

# An Optimal Way To Find And Evaluate OPI of Student Using Data Mining

Ashish Patil<sup>1</sup>, Aiysha Khairatkhan<sup>2</sup>, Apurva Nangare<sup>3</sup>, Priyanka Ghorpade<sup>4</sup>

<sup>1,2,3,4</sup>Dept of Computer science and engineering

<sup>1,2,3,4</sup> AGTI's Daulatrao Aher College of Engineering Karad, Maharashtra, India

**Abstract-** Education can be utilized as a tool to face many problems, overcome many hurdles in life. The knowledge obtained from education helps to enhance opportunities in one's Employment development. To extract useful information from the knowledge obtained, Educational Data Mining is widely used. Educational data mining provides the process of applying different data mining tools and techniques to analyze and visualize the data of an institution and can be used to discover a unique pattern of students academic performance and behavior. The present work intends to enhance students academic performance in data mining techniques. There is need to calculate API using the Behavioral Index as well as the academic performance. Academic performance achievement is the level of achievements of the students educational goal that can be measured and tested through examination. However, the academic performance achievements varies as different kind of students may have different level of performance achievements. Where as, the behavioral performance achievement varies with the personality and behavioral techniques. These available students data could be extracted to produce academic useful information along with the Behavioral index. Hence, OPI i. e overall performance index will be calculated. we provide the basic information of student as well as the student has to give the two tests Academic test and the personality test. On the basis of these two tests the overall performance of the student is calculated. The project will also help to find out the weak students and the strong students. Various factors of the student are considered (such as the occupation of their parents, education of their parents, family relation etc) there are 35 factors of student for calculating the overall performance of the students.

**Keywords-** Academic performance index(API), behavior performance index(BPI), Overall performance index(OPI), Naive Bayes, data mining.

## I. INTRODUCTION

The analysis aims to firstly implement an automated system which just requires the data-set of students, and then the system classifies the students automatically into two classes which are pass and fail reducing the human work.

Secondly, building classification algorithms on educational environments helps to identify the students who need special tutoring or counseling from the school/college. The higher authorities of an institution can use such classification models to improve students' performance according to the data-set. The proposed system predicts students' academic performance and the factors which affect performance failure. Building such classifiers helps an educational institution to get the picture of their educational level, can compare their progress with other educational institutions and finally, guide students for their better future.

## II. LITERATURE REVIEW

In several studies, Association rule based DM approach has addressed input variables such as sex, age and performance over past years and the proposed system has outperformed traditional allocation procedure. They have used many approach such as neural networks and decision tree (94% combined accuracy), binary classification (72% accuracy) [1].

The best result was obtained by Naïve Bayes classification. The authors adopted regression approach to predict math skills based on score obtained by individuals. Most of the students join the public schools for free education. There were some core courses which share a common language like in other countries. 170 International Conference on Computing, Communication and Automation (ICCCA2017) 2 The grading point is scaled up to 20, where 0 is the lowest and 20 is the perfect score. During school year, students were evaluated in three periods and last evaluation.

In [2], Naïve Bayes Classification is used to build a model in which probability distribution function is computed to take care of continuous data. In order to increase the accuracy of the model, optimal equal width binning for discretization is introduced. Furthermore, to increase the accuracy of the model classes are balanced.

In [3], two classifiers namely, Naïve Bayes and J48, are used by considering the data from the UCI Machine Learning Repository. Analysis for these algorithms are

performed using WEKA tool and the accuracy of the models are increased using discretization of continuous features.

In [4], many classification algorithms are used and one among them is NB classifier. The students are classified into four classes A, B, C and F where these classes are labels. The entire data-set is used to build the classifier and then Bootstrap method is applied to enhance the accuracy of the each classifier.

In [5], Bayesian Network is used to classification the students based on marks scored by them. Model is built using training data-set and to compare the relative performance test data-set is used. 10-fold-cross validation is used for model evaluation.

In [6], a technique is used to predict the students' performance by combining three classification algorithms, Naive Bayes, the 1-NN and the WINNOW using the voting methodology.

