Published by :
http://www.ijert.org

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 6 Issue 05, May - 2017

Evaluating Student's Performance using K-Means Clustering

Mr. Shashikant Pradip Borgavakar
Research Scholar: C.S.E
Swami Vivekanand College of Engineering
Indore, India

Mr. Amit Shrivastava
Asst. Professor: Computer Science & Engineering
Swami Vivekanand College of Engineering
Indore, India

***Abstract*— Data Clustering is the task of grouping a set of objects in such a way that objects in the same group are more similar to each other than to those in other groups. In this paper data clustering is used as k-means clustering to evaluate student performance. Evaluating student performance on basis of class test, mid test and final test. As we get cluster of student on this basis of student marks will help to reduce ratio of fail student. This information will help professor to student fail chance before final exam..**

Keywords- k-means, Database, academic performance etc.

INTRODUCTION

Data clustering is a process of extracting previously unknown, valid, positional useful and hidden patterns from large data sets (Connolly, 1999). The amount of data stored in educational databases is increasing rapidly. Clustering technique is most widely used technique for future prediction. The main goal of clustering is to partition students into homogeneous groups according to their characteristics and abilities (Kifaya, 2009). These applications can help both instructor and student to enhance the education quality. This study makes use of cluster analysis to segment students into groups according to their characteristics.

I. LITERATURE SURVEY

Research Paper	Improving the Accuracy and Efficiency of the k-means Clustering Algorithm	An Iterative Improved k-means Clustering	Refining Initial Points for K-Means Clustering	Comparision of various clustering algorithms
Problem being addressed	Lower accuracy and efficiency	Number of Iterations are Less	Estimate is fairly unstable due to elements of the tails appearing in the sample	Which clustering algorithm is best
Importance of the problem	algorithm requires a time complexity	Total number of iterations required by k-means and improved k-means is much larger	Importance of the problem of having a good initial points	Way of Process
Gap in the prior work	Accuracy and Efficiency is most complicated to reducing	Check multiple iterations	To finding Initial Points	Finding algorithm
Specific research questions or research objective	To Overcome the problem of Accuracy and Efficiency	This paper presented iterative improved k-means clustering algorithm that makes the k-means more efficient and produce good quality clusters	A fast and efficient algorithm for refining an initial starting point for a general class of clustering algorithms has been presented	data mining is that to discover the data and patterns and store it in an understandable form
Broad outline of how the author solved the problem	Using K-Means clustering Algorithm and The enhanced Method	Iteration improve k-means cluster algorithm	Using Clustering Cluster	Applied DBSCAN and OPTICS algorithms
Details of implementation of procedure	Phase 1 of the enhanced algorithm requires a time complexity of O(n2) for finding the initial centroids, as the maximum time required here is for computing the distances between each data point and all other data-points in the set	Dividing number of parts thencalculate centres and decide memership of patterns then rrepeat same steps	Results on Real Word Data	All clustering algorithm process and find
Key contribution of the paper claimed by the author.	define k centroids, one for each cluster	iterative improved k-means clustering algorithm	Clustering Clusters	K-Means clustering Algorithm

II. DATA CLUSTERING

Data Clustering is unsupervised and statistical data analysis technique. It is used to classify the same data into a homogeneous group. It is used to operate on a large data-set to discover hidden pattern and relationship helps to make decision quickly and efficiently. In a word, Cluster analysis is used to segment a large set of data into subsets called clusters. Each cluster is a collection of data objects that are similar to one another are placed within the same cluster but are dissimilar to objects in other clusters.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 6 Issue 05, May - 2017

III. CLUSTERING IN HIGHER EDUCATION

Education is an essential element for the progression and betterment of a country. Education makes a people perfect by which he/she can participate in any progressive work for the country. Education makes a country civilized and well-mannered. Clustering in higher education means it classifies the student by their academic performance. Lack of deep and enough knowledge in higher educational system may prevent system management to achieve quality objectives, data clustering methodology can help bridging this knowledge gaps in higher education system.

IV. PROPOSED MODEL

In university academic performance are measured by internal and external assessment. Internal assessments are class test marks, lab performance, assignment, quiz, attendance. External assessments are previous semester grade and final semester grade. So, by taking the internal assessment and previous exam grade and by using data clustering technique we can predict what will be the final grade of a student.

1. If prev-grade=high, quiz=good,assignment=complete, lab-performance=good ,class-test=good, attendance=regular and then final-grade=good
2. If prev-grade=average, quiz=good, assignment=incomplete lab-performance=good Class-test=average and atendance=regular then final-grade= average
3. If prev-grade=low, quiz=average, assignment=incomplete, lab-performance= poor mid-term=low and attendance=irregular then final-grade=low.

The proposed model try to identify the weak students before final exam in order to save them from serious harm. Teachers can take appropriate steps at right time to improve the performance of student in final exam.

V. K-MEANS CLUSTERING ALGORITHM

K-means is an old and widely used technique in clustering method. Here, k-means is applied to the processed data to get valuable information .The pseudo-code of k-means clustering is given below.

Step 1: Accept the number of clusters to group data into and the dataset to cluster as input values

Step 2: Initialize the first K clusters - Take first k instances or - Take Random sampling of k elements

Step 3: Calculate the arithmetic means of each cluster formed in the dataset.

Step 4: K-means assigns each record in the dataset to only one of the initial clusters - Each record is assigned to the nearest cluster using a measure of distance (e.g Euclidean distance).

