

# A Review on Mining Students’ Data for Performance Prediction

Ms. Kavyashree K R<sup>1</sup>, Mrs. Lakshmi Durga<sup>2</sup>

P G Scholar, Department of Computer Science and Engineering, Maharaja Institute of Technology (MIT),  
Mysore, Karnataka, India<sup>1</sup>

Assistant Professor, Department of Computer Science and Engineering, Maharaja Institute of Technology (MIT),  
Mysore, Karnataka, India<sup>2</sup>

**Abstract:** A country’s growth is strongly measured by the quality of its education system. Education sector has witnessed sea change in its functioning. Today it is recognized as an industry and as an industry it is facing challenges. The challenges of higher education being decrease in students’ success rate and their leaving the course without completion. An early prediction of students’ failure may help the management provide timely counselling as well as coaching to increase the success rate and student retention. Data mining are widely used in educational field to find new hidden patterns from student’s data which are used to understand the problem. Classification is one of the prediction type classifiers that classifies data based on the training set and uses the pattern to classify a new data. Aim of the project is to develop an internetworking application that uses data mining technique to predict the students’ performance based on their behaviour. This paper explores the link between emotional skills of the students along with the socio economic and previous academic performance parameters using Naive Bayes Classifier technique.

**Keywords:** Educational Data Mining (EDM), Classification, Data Mining, Prediction.

## I. INTRODUCTION

Higher Education of Students has a direct impact on the work force provided to the industry and hence it directly affects the economy of the country. In educational institutions quality of the education is judged by the success rate of the students and to what extent the institute is capable of retaining its students. Student’s academic performance is based upon the diverse factors like personal, social, psychological and other environmental factors. Predicting Students’ performance can help identify the students who are at risk of failure and thus management can provide timely help and take essential steps to coach the students to improve his/her performance

Data mining technique have been applied to predict the academic performance of the students based on their socio-economic condition and previous academic Performances. Classification is one of the data mining technique of predictive types that classifies data (Constructs a pattern) based on the training set and use the pattern to classify a new data (testing set). Classification maps the data into predefined sets or groups of classes. It is often referred to as supervised learning because the classes are determined before examining the data.

Patterns that are discovered by Data Mining methods from educational data can be used to enhance decision making in terms of identifying students at risk, decreasing student drop-out rate, increasing student’s success and increasing student’s learning outcome.

The main objectives of this study are:

1. Identification of different factors which affects a student’s learning behaviour and performance during academic career.
2. Construction of a prediction model using classification data mining technique on the basis of identified predictive variables.
3. To extract valid information from existing students to manage relationships with upcoming students.
4. To improve the performance of the student.
5. Validation of the developed model for higher education students studying in Universities/ Institutions.

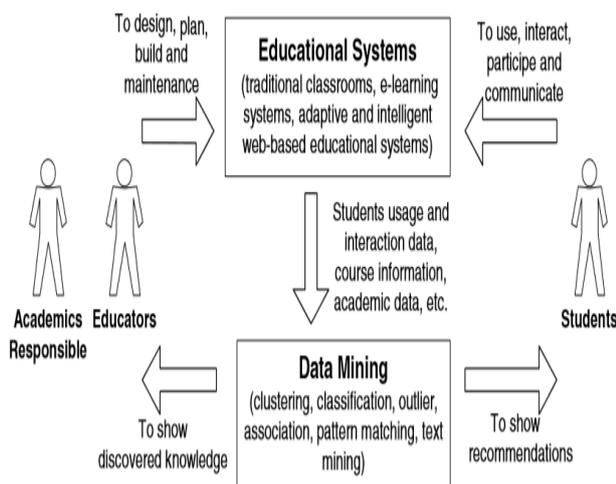


Fig 1- the cycle of applying Data Mining in Education System

## II. DATA MINING TECHNIQUES

Most cited literature survey papers in Educational Data Mining have made use of various classification techniques in order to extract hidden information. It can be categorized into four groups:

1. Decision trees – The algorithms used are ID3, CART, C4.5, Rep tree, Random tree, CHIAD. Decision tree techniques are simple to understand and interpret, It allows addition of new possible scenarios, It helps to determine worst, best and expected values for different scenarios, It can be combined with other decision tree techniques to generate rules easily. This technique has many disadvantages as the number of training data increases like over fitting. It does not handle numeric data and pruning may become cumbersome.
2. Bayesian Classifier- It includes Naive Bayes algorithm and its variants. This technique is simple and easy to understand, requires small amount of training data to estimate the parameters, Fast Space efficient, Insensitive to irrelevant features and handles both real and discrete data well. The major disadvantage of this technique is that it assumes conditional independence among the attributes.
3. Neural Networks- It includes algorithm like Multi-layer Perceptron. This technique is a generalized method, works well with noise. But it does not scale well from small research system to large real-time system. It is computationally expensive and it does not guarantee a solution. It mainly depends on the algorithm used to train the system.
4. Clustering Techniques- It includes algorithms like K-means, Nearest Neighbour etc. This technique is relatively scalable and simple. But they are highly sensitive to noise and outliers. The proposed system is the real time, web based application which has to take any number of training data as an input. So Bayesian classifiers are best suited as it easy to implement and insensitive to irrelevant features.

## III. DATA MINING PROCESSES

The major objective of the proposed methodology is to build a performance prediction model that classifies a student's performance as BAVG, AVG, ABVG and EXCL. The classifiers, has been built by combining the Standard process for data mining that includes Business Understanding, Data understanding, Data preparation, Modelling and Classification. The System Architecture diagram is as shown below.

### 1. Data Understanding

The data of students from various institutions is collected through a structured questionnaire. It can also be collected at the time of college registration. Administrator is responsible for collecting this student's information.

### 2. Data-Pre-processing

The data collected is saved in tabular form. The cleaning process requires eliminating data with missing values, correcting inconsistent data, identifying outliers, as well as removing duplicates etc.

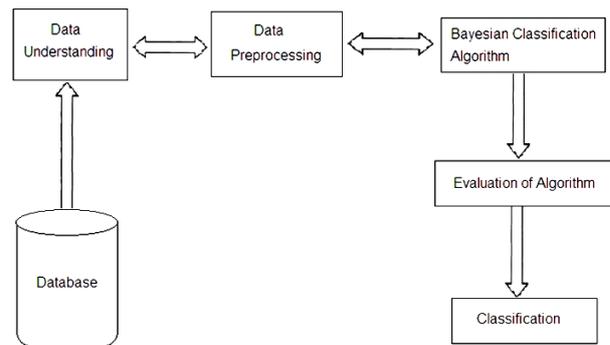


Fig 2- System Architecture

### 3. Modelling

This phase uses data mining technique to identify patterns within the training data. This phase include Selecting techniques, Designing tests, Building models, Assessing models, Generating rules etc.

### 4. Classification

The proposed model predicts the students' individual academic performance based on factors and classifies them among these four groups, namely BAVG (<60%), AVG (60% to less than 70%), ABVG (70% to less than 79%) and EXCL (>=80%).

TABLE I. Attributes Description

Attribute Name	Values	Description
GENDER	Male, Female	Gender
FE	Midschool, Inter, Grad, Postgrad	Father's Education
ME	Midschool, Inter, Grad, Postgrad	Mother's Education
FO	Govjob, Pvtjob, Business	Father's Occupation
MO	Govjob, Pvtjob, Business ,Housewife	Mothers Occupation
FI	MIG,JHG,LJG,VHIG	Annual Family Income
LOAN	Yes, No	Educational loan at any level of education
EARLYLIFE	Metro, City, Village	15 years of life spent
MEDIUM	English, Other	Medium of instruction at school level.
TENTH	BLAVG, AVG, ABAVG, EXCL	% marks in 10th
TWELVTH	BLAVG, AVG, ABAVG, EXCL	% marks in 12th
GRAD	BLAVG, AVG, ABAVG, EXCL	% marks in Graduation
FIRST_SEM	BLAVG, AVG, ABAVG, EXCL	% marks in 1 <sup>st</sup> Semester of MCA
SECSEM	BLAVG, AVG, ABAVG, EXCL	% marks in 2 <sup>nd</sup> Semester of MCA
THIRDSEM	BLAVG, AVG, ABAVG, EXCL	% marks in 3 <sup>rd</sup> Semester of MCA
GRADDEGTYPE	Regular, Distance	Type of Graduation Degree
GRADDEGSTREAM	CS, NCS	Graduation Degree Stream
GAPYEAR	Yes, No	Gap year in education
ACADEMICHRS	INSUF, SUF, OPTIMAL	Hours spent on academic activities
ASSERTION	D, S,E*	Assertiveness of the student
EMPATHY	D, S,E	Empathy of the student
DECISIONMAKING	D, S,E	Decision making ability of the student
LEADERSHIP	D, S,E	Leadership ability of the student
DRIVE	D, S,E	Drive of the student
STRESSMGMT	D, S,E	Stress management skill of the student

