

# **Proyecto Integrado V – Línea de Énfasis**

## **Entrega - 1**

German Darío Arbeláez Ríos

Juan Carlos Presiga Montoya

GRUPO 13

PREICA2501B020128

**Andrés Felipe Callejas**



MEDELLÍN

Mayo 2025

## Justificación

Para la actividad de modelado y predicción se escogió la métrica de Random Forest. La aplicación se implementa con un entrenamiento de la métrica, el registro de los datos a través del logger y la utilización del modelo lo cual nos proporciona unas medidas de importancia de características, es un modelo robusto y realiza un manejo de relaciones no lineales.

El RandomForest es un ensamble de múltiples árboles de decisión, lo que hace que sea menos propenso al sobreajuste que escogiendo un solo árbol.

Así mismo puede tomar relaciones no lineales y complejas tomando las variables predictoras y la de objetivo.

Igualmente, permite identificar fácilmente las variables que son mas relevantes para la predicción.

Las características que se utilizaron son las siguientes;

- Precio de cierre (close)
- Precio de apertura (open)
- Precio máximo (high)
- Precio mínimo (low)

- Volumen de negociación (volume)
- Medias móviles de precios (por ejemplo, price\_ma5, price\_ma20)
- Medias móviles de volumen (por ejemplo, volume\_ma5, volume\_ma20)
- Rango de precios diario: (high - low) / close
- Retorno diario: close.pct\_change()

Al utilizar las características relacionadas anteriormente randomforest es ideal por que permite mejorar la capacidad predictiva y la robustez del modelo, también reduce el riesgo que el modelo sea muy simple o se pierdan patrones importantes.

## COMPARACIÓN DE PREDICCIÓN CON RELACION A LA BD

Predicción para fecha determinada

```
2025-05-25 15:08:03,062 - DataCollector - INFO - Modelo entrenado exitosamente
2025-05-25 15:08:03,063 - DataCollector - INFO - Realizando predicciones...
2025-05-25 15:08:03,071 - DataCollector - INFO - Modelo cargado exitosamente
2025-05-25 15:08:03,106 - DataCollector - INFO - Última predicción de volatilidad: 34.2850 para la fecha 2025-05-22 00:00:00
2025-05-25 15:08:03,116 - DataCollector - INFO - Predicciones guardadas en 'predictions.csv'
```

Datos reales de la BD



The screenshot shows a database interface with a table of historical data. The table has columns for 'volatility', 'SMA\_20', and 'EMA'. The 'volatility' column is highlighted with a red box, showing a value of 33.2136 for row 250. The 'SMA\_20' column shows a value of 965.355 for row 250. The 'EMA' column shows a value of 968.73 for row 250. The interface also includes a search bar, a filter button, and a 'Actualizar a PRO' button.

	volatility	SMA_20	EMA
247	23.2583	956.105	
248	30.3798	961.55	
250	33.2136	965.355	
250	32.9675	968.73	

Se observa que la predicción se acerca bastante al dato real para esa fecha