

Proyecto Integrado V - Línea de Énfasis

Reporte Final

Presentado por:

Nikol Tamayo Rua

Juliana María Peña Suarez

Docente:

Andres Felipe Callejas Jaramillo

Curso:

Proyecto Integrado V

Programa Ingeniería de Software y Datos

Facultad Ingenierías y Ciencias Agropecuarias

Institución Universitaria Digital de Antioquia

2025

Tabla de Contenido

Resumen	3
Introducción.....	6
Objetivos	7
Metodología.....	9
KPIs Definidos.....	16
Bibliografía	22

Resumen

El presente desarrollo tiene como propósito central el análisis automatizado del comportamiento histórico de las acciones de META Platforms Inc., a través de una arquitectura modular basada en Python. En su primera etapa, se construyó un sistema robusto capaz de recolectar datos directamente desde Yahoo Finanzas mediante técnicas de web scraping, con el fin de alimentar procesos de análisis financiero enfocados en la toma de decisiones informadas.

En la segunda etapa, el proyecto evolucionó desde un tratamiento básico de datos hacia un enfoque integral de análisis predictivo, automatización y visualización avanzada. El dataset original fue enriquecido con nuevos indicadores clave, tales como media móvil, volatilidad, retornos acumulados y variaciones porcentuales, lo que permitió una comprensión más profunda y robusta de la serie temporal de precios de META Platforms Inc.

Se desarrollaron modelos estadísticos como ARIMA y SARIMA, destacándose este último por su capacidad para capturar patrones estacionales y tendencias inherentes en los datos. El modelo SARIMA fue implementado dentro de un módulo específico (modeller.py), que incluye funciones para entrenamiento, predicción y almacenamiento

del modelo en un archivo serializado (model.pkl), facilitando su reutilización y actualización continua.

Para evaluar el desempeño predictivo, se aplicaron múltiples métricas complementarias:

- **MAE (Error Absoluto Medio):** Proporciona una medida directa del error promedio, expresando en unidades de precio cuánto se desvía, en promedio, la predicción respecto al valor real.
- **RMSE (Raíz del Error Cuadrático Medio):** Penaliza más los errores grandes, siendo especialmente útil en contextos financieros donde los desvíos significativos tienen mayor impacto.
- **R² (Coeficiente de Determinación):** Indica el porcentaje de variabilidad en la serie que el modelo logra explicar, reflejando su capacidad predictiva global.
- **MAPE (Error Porcentual Absoluto Medio):** Expresa el error promedio en términos porcentuales, facilitando la interpretación relativa de la precisión del modelo.

Estas métricas combinadas ofrecieron una evaluación completa, equilibrando precisión absoluta, sensibilidad a errores grandes, y capacidad explicativa del modelo. Finalmente, se construyó un dashboard interactivo con Streamlit incluyendo cinco KPIs relevantes, facilitando la exploración dinámica del comportamiento de las acciones de META.

Introducción

Este proyecto busca poder realizar un análisis del comportamiento histórico de acciones de META Platforms Inc. (anteriormente Facebook), una de las compañías tecnológicas más influyentes y representativas del mercado bursátil global. Desde su salida a bolsa en mayo de 2012, META ha sido protagonista en el ámbito digital, destacando por su constante innovación en redes sociales, publicidad digital, y recientemente, en la construcción del metaverso.

Analizar el comportamiento histórico de sus acciones permite identificar patrones, tendencias y reacciones del mercado ante diversos factores económicos, tecnológicos, regulatorios y sociales. Este tipo de análisis resulta fundamental tanto para inversionistas individuales como para instituciones interesados en evaluar oportunidades de inversión y riesgos asociados.

Mediante la técnica de web scraping se busca automatizar el proceso de extracción y recolección de datos históricos para obtener el precio de las acciones de META desde la plataforma Yahoo Finanzas, almacenando la información de manera estructurada, para posteriormente se transformada y obtener unos KPI claves que se visualizaran en Streamlit para comprender mejor este comportamiento.

Objetivos

Objetivo General:

Desarrollar una herramienta automatizada que permita respaldar los datos históricos utilizando el indicador financiero META, enriqueciendo la información mediante su estructuración y visualización, con el fin de modelar su comportamiento para apoyar la toma de decisiones informadas.

