# Integrantes

* Arias Victoria
* Gretter Alejandro
* Molina Juan Ignacio

# Descripción de los datos y variables de entrada

El objetivo del trabajo es desarrollar un sistema de predicción del precio de cierre diario de Bitcoin (BTC-USD) para los próximos 7 días, utilizando datos históricos financieros. Se busca aplicar un flujo completo de Machine Learning que incluya: descarga de datos, preprocesamiento, creación de features, entrenamiento de modelos y generación automática de predicciones diarias.

El foco del trabajo está en la metodología, no en la precisión absoluta de los precios.

# Datos y variables

Se utilizaron las siguientes fuentes de datos:

|  |  |
| --- | --- |
| **Fuente** | **Variable principal** |
| Yahoo Finance | Precio de Bitcoin (BTC-USD) |
| Yahoo Finance | Oro (GOLD) |
| Yahoo Finance | S&P 500 Index |
| FRED | Tasa de fondos federales (FED Funds) |

Variables de entrada (feactures) principales:

* Promedios móviles y desviaciones estándar (7 días) del precio BTC
* Retornos diarios (return\_1d)
* Variables derivadas de los precios de oro, S&P500 y FED Funds Rate

Se eliminó la variable de Google Trends (multiTimeline) porque **ya no se utiliza** en la última versión del proyecto

# Preprocesamiento

# Normalización y limpieza: se completaron valores faltantes con ffill.

# Ventanas de predicción: se construyeron matrices X y y para predicción de 7 días, desplazando la ventana en 1 día cada vez.

* 1. Matriz final:
     1. X con todas las features (excluyendo columnas de fechas y precios futuros)
     2. y con los 7 valores futuros del precio BTC por registro

*Archivos generados:*

* X.npy, y.npy, dates.npy (matrices de features, targets y fechas)
* dataset\_completo\_mejorado.csv (dataset procesado)

# Modelos evaluados y búsqueda de hiperparámetros

# Se implementaron y compararon tres modelos multisalida: Estrategia de predicción de 7 días

* **Ridge Regression** – baseline lineal regularizado.
* **Random Forest Regressor** – robusto frente a ruido y no linealidades.
* **LightGBM Regressor** – modelo de boosting, mejor desempeño y estabilidad.

*Entrenamiento:*

* Dividido en train/test (80% / 20%).
* Validación temporal para mantener orden cronológico.
* Se calcularon métricas MAE y RMSE por día de horizonte.

*Resultado***:** LightGBM multisalida mostró el menor error global, mientras que Random Forest mejora significativamente sobre Ridge.

# Generación de predicciones diarias

* Script predict\_7days.py toma la **última ventana de entrada del dataset** y genera predicciones para los próximos 7 días **desde el día siguiente a la ejecución**
* Se guardan los resultados en data/processed/predicciones\_diarias/predicciones\_7dias\_YYYY-MM-DD.csv y también se imprimen en terminal.
* Cada predicción incluye la **fecha de predicción** y el **precio estimado de cierre**.

*Ejemplo de salida CSV:*

Date Predicted\_Close

2025-10-17 103619.52

2025-10-18 107324.86

2025-10-19 105909.22

…… …….

# Evaluación de errores

* Se reportan métricas **MAE** y **RMSE** para cada día del horizonte de predicción.
* Error esperado: **3–5% del precio diario** para horizontes de 1 a 3 días, aumentando hacia el día 7.
* Evaluación histórica realizada sobre el conjunto de test.

# Conclusiones

* Se implementó un flujo completo y reproducible de predicción de precios BTC.
* El sistema incorpora **múltiples fuentes de información** y genera predicciones **automáticas diarias**.
* Se compararon tres modelos multisalida, seleccionando LightGBM como el más eficiente.
* La metodología permite la actualización diaria del modelo y predicciones a 7 días con un sistema estructurado