

Juan Diego Sánchez

00344455

2025-03-30

Proyecto #1

1.- Business Understanding (Entendimiento del Negocio/ Problema)

Objetivo Principal

- Realizar un análisis de sentimiento sobre los tweets publicados por Elon Musk desde el 2022-07-05 hasta el 2023-06-13 y relacionar su impacto en la variación del precio por acción de Tesla (TSLA)

Preguntas de Negocio

- ¿Existe correlación entre el sentimiento expresado en los tweets y movimientos significativos en el mercado de valores?
- ¿Se puede modelar la variación futura del precio de la acción de Tesla en función del sentimiento expresado en sus publicaciones?

2.- Data Understanding (Entendimiento de los Datos)

Se trabajará con un dataset publicado en kaggle.com de la cual se extraerá el texto del tweet, fecha de publicación para luego ser procesados y clasificados en función de un análisis de sentimiento. Por el dato del mercado de valores se extraerá el valor de la acción en USD al cierre junto con el volumen de transacciones realizadas durante el mismo periodo de tiempo

3.- Data Preparation (Preparación de los Datos)

Se inicia la preparación de los datos revisando si existen valores duplicados o vacíos. Se concluye que dentro de las características a ser evaluadas no existen datos que puedan afectar el análisis. A posterior, se ajusta el formato de la columna fecha a datetime.

En el caso del dataset de tweets se procede a homogeneizar el contenido del texto; se elimina los URLs, menciones, hashtags, signos de puntuación, se convierte todo a minúsculas y se eliminan todos los pronombres que a posterior podrían llegar a desdibujar el análisis. De igual modo, se crean nuevas filas con el año, mes y día para posteriores análisis. Se calcula la intensidad de sentimiento obteniendo un número que oscila del -1 a 1 que será estandarizado en clasificadores, e.g: negativo, parcialmente negativo, neutral, parcialmente positivo y positivo.

Finalmente, se realiza un merge para juntar ambos datasets usando como llave la fecha. A partir de lo cual se procede a realizar graficas para poder analizar de mejor manera el dataset, se empieza por generar un diagrama de líneas en el mismo gráfico, pero distintos ejes el sentimiento y el precio de cierre, también se gráfica el volumen diario para buscar correlaciones entre estas variables. Se gráfica de igual manera la distribución de sentimientos agrupada por mes, un scatterplot relacionando el sentimiento y el cambio porcentual en el precio y un boxplot distribuyendo el sentimiento según el volumen previamente categorizado usando la mediana como umbral.