© INTERNATIONAL JOURNAL FOR RESEARCH PUBLICATION & SEMINAR
ISSN: 2278-6848 | Volume: 07 Issue: 09 | October – December 2016
Paper is available at www.jrps.in | Email : info@jrps.in

An Analysis of Enhancement in K-Means Clustering

[1]Preeti saini, Research Scholar, Department of CSE, JIET Jind. preetisaini06@gmail.com
[2]Sapna aggarwal, Assistant professor, Department of CSE, JIET Jind. Sapna.ruby@gmail.com

ABSTRACTS: Today in modern era, everyone has to retrieve the large amount of data from a vast collection of data. This process of retrieving useful data in understandable form is data mining. Big data[6] is a term for data sets that are so large or complex that old data processing applications are insufficient. Accuracy in big data might lead to more confident decision making, & better decisions could result in greater operational efficiency, cost reduction & reduced risk. Various algorithms[5] & techniques like Classification, Regression, Artificial Intelligence, Neural Networks, Association Rules, Decision Trees, Algorithm, Nearest Neighbour approach are used for knowledge discovery from databases. Clustering is an important data analytic technique which have a significant role in data mining application. Clustering is the technique of arranging a set of similar objects into a group. Partition based clustering is an important clustering technique. This technique is centroid based technique in which data points splits into k partition and each partition represents a cluster. A widely used partition based clustering algorithm is k- means clustering algorithm. But this method has problem of empty cluster. The problems could be reduced by using an enhanced algorithm. In this paper, we have analysis of the old k-means algorithm and an enhanced k-means algorithm.

Keywords: Clustering, K-Mean, Data mining.

[1]Introduction

Data mining[4] is an interdisciplinary subfield of computer science. It is computational process of recognize style patterns in large data sets involving methods at intersection of artificial intelligence, statistics, & database systems.[1] overall goal of data mining process is to extract information from a data set & transform it into an able to be understand structure for further use. Aside from raw analysis[1] step, it involves database & data management data pre-processing, model & inference considerations, interestingness metrics, complicated thing considerations, post-processing of discovered structures & online updating. Data mining is analysis step of "knowledge disc losable in databases" process, or KDD.

The term is a misnomer, because goal is extraction of patterns & knowledge from large amounts of data, not extraction (mining) of data itself. It also is a buzzword [&] is frequently applied to any form of large-scale data or information processing (collection, warehousing, analysis, & statistics) as well as any application of computer decision including artificial intelligence, machine learning, & business intelligence. book Data mining[1]: Special machine learning tools & techniques with Java was originally to be named just Practical machine learning, & term data mining was only added for marketing reasons. Often more general terms data analysis & analytics – or, when referring to real approach, artificial intelligence & machine learning – are more appropriate.

The actual data mining[5] task is arrangement or semi-automatic analysis of large quantities of data to extract previously unknown, interesting style such as groups of data records unusual records (anomaly detection), & dependencies.

[2]Clustering