subconjuntos de entrenamiento y prueba (10 h).
- 2.7. Evaluación del esquema de ventanas deslizantes (60→15/20) (10 h).
3. Diseño, entrenamiento y evaluación de modelos (240 h)
 - 3.1. Entrenamiento y evaluación de modelos benchmark (ARIMA, GARCH) (20 h).
 - 3.2. Entrenamiento y tuning de Random Forest como modelo base multivariado (30 h).
 - 3.3. Entrenamiento y optimización de XGBoost con búsqueda de hiperparámetros (40 h).
 - 3.4. Entrenamiento de MLP con features planos e ingeniería derivada (30 h).
 - 3.5. Implementación y ajuste de LSTM sobre secuencias temporales (40 h).
 - 3.6. Exploración y entrenamiento de TCN como alternativa secuencial a LSTM (30 h).
 - 3.7. Implementación y ajuste de modelo Transformer para series temporales multivariadas (30 h).
 - 3.8. Exploración opcional de modelos híbridos o combinados (LSTM + XGBoost, etc.) (20 h).
4. Validación, *backtesting* y análisis de resultados (90 h)
 - 4.1. Implementación del módulo de backtesting (30 h).
 - 4.2. Simulación de escenarios operativos y decisiones basadas en señales (30 h).
 - 4.3. Análisis de métricas estadísticas (MAE, MSE, MAPE, precisión direccional) (15 h).
 - 4.4. Evaluación del rendimiento financiero del modelo (ROI, drawdown, señales) (15 h).
5. Documentación, presentación y cierre (80 h)
 - 5.1. Documentación técnica del código, notebooks y experimentos (30 h).
 - 5.2. Redacción de la memoria técnica del proyecto (30 h).
 - 5.3. Preparación de presentación final y defensa oral (10 h).
 - 5.4. Revisión, correcciones y cierre formal del proyecto (10 h).

Cantidad total de horas: 600 h.

10. Diagrama de Activity On Node

Las figuras 2 y 3 presentan el diagrama de Activity On Node (AoN) correspondiente al proyecto, elaborado en base al desglose de tareas (WBS) definido previamente. Los tiempos indicados para cada actividad están expresados en horas. El camino crítico, representado mediante flechas negras de trazo grueso, identifica la secuencia de tareas que determinan la duración mínima del proyecto. La duración total estimada de dicho camino crítico es de 270 horas.

El camino crítico es la ruta más larga a través del diagrama de red del proyecto, considerando las dependencias entre tareas. Se compone de aquellas actividades que, en caso de sufrir demoras, impactarían directamente en la fecha final de entrega. Por esta razón, su identificación resulta fundamental para la gestión del cronograma y la priorización de recursos.

