Abstract

Over the years, the development of the film industry has become increasingly rapid and in great demand. Easy access to enjoy a movie is one of the factors that make movie fans increase. Film reviews are a place for film connoisseurs to submit a comment, either positive or negative. One of the techniques used to categorize whether the film reviews are classified as positive or negative is sentiment analysis or opinion mining. However, the large number of features contained in the dataset will become a problem in the classification process, resulting in decreased performance in the classification process because many features are less relevant. So, this study uses the Gini Index feature selection with the Support Vector Machine classification method using a film review dataset with a comparison of training data and test data of 70:30. The results of using the Gini index feature selection obtained results of 89.21% with a feature reduction of 25,242 from the total feature selection of 30,242, compared to without using feature selection which obtained an F1-Score of 86.1%.

Keywords: sentiment analysis, gini index, support vector machine, moview reviews