Data Mining is used in Education System

B.Venkataratnam[1*], G. Sravanthi[2], C. Deepa[3]

[1,2,3]Department of Computers, Avinash Degree College, Osmania University, Hyderabad, India

Corresponding Author: venkatarathnam@gmail.com , Tel.: +91-9177421480

Abstract—Data mining is the process of sorting through large datasets to identify patterns and establish the relationship to solve problems through data analysis. Data mining tools allows enterprises to predict future trends. One of the prominent applications of data mining tools is in the field of education. Educational data mining, also known as EDM is an emerging, multidisciplinary area of research, concerned with studying and analyzing data from various educational databases. It helps in extracting meaningful information from large repositories using various data mining methods to enable data-driven decision making for improving the current educational practice. One of the biggest challenges faced by higher education today is predicting the paths of students. Data mining is a better tool to predict the results of the student. Educational Data Mining (EDM) helps in a big way to answer the issues of predictions and profiling of not only the students but other stake holders of education sectors. Thispaper focus on using various data mining tools like Association, Clustering, and Decision tree to analyze the large volume of data stored in the University databases to track the academic results of the students.

Keywords— Data Mining, EDM,Analyze, Decision, Clustering

I. INTRODUCTION

Educational data mining (EDM) describes a research field concerned with the applications of data mining, machine learning and statistics to information generated from educational settings (eg. Universities and intelligent tutoring systems). It tends to focus on developing new tools for discovering patterns in data. These patterns are generally about the micro concepts involved in learning.EDM generally emphasizes reducing learning into small components that can be analyzed and then influenced by software that adapts to the students. Students data collected during admissions are being explored to develop predictive models by applying educational data mining that classify data or find relationships. Decision tree gives possible outcomes and are used to predict students performance. Clustering finds group of objects so that objects that are in a cluster are more like each other to objects in another cluster. Predicting student's performance becomes easier with the help of these tools.

Researchers attempted to classify students into different clusters with dissimilar risks in exam failure, but also to detect with realistic accuracy what and how much the students know, in order to deduce specific learning gaps.In order to understand how and why data mining works, it's important to understand a few fundamental concepts. First, data mining relies on four essential methods: Classification, categorization, estimation, and visualization.Classification identifies associations and clusters, and separates subjects under study. Categorization uses rule induction algorithms to handle categorical outcomes, such as "persist" or "dropout," and "transfer" or "stay". Estimation includes predictive functions or likelihood and deals with continuous outcome variables, such as GPA and salary level. Visualization uses interactive graphs to demonstrate mathematically induced rules and scores, and is far more sophisticated than pie or bar charts. Visualization is used primarily to depict three-dimensional geographic locations of mathematical coordinates. Higher education institutions can use classification, for a comprehensive analysis of student characteristics, or use estimation to predict the likelihood of a variety of outcomes, such as transferability, persistence, retention, and course success.

Phases of Data Mining: Data mining is an iterative process that typically involves the following phases:

- Problem definition
- Data exploration
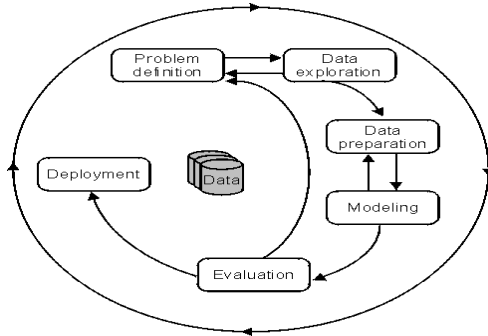- Data preparation
- Modeling
- Evaluation

Fig 1 Data Mining Process

II. RELATED WORK

Student'sperformance: There can be various measures to judge students performance. Several DM techniques have been used in this task, although association rule mining has been the most common.

A student from particular branch is assured for recommendations of both classification and clusteringtechniques. This process needs an educational dataset to initiate the task. This framework is used torepresent the student academic and non-academic attributes in higher education to hold completeinformation about students.In this experiment we collected data of some students from Osmania University. In the first step we clean and integratedata. For our problem we chosen eight attributes these were converted into its equivalent values which are given belowin the table.

Table 1: Selected Attributes

S.No.	Attributes	Description
1	Name	Vamsi, Vishnavi, Others
2	Gender	Male, Female
3	Attendance	Good, average, below average
4	I year	Promoted, detained
5	II year	Promoted, detained
6	III year	Promoted, detained
7	Final Result	Distinction, Pass, Fail

III. METHODOLAOGY

A. Classification Phase
To discover an efficient classifier in the educational dataset the classification algorithm is used. Therecommended course must be viewed for the student is the main role of classifier. The current level failed student records are removed in this classification phase. The different classification algorithms were applied by using this training dataset with course attribute to improve the student performance.

B. Clustering Phase
Based on marks resemblance the student records are partition into number of clusters by applyingclustering algorithm on educational dataset. The number of clusters is decided and previous gradeattributes are removed in this phase. To identify the clusters and distribution of the each course percentage is identified using K-means algorithm.

