Abstract

Today, Hoax is developing very massively. Various media used as a place for Hoax to circulate, one of which is in the news. Many studies have detected Hoax news using text classification, but only using one classification model. In this study, two classification models are used, namely Naïve Bayes and Support Vector Machines. This study uses a dataset containing 1,381 records consisting of Hoax and No Hoax news in English. Furthermore, the dataset is preprocessed first. Then the preprocessed data is processed in the word weighting process. The results from word weighting are then processed into the Naïve Bayes classification model and Support Vector Machine. The classification model validation resulted in an accuracy of 96.21% for Naïve Bayes and 97.22% for the Support Vector Machine for dataset distribution (70% Trainset and 30% Testset). Whereas the distribution of datasets (60% Trainset and 40% Testset) yields an accuracy of 94.50% for Naïve Bayes and 96.02% for Support Vector Machines. The differences in the accuracy of the two classification models are due to differences in the way they work and the distribution of datasets. The Naïve Bayes classification model uses probability methods with unrelated (stand-alone) classes. Due to the interrelations between classes, it causes a decrease in the value of accuracy. Whereas the Support Vector Machine classification model uses dimensioned space to classify Classes. So Class can be related to each other. This causes the accuracy value can be maximized. Both of these models are Supervised Learning. This means that both of these algorithm models require learning in order to be able to do classification. In this case, the data used for learning is called Trainset. The more data in the train it will make learning algorithms better. Based on the results of the study, the Support Vector Machine classification model produces better accuracy than Naïve Bayes.

Keywords: News, Hoax, Classification, Naïve Bayes, Support Vector Machine