



**Business
School**

MÁSTER EN DATA SCIENCE AND BUSINESS ANALYTICS

PREDICCIÓN DE PRECIOS DE BITCOIN USANDO ANALISIS DE SENTIMENTOS DE NOTICAS FINANCIERAS

Violeta Chamosa Fondenvila

INTRODUCCIÓN

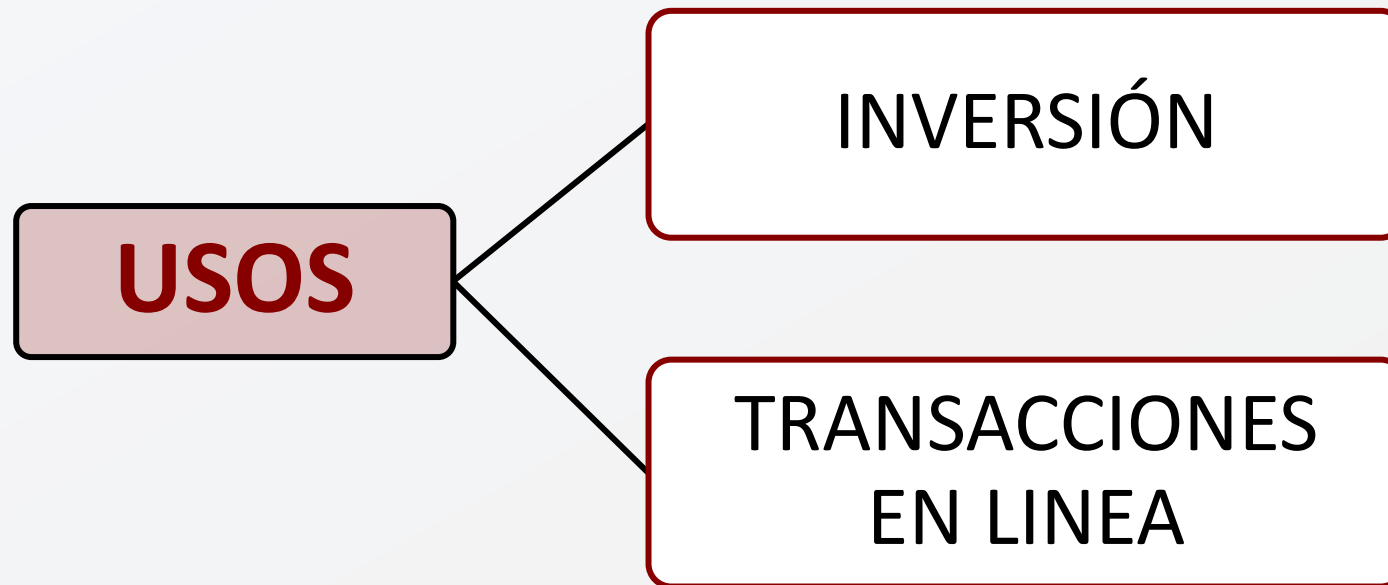


OBJETIVO: desarrollar un modelo preciso de predicción de precios de Bitcoin combinando la técnicas de **procesamiento del lenguaje natural** sobre noticias financieras y modelos de **aprendizaje automático**

¿QUÉ ES BITCOIN?



PREDICCIÓN DE PRECIOS DE BITCOIN



PREDICCIÓN DE PRECIOS DE BITCOIN

Distribución del uso de distintos modelos



APRENDIZAJE AUTOMÁTICO

APRENDIZAJE AUTOMÁTICO

Es una rama de la inteligencia artificial que permite a las máquinas aprender y tomar decisiones basadas en datos sin haber sido programadas para ello explícitamente

SUPERVISADO

NO SUPERVISADO

POR REFUERZO

APRENDIZAJE AUTOMÁTICO

Datos etiquetados

| Sexo (X1) | Peso (X2) | Altura (Y) |
|-----------|-----------|------------|
| F | 62 | 172 |
| M | 89 | 183 |
| F | 55 | 164 |
| M | 74 | 170 |
| M | 79 | 176 |

