Scientific Journal of Impact Factor (SJIF): 4.14

e-ISSN (O): 2348-4470
p-ISSN (P): 2348-6406

International Journal of Advance Engineering and Research Development

Volume 3, Issue 12, December -2016

Distributed Clustering Technique of Clustering the Data

Data Mining

Abhishek Masiwal[1]

[1]*Computer Science and Engineering Department, Inderprastha Engineering College,*

Abstract – *Today data mining is very important to group similar data together. In this paper, we have given various methods to categorize similar data. There are various techniques to find similarity among various attributes of data and to cluster them all together.*

Keywords- *Data Mining*

I. INTRODUCTION

The main function of data mining is to apply various methods and algorithms to find similar patterns among the data. It focuses on discovering new techniques and tools to mine the data. This helps us in grouping of the similar data showing same patterns and finally removing the irrelevant data.

So, data mining can be implemented anywhere so as to obtain similar characteristics among the data.

II. SCOPE

Here, with the help of data mining the attributes of student's performance can be clustered and then these clusters defined the performance of students. The clusters helps us in identifying the performance of the students using the data of students. This helps us to identify the students with low academic results and how their performance can be improved using various others methods.

III. RELATED WORK

Today data mining is widely used in the field of education like colleges, universities, various institutions etc. so as to mine the important data and for the purpose of studying. And today the process of doing data mining is widely recognized and thus, gaining popularity.

Hoe [1] used different analyzing techniques to find the patterns among the data of students used for data mining.
Suchitra [2] used various algorithms for predicting the performances of the students.

The mining of the student's data used various artificial neural networks to find the similarity between the student's data so as to group them together.

J. James Manharans [3] used the algorithm of K-Means Clustering to identify the performances of the students. By using this algorithm we can more efficient results measuring the performances of the students. I used various methods to find out distinguished patterns among the student's scores.

IV. PROBLEM

To identify the performance of student's and to improve their performances. Here, improving the performance is the major challenge for us.
I applied the algorithm of clustering on the distributed environment.
Thus, the problem definition is to improve the efficiency and accuracy of the distributed environment.

V. WORK PROPOSED

A. COLLECTION OF DATA

I collected the data from the database system of my college Inderprastha Engineering College. The records consist of the academic records of students in the sessional as well as the semester exams. I also got the records of the student's attendance in their classes and also the percentage scored by them in their exams.

B. SELECTING AND TRANSFORMING DATA

We took only that data that which we have to mine. Some variable information were taken from the database and some were derived and is listed below in Table 1.

Table 1: Attributes selected for data mining

Attributes	Description	Values
MP	Marks of previous semester	1st >60% 2nd >45% and < 60% 3rd < 45%
ATTEN	Attendance	Low, Medium and High
PL	Performance in Lab	Good or Bad
MS	Marks in Sessional	Good, Average or Bad
ME	Marks in End Semester	1st >60% 2nd >45% and < 60% 3rd < 45%
GP	General Proficiency	Yes, No
AM	Average marks in all semester	Good, Bad or Average

Domain values of the variables defined are-

MP- MP means marks of previous semester of a student. It is divided into three class values: First > 60%, Second > 45% and 60% and Third < 45 %.

ATTEN- ATTEN means the attendance of the students and is divided into three classes namely: Low < 50%, Medium <75% but >50% or High >75%.

PL- PL means marks in the practical and is divided into two classes good or bad, where Good> 75% and Bad < 75%.

MS- MS means marks in the sessional and is split into three classes namely Good> 90%, Average< 90% and > 50% and bad < 50%.

ME- ME denotes the marks in the end semester and has three categories: First > 60%, Second > 45% but < 60% and Third < 45%.

