



4-4-2025

Caso Banco Central

Aprendizaje automático



Eduardo Tejada Moreta
20240276

Introducción 2

Recursos locales empleados	3
Recursos en la nube que usarán. Módulos de Azure ML que emplearán y por qué.	4
Infraestructura.....	6
Diagrama de flujo del pipeline	7
Diagrama de flujo del modelo de predicción	8
Procedimiento realizado para resolver el problema.	9
Método de recolección de datos y formato de salida propuesto de la información.	11
Costos de la propuesta.....	12
Métricas de efectividad del sistema propuesto.....	13
Dificultades al realizar el proyecto	14
Guiade uso de la demo realizada	14
Posibles mejoras de la propuesta.....	17
Riesgos y limitaciones	17

Introducción

El Banco Central tiene un banco de datos con miles de tablas de información financiera del país recopilada por cientos de estadistas a nivel nacional y ofrecidas por diversas instituciones financieras y del estado como el Ministerio de Hacienda. Al banco le interesa desarrollar un sistema modular de regresión que permita hacer pronósticos y estimaciones de estados futuros predefinidos en las distintas tablas con que ellos cuentan para ofrecer este proceso como un servicio de predicción que le produzca beneficios extra

Recursos locales empleados

En este proyecto el único recurso local utilizado es para el despliegue de la interfaz gráfica, para el uso de manera sencillas.

Recursos de hardware utilizados (estos no son los necesarios para la ejecución de la aplicación):

- CPU: 3.20ghz de base, 8 núcleos, 16 logios
- RAM: 16gb

Recursos de software:

- SO: Windows 11
- Lenguaje: Python 12
- Dependencias:
 - Streamlit (última versión estable),
 - Pandas (UVE),
 - Plotly (UVE)

Recursos en la nube que usarán. Módulos de Azure ML que emplearán y por qué.

El modulo usado para este proyecto, es perteneciente a la sección de Azure ml **Creación**, el cual se denomina **Notebook** y de la sección **Recursos** elegimos **Datos**

Este me permite ejecutar mis diferentes notebooks (o archivos .py) en una máquina virtual orientada al área del ML (**Proceso**).

Se uso esta forma debido a la naturaleza del proceso, la actualización constante del modelo nos impide entrenarlo solo una vez o cada cierto tiempo e ir utilizando el modelo entrenado lo que hace que usar un modulo especifico para el entrenamiento, hace que se consuman mas recursos y a la vez mayor costo monetario, otro punto es el diseño del sistema lo cual hablaremos a continuación.

Propiedades del recurso

Estado

 Detenido

Última operación

Detenido a las Apr 2, 2025 11:05 AM: Correcto

Tamaño de la máquina virtual

Standard_DS11_v2 (2 núcleos, 14 GB de RAM y disco de 28 GB)

Unidad de procesamiento

CPU: [Memory optimized](#)

Costo estimado

0.15 \$/h (cuando se ejecuta)

Almacenamiento de datos adicional

--

Aplicaciones

JupyterLab Jupyter VS Code (Web) VS Code (escritorio) ...

Fecha de creación

3/30/2025, 1:13:23 PM

Acceso SSH

Disabled

Acceso a la raíz

Habilitado

El sistema lleva un flujo continuo y automático por cada solicitud, el proceso manual es mínimo, cada solicitud desde la interfaz de usuario realiza el pipeline internamente facilitando el procesamiento por parte del modelo. Debido a este punto utilizamos el Notebook y lo ejecutamos y un proceso, lo cual hace que los recursos consumidos se

vean reducido y debido también a que el entrenamiento de este modelo no requiere un alto consumo de recursos.

