Abstract

The automated categorization (classification) of texts into predefined categories is one of the widely explored fields of research in text mining. Now-a-days, availability of digital data is very high, and to manage them in predefined categories has become a challenging task. Machine learning technique is an approach by which we can train automated classifier to classify the documents with minimum human assistance. This paper discusses the Naïve Bayes, Rocchio, k-Nearest Neighborhood and Support Vector Machine methods within machine learning paradigm for automated text categorization of given documents in predefined categories.

Index Terms

Computer Science

Information Systems

Keywords

Machine learning  Text categorization.