Datos de
entrenamiento

Algoritmo de
aprendizaje

MODELO

PREDICCIÓN

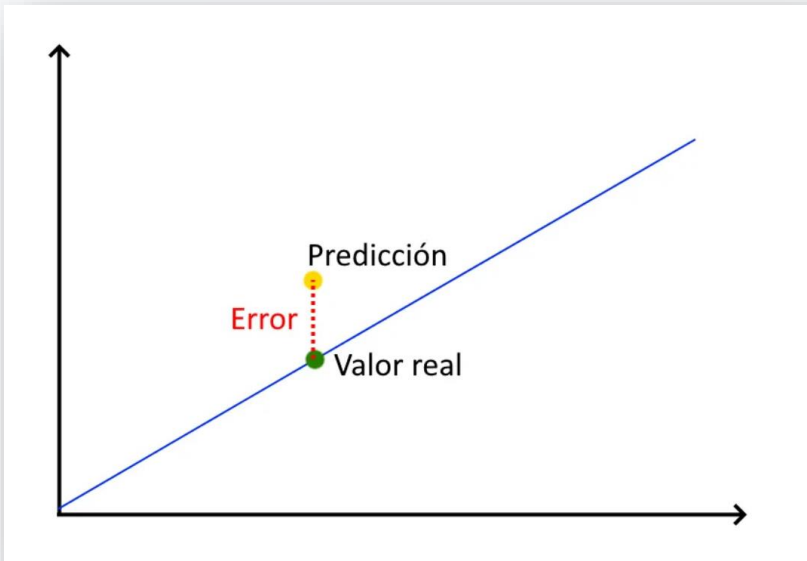
| Sexo (X1) | Peso (X2) | Altura (Y) |
|-----------|-----------|------------|
| F | 66 | 176 |
| M | 78 | 179 |

Datos de
prueba

| Sexo (X1) | Peso (X2) | Altura (Y) |
|-----------|-----------|------------|
| F | 66 | 175 |
| M | 78 | 180 |

MÉTRICAS DE EVALUACIÓN

Métricas en problemas de regresión



Error Absoluto Medio (MAE)

Error Cuadrático Medio (MSE)

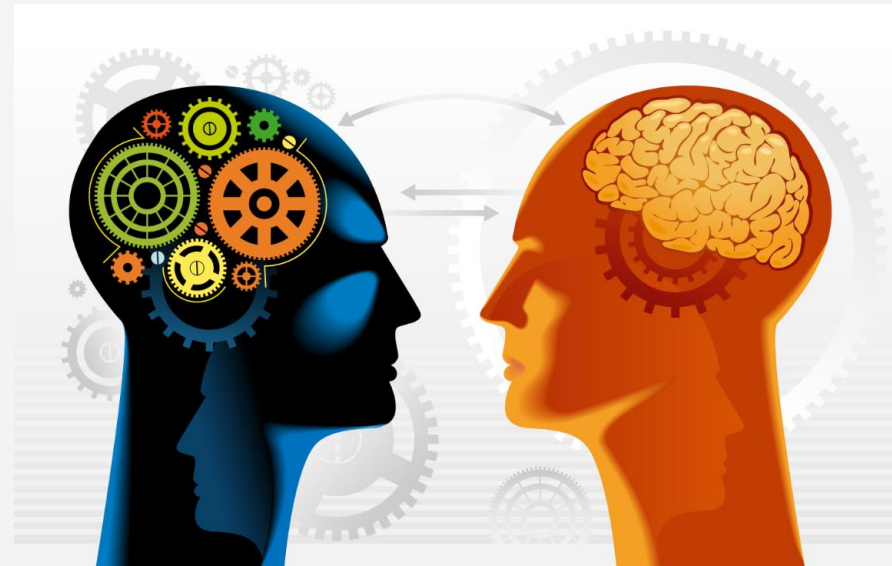
Raíz Cuadrada del Error Cuadrático Medio (RMSE)

Error Porcentual Absoluto Medio (MAPE)

