International Journal of Electrical, Electronics and Computer Engineering **4**(2): 162-167(2015)

ISSN No. (Online): 2277-2626

Analyzing student learning performance using data classification: A Review

Ruchi Jain*, Kailash Patidar** and Megha Jain***

**Research Scholar, Sri Satya Sai Institute of Science & Technology, Bhopal, (MP), INDIA*

***HOD (CSE/IT) Sri Satya Sai Institute of Science & Technology, Bhopal, (MP), INDIA*

**** Asst. Professor, Sri Satya Sai Institute of Science & Technology, Bhopal, (MP), INDIA*

(Corresponding author: Ruchi Jain)

(Received 23 October, 2015 Accepted 12 December, 2015)

(Published by Research Trend, Website: www.researchtrend.net)

ABSTRACT: The performance of students in higher education is a most challenge task day by day in academic as well as in other curricular activities. Today's everybody can find the huge amount of data stored in the web technology, but they don't relies that what data is suitable? As they all know that internet technology is growing as much as faster, but the learning approach of students are not up to the mark. That type of database contains widely open or secret information to improve student performance. In this paper we have described various approaches of student learning and performance enhancement and we have tried to solve their learning performance approach by using well known entropy and gain based data mining technique to evaluation of performance. While the evaluating the students performance then there is typical task is database and its feature selection. So in this paper we have proposed and approach for feature selection by using highest gain of the particular attribute, after that classification process will be done.

Keywords: Classification, Educational Data mining, Navi bays, Entropy, Information Gain etc.

I. INTRODUCTION

In this paper, we will be analyzing role of data mining in education sector. This area has emerged as field of research and is termed as Educational data mining (also referred to as EDM).

Several nations are using EDM to enhance their education sector so that they can lead to produce qualified and educated citizens who can elevate their economy and value of their nations.

In our country various students are enrolled for various courses but do they deserve to be in that field or which stream or sector will be more appropriate for them? As for various degree courses different examinations are conducted to pursue that field but if they are unable to qualify, then which other stream should be suitable for them, must be recommended. For this purpose instead of conducting different examination for higher studies, the one general examination should be conducted and the result of this examination will go to decide as to which student should opt which field or stream.

The student's performance is a primary concern in education environment. To analyze the academic performance factors like personnel, social, psychological etc. will have to be considered. For analyzing the students' performance the attributes of the student having a higher academic performance will be compared with the student having lower academic performance.

EDM can be considered as data driven decision making practice which improves the current educational system and learning material. EDM applies the data mining techniques to the dataset derived from the educational system and provides the solution to the various educational questions. There exists relationship between different data types and end users of the educational environment to gain better understanding of student and their learning context. In EDM new data patterns are revealed and new models or algorithm are developed so that various issues and challenges are identified from large datasets using different data mining techniques. Large datasets results in huge repository of data on student learning information and teacher-student interactions. These data repositories are merged with theory for the computational approaches which transforms instructions to learners' need in order to provide better appreciation of students' learning abilities. Thus data flow in EDM can be summarized as follows:-

