



## A View Point of Enhancing the Quality of Education by Applying Data Mining Techniques

Priya Babbar

Research Scholar,

Department of Computer Science Engineering,  
Lovely Professional University, Phagwara, (Punjab)  
Email: [priyababbar799@gmail.com](mailto:priyababbar799@gmail.com)

Barjinder Singh

Assistant Professor,

Department of Computer Science Engineering,  
Lovely Professional University, Phagwara, (Punjab)  
Email: [barjinder.16469@lpu.co.in](mailto:barjinder.16469@lpu.co.in)

**Abstract**— Education is the process of giving knowledge to students and then evaluating the performance of students with the help of different techniques. For this purpose we use Educational Data Mining. It is a field that discovers the knowledge from the large amount of Educational data. The main purpose of Educational Data Mining is to find the appropriate pattern of educational data by mining so that there is improvement of qualification of education. This paper identifies the factors that are linked with those students whose academic performance is not good and improve the performance of students by identifying slow learners. This paper includes the various attribute selection methods to select those attributes that can improve the performance of students like family background, extra classes, age factor, CGPA or percentage of higher secondary education, frequency of exercise, time spend in exercise, score obtained in analytical questions, descriptive questions etc. and discard the unaffected attributes. For this purpose ranking algorithms will be used to remove those parameters. Various prediction algorithms will be used to predict the student's behavior of pass or fail and then more affective attributes will be applied to those students who will be going to fail to improve their performance.

**Keywords**— EDM (Educational Data Mining), Clustering, Classification, Prediction

### I. INTRODUCTION

Predicting student performance is important for decision makers in the field of education.[8] To predict data mining technique is applied in education called educational data mining. With the help of prediction it is possible to support various important future actions that impact on development of students. In this research we use the concept of Feature selection or attribute selection in data mining. Feature selection technique finds best results by minimizing parameters/ attributes used in data collection. As we know that some student's performance is not same during the year. Some of them will improve and through our study we will find these students that can do better than their previous results. This can be done through analysis of their results with the help of applying different attributes to the students. Research is to enhance the student performance. With the use of less parameters/ attributes the performance will also increased and efficiency will improved. [1]

There are two types of cryptography algorithm that are given below:

#### a. Symmetric key cryptography algorithm

© 2015-19, IJARCS All Rights Reserved

#### b. Asymmetric key cryptography algorithm

### II. METHOD

Data will be collected from different sources as well as set of factors that can affect the performance of students. Then dataset will be prepared by applying technique of attribute selection and data cleaning. At the stage of data mining, Data Mining algorithms are applied to predict student failure like a classification problem. To do this, we will use classification algorithms that are based on rules and decision trees. At the interpretation level, the factors affecting performance and how they are related are considered and applied. [2]

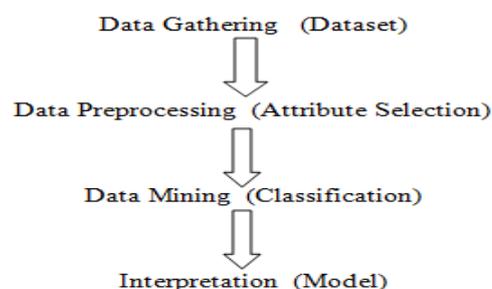


Fig 1: Method of processing

### III. ALGORITHMS USED [3]

#### a) Association rule algorithm

It mainly deals with search statistical relations between objects in dataset. It finds how events aggregate together. Association rules are created by using if-then patterns of data. These are used to analyze and predicting the behavior of students.

#### b) Classification algorithm

It can describe or classify objects related to dataset into predefined set of classes. It is supervised learning approach. It includes objects in dataset used to understand existing objects and predict behavior of new objects.

c) *Clustering algorithm*

It is collection of objects of similar type in one group. The cluster provides us better results. K-means clustering algorithm will be used in our study to result in cluster of students who have similar learning patterns.

d) *C-means clustering algorithm*

To identify the more significant variables that can affect and make impact on the performance of the students we will apply c- means clustering algorithm.[5] C-Means Clustering algorithm is a method of clustering that will allow one piece of data to belong to two or more clusters. This method is used for pattern recognition.

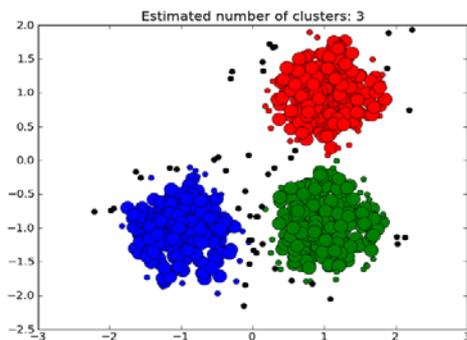


Fig 2: Clustering Algorithm

e) *Inductive and Deductive learning*

Machine learning in mainly classify into two different types. In deductive learning, we learn something with existing knowledge and produce some new knowledge from existing knowledge. In inductive learning rules and patterns are extracted from large datasets. In clustering partition the dataset in to subsets for optimization.