PROCESAMIENTO DEL LENGUAJE NATURAL

PROCESAMIENTO DEL LENGUAJE NATURAL

Es una rama de la inteligencia artificial que trata de enseñar a los ordenadores a entender y trabajar con el lenguaje humano



DESARROLLO DEL PROYECTO

RECOLECCIÓN DE DATOS



```
graph TD; A[RECOLECCIÓN DE DATOS] --> B[FEATURE ENGINEERING]; B --> C[PREPROCESAMIENTO DE DATOS]; C --> D[ESTUDIO DE VARIABLES]; D --> E[APLICACIÓN DE LOS MODELOS];
```

The diagram illustrates a five-step project development process. The steps are represented by horizontal bars of increasing length and decreasing vertical position, creating a staircase effect. The first bar, 'RECOLECCIÓN DE DATOS', is grey. The subsequent four bars, 'FEATURE ENGINEERING', 'PREPROCESAMIENTO DE DATOS', 'ESTUDIO DE VARIABLES', and 'APLICACIÓN DE LOS MODELOS', are in shades of brown and orange. Downward-pointing arrows connect the end of each bar to the start of the next, indicating a sequential flow.

FEATURE ENGINEERING

PREPROCESAMIENTO DE DATOS

ESTUDIO DE VARIABLES

APLICACIÓN DE LOS MODELOS

RECOLECCIÓN DE DATOS

HISTORICO DE PRECIOS

- API: YFinance

| | Open | High | Low | Close | Volume |
|------------|--------------|--------------|--------------|--------------|-------------|
| Date | | | | | |
| 2014-09-17 | 465.864014 | 468.174011 | 452.421997 | 457.334015 | 21056800 |
| 2014-09-18 | 456.859985 | 456.859985 | 413.104004 | 424.440002 | 34483200 |
| 2014-09-19 | 424.102997 | 427.834991 | 384.532013 | 394.795990 | 37919700 |
| 2014-09-20 | 394.673004 | 423.295990 | 389.882996 | 408.903992 | 36863600 |
| 2014-09-21 | 408.084991 | 412.425995 | 393.181000 | 398.821014 | 26580100 |
| ... | ... | ... | ... | ... | ... |
| 2023-10-06 | 27412.123047 | 28252.537109 | 27215.552734 | 27946.597656 | 13492391599 |
| 2023-10-07 | 27946.781250 | 28028.091797 | 27870.423828 | 27968.839844 | 6553044316 |
| 2023-10-08 | 27971.677734 | 28102.169922 | 27740.662109 | 27935.089844 | 7916875290 |
| 2023-10-09 | 27934.472656 | 27989.470703 | 27302.562500 | 27583.677734 | 12007668568 |
| 2023-10-10 | 27599.025391 | 27715.847656 | 27410.812500 | 27431.419922 | 9899147264 |

3311 rows × 5 columns

ARTICULOS FINANCIEROS

- web scraping

| | articles |
|------------|---|
| Date | |
| 2018-02-25 | Original Pizza Day Purchaser Does It Again W... |
| 2018-02-26 | Bitcoin Pizza Day 2 How A Lightning Payment... |
| 2018-02-27 | Rapper 50 Cent Who Bragged About Owning Bit... |
| 2018-02-28 | This Is Who Controls Bitcoin British Man ... |
| 2018-03-01 | Bitcoin makes inroads in LA s residential re... |
| ... | ... |
| 2023-02-20 | Bitcoin regains 25K amid hope record China ... |
| 2023-02-21 | Bitcoin active addresses concern analyst d... |
| 2023-02-22 | Bitcoin Ethereum Technical Analysis BTC Fa... |
| 2023-02-23 | Bitcoin bears attempt to pin BTC price under... |
| 2023-02-24 | Bitcoin on chain data highlights key similar... |