### III. MOTIVATION

This study aims to firstly implement an automated system which just requires the data-set of students, and then the system classifies the students automatically into two classes which are pass and fail reducing the human work. Secondly, building classification algorithms on educational environments helps to identify the students who need special tutoring or counseling from the school. The higher authorities of an institution can use such classification models to improve students' performance according to the data-set. The proposed system predicts students' academic performance and the factors which affect performance failure. Building such classifiers helps an educational institution to get the picture of their educational level, can compare their progress with other educational institutions and finally, guide students for their better future.

### IV. PROBLEM STATEMENT

Education plays an important role in the development of the country and a key factor for achieving long term economic progress. To Analyze / Identifying weak students and strong students is difficult regarding knowledge with respect to outcome based education.

The Statistics of literacy shows the percentage of failure rates due to, dropping out from the schools and student failing in a particular subject.

### V. DATA DESCRIPTION

The data-set considered consists of 395 tuples and 34 attributes [7]. Each tuple represents the attribute values of a student or it provides the details of the student in terms of academic performance and social behaviour.

Attribute	Description (Domain)
Age	Student's Age(numeric: from 15 to 22)
Address	Student's Home address(binary: urban or rural)
Mother Edu	Mother's Education(numeric from 0 to 4 <sup>th</sup> )
Father Edu	Father's Education(numeric from 0 to 4 <sup>th</sup> )
Father job Guardian	Father's job Student's guardian(nominal(mother, father, or other))
Family Edu support	Family educational support
Family size	Family size(numeric from 0 to many)
Travel time	Home to school travel time
Study time	Weekly study time(numeric: in hours)
Failures	Number of past class failures(numeric: 0 to n)
Activities	Extra curricular Activities(binary: yes or no)
Dalc	Workday alcohol consumption(binary: yes or no)
goout	Going out with friends(numeric)
Free time	Free time after school(numeric)
Health	Current Health status

Fig.1: The details of a student which forms the data-set

### VI. SYSTEM ARCHITECTURE

The system architecture depicts that overall performance of the system. All the basic data is taken as the input from the student and applying the classification models on it and the Academic Performance index is generated. The Behavioural performance index is also generated separately and by combining both the Academic Performance Index and Behavioural Performance Index the Overall Performance Index is generated.

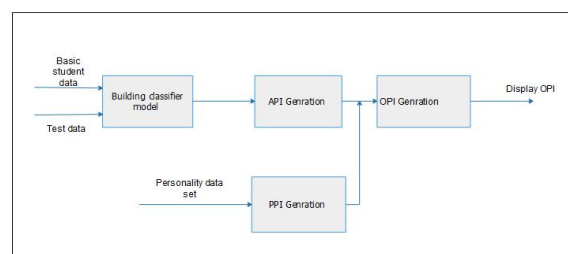


Fig.2: System Architecture

**VII. NAIVE BAYES CLASSIFICATION**

Classification is a process which involves separation of classes based on extracted features. After the classification, classes formed will be distinct from each other. Different classes will have different features. The patterns found in the training data-set play an important role to build the classifier. There are many classification algorithms like k-nearest neighbours, decision tree learning, support vector machine, naive bayes and neural networks which can be used according to the requirements of an application. In the proposed system, Naive Bayes classifier is used. Naive Bayes classifiers are statistical classifiers. Given a tuple, this classifier can predict to which particular class a tuple belongs to. A Naive Bayesian classifier is based on Bayes theorem [8] [9].

Bayes theorem provide a way for calculating the posterior probability, P(c|x), from P(c), P(x) and P(x|c). Naive Bayes classifier assumes that the effect of the value of a predictor (x) on a given class (c) is independent of the values of other predictors. This assumption is called class conditional independence.

Bayes theorem is given by;

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

$$P(c|x) = P(x_1|c) \times P(x_2|c) \times \dots \times P(x_n|c) \times P(c)$$

$$P(x_n|c) \times P(c)$$

where,

P(c|x) is the posterior probability of the class for given attribute

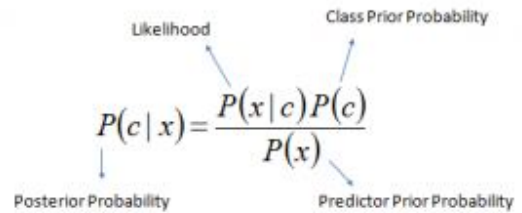
P(c) is the prior probability of class

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

P(x) is the prior probability of predictor.