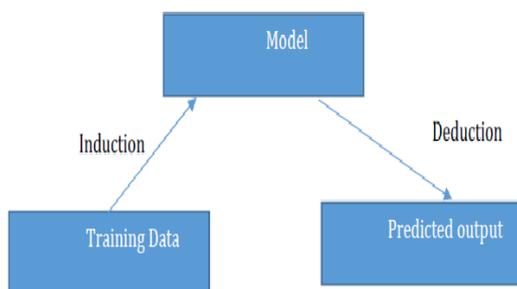


Fig 3: Inductive and Deductive Learning

IV. HYBRID APPROACH OF CLUSTERING AND DECISION TREE

This research will check the effectiveness of decision tree and clustering algorithms by applying them to large scale data set. Classification methods try to find those students who are likely to fail or need more attention. Clustering methods try to make cluster of

students according to their knowledge of subjects. All available data will integrate into single dataset and irrelevant data will be excluded i.e. data with incomplete information about student.

V. ATTRIBUTES USED: [6]

Student’s Sex, Institution at high level, Type of board, Medium of instruction, Type of school, Private tuition, Area at school level, Student having mobile, Computer at home, Student having net access, Internal grade of student, Attendance count Based on School Attendance, Whether qualified or not, Family background, Extra classes, Age factor, CGPA or percentage of higher secondary education, Frequency of exercise, Time spend in exercise, Score obtained in analytical questions, Descriptive questions, 10<sup>th</sup> grade marks, board of study, year drops, backlogs, attendance, class test scores[7] etc.

V. EVALUATION OF PERFORMANCE

Once clusters of similar data are made, the marks of students will changed to discrete values from continuous values like excellent(90-100), very good(80-90) etc.[4] Again clusters will be divided into training datasets. Various attributes will be applied to these training datasets and performance will be evaluated. Again feature selection technique will be applied and less affecting attributes will be eliminated from datasets. Those which attributes are correlated also removed by this technique. Again performance of students will be evaluated. Attributes are selected by different attribute selection algorithms. Algorithms giving same effective attributes will be selected for our study. In this way we will reduce the attributes. With less attributes the efficiency as well as performance will be increased.

VI. ACKNOWLEDGMENT

All praise in the name of almighty God, who give us in the darkness and help in difficulties. The dissertation is the result of full semester of work whereby I have been accompanied and supported by many people. It is a pleasant aspect to that I have the opportunity to express my gratitude for all of them. I am also extremely indebted to my guide Mr. Barjinder Singh (Assistant professor, Department of Computer science, Lovely Professional University, Phagwara). I am very much thankful to Mr. Barjinder Singh for picking me as a student at the critical stage of my masters. I warmly thank him for his valuable advice, constructive criticism and his extensive discussions around my work.

REFERENCES

[1] Hind Almayan, Waheeda Al Mayyan, ‘Improving Accuracy of Students’ Final Grade Prediction Model Using PSO’ 6th International Conference on Information Communication and Management, 2016 pp. 35-39  
 [2] Carlos Márquez-Vera, Cristóbal Romero Morales, and Sebastián Ventura Soto ‘Predicting School Failure and Dropout by Using

- Data Mining Techniques' IEEE Journal Of Latin-American Learning Technologies, Vol. 8, No. 1, February, 201, pp. 7-14
- [3] Data Mining: Concepts and Techniques Second Edition Jiawei Han *University of Illinois at Urbana-Champaign* Micheline Kamber
- [4] Md. Fahim Sikder, Md. Jamal Uddin and Sajal Halder 'Predicting student's yearly performance using Neural Network' 5th International Conference on Informatics, Electronics and Vision (ICIEV), 2016 pp. 524-529
- [5] Ashish Dutt, Saeed Aghabozrgi, Maizatul Akmal Binti Ismail, and Hamidreza Mahrooian 'Clustering Algorithms Applied in Educational Data Mining' International Journal of Information and Electronics Engineering, 2015, pp. 112-116
- [6] Parneet Kaura, Manpreet Singh, Gurpreet Singh Josanc (ICRTC) 'Classification and prediction based data mining algorithms to predict slow learners in education sector' 3rd International Conference on Recent Trends in Computing, 2015, pp. 500-508
- [7] John Jacob, Kavya Jha, Paarth Kotak, Shubha Puthran 'Educational Data Mining Techniques and their Applications' International Conference on Green Computing and Internet of Things (ICGCIoT), 2015, pp. 1344-1348
- [8] Norlida Buniyamin, Usamah bin Mat, Pauziah Mohd Arshad 'Educational Data Mining for Prediction and Classification of Engineering Students Achievement' IEEE 7th International Conference on Engineering Education (ICEED), 2015, pp. 49-53