Suscripción

Cuota dedicada

2

núcleos usados

18

núcleos disponibles

0

Número de núcleos dedicados que puede usar para su suscripción

Infraestructura

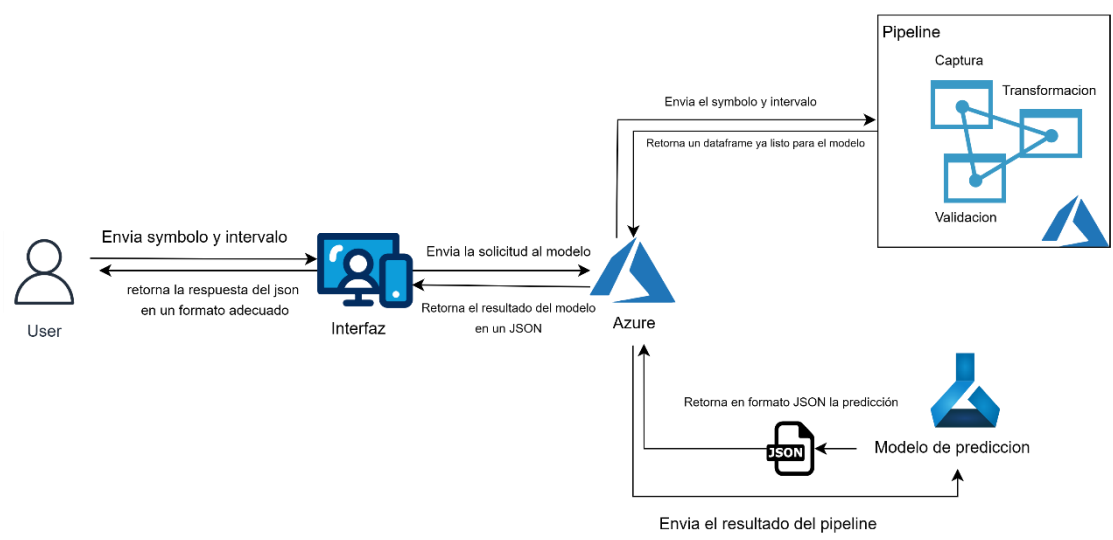


Diagrama de flujo del pipeline

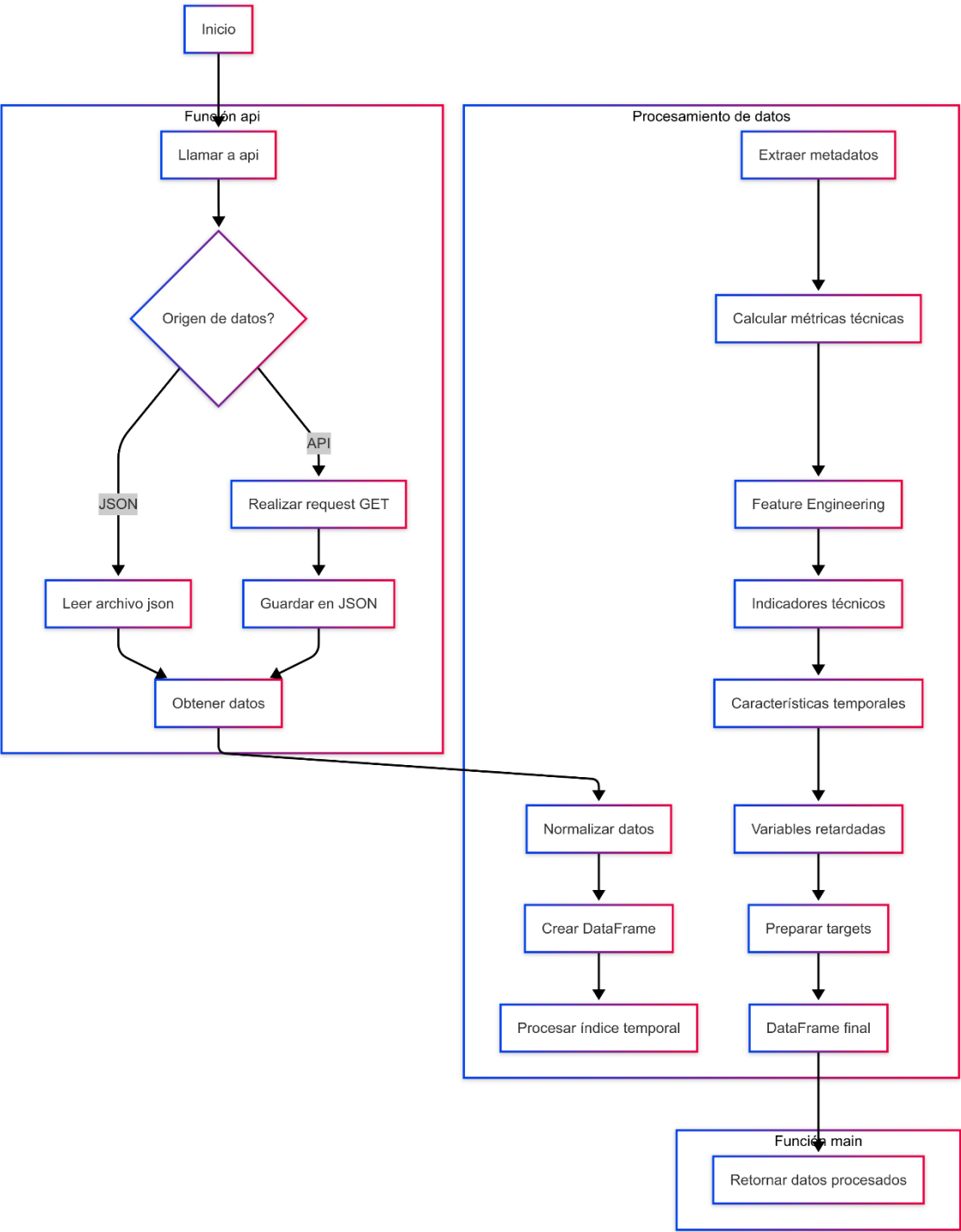
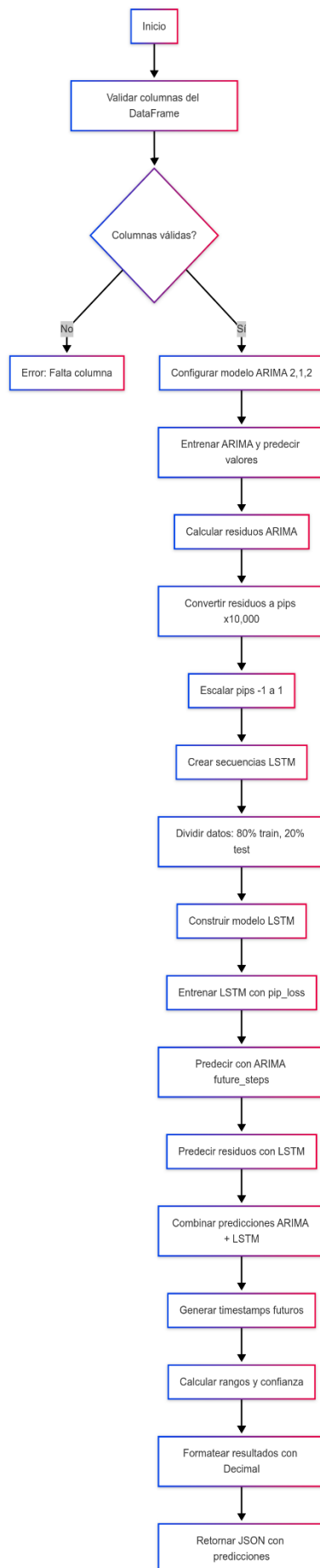


Diagrama de flujo del modelo de predicción



Procedimiento realizado para resolver el problema.

1. Identificación del objetivo:

Predecir valores de activos en bolsa, considerando su volatilidad inherente y factores externos (políticos, especulativos, alta frecuencia de transacciones).

2. Análisis del problema central:

- Las predicciones financieras son intrínsecamente inexactas por:
 - **Eventos políticos:** Decisiones gubernamentales, conflictos geopolíticos.
 - **Especulaciones:** Rumores o expectativas del mercado no siempre ligadas a fundamentos.
 - **Acciones masivas de traders:** Movimientos bruscos por operaciones de alto volumen.

3. Elección de intervalos temporales como eje estratégico:

- **Intervalos cortos (5M, 15M, 1H):**
 - Generan **más datos** (ej.: 8 horas en intervalos de 5 minutos = 96 puntos vs. 1 día = 1 punto).
 - Permiten detectar **patrones operativos** (comportamiento de traders, reacciones a noticias en tiempo real).
 - Aunque son más volátiles, contienen información valiosa para modelos de secuencias (LSTM).
- **Intervalos largos (1D, 1W):**
 - Reflejan tendencias macro, pero pierden detalles críticos para el trading activo.

