# "Importance of Data Mining in Higher Education System"

[1]Bhise R.B., [2]Thorat S.S., [3]Supekar A.K.
*[1,3]Department of Physics, B. J. College, Ale, (Pune ), MS.*
*[2]Department of BCA, B. J. College, Ale, (Pune ), MS.*

***Abstract:*** *Data Mining is used to extract meaningful information and to developed significant relationship among variables stored in large data warehouse [1]. "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" as defined by the educational data mining community [2]. Education is an essential element for the progress of country. Mining in educational environment is called educational data mining. It is concerned with developing new methods to discover knowledge from educational database [3]. Educational data mining provides a set of techniques, which can help the educational system to overcome these issues [4]. The objective of this research is to introduce educational data mining by describing step by step process using technique of K-means (Clustering Methods). The student evaluation factors like mid-term and final exam assignment are studied. This study will help the teacher to reduced drop-out ratio to a significant level and improve the performance of students.*
***Key Words:*** *Data Mining, Educational Data Mining, K-Means, Clustering, Higher Education System.*

## 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 academic career using K-Means (Clustering) in an educational institute.

## II. Data mining in higher education system:

Education is an essential element for the betterment and progress of a country. It enables the people of a country civilized and well mannered. Mining in educational environment is called Educational Data mining, concern with developing new methods to discover knowledge from educational database in order to analyze student's trends and behaviors towards education. Lack of deep and enough knowledge in higher educational system may prevents system management to achieve quality objectives, data mining methodology can help bridging this knowledge gaps in higher education system.
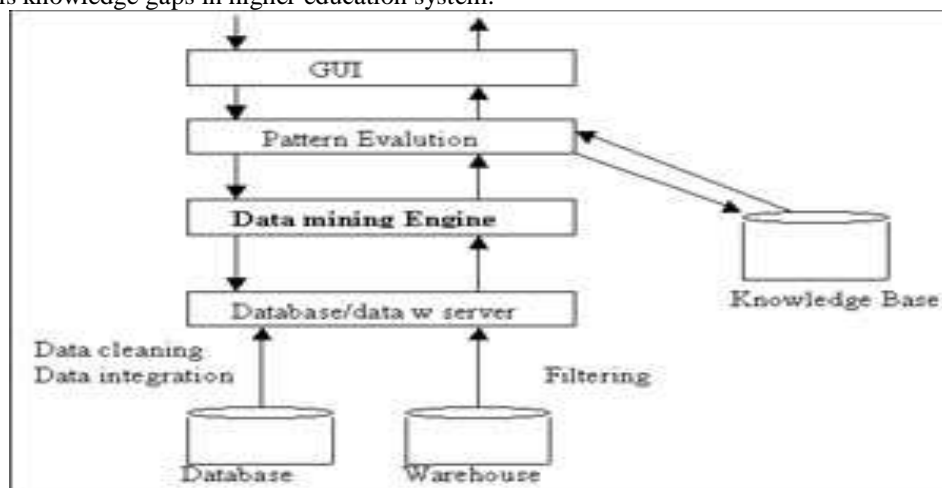
### III. K-means (Clustering):

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 classification are shown in fig.
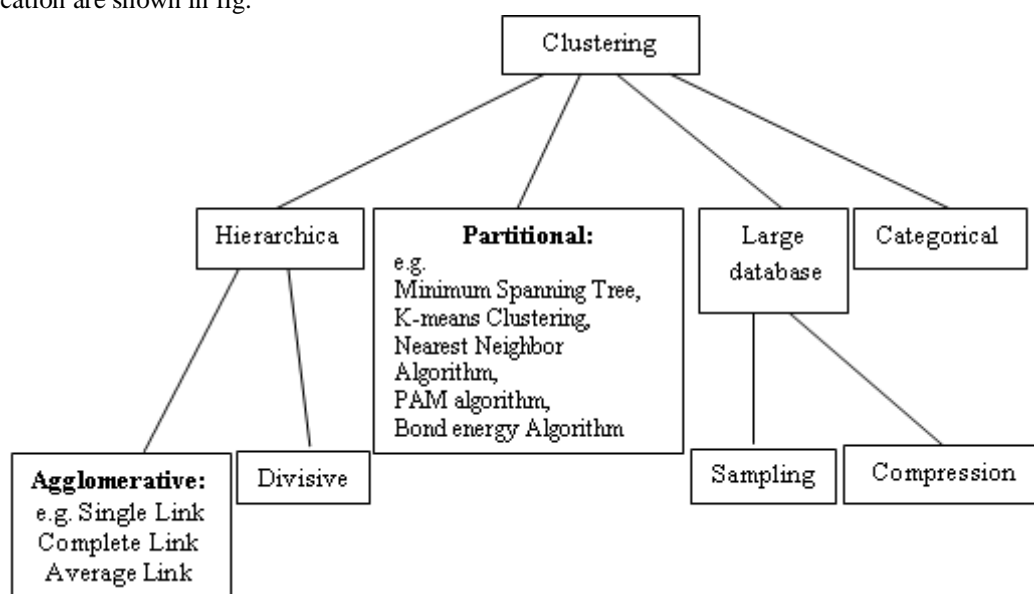


In educational data mining clustering has been used to group the student according their behavior e.g. clustering can be used to distinguished claver student from dull student according to their performance in exam.

