Proyecto Integrado V - Linea de Énfasis (Entrega 1)Proyecto Integrado V - PREICA2501B020128

Programa

Ingeniería de Software y Datos

Profesor

Andrés Callejas

Estudiante

Miguel ángel Leal Beltrán

Universidad

IUDigital de Antioquia

Bogotá 2025

Contenido

[Resumen 2](#_Toc197892360)

[Introducción 3](#_Toc197892361)

[Metodología 4](#_Toc197892362)

[Resultados y Discusión 5](#_Toc197892363)

[Conclusiones 6](#_Toc197892364)

[Referencias 6](#_Toc197892365)

### Resumen

Este informe documenta el desarrollo de un sistema automatizado para la recolección, procesamiento y almacenamiento de datos históricos del Índice de Volatilidad CBOE (VIX), comúnmente conocido como el "índice del miedo". El sistema implementa un enfoque de programación orientada a objetos utilizando Python, con capacidades de recuperación ante fallos mediante múltiples métodos de obtención de datos. La solución incluye un pipeline completo de datos que se ejecuta diariamente mediante GitHub Actions, garantizando la actualización continua del conjunto de datos y permitiendo su enriquecimiento con métricas técnicas relevantes. El sistema mantiene la trazabilidad de las operaciones realizadas mediante un sistema de logs estructurado y almacena los resultados tanto en formato CSV para análisis posteriores como en un repositorio de control de versiones que mantiene el historial de cambios y actualizaciones.

### Introducción

El Índice de Volatilidad CBOE (VIX) es un indicador económico crucial que mide la volatilidad implícita del mercado bursátil estadounidense, específicamente del índice S&P 500, para los próximos 30 días. Creado por el Chicago Board Options Exchange (CBOE), el VIX se calcula a partir de los precios de opciones del S&P 500 y refleja las expectativas del mercado sobre la volatilidad futura. A menudo denominado "índice del miedo", valores elevados del VIX (generalmente por encima de 30) indican períodos de alta incertidumbre y estrés en los mercados, mientras que valores bajos (por debajo de 20) sugieren condiciones de mercado más estables y optimistas.

El seguimiento y análisis del VIX resulta fundamental para diversas aplicaciones financieras:

* Gestión de riesgos y estrategias de cobertura
* Identificación de oportunidades de inversión durante períodos de alta volatilidad
* Predicción de posibles cambios de tendencia en el mercado
* Análisis macroeconómico y de sentimiento de mercado

Sin embargo, la recolección manual de estos datos presenta numerosos desafíos, incluyendo:

* Necesidad de intervención humana regular
* Posibles inconsistencias en la frecuencia de actualización
* Dificultad para mantener registros históricos completos
* Limitaciones para realizar un seguimiento preciso de cambios a lo largo del tiempo

El presente proyecto aborda estos desafíos mediante la automatización de la recolección de datos históricos del VIX, utilizando técnicas de programación orientada a objetos y prácticas modernas de ingeniería de software. El sistema desarrollado garantiza la persistencia y trazabilidad de los datos, permitiendo análisis continuos y consistentes a lo largo del tiempo.

### Metodología

#### Arquitectura del Sistema

El sistema se ha desarrollado utilizando una arquitectura modular basada en clases con responsabilidades claramente definidas, siguiendo principios de programación orientada a objetos. La arquitectura se divide en tres componentes principales:

1. **Sistema de Logging (Logger)**: Implementado como una clase independiente, proporciona capacidades de registro estructurado para documentar todas las operaciones del sistema y facilitar la depuración y el seguimiento de la ejecución.
2. **Colector de Datos (Collector)**: Responsable de la obtención de datos históricos del VIX desde Yahoo Finance, implementa múltiples estrategias de recolección para garantizar la robustez y continuidad del servicio, incluso ante fallos temporales de las API.
3. **Procesador de Enriquecimiento (Enricher)**: Calcula indicadores técnicos adicionales a partir de los datos crudos, añadiendo valor analítico al conjunto de datos mediante métricas comúnmente utilizadas en análisis financiero.

#### Tecnologías Utilizadas

El sistema se ha implementado utilizando las siguientes tecnologías:

* **Python 3.9**: Lenguaje de programación principal, elegido por su amplio ecosistema para análisis de datos.
* **pandas**: Biblioteca para manipulación y análisis de datos tabulares.
* **pandas\_datareader**: Interfaz para acceder a datos financieros desde diversas fuentes.
* **yfinance**: Biblioteca alternativa para obtener datos de Yahoo Finance.
* **numpy**: Soporte para operaciones matemáticas avanzadas.
* **GitHub Actions**: Plataforma de CI/CD utilizada para automatizar la ejecución diaria del sistema.
* **Control de versiones Git**: Permite mantener un historial completo de las actualizaciones de datos.