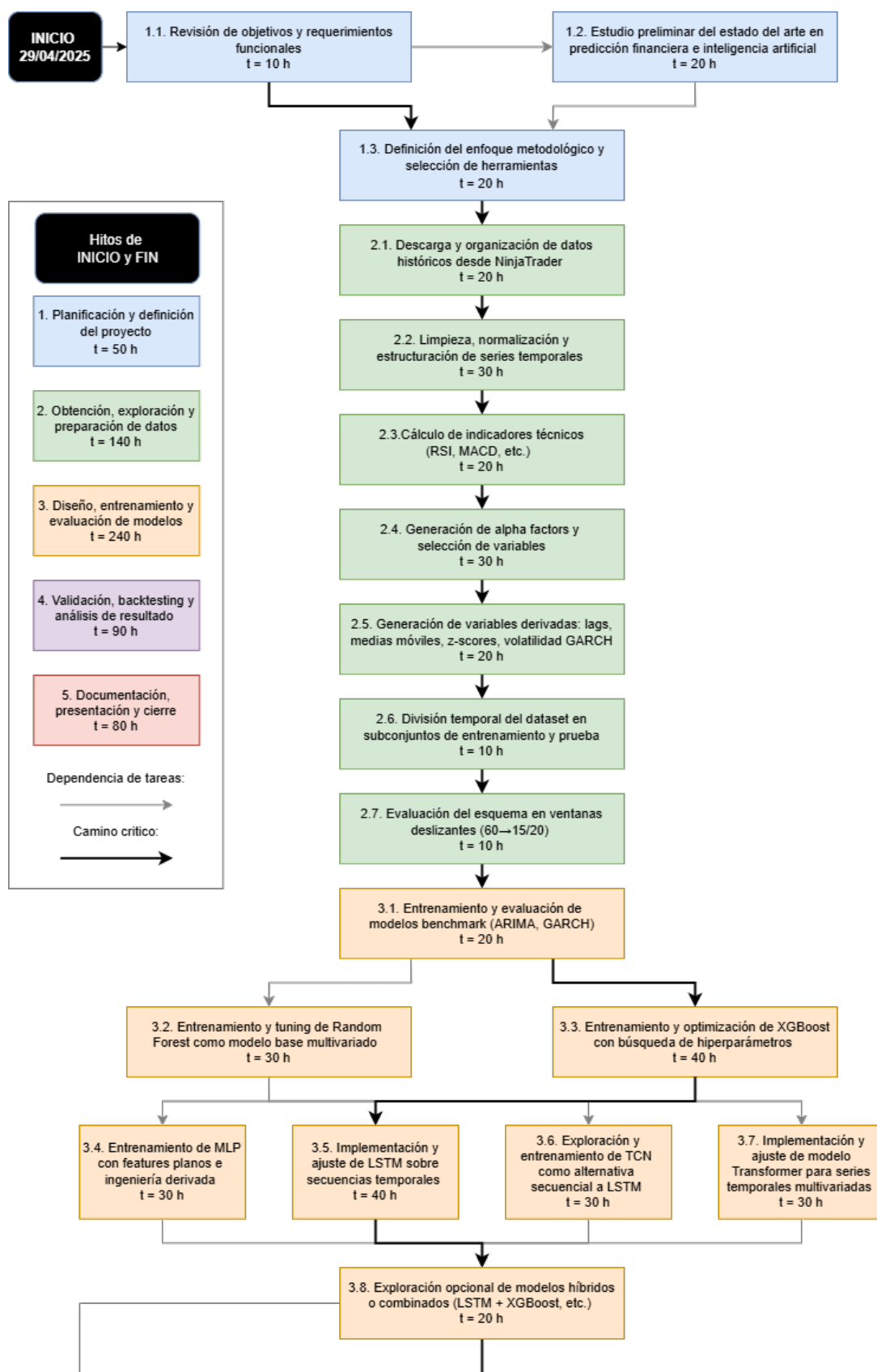


Figura 2. Diagrama de *Activity on Node* (Inicio).