Objetivos Específicos:

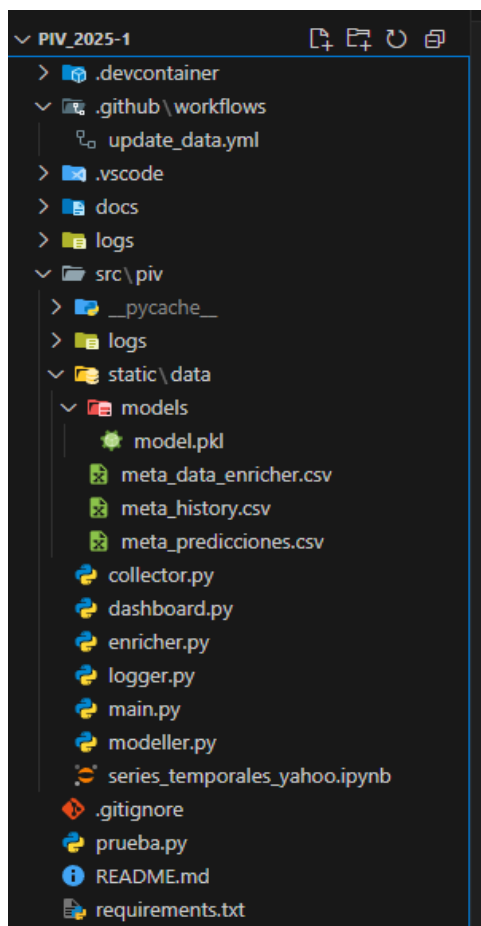
- Recolectar y consolidar datos históricos del precio de las acciones de Meta Platforms Inc. desde plataformas financieras como Yahoo Finanzas, mediante herramientas de scraping.
- Preprocesar, limpiar y almacenar los datos recolectados en formatos estructurados (como archivos .csv) para garantizar su integridad, consistencia y disponibilidad para el análisis posterior.
- Calcular indicadores clave de desempeño (KPIs) tales como precios promedio, valores máximos y mínimos, variaciones porcentuales y otras métricas relevantes para la evaluación del comportamiento bursátil de la empresa.
- Aplicar un modelo de series temporales (ARIMA) para analizar tendencias, estacionalidades y realizar pronósticos sobre la evolución futura del precio de las acciones de Meta Platforms.
- Diseñar un dashboard interactivo en Streamlit, que integre visualizaciones dinámicas (gráficos, tablas y KPIs) para facilitar la interpretación de los resultados tanto exploratorios como predictivos.

- Documentar y presentar el proyecto completo a través de un informe técnico bajo normas APA y un video explicativo que detalle los métodos utilizados, hallazgos clave y recomendaciones.

Metodología

La metodología empleada en este proyecto se centra en el diseño de un sistema automatizado y modular para la recolección, limpieza, almacenamiento y visualización de datos históricos del activo financiero META (acciones de Meta Platforms Inc.), disponible en Yahoo Finanzas. Este sistema fue implementado en Python bajo un enfoque orientado a objetos (OOP), asegurando la escalabilidad, mantenibilidad y separación de responsabilidades del código.

Arquitectura del Sistema



El desarrollo del proyecto PIV_2025-1 se estructuró en siete etapas secuenciales e integradas, combinando técnicas de ingeniería de datos, análisis financiero, modelado estadístico y visualización interactiva. A continuación, se describen cada una de ellas:

1. Estructura del proyecto.

El proyecto está organizado en un sistema modular dentro del directorio `src/piv/`, dividiéndose en los siguientes componentes:

- **Logger (`logger.py`):** Documenta tanto los eventos críticos como los tiempos de ejecución, errores y validaciones.
- **Collector (`collector.py`):** Encargado de la extracción de datos desde Yahoo Finanzas utilizando scraping controlado.
- **Enricher (`enricher.py`):** Almacena los diferentes cálculos e indicadores financieros derivados.
- **Modeller (`modeller.py`):** Contiene la lógica de entrenamiento y predicción utilizando el modelo ARIMA.
- **Dashboard (`dashboard.py`):** Interfaz visual e interactiva desarrollada en Streamlit.
- **Main (`main.py`):** Orquestador principal que automatiza y encadena todo el flujo de ejecución del proyecto.

Además, los datos y modelos se almacenan organizadamente bajo `static/data/`, y los registros de logs se consolidan en el directorio `logs/`.

2. Extracción de Datos

La recolección de datos se realizó desde Yahoo Finanzas a través de técnicas de web scraping utilizando las bibliotecas requests y BeautifulSoup. El proceso incluyó:

- ✓ Conexión directa a la URL oficial del histórico de acciones de Meta Platforms Inc.
- ✓ Aplicación de headers personalizados para evitar bloqueos del servidor.
- ✓ Parsing del HTML para identificar y extraer la tabla de datos históricos mediante selectores CSS.
- ✓ Normalización y estructuración de los datos extraídos en un DataFrame.