4. Selección del modelo híbrido (LSTM + ARIMA):

- **Por qué LSTM en intervalos cortos:**
 - Necesita **grandes volúmenes de datos** para aprender patrones secuenciales.
 - Un día en intervalos de 5M = 288 datos → suficiente para entrenamiento.
 - Un día en intervalos de 1D = 1 dato → insuficiente para la LSTM.
- **Por qué ARIMA:**

- Complementa la LSTM capturando **tendencias generales** (ej.: dirección alcista/bajista en 1D).
- Mitiga el "ruido" de los intervalos cortos al aportar estabilidad histórica.

5. Limitaciones clave:

- Los eventos imprevistos (ej.: un tweet de un CEO influyente) pueden invalidar las predicciones.
- La volatilidad en intervalos micro (5M) exige ajustes constantes del modelo.

ARIMA: [▷ Modelo ARIMA](#)

LSTM: [¿Qué son las Redes LSTM? | Codificando Bits](#)

Método de recolección de datos y formato de salida propuesto de la información.

La recolección de datos financieros, especialmente de activos que cotizan en bolsa puede ser algo complicado debido a la naturalidad de este y el medio por el cual se quiera consumir, como el caso de uso fue de un sistema lo ideal es que consumiera una API

La API utilizada es de la empresa; ALPHA VANTAGE, la cual brinda una API free con ciertas limitaciones, y una versión premium en el cual ambas de pueden retornar un json o Excel, en este caso se optó por un JSON.

Docs: [Free Stock APIs in JSON & Excel | Alpha Vantage](#)

Luego que capturamos esta información debemos transformarla para que los modelos puedan procesarlo correctamente.

Hay diferentes variables de importancia para el modelo las cuales se mencionarán a continuación:

Variables básicas:

Open: Precio de apertura del período. High: Precio máximo. Low: Precio mínimo.
Close: Precio de cierre (usualmente se usa como variable objetivo o target). Volume: Volumen de transacciones.

Variables Derivadas y Técnicas

Retorno porcentual entre el cierre actual y el anterior. Diferencia entre el precio de cierre y el precio de apertura (o el promedio de [High, Low]).

Medias Móviles: Media móvil simple (SMA) y/o exponencial (EMA) de diferentes ventanas (por ejemplo, 5, 10, 20 períodos).

Indicadores Técnicos:

RSI (Índice de Fuerza Relativa). MACD (Moving Average Convergence Divergence).
Bandas de Bollinger. Volatilidad (por ejemplo, desviación estándar de los precios en una ventana determinada).

Variables de Tiempo y Estacionalidad:

Fecha/Hora: Extraer características como la hora del día, el día de la semana o incluso indicadores de estacionalidad (aunque en intradía, el patrón puede repetirse a lo largo del día).