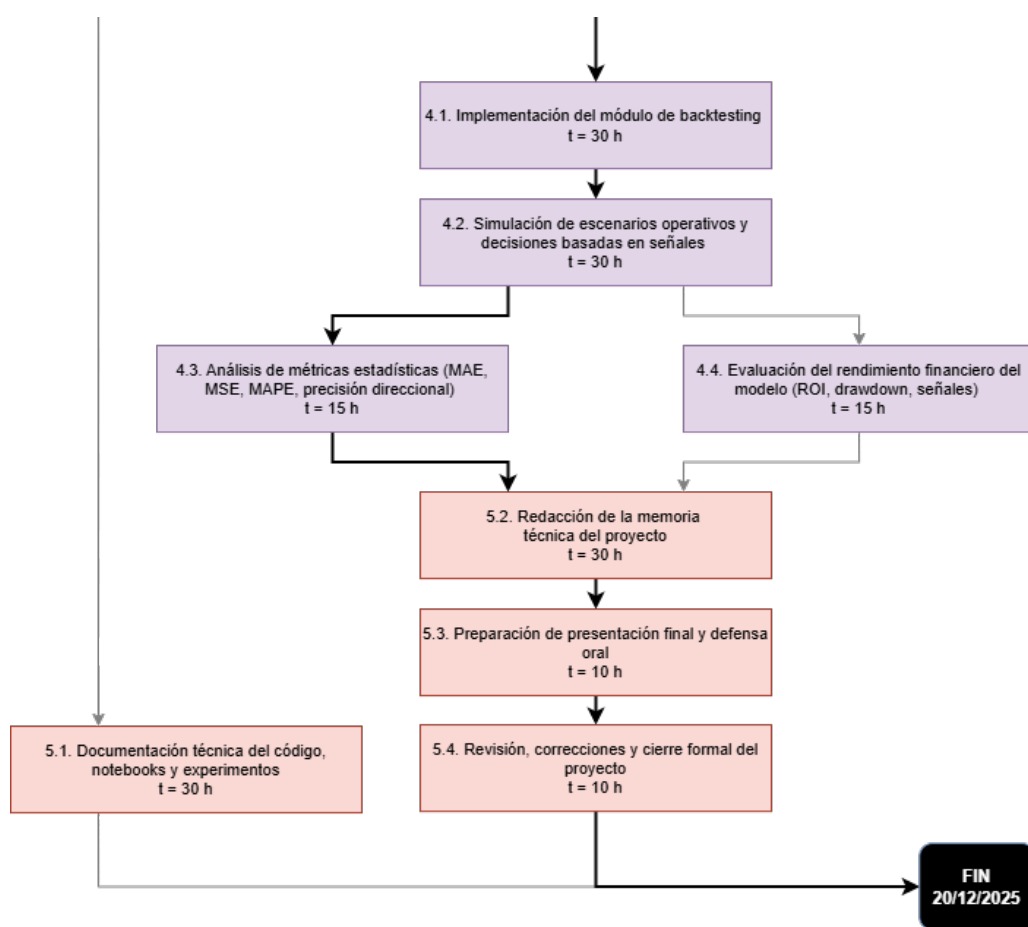


Figura 3. Diagrama de *Activity on Node* (Final).

11. Diagrama de Gantt

A continuación, se presenta el diagrama de Gantt correspondiente a la planificación del proyecto. Para su elaboración, se tomó como referencia el desglose estructurado del trabajo (WBS) previamente definido, lo cual permitió organizar cronológicamente las tareas e identificar sus respectivas duraciones, dependencias y secuencia lógica de ejecución.

Con el fin de facilitar la comprensión y mejorar la legibilidad del cronograma, el diagrama ha sido dividido en dos componentes complementarios. En primer lugar, se muestra la tabla de planificación (Cuadro 1), donde se indican los códigos de tareas, sus descripciones, la duración estimada en horas, y las fechas asignadas de inicio y finalización. Posteriormente, se presenta el diagrama de barras de Gantt (Figura 5), que representa gráficamente el avance temporal de cada tarea, permitiendo visualizar la superposición de actividades, las fases del proyecto y la secuencia general.

La semana laboral fue configurada de lunes a sábado, con una jornada de 3 horas diarias, en concordancia con la dedicación estimada y la disponibilidad del responsable del proyecto. El cronograma completo abarca el período comprendido entre el 29 de abril y el 20 de diciembre de 2025, dentro del cual se distribuyen las 600 horas previstas de trabajo, cumpliendo con los objetivos y tiempos establecidos en el Acta de Constitución.

WBS	Tarea	Duración (h)	Inicio	Fin
1.1	Revisión de objetivos	10	2025-04-29	2025-05-02
1.2	Estudio del estado del arte	20	2025-05-03	2025-05-10
1.3	Definición metodológica y herramientas	20	2025-05-12	2025-05-19
2.1	Descarga de datos	20	2025-05-20	2025-05-27
2.2	Limpieza y normalización	30	2025-05-28	2025-06-07
2.3	Cálculo de indicadores técnicos	20	2025-06-09	2025-06-16
2.4	Generación de alpha factors	30	2025-06-17	2025-06-27
2.5	Variables derivadas	20	2025-06-28	2025-07-05
2.6	División en conjuntos	10	2025-07-07	2025-07-09
2.7	Evaluación de ventanas	10	2025-07-10	2025-07-12
3.1	Modelos benchmark	20	2025-07-14	2025-07-21
3.2	Random Forest	30	2025-07-22	2025-08-01
3.3	XGBoost	40	2025-08-02	2025-08-15
3.4	MLP	30	2025-08-16	2025-08-26
3.5	LSTM	40	2025-08-27	2025-09-09
3.6	TCN	30	2025-09-18	2025-09-29
3.7	Transformer	30	2025-09-30	2025-10-10
3.8	Modelos híbridos	20	2025-10-11	2025-10-18
4.1	Backtesting	30	2025-10-20	2025-10-30
4.2	Simulación de señales	30	2025-10-31	2025-11-11
4.3	Métricas estadísticas	15	2025-11-12	2025-11-17
4.4	Evaluación financiera	15	2025-11-18	2025-11-22
5.1	Documentación técnica	30	2025-11-24	2025-12-04
5.2	Redacción de la memoria	30	2025-12-05	2025-12-16
5.3	Preparación de defensa	10	2025-12-17	2025-12-20

Cuadro 1. Cronograma final del proyecto con jornadas de lunes a sábado (3 h/día).

