

PROYECTO INTEGRADO V - LINEA DE ÉNFASIS

Automatización de la recolección de datos históricos financieros del Grupo Aval mediante Github Actions y Python

Estudiante:

Jean Carlos Páez Ramírez (Grupo: Preica2501b020128)

Docente:

Andrés Felipe Callejas Jaramillo

Especialista en Analítica de Datos

Curso:

Proyecto Integrado V - Línea de Énfasis

Programa de Ingeniería en Desarrollo de Software y Datos

Facultad Ingeniería y Ciencias Agropecuarias

Institución Universitaria Digital de Antioquia

Contenido

Resumen.....	3
Introducción	4
Metodología	5
Infraestructura Técnica	7
Resultados Técnicos y Modelo Predictivo.....	8
Conclusiones	9
Bibliografía	10

Resumen

Esta propuesta presenta el desarrollo de un sistema profesional automatizado para la recolección, almacenamiento, enriquecimiento y trazabilidad de datos bursátiles del Grupo Aval, el conglomerado financiero más grande de Colombia. A través del uso de Python, GitHub Actions y técnicas de análisis técnico, se logra una solución robusta que permite la actualización diaria de datos financieros desde Yahoo Finance, el procesamiento mediante programación orientada a objetos (OOP), y el registro estructurado de logs para garantizar la transparencia operativa. Esta herramienta establece una base sólida para el análisis de indicadores financieros, modelado predictivo y toma de decisiones estratégicas en tiempo real.

Palabras clave: Grupo Aval, automatización, GitHub Actions, yfinance, Python, recolección de datos históricos, análisis económico.

Introducción

El análisis de series de tiempo financieras constituye un elemento fundamental en la toma de decisiones dentro del ámbito económico y empresarial. En este contexto, contar con datos históricos confiables, actualizados y fácilmente accesibles es esencial para elaborar modelos predictivos y realizar estudios de comportamiento de activos bursátiles. Grupo Aval Acciones y Valores S.A. (AVAL) representa una de las entidades más relevantes del sistema financiero colombiano. Su comportamiento bursátil ha sido afectado por eventos significativos como la pandemia del COVID-19 en 2020, la escisión y venta de BAC Credomatic en 2022, y las implicaciones reputacionales derivadas del caso Odebrecht en 2023.

Frente a este contexto de alta volatilidad e incertidumbre, esta solución automatizada permite mantener una trazabilidad diaria sobre la evolución del activo, facilitando el análisis técnico y la predicción de precios con enfoque estratégico. El sistema se construye con tecnologías de código abierto y se integra a flujos de trabajo continuos en la nube para operar sin intervención humana.

Metodología

El desarrollo del proyecto se llevó a cabo siguiendo una metodología basada en la implementación iterativa de componentes funcionales, estructurada de la siguiente manera:

Selección del activo financiero: Se eligió el Grupo Aval (símbolo bursátil AVAL) como objeto de análisis, dada su relevancia en el mercado financiero colombiano. Los datos históricos de AVAL se obtuvieron desde Yahoo Finanzas, una fuente confiable y ampliamente utilizada para información financiera (Yahoo Finance, s.f.).

Desarrollo del colector de datos: Se implementó un script en Python utilizando las librerías `yfinance` y `pandas`. Este script obtiene datos históricos de AVAL desde Yahoo Finanzas y los almacena en un archivo `historical.csv`. La biblioteca `pandas` proporciona estructuras de datos de alto rendimiento y herramientas de análisis fáciles de usar para el lenguaje de programación Python (Reback et al., 2020).

Módulo de logging personalizado: Para garantizar la trazabilidad del proceso, se desarrolló un sistema de registros basado en la librería `logging`, con un manejador personalizado (`csv_logger.py`) que guarda cada ejecución en un archivo CSV (`log_data.csv`). Este enfoque permite mantener un historial detallado de las ejecuciones, facilitando la auditoría y el diagnóstico de posibles fallos (Python Software Foundation, 2020).

