

Abstract

The rapid digital transformation has driven the increased production and consumption of film reviews on the internet, which is a valuable source of data in analyzing public opinion. This study aims to classify movie review sentiment from the IMDb dataset into positive and negative categories using the Support Vector Machine (SVM) algorithm. A total of 50,000 review data was used in this study, which was balanced between positive and negative classes. The analysis process begins with the pre-processing stages of data in the form of case folding, tokenization, stopword removal, and stemming. Furthermore, feature extraction was carried out using the TF-IDF method in two schemes, namely unigram and bigram. The classification model was then built with two types of SVM kernels, namely linear and sigmoid kernels, and evaluated using accuracy, precision, recall, and F1-score metrics. Based on the test results, the sigmoid kernel showed superior performance with an accuracy value of 86% and an F1-score of 86.12%. Meanwhile, TF-IDF Bigram produces the highest F1-score compared to Unigram. The results show that kernel selection and feature schema have a significant influence on classification accuracy, and confirm that the combination of sigmoid kernel and TF-IDF Bigram is the most optimal configuration in the sentiment analysis task for film reviews.

Keywords: Sentiment Analysis, Movie Reviews, Support Vector Machine, TF-IDF,IMDB