

# Analysis of Student Result Using Clustering Techniques

P.Veeramuthu<sup>#1</sup>, Dr.R.Periyasamy<sup>#2</sup>, V.Sugasini<sup>#3</sup>

<sup>#PG and Research Department of Computer Science  
Nehru Memorial College,  
Puthanam Patti-621007</sup>

**Abstract**— One of the main concerns of higher educational system is evaluating and enhancing the educational organization. Educational data mining is an emerging discipline concern with developing methods for exploring the unique type of data that come from educational setting and using those methods to better understand students and the setting which they learn in. For better achieving this quality objective we think that these organizations need deep and enough knowledge to better do assessing, evaluating, planning and decision making process. Most of the required knowledge can be extracted from the historical and operational data that reside in the educational organization's databases. Data mining techniques are analysis tools that can be used to extract meaningful knowledge from large data sets. This paper is designed to present and justify the capabilities of data mining in the context of higher educational system. The main contribution of the paper is an analysis model that represents the data mining processes.

**Keywords**—Data Mining, Educational Data Mining, Clustering, Higher Education System and Knowledge Discovery Data.

## I. INTRODUCTION

Now a day's large quantities of data is being accumulated. Data mining is the process of discovering interesting knowledge from large amount of data stored in database, database warehouse or other information responsibility [5]. The educational system in India is currently facing several issues such as identifying students need, personalization of training and predicting quality of student interactions. Educational data mining (EDM) provides a set of techniques which can help educational system to overcome this issue in order to improve Learning experience of students as well as increase their profits [6]. Manual data analysis has been around for sometimes now, but it creates bottleneck for large data analysis. The transition won't occur automatically; in this case, we need the data mining. Data mining software allow user to analyzed data from different dimensions categorized it and summarized the relationship, identified during mining process [7]

This study aims to analyze how different factor affect a Students learning behavior and performance using Clustering for Data Mining Techniques.

## II. CATEGORIZATION OF DATA MINING SYSTEMS:

There are many data mining [DM] systems [3] [5] available which are specialized systems dedicated to a given data source or confined to limited DM functionalities, that are versatile and comprehensive. Some DM systems are according to the type of data handled such as spatial data,

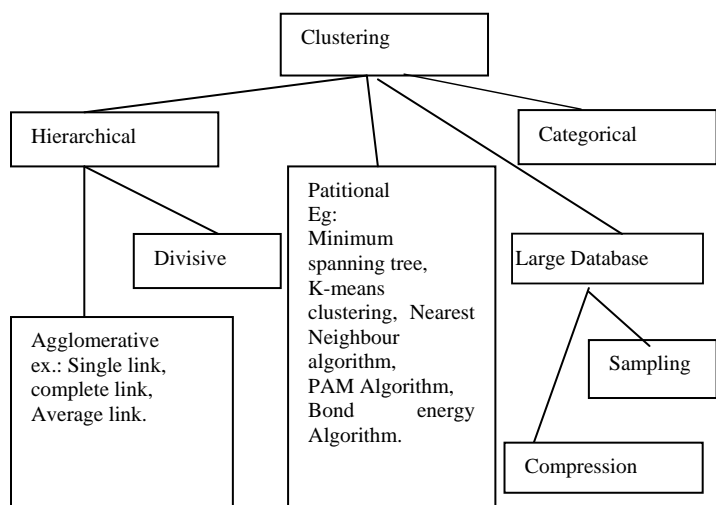
multimedia data, time-series data, text data, World Wide Web etc., some DM systems are based on the data model involved such as relational database, object-oriented database, data warehouse, transactional etc., Some DM systems are based on the kind of knowledge discovered or the functionalities such as characterization, discrimination, association, classification, clustering etc. Some systems tend to be comprehensive systems offering several DM functionalities together. DM systems employ and provide different techniques, classification according to the data analysis approach used such as machine learning, neural networks, genetic algorithms, statistics, visualization, data base or data warehouse-oriented etc. The classification can also take into account the degree of user interaction involved in the DM process such as query-driven systems, interactive exploratory systems, or autonomous systems.