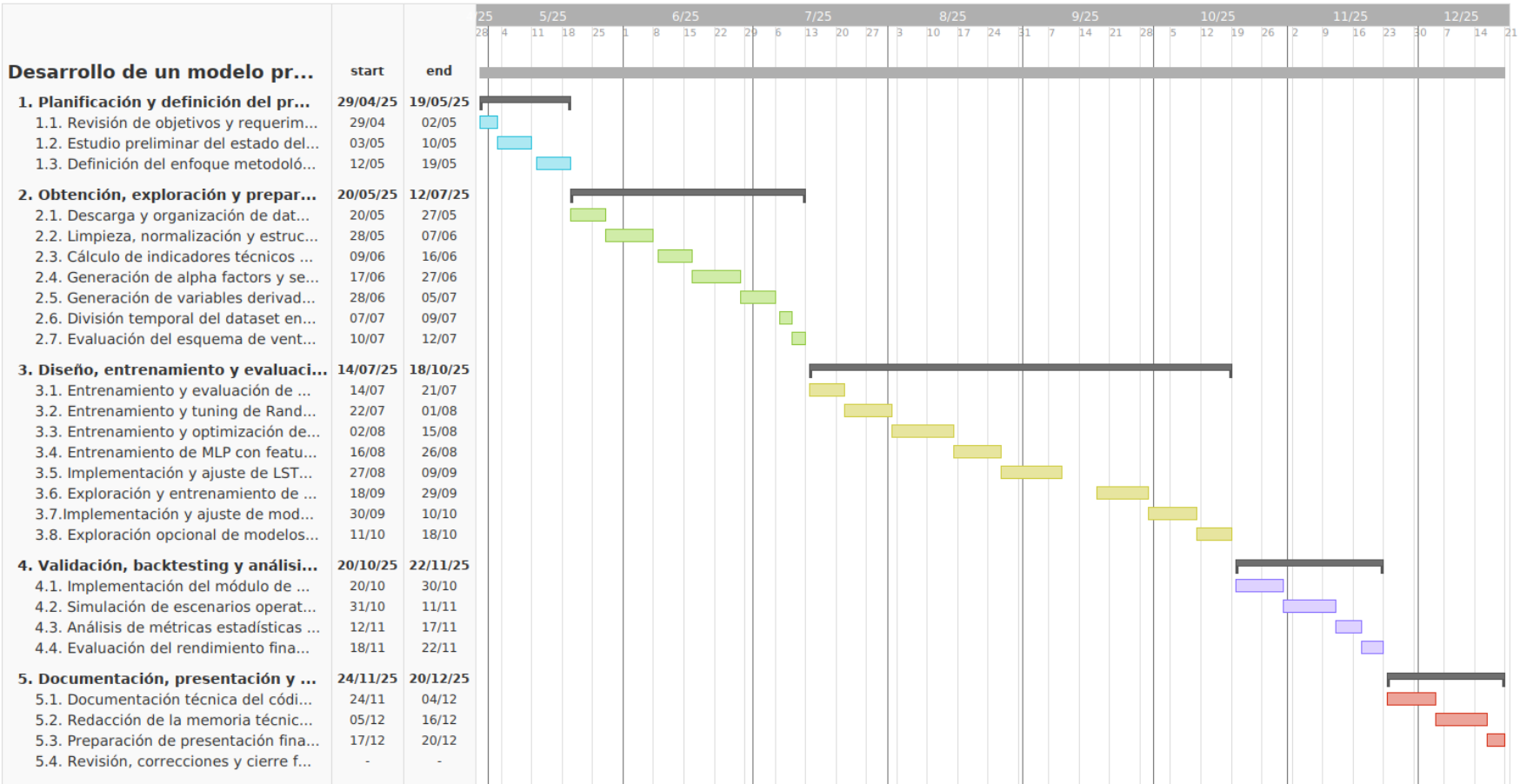


Figura 4. Diagrama de Gantt.