Automatización con GitHub Actions: Se configuró un flujo de trabajo (`update_data.yml`) que ejecuta el script diariamente a las 12:00 UTC. Este flujo permite la descarga automática de nuevos datos, la actualización del archivo histórico y el registro de la ejecución en los logs. GitHub Actions permite automatizar, personalizar y ejecutar flujos de trabajo de desarrollo de software directamente en el repositorio de GitHub (GitHub, s.f.).

Estructura del proyecto: El repositorio se organizó con una estructura modular clara y documentada, incluyendo el archivo `setup.py` para facilitar la instalación del proyecto como paquete.

Esta metodología permitió el diseño de una solución escalable y fácilmente integrable en proyectos de análisis financiero o académico.

Infraestructura Técnica

- ✓ **Lenguaje principal:** Python 3.10
- ✓ **Bibliotecas clave:** yfinance, pandas, logging, streamlit, statsmodels
- ✓ **Automatización CI/CD:** GitHub Actions
- ✓ **Logging de trazabilidad:** ejecución con registro dual (.log y .csv)
- ✓ **Organización del repositorio:** src/ (código), docs/, workflows/, dashboard.py
- ✓ **Dashboard de visualización:** Streamlit con indicadores técnicos y resultados del modelo

ARIMA

Resultados Técnicos y Modelo Predictivo

El modelo predictivo implementado fue un ARIMA(3,1,1), entrenado sobre precios ajustados del activo AVAL. Se obtuvo un desempeño destacado en las métricas clave:

- ✓ **MAE:** 7.83 → Error promedio absoluto.
- ✓ **RMSE:** 11.14 → Penalización de errores grandes.
- ✓ **MAPE:** 2.48% → Error porcentual bajo.
- ✓ **R²:** 0.921 → El modelo explica más del 92% de la variabilidad observada.

Los principales indicadores técnicos incluidos en el dashboard son:

- ✓ **RSI (Relative Strength Index):** detección de sobrecompra o sobreventa.
- ✓ **Cruces de medias móviles (SMA 21 y 50):** señales de entrada/salida.
- ✓ **Bandas de Bollinger:** rangos de volatilidad.
- ✓ **Volatilidad 7 días.**
- ✓ **Retorno acumulado con dividendos reinvertidos.**

Conclusiones

El desarrollo de Grupo Aval Tracker representa una solución técnica y estratégica frente a los desafíos del análisis financiero moderno. La automatización total del proceso —desde la recolección de datos hasta la visualización de resultados— permite operar sin intervención humana, garantizando eficiencia, trazabilidad y actualización continua.

El modelo predictivo ARIMA(3,1,1), respaldado por métricas como un R^2 de 0.921 y un MAPE inferior al 2.5 %, ofrece una precisión sólida para anticipar el comportamiento del precio bursátil en el corto plazo. Esto proporciona una herramienta valiosa para inversionistas, analistas financieros y tomadores de decisión que requieren señales confiables y en tiempo real.

La integración de indicadores técnicos como RSI, medias móviles y volatilidad mejora el análisis contextual de la acción de Grupo Aval, permitiendo identificar puntos de entrada o salida y entender la dinámica del activo bajo distintas condiciones del mercado. Además, el diseño modular y escalable del sistema facilita su adaptación a otros activos o sectores, consolidando esta propuesta como una base sólida para soluciones avanzadas en analítica bursátil.

Bibliografía

Danta Analytics. (s.f.). APA - Automatización de rutinas de análisis. Recuperado de <https://dantaanalytics.com/servicios/apa/>

GitHub. (s.f.). GitHub Actions documentation. Recuperado de <https://docs.github.com/actions>

Python Software Foundation. (2020). logging — Logging facility for Python. Recuperado de <https://docs.python.org/3/library/logging.html>

Reback, J., McKinney, W., & Wes McKinney. (2020). pandas: powerful Python data analysis toolkit. Recuperado de <https://pandas.pydata.org/>

Yahoo Finance. (s.f.). Yahoo Finance - AVAL. Recuperado de <https://es-us.finanzas.yahoo.com/quote/AVAL/>