

## DATA MINING IN EDUCATIONAL SECTOR

---

**Prof. Sandhya Pandey**  
S.I.A. College of Higher Education, Mumbai

---

### ABSTRACT

*Data Mining refers to extracting knowledge from large amount of data. It is a powerful new technology with great potential. Data mining tools predict future trends and behaviours. For promotion of superior education, EDM (Educational Data Mining) plays important role. EDM describes a research field concerned with the application of data mining and statistics to generate useful information for educational field.*

*In the last decade, number of Higher Education Institutions grown rapidly in India. These Institutions focused on the strength of students. Education sector has a lot of data that can produce valuable and relevant information. This knowledge can be used to increase the quality of education. This paper is focused on Educational Data Mining in Higher Education for improvement of student's performance.*

**Keywords:** EDM, Higher Education, Students Performance

### OBJECTIVES

Importance of Educational Data Mining in Higher Education Student's Improvement.

#### **Methodology**

The secondary sources of data and related study were used.

### INTRODUCTION

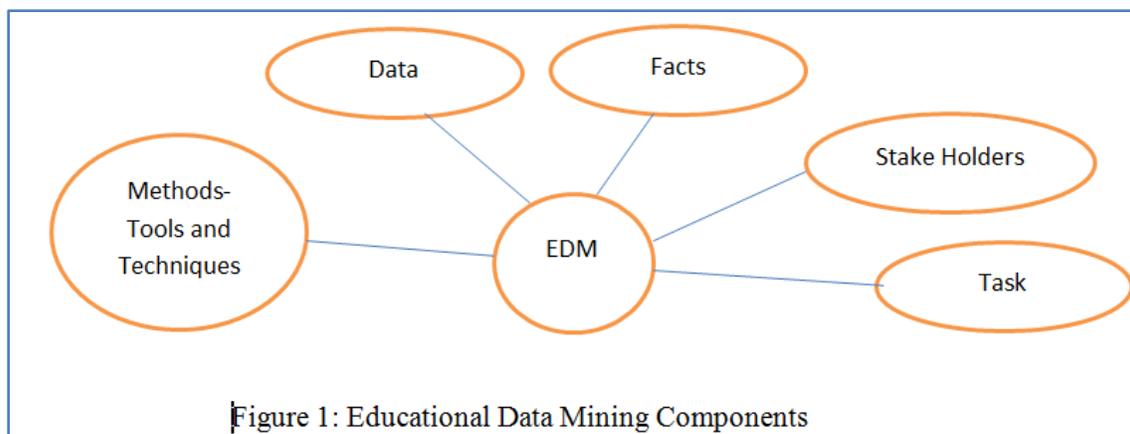
Data mining is a type of sorting technique which is used to extract hidden information from large databases. Data mining concepts and methods can be applied in various fields like marketing, medicine, real estate, customer relationship management, engineering, web mining, etc. Education field is very important area in terms of any country's growth and development.

Educational Data Mining (EDM) is a new emerging technique that can be applied on the data related to the field of education. Educational Data Mining in Higher Education can be used to improve graduate student's performance, and overcome the problem of low grades of graduate students. To achieve this, data could be collected with different sources, pre-processing can be done and data mining methods can be applied to prepare results to improve performance of students.

The discovered knowledge can be used to offer a helpful and productive recommendations to the higher education institutes to enhance their decision making process, to improve student's academic performance and trim down failure rate.

#### **Components of Educational Data Mining**

Data mining in higher education is a recent exploring field and this area of research is achieving popularity because of its potentials to educational institutes. For implementation of educational data mining following components can be used:



### **Steps for EDM:**

Using these components some steps can be followed for implementation of Educational data mining.

1. Problem Statement
2. Collecting data sets from educational environment
3. Data Classification and clustering
4. Data Mining
5. Result evaluation

#### **1. Problem Statement:**

Student's educational performance in higher education plays vital role in any institution. It is a continual process for formation of Vision and Mission of an institution. For this purpose Educational Data Mining in Higher Educational Institute can play important role and Institutions can track some steps for their student's improvements.

