Abstract

In data mining, classification is the way to splits the data into several dependent and independent regions and each region refer as a class. There are different kinds of classifier uses to accomplish classification task. Moreover classification is bounded in case of classifying of text documents. The motives of the work which a present in the article is to evaluate multiclass document classification and to learn achieve accuracy of classification in the case of text documents. Naive Bayes approach is used to deal with the problem of document classification via a deceptively simplistic model. The Naive Bayes approach is applied in Flat (linear) and hierarchical manner for improving the efficiency of classification model. It has been found that Hierarchical Classification technique is more effective than Flat classification. It also performs better in case of multi-label document classification. In contrast to retrospect we observe significant increase in the generation of data each day. And hence with the advent of smarter technologies, data is required to be classified and sorted before framing out decisions from it. There are so many techniques available for classifying documents into various categories or labels. Data mining is the process of non-trivial extraction of novel, implicit, and
actionable knowledge from large data sets.

References

1. Shweta Joshi. "Categorizing the Document Using Multi Class Classification in Data Mining", 2011 International Conference on Computational Intelligence and Communication Networks, 10/2011


Index Terms

Computer Science Artificial Intelligence

Keywords

Data Mining, Mining Techniques, Classification, Document Classification, Naïve Bayes
Classifier.