Abstract

Text Classification is an important problem in text mining used to categorize an undefined label. In this work, various classification models have been evaluated after pre-processing of the text dataset. The pre-processing steps include tokenization, stop word removal and stemming, after which different term weight scheme have also been implemented. Various pruning techniques have also been implemented to get the maximum count of the terms. Based on this analysis, we summarized that Naïve Bayes method gives the highest accuracy while comparing with other state of the art text classifiers.

References


Index Terms

Computer Science

Data Mining
Keywords

Text Classification Models, Pruning Methods, Vector Space Model, Absolute Pruning