Los datos fueron almacenados en el archivo meta_history.csv, ubicado en static/data/, permitiendo persistencia y reutilización sin necesidad de repetir la extracción.

3. Procesamiento y Limpieza de Datos

Una vez obtenidos los datos brutos, se implementa un proceso de limpieza y transformación que incluye:

- ✓ Normalización de nombres de columnas: Se identifican y renombran las columnas para mantener una nomenclatura coherente y en español, independientemente del idioma original.
- ✓ Manejo de valores nulos y duplicados: Eliminación de registros incompletos o inconsistentes.

- ✓ Limpieza de valores numéricos: Se eliminan caracteres no numéricos y se convierten los valores a sus tipos de datos apropiados utilizando expresiones regulares.
- ✓ Manejo de fechas: Se convierten las cadenas de texto a objetos de fecha para facilitar los análisis temporales.

Este proceso permitió construir una base de datos sólida y limpia para su análisis posterior.

4. Enriquecimiento de Datos

En la etapa de enriquecimiento, implementada en `enricher.py`, se generaron nuevas variables a partir de la serie temporal original. Estas variables incluyeron:

- Tasa de variación diaria
- Media móvil de 5 días
- Volatilidad móvil (desviación estándar móvil)
- Retorno acumulado
- Día de la semana (como variable categórica)

Estas características permitieron obtener una visión más profunda del comportamiento de la acción y aportaron valor al proceso de modelado predictivo. Los datos enriquecidos se guardaron en `meta_data_enricher.csv`.

5. Modelado Predictivo Desacoplado

El modelado fue desarrollado en el módulo `modeller.py`, con un enfoque desacoplado para facilitar mantenimiento y reutilización:

- **Entrenamiento:** Se utilizó un modelo ARIMA, adecuado para capturar patrones estacionales, tendencias y ruido en series financieras. El modelo entrenado fue serializado en el archivo `model.pkl`.
- **Predicción:** Mediante la función `predecir()`, el modelo puede cargar el archivo `model.pkl` y generar proyecciones para periodos futuros. Los resultados se exportan a `meta_predicciones.csv`.

La elección del modelo ARIMA se justificó por su capacidad para capturar componentes estacionales, tendencias y ruido en series temporales financieras.

6. Evaluación del Modelo

Se aplicaron las siguientes métricas complementarias para evaluar el desempeño del modelo:

- **MAE (Mean Absolute Error):** Para evaluar el error medio sin importar la dirección del sesgo.

- RMSE (Root Mean Squared Error): Para penalizar errores más grandes de forma más agresiva.
- MAPE (Mean Absolute Percentage Error): Para facilitar la interpretación relativa del error.
- R^2 (Coeficiente de determinación): Para medir la capacidad explicativa del modelo.

Estas métricas fueron seleccionadas por su valor complementario en términos de precisión absoluta, penalización de errores significativos, capacidad explicativa y facilidad de interpretación relativa.

7. Visualización Interactiva en Streamlit

Como producto final del análisis, se desarrolló un dashboard interactivo usando la biblioteca Streamlit. Este dashboard permite:

- Visualizar los precios históricos y proyectados de las acciones de Meta.
- Explorar los indicadores financieros enriquecidos.
- Analizar las predicciones del modelo de forma dinámica y accesible.

El dashboard constituye una herramienta intuitiva para stakeholders y usuarios no técnicos, facilitando la comprensión del comportamiento bursátil de la empresa.

8. Documentación y presentación

Todo el proceso fue documentado de forma rigurosa y almacenado en /docs/.

La documentación incluye:

- Un informe técnico bajo normas APA, detallando cada etapa del análisis.
- Video explicativo con soporte visual para presentar los resultados, decisiones metodológicas y hallazgos clave del proyecto.
- Presentación que completa a la descripción del proyecto junto con el acceso directo a las visualizaciones de los resultados

