Abstract

Forest fires and haze are natural disaster phenomena that are of national and international concern. Forest fires in Indonesia have become a frequently discussed topic not only for Indonesians but also for people in other countries that are affected by the event as well, especially on social media such as Twitter. Twitter is one of the most popular social media in the world with 24 million active users in Indonesia, making it one of the main platforms for Indonesians to discuss a particular topic. Opinions that are expressed in a discussion related to the topic of forest fires that happened in Indonesia on Twitter social media can be utilized to analyze the sentiment of the Indonesian people towards that event. There are two approaches that can be used for sentiment analysis, namely the lexicon approach and the machine learning approach. One of the commonly used methods is the Support Vector Machine (SVM), a method based on machine learning. SVM is a method that has the ability to linearly separate data, making it a widely used algorithm in classification problems. Therefore, this study analyzes Twitter sentiment towards forest fires and haze using the SVM method. The constructed SVM model can accurately predict the sentiment of Twitter users towards forest fires and haze in Indonesia. The tests showed that the SVM model trained using the upsampled dataset achieved the highest performance. Three test scenarios were conducted by varying the data's split ratio, applying the upsampling resampling technique, and performing hyperparameter tuning. The test results indicate that the SVM model trained with a dataset after upsampling and a split ratio of 85:15 achieved the highest sentiment prediction performance with an f1-score of 0.854, precision of 0.855, and recall of 0.855.

Keywords: forest fires, haze, Twitter, sentiment analysis, SVM, machine learning