C. Request an Output from the System
This phase will describe the procedure about how the data will be entered by the new student in thesystem. Here the system will read the data and check Whether the data will be valid for the process ornot. The classification phase will predict the department with certain norms and regulations whichis already declared

IV.RESULTS AND DISCUSSION

The analysis of results in a particular institution(i.e. Universities, affiliated colleges) using excel sheets is time consuming process, that needs a deep understanding of statistical function in excel. This process needs large amount of time to finish the result analysis and to generate thee charts. This section represents the results generated from the study of EDM. The attributes were used according to their result percentage. The result was evaluated base on decision tree. From the decision tree we are easily idntify thee weak sstudents and whose chances of fail or maximum. After identifying weak sstudentss we can work hard on that studnt to minimize the failure result and we can imporve overall result and performance of the student.

Table 2: Result Analysis

| S.No	Name	Gender	Attendance	I Year	II Year	III Year	Result
1	S VAISHANAVI	Female	Good	Promoted	Promoted	Promoted	Distinction
2	SHAIK TARANNUM	Female	Good	Promoted	Promoted	Promoted	Pass
3	KAKI TEJASWI	Female	Good	Promoted	Promoted	Promoted	Pass
4	SHAIK SIMRAN	Female	Average	Promoted	Promoted	Promoted	Distinction
5	MOHD AZEEM ALI	Male	Good	Detained	Promoted	Promoted	Pass
6	GAUTHAMI	Female	Below Average	Promoted	Promoted	Promoted	Pass
7	ERRA NAVANEETHBHATT	Male	Good	Promoted	Promoted	Promoted	Distinction
8	CHITTIPOLU SRAVANI	Female	Average	Promoted	Promoted	Promoted	Pass
9	SRILODHA SRUTHI	Female	Average	Promoted	Promoted	Promoted	Pass
10	KUNCHAM SAI VARSHA	Female	Good	Promoted	Promoted	Promoted	Distinction
11	KARNATI PRIYANKA	Female	Good	Promoted	Detained	Promoted	Pass
12	AAMINA AKHTER	Female	Below Average	Promoted	Promoted	Promoted	Pass
13	SHAGA SANDEEP	Male	Below Average	Promoted	Promoted	Detained	Distinction
14	DUVVURI SIVANI	Female	Average	Promoted	Promoted	Promoted	Pass
15	V VAISHNAVI	Female	Good	Promoted	Promoted	Promoted	Pass
16	PARIYADA BHARATH	Male	Below Average	Promoted	Promoted	Promoted	Distinction
17	GUGGULLA KULLAI REDDY	Male	Below Average	Promoted	Promoted	Promoted	Pass
18	KATKURI SHIRISHA	Female	Good	Promoted	Promoted	Promoted	Fail
19	SAHITI M	Female	Good	Detained	Detained	Detained	Fail
20	JAKKA KARTHIK	Male	Good	Promoted	Promoted	Promoted	Pass
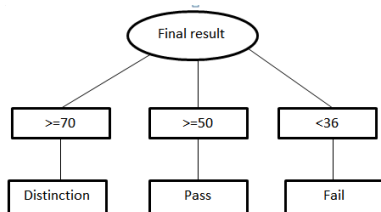
Fig 2 Decision Tree

V.CONCLUSION

This paper discusses about Educational Data Mining, it's applications and techniques to be used in EDM. The various data mining algorithms that can support education system via generating valuable information are discussed. Data mining techniques can be very helpful in areas like prediction of academic results, student's placement and to improve students' academic results in higher institutions. The Data mining techniques also used in analyzing the student's academic results in different aspects as well as predicting the reason. Once data retrieved from the relevant resources, the Decision Tree can be applied to categorize the data. The feature selection of attributes from large data set plays an important role in efficiency of result analysis using Decision Tree.

VI.REFERNCES

[1]. Brijesh Kumar Baradway and saurabhPal,"Mining Educational Data to Analyze Students Performance", Int. Journal of Advances in Computer Science and Applications, Vol. 2(6) ,pp. 63- 69, 2011.
[2]. Ajith P,,Tejaswi B, Sai MSS. Rule Mining Framework for Students Performance Evaluation. International Journal of Soft Computing and Engineering, Vol.2(6), 2013.
[3]. GalitShmueli, Nitin R. Patel, Peter C. Bruce, Data Mining In Excel.
[4]. Amritpal Kaur, Use of Data Mining in Education Sector.
[5]. Dr. MohdMaqsood Ali, Roleof Data Mining in Educatin Sector.
[6]. Vasile Paul Bresfelean, Data Mining Applications in Higher Education and Academic Intelligence Management.
[7]. B.Namratha, Nitesh Sharma. Educational Data Mining – Applications and Techniques.
[8]. Himanshi, Komal Kumar Bhatia, Prediction Model for Under-Graduating Student's Salary Using Data Mining Techniques. IJSRNSC Vol.6, Issue-2, April.2018.
[9]. Ajay Vrma, Y.S. Chouhan. Recent Methodologies for Improving and Evaluating Academic Performance. ISROSET Vol.3, Issue.2

Authors Profile

Mr B Venkataratnam completed MCA from Kakatiya university, pursuing Phd from osmania university.Currently working as Associate professor in Avinash degree college.

Ms G Sravanthi completed MTech from JNTUH,MCA from Osmania university. Currently working as Assistant professor in Avinash degree college.

Ms C Deepa completed MCA from Osmania university. Currently working as Associate professor in Avinash degree college.