Naive Bayes classifier’s assume that the effects of the value of the predictor (x) on the given class (c) is independent of the values of the other predictors. This assumption is called class conditional independence. In order to calculate the posterior probability, a table called frequency table is constructed for each feature (attribute) against a particular class. Then the frequency tables obtained are converted into probability tables and Naive Bayesian (Eq. 1) is used to calculate posterior probability for each class. The outcome of this computation is the class with the highest posterior probability. A student will be added to the class which has the highest posterior probability.

Navie Bayes theorem provide the way for calculating posterior probability P(c|x) from P(c), P(x) and P(x|c). Look at the equation below:



$$P(c | X) = P(x_1 | c) \times P(x_2 | c) \times \dots \times P(x_n | c) \times P(c)$$

Above,

- P(c|x) is the posterior probability of class (c, target) given predictor (x, attributes).
- 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.

**VIII. PROPOSED SYSTEM**

In our system there are many modules, but the Admin module and the student module are two main modules. Firstly if the student is new to the system he has to do registration in the system. Then that student can login in to the system give the appropriate test i. e: Academic Test, Behavior Test, and G.K Test.

The result is generated after successfully giving the test and Overall Performance Index is calculated and displayed to the individual student as well as to the Admin side.

The admin has all the authority of generating the questions and also he can see the individual report of the student as well as the overall all report of all students.

**IX. GOALS AND OBJECTIVES**

- To optimizing and analyzing the student performance.
- To Evaluating student behavior based on the basis of attributes using data mining algorithm.
- To analysis of student based on API using data analytics.
- To analyze and classify student into various categories based on API’s using classification algorithm.

## X. MODULES AND THEIR FUNCTIONALITIES

### Student and Admin Registration:

Students as well as admin will register into the system and they will get the login id and password. They can login into the system for giving tests. Also students can check their own performance. Admin will have the report of the students.

### Detail information of student and Admin:

The Students have to give their detailed information of their family details, relation in the family, etc. All the information given by the student is useful for analyzing the student behavior and performance.

### Academic details and Technical Test:

The Academic details of the students contain their percentage and technical details. Students are supposed to give the technical test according to their branches.

### Personality Test:

Personality of the student will be checked by taking a personality test. The personality test contains all the personality development questions. The ethics and behavior of students will also plays an important role in student performance.

### OPI calculation:

Overall Performance Index of the student is calculated by are as follows:

$$OPI=API+PPI$$

Where,

OPI=Overall Performance Index

API=Academic Performance Index

PPI=Personality Performance Index

### Analysis of OPI with respect to API:

Overall performance of the student will be calculated by the above formula w. r. t Academic Performance Index and Personality Performance Index.

## XI. BENEFITS OF THIS SYSTEM

- **In the various organizations:**

In various organizations this might be used for the performance calculation of their students.

- **In all the universities:**

In all the universities may use this system for calculating the student performance and as well as to give outcome based education.

## XII. CONCLUSION

Education is very important in today's generation and methods to analyze the education system in school and to predict for the advancement of institution is very essential. One of the other data mining techniques i.e., classification is an interesting topic to the researchers as it is accurately and efficiently classifies the data for knowledge discovery. This student's performance analysis system will analyze the performance of the students. This system will also identify the weak students and strong students. Student performance can be influenced by different types of attributes On the basis of the various different attributes of the student, the performance is analyzed and report is generated. Student performance analysis system also useful for the outcome based education.

## XIII. ACKNOWLEDGMENT

Special appreciation goes to Prof. Ashish N.Patil Head of Department of Computer Science & Engineering, Dr. Ashok Gujar Technical Institute's Dr. Daulatrao Aher College of Engineering, Karad for his constant support and guidance throughout the course of our work. His sincerity, thoroughness and perseverance have been a constant source of inspiration for us.