### IV. 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.), B.J. College, Ale, in Nov-2012.

The programming environment use for application was Visual studio 2008 for building data mining model it was compatible with MYSQL, in which data was stored.

**4.1. 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 DMX 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.

**4.2  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 clustering algorithm by MacQueen in 1967

**4.3 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.

5.  **Until** The centroids don't change.

# V.     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: 5.1 Number and percentage of students regarding to class obtained

| Class | Marks | No. of Students | Percentage |
|---|---|---|---|
| Fail | > 40 | 6 | 5 |
| Pass Class | 40<percentage<50 | 14 | 11.67 |
| Second Class | 50<percentage<55 | 23 | 19.17 |
| Higher Second Class | 55<percentage<60 | 17 | 14.17 |
| First class | 60<percentage<70 | 47 | 39.16 |
| First class with distinction | 70 <= | 13 | 10.83 |

Table: 5.2 Number of students regarding class

| Class | No. of Students |
|---|---|
| High | 13 |
| Medium | 87 |
| Low | 20 |

Table: 5.3 Percentage of students

| Class | Marks | No. of Students | Percentage |
|---|---|---|---|
| Passed | > = 40 | 114 | 95 |
| Failed | < 40 | 6 | 5 |

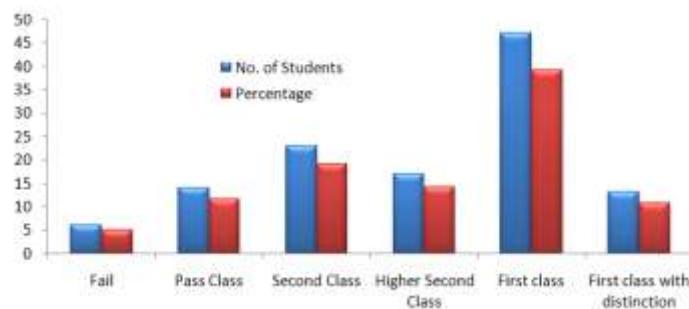The Graphical Presentation of Result is shows as follows



Fig. 5.1 Number and  percentage of students regarding to class obtained
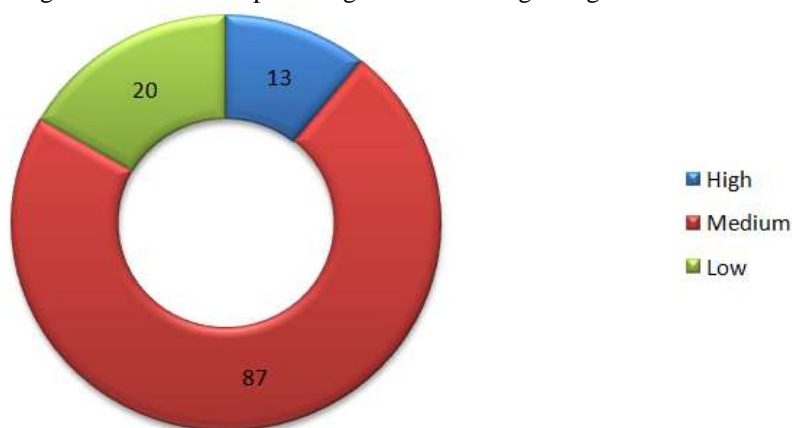


Fig.5.2 Percentage of students

# VI.     Conclusion:

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 a instructor as well as for students.

**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]   Sunita B. Aher,(2011), data mining in education system using WEKA, International J. of Computer Applications, pp20-25.
[3]   Alaa el- Halees,(2009), Mining Students Data to Analyze e-learning behavior. A case study.
[4]   Connolly T., C. Begg et al,(1999) Database System: A practical approach to design, Implementation and management( 3 rd edition), Harlow;  Addison-Wesley,687.
[5]   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)
[6]   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.
[7]   ZhaoHui. Maclennan.J, (2005). Data Mining with SQL Server 2005 Wihely Publishing, Inc.