1826 rows × 1 columns

FEATURE ENGINEERING

HISTORICO DE PRECIOS

- Indicadores técnicos

Panda_ta

```
1 data = data.assign(  
2     SMA_14 = data['Close'].rolling(14).mean(),  
3     EMA_26 = data['Close'].ewm(span=26, adjust=False).mean(),  
4     EMA_12 = data['Close'].ewm(span=12, adjust=False).mean(),  
5     MACD = macd.MACD_12_26_9,  
6     MACDh = macd.MACDh_12_26_9,  
7     MACDs = macd.MACDs_12_26_9,  
8     CCI = data.index.map(cci),  
9     ADX = data.index.map(adx.ADX_14),  
10    RSI = data.index.map(rsi),  
11    ATR = data.index.map(atr),  
12    OBV = data.index.map(obv),  
13    CMF = data.index.map(cmf),  
14    )  
15
```

ARTICULOS FINANCIEROS

- Análisis de sentimientos

TextBlob: polaridad y subjetividad

Vader Sentiment Analyzer: positividad, negatividad, neutralidad y compuesto

| | articles | Subjectivity | Polarity | sentiment | negative | positive | neutral |
|------------|---|--------------|-----------|-----------|----------|----------|---------|
| date | | | | | | | |
| 2018-02-25 | Original Pizza Day Purchaser Does It Again W... | 0.441667 | 0.220833 | 0.7788 | 0.021 | 0.087 | 0.892 |
| 2018-02-26 | Bitcoin Pizza Day 2 How A Lightning Payment... | 0.446667 | 0.010000 | -0.6597 | 0.080 | 0.037 | 0.883 |
| 2018-02-27 | Rapper 50 Cent Who Bragged About Owning Bit... | 0.518506 | 0.001623 | -0.6705 | 0.073 | 0.034 | 0.892 |
| 2018-02-28 | This Is Who Controls Bitcoin British Man ... | 0.459091 | 0.039394 | 0.4939 | 0.038 | 0.093 | 0.869 |
| 2018-03-01 | Bitcoin makes inroads in LA's residential re... | 0.335000 | -0.083333 | -0.1543 | 0.097 | 0.081 | 0.822 |
| ... | ... | ... | ... | ... | ... | ... | ... |
| 2023-02-20 | Bitcoin regains 25K amid hope record China ... | 0.575510 | 0.102041 | 0.1952 | 0.113 | 0.136 | 0.751 |
| 2023-02-21 | Bitcoin active addresses concern analyst d... | 0.350000 | 0.083333 | 0.8225 | 0.035 | 0.124 | 0.841 |
| 2023-02-22 | Bitcoin Ethereum Technical Analysis BTC Fa... | 0.283333 | 0.166667 | -0.1027 | 0.052 | 0.040 | 0.908 |
| 2023-02-23 | Bitcoin bears attempt to pin BTC price under... | 0.334407 | 0.057197 | -0.2382 | 0.072 | 0.073 | 0.855 |
| 2023-02-24 | Bitcoin on chain data highlights key similar... | 0.733333 | 0.250000 | 0.7960 | 0.000 | 0.083 | 0.917 |

