

Abstract

The film industry has always developed rapidly over time. There are many considerations to determine a quality film, film reviews are one of the factors that have an important role in determining a quality film. The IMDb site is one of the platforms used to accommodate one's sentiments towards films. Machine learning methods can be used to make it easier for us to summarize or analyze the many opinions that exist on the platform. In this final project, a sentiment analysis system was built using the KNN machine learning method by combining the extraction features of TF-IDF and Lexicon SentiWordNet. The data used in this final project is 2000 film review data from the IMDb website. The final result of this research is an analysis of combining the TF-IDF feature extraction method with Lexicon SentiWordnet and testing application selection feature Information Gain (IG). From the results of experiments, combining the TFIDF extraction feature with Lexicon SentiWordnet has an accuracy result that is not higher than using only the TF-IDF extraction feature, which is 73.31%, and the use of the IG selection feature with the right threshold is able to optimize the performance results.

Keywords: sentiment analysis, KNN, Lexicon SentiwordNet, TF-IDF, movie review.