

Informe

Realizamos un modelo de inteligencia artificial que es capaz de analizar sentimientos a través del lenguaje natural, a raíz de oraciones o comentarios puede determinar si el comentario o la opinión es positiva, neutral o negativa.

Utilizamos un DATASET que comentarios de Twitter, cargamos los datos utilizando la librería de pandas.

Creamos una función para el proceso de datos que nos permite colocar todas las letras en minúsculas, nos elimina los caracteres especiales, las palabras vacías y usamos la lematización para tener más precisión.

Las etiquetas que nos identifican si el pensamiento es neutral, negativo o positivo la convertimos a números, asignando el 0 a los negativos, 1 a los neutros y 2 a los positivos y lo guardamos en una nueva columna llamada 'label'.

Realizamos una vectorización.

Dividimos el conjunto de datos entre la data de entrenamiento y data de prueba utilizando la librería de train_test_split.

Entrenamos el modelo, con una interacción de 1000 que fue con lo que mejor resultados obtuvimos.

Realizamos una evaluación con la data de prueba, que fue la data que el no observo.

Y tenemos los resultados:

- Tenemos una precisión de un 83% lo que indica que nuestro modelo es bueno.
- En cuanto a las predicciones negativas el 84% de ellas fueron correctas, el 77% de las neutras fueron correctas y el 90% de las positivas fueron correctas.
- En cuanto a recall el modelo solo detecto el 62% de los textos realmente negativos demostrando que hay un sesgo, detecta casi todos los comentarios neutro con un 97% y es bueno detectando los comentarios positivos con un 83%.