12. Presupuesto detallado del proyecto

A continuación se presenta el presupuesto estimado para el desarrollo del proyecto, expresado en dólares estadounidenses. El mismo se encuentra desglosado en costos directos, que incluyen el tiempo dedicado al desarrollo y servicios complementarios, y costos indirectos, asociados a recursos generales necesarios para la ejecución. Se estima una dedicación total de 600 horas.

COSTOS DIRECTOS			
Descripción	Cantidad	Valor unitario [USD]	Valor total [USD]
Horas de ingeniería	600 horas	25	15000
Servicios Colab Pro (mensual)	7 meses	10	70
SUBTOTAL			15070
COSTOS INDIRECTOS			
Descripción	Cantidad	Valor unitario [USD]	Valor total [USD]
Electricidad	7 meses	30	210
Internet	7 meses	30	210
Espacio de trabajo	7 meses	20	140
SUBTOTAL			560
TOTAL			15630

13. Gestión de riesgos

a) Identificación de los riesgos y estimación de sus consecuencias

Riesgo 1: **Insuficiencia** de datos históricos representativos para entrenar el modelo.

- Severidad (S): 9
Si el conjunto de datos no representa adecuadamente distintos escenarios de mercado (alta volatilidad, caídas bruscas, períodos estables), el modelo podría generalizar mal y producir señales erróneas, comprometiendo todo el proyecto.
- Probabilidad de ocurrencia (O): 3
Ya se dispone de un dataset extenso (1200 días de datos minuto a minuto), pero no se ha verificado aún la cobertura completa de condiciones extremas de mercado.

Riesgo 2: Sobreajuste (overfitting) del modelo a los datos de entrenamiento.

- Severidad (S): 8
El modelo podría tener un rendimiento excelente en validación offline pero fallar en escenarios reales, generando señales incorrectas en operación real o simulada.
- Probabilidad de ocurrencia (O): 5
Debido a la alta capacidad de las redes neuronales profundas y la complejidad del mercado, este riesgo es bastante probable si no se controlan bien las métricas y el proceso de validación cruzada.

Riesgo 3: Incompatibilidades técnicas al integrar fuentes de datos en tiempo real (NinjaTrader u otras).

- Severidad (S): 7
La imposibilidad de conectar o sincronizar correctamente los datos podría limitar la evolución del proyecto hacia un sistema operativo real.
- Probabilidad de ocurrencia (O): 4
Las APIs de datos reales suelen tener restricciones de uso, latencia o formatos no documentados. Aunque esta etapa es exploratoria, podría entorpecer avances si no se resuelve.

Riesgo 4: Dificultades para interpretar los resultados del modelo y sus decisiones.

- Severidad (S): 6
Si el modelo es una “caja negra”, será difícil justificar sus decisiones o ajustar parámetros. Esto impactaría negativamente en la confianza del operador y la utilidad práctica del sistema.
- Probabilidad de ocurrencia (O): 6
Es común en modelos complejos como LSTM o XGBoost si no se incorporan herramientas de interpretabilidad desde el inicio.

Riesgo 5: Limitaciones en el hardware disponible para entrenamiento eficiente.

- Severidad (S): 5
El tiempo de entrenamiento puede volverse excesivo o incluso impedir el ajuste de ciertos modelos. Esto afectaría la posibilidad de explorar variantes.
- Probabilidad de ocurrencia (O): 6
Actualmente no se dispone de GPU funcional en entornos locales, lo que limita la eficiencia, aunque existen opciones como Colab o alquiler de cómputo en la nube.

b) Tabla de gestión de riesgos:

Riesgo	S	O	RPN	S*	O*	RPN*
1. Insuficiencia de datos representativos	9	3	27	6	2	12
2. Sobreajuste del modelo (overfitting)	8	5	40	6	3	18
3. Problemas técnicos con APIs de datos en tiempo real	7	4	28	5	2	10
4. Dificultad para interpretar las decisiones del modelo	6	6	36	4	3	12
5. Limitaciones de hardware para entrenamiento	5	6	30	4	3	12

Criterio adoptado:

Se tomarán medidas de mitigación en los riesgos cuyos números de RPN sean mayores a 25.

c) Plan de mitigación de los riesgos que originalmente excedían el RPN máximo establecido:

Riesgo 1: **Se** ampliará el dataset histórico incorporando más jornadas completas de datos minuto a minuto del índice MNQ, priorizando que incluyan distintos contextos de mercado (alcistas, bajistas y laterales). Se implementará una etapa de análisis de diversidad temporal para asegurar que el conjunto de datos refleje diferentes condiciones de volatilidad y comportamiento del precio, sin necesidad de recurrir a variables externas. También se realizará una selección estratégica de días representativos para las etapas de validación y testeo.