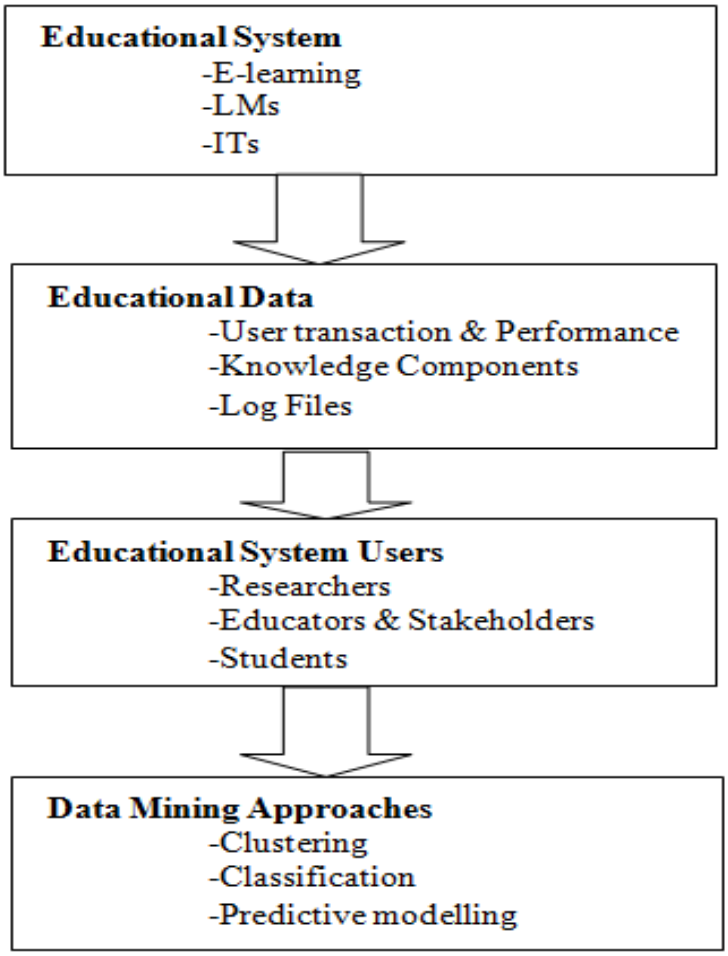

Fig.1. Application of Data mining in Education System [1].

Fig. 1 shows that, Initially there are various educational systems in the system viz e-learning, learning machines and when the system is chosen then the data is extracted which can be KDD, User transaction & Performance or log files. Now the extracted data can be used by the various end users :[1]

Learners:- They basically analyze the students and give them feedback on their work and enhance their learning performance.

Educators:- They generally analyze themselves to gain better understanding of their students' learning process and enhance their teaching performance.

Researchers:- They choose the best data mining algorithms to analyze various educational datasets.

Stake Holders:- Their task is to organize the institutional resources so that these resources can be optimally used.

This process distinctly reveals the role of end users in different ways. These end users achieves the EDM goals [1] which can be recognized as follows :-

1. Developing a Student Model by studying detailed information of a student such as knowledge, skills, motivation, meta-cognition, attitudes, experiences, satisfaction, learning process and styles and preferences.
2. Predicting students' future learning performance and outcomes based on data from course activities.
3. Analyzing the behavior of learner who guide the student properly.
4. Consulting the stake holders frequently.
5. Generating or improving the domain models that characterize the optimal instructional sequences.
6.studying the effects of different kinds pedagogical support that can be provided by learning software.
7. Implementation of scientific knowledge which is developed through the computational models and incorporates the model of the student, the domain and the software pedagogy.

