

Carlos Cabruja - Data Science

<https://nuwe.io/users/carloscabruja>

28 MAYO 2022 / ON-LINE

CAIXABANK TECH HACKATHON





REACT



DATA SCIENCE



#TECH4U

 CaixaBank  NUWE

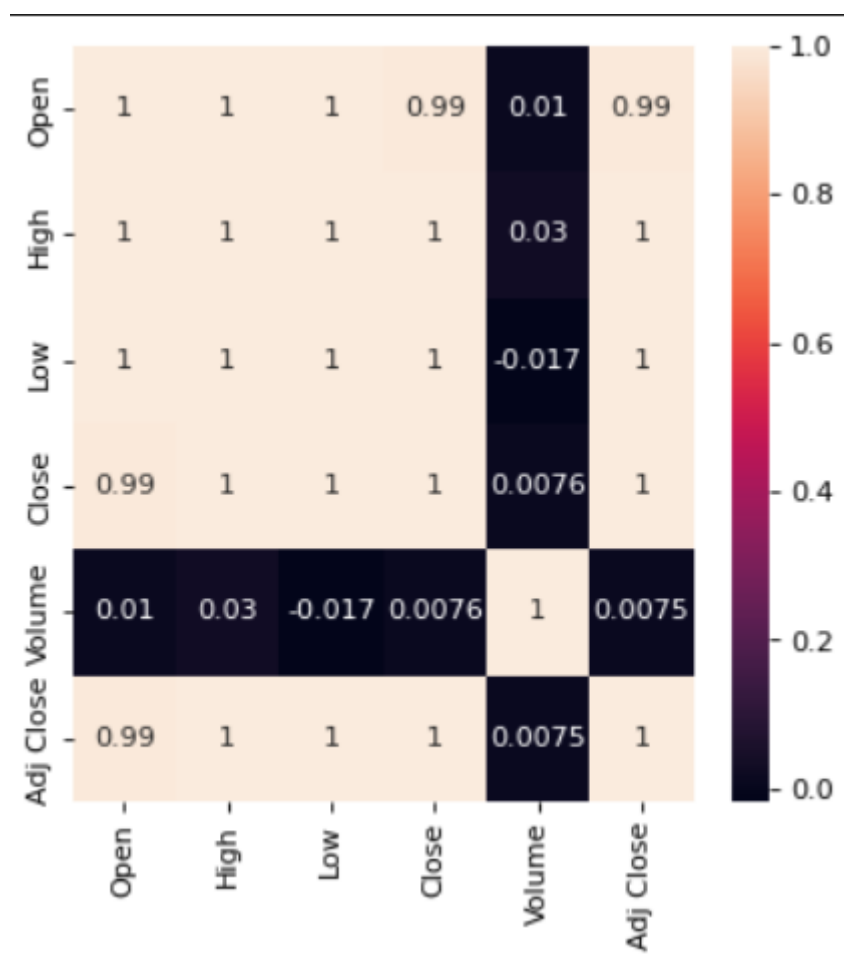
Resumen del Análisis Exploratorio

Se cargaron los datos y se unieron los datos de entrenamiento y de prueba con los datos de los tweets, para ello se usó una herramienta para el análisis de sentimiento a través de procesamiento de texto llamada “VADER” (*Valence Aware Dictionary and sEntiment Reasoner*).

Al unir los datos suministrados por los tweets, se decide no trabajar con datos menores al 2015 (Fecha de los datos de los tweets) para estudiar en grado de influencia que puede tener los tweets nuestra variable Target.

Se descubre que existe una alta correlación entre las variables indicadores y nos quedamos con las variables que hacen más sentido que son:

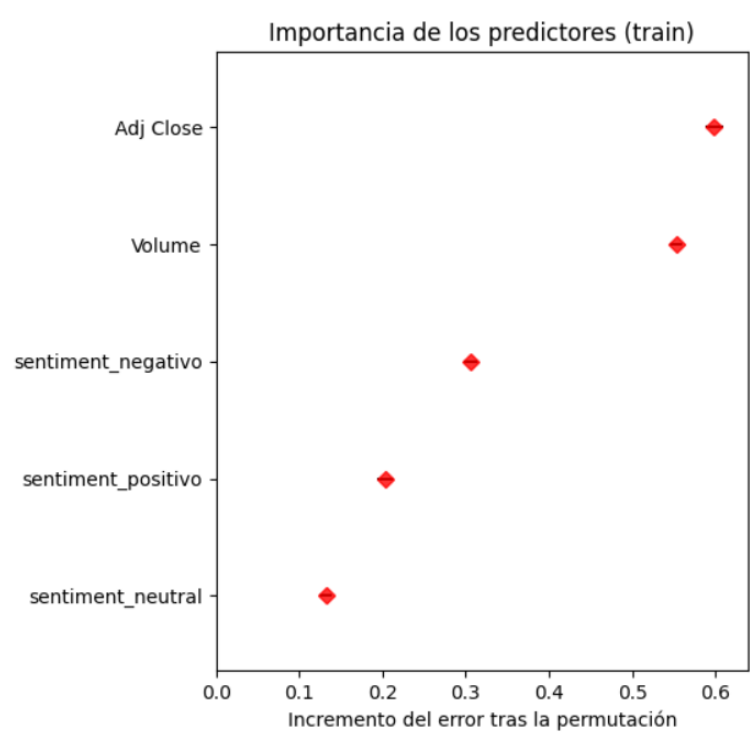
- * Adj close: precio de cierre ajustado por splits y distribuciones de dividendos o plusvalías.
- * Volume: El número físico de acciones negociadas del índice bursátil.
- * Target: El valor de la variable target.
- * Sentiment: El valor de la variable sentimiento.



Para la modelización de nuestro problema nos valdremos de los modelos supervisados de regresión y clasificación, los más populares son *Random Forest* y *Gradient Boosting*.

Modelos que simplifcadamente podemos entender como dadas ciertas condiciones “X”, “Y” es el resultado, un poco parecido a nuestro objetivo que es ver si dado que el precio ajustado, el volumen de la transacción y el sentimiento general de los tweets, el precio de cierre de los 3 días siguientes es mayor o menor que el precio de cierre del día actual.

Vemos que los modelos resultan tener un *accuracy* estimado muy parecido alrededor del **57%**, por lo que no resulta ser muy confiable, veamos en cuanto le aporta los tweets a nuestro modelo con mayor *accuracy*



Al final parece ser que nuestra variable sentimiento no aporta mucho valor a nuestro modelo, pero es interesante ver que de los 3 sentimientos, el que más aporta es el sentimiento negativo, habría que recabar más datos, midiendo el volumen de los tweets, a ver si realmente pueden llegar a cambiar la tendencia del indicador.