## III. DATA MINING IN HIGHER EDUCATION SYSTEM:

Today the important challenge that higher education faces, is reaching a stage to facilitate the universities in having more efficient, effective and accurate educational processes. Data mining is considered as the most suited technology appropriate in giving additional insight into the lecturer, student, alumni, manager, and other educational staff behavior and acting as an active automated assistant in helping them for making better decisions on their educational activities. As discussed before, lack of deep and enough knowledge in higher educational system may prevent system management to achieve their quality objectives. Data mining technology can help bridging this knowledge gaps in higher educational system. Therefore the hidden patterns, association and anomalies, which are discovered by some data mining techniques, can be used to improve the effectiveness, efficiency and the speed of the processes. As a result, this improvement may bring a lot of advantages to the higher educational system such as maximizing educational system efficiency, decreasing student's drop-out rate, increasing student's promotion rate, increasing student's retention rate, increasing student's transition rate, increasing educational improvement ratio, increasing student's success, increasing student's learning outcome, and reducing the cost of system processes. In order to achieve the above quality improvement, we need a data mining system that can provide the needed knowledge and insights for the decision makers in the higher educational system. We have analyzed such a system and in the next section we will present and describe an analysis model for this system.

#### IV. CLUSTERING

Data mining software allows the user to analyze data from different dimensions categorize it and summarize the relationship. Identify during mining process. Data mining techniques are use to operate on large volume of data to discover hidden pattern and relationship helpful in decision making. Different data mining techniques are used in the field of education. Cluster analysis used to segment a large set of data into subsets. Each cluster is collection of data objects that are similar to another placed within the same cluster but dissimilar to objects in other cluster. Clustering is one of the basic techniques often used in analyzing data sets. This study makes use of cluster analysis to segment students into groups according to their characteristics. Clustering can be considered the most important unsupervised learning technique. Clustering and its



#### V. APPLICATION SOFTWARE:

In this study, data gathered from college students was analyzed using a data mining technique namely k-means clustering. The data set used in this study was obtained from department of Bachelor of computer Application (B.C.A.), N.M.C College, Trichy, in Nov-2013. The programming environment use for application was Visual studio 2008 for building data mining model it was compatible with SQL, in which data was stored.

##### 5.1. Preparation:

In this step data stored in different tables was joined in a single table after joining process errors were removed. For this model makes prediction about fail and pass ratio of student based on performance in exam. Model was developed using queries available in visual studio 2008.

Case 1 – If mid-term grade = Low, Internal Exam Grade = Low, Practical work = poor, Final Term = Average, then final grade = Low.

Case 2 – If mid-term grade = Average, Internal Exam Grade = Average, Practical work = Good, Final Term = Average, then final grade = Average.

Case 3 – If mid-term grade = High, Internal Exam Grade = Good, Practical work = High, Final Term = High, then final grade = Good.

#### 5.2 Implementation of mining model:

In this steps k-means clustering algorithm was applied to the proposed data and get valuable information, k- means is an old and most widely used by clustering algorithm.

#### 5.3 Algorithm 1 Basic K-means Algorithm:

1. Select K points as the initial centroids.
2. Repeat.
3. From K- cluster by assigning all points to the closest centroids.
4. Recomputed the centroid of each cluster.

#### VI. RESULTS

We grouped the students regarding their final grades in three ways

1. Assign possible labels that are the same as number of possible grades
2. Group the students in three classes “High”, “Medium”, “Low”.
3. Categorized the students with one of two class labels “Passed” for marks greater than or equal to 40 and “Failed” for marks less than 40

**Table: 1** Number and percentage of students regarding to class obtained

CLASS	MARKS	NO. OF STUDENTS	PERCENTAGE
S-GRADE	90-100	2	5%
A -GRADE	80-89	3	7%
B-GRADE	70-79	9	23%
C-GRADE	60-69	10	26%
D-GRADE	50-59	2	5%
E-GRADE	40-49	2	5%
U-GRADE	< 40	10	26%