- Severidad (S^*): 6
Al contar con un volumen mayor y mejor distribuido de datos históricos OHLCV, se reduce el riesgo de que el modelo aprenda patrones sesgados o poco generalizables.
- Probabilidad de ocurrencia (O^*): 2
La probabilidad disminuye al incorporar nuevas jornadas históricas dentro del mismo tipo de datos (sin necesidad de nuevas fuentes), lo cual es viable con los recursos disponibles.

Riesgo 2: Se aplicarán técnicas de regularización como dropout y early stopping, además de validación cruzada diaria. También se incluirá una etapa de backtesting detallada para detectar desajustes.

- Severidad (S^*): 6
La severidad se reduce al tener mecanismos para detectar y limitar el sobreajuste antes de usar el modelo en producción.
- Probabilidad de ocurrencia (O^*): 3
La probabilidad disminuye gracias a buenas prácticas de entrenamiento y validación robusta.

Riesgo 3: Se hará una prueba piloto con datos históricos simulados en pseudo-streaming antes de cualquier integración. Se evaluarán varias opciones de proveedor (NinjaTrader, Yahoo Finance, Alpaca) para asegurar compatibilidad.

- Severidad (S^*): 5
La severidad baja al no ser esta funcionalidad crítica para el proyecto inicial.
- Probabilidad de ocurrencia (O^*): 2
La ocurrencia se reduce al planificar una fase de integración opcional y escalonada.

Riesgo 4: Se integrarán herramientas de interpretabilidad como SHAP o LIME, especialmente si se emplean modelos como XGBoost o redes neuronales. Esto permitirá entender qué variables influyen en las predicciones y construir confianza.

- Severidad (S^*): 4
La posibilidad de analizar decisiones del modelo reduce el impacto negativo de una falta de transparencia.
- Probabilidad de ocurrencia (O^*): 3
La probabilidad se reduce al planificar desde el inicio la implementación de herramientas explicativas.

Riesgo 5: Se utilizarán servicios en la nube (Google Colab) para entrenamiento intensivo. También se reducirán cargas computacionales mediante ingeniería de features eficiente y reducción de dimensionalidad.

- Severidad (S^*): 4
El impacto se mitiga al contar con alternativas de cómputo externas sin grandes inversiones.
- Probabilidad de ocurrencia (O^*): 3
La probabilidad disminuye al usar recursos en la nube y reducir requerimientos en etapas tempranas.

14. Gestión de la calidad

A **continuación** se presentan diez requerimientos clave relacionados con el desarrollo del modelo predictivo para el índice MNQ. Para cada uno se detallan las acciones de verificación (análisis interno del modelo y su implementación) y validación (evaluación basada en el comportamiento observado de las predicciones) que permiten asegurar su cumplimiento.

- Req #1: El modelo debe generar predicciones sobre el retorno futuro del índice MNQ.
 - Verificación: se controla que el target esté correctamente calculado a partir de los precios históricos (por ejemplo, retorno acumulado a 15 minutos), y que las funciones utilicen solo datos previos al punto de predicción.
 - Validación: se evalúan las predicciones generadas en distintas jornadas, verificando si siguen el comportamiento general del retorno real.
- Req #2: El modelo debe alcanzar un MAE menor a **0.4** en el conjunto de validación.
 - Verificación: se ejecutan múltiples corridas sobre los conjuntos de validación y testeo, registrando el MAE y su variabilidad.
 - Validación: se analiza si el error medio se mantiene dentro de márgenes aceptables en distintos escenarios de mercado.
- Req #3: El modelo debe generalizar correctamente en contextos de mercado alcistas, bajistas y laterales.
 - Verificación: se etiqueta cada jornada por su tipo de comportamiento predominante y se calcula el rendimiento del modelo por separado para cada grupo.
 - Validación: se observa si el modelo mantiene un desempeño estable y no presenta deterioro significativo en ninguno de los contextos.
- Req #4: Los datos utilizados deben representar fielmente la dinámica intradía del índice MNQ.
 - Verificación: se revisa la cobertura horaria del dataset, la ausencia de valores nulos y la coherencia entre columnas (OHLCV).
 - Validación: se realiza un análisis gráfico y estadístico para confirmar la diversidad temporal y estructural de las jornadas incluidas.
- Req #5: El modelo debe utilizar únicamente variables disponibles hasta el momento de la predicción.
 - Verificación: se audita el pipeline de features para garantizar que no haya uso de datos futuros (leakage).
 - Validación: se comprueba que todas las transformaciones se basan en ventanas temporales anteriores al instante de predicción.
- Req #6: El modelo debe ser evaluado con una metodología de validación adecuada para series temporales.

