Abstract

Sentiment analysis is crucial for understanding user opinions and reactions towards products or services. Movie reviews encompass complexities involving various aspects such as plot, acting, and visuals. Understanding sentiment based on these reviews can provide valuable insights for filmmakers, directors, and policymakers. However, sentiment analysis of movie reviews is not necessarily easy due to potential challenges. These challenges may include the use of unstructured review language, inappropriate feature selection, complex datasets, or computationally intensive classification methods.

In this research, the IMDb dataset containing 50,000 review data points along with sentiment labels (positive, negative) is utilized. The classification algorithm used is Naïve Bayes, specifically Multinomial Naïve Bayes. This algorithm is supported by vectorization using TF-IDF and feature selection using chi-squared.

The research results indicate that the Multinomial Naïve Bayes model with optimal TF-IDF parameters achieves an accuracy of 86.14%. The model's performance improves to 87.71% after implementing chi-squared with optimal parameters.

This study provides insights into techniques and challenges in sentiment analysis of IMDb movie reviews using the Multinomial Naïve Bayes classification method. It also tests preprocessing parameters to determine and evaluate the resulting performance.

Keywords: Multinomial Naïve Bayes, TF-IDF, chi-squared, accuracy, IMDb