

1. Exploración y análisis de los datos:

- 1.1. Abrir la base de datos y explorarla;
- 1.2. Mirar los valores nulos y que procedemos a cambiar los valores de cierre por el valor de cierre previo y poner el volumen a 0
- 1.3. Hacemos una exploración mínima para observar el comportamiento del índice al largo del tiempo y ver que no se dislumbra un comportamiento cíclico.
- 1.4. Generamos columnas nuevas que incluyen la información de la oscilación máxima diaria y la variación de la acción de apertura a cierre. Guardamos diferentes dataframes para valorar la aportación.

2. Cribado de modelos y dataframes con Pycaret:

Se determina que el mejor método es el y el mejor dataframe incluye todas las columnas añadidas.

3. Optimización del mejor modelo con el mejor dataframe

Se determina que el mejor método es el y el mejor dataframe incluye todas las columnas añadidas.

4. Generación de la predicción para test_X

- 4.1. Modificación del test_X al igual que el dataframe utilizado para entrenar el modelo
- 4.2. Rellenamos los valores nulos aprovechando que se puede juntar con la tabla anterior.
- 4.3. para utilizarlo con el modelo optimizado.

5. Pasos propuestos (no realizados por falta de tiempo):

- Optimizar mejor el modelo predictivo (con Optuna)
- Probar otros modelos de los top3 y también algún sencillo de red neuronal.
- Añadir la información que puede aportar la base de datos de Tweets al data set, i volver a probar diferentes modelos. *Tenía la idea que el número de tweets por día ya es un indicador (si es alto respecto al promedio, indica miedo o complacencia porque va muy mal o muy bien, normalmente lo primero). Por otro lado, podría hacer un análisis de sentimiento de las frases y los iconos de los tweets con la librería pysentimiento.*

Gràcias por todo,

Saludos cordiales,

Marc Humet Montada