Abstract

Data mining is the process of extracting knowledge from the huge amount of data. The data can be stored in databases and information repositories. Data mining task can be divided into two models descriptive and predictive model. In the Predictive model, we can predict the values from a different set of sample data, they are classified into three types such as classification, regression and time series. The descriptive model enables us to determine patterns in a sample data and sub-divided into clustering, summarization and association rules. Clustering creates a group of classes based on the patterns and relationship between the data. There is different types of clustering algorithms partition, density based algorithm. In this paper, algorithms are analyzing and comparing the various clustering algorithm by using WEKA tool to find out which algorithm will be more comfortable for the users for performing clustering algorithm. This present the application's of data minning WEKA tool it provide the cluster's huge data set and clustering that provide making hand in the optimizing in search engine.

References


7. Z. Huang,“Extensions to the k-means algorithm for clustering large data sets with categorical values”. Data Mining and Knowledge Discovery,2:283–304, 1998.


15. Sanjoy Dasgupta – Performance guarantees for hierarchical clustering Department of Computer Science and Engineering University of California, San Diego.


21. Madjid Khalilian, Norwati Mustapha, MD Nasir Suliman, MD Ali Mamat, "K-Means Based


Index Terms

Computer Science

Algorithms

Keywords

Data Mining, Clustering algorithms, K-mean, LVQ, SOM, cobweb, WEKA