#### Estrategia de Recolección de Datos

El sistema implementa una estrategia de recolección robusta con múltiples capas de contingencia:

1. **Método Principal (pandas\_datareader)**: Intenta obtener datos históricos del VIX utilizando la biblioteca pandas\_datareader, que proporciona una interfaz estándar para acceder a datos financieros.
2. **Método Alternativo (yfinance)**: En caso de fallo del método principal, el sistema recurre automáticamente a la biblioteca yfinance, que implementa técnicas diferentes para acceder a los datos de Yahoo Finance.
3. **Datos Sintéticos de Respaldo**: Como última opción, si ambos métodos fallan, el sistema puede generar datos sintéticos basados en características estadísticas históricas del VIX, permitiendo pruebas y desarrollo continuos.

#### Procesamiento y Enriquecimiento de Datos

Una vez obtenidos los datos históricos básicos (fecha, apertura, cierre, máximo, mínimo, volumen), el sistema calcula métricas financieras adicionales que aportan valor analítico:

* **Variación Diaria (%)**: Cambio porcentual diario en el precio de cierre.
* **Medias Móviles (MA5, MA20)**: Tendencias a corto y medio plazo (5 y 20 días).
* **Volatilidad**: Desviación estándar móvil de 20 días para medir la estabilidad.
* **Índice de Fuerza Relativa (RSI)**: Indicador de momentum para identificar condiciones de sobrecompra o sobreventa.
* **Clasificación de Nivel**: Categorización del valor del VIX como Alto, Normal o Bajo basado en percentiles históricos.

#### Persistencia y Automatización

Los datos recolectados y enriquecidos se almacenan en dos formatos CSV:

* vix\_data.csv: Datos históricos básicos obtenidos directamente de la fuente.
* vix\_data\_enricher.csv: Conjunto de datos enriquecido con métricas adicionales.

La automatización de la recolección se implementa mediante GitHub Actions, configurado para ejecutar el sistema diariamente a medianoche (UTC). Esto garantiza actualizaciones regulares del conjunto de datos sin intervención manual.

### Resultados y Discusión

La implementación del sistema ha demostrado varias ventajas significativas sobre los métodos manuales de recolección:

1. **Robustez**: La estrategia de múltiples métodos de recolección garantiza la continuidad del servicio incluso cuando una fuente de datos experimenta problemas temporales.
2. **Consistencia**: La automatización mediante GitHub Actions asegura que los datos se actualicen regularmente sin intervención humana, manteniendo la integridad de la serie temporal.
3. **Trazabilidad**: El sistema de logging detallado proporciona un registro completo de todas las operaciones, facilitando la depuración y el análisis de posibles problemas.
4. **Valor Analítico**: El enriquecimiento de datos con métricas financieras adicionales amplía significativamente las posibilidades de análisis posteriores.
5. **Escalabilidad**: La arquitectura modular del sistema permite extenderlo fácilmente para incluir nuevos indicadores o fuentes de datos en el futuro.

El análisis preliminar de los datos recolectados revela patrones interesantes en la volatilidad del mercado durante el último año, con períodos de alta volatilidad coincidentes con eventos económicos significativos. Las métricas calculadas, como el RSI, proporcionan perspectivas adicionales sobre posibles puntos de inflexión en el mercado.

### Conclusiones

El presente proyecto ha demostrado la viabilidad y las ventajas de implementar un sistema automatizado para la recolección y enriquecimiento de datos históricos del Índice VIX. La arquitectura orientada a objetos, combinada con técnicas de programación defensiva y estrategias de contingencia, ha resultado en un sistema robusto capaz de operar de manera autónoma.

El enfoque de automatización mediante GitHub Actions proporciona una solución elegante y sin costes adicionales para la ejecución periódica del sistema, mientras que la persistencia en CSV y el control de versiones Git garantizan la accesibilidad y trazabilidad de los datos recolectados.

Las métricas de enriquecimiento implementadas transforman los datos crudos en información accionable, facilitando análisis técnicos más sofisticados y proporcionando una base sólida para desarrollos futuros, como implementaciones de aprendizaje automático para predicción de volatilidad o sistemas de alerta basados en umbrales predefinidos.

### Referencias

1. Chicago Board Options Exchange. (2023). VIX Index - CBOE Volatility Index. <https://www.cboe.com/tradable_products/vix/>
2. McKinney, W. (2017). Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython. O'Reilly Media.
3. Yahoo Finance. (2023). ^VIX - CBOE Volatility Index. <https://finance.yahoo.com/quote/%5EVIX/history/>
4. Pandas Development Team. (2023). pandas-datareader: Data Readers extracted from the pandas codebase. <https://pandas-datareader.readthedocs.io/>
5. Python Software Foundation. (2023). venv — Creation of virtual environments. <https://docs.python.org/3/library/venv.html>
6. GitHub. (2023). GitHub Actions Documentation. <https://docs.github.com/en/actions>