

Abstract

Many films can be an option to watch, but not necessarily films that are of good quality. Film reviews are very important because they are able to provide information about the quality of a film so that it can be used as a reference for watching a film. Due to the large amount of data regarding film reviews and it takes a long time to read them manually one by one, a system that can perform sentiment analysis in a short time is needed. Machine learning methods such as K-Nearest Neighbor (KNN) can be used to perform sentiment analysis, but have poor performance if the features used are not well defined. To determine the features, the use of feature extraction and feature selection needs to be used. In this study, Term Frequency-Inverse Document Frequency (TF-IDF) is used for feature extraction and chi-square is used for feature selection. With a combination of KNN methods, TF-IDF feature extraction, and chi-square feature selection, it can provide the best performance evaluation results at $K = 37$ with an accuracy value of 80.75% accuracy, 83.24% precision, 77.00% recall, and F1-measure. 80.00%.

Keywords: sentiment analysis, TF-IDF, chi-square, KNN