

Proceso que se ha seguido

1. Se han importado las principales librerías.
2. Se ha hecho un estudio de las variables para identificar los 'outliers' y la distribución de la muestra.
3. Dado que para un modelo no importa mucho y tras ver que tampoco había tanta diferencia entre el cierre ajustado y el normal, se ha calculado todo con el cierre ajustado. Según la teoría del trading se ha valorado el volumen como la fuerza del mercado. Es decir, a mayor volumen mayor confirmación de la tendencia.
4. Se han tratado los datos como si fueran velas con su mecha y cuerpo que corresponde a los valores de High and Low, Open, Close. Y el volumen era la variable que daba fuerza a los precios.
5. Se ha hecho pruebas con distintos juegos de variables calculando las mechas y el cuerpo de las velas. Interpretado los distintos patrones que hay de velas.

Proceso que se ha seguido

6. Se ha discretizado la variable “Volume” para que la muestra sea más sencilla de interpretar para el model
7. Con todos los datos ajustados se han pensado en diseñar 3 modelos, random forest, KNN y stacking classifier
8. El formato de los tweets era muy complejo de analizar y la información que había dentro de ella no se ha considerado relevante para el análisis.
9. La variable fecha tampoco se ha considerado relevante para el análisis.