Abstract

There are abundant reserves of text document in electronic form, and is getting bigger each day. Those documents represent a bulk of information, those of which can be easily accessed. To find any value in this huge collection required some kind of organization; most of the task in document organization can be automated via text classification. Accuracy and the level of understanding of the system used can increase the function significantly.

This final assignment is a research to improve the performance of Naive Bayes Classifier in multiclass classification with the implementation of Error-Correcting output coding (ECOC), these methods are used to classify news documents into news class classification. The datasets used are news in bahasa Indonesia taken from several news website.

The research shows that the usage of ECOC can improve classification accuracy within the usage of Naive Bayes Classifier up to 66%. This improvement in accuracy was achieved in expense of increased training time, which coincides linearly with the code length used. In this final assignment, the research in ECOC Classifier was done with varying code length, in search of the threshold which resulted in the best accuracy in comparison with normal Naive Bayes Method.

Keywords*: text mining, text classification, multiclass, naïve bayes, error-correcting output coding (ECOC), indonesian news*.