II. APPLICATION OF DATA MINING IN EDUCATIONAL SYSTEMS

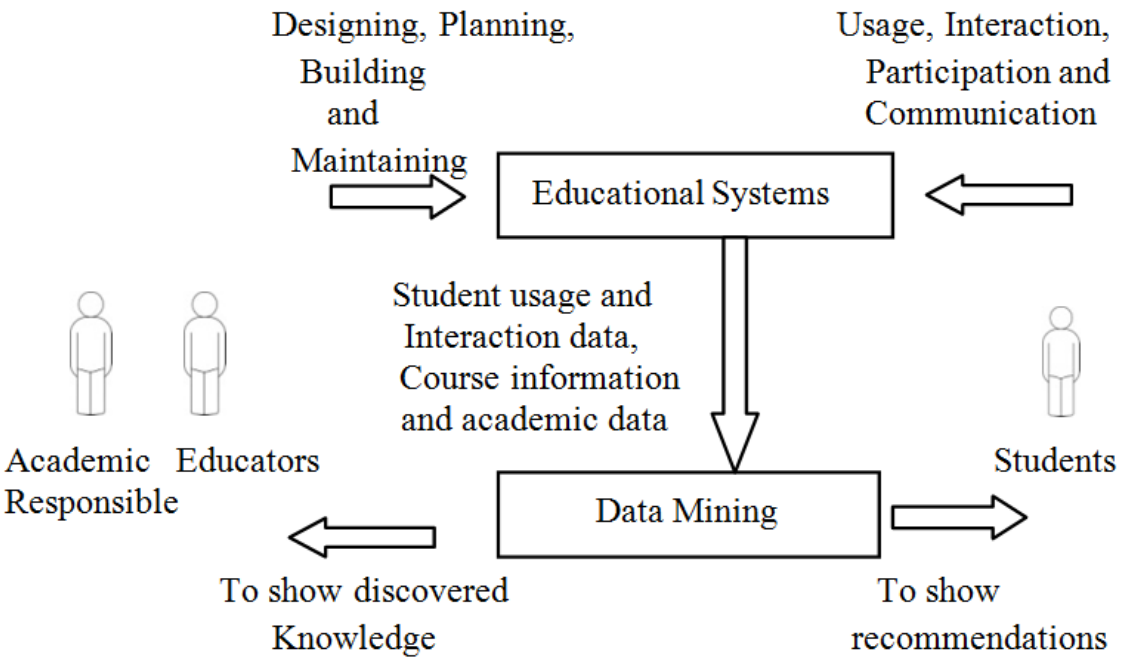

Fig. 2. Cycle of applying data mining in educational systems [2].

Figure 2, shows about various end users, on one end there are educators and academic responsible person who are for designing, planning, building and maintaining educational phenomenon's to the systems like traditional classrooms, e-learning systems, adaptive and intelligent web based educational systems and these systems are used by the students to use, interact, participate and communicate with this systems then the outcome of this system will generate some parameters for student usage, interaction data, course information and academic data which will undergo the process of either clustering, classification, outlier detection, association ,pattern matching, text mining and the final result for the educators will be discovered knowledge and there will be recommendations for the students.

III. MODELS FOR EDUCATIONAL DATA MINING

The data models can be of two types [3] :-

Predictive Models:- This model generates the prediction about values of data considering known results and other historical data. It predicts new values of the properties. Predictive model includes many different algorithms like Classification, Regression, Time Series Analysis, Prediction.

Descriptive Models:- This model explores patterns or relationships in data .It identifies the properties of data examined. Descriptive model includes Clustering, Summarization, Association rules, Sequence Discovery. We will be applying the association rules mining to predict the relationships among various attributes which will reveal the hidden patterns of student with good and bad performance.

Classification:- it is the process of finding a model (or function) that describes and distinguishes data classes or concepts, for the purpose of being able to use the model to predict the class of objects whose class label is unknown .It is a two step process:-

(a) Model Construction

(b) Model Usage

In this first step the model is built from the training set and in the next step model is used for classification.

IV. RESEARCH METHODOLOGY [4]

In this study we will be considering Naives Bayes classification in order to classify the student according to its characteristics considered. This theorem is very easy to apply and can be easily interpret. It can be readily applied to huge dataset.

Basically, it uses the knowledge of prior events to predict the future events. It deals with probability inference which is applied for decision making and inferential statistics. This process uses Baye's Theorem which provides the way to calculate posterior probability. The values of training dataset is known as predictors. This theorem demonstrates the independence assumptions between predictors.

The concept of conditional probability should be known, the conditional probability means that if the event has already occurred (which is known as prior probability) then the probability of future event can be determined (which is known as posterior probability).

The formula for calculating the posterior probability is :-

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

where

c is the target class and x is the predictor

$P(c|x)$ is the posterior probability of *class* (*target*) given *predictor* (*attribute*).

$P(c)$ is the prior probability of *class*.

$P(x|c)$ is the likelihood which is the probability of *predictor* given *class*.

$P(x)$ is the prior probability of *predictor*.

V. DATASET FOR THE EDM [5]

In this paper, certain attributes which may be used to predict the performance of student. Thus the dataset can be summarized as follows :-

Evaluation Criteria	Description	Expected Values
Sex	Gender of the student	{male, female}
Cat	Category of the student	{unreserved, OBC, SC, ST, Minority}
G_HS	High school grade	{A+,A,B+,B,C+,C,F}
G_SS	Higher school grade	{A+,A,B+,B,C+,C,F}
Med	Teaching Medium	{English, Hindi}
P_Col	Profile of College	{Good,Bad}
F_Qual	Father's Qualification	{illiterate,elemantry,secondary,UG,PG,Doctrate,NA}
M_Qual	Mother's Qualification	{illiterate,elemantry,secondary,UG,PG,Doctrate,NA}
F_Occ	Father's Occupation	{Business,Service,Agriculture,Retired,NA}
M_Occ	Mother's Occupation	{Business,Service,House wife(HW),Retired,NA}
Tutorial	Marks obtained in Tutorial Examination	{Good, Average,Poor}
Attendance %	Attendance of the student during the academic year	{Excellent,Very Good,Good,Poor}