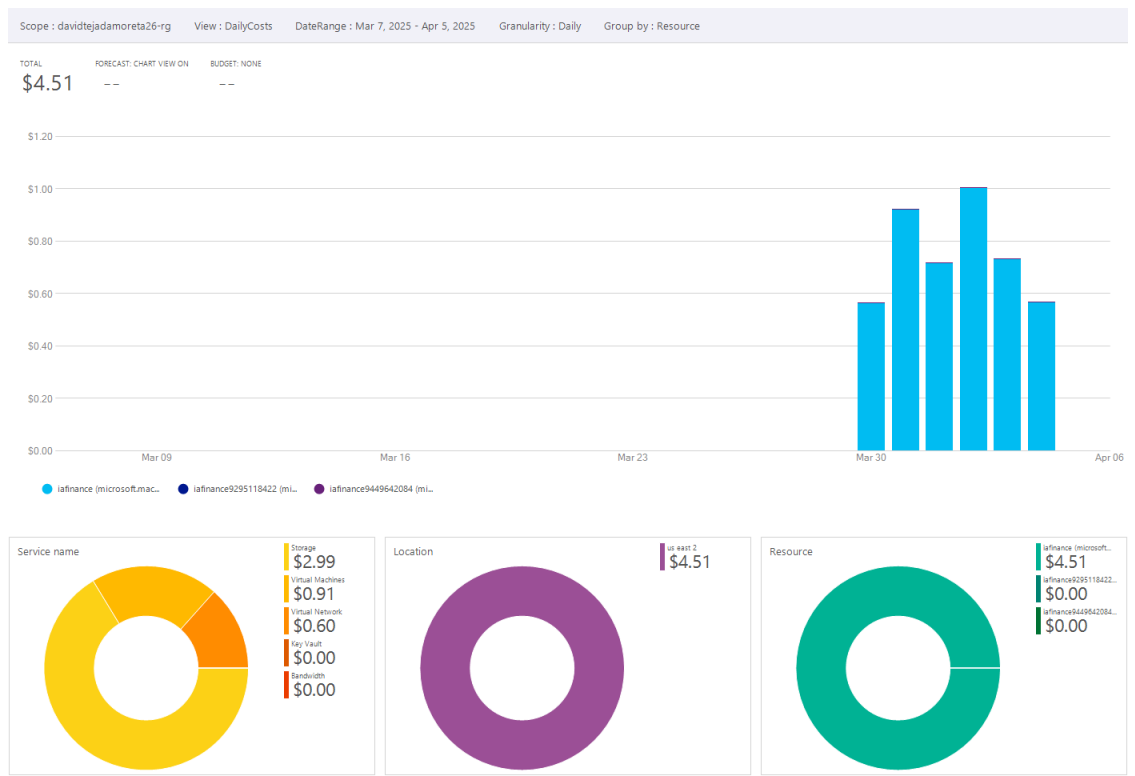
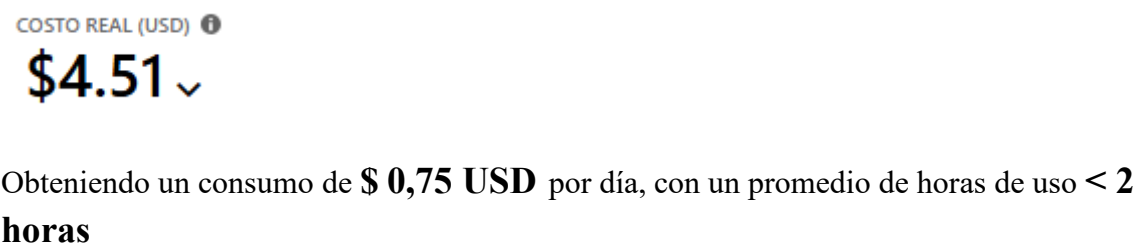
Lag Features: Incorporar valores retrasados de las variables principales para que el modelo pueda aprender de la secuencia histórica.

El formato de salida es una lista de JSON con cada predicción, en el intervalo dado, el máximo y mínimo del valor del precio.

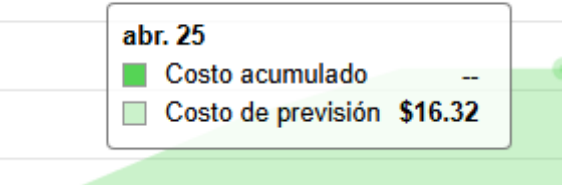
Costos de la propuesta.

El costo de la propuesta es dependiente al uso debido a que el modelo es entrenado en la nube con Azure.

El costo de este recurso es dependiendo a su uso, se usaron storage, virtual machines y virtual network y en un lapso de **6 días** tuvo un costo de:



En el lapso de un mes se estima un total de:



En el apartado de la API, hace un total de **45 USD** mensual, que diariamente nos da un total **1,41 USD**

Concepto	Descripción	Costo Diario (USD)	Costo Mensual Estimado (USD)
Entrenamiento en Azure	Uso de Storage, Virtual Machines y Virtual Network durante < 2 horas/día	\$0.75	\$22.50 (basado en 30 días)
Consumo de la API	Servicio de API con uso constante	\$1.41	\$45.00
Total, estimado mensual	Suma de todos los costos estimados	\$2.16	\$67.50

