Abstract

Classification on text mining has many benefits, one of which is determining a spam or ham type message. This research is motivated by the number of SMS messages that contain fraud, links that contain malware, calls to bet and many other types. Text messaging with an example like that is a spam type sms message. So from that in this study a system will be developed that can classify spam SMS messages and those that don't. The system built will use the Support Vector Machine classification method and the Singular Value Decomposition feature selection method using the English-language dataset. The results of this study obtain the highest accuracy, the fastest execution time, and the limit of the amount of data that can be executed without using the Singular Value Decomposition feature selection method. The highest F1-score is equal to 90,95% using the Gaussian RBF kernel with a machine that has been added to the Singular Value Decomposition feature selection. The fastest execution time was 8,16 seconds by using Linear kernel with nv variable of 100. While the limit of the most amount of data that can be executed without using the Singular Value Decomposition feature selection method is 2500 data from a total dataset of 5572 data.

Keywords: Support Vector Machine, Singular Value Decomposition. Text Classification, Spam Message