



## An Intelligent System to predict Students academic performance using Data Mining

Liya Treasa Kunjumon<sup>1</sup>, Sharon Shaji<sup>2</sup>, Saffi Treasa Saji<sup>3</sup>, Thasneem Naushad<sup>4</sup>, Neena Joseph<sup>5</sup>

Mangalam College Of Engineering ,India, liya4liya97@gmail.com

Mangalam College Of Engineering, India, sharonshaji998@gmail.com

Mangalam College Of Engineering , India, saffitreesa97@gmail.com

Mangalam College Of Engineering ,India, thasneemnaushad2001@gmail.com

Mangalam College Of Engineering India, neena.joseph@mangalam.in

### ABSTRACT

Educational database contains a large volume of data where prediction of student's academic success is a challenging factor. Growth of an institution is based on the academic levels of student. Data mining can be an effective method for predicting the success. Classification task is used for the evaluation. Extracted knowledge describes the student's performance and helps to identify student's who need special attention. It brings benefits to students, teachers and for educational institutions also.

**Key words:** Educational database, Data mining, Classification

### 1 .INTRODUCTION

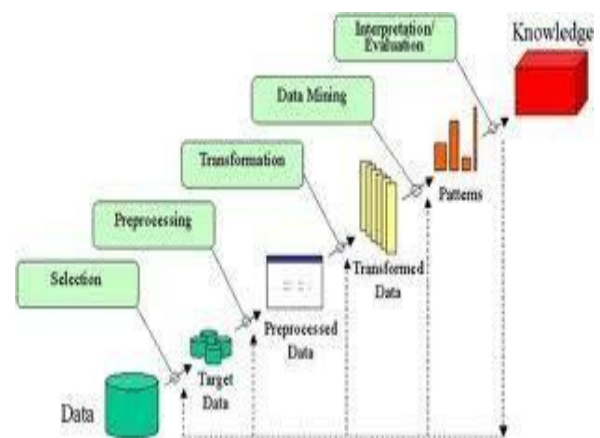
There is a huge amount of data available in the information industry. There is no use until it is converted into useful information. It is necessary to analyze this huge amount of data and extract useful data from it. This extraction of information from huge set of data is known as Data Mining.

Data mining involves Data Cleaning, Data Integration, Data Transformation, Data mining Pattern Evaluation and Data Presentation [1]. Data mining can be used in many complex real world situations. Data mining applications include classification, estimation, Association rule mining, clustering, sequential pattern discovery [1]. Other applications include Business data analytics ,Text mining, Web mining, Social network data analytics .Data mining applications ranges from business to medicine to engineering.

Education is an area where data mining have an immense applications. .Nowadays educational institutions stores the information in both structured and unstructured format, creates challenges to evaluate this data [4].

Success and prestige of an institution is based on the academic levels of students. NBA accreditation is awarded based on the quality assessment in technical education sector. Students quality can be assessed using

data mining technique. This technique predicts whether a student will pass or fail in examination, helps the faculty and institution to give more training for the students who are not performing well. This enhance student's success.



**Figure 1:** Data Mining Life Cycle

Various factors that are taken into consideration for prediction includes Family background, 10<sup>th</sup>, 12<sup>th</sup> marks, internal assessment, assignments attendance, hours spend for studying behavior, sentimental analysis and so on [3]. All these factors are very important in the analysis and gives an accurate result in prediction. This prediction helps to improve the quality of students and institution, gives the relationship of the factors with students performance[2] and also helps to acquire NBA(National Board of Accreditation). The section II deals with related work, Section III explains the methodologies in the proposed system, Section IV explains the result of the prediction and Section V gives the conclusion.

### 2. RELATED WORK

The fundamental techniques for classification are decision tree classifier, neural networks, rule based classifier and Lazy based classifier. Datasets utilized within the classification algorithm ought to be clear and can be pre-processed for handling missing or redundant attributes. The data are to be handled with efficiency to induce the best outcome from the Data Mining process

[5]. The analysis and interpretation of classification algorithm is time consuming.

A survey has done from 1995 to 2005 on the application of data mining to the traditional educational system. After analysing the system the data are pre-processed and have applied various data mining techniques. Mainly the survey was based on web based learning [6], able to provide learning profile based on the behaviour of students.

Decision tree are commonly used for gaining information for the purpose of decision -making. Decision tree starts with a root node on which it is for users to take actions [7]. The final result is a decision tree in which each branch represents a possible scenario of decision and its outcome.

