

# Predicting Student Performance Using Machine Learning

Wable Nikita Arun<sup>1</sup>, Dhamdhare Rituja Vasant<sup>1</sup>, Bhor Ankita Bhanudas<sup>1</sup>, Prof.A.S.Kharat<sup>2</sup>

BE Student, Department of Computer Engineering, Jaihind College of Engineering, Kuran, Maharashtra, India<sup>1</sup>

Professor, Department of Computer Engineering, Jaihind College of Engineering, Kuran, Maharashtra, India<sup>2</sup>

**ABSTRACT**-Predicting Students performance beforehand can be very beneficial for educational institutions to improve their teaching quality. This paper proposes to predict students performance by considering their academic details. Educational organizations are unique and play utmost significant role for the development of any country. As Education transforms the lives of individuals, families, communities, societies, countries and ultimately the world! This is why we live comfortable lives today. Now a day's education is not limited to only the classroom teaching but it goes beyond that like Online Education System, Web-based Education System, Seminars, Workshops, MOOC course. becomes It's more challenging to Predict student's performance because of the huge bulks of data stored in the environments of Educational databases, Learning Management databases. Students' performance can be evaluated with the help of various available techniques.It is evolving area of study that emphasizes on various techniques like classification, prediction, feature selection. It is employed on learning records or data related to education to predict the students' performance and learning behavior by extracting the hidden knowledge.

**KEYWORDS:** College Education, Machine Learning, Result Prediction, supervised learning.

## I. INTRODUCTION

Today every educational institution handles and deals with large amount of student data which can be beneficial for a number of reasons. One of the important application of such data is predicting student performance. Such a prediction can be useful not only for the students but also for teachers/mentors. Mentors can provide special assistance to the students who are on the verge of failing. In order to determine which category a student lies, such data can be quite helpful. This application can be used by any prominent school or colleges. It can be used to predict the pointer ranges or percentage range for future semester exams. These ranges can be predicted using a number of data mining algorithms such as classification algorithms, rule-based algorithms, ensemble methods, and neural networks. The main aim of this project is the selection of features that show a strong relationship with a target attribute that is to be predicted from a high dimensional data-set.

### 1.1 Background

We have evaluated and compared the number of algorithms such as decision tree, random forest, support vector machine, naive Bayes and neural networks by applying them on the data-set The rest of the paper provides an explanation on nature of neural networks along with the results of our evaluation. Machine learning is used for analyzing data based on past experience and predicting future performance .Reinforcement machine learning algorithms is a branch of artificial intelligence .It automatically determines the behaviour of environment and maximizes its performance .In this algorithm agent is used for taking decision. With the help of agent best action is selected .It gives feedback by adapting the environment. It is an automated learning system .In this system we give input in the terms of marks and attendance. The intention of our system is to analyses and predict performance of student that will help to improve the marks of that student. In this paper this system is implemented for helping student to improve their academic performance.

### 1.2 Motivation

Advertisement act as a vital role in every industry and it helps in growth of the industry. The advertisement that are published are not seen properly by the user just because of insufficient clicking.We don't have any platform than can provide us the standard web-page depth that is visited by each and every user that logs in your website.

### 1.3 Problem Statement

In any form of higher education it is necessary to predict student's academic performance. There are two reasons for this: it is essential to identify which set of students would do well in semester end examination so that they can be awarded scholarships and more importantly to identify the students who may fail in semester end examinations viewed by user. The current system is maintaining academics records manually. Manual maintenance of records involves burden and it is quite tedious task. In general existing system there is no security.

## II. RELATED WORK

Thai Nghe et al [3] has used DT and BN classes of MLAs for predicting the undergraduate and post graduate results of two universities in Thailand. The total number of student records used for this prediction is 20492 and 932 respectively. Algorithms used for this prediction are C4.5, MSP and Naïve Bayes. They concluded that for all classes of predictions DT yields better results than BN by 3 to 12%. Resampling was used to improve the prediction accuracy.

Kotsiantis et al [4] described a model to predict student results for a distance learning course in Hellenic Open University. Predictions were done on the basis of marks obtained in written assignments. The algorithms used for this prediction are C4.5, Naïve Bayesian Network (NBN), Back Propagation (BP), 3-Nearest Neighborhood (3-NN) and Sequential Minimal Optimization (SMO). A set of 510 students of the university was chosen for experimental purpose. It was found that the NBN algorithm generates the best results (accuracy 72.48%).

Ramaswami et al [5] developed a predictive data mining model for student performance to identify the factors causing poor performance in higher secondary examination in TamilNadu. A data set for 772 students collected from regular students and school offices were used for this prediction. Algorithm used for this prediction is Chi-Square Automatic Interaction Detection (CHAID) DT. This tree was used to generate a set of decision rules used for predicting student grades. The overall prediction accuracy achieved was 44.69%.

Menaei-Bidgoli et al [6] applied data mining algorithms on "logged data" in a educational web based learning system. The system was tested with a data set of 227 students enrolled in a physics course in Michigan State University. Classification was initially performed using Quadratic BN, 1-NN, Prazen Window, Multilayer Perceptron (MLP) and C5.0 DT. It was seen that combining these classifiers increases prediction accuracy. Genetic Algorithms (GA) were further used to improve prediction accuracy by 10%.

Kovacic [7] explores the "socio-demographic" and "study environment" factors that results in student dropout in a polytechnic college in New Zealand. He uses student enrollment data like age, gender, ethnicity for this purpose. The total number of student records used for purpose was 450. Algorithms used for this prediction are CHAID and Classification and Regression Trees (CART). It was found that CART obtained a higher degree of accuracy (60.5%). Based on the results of Confusion Matrix and ROC curve he concluded that decision trees based on enrollment data alone are not sufficient to classify students accurately.

Karamohzis and Vrettos [8] have used ANN for predicting student graduation outcomes at Waubensee College. The prediction model was constructed from a profile of 1407 students of which 1100 were used for training and 307 were used for testing purpose. The average predictive efficiency for training and test sets were 77% and 68% respectively.

## III. PROPOSED METHOD

A machine learning Approach for tracking and predicting student performance in degree programs in this paper we were only tracking the degree programs and not predicting the overall performance of students and also they were using a data driven approach and clustering method based on probabilistic matrix factorization this is quite difficult to predict the student performance and ensemble learning algorithms were used which is hard to implement and complex data set is allowed. The proposed system automatically handles students data to predict their performance. System predict the defaulter and non-performing students.

### 3.1 Architecture

