

Abstract

At this time, the world film industry is growing quite rapidly and loved by the world community. To find out whether a film is interesting or not, we have to see movie reviews from the reviewers of that film. Some reviews may look clearly for categorized into positive or negative reviews, but there are some reviews that are not clearly categorized. One of the automation techniques to help classify reviews is sentiment analysis. K-Nearest Neighbor (KNN) is a fairly simple classification method. However, K-Nearest Neighbor has problems related to high feature dimensions. To overcome this problem, this study conducted a sentiment analysis of movie reviews using the K-Nearest Neighbor method and Gini Index feature selection on Internet Movie Database (IMDB) dataset which was divided into positive sentiment and negative sentiment. Feature selection using the Gini Index can reduce high feature dimensions to improve K-Nearest Neighbor performance. By paying attention to the value of k and the Gini Index threshold, the average accuracy of using the K-Nearest Neighbor method without feature selection is 79,6% and the best average accuracy using the K-Nearest Neighbor method with the Gini Index feature selection is 82,1%. This proves that the Gini Index feature selection can improve the performance of the K-Nearest Neighbor method.

Keywords: movie reviews, sentiment analysis, K-Nearest Neighbor, Gini Index, feature selection