1826 rows x 7 columns

PREPROCESAMIENTO DE DATOS

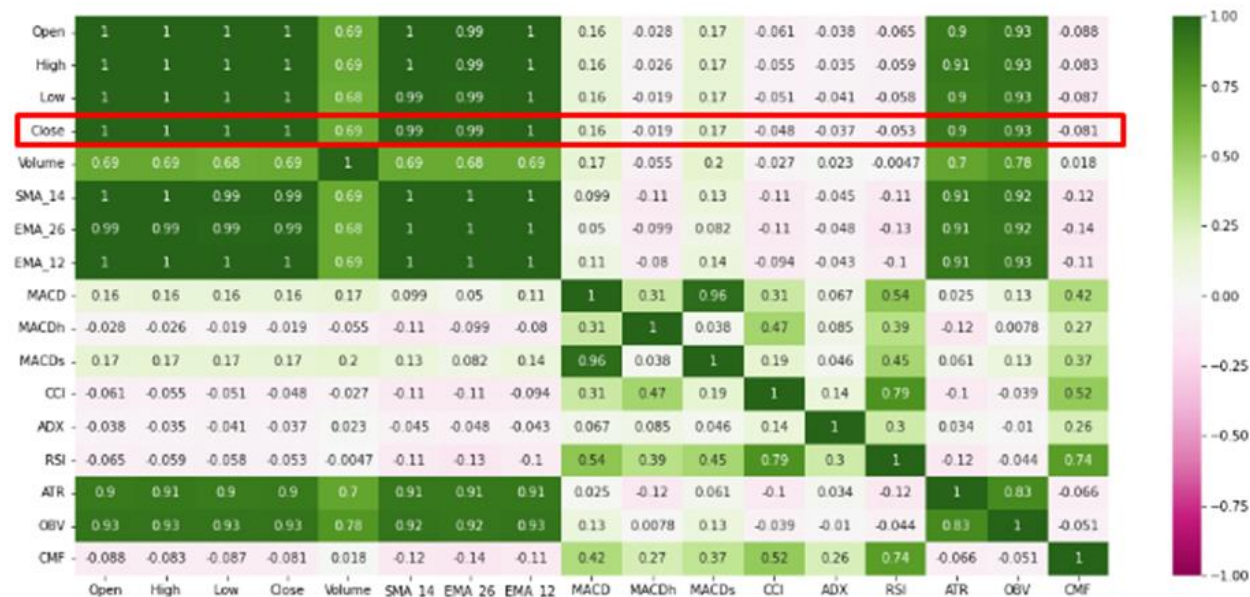


ESTUDIO DE VARIABLES

ANALISIS DE LA CORRELACIÓN

```
1 fig = plt.subplots(figsize=(18,8))
2 sns.heatmap(df.corr(), cmap="PiYG", vmin=-1.0, vmax=1.0, annot=True)
3
```

