

Proyecto Integrado V - Línea de Énfasis (Entrega 2)

Por

Shara Milena Mosquera Goez

Victor Manuel Calle Londoño

Docente

Andres Felipe Callejas

CURSO

Proyecto Integrado V

INSTITUCIÓN UNIVERSITARIA DIGITAL DE ANTIOQUIA

MEDELLÍN

2025

Datos históricos Riot Platforms, Inc.

Resumen

El presente documento describe el proceso de automatización de la recolección continua de datos históricos del indicador económico de Riot Platforms, Inc. La implementación de esta solución garantiza la obtención, almacenamiento y mantenimiento de los datos recopilados de manera confiable y precisa. El sistema está diseñado para operar en un entorno de control de versiones, asegurando que cada modificación y actualización esté debidamente registrada. Además, se ha establecido una documentación inicial clara que facilita la comprensión del proceso, su configuración y su mantenimiento. Esta iniciativa tiene como objetivo principal optimizar el acceso y la gestión de los datos históricos, permitiendo un análisis eficiente y basado en información actualizada.

Introducción

En el ámbito de la gestión y análisis de datos, la capacidad de automatizar la recolección de información histórica se ha convertido en un factor crucial para las empresas que buscan tomar decisiones basadas en datos. En este contexto, Riot Platforms, Inc. ha sido seleccionada como la empresa objetivo para desarrollar un sistema de recolección continua de datos históricos.

Riot Platforms, Inc. es una compañía estadounidense dedicada principalmente a la minería de Bitcoin y a la provisión de infraestructura tecnológica para este sector. Con sede en Castle Rock, Colorado, la empresa opera instalaciones de minería a gran

escala en Texas y Kentucky. Además, posee capacidades de ingeniería y manufactura de equipos eléctricos especializados, dirigidos tanto a la industria de centros de datos como a otros sectores críticos como generación de energía y servicios públicos. Su modelo de negocio se basa en una estrategia de integración vertical, lo que le permite gestionar desde la producción y diseño de equipos hasta la operación de sus instalaciones mineras. Esta estructura le brinda una ventaja competitiva al optimizar costos y eficiencia en sus operaciones.

Este proyecto se centra en garantizar que los datos de Riot Platforms, Inc. se obtengan de manera precisa, se almacenen de forma segura y se mantenga un registro completo de su evolución mediante el uso de un sistema de control de versiones. Adicionalmente, la solución incluye una documentación inicial clara que facilita la comprensión del proceso, asegurando que cualquier usuario o administrador pueda gestionar y mantener el sistema de manera efectiva. Esta iniciativa contribuye a mejorar la eficiencia en el manejo de la información y permite un análisis más robusto de los datos históricos recopilados.

Metodología

Para cumplir con el objetivo de automatizar la recolección continua de datos históricos del indicador económico relacionado con Riot Platforms, Inc., se diseñó una metodología orientada a garantizar la persistencia y trazabilidad. A continuación, se describen cada una de estas etapas:

1. Recolección de Datos:

La información histórica de Riot Platforms, Inc, se extrae del portal Yahoo Finance utilizando técnicas de web scraping, obteniendo datos relevantes y garantizando su disponibilidad para procesos posteriores de análisis.

2. Monitoreo de eventos:

Se incorporan loggers que realizan el registro de eventos para facilitar el monitoreo y la trazabilidad del proceso de recolección, permitiendo detectar y documentar posibles errores o incidencias durante su ejecución.

3. Automatización del Proceso:

Para asegurar una recolección continua y sin intervención manual, se implementó un flujo de trabajo automatizado utilizando GitHub Actions, una herramienta de integración continua (CI/CD) que permite ejecutar tareas programadas directamente desde el repositorio.

4. Persistencia y Almacenamiento:

Los datos recolectados son almacenados localmente en archivos CSV organizados dentro del proyecto. Esta estrategia garantiza la persistencia de los datos históricos, facilitando su posterior análisis y validación. El sistema de carpetas asegura una estructura limpia y reutilizable.

5. Trazabilidad y Control de Versiones:

El proyecto está versionado en un sistema de control de versiones (Git), lo cual permite llevar un historial detallado de los cambios realizados tanto en los scripts de recolección como en los archivos de datos. Esto asegura la trazabilidad y facilita auditorías o restauraciones en caso de errores.