\*D, S, E represent need to develop, need to strengthen, need to enhance respectively

#### IV. CONCLUSION

Today academic success of students of any professional Institution has become the major issue for the management. An early prediction of students at risk of poor performance helps the management take timely action to improve their performance through extra coaching and counseling. Employing Data mining techniques like classification can help develop a decision support system to help authorities identify the weak students and take timely measures. Effects of emotional quotient and other parameters have been taken into account when developing a model.

#### REFERENCES

- [1] Pooja Thakar, Anil Mehta- "Performance Analysis and Prediction in Educational Data Mining: A Research Travelogue." International Journal of Computer Application, January 2015.
- [2] Nikitaben Shelke, Shriniwas Gadage, Savitribai Phule Pune University, Maharashtra, India- "A Survey of Data Mining Approaches in Performance Analysis and Evaluation". International Journal of Advanced Research in Computer Science and Software Engineering - Volume 5, Issue 4, 2015
- [3] C Romero, S Ventura, University of Cardoba, Spain- "Educational Data Mining: A survey from 1995 to 2005". Expert systems with Applications-2007
- [4] A. Dinesh Kumar, Dr. V. Radhika, – "A Survey on Predicting Student Performance "(IJCSIT) International Journal of Computer Science and Information Technologies, 2014.
- [5] Dorina Kabakcheiva, Sofia University-"Predicting Student Performance by Using Data Mining Methods for Classification"- Bulgarian academy of sciences, Cybernetics and information technologies, 2013.
- [6] Azwa Abdul Aziz, Nur Hafieza Ismail, Fadhilah Ahmad- "Mining Students' Academic Performance" - Journal of Theoretical and Applied Information Technology 31st July 2013.
- [7] V. Ramesh Assistant, P. Parkavi, K. Ramar -"Predicting Student Performance: A Statistical and Data Mining Approach "- International Journal of Computer Applications 2013
- [8] Surjeet Kumar Yadav, Brijesh Bharadwaj, Saurabh Pal-"Data Mining Applications: A comparative Study for Predicting Student's performance"- International journal of innovative technology & creative engineering December 2013
- [9] Surjeet Kumar Yadav, Saurabh Pal- "Data Mining: A Prediction for Performance Improvement of Engineering Students using Classification"- World of Computer Science and Information Technology Journal (WCSIT) 2012
- [10] Performance Brijesh Kumar Baradwaj ,Saurabh Pal Purvanchal University, India-"Mining Educational Data to Analyze Students"- (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 2, No. 6, 2011.
- [11] Brijesh Kumar Bhardwaj, Saurabh Pal- "Data Mining: A prediction for performance improvement using classification" - (IJCSIS) International Journal of Computer Science and Information Security, Vol. 9, No. 4, April 2011
- [12] Emmanuel N. Ogor -"Student Academic Performance Monitoring and Evaluation Using Data Mining Techniques"-Fourth Congress of Electronics, Robotics and Automotive Mechanics, 2007.
- [13] R. R. Kabra, R. S. Bichkar G. H. Raison Pune, India-"Performance Prediction of Engineering Students using Decision"-International Journal of Computer Application, Volume 36- No.11, December 2011.
- [14] M. Ramaswami and R. Bhaskaran- "A CHAID Based Performance Prediction Model in Educational Data Mining"- IJCSI International Journal of Computer Science Issues, Vol. 7, January 2010.
- [15] RYAN S.J.D. BAKER, KALINA YACEF- "The State of Educational Data Mining in 2009: A Review and Future Visions"- Journal of Educational Data Mining, Article 1, Vol 1, No 1, Fall 2009