Métricas de efectividad del sistema propuesto

En una prueba de 25 consultas se obtuvo lo siguiente:

solicitudes	Símbolo	Tipo de Activo	Intervalo de Predicción	Nivel de Predicción (Condición Normal)	Nivel de Predicción (Caída Política)	Min/Max Respeto
25	IBM	Acción	10 minutos	~80%	N/A*	100%
25	AUDCAD	Divisa	10 minutos	~70%	~50%	100%
25	USDMXN	Divisa	10 minutos	~70%	~50%	100%

Otra medida para la efectividad del sistema es mediante el valor “confidence” del json, que establece el nivel de confianza.

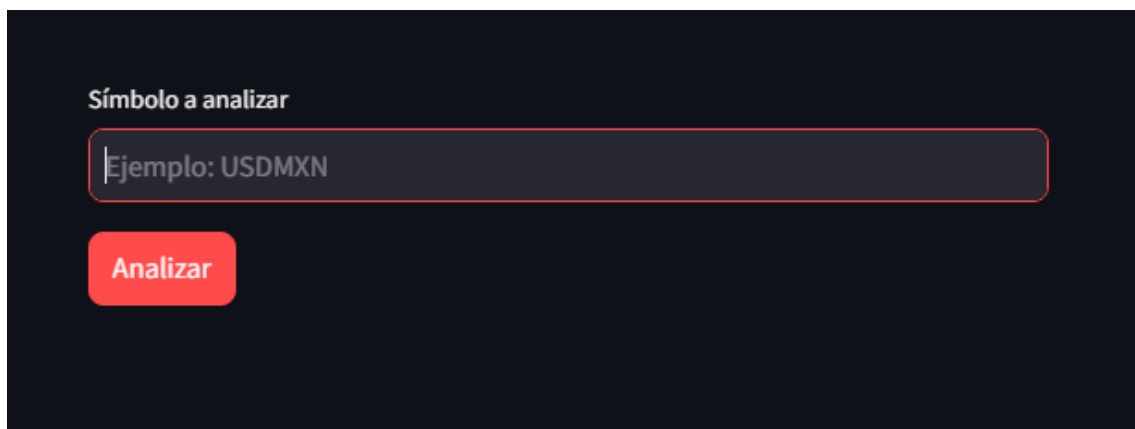
Dificultades al realizar el proyecto

La principal dificultad de este proyecto es el hacer algo muy volátil como la bolsa de valores, en algo predecible, el buscar un punto en el cual sea más fácil la predicción por parte de un algoritmo.

Otra dificultad fue el extraer en vivo la fuente de datos rica en variables para que nuestro modelo se entrene con éxito debido a que estos recursos financieros mayormente son de pagas.

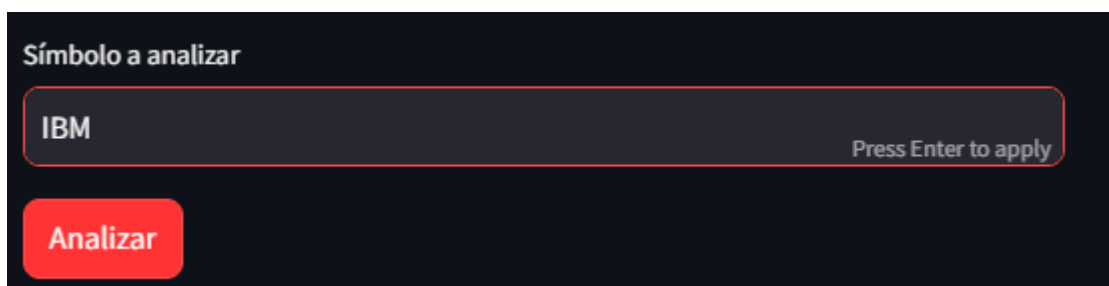
Guía de uso de la demo realizada

1-Agregamos el símbolo a analizar debe agregarse en mayúscula y pegado, aunque sea un par de divisas.