VI. FACULTY PERFORMANCE DATASET [6]

Evaluation Criteria	Description	Expected Values
C_Qual	Current Qualification	{PG,Ph.D,Research Scholar}
Specialization	Specialization field	{IT, CS, Business, English, Engineering, Mathematics, Drafting, NA}
T_Exp	Teaching Experience	{0-1,1-5,5-10,10-above}
Uni_Type	Type of University	{Public,Private}
Indus_Exp	Experience in any Industry	{Yes,No}
Work_Exp	Working Experience	{Academic-only,Industry-only,Both}
R_Prev_Job	Rank/Title in Previous Job	{Non_teaching,Professor,Managerial,NA}
Sal_pref	Salary Preference	{<15000,15000-20000,20000-30000,30000-50000,>50000}
Perform_Prev_job	Performance in previous job	{Excellent,Very Good,Good,Poor}
Human_Relate	Human Relationship	{Good_Interpersonal, }

VII. EXPECTED OUTCOME [7]

The outcome of this study will describe significant and essential learning that learners' have gained. Once the model is generated, it is necessary to verify its accuracy. The accuracy of the model can be calculated on the basis of certain parameters like Recall, Accuracy, Precision, F-measure. These parameters can be estimated using confusion matrix which is explained as follows :-

	Predicted	
Actual	A:Hits	B:misses
	C:False Alarms	D:Correct Rejections

Recall is the proportion of positive cases that were correctly identified. Recall is also termed as Sensitivity or True Positive Rate (TPR). It can be expressed as

Recall (R) = $\frac{A}{A+B}$

Accuracy is the proportion of the total number of predictions that were correct. The following equation can determine accuracy :

Accuracy (AC) = $\frac{A+D}{A+B+C+D}$

Precision is the proportion of the predicted positive cases that were correct. It can be calculated using the equation,

Precision (P) = $\frac{A}{A+C}$

F-measure [8] computes some average of the information retrieval precisions and recall metrics. It can be calculated as harmonic mean of precision and recall

F-measure = $2 \times \frac{precision \times recall}{precision+recall}$

VIII. CONCLUSION

In this paper, they have discussed various approaches of educational data mining to improve learning of students and enhancement of their performance. They have also proposed an ideal approach for performance evolution of student using navi bays, before classification the initial process will be done where entropy will be calculated for each domain, after successful calculation of it, the gain will be calculated by standard formula.

On the basis of higher gain attribute selection process will be completed. The main aim of this research is that the enhancement of learning approach by using data mining technique as well as machine learning approach.

REFERENCES

[1]. Sedigah Abbasnasab Saradeh, Mohd. Rashid Mohd. Saad, Abdul Jalil Othman, Rosalam Che Me; "Enhancing Education Quality Using Educational Data " ; *Scholars Journal of Arts, Humanities and Social Sciences*.

[2]. Edin Osmanbegovic, Mirza Suljic (2012). "Data Mining Approach For Predicting Student Performance"; *Journal of Economics and Business*, Vol. **10**, Issue 1, May 2012.

[3]. Sayali Rajesh Suyal, Mohini, Mukund Mohod (2014). "Quality Improvisation of Student Performance Using Data Mining Techniques" ; *International Journal of Scientific and Research Publications*, Volume **4**, Issue 4, April 2014.

[4]. http://www.saedsayad.com/naive_bayesian.htm

[5]. Ajay Kumar Pal, Saurabh Pal (2013). "Data Mining Techniques in EDM for Predicting the Performance of Students"; *International Journal of Computer and Information Technology*,Volume **02**-Issue 06,November 2013.

[6]. Roxanne A. Ancheta, Rosmina Joy M. Cabautan, Bartolome T. Tanuilig Lorena, W. Rabago (2012). "Predicting Faculty Development Trainings and Performance Using Rule-Based Classification Algorithm"; *Asian Journal of Computer Science and Information Technology* **2**: 7203-209.

[7]. http://www.seas.gwu.edu/~bell/csci243/lectures /performance.pdf

[8]. http://www.cs.cornell.edu/courses/cs678/2006sp /performance_measures.4up.pdf