

Proyecto Recuperación de Información: Análisis de Sentimientos

En diferentes sitios de comercio electrónico los clientes pueden evaluar productos para auxiliar a otros clientes que desean comprar el producto, estableciendo la evaluación mediante estrellas, con una escala de 1 a 5, siendo uno la más baja y 5 la más alta. Se ha observado que en los sitios que proporcionan esta facilidad a los clientes de manera adecuada, los vendedores aumentarán sus ventas, mientras que los clientes completarán su proceso de compra sin ningún problema.

Para este ejercicio se les proporciona un archivo con 4915 reseñas de productos (reviewText), y sus correspondientes evaluaciones (overall), indicadas con un el número de estrellas.

1. Procese las reseñas y de acuerdo con las estrellas defina cuales serán las que considerarán como positivas y negativas, así como si elimina algunas.
2. Genere una gráfica de pastel indicando el número de opiniones positivas, negativas y eliminadas (si existen).
3. Preprocese las reseñas y genere el vocabulario. Guárdelo en un archivo y muestre en pantalla su longitud además de escribir los 50 términos más comunes.
4. Identifique las opiniones positivas de acuerdo con el criterio definido y etiquete las reseñas con positive y negative.
5. Creen una nube de palabras con las 500 más comunes en las reseñas positivas y 500 en las negativas.
6. Para la clasificación utilicen 80% de los datos para entrenar y 20% para probar
7. Utilicen los SVM, C4.5 y Regresión logística como para clasificar
8. Muestre la matriz de confusión, presión, recuerdo y medida f, así como la exactitud de la clasificación para cada algoritmo.
9. Escriba su reporte y súbalo junto con el código desarrollado a Teams.

Se recomienda realizar la programación en Python, siéntanse libres de utilizar bibliotecas tal como: sklearn, collections, Counter(), WordCloud(), CountVectorizer(), nltk, matplotlib.pyplot, es decir podrán utilizar las bibliotecas de Python que juzguen necesarias.