PERFORMANCE OF STUDENT PREDICTION

[1]Mohammed Afzal Ahamed; [2]Prayuk Chaisanit; [3]Prof. Mahesh T R

[1]Department of CSE, School of Engineering and Technology – Jain University, India
[1] Afzalahamed511@gmail.com, [2] mpc.chail@gmail.com, [3] maheshtr.1978@gmail.com

Abstract— Data mining has been successfully implemented in the business and technology world in these days, its use in high education and still relatively new. In the term of education data mining would help the institution to come up with their student performance including academic performance, attendance and if the candidates participated in any activities. Using data mining the aim was to develop a model which can derive the conclusion on students' enrolment behaviour. Different methods and techniques of data mining were compared during prediction of students' enrolment. This paper is supporting the technique that will help the institute to analyst the prediction of admission in which department student's intent to enrol in this institution. It contributes the techniques that would help to predict the performance of students by using the attendant, their work performance, and their behaviours in classes also their performing in the test in each month. The system also proposes an enhanced data mining technique like Bayes algorithm which helps to predict the performance of student by their given data also Cluster technique to separate each student's performance in a different group.

Keywords— Bayes, Naive Bayes, Clustering, Data analysing, Classification algorithm

I. INTRODUCTION

Data mining is one of the methods that can be implemented in higher educational institutions for student enrolment and to predict student performance. By using data mining techniques institutions can aim to improve the quality of education. A world's growth is strongly measured by the quality of its education system. Education sector, across the earth has witnessed sea change in its functioning. Today it is recognized as an industry and like any other industry it is facing challenges, the major challenges of higher education being decrease in students' success rate and their leaving a course without completion. An early prediction of students' failure may help to provide timely counseling as well coaching to increase success rate and student retention. We used different classification techniques to build a performance prediction model based on students' social integration, academic integration, and various emotional skills which have not been considered so far. This paper comprises the use of decision tree, Naive Bayes, Support Vector Machine techniques and algorithms based on them. These algorithms were used to study and recognize the space in prevailing examining techniques for analyzing of scholar's performance. Student performance can be predicted considering different factors of his past academic records to enroll a student in

particular department. Educational institutes focus on generating graduates with good academic performances as well as extra-curricular activities. They need to keep track on how the student is performing in particular attributes like students last semester grades, internal assessment, lab assignments, attendance, final semester marks, etc.

We are actually trying to enhance student's acquirement and successive more effectively in a way using data mining techniques educational.

1.1 Problem Statement

Using data mining the aim was to develop a model which can derive the conclusion on students' performance. Different methods and techniques of data mining were compared during prediction of students' enrollment. This paper is supporting the technique that will help the institute to analyst the prediction of students' performance in which department student's intent to enroll in this institution using previous academic data of student. It contributes the techniques that would help to predict the performance of students by using the students' last semester grades, internal assessment, lab assignments, attendance, final semester marks, etc.

Preliminary education adds to a nation's literacy rate but higher education has a direct impact on the work force being provided to the industry and hence directly affects the economy. Lots of Institutions of higher learning have been set up across India. However, the quality of education is judged by the success rate of students and to what extent an institute is capable of retaining its students. 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 performance. Data mining techniques had been applied to predict the academic performances of the students based on previous academic performances and their sociological-economic condition.

This paper explores the link between emotional skills of the students along with sociological economic and previous academic performances parameters to predict academic performances using data mining techniques. The emotional skills like assertions, leaderships, stresses management etc. are obtained, using standard Emotional Skill assessment process ESAP.

Higher educational institutions consider Student's performance as one of its most crucial part. This is because of the fact that most of the educational institutions are based on the best record of academic performances.

By using Educational data mining techniques, the educational institutions can have the idea before the starting of the new semester and can have informed decision so it will help them to effectively deal with all problems faced by the students while performing academically or in their personal life as will be already known to them. Large volumes of data are analyzed using educational data mining techniques so as to find different trends and patterns to predict the student performance.

There are multiple data classification techniques used for predicting the results each one having its own advantages and disadvantages. The system also proposes an enhanced data mining technique like Naive Bayes and Apriority Algorithm which help to predict the performance of student by their given data also Cluster technique to separate each student's performance in a different group.

1.2 Limitations of the Current Work

Many techniques have been proposed with improved technologies. But every system has some limitations which can be improved such as:

- Institutions should be able to collect valid data from the sources to get valid output data so that it can be analyze easily.
- Skilled person is requiring analyzing the data. Tools present for data mining are powerful. But they require a skilled person to process the data using these tools to prepare the data and understand the output.
- It can be only used in higher studies institutions. Small scale educational institutions cannot implement this technology due to their fewer intakes of students.
- The technology proposed is not always cheap.

