

Unidad 1. Evidencia de aprendizaje (EA1)

Proyecto Integrado V - Línea de Énfasis

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

Tabla de Contenido

Resumen	3
Introducción	5
Objetivos	6
Metodología	7
Bibliografía	13

Resumen

El presente desarrollado 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 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, balanceando precisión absoluta, sensibilidad a errores grandes, y capacidad explicativa del modelo. Finalmente, se construyó un dashboard interactivo en Power BI con cinco KPI 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 Power BI para comprender mejor este comportamiento.

Objetivos

Objetivo General:

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

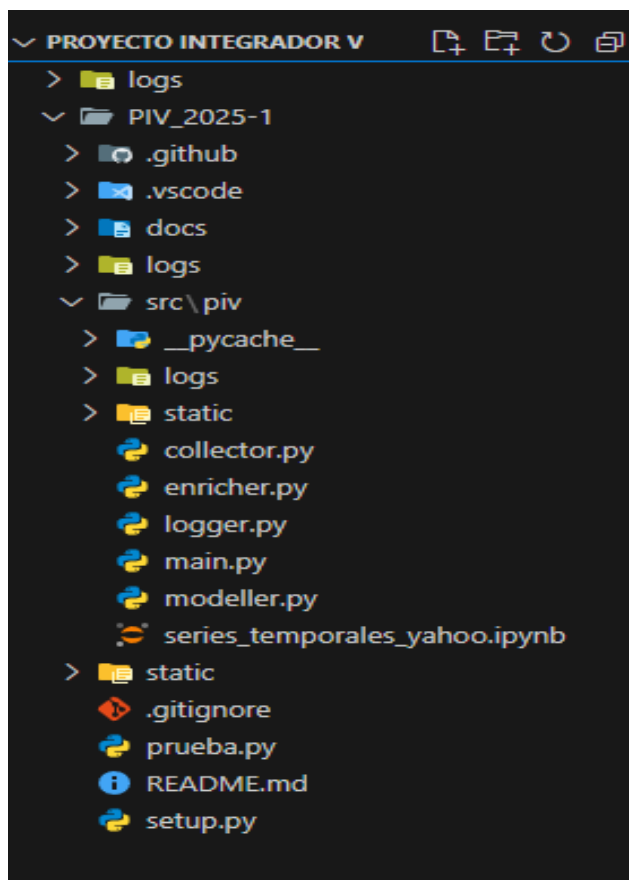
Objetivos Específicos:

- Obtener los datos históricos del precio de las acciones de META directamente desde Yahoo Finanzas.
- Preprocesar y almacenar los datos recolectados en un archivo .csv y una base de datos SQLite sin pérdida de histórico.
- Calcular indicadores clave de desempeño (KPI) como precios promedios, máximos, mínimos y variaciones porcentuales.
- Presentar los resultados obtenidos en formato tabular y visual mediante gráficos, con soporte en un video explicativo y un informe en formato APA.

Metodología

La metodología empleada en este proyecto se centra en el diseño de un sistema automatizado y modular para la recolección, limpieza, almacenamiento y trazabilidad 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



1. El proyecto se estructura en tres módulos principales:

- Módulo de Registro (Logger): Proporciona capacidades de registro detallado para el seguimiento de la ejecución y la identificación de posibles errores.
- Módulo de Recolección (Collector): Responsable de la extracción de datos mediante técnicas de web scraping.
- Script Principal (Main): Orquesta la ejecución de los componentes y realiza transformaciones adicionales sobre los datos obtenidos.

2. Proceso de Extracción de Datos

La extracción de datos se realiza mediante técnicas de web scraping utilizando las bibliotecas requests y BeautifulSoup de Python. El proceso sigue estos pasos:

- ✓ Se establece una conexión con la URL de Yahoo Finance para META (<https://es.finance.yahoo.com/quote/META/history/>).
- ✓ Se utiliza un encabezado de usuario personalizado para evitar restricciones de acceso.
- ✓ Se analiza el contenido HTML para localizar la tabla de datos históricos mediante selectores CSS específicos.
- ✓ Se extraen los encabezados y filas de la tabla, preservando la estructura original de los datos.

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.
- ✓ 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 análisis temporales.
- ✓ Filtrado de Datos Inconsistentes: Se eliminan filas con valores faltantes o incorrectos en campos críticos.

Persistencia de Datos

Los datos procesados se almacenan en formato CSV en una ubicación predefinida (`src/edu_piv/static/data/meta_history.csv`), lo que permite su posterior utilización para análisis y visualizaciones sin necesidad de repetir el proceso de extracción.

Registro y Monitoreo

Todo el proceso es monitoreado mediante un sistema de registro que documenta:

- Tiempo de ejecución
- Éxitos y fracasos en la recolección de datos
- Dimensiones y características de los conjuntos de datos obtenidos
- Errores y excepciones producidos durante la ejecución

Este enfoque metodológico garantiza la reproducibilidad del proceso y facilita la detección y corrección de errores, permitiendo un refinamiento continuo del sistema de extracción y análisis de datos.

4. Exploración y Enriquecimiento de Datos

Tras la recolección, se realizó una exploración de los datos históricos de META, identificando tendencias generales, presencia de outliers, valores nulos y patrones temporales.

Luego, se procedió al enriquecimiento de la serie temporal, implementado en el módulo `enricher.py`. Se generaron nuevas variables derivadas, incluyendo:

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

El resultado se consolidó en el dataset `meta_data_enricher.csv`, que fue utilizado tanto para el modelado como para la visualización.

5. Modelado Predictivo Desacoplado

Se diseñó un sistema de modelado desacoplado en el módulo `modeller.py`, que contiene las funciones:

- `entrenar()`: realiza el entrenamiento del modelo SARIMA, serializa el artefacto y lo guarda como `model.pkl` en el directorio `static/models/`.
- `predecir()`: permite cargar el modelo previamente entrenado y generar predicciones sobre datos futuros.

La elección del modelo SARIMA 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)
- RMSE (Root Mean Squared Error)
- R^2 (Coeficiente de Determinación)
- MAPE (Mean Absolute Percentage Error)

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 Power BI

Como culminación del ciclo analítico, se diseñó un dashboard interactivo en Power BI que permite explorar visualmente la evolución de las acciones de META y comprender el comportamiento del activo a través de indicadores clave.

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.