Abstract

E-mail spam is an unsolicited message which usually goes into our e-mail account automatically and sent in bulk. The contents of this e-mail spam can be of various kinds, such as advertising, pornography, as well as news of fraud. Sending an e-mail spam to our e-mail account can cause a negative impact. Filtering spam e-mails manually can spend time because the recipient must be distinguished which the e-mails are regular and e-mail spam. Therefore, made a system for classifying e-mails which are regular and e-mail spam

In this final project, Support Vector Machine and Naive Bayes are implemented as classifier. Further examination of the results of the accuracy, error rate, reproduction ratio value, and relevance ratio value. Then performed an analysis of the results of testing the system by using Support Vector Machine and Naive Bayes.

The research in this final project produces an accuracy value of 92.67058%, the error rate of 7.32942%, reproduction ratio of 93.27662%, and the relevance ratio of 94.59148% with Linear SVM. While with the Naïve Bayes yield value of 81.89082 % accuracy, error rate of 18.10918 %, reproduction ratio of 72.83212 %, and relevance ratio of 96.57188 %.

Keywords: *classification, e-mail spam, support vector machine* (SVM)*, naïve bayes*