## REFERENCES

- [1] <https://searchsqlserver.techtarget.com/definition/data-mining>
- [2] <https://www.analyticsvidhya.com/blog/2017/09/naive-Zayes-explained>
- [3] <https://www.techopedia.com/definition/1181/data-mining>
- [4] <https://data-flair.training/blogs/classification-algorithms/>
- [5] <https://www.semanticscholar.org/paper/Classification-algorithm-in-Data-mining-%3A-An-Neelamegam>
- [6] Ahmed, A. B. E, Ibrahim S. E. "Data Mining: A prediction for Student's Performance Using Classification Method." World Journal of Computer Application and Technology Volume 2(2) (2014).
- [7] Syed Tanveer Jishan, Raisul Islam Rashu, Naheena Haque and Rashedur M Rahman. "Improving accuracy of students' final grade prediction model using optimal equal

- width binning and synthetic minority over-sampling technique”, Springer Open Journal, (2015).
- [8] Kayah, F. “Discretizing Continuous Features for Naive Bayes and C4. Classifiers”. University of Maryland publications: College Park, MD, USA.
- [9] S. Taruna, Mrinal Pandey, “ An empirical analysis of classification techniques for predicting academic performance”, IEEE Advances Computing Conference (2004).
- [10] Varsha Namdeo, Anju Singh, Divakar Singh and Dr. R.C Jain, “Result Analysis Using Classification Techniques”, International Journal of Computer Applications (0975 - 8887) Volume 1 – No. 22. (2010)
- [11] S. Kotsiantis, K. Patriarcheas, M. Xenos, “A combinational incremental ensemble of classifiers as a technique for predicting students’ performance in distance education”, Knowledge-Based Systems 23 Elsevier, 529–535, (2010).
- [12] “Student Academic Performance and Social Behavior Predictor using Data Mining Techniques” by Suhas S Athani, Sharath A Kodli, Mayur N Banavasi, P. G. Sunitha Hiremath. International Conference on Computing, Communication and Automation (ICCCA2017).
- [13] Meghna Khatri. “A Survey of Naïve Bayesian Algorithms for Similarity in Recommendation Systems”. International Journal of Advanced Research in Computer Science and Software Engineering, Volume 2, Issue 5, May 2012.
- [14] Bekele, R., Menzel, W. “A bayesian approach to predict performance of a student (BAPPS): A Case with Ethiopian Students”. Journal of Information Science (2013).
- [15] “Data Mining: Prediction for Performance Improvement of Graduate Students using Classification” by Kamal Bunkar, Rajesh Bunkar, Umesh Kumar, Singh Bhupendra Pandya. IJCSI International Journal of Computer Science Issues, Vol. 7, Issue 1, No. 1, 2010.
- [16] “Predicting Academic Performance of Student Using Classification Techniques” by Sagardeep Roy, Anchal Garg. 2017 4th IEEE Uttar Pradesh Section International Conference on Electrical, Computer and Electronics (UPCON) GLA University, Mathura, Oct 26-28, 2017
- [17] “Student Performance Analysis System (SPAS) by Chew Li Sa, Dayang Hanani bt. Abang Ibrahim, Emmy Dahliana Hossain, Mohammad bin Hossin. International Journal of Emerging Technology and Advanced Engineering, Vol. 2, No. 7, 2012, pp.478-482.
- [18] “Analysis of students' performance evaluation using classification techniques” by V. Shanmuga rajeshwari, R. Lawrance. 2016 International Conference on Computing Technologies and Intelligent Data Engineering (ICCTIDE'16)
- [19] “Thrust Area in Data Science-Big Data and Data Analytics”,
- [20] “A Big Data Analytics - Challenges With-in New Data, Metadata Management & Analysis Platforms”, A National Conference on Innovative Applications and Research in Computer Science and Engineering (NCIARCSE-2017) at AGTI,s DACOE, Karad in association with International Advanced Research Journal in Science, Engineering and Technology (IARJSET) ISSN (Online) 2393-8021, ISSN (Print) 2394-1588, Vol. 4, Special Issue 4, January 2017.
- [21] “Challenges within Big Data and Big Data Analysis”, International Journal for Scientific Research & Development (IJSRD), Vol. 6 Issue 01, March 2018