

Abstract

The development of Internet technology is very fast to make the amount of information of text documents is increasing. Therefore, the method is needed to find information through the categorization process. But the high-dimensional data can interfere with the performance results of categorization. Therefore feature selection is needed greatly affect categorization is feature selection. There are several algorithms in feature selection, one of which is ambiguity Measure. In this final report implements AM (ambiguity Measure) feature selection algorithm, to analyze the results of feature selection using a threshold value for selecting these features that influence the categorization process. Then the observed value of precision and recall using naïve bayess algorithm contained in the Weka tool. After doing the experiment by setting the threshold value for feature selection show the higher set threshold value then number of features selected by the system decrease, but the performance of the categorization increases. Categorization performance reaches the highest value when the specified threshold with 0.95. Then compare the accuracy of the dataset prior to feature selection and feature datasets after selection, the results yielded accuracy after feature selection is higher than the dataset prior to feature selection.

Keyword : *feature selection, text categorization, Ambiguity Measure*