1.3 Objectives

- Analyze and classify the student data to divide them in different groups.
- Predict the student performance with the given data and to classify students in different groups.
- Considering various factors in that data so that it will be easy to classify students based on the given data.
- Analyzing the root cause of students with low performance.

II. LITERATURE SURVEY

Several research studies have dealt with the prediction of students' performance in different levels. Data mining is also known as the process of Discovery of Knowledge which refers to extracting or mining information from huge bunch of data. It helps in determining fascinating knowledge such as anomalies, changes, associations, patterns

And important structures from huge volumes of information stored inside several different kinds of databases as in data warehouses or other information repositories available. It's been popularly used nowadays due to the availability of very huge volumes of data in electronic form and there is a need for converting such data into useful information & knowledge for large applications. Decision Support, Artificial Intelligence, Machine Learning, Statistics and Database Systems and Business Management are some of the fields using its applications.

The existences of some issues are implicitly / explicitly stated in the processes of predicting the future achievements of students. Making the motivation for some recommendations that can be taken when considering different EDM techniques to have better results. The algorithm can deal with the certain type of features, utilize influencing factors that affect classification accuracy, therefore, the preprocessing steps are time-consuming and effort-intensive. Feeding algorithm with students' a priori knowledge (i.e., previous semester grades, daily exams, and tests) can improve accuracy.

Using algorithm that has the ability to handle outliers. Taking into its consideration dataset scalability. The result of the algorithm must be interpreted to understandable rules. Using of methods that can deal with misclassification problem resulted from using an ambiguous feature, for example, fuzzy reasonable. Data collected manually through questionnaire is time-consuming thus student admission data in learning system is recommended. The most important issue in predicting student's early failure or success is to determine and identify the most influencing factors that affect student during their study. This is done by analyzing the relationship between students behavioral and their scores.
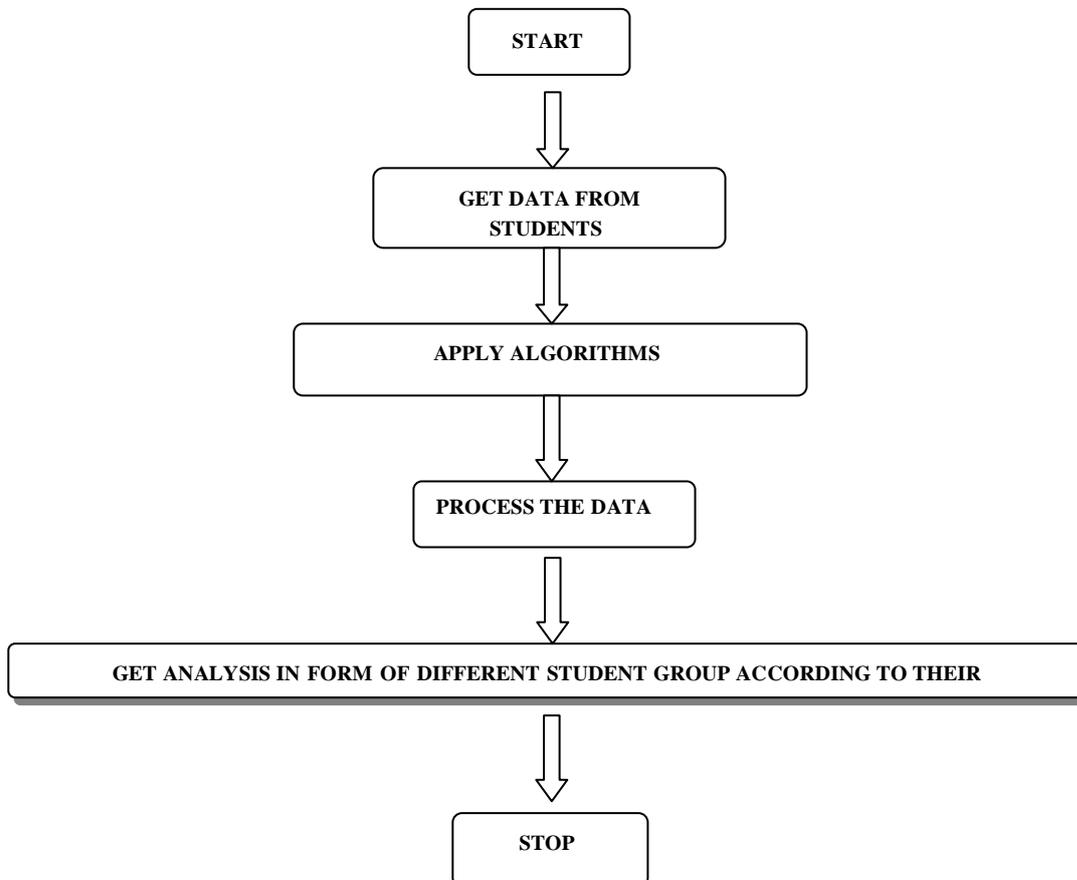
III.   DESIGN METHODOLOGY

FIG.3.1- FLOW DIAGRAM

47

We use different types of algorithm in our system to analyst data.

