Abstract

This paper describes how data mining techniques are used in Hadoop for cloud data where it is an open source implementation. Extraction of useful information from raw data is always referred by the term DM. The techniques of DM are integrated into the normal day-to-day life has become very popular. Data mining are useful to improve the efficiency for the reduction of cost in the businesses field. In the cloud computing paradigm, the applications and techniques of data mining are most wanted. The users can retrieve meaningful information from virtually integrated data warehouse and it is allowed by implementing the data mining in cloud computing for reducing the cost of storage and infrastructure. This paper aims at predicts the churn customer in telecommunication industry where the dataset is stored in cloud and implemented using data mining techniques in Hadoop. In this paper, classification is used to analyze the dataset of telecommunication industry and classify the numerical and text data and predict the churners who are likely to switch from current network, and the clustering is used to group the result of the classification from the given data set for the best prediction of numerical and text data together in Hadoop. Hadoop is an environment easy to implement the classification; clustering techniques and cloud are used to store the data set for the industry.
References

- Ruxandra Stefania Petre, Datamining in Cloud Computing, Database system journals, pp. 67-70.
- Jay Gholap, Performance Tuning of J48 algorithm for Prediction of Soil Fertility, pp. 1-5.
- Himel Dev, Tanmoy Sen, Madhusudan Basak and Mohammed Eunus Ali, An Approach to Protect the privacy of Cloud Data from Data Mining based Attacks, pp. 1-9
- Dr. Rajni Jain, Introduction to Data Mining Techniques, pp. 1-11.
- Introduction to Cloud Computing and Hadoop, pp. 1-4.
- Zhexue Huang, Extension to K-Means Algorithm for Clustering Large Data Sets with Categorical Values, pp. 283-304.

Index Terms

Computer Science
Distributed System

Keywords
Datamining  Hadoop  HDFS  Map/Reduce.