<AxesSubplot:>

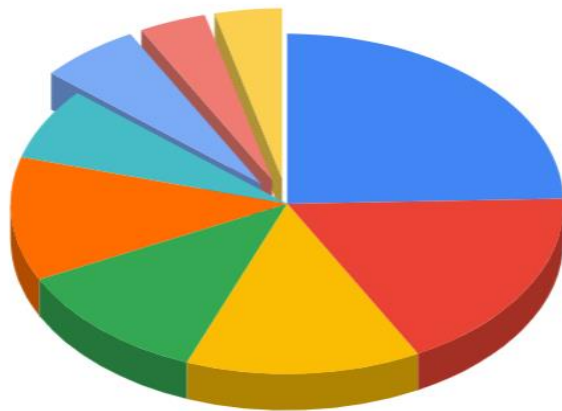


APLICACIÓN DE LOS MODELOS

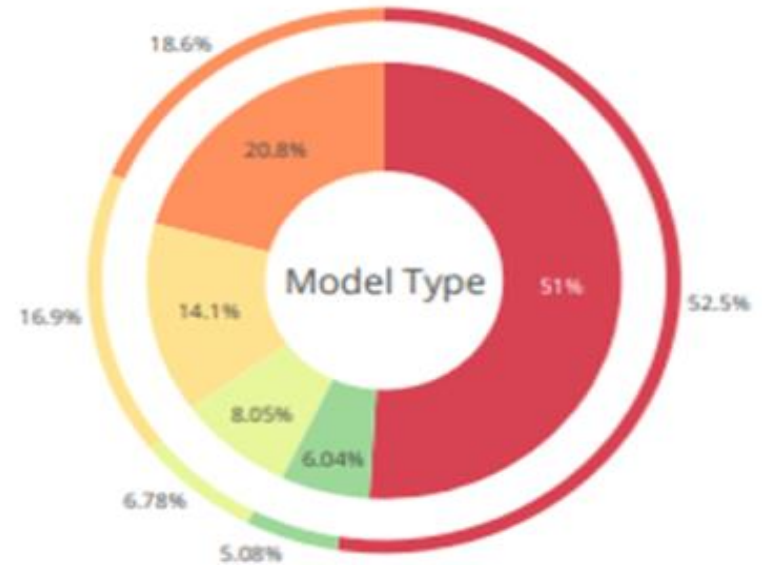
ELECCIÓN DE LOS MODELOS

- ARIMA
- LSTM
- XGBoost
- GRU

Distribución del uso de distintos modelos



Publication Count in Model Type



APLICACIÓN DE LOS MODELOS

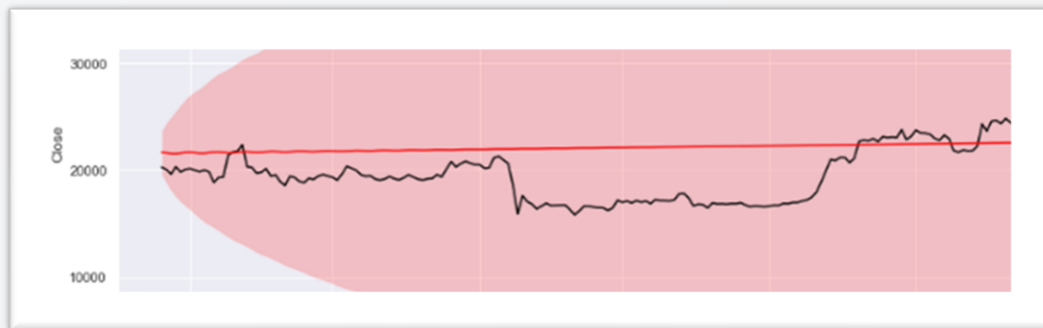
IMPLEMENTACIÓN GENERAL DE LOS MODELOS



APLICACIÓN DE LOS MODELOS

ARIMA

Modelo de Medias Móviles Autorregresivas Integradas es un modelo estadístico usado en econometría para el análisis y predicción de series temporales



Importar librerías

Dividir datos (entrenamiento y prueba)

Seleccionar hiperparámetros del modelo

Entrenar el modelo

Aplicar el modelo en el conjunto de prueba

Mostrar los resultados gráficamente

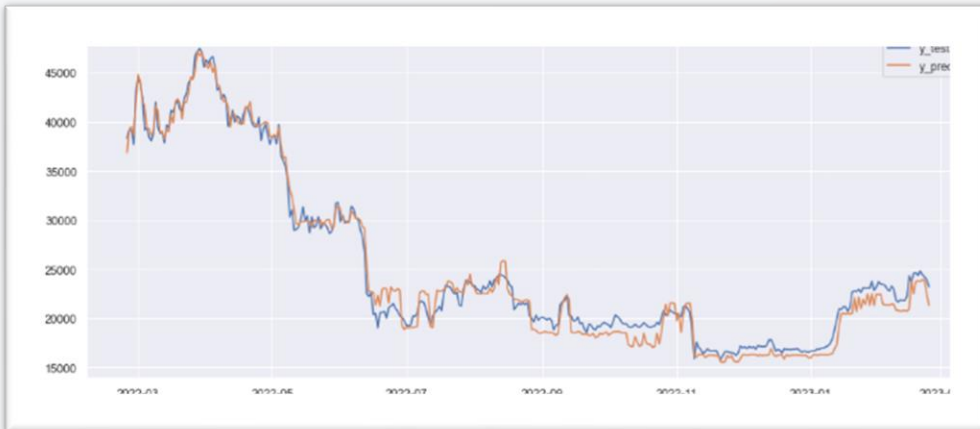
Cálculo de métricas de evaluación

APLICACIÓN DE LOS MODELOS

XGBOOST

Algoritmo de aprendizaje automático
ensambla modelos simples para obtener la
predicción final.

Gran velocidad y rendimiento



Importar librerías

Dividir datos (entrenamiento y prueba)