Classification algorithm: is a data mining technique that helps us to map data into predefined category. It is a supervised learning technique which needs categorized

Training data so it can create rules for categorizing test data into pre-arranged category.

NAÏVE BAYES: model assumes that all variables contribute toward classification and that they are mutually independent.

In other words, It assumes that variables are not correlated. This is unrealistic assumption for most of the datasets, however it leads to a simple prediction framework that gives surprisingly good result in many practical cases.

Clustering algorithm: generally means grouping certain set of components in a way that the components in the same cluster are more similar to each other than to those in other categories. Several fields like pattern recognition, image analysis, machine learning and information retrieval refer to this as a common technique for statistical data analysis.

In process of implementation of clustering different classes can be discovered from the data and are examples of unknown apriori. As our main aim is to analyze students' performance into any of the predefined level Good and Bad, for which clustering was not appropriate, so we have used classification method instead of clustering method proper data mining techniques are used to analyzing the existing components and then classifying them in order to provide relevant results or outcomes. Hence if all factors and components are considered for the analysis, it can effectively increase the prediction model accuracy. The most important issue in predicting student's early failure or success is to determine and identify the most influencing factors that affect student during their study. This is done by analyzing the relationship between students behavioral and their scores.

A. Models

We are using three kinds of classification models so as to learn the predictive function which is required. The models are used for experimental analysis. They are selected on the basis of their frequent usage in the existing literature. The list of methods is as follows: 1) Decision Tree A decision tree is a tree in which each branch node will represent a choice between several alternatives and each leaf node will represent a decision. A decision tree is commonly used for obtaining information so as to fulfils the purpose of decision making. Decision tree starts from a root node which is there for users to take actions. From root node users split each and every node recursively into different nodes according to decision tree learning algorithm. The final result is a decision tree where each branch represents a possible context of the decision and its outcome.

B. Bayes

Bayes algorithm is actually based on the probability theory, i.e. the Bayesian theorem and is a simple classification method. It is named as naive because it solves problems based on two critical assumptions: it assumes that there are zero hidden components that will affect the process of analyzing and it supposes that the prognostic components are conditionally independent with similar classification. This classifier provides an efficient algorithm for data classification and it represents the promising approach to the discovery of knowledge.

C. Data Preparations

For experimental purpose, the data of graduate and undergraduate students have been collected from different universities during the period (2017 to 2018) through a questionnaire survey. Once we got the details of all the students, we divided the training dataset, considering various feasible dividing components, i.e. the components which will have a major effect on the students' performance. Preprocessing is applied to obtain the most relevant characteristics of students. After removing inconsistencies and duplications in the dataset, we considered student instances for experiments. The student's performance model was created, where performance is measured with the performances in the areas such as "Academic", "Behavior", "Extra- Curricular", and "Placement".

IV.   PROPOSED SYSTEM

There is an important factor that would help and affect the prediction of students
Performance in the institution there are

1. High school percentages

High school marks are one of the most important factors to help the institution to predict the performance of their students the researchers have found that the student who have done well in theirs intend to perform well in their undergraduate programs due to their studies behavior since there were small.

2. First year of undergraduate programs percentage

first year mark is the important factor because it is showing the actual behaviors of student seriousness in their studies because it is the time that before you go for the different departments it have included all the basic of each subject to show the students weaknesses and strength so in the 2nd year student can decide their way to go and also showing how the student attitude in their studies.

3. Student's Activities achievement

The research had shown that most of the student who is not regular in classes but they have been participating in some events or competitions have been doing well in their academics because they have learn something more than what actually had been teaching in the class they have faced many situations and have to solve the problems themselves so they would understand the problems and solve it in the better way they also can understand the subject more faster than people who just sitting and attending regular classes.

4. Student behaviors

Students behaviors became the important factors of student's performance in academic due to the way they have representing themselves to the teachers, they attendant for the class and also the family background these all are important factors for the student to present their studies proposal. We have seen that the student who have maintained their class discipline have been doing well in the academic performance and also have respectful to the teachers and students. Regulation for classes attending of the student have shown the great result in the academic performance, the students who attending the class regularly could have done well in theirs exam and also their work because when you attending the class you can understand and get to know things 70-80% and the left is when you reread and study after the class will help you to understand more better and get the perfect results but most of students who has missed many classes would be disadvantaged of most topics and contents of the subject that have been taught.

