Abstract

SENTIMENT ANALYSIS - Amazon Customer Reviews

Abstract

The rapid growth of e-commerce platforms has resulted in a plethora of customer reviews that serve as valuable resources for both businesses and consumers. However, manually analyzing these reviews for sentiment is a huge task. This project aimed to automate the sentiment analysis process of customer reviews by using Natural Language Processing (NLP) and Machine Learning techniques. Specifically, the project focuses on Amazon product reviews.

The dataset comprises over 130,000 customer reviews, each with associated star ratings. The text data undergoes a series of preprocessing steps, including tokenization, removal of stop words, and text normalization. The cleaned reviews are then transformed into numerical vectors using the Term Frequency-Inverse Document Frequency (TF-IDF) technique. Labels for sentiment ('Positive,' 'Neutral,' 'Negative') are derived from the star ratings.

To address class imbalance, Random Oversampling is applied to balance out the distribution of different sentiment labels. A Logistic Regression model is trained on this balanced dataset, and hyperparameter tuning is conducted using Grid Search with Cross-Validation. The final model, which was tuned, achieved an accuracy of approximately 75%, with improved F1-scores for the 'Negative' and 'Neutral' classes

Although the model showed promising results, further improvements are still possible throug, feature engineering, and the application of more complex algorithms

-S. Saini