

Proyecto Integrado V - Linea de Énfasis (Entrega 2)

Estudiante:

Sergio Raúl Bustos

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

Junio

2025

Resumen

Este proyecto tiene como objetivo desarrollar un sistema automatizado para la recopilación y almacenamiento de datos históricos del valor de las acciones de Apple Inc. (AAPL), desde su debut en la bolsa el 12 de diciembre de 1980 hasta la fecha actual. La obtención de los datos se llevó a cabo a través del sitio web Yahoo Finanzas, utilizando para ello Python y técnicas de Web Scraping. El procesamiento de los datos se realizó con la biblioteca Pandas, y los resultados fueron guardados tanto en archivos CSV como en una base de datos SQLite. Como producto final, se construyó una aplicación funcional que permite consultar y almacenar esta información de forma automatizada, organizada en torno a un flujo de ejecución principal (Main), con capacidad de integrarse en entornos productivos mediante herramientas de integración continua como GitHub Actions.

Introducción

En este proyecto se presenta un sistema automatizado para la recopilación, almacenamiento y análisis de datos históricos del precio de las acciones de Apple Inc. (AAPL) en el mercado bursátil. El objetivo principal es facilitar el acceso y tratamiento de esta información con fines de análisis financiero, permitiendo comprender de forma más detallada la evolución y comportamiento de la empresa en el mercado a lo largo del tiempo. La motivación radica en el creciente interés por el análisis de datos históricos como base para la toma de decisiones estratégicas, estudios de tendencias y desarrollo de modelos predictivos en el ámbito bursátil.

Objetivo General

Desarrollar una interfaz interactiva en formato Jupyter Notebook que permita la visualización, análisis y exploración dinámica de los indicadores clave de desempeño (KPIs) financieros de las acciones de Apple Inc., facilitando la interpretación de métricas como tasa de variación, medias móviles, volatilidad, retorno acumulado y desviación estándar, para apoyar la toma de decisiones basada en datos históricos.

Objetivos Específicos

1. Implementar la extracción y transformación de datos históricos de Apple Inc. desde Yahoo Finanzas, asegurando la calidad y limpieza de la información para su análisis.
2. Calcular y enriquecer el dataset con un conjunto mínimo de cinco KPIs relevantes: tasa de variación diaria, medias móviles a corto y largo plazo, volatilidad a distintos horizontes, retorno acumulado y desviación estándar.
3. Desarrollar gráficos interactivos y visualizaciones dinámicas en el notebook para representar cada KPI, permitiendo al usuario explorar patrones temporales y relaciones entre indicadores.

Metodología

La metodología empleada en esta etapa inicial del proyecto consistió en la creación de un script automatizado enfocado en la extracción, transformación y almacenamiento de datos históricos de las acciones de Apple Inc. (AAPL). Este proceso se dividió en las siguientes fases:

- **Obtención de datos:** Se utilizó la librería requests para realizar peticiones HTTP al portal Yahoo Finanzas, incorporando un encabezado User-Agent que imita la actividad de un navegador web. El contenido HTML recuperado fue interpretado con la ayuda de BeautifulSoup, permitiendo ubicar y extraer la tabla correspondiente a los registros históricos del precio de las acciones de Apple.
- **Tratamiento de datos:** Posteriormente, los datos fueron estructurados en un DataFrame mediante la biblioteca Pandas. Durante esta etapa se procedió al ajuste de nombres de columnas, verificación de integridad de los datos y eliminación de registros vacíos o incorrectos, asegurando así un formato limpio y apto para análisis posteriores.
- **Monitoreo de ejecución:** Se integró un sistema de registro personalizado mediante la librería logging, con el fin de documentar la actividad del programa. Este sistema captura eventos clave como fallos de conexión, inicialización de clases, y la correcta creación de los archivos procesados.
- **Persistencia de la información:** Los datos ya transformados se almacenaron en dos formatos: como archivos .csv y dentro de una base de datos SQLite. Ambos tipos de archivos se guardan automáticamente en directorios generados por el sistema, específicamente en las carpetas data y static.
- **Diseño modular del sistema:** La aplicación fue construida con una estructura modular, distribuyendo sus funcionalidades en clases independientes: **Logger**, **Collector** y **main**. Esta arquitectura mejora la legibilidad del código, facilita las pruebas unitarias y permite futuras ampliaciones. La ejecución general se controla desde el módulo main, el cual también gestiona la creación de carpetas y la generación de registros de actividad.
- El archivo **Modeler** implementa una clase encargada de entrenar y guardar un modelo de regresión basado en Random Forest para predecir el precio de cierre de acciones de Apple Inc., utilizando variables técnicas y de volatilidad; además, permite generar predicciones a partir de nuevos datos.

- **transformacion** implementa una clase encargada de preparar y enriquecer un conjunto de datos financieros. Su función principal realiza la limpieza de datos originales, corrigiendo duplicados, nulos y formatos erróneos; calcula indicadores técnicos como retornos logarítmicos diarios, medias móviles y volatilidad a diferentes horizontes temporales; elimina registros incompletos en las métricas clave y finalmente completa valores faltantes en la tasa de interés de referencia, garantizando así un dataset robusto y listo para análisis o modelado predictivo.

Transformación y enriquecimiento de datos

Una vez recopilada la información histórica desde Yahoo Finanzas, se llevó a cabo un proceso de transformación y enriquecimiento de los datos asociados a Apple Inc. con el objetivo de optimizar su análisis y facilitar la generación de modelos predictivos. Este tratamiento de datos permite representar indicadores clave de desempeño (KPI) a través de archivo ipynb, donde se analiza el comportamiento de las acciones de Apple.

Para mejorar la capacidad analítica, se incorporaron las siguientes métricas:

- **Retorno logarítmico diario (retorno_log_diario):** Permite cuantificar el cambio porcentual entre precios de cierre consecutivos, capturando variaciones relativas de forma simétrica y minimizando la distorsión por valores atípicos.
- **Media móvil de 7 días (media_movil_7d):** Suaviza las oscilaciones diarias mediante el promedio de los últimos siete días, facilitando la identificación de tendencias de corto plazo.
- **Media móvil de 30 días (media_movil_30d):** Calcula el promedio de cierre de los últimos 30 días, ofreciendo una visión de tendencias más prolongadas.
- **Volatilidad de 7 días (volatilidad_7d) y de 30 días (volatilidad_30d):** Miden la dispersión de los precios respecto a su media en ambos horizontes temporales, proporcionando una evaluación del riesgo y la estabilidad del mercado en el corto y largo plazo, respectivamente.