Comparison of various data mining techniques for prediction is done. The data set is taken from the prestigious Bulgarian university. Several well-known data mining classification algorithms including rule learner, decision tree classifier, neural network and Nearest Neighbour classifier are applied on the dataset. As per the study the highest accuracy is achieved by neural network of 73.59% followed by decision tree model and KNN classifier [8].

A study was conducted based upon the psychological and environmental factor and is predicted by different educational data mining techniques [9]. The main objective of Educational Institution is to provide the best quality education and to improve the behavior of the students [6]. The most often used techniques for this type of goal are classification, clustering and association. Classification is the one of the most popular technique in data mining. Predefined classes are predicted using classifier model.

To identify the students who are at the edge of the performance level in a Post Graduate programme. We collected the input data from the students of both first year and second or final year students. Since, the number of students are always less than the Under Graduate level, the availed data samples count was also limited to one hundred students. The data were fed into a system for portioning and classification [5], implemented the rules that are defined in SVM algorithm to predict the students final grade [10]. This led to reduction in the accuracy, since the samples rate was reduced.

The academic achievement of higher secondary school education is a deciding factor in the life of any student. The work, propose the application of an Artificial Neural Network [11] for predicting student's performance at the

final examinations in the course of Mathematics. Aim is to identify the best training algorithm for constructing an accurate prediction model. Neural network classifier in a user-friendly software tool for the prediction of student's performance in order to making this task easier for educators to identify weak students with learning problems in time.

### 3. PROPOSED SYSTEM

Data mining is a knowledge discovery process from a huge volume database. This mechanism works only in large data set where the student academic performance is evaluated. The Naive Bayes algorithm is a simple probabilistic classifier which is based on Bayes theorem with strong and naive Independence assumptions [11].

#### Naive Bayes Algorithm

**Step 1:** Scan the student data set

**Step 2:** Calculate the probability of each attribute value.  $[n, n_c, m, p]$

**Step 3:** Apply the formulae

$$P(\text{attribute value}(a_i)/\text{subject value}(v_j)) = \frac{n_c + mp}{(n+m)}$$

Where:

$n$  = the number of training data item for which  $v=v_j$

$n_c$  = number of examples for which  $v=v_j$  and  $a=a_i$

$p$  = a priori estimate for  $P(a_i, v_j)$

$m$  = the parallel size of the sample

**Step 4:** Multiply the probabilities by  $p$

**Step 5:** Compare the values and classify the attribute values to one of the predefined set of class.

There are many features affecting the student performance. The factors considered are learning style family background, parents education, schooling education, assignments, attendance percentage, unit test performance, previous university result, hours spend for study, social media addiction [3].

Many socio economic, non-academic and academic factors influence the performance of the students. Student's data was collected, pre-processed and data mining techniques was applied to discover classification, all tasks knowledge was extracted that describes students' behaviour. These evaluation helps to improve the academics and behavior of the student. Data mining techniques extract hidden knowledge.

The extracted knowledge helps the institutions to improve their teaching method. This will effect overall performance of the student as well as the institution. Performance prediction model based on data mining techniques along with behavioral features can improve the accuracy of the prediction. Behaviour of the students are determined from the data collected from mentors and

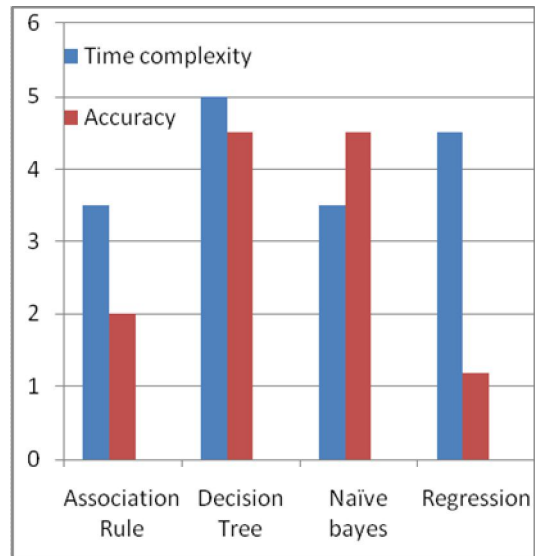
class-in-charge. Mentors are able to give the data like performance of student in their academics as well as their behavior. Mentors also understands the problems faced by the students in their studies and also in their personal life. Class-in-charge will be able to give the information regarding the overall behavior of the student like whether he/she is calm, arrogant ,hyperactive, and so on.