Family background in this the research have shown that the students whose family have issues like parent's unemployment, parents Income, domestic violence and parent education. The student whose parents is unemployment is intended to misbehavior due to uneducated or if they parents is the business people who is actual rich would behave like spending and throwing money on everything they want to achieve something they need also students whose parents is have less income intended to make themselves to work even during class to overcome their family need and survival so they may miss classes and could not follow up to the point that have been taught in classes so it would affect their performance and the most issues is domestic violence and parents education is main of all because the parents who has less educated would then help their child to solve some problems or give them an advice to do or not to do something for the students in their academic and the domestic violence is making the students to misbehave and showing bad attitude to the people around to uneducated and the family problem and the way grown from the people surrounding them.

Table number 4.1

CATEGORY	NAME	DESCRIPTION	POSSIBLE VALUES
ACADEMIC	AttendanceA80 MarkA80 MarkA40 MarkA0 Understanding	Attendance above 80% Marks above 80% Marks above 40% Marks above a 0% Understandings of things	A -YES; B - NO A -YES; B - NO A -YES; B - NO A -YES; B - NO A -YES; B - NO

BEHAVIOUR	ParentStudy ParentEmployed FamilyIssue FailToleranceCapacity FightwithFriends SpendMoneyUseful BadHabitfromFriend BadHabitfromFamily PoliceCompliant	ParentStudy ParentEmployed FamilyIssue Students tolerance for failures Get in flight with friends Spends money to buy needed items Got a bad habit from friend Got a bad habit from family Has police complaint	A – YES; B – NO A – YES; B – NO A – YES; B – NO A – YES; B – NO A – YES; B – NO A – YES; B – NO A – YES; B – NO A – YES; B – NO A – YES; B – NO
PLACEMENT AND EXTRA-CURRICULAR	SportInterest PartTimeJob SocialServices PersonalInterviews Aptitude	Interesting in the sport Does part time job or internship? Does social services Personal interviews plastic Aptitude plastic	A – YES; B – NO A – YES; B – NO A – YES; B – NO A – YES; B – NO A – YES; B – NO

V. CONCLUSIONS

In the present article, an application is proposed that can be used to know the student's behavior and their academic performance to predict the possible of percentages that they could get for each year and final results that may appear after theirs graduate from the institution.

It is actually the collection of the candidate's previous years from their 10th to 12th STD and each year results in undergraduate programs also their behaviors in the class, theirs attitudes and any achievement to find out the possible that the students can be performed each year and their final results. This would help the institution to find out the students' weaknesses and to find the solution to bring them overcome to perform well in academic also it helps for student to see their strength and weaknesses so they would know where to work out to achieve the goals.

REFERENCES

[1]. Astha Soni, Vivek Kumar, and Rajwant Kaur, D. Hemavathi , "PREDICTING STUDENT PERFORMANCE USING DATA MINING TECHNIQUES," Kancheepuram Dt., Tamil Nadu, INDIA,Volume 119 No. 12 2018,pp. 221-227

[2]. Saja Taha Ahmed, Prof. Dr. Rafah shihab Al-Hamdani, and Dr.Muayad Sadik Croock ,"Studying of Educational Data Mining Techniques", ISSN: 2350-0328,Vol. 5, Issue 5 , May 2018

[3]. Elakia, Gayathri, Aarthi, and Naren J." Application of Data Mining in Educational Database for Predicting Behavioural Patterns of the Students", SASTRA UNIVERSITY, India, ISSN: 0976-9646, Vol. 5 (3), 2014, 4649-4652

[4]. Heena Sabnani1 Mayur More, Prashant Kudale, and Prof. Surekha Janrao," Prediction of Student Enrolment Using Data Mining Techniques", Terna Engineering College, Nerul, Navi Mumbai, e-ISSN: 2395-0056, p-ISSN: 2395-0072, Volume: 05 Issue: 04 | Apr-2018.

[5]. Maryam Zaffar, K.S. Savita, Manzoor Ahmed Hashmani, and Syed Sajjad Hussain Rizvi "A Study of Feature Selection Algorithms for Predicting Students Academic Performance", Vol. 9, No. 5, 2018.

[6]. Shanthini, A., G Vinodhini and R.M. Chandrasekaran" Predicting Students' Academic Performance in the University Using Meta Decision Tree Classifiers" SRM University, Chennai, India, 2018, 14 (5): 654-662.