

Práctica Final - Análisis de Sentimiento con NLP

Resumen Ejecutivo

Este informe resume el desarrollo y evaluación de un modelo de análisis de sentimiento aplicado a reseñas de productos de Amazon. Se utilizaron múltiples técnicas de preprocesamiento, vectorización TF-IDF y modelos supervisados como Regresión Logística, Naive Bayes, SVM y Random Forest para clasificar sentimientos positivos y negativos. A continuación, se presentan los detalles técnicos, resultados obtenidos y conclusiones basadas en métricas de evaluación.

1. Preprocesamiento

Se eliminaron signos de puntuación, se transformaron números a palabras y se eliminaron stopwords usando NLTK. Además, se aplicó lematización para normalizar las palabras a su forma base, y se utilizó TF-IDF como vectorizador para convertir texto a valores numéricos.

2. Modelos Entrenados

Se entrenaron cuatro modelos distintos con los mismos datos preprocesados:

- Regresión Logística
- Naive Bayes Multinomial
- Support Vector Machine (SVM)
- Random Forest Classifier

Comparativa de Modelos

| Modelo | Accuracy Test | F1 Score Promedio |
|---------------------|---------------|-------------------|
| Logistic Regression | 0.928 | 0.93 |
| Naive Bayes | 0.935 | 0.93 |
| SVM | 0.933 | 0.93 |
| Random Forest | 0.902 | 0.9 |

3. Conclusiones

Todos los modelos obtuvieron buenos resultados, destacando Naive Bayes y SVM con una precisión superior al 93%. Sin embargo, el modelo más balanceado fue la Regresión Logística, con excelente

Práctica Final - Análisis de Sentimiento con NLP

rendimiento tanto en validación como en test.

Por lo tanto, se recomienda utilizar Regresión Logística como modelo final para este proyecto de clasificación de sentimientos en reseñas.