IJARSCT

ISSN (Online) 2581-9429

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

Impact Factor 6.252

Volume 2, Issue 1, July 2022

Data Analysis using Data mining Techniques as K Means and Apriori Algorithms

Prof. Snehal Dhane
Assistant Professor
Sinhgad Institute of Business Administration And Research (Sibar), Kondhwa (Bk.), Pune, Maharashtra, India
snehal.dhanee@gmail.com

Abstract: *Data is major concern in today's digitized world. Data mining is an important method to increase efficiency, discover hidden (novel), useful, valid and understandable knowledge from a massive database. It is the process of analysing data from different perspective and summarizing the data into a useful identical format of information that can be used to predict future trends or performances. The ultimate goal of data mining is to recognize pattern full information and predictions. The most famous algorithms are K means and Apriori algorithms. Such algorithms were created to extract what is called nugget of knowledge from large sets of data. There are several different methodologies to approach this problem: classification, association rule, clustering etc.*

Keywords: Data mining, K means algorithm, apriori algorithm

I. INTRODUCTION

In today's world the abundant data are being collected, stored, and retrieved from databases from every click. Also it is not exact to keep growing the size of database year after year. Different organizations are rich in databases and it is not difficult to find the data. But there is invaluable information and knowledge "hidden" in databases which cannot be extracted without automatic techniques

II. K-MEANS AND APRIORI ALGORITHMS

K-Means and apriori algorithms are the popular techniques of data mining. This k-Means clustering algorithm help to us observe massive amount of data and organize the massive amount of data in homogeneous clusters. The Apriori algorithm is used for mining frequent item sets and devising association rules from a transactional database. The parameters "support" and "confidence" are used. Support refers to items' frequency of occurrence; confidence is a conditional probability. Items in a transaction form an item set. The aim of K-Means clustering algorithm is to calculate number of distance between the tested items and their related cancroids and tries to minimize theses values for each set of cancroids.
K= The number of items.
Mean =average number of items which grouped into similar clusters with the closest mean /centroid.
In other words we can say that the k-means algorithm is a simple iterative process to partition a known dataset into a user specified number of clusters.

2.1 K Means Algorithm

1. Place K points in the space represented by the objects that are being clustered. These points represent initial group centroids.
2. Assign each object to the group that has the closest centroid.
3. When all objects have been assigned, recalculate the positions of the K centroids.

Repeat Steps 2 and 3 until the centroids no longer move. This produces a separation of the objects into groups

DOI 10.48175/IJARSCT-5734

2.2 Apriori Algorithm

Apriori algorithm is a sequence of steps to be followed to find the most frequent itemset in the given database. This data mining technique follows the join and the prune steps iteratively until the most frequent itemset is achieved. A minimum support threshold is given in the problem or it is assumed by the user.

III. COMPARATIVE STUDY OF DIFFERENT DATA MINING TECHNIQUES

In this paper, the dataset was taken from one of the private restaurants databases and continued by processing data at the data pre-processing stage by transforming data. Then the pre-processing data is applied to approach 1 and approach 2. In approach 1 the data will be managed using apriori algorithm. While in approach 2 the data is grouped first using the k-means clustering algorithm then the grouping result data is applied to the apriori algorithm. Both of these approaches will each produce the Rule Mining Association. The results of Approach 1 and Approach 2 will be analyzed to determine the impact of grouping data on apriori algorithms based on computation time comparisons.

IV. CONCLUSION

Data mining techniques are used to create data models and test them as well. It is usually a framework like R studio or Tableau with a suite of programs to help build and test a data model. After comparing the apriori algorithm and the combination of the K-means + Apriori algorithm on the dataset then the computation time is calculated so that the results of the combination of the K-Means algorithm and the Apriori algorithm are more detailed and complete when compared with the results obtained by applying the Apriori algorithm only.

REFERENCES

[1]. A Review: Comparative Study of Various Clustering Techniques in Data Mining Aastha Joshi Student of Masters of Technology, Department of Computer Science and Engineering Sri Guru Granth Sahib World University, Fatehgarh Sahib, Punjab, India, Rajneet Kaur Assistant Professor, Department of Computer Science and Engineering, Sri Guru Granth Sahib World University, Fatehgarh Sahib, Punjab, India,

[2]. http://www.computerscijournal.org/vol8no1/a-comparative-study-of-classification-techniques-in-data-mining-algorithms/

DOI 10.48175/IJARSCT-5734