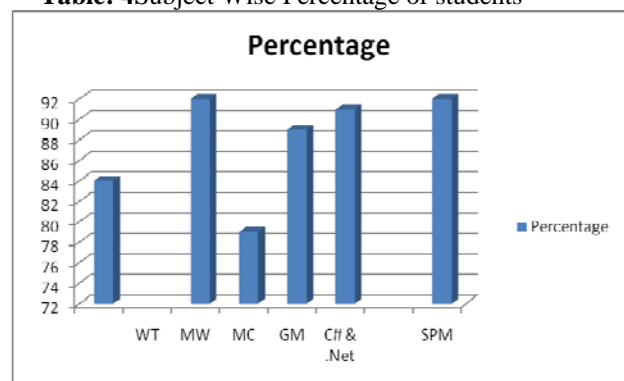
**Table:2** Number Of Students Regarding Class

CATEGORIES	NO.OF.STUDENTS
HIGH	22
MEDIUM	12
LOW	4

**Table: 3** Percentage of students

Class	Marks	No. Of.Students	Percentage
Passed	40 <= Percentage	28	73%
Failed	40 > Percentage	10	26%

**Table: 4** Subject Wise Percentage of students



## VII. CONCLUSION:

This paper has been an effort in providing the motivation toward advancing the traditional educational process via data mining technology. The model is also presented as a guideline for higher educational system to improve their decision-making processes. It can be used to analyze the existing work, identifying existing gaps and further works. The researchers may use the model to identify the existing area of research in the field of data mining in higher educational system. In this study we make a use of data mining process in a student's database using K-means clustering algorithm to predict students result. We hope that the information generated after the implementation of data mining technique may be helpful for an instructor as well as for students.

## VIII. FUTURE WORK:

For future work we redefine our techniques in order to get more valuable and accurate outputs useful for instructors to improve the students learning outcomes. Some different software's may be utilize while at the sometimes various factors will be used.

## REFERENCE:

- [1]. Shaeela Ayesha,(2010),data mining model for higher education system, J. of scientific research, vol -43, pp24-29.
- [2]. C. Romero, S. Ventura "Educational data Mining: A Survey from 1995 to 2005", Expert Systems with Applications (33), pp. 135-146, 2007
- [3]. Shaeela Ayesha, Tasleem Mustafa, Ahsan Raza Sattar, M. Inayat Khan, "Data Mining Model for Higher Education System", European Journal of Scientific Research, Vol.43, No.1, pp.24-29, 2010
- [4]. K. H. Rashan, Anushka Peiris, "Data Mining Applications in the Education Sector", MSIT, Carnegie Mellon University, retrieved on 28/01/2011
- [5]. Han Jiawei, Micheline Kamber, Data Mining: Concepts and Technique. Morgan Kaufmann Publishers,2000
- [6]. W.M.R. Tissera, R.I. Athauda, H. C. Fernando "Discovery of Strongly Related Subjects in the Undergraduate Syllabi using Data Mining",IEEE International Conference on Information Acquisition, 2006
- [7]. Sri Lanka Institute of Information Technology, <http://www.sliit.lk/> , retrieved on 28/02/2011
- [8]. Sun Hongjie, "Research on Student Learning Result System based on Data Mining", IJCSNS International Journal of Computer Science and Network Security, Vol.10, No. 4, April 2010
- [9]. Academy Connection – Training Resources In html, <http://www.cisco.com/web/learning/netacad/index>, December 28th, 2005.
- [10]. Wayne Smith, "Applying Data Mining to Scheduling Courses at a University", Communications of the Association for Information Systems, Vol. 16, Article 23, 2005
- [11]. Firdhous, M. F. M. (2010). Automating Legal Research through Data Mining. International Journal of Advanced Computer Science and Applications - IJACSA, 1(6), 9-16.
- [12]. Jadhav, R. J. (2011). Churn Prediction in Telecommunication Using Data Mining Technology. International Journal of Advanced Computer Science and Applications - IJACSA, 2(2), 17-19.