

# Educational Data Mining to Analyze Students Performance – Concept Plan

Prajakta Akerkar<sup>1</sup>, Yashoda Bhat<sup>2</sup>, Mona Deshmukh<sup>3</sup>

<sup>1,2</sup>Student, VES Institute of Technology

<sup>3</sup> Professor, Dept. MCA, VES Institute of Technology, Maharashtra, India

\*\*\*

**Abstract** – Data mining is of great significance in the business world as it aids in decision making and gaining insights in the data which is stored by the organizations. Educational institutions store a lot of data related to the students which is retrieved as and when required by the management but such kind of retrieval does not provide any insights into the data and it is extremely tedious for any human to analyse and derive any decisions from that information. So the purpose of this paper is to suggest various techniques that can be used to analyse and derive insights from the existing data.

**Key Words:** Educational data mining (EDM), Student’s Performance, K-means, Naïve Bayesian, Clustering, Classification

## 1. INTRODUCTION

Computers can store, process and retrieve any type of information like text, numbers, images, etc. Data mining is a knowledge discovery process. Mining in the field of education is known as Educational Data Mining. Student’s attendance in class, family income, mother’s qualification, current knowledge and motivation greatly influence their performance in exams. [2]

Understanding and analyzing the reasons of poor performance is necessary because, that should be avoided by the institutions to have a good reputation ahead and not just poor performance the reasons for good performance should also be known so that it can be repeated ahead.

There are many data mining algorithms available which can be applied to the raw data to get the necessary results but the objective of this paper is to suggest an algorithm theoretically which can be applied perfectly to the data to get the results. Some of the classic EDM problems are stated as follows [7]:-

- Classification
- Clustering
- Frequency pattern mining
- Emerging pattern mining
- Visual analytics

- Predictive Modeling.

Major applications of EDM are :

- Developing concept maps
- Social network analysis
- Analysis and visualization of data
- Predicting student performance
- Recommendations for students, teachers and educational institutions
- Grouping students.

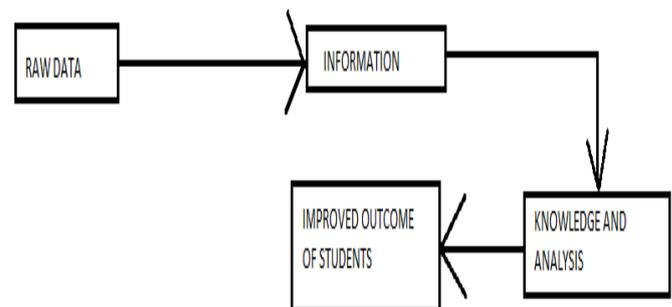


Fig -1: EDM Flow

## 2. METHODOLOGY

1. Identifying stakeholders

Considering all the people involved in education the stakeholders can be divided into three groups as follows:-

**1. Primary:** - These set of people are directly involved in teaching and learning.

Example – Teachers and Students

**2. Secondary:** - These set of people are indirectly involved in the process of teaching and learning they mostly contribute to the growth of the educational institution

Example – Alumni, Parents, Trustees

**Hybrid:** - These people are involved in the administrative process. They mostly do decision making for the institution.

Example – Non-teaching staff, educational planners.

### 3. Collecting data

Related student data can be collected from the following sources:

- Several Learning management systems (LMS) track students information as to when the students has accessed the learning object, how many times it has been accessed and the time spent on the learning object each time a student access it. [5]
- Intelligent tutoring system record data every time a student submits a solution to a particular problem. The time of submission, whether the answer is correct or not is captured by the ITS.
- Offline data: - Here the data is typically captured from the classroom interaction, student’s attendance, course information, grades.

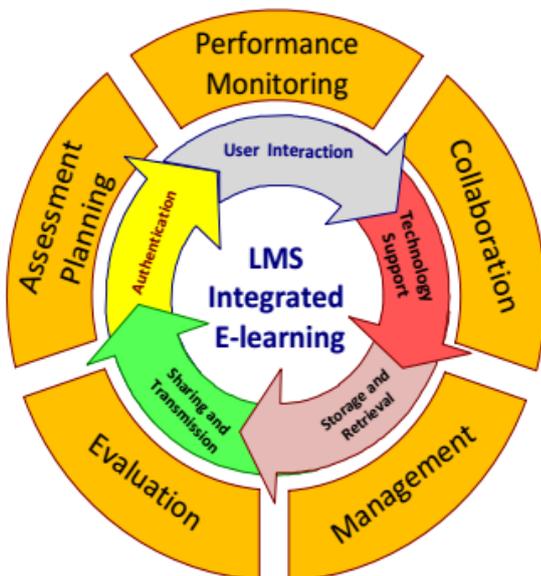


Fig -2 : Integrated Learning Management Systems [5]

In this step gather only those fields which are required for the EDM process these fields are called student related variables. The judgement parameters not only give an idea about the marks/grades gained by the student but also the overall personality of the student.

Following student related variables can be considered:-

Table -1: EDM Judgement Parameters

Sr No.	Judgement Parameters
1	Gender
2	Xth grade/%
3	XIIth grade/%
4	Family income
5	Father's qualification
6	Father's occupation
7	Mother's qualification

8	Mother's occupation
9	Siblings (if any)
10	Siblings qualification
11	Siblings occupation
12	Preferred time to study (Morning/Afternoon/Night)
13	Interested in higher studies
14	Caste
15	School type (Girls/Boys/Co-ed)
16	Do you have a cell phone
17	Are you active on social network?
18	Hobbies
19	Favorite subject
20	Where do you stay? (area)
21	Time spent on travelling everyday
22	Do you submit assignments on time
23	Attendance in class
24	Attentiveness in class
25	Do you make notes in class?
26	How many reference books do you refer for a subject?

