

## **Abstract**

Words in a document that frequently occurring but meaningless in categorization are called as stopwords. Words that categorize as stopwords do not contribute for categorization, they should be remove during indexing before categorization process. However, using a single fixed stopword list across different document collection could be detrimental the performansi of categorization.

On this final project, Term-Based Random sampling is used as approach to provide stopword list automatically for document collection that is processed. This approach, define how informative a term is. So, it's enable us to define a stopword list automatically. The stopword list that is produced will evaluate by categorization step for document that use this stopword list.

This approach will evaluate by comparing the result of categorization that produce by use preprocessing using stopword list that produce by this approach with the performansi using stopword Salton and Buckley I, Salton and Buckley II, Google's stopword, default English Stopword, and also comparing with categorization's performansi without using stopword.

From the evaluation, better stopword list is coming from stopword list that produce by using Term-Based Random Sampling Approach. The accuracy is 88,24%.

### ***Keyword***

Categorization, term, stopword