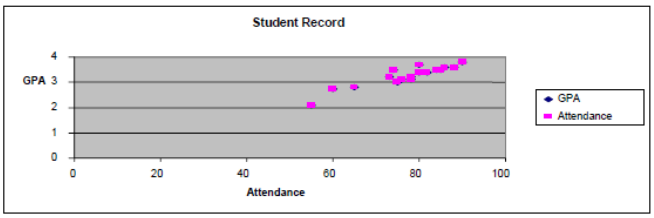
Step 5: K-means re-assigns each record in the dataset to the most similar cluster and re-calculates the arithmetic mean of all the clusters in the dataset.

Fig. 1 Generalized Pseudocode of Traditional k-means.

VI. RESULT AND DISCUSSION

The model produced following results:

Graph.1: Shows the relationship between GPA and Attendance ratio.



A. Data Arrangement in tables

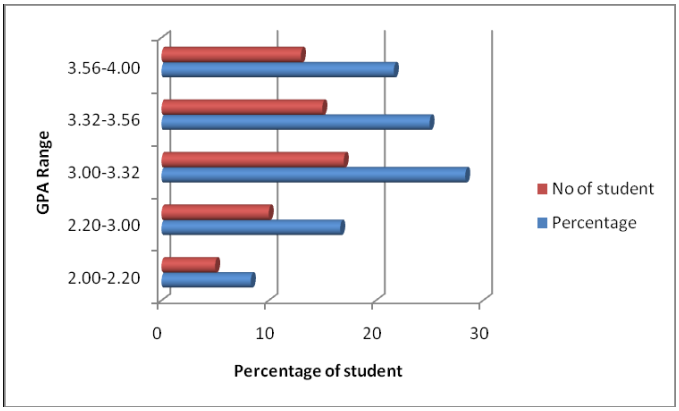
We grouped the students regarding their final grades in several ways 3 of which are: · Assign possible labels that are same as number of possible grades. · Group the students in three classes "High" "Medium" and "Low". · Categorized the students with one of two class labels "Passed" for grade above 2.20 and "Failed" for grade less than or equal to 2.20

Table 1

Class	GPA	No of student	Percentage
1	2.00-2.20 2.20-	5	8.33
2	3.00 3.00-3.32	10	16.67
3	3.32-3.56 3.56-	17	28.33
4	4.0	15	25
5		13	21.67

Here, I cluster student among their GPA, that means, from GPA 2.00- 2.20 we have 8.33% student. From 2.20-3.00 student percentage is 16.67%. From 3.00-3.32 we have 28.33%. From 3.32-3.56 percentage is 25% .The percentage is 21.67% between GPA 3.56-4.00. The graphical representation of GPA and the percentage of student's among the student is given below.

Graph 2: Number and percentage of students regarding to GPA

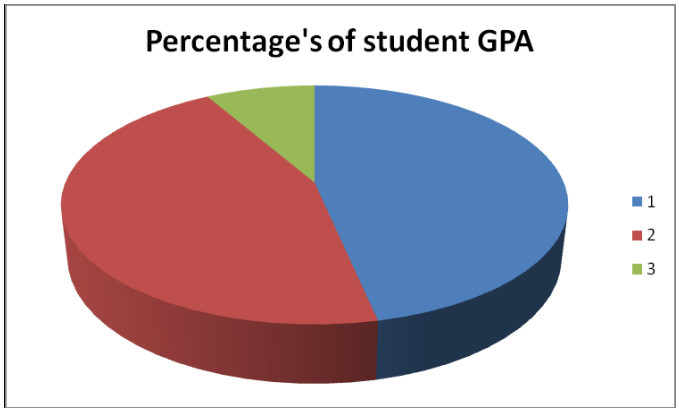


International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 6 Issue 05, May - 2017

Class	GPA	No of student	Percentage
High	>=3.50	28	46.67
Medium	2.20<=GPA<3.5	27	45
Low	<=2.20	5	8.33

After clustering the student, we group the student into three categories. One is High, second is Medium, and the last one is Low. Graphical representation of these three categories is given below:

Graph 3: Shows the percentage of students getting high, medium and low GPA



REFERENCES

[1] Alaa el-Halees (2009) Mining Students Data to Analyze e-Learning Behavior: A Case Study.
[2] Behrouz.et.al., (2003) Predicting Student Performance: An Application of Data Mining Methods With The Educational Web-Based System Lon-CAPA © 2003 IEEE, Boulder, CO.
[3] Connolly T., C. Begg and A. Strachan (1999) Database Systems: A Practical Approach to Design, Implementation, and Management (3rd Ed.). Harlow: Addison-Wesley.687
[4] Erdogan and Timor (2005) A data mining application in a student database. Journal of Aeronautic and Space Technologies July 2005 Volume 2 Number 2 (53-57)
[5] Galit.et.al (2007)Examining online learning processes based on log files analysis: a case study. Research, Refelection and Innovations in Integrating ICT in Education.
[6] Henrik (2001) Clustering as a Data Mining Method in a Web-based System for Thoracic Surgery: © 2001
[7] Han,J. and Kamber, M., (2006) "Data Mining: Concepts and Techniques", 2nd edition. The Morgan Kaufmann Series in Data Management Systems, Jim Gray, Series Editor.
[8] Kifaya(2009) Mining student evaluation using associative classification and clustering. Communications of the IBIMA vol. 11 IISN 1943-7765.
[9] ZhaoHui. Maclennan.J, (2005). Data Mining with SQL Server 2005 Wihely Publishing, Inc