

CaixaBank Tech- Hackathon

Reto Data Science - IBEX35

Marina Boyero Torrente

1. Preprocesamiento de los datos

1.1 Limpieza

- Eliminación de valores perdidos (NaN)

1.2 Análisis de correlación entre variables



- Se observa que la correlación entre las variables 'Open', 'High', 'Low', 'Close' y 'AdjClose' es muy alta!
- Se mantiene solo 'Close', dado que se quiere predecir si esta variable será mayor o menor en 3 días
- Se evita así problema de **multicolinealidad**

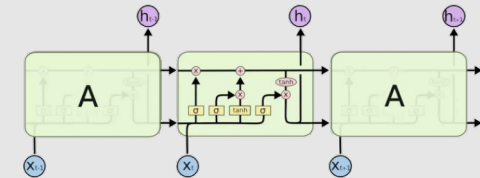
1.3 Extracción de variables relevantes

- No se considera 'Volume' por tomar valor 0 durante varios años, probablemente porque no se registraron datos en estas fechas.
- Se utiliza, por tanto, el **precio de cierre ('Close')** y, como se trata de un problema de series temporales, se utiliza la **predicción (0/1) de los 3 días anteriores**

2. Algoritmo

2.1 Selección del algoritmo

- Al ser un problema de predicción de series temporales, se utiliza la red **LSTM** (Long-Short Term Memory).
- Red neuronal recurrente que tiene la **capacidad de recordar información/eventos pasados** y utilizarlos para predecir valores futuros.



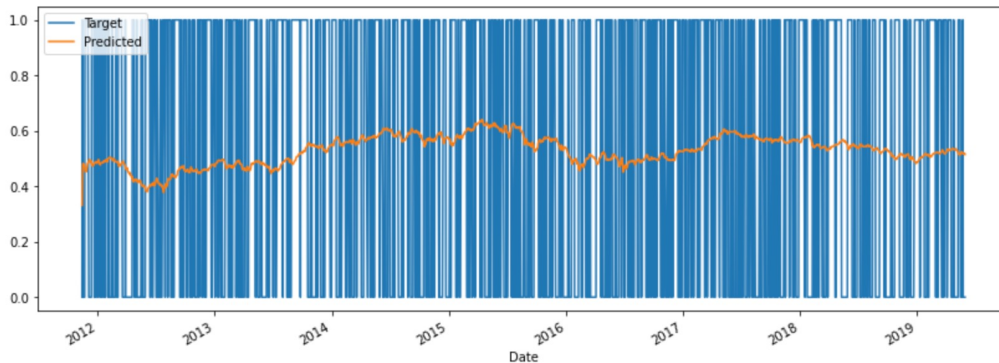
2.2 Selección de los hiperparámetros del algoritmo

- N° de neuronas = 100
- N° de capas ocultas = 1
- N° de capas de salida = 1
- Función de activación = relu
- Función de pérdida = mse
- Optimizador = 'Adam'

3. Entrenamiento

3.1 Separación del conjunto de entrenamiento en entrenamiento (70%) y validación (30%)

- Objetivo: **evaluar** “performance” del modelo

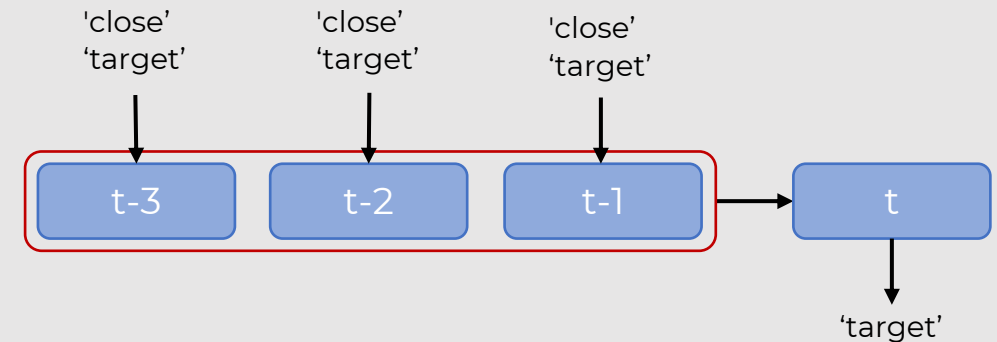


- Rmse = 0.5
- Modelo muy mejorable → Jugar con hiperparámetros de la red para obtener mejores resultados

4. Predicción

4.1 Sobre el conjunto de test

- Predicción inicial en base a los 3 últimos días de entrenamiento
- Cada predicción se realiza sobre los datos de los 3 días anteriores



4.2 Se exportan predicciones

- En formatos requeridos