- Verificación: se implementa validación cruzada por bloques de tiempo o separación por jornadas completas, sin entrecruzamiento.
 - Validación: se observan los resultados de error bajo distintos esquemas para confirmar que el modelo no depende de un único conjunto.
- Req #7: El modelo debe ser interpretable al menos parcialmente mediante técnicas de análisis de variables.
 - Verificación: se aplica SHAP o LIME para estimar la contribución de cada variable a la predicción.
 - Validación: se analizan los factores predominantes y se evalúa si están alineados con comportamientos financieros conocidos.
 - Req #8: El proceso de entrenamiento debe ser reproducible y estar correctamente documentado.
 - Verificación: se utiliza un entorno controlado (por ejemplo, entorno Conda y scripts automatizados) para asegurar consistencia entre ejecuciones.
 - Validación: se ejecuta el pipeline completo en un entorno limpio y se verifica que los resultados obtenidos coincidan con los esperados.
 - Req #9: El modelo debe ser comparado contra al menos un modelo base tradicional.
 - Verificación: se entrena un modelo ARIMA u otro baseline clásico bajo condiciones similares y se comparan métricas como MAE o RMSE.
 - Validación: se analiza si el modelo predictivo propuesto presenta mejoras sustanciales respecto del modelo base.
 - Req #10: El modelo debe incluir un análisis de sensibilidad respecto a sus hiperparámetros principales.
 - Verificación: se explora sistemáticamente el impacto de los principales hiperparámetros sobre el error mediante grid search o técnica equivalente.
 - Validación: se examina si pequeñas variaciones afectan significativamente el rendimiento, para evaluar la robustez del modelo.

15. Procesos de cierre

Para garantizar un cierre ordenado y profesional del proyecto, se llevarán a cabo las siguientes actividades:

- Revisión del cumplimiento de objetivos y requerimientos: se analizará en detalle si el modelo predictivo desarrollado cumple con los objetivos definidos inicialmente, así como con los requerimientos funcionales y no funcionales establecidos. Esta revisión incluirá un contraste entre el alcance planificado y el producto final entregado, validando las métricas de desempeño y utilidad práctica.

- Evaluación del cronograma y uso de recursos: se compararán los tiempos estimados en el plan de proyecto con los tiempos reales dedicados a cada etapa. Asimismo, se evaluará el uso de los recursos asignados, identificando posibles desvíos y sus causas.
- Identificación de técnicas y herramientas utilizadas: se documentarán las estrategias de modelado, preprocesamiento, evaluación y validación aplicadas a lo largo del proyecto. Se destacarán aquellas técnicas que resultaron más efectivas y se registrarán también aquellas que no aportaron valor en este caso específico, como parte del aprendizaje acumulado.
- Memoria de dificultades y soluciones aplicadas: se elaborará un informe sobre los principales obstáculos encontrados durante el desarrollo del modelo, incluyendo problemas técnicos, limitaciones de datos o desafíos computacionales, y se detallarán las soluciones adoptadas.
- Lecciones aprendidas: se generará un informe específico con las lecciones aprendidas durante el desarrollo del sistema de predicción. Este documento incluirá recomendaciones para futuros proyectos similares, identificando buenas prácticas, errores a evitar y oportunidades de mejora.
- Agradecimientos: se incluirá una sección de agradecimientos en la memoria técnica final, reconociendo el acompañamiento académico recibido, así como los aportes técnicos y motivacionales brindados por el equipo docente y colaboradores externos.
- Presentación y defensa pública del trabajo: se realizará una exposición formal del proyecto ante el comité académico de la carrera, presentando los resultados obtenidos, el análisis técnico, las decisiones tomadas y la justificación de las mismas. Esta presentación será coordinada por la FIUBA y estará abierta a los interesados.