Thus a survey can be conducted based upon the above variables which can be used to judge a student completely. To classify the students there have to be a predefined set of classes as mentioned below:-

- Poor
- Satisfactory
- Average
- Good
- Excellent

### 4. TECHNIQUES OF IMPLEMENTATION

#### A. Clustering Algorithm

Clustering means grouping data into groups of similar objects. It is significant in information retrieval and text mining, scientific data exploration, web analysis, and many more.

It is unsupervised and statistical data analysis technique. Cluster analysis is used to break down a large data set into small subsets called as clusters. Each cluster is a collection of similar objects.

Application of clustering to Educational Data Mining:-

K-means is the best clustering algorithm in data mining . It proposes to partition 'n' objects into 'k' clusters. The objective of this technique is to minimize the squared error function or total intra-cluster variance. Let  $X = \{x_1, x_2, x_3, \dots, x_n\}$  be the set of data points and  $V = \{v_1, v_2, v_3, \dots, v_n\}$  be the set of centers.

- 1) Randomly select “c” cluster centers.
- 2) Calculate the distance between each data point and cluster centers using the Euclidean formula.
- 3) Assign the data point to the cluster center whose distance from the cluster center is minimum of all the cluster centers.
- 4) Recalculate the new cluster center using :

$$v_i = (1/c_i) \sum_{j=1}^{c_i} x_j$$

Where, ‘ci’ is the number of data points in ith cluster.

- 5) Recalculate the distance between each data point and new obtained cluster centers.
- 6) If no data point was reassigned then stop, otherwise repeat from step (3). [4,7]

## B. Classification Algorithm

Classification is the form of data mining that defines models having important data classes. Usually decision tree classification is employed in this technique. In this test data sets are used to find the accuracy of the classification rules. If the accuracy is acceptable the rules can be applied to the new data tuples. [7]

Naïve Bayesian Algorithm is a classification technique based on Bayes Theorem with an assumption of independence among the predictors.

That means it assumes that the presence of a particular feature in a class is unrelated to the presence of any other feature.

This algorithm is easy to build and very useful in classifying large data sets.

Bayes theorem provides a way of calculating posterior probability  $P(c|x)$  from  $P(c)$ ,  $P(x)$  and  $P(x|c)$ . Following is the equation:-

$$P(c|x) = \frac{P(x|c) P(c)}{P(x)}$$

$$P(c|x) = P(x_1|c) \times P(x_2|c) \times \dots \times P(x_n|c) \times P(c)$$

This technique works in large student database to evaluate the results. [1]

## 4. CONCLUSION

This research paper given the major idea of creating a concept plan for every student and analyzing each and every student completely through the judgement parameters.

Thus this paper gives a theoretical concept of educational data mining using by using K-means clustering and Bayes algorithm to collect the data by conducting a survey, apply the algorithms and insights for every student. After the students are classified further action can be taken on each and every class of students to enhance their performance and provide more detailed guidance to them.

## 5. FUTURE WORK

Future enhancements includes implementing this work in XL Miner tool or KEEL.

## ACKNOWLEDGEMENT

This acknowledgment is a small effort to express my gratitude to all those who have assisted us during the course of preparing this paper.

We are greatly indebted to express immense pleasure and sense of gratitude towards our guide and mentor Prof. Mona Deshmukh for her constant support and valuable encouragement.

We express our heart-felt gratitude to Mr. M. Prabhanath Nair for his timely inputs.

## REFERENCES

- [1] Prediction of Students Performance using Educational Data Mining Ms.Tismy Devasia1, Ms.Vinushree T P2, Mr.Vinayak Hegde3 Department of Computer Science Amrita Vishwa Vidyapeeth University, Mysuru Campus.
- [2] Evolutionary Algorithm Based Rule(s) Generation for Personalized Courseware Construction in Educational Data Mining Sreenath. K1, Jeyakumar. G2 Department of Computer Science and Engineering, Amrita School of Engineering, Coimbatore Amrita Vishwa Vidyapeetham, Amrita University, India 1
- [3] Comparison of applications for educational data mining in Engineering Education Diego Buenaño Fernández Facultad de Ingenierías y Ciencias Agropecuarias Universidad de las Américas Quito, Ecuador Sergio Luján-Mora Department of Software and Computing Systems University of Alicante
- [4] A Systematic Review on Educational Data Mining Ashish Dutt, Maizatul Akmar Ismail, and Tutut Herawan.
- [5] Educational Data Mining and Big Data Framework for e-Learning Environment Prakash Kumar Udupi1, Nisha Sharma2, S K Jha3 Department of Computing, Middle East College, Muscat, NIMT, Kurukshetra, Haryana, India, AIIT, Amity University, Noida, India
- [6] Data ware house design for Educational Data Mining Oswaldo Moscoso-Zea1, Andres-Sampedro1, and Sergio Luján-Mora2 1Equinoctial Technological University, Faculty of Engineering, Quito, Ecuador 2University of Alicante,

Department of Software and Computing Systems, Alicante,  
Spain.

[7] Educational Data Mining –Applications and Techniques  
B.Namratha Assistant Professor, Department of Information  
Technology, Anurag Group of Institutions,RR Dist.,  
Hyderabad, Telangana, India Niteesha Sharma Assistant  
Professor, Department of Information Technology, Anurag  
Group of Institutions,RR Dist., Hyderabad, Telangana, India