

Abstract

The rise of beauty products in recent times can make consumers hesitate to choose a beauty product, especially for women. Beauty product reviews have become a very valuable source of information for consumers in making decisions to purchase a product in improving their products and marketing strategies. The process of sentiment analysis on negative and positive beauty product reviews will be classified one by one. Therefore, in this study, sentiment analysis was applied to the beauty product review data using the K-Nearest Neighbor (KNN) method to find the best k in the case of this study. The dataset used will be pre-processed with case folding, noise removal, tokenization, stemming, stopword removal and slang words, for feature extraction using Term Frequency Inverse Document Frequency (TF-IDF) to calculate the weight of a word in the document, and the feature selection method uses Chi-Square which aims to select the features needed to increase the accuracy value. In this study, the best accuracy value was 71% of the data classified using KNN with a k value of 50 and the model on feature selection with 76 features.

Keywords: K-Nearest Neighbors (KNN), Term Frequency Inverse Document Frequency (TFIDF), Chi-Square, Sentiment Analysis, Product Reviews