#### **2. Collecting data sets from educational environment:**

For educational data mining in any institution data sets can be collected for final year students from their previous records of all the areas related to their performance like academic details of grade 10<sup>th</sup>, 12<sup>th</sup> and previous semester details, extra-curricular activities details etc. This data collection can be creating using data warehouse technique.

In present day's educational system, a student's performance is determined by the internal assessment and end semester examination. The internal assessment is carried out by the teacher based upon student's performance in educational activities such as class test, seminar, assignments, and general proficiency. The end semester examination is one that is scored by the student in semester examination. Each student has to get minimum marks to pass a semester in internal as well as end semester examination.

**STUDENT RELATED DATA**

Variable	Description	Possible Values
10 <sup>th</sup>	Grade 10 <sup>th</sup> marks	>75%,>60%,>48%,<48
12 <sup>th</sup> / Diploma	Grade 12 <sup>th</sup> marks/Diploma	>75%,>60%,>48%,<48
CTG	Class Test Grade	Poor , Average, Good
ATT	Attendance	Poor , Average, Good
ASS	Assignments	Poor , Average, Good
FSG	First Semester Grade	O,A,B,C,D,E,F
SSG	Second Semester Grade	O,A,B,C,D,E,F
TSG	Third Semester Grade	O,A,B,C,D,E,F
LSG	Last Semester Grade	O,A,B,C,D,E,F
GP	General Proficiency	Good, Average, Poor

**3. Data Classification and clustering:**

Classification is the most commonly applied data mining technique, which employs a set of pre-classified examples to develop a model that can classify the of records at large. The data classification process involves learning and classification. Using this all student’s data can be classified. This technique is useful for success analysis with low, medium, high risk students.

Clustering can be used as identification of similar classes of objects. By using clustering techniques we can further identify and discover overall distribution pattern and correlations among data attributes. For example:

If Lower\_class\_grade=good and Higher\_class\_grade=good and ATT,ASS=Good then Topper  
 If Lower\_class\_grade=poor and Higher Class-grade=good and ATT,ASS=Averagethen Average  
 If Lower\_class\_grade=poor and Higher\_class\_grade=poor and ATT,ASS=Poorthen Below Average/need improvement.

**4. Data Mining Methodology:**

After data collection and Clustering any data mining tool can be applied. Firstly, minimum support is applied to find all frequent item sets in a database. Secondly, these frequent item sets and data assurance are used to form rules.

Finding all frequent item sets in a database is difficult since it involves searching all possible item sets (item combinations). After creation of data sets next step is forming a rule which is comparatively easy.

**5. Results Evaluation:**

After implementation of Data Mining Methodology results can be generated and based on the results improvement steps can be performed for graduate students.

## **CONCLUSION**

In this paper, task is used on student's previous performance to predict the students division. This type of study will help to the students and the teachers to improve the knowledge of the students further. It can also be useful to identify those students which needed special attention to reduce failure and taking appropriate action for the next semester examination.

Subsequently Educational Data Mining in Higher Education can be used to collect relevant and actionable data of learners and used for improvement of student's grade. Teachers and management person can also be beneficiaries to prepare strategies with student's bright future.

## **REFERENCES**

- [1] <http://arxiv.org/ftp/arxiv/papers/1201/1201.3417.pdf>
- [2] <http://airccse.org/journal/ijdms/papers/5313ijdms04.pdf>
- [3] [http://esjournals.org/journaloftechnology/archive/vol2no2/vol2no2\\_7.pdf](http://esjournals.org/journaloftechnology/archive/vol2no2/vol2no2_7.pdf)
- [4] [http://www.ijarcsse.com/docs/papers/Volume\\_3/8\\_August2013/V3I8-0105.pdf](http://www.ijarcsse.com/docs/papers/Volume_3/8_August2013/V3I8-0105.pdf)