Desarrollo y validación del modelo predictivo

Justificación del modelo

Para la predicción del precio de cierre ajustado de Riot Platforms, Inc., se eligió el modelo **Random Forest Regressor** por su capacidad para ofrecer resultados precisos en problemas de regresión sin requerir supuestos estrictos sobre la distribución de los datos o la relación entre variables.

Este modelo destaca por:

- Su **robustez frente al ruido y valores atípicos**, ya que al combinar múltiples árboles de decisión reduce el impacto de datos extremos.
- Su **flexibilidad para capturar relaciones complejas** entre las variables, lo que lo hace adecuado para entornos financieros donde los comportamientos no siempre son lineales o predecibles.
- Su **eficiencia práctica y facilidad de uso**, ya que puede implementarse rápidamente mediante bibliotecas consolidadas como [scikit-learn](#), permitiendo una integración fluida en entornos desacoplados y automatizados.

La elección de este modelo se alinea con el enfoque del proyecto: obtener predicciones confiables con un modelo interpretable, escalable y de fácil mantenimiento dentro de un flujo automatizado de trabajo.

Justificación de las métricas

Se utilizaron dos métricas principales para evaluar el desempeño del modelo:

- **MAE (Mean Absolute Error):** proporciona una medida directa y fácilmente interpretable del error promedio. Es robusta frente a valores atípicos. En este proyecto, se obtuvo un MAE de **0.0247**, lo que indica que el modelo, en promedio, se equivoca en **2.47 céntimos de dólar**, una desviación considerada aceptable para el tipo de indicador evaluado.
- **RMSE (Root Mean Squared Error):** fue utilizada como complemento al MAE, ya que penaliza errores grandes con mayor intensidad. Se obtuvo un RMSE de **0.0373**, es decir, una “distancia típica” al valor real de **3.73 céntimos**, lo cual indica que no hay errores extremos que comprometan la estabilidad del modelo.

El uso combinado de ambas métricas proporciona una visión más completa del rendimiento del modelo, equilibrando precisión general con sensibilidad a errores significativos.

Resultados obtenidos

Los resultados indican que el modelo desarrollado presenta un comportamiento estable y confiable para la predicción del precio de cierre ajustado. Las métricas obtenidas confirman que el margen de error se mantiene dentro de rangos adecuados, lo que valida tanto la estrategia de modelado como el enriquecimiento de datos aplicado previamente.

Visualización en Streamlit

Se desarrolló un dashboard en Streamlit que muestra gráficamente los resultados del modelado, el enriquecimiento y la representación de los indicadores claves como:

- Media móvil
- Tasa de variación diaria
- Retorno acumulado
- Volatilidad
- Desviación estándar

El análisis del dashboard se puede ver en el siguiente enlace

<https://proyectointegrador5.streamlit.app/>

Conclusión

Este proyecto permitió implementar con éxito un sistema automatizado para la recolección de datos históricos de Riot Platforms, Inc., asegurando su almacenamiento estructurado, trazabilidad mediante control de versiones, y documentación clara del proceso. Asimismo, se desarrolló un modelo predictivo eficaz que fue validado con métricas apropiadas, reflejando un desempeño confiable y coherente con los objetivos planteados.

La integración de herramientas como GitHub Actions para la automatización y Random Forest para la predicción demuestra un enfoque sólido en la aplicación de técnicas de ingeniería y ciencia de datos. En conjunto, este trabajo evidencia competencias en automatización, modelado, validación y visualización de datos, aportando una base técnica robusta para futuras ampliaciones o aplicaciones en contextos similares.

Bibliografía

Yahoo Finance. (s.f.). *Yahoo Finanzas*. <https://finance.yahoo.com/>

Python Software Foundation. (s.f.). *Python 3 documentation*. <https://docs.python.org/3>

GitHub. (s.f.). *GitHub Documentation*. <https://docs.github.com/>

GitHub. (s.f.). *GitHub Actions Documentation*. <https://docs.github.com/actions>

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, É. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12, 2825-2830. <https://scikit-learn.org/>

Seabold, S., & Perktold, J. (2010). *Statsmodels: Econometric and statistical modeling with Python*. <https://www.statsmodels.org/stable>

Streamlit Inc. (s.f.). *Streamlit documentation*. <https://docs.streamlit.io/>

OpenAI. (2025). *ChatGPT* (versión GPT-4). <https://chat.openai.com/>