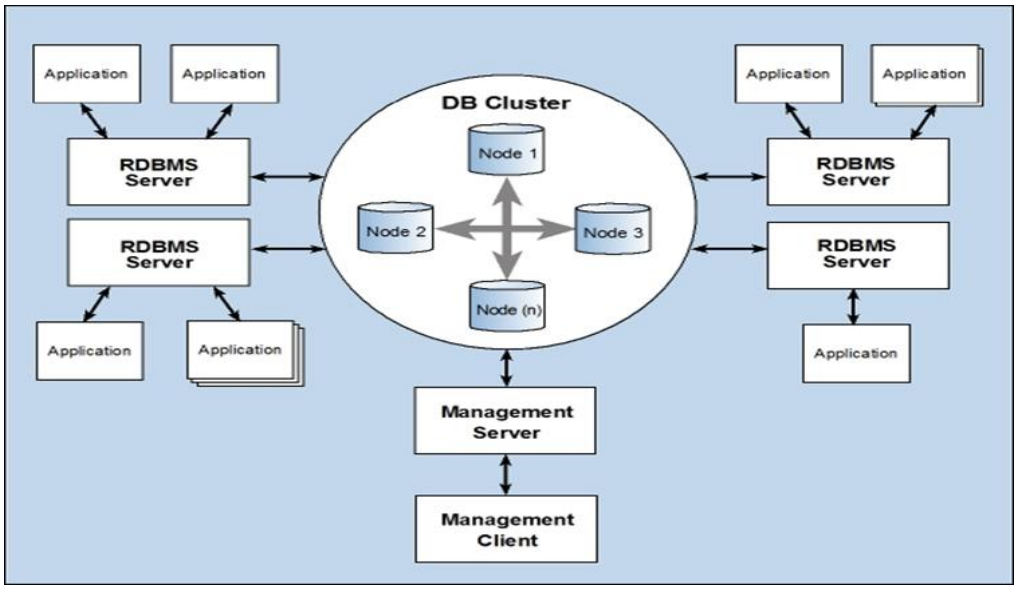
GP- GP denotes General Proficiency and has two classes Yes- where the student has been awarded the GP and No- where the student has not been awarded the GP.

AM- AM means Average Marks in all Semester and has two classes: Good- where the average >60% and Bad- where average < 60%.

VI. FRAMEWORK PROPOSED

We apply the following steps to our system-

The Clustering algorithm is applied to the local nodes as shown in the figure 1. A cluster can be called as a single object. The nodes represents the cluster which consist of the similar data groups clubbed together having similar characteristics.



1. Data Clustering Image

Steps for the Work are-

1. The user enter the data on the Application.
2. The server will generate a request according to the user.
3. The nodes represented in the figure are the individual clusters having similar data.
4. The cluster with the required data is sent to the RDBMS Server.
5. The server contains the address of all the nodes.
6. The server then send the obtained result to the user in this way-

$$X= N1+N2+N3$$

Where X= Final Result

N1- node1 result, N2- node2 result and N3- node3 result

VII. CONCLUSION

Here, we used the clustering technique to group the similar data having similar characteristics. The study will help us to identify the students with good academic records, the students with average or bad academic records. It will also help us to find whether the students are scoring more in end semesters or sessional.

We will get to know about the attendance of students who are regularly or not regularly attending their classes. This will help us in understanding the reason behind the low scores off the students and to identify the various ways in improving their performances on the academic front. It will also improve the interaction of teachers with their students. We can find out the reason why some students are scoring more and some are scoring less and can bridge the gap between them by appropriate measures.
This helped us in the grouping of similar data and thus the obtained results can be used for our future purposes also.

VIII. REFERENCES

[1] HOE, Ah mad, Tam chin Hooi; Shan Mug am m; Gunasegaram, S.S.; Cob, Z.C. Ramasamy" Analyzing students records to identify patterns of students "performance".

[2] James Manoharans, Dr. S. HariGanesh, M.LovelinPonnFelciah, A.K. ShafreenBanu"Discovering"Discovering Students" Academic Performance Based on GPA using K-Means Clustering Algorithm"IEEE2014.

[3] Suchitraborkar K. Rajeswari "Attributes Data Mining and Artificial Neural Networks" International Journal of Computer Applications (0975-8887) Volume 86- No 10, January 2014.

