

Abstract

Twitter is one of the social media by Indonesians. In Twitter there are many crimes that occur most often which is one of the information or opinions from the public, one of the events or the election program of the West Java Governor 2018. The election of the Governor of West Java 2018 is followed by candidates from political figures who are competent in their fields, namely Ridwan Kamil, Deddy Mizwar, Sudrajat and TB Hasanuddin. Instead of these events or events, opinions or sentiments can be used in the form of positive, negative or neutral sentiments by using Sentiment Analysis. And data sourced from tweets or tweets about candidates for Governor in the West Java 2018 Governor Election on Twitter by using keywords of candidate names such as "ridwankamil", "deddymizwar", "sudrajat", and "tbhasanuddin", while for the process classification of classes, either in the form of positive, negative or informal classes. The Naïve Bayes Classifier algorithm and also the TF-IDF weighting. So the purpose of this research is to look at the direction from the point of view of those who elect the Governor in the West Java Governor Election in 2018 by taking into account the number of positive, negative and neutral classes per the name of the candidate Governor. And for the scenario used in this study is the use of stopword and without keywords and N-gram features such as Unigram, Bigram, Trigram also uses training composition data of 60%, 70% and 80% in testing and classifying sentiment classes. The results obtained in this study for the amount of accuracy on the use of stopword that is equal to 97.92% obtained from TB Hasanuddin data on the Trigram feature and the composition of the training data is 60%. Then to get data from Deddy Mizwar's data with the Trigram feature and the percentage of training data is 60% with the accuracy of 72.92%. As for the calculation of accuracy in use without stopword, the highest accuracy is obtained from TB Hasanuddin data with Trigram features and the percentage of training data is 60% with an accuracy of 97.53%. And for the smallest accuracy on the use without stopword, it is obtained from Deddy Mizwar data with the Bigram feature and the percentage of training data is 70% with an accuracy of 72.22%. And the results from the sentiment testing for the highest positive class were obtained from the data of Sudrajat data with Trigram feature and the percentage of testing data was 40% with the number of positive class values on the classification was 104 and accuracy was 95%. Then to find out the exact amount of data in the highest neutral class from TB Hasanuddin data with Trigram features and the percentage of testing data is 40% with the total value of the neutral class is 203 classes and accuracy is 95.83%. While for the negative class category with a large amount of data from Deddy Mizwar data with Trigram features and 40% reduction in testing data with a negative class value of 91 classes and an accuracy of 77.5%.

Keywords: Twitter, Sentiment Analysis, Machine Learning, TF-IDF, Naïve Bayes Classifier (NBC)