KPIs Definidos

1. Retorno Diario (%)

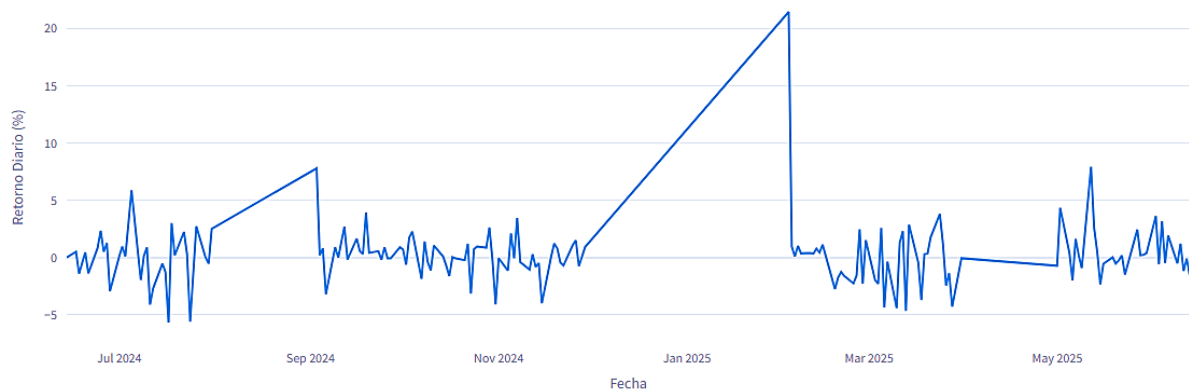
¿Qué busca?

Evaluar el comportamiento diario de la acción de META para identificar la rentabilidad inmediata y la sensibilidad del precio a eventos recientes.

¿Qué mide?

La variación porcentual entre el precio de cierre de un día respecto al cierre del día anterior.

Retorno Diario (%) - Evolución Temporal



Análisis del grafico

Se observan picos positivos y negativos que superan el $\pm 5\%$, con un pico extraordinario cercano al 20% a inicios de 2025. Esto indica que hubo algún evento importante (como lanzamiento de un nuevo modelo de IA integrado, cambios en la directiva, o hechos macroeconómicos) que generó un fuerte impacto en el valor de la

acción. Posteriormente, el retorno diario muestra mayor estabilidad, aunque continúa habiendo movimientos bruscos esporádicos.

2. Tasa de Variación Apertura-Cierre (%)

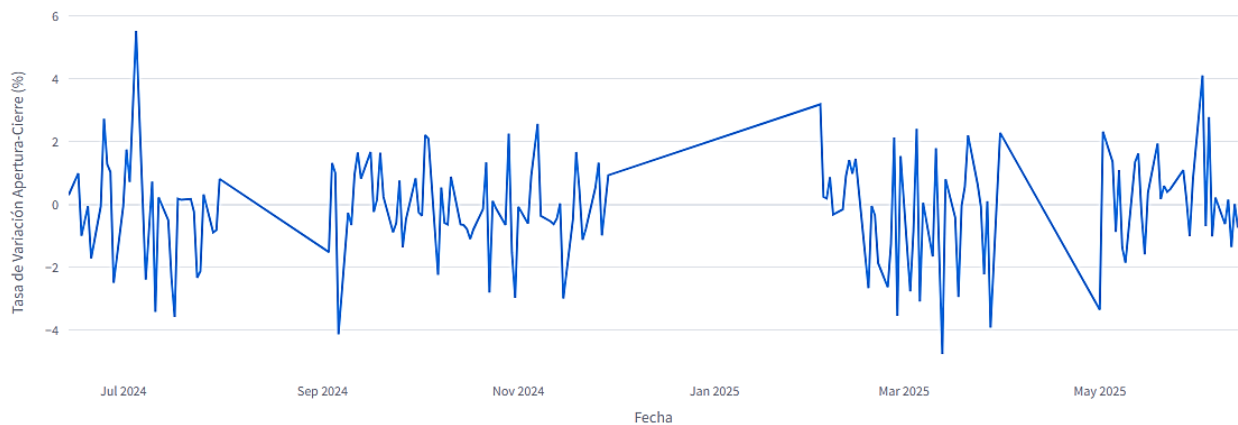
¿Qué busca?

Detectar la dinámica intradía del mercado, observando cómo varía el precio desde la apertura hasta el cierre en un mismo día bursátil.

¿Qué mide?

La diferencia porcentual entre el precio de apertura y el de cierre del mismo día. Indica si la acción ganó o perdió valor durante la jornada.

Tasa de Variación Apertura-Cierre (%) - Evolución Temporal



Análisis del grafico

Se identifican picos positivos y negativos superiores al 4%, señalando jornadas con alta especulación o noticias de impacto durante el día. Un fuerte movimiento positivo

puede reflejar una corrección optimista después de una apertura a la baja, y viceversa. Este indicador nos permitió ver qué tan reactivas son las decisiones de los inversionistas en el transcurso del día.

3. Retorno Acumulado (%)

¿Qué busca?

Mostrar la rentabilidad total acumulada de la acción META en el periodo analizado, evaluando si ha generado ganancia o pérdida a lo largo del tiempo.

¿Qué mide?

La suma acumulada de los retornos diarios desde el punto de partida. Se expresa en porcentaje.