**Table 1:** Features Analyzed

S.No	Description	Possible values
1	Student’s gender	Male, Female
2	Student’s category	OBOC,SC/ST,GN
3	Medium of teaching	English, Malayalam
4	Where do you stay	Hostel, PG, Home
5	Father’s occupation	Farmer, Business, Service, retired, Not applicable
6	Mothers occupation	Farmer, Business, Service, retired, Not applicable
7	10 <sup>th</sup> % , Board	<45,45-60,61-80.>81
8	12 <sup>th</sup> % , Board	<45,45-60,61-80.>81
9	Attendance %	<75, >75
10	Hours for study	<2 , 2-4, >4
11	Behavior	Calm, Arrogant
12	Current Btech CGPA	<6, 6-7, 7-8, >8
13	Internal assessment %	>45, 45-60, 60-80, >80
14	Usage of Internet	Yes, No

**4. RESULT**

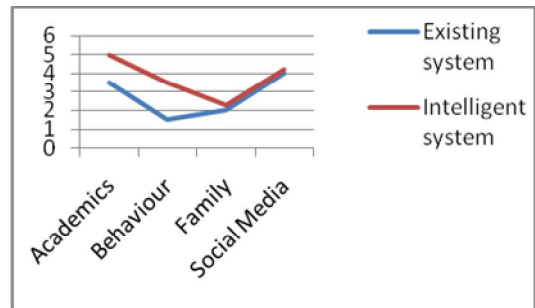
The experiment results that there is a strong relationship between the academic features and behavioral features for the prediction. Fig 2 shows Naïve Bayesian algorithm provides more accuracy and less time complexity compared with others algorithms.



**Figure 2:** Comparison

The result gives the list of students who needs special attention in their academics which helps the faculty to give more attention which will lead to increase in their student’s academic level.

Fig 3 shows that this intelligent system increased the level of accuracy of prediction compared to the existing system with the consideration of behavioral features along with the academic features



**Figure 3:** Comparison of Existing System and Proposed System

**5. CONCLUSION**

Academic achievement of an institution manly depends on student performance. Student’s performance prediction model along with their behavioural features will give accurate results. These types of features are related to the learners interactivity with the learning system. The performance of student’s predictive model is based on Naive Bayesian classifier which reveals that there is a strong relationship between the learning behavior and their academic achievement .The addition of behavioural features improved the prediction .This prediction can help educators to understand and identify

weak students and which in turn helps to improve the learning system.

## REFERENCES

- [1] Rupali, Gaurav Gupta "Data mining techniques, Applications and Issues" International Journal of Advanced Research in Computer Science & Electronics Engineering(IJARCSEE) vol 2, Issue 2, Feb2013
- [2] Amjad Abusaa "Educational Data mining & student's performance prediction" International Journal of Advanced computer science and Applications vol 7, Nov 5 2016  
<https://doi.org/10.14569/IJACSA.2016.070531>
- [3] V Ramesh, P Thenmozhi, Dr K Ramar "Study of Influencing factors of Academic Performance of students: A Data mining Approach", International Journal of scientific& Engineering Research , volume 3,Issue 7,July 2012
- [4]Sedighen Abhasnasab Sardarh, Rosalam Cheme "Enhancing Education Quality using Educational data "
- [5] M.S. Mythili1 and A.R.Mohamed Shanavas , "An analysis of students' Performance using classification algorithms ", IOSR-JCE, Volume 16, iss1, Jan. 2014.  
<https://doi.org/10.9790/0661-16136369>
- [6] C. Romero and S. Ventura, "Educational data mining: A survey from1995 to 2005," Expert systems with applications, vol. 33, no. 1, pp. 135-146,2007.  
<https://doi.org/10.1016/j.eswa.2006.04.005>
- [7] Baradwaj, Brijesh Kumar, and Saurabh Pal. Mining Educational Data to Analyze Students' Performance. Arxiv preprint arxiv: 1201.3417, 2012.
- [8] Kabakchieva, D., Stefanova, K., Kisimov, V. (2011). Analyzing University Data for Determining Student Profiles and Predicting Performance. Conference Proceedings of the 4th International Conference .on Educational Data Mining (EDM 2011), 6-8 July 2011, Eindhoven, The Netherlands, pp.347-348.
- [9] JA.Dinesh Kumar and V.Radhika, "A survey on predicting student performance", International Journal of Computer Science and Information Technologies, Vol. 5, 2014.
- [10] K B Eashwar, R Venkatesan "Students Performance Prediction using SVM" International Journal of Mechanical Engineering & Technology vol8,Issue 11,Nov 2017
- [11]Vinaya Patil, 2 Shiwani Suryawanshi, 3Mayur Saner, 4ViplavPatil, 5Bhushan Sarode "Student Performance Prediction using Classification Data Mining Techniques"