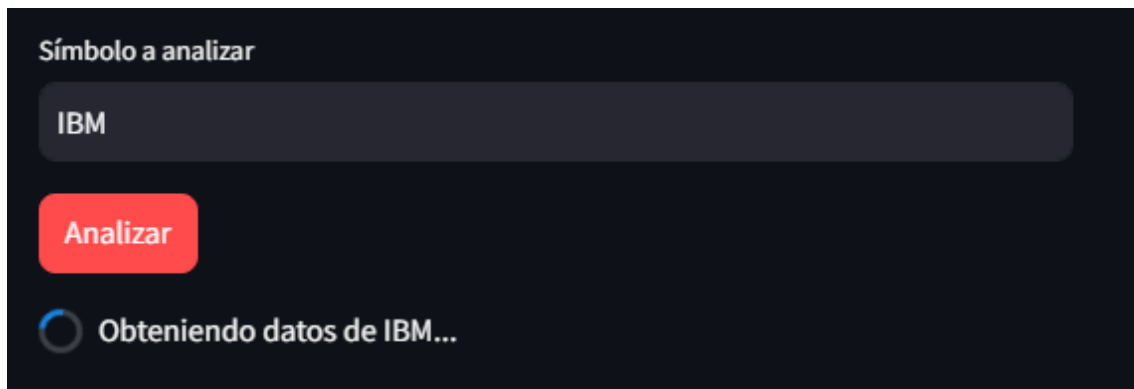
The screenshot shows a dark-themed interface. At the top, the text "Símbolo a analizar" is displayed. Below it is a text input field containing the placeholder text "Ejemplo: USDMXN". Below the input field is a red button with the text "Analizar" in white.

2-Procederemos a darle a Analizar.



The screenshot shows the same dark-themed interface. The text input field now contains the text "IBM". To the right of the input field, the text "Press Enter to apply" is visible. Below the input field is the same red "Analizar" button.

3-Esperamos el tiempo de carga.



The screenshot shows a dark-themed web interface. At the top, the text "Símbolo a analizar" is displayed. Below it is a dark input field containing the text "IBM". Underneath the input field is a red button with the white text "Analizar". At the bottom of the interface, there is a loading indicator consisting of a blue circular progress bar and the text "Obteniendo datos de IBM...".

4-Visualizar el resultado, el timestamp es el tiempo del activo, predicted, el valor de este en el timestamp dado, predicted_min es valor mínimo y el predicted_max es el valor máximo, la tabla puede descargarse en formato CSV.

Resultado Detallado

	timestamp	predicted	predicted_min	predicted_max
0	2025-03-28 20:00:00	242.7447	242.7437	242.7457
1	2025-03-28 20:05:00	242.7493	242.7483	242.7503
2	2025-03-28 20:10:00	242.7489	242.7479	242.7499
3	2025-03-28 20:15:00	242.7475	242.7465	242.7485
4	2025-03-28 20:20:00	242.7462	242.7452	242.7472
5	2025-03-28 20:25:00	242.7465	242.7455	242.7475
6	2025-03-28 20:30:00	242.7465	242.7455	242.7475
7	2025-03-28 20:35:00	242.7465	242.7455	242.7475

Análisis de Precisión Decimal



Posibles mejoras de la propuesta

Dependiendo del alcance que se quiere llegar de este se pueden hacer múltiples mejoras, pero una de estas mejoras es exponer el modelo como un servicio API, para poder interoperar con diferentes aplicativos.

1-Definir claramente que variables estadísticas nos podrían ayudar con la predicción.

2-Generar nuestro propio dataset a partir de las consultas a la API, para utilizarlo en otros modelos.

3-Inclusion de API de openia o Deepseek, para hacer una predicción contextual y tener una respuesta mas valida que solo una respuesta cuantitativa.

Riesgos y limitaciones

Los riesgos van de parte de la volatilidad, estas predicciones tienen y van a seguir teniendo márgenes de error, debido a la naturaleza del área como son los activos financieros. La respuesta nunca va a hacer 100% certeza.

Las limitaciones que estas traen, es que el nivel de eficiencia y efectividad va a acorde a la personalización, lo cual hace que sea recomendable que el uso de esta sea de una persona con conocimientos en finanzas o ML