Seleccionar hiperparámetros del modelo

Entrenar el modelo

Aplicar el modelo en el conjunto de prueba

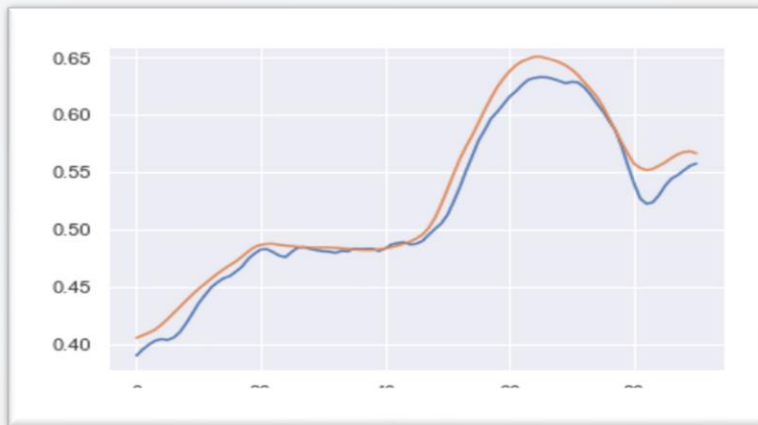
Mostrar los resultados gráficamente

Cálculo de métricas de evaluación

APLICACIÓN DE LOS MODELOS

LSTM

Red neuronal recurrente que usa una célula de memoria para recordar información a largo y corto plazo, captura dependencias a largo plazo



Importar librerías

↓
Escalar los datos

↓
Dividir datos (entrenamiento, validación y prueba)

↓
Definir y compilar el modelo

↓
Entrenar el modelo

↓
Aplicar el modelo en el conjunto de prueba

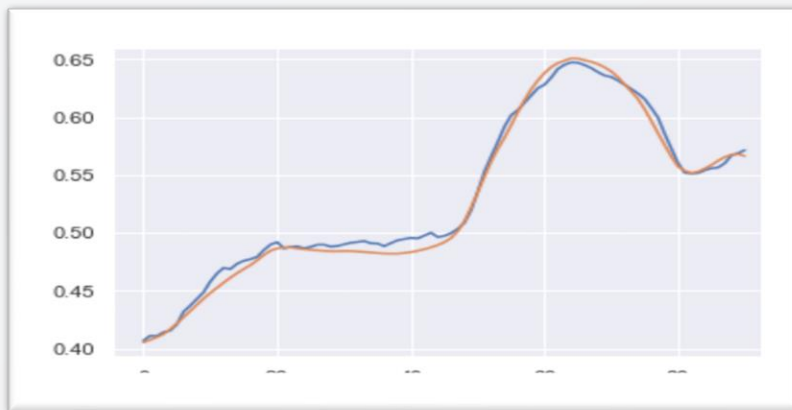
↓
Mostrar los resultados gráficamente

↓
Cálculo de métricas de evaluación

APLICACIÓN DE LOS MODELOS

GRU

Unidad de procesamiento en redes neuronales recurrentes que recuerda información del medio plazo. Tiene una estructura más simple que el LSTM



Importar librerías

↓
Escalar los datos

↓
Dividir datos (entrenamiento, validación y prueba)

↓
Definir y compilar el modelo

↓
Entrenar el modelo

↓
Aplicar el modelo en el conjunto de prueba

↓
Mostrar los resultados gráficamente

↓
Cálculo de métricas de evaluación

CONCLUSIONES

| MODEL Dataset | ARIMA | XGBOOST | LSTM | GRU |
|---|--|--|--|---|
| Dataset del histórico de precios | MSE: 73105534.9 MAE: 7511.206 RMSE: 8550.177 MAPE: 0.276 | MSE: 14679766.98 MAE: 3539.652 RMSE: 3831.4184 MAPE: 0.126 | MSE: 0.00024 MAE: 0.01343 RMSE: 0.0157 MAPE: 0.032 | MSE: 0.00017 MAE: 0.0107 RMSE: 0.013 MAPE: 0.026 |
| Dataset con indicadores técnicos y análisis de sentimientos | MSE: 12106086.8 MAE: 2923.815 RMSE: 3479.380 MAPE: 0.163 | MSE: 1179253.645 MAE: 883.043 RMSE: 1085.934 MAPE: 0.039 | MSE: 0.00019 MAE: 0.01124 RMSE: 0.0138 MAPE: 0.021 | MSE: 4.131e-05 MAE: 0.0053 RMSE: 0.0064 MAPE: 0.010 |



Business
School

MÁSTER EN DATA SCIENCE AND BUSINESS ANALYTICS

MUCHAS GRACIAS

Violeta Chamosa Fondenvila