Análisis del grafico

Se observa un crecimiento sostenido hasta superar el 40% de retorno acumulado a inicios de 2025. Luego se presenta una caída significativa en el segundo trimestre,

seguida de una nueva recuperación. Esta métrica es útil para visualizar el panorama general del desempeño y ayuda a detectar tendencias de crecimiento o correcciones sostenidas.

4. Media Móvil 5 Días (\$)

¿Qué busca?

Suavizar las fluctuaciones diarias para identificar tendencias generales de precios en el corto plazo.

¿Qué mide?

El promedio de los precios de cierre de los últimos 5 días. Ayuda a reducir el “ruido” diario y resaltar la tendencia.

Media Móvil 5 Días (\$) - Evolución Temporal



Análisis del grafico

Se evidencia una tendencia positiva general con puntos de consolidación y corrección. El pico más alto ocurre entre febrero y marzo de 2025, seguido de una leve caída y posterior recuperación. Este comportamiento refleja que, a pesar de los altibajos diarios, el activo presenta momentos de confianza sostenida en el mercado.


5. Volatilidad (Rolling 5D)

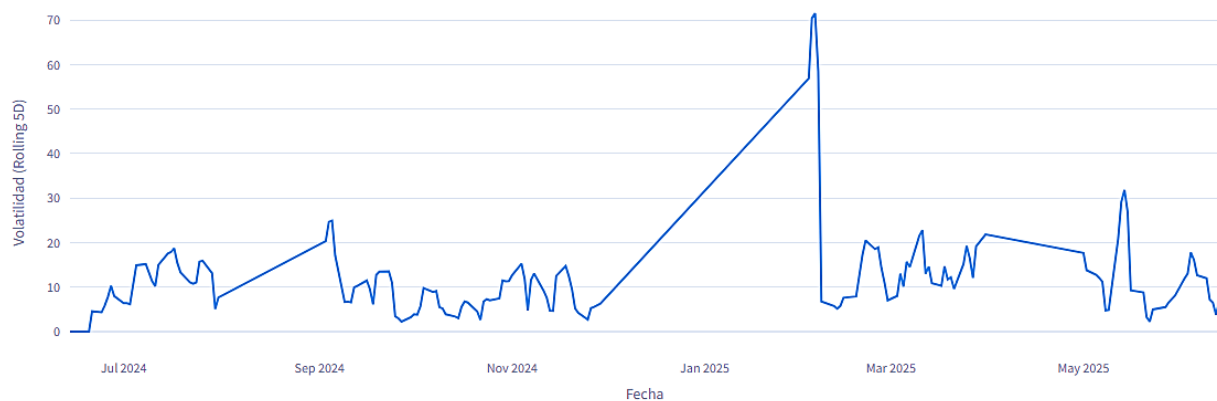
¿Qué busca?

Medir el riesgo o inestabilidad del activo en ventanas móviles de 5 días, para entender cuán variable o predecible es su comportamiento.

¿Qué mide?

La desviación estándar de los retornos diarios en una ventana de 5 días. Cuanto mayor sea la volatilidad, mayor es la incertidumbre y el riesgo percibido por los inversores.

 Volatilidad (Rolling 5D) - Evolución Temporal



Análisis del grafico

Se observa un pico marcado a inicios de 2025, con valores cercanos al 70%, lo cual coincide con eventos donde también hubo grandes retornos diarios. La alta volatilidad sugiere incertidumbre o expectativas divididas entre los actores del mercado. Posteriormente, el gráfico muestra una reducción progresiva en la volatilidad, lo cual puede indicar mayor estabilidad y confianza en el comportamiento del activo.

Link directo dashboard: <https://piv2025-1-gr8ygvipouohoksumtpf2s.streamlit.app/>

Bibliografía

Scrapy. (2025). Scrapy Documentation. Recuperado de <https://docs.scrapy.org/>

Python Software Foundation. (2025). Python (Versión 3.10) [Software].
Recuperado de <https://www.python.org/>

Meta Platforms Inc. (2025). META - Meta Platforms, Inc. Historical Data. Yahoo Finance. <https://finance.yahoo.com/quote/META/history>

BeautifulSoup. (2025). BeautifulSoup Documentation. Recuperado de <https://www.crummy.com/software/BeautifulSoup/>

Git. (2025). Git Documentation. Recuperado de <https://git-scm.com/doc>

Microsoft. (2025). Visual Studio Code Documentation. Recuperado de <https://code.visualstudio.com/doc>

McKinney, W. (2022). Python for Data Analysis (3ra ed.). O'Reilly Media.

Cedeño, L. (2021). Minería de datos aplicada con Python. Alfaomega Grupo Editor.