

## ***Integrantes***

- Arias Victoria
- Gretter Alejandro
- Molina Juan Ignacio

### ***1. Descripción de los datos y variables de entrada***

Los datos fueron obtenidos desde YahooFinance utilizando la librería yfinance, bajo el ticker BTC-USD. Se trabajó con datos diarios del precio de Bitcoin, tomando como variable de salida el precio de cierre ajustado (Adj Close).

#### **Variables de entrada seleccionadas:**

Precio de apertura (Open), máximo (High), mínimo (Low) y volumen (Volume). Retornos diarios. Ventanas móviles: media y desviación estándar de 7, 14 y 30 días. Desplazamientos temporales (lags) de 1 a 7 días. Indicadores externos agregados: precios del oro y del índice S&P500; para explorar correlaciones macroeconómicas.

### ***2. Preprocesamiento***

- a. Eliminación de valores nulos y duplicados.
- b. Normalización de variables mediante escalado MinMaxScaler.
- c. División del dataset en entrenamiento (70%), validación (15%) y prueba (15%), preservando el orden temporal.
- d. Construcción de un dataset supervisado multisalida, donde cada instancia predice los precios de cierre de los próximos 7 días.
- e. Guardado de los conjuntos procesados en la carpeta data/processed/ para su uso posterior.

### ***3. Modelos evaluados y búsqueda de hiperparámetros***

Se evaluaron tres modelos principales, realizando búsquedas de hiper parámetros con GridSearchCV y Randomized Search CV:

★ **Regresión Ridge (Lineal Regularizada):** parámetro alpha.

★ **Random Forest Regressor:** parámetros n\_estimators, max\_depth, min\_samples\_split.

★ **LightGBM Regressor (MultiOutput):** parámetros `num_leaves`, `learning_rate`, `n_estimators`.

El mejor desempeño se obtuvo con Random Forest multisalida, que presentó una buena relación entre error, estabilidad y facilidad de entrenamiento.

#### 4. Estrategia de predicción de 7 días

El sistema final utiliza un modelo multisalida (una sola instancia del modelo predice simultáneamente los valores de los próximos 7 días:  $T+1 \dots T+7$ ).

Esto permite evitar la acumulación de error típica de los enfoques autorregresivos.

El script `predict_7days.py` carga el modelo entrenado y el último conjunto de datos procesado, generando automáticamente las predicciones para los 7 días siguientes, que se guardan en `data/processed/prediccion_7dias.csv` con el siguiente formato:

Date,Predicted\_Close

2025-10-14,118841.34

2025-10-15,118373.67

2025-10-16,117597.73

2025-10-17,116018.85

2025-10-18,115992.55

2025-10-19,114615.17

2025-10-20,114774.30

#### 5. Evaluación y error esperable

El rendimiento se midió con MAE (Mean Absolute Error) y RMSE(Root Mean Squared Error):

| Modelo           | MAE   | RMSE       | Observaciones       |
|------------------|-------|------------|---------------------|
| Ridge Regression | Alto  | Alto       | Usado como baseline |
| Random Forest    | Medio | Medio-bajo | Buen equilibrio     |

|          |      |          |                                       |
|----------|------|----------|---------------------------------------|
| LightGBM | Bajo | Más bajo | Ligera mejora, pero mayor complejidad |
|----------|------|----------|---------------------------------------|

El error esperable se encuentra en torno al 3–5% del precio diario en horizontes de 1–3 días, aumentando gradualmente hacia el día 7.

## 6. Conclusiones

Se desarrolló un sistema de machine learning que predice el precio de cierre de Bitcoin para los próximos 7 días, cumpliendo con los requerimientos planteados.

El modelo Random Forest multisalida fue seleccionado para producción por su estabilidad y buena generalización.

El proyecto permitió aplicar de forma integral los contenidos de la materia: análisis de series temporales, ingeniería de características, evaluación de modelos y automatización del flujo predictivo.

La precisión no fue el objetivo central, sino la correcta aplicación de la metodología y la reproducibilidad del proceso de aprendizaje automático.