

Sistema de Análisis de Sentimientos con LSTM y Word2Vec para Clasificación de Reseñas

Franklin Manuel Manjarres

El proceso inicia con la instalación y ajuste de dependencias críticas (`NumPy`, `TensorFlow`, `Gensim`), donde se resuelven conflictos de versiones. Luego, se carga un *dataset* de 1,600 reseñas hoteleras desde *GitHub*, se limpian los textos (eliminación de *stopwords*, lematización) y se tokenizan las secuencias. Los datos se dividen en entrenamiento (80 %) y prueba (20 %), aplicando *padding* para estandarizar la longitud de las entradas a 500 tokens.

Se construye un modelo con *embeddings* preentrenados de `Word2Vec` (*Google News*, 300 dimensiones) no entrenables, seguido de una capa `LSTM` bidireccional con *dropout* para prevenir sobreajuste. Tras 6 épocas de entrenamiento (*early stopping* en `val_loss`), el modelo alcanza un 92 % de precisión en el *set* de prueba, con métricas balanceadas (F1-score: 0.92) y una matriz de confusión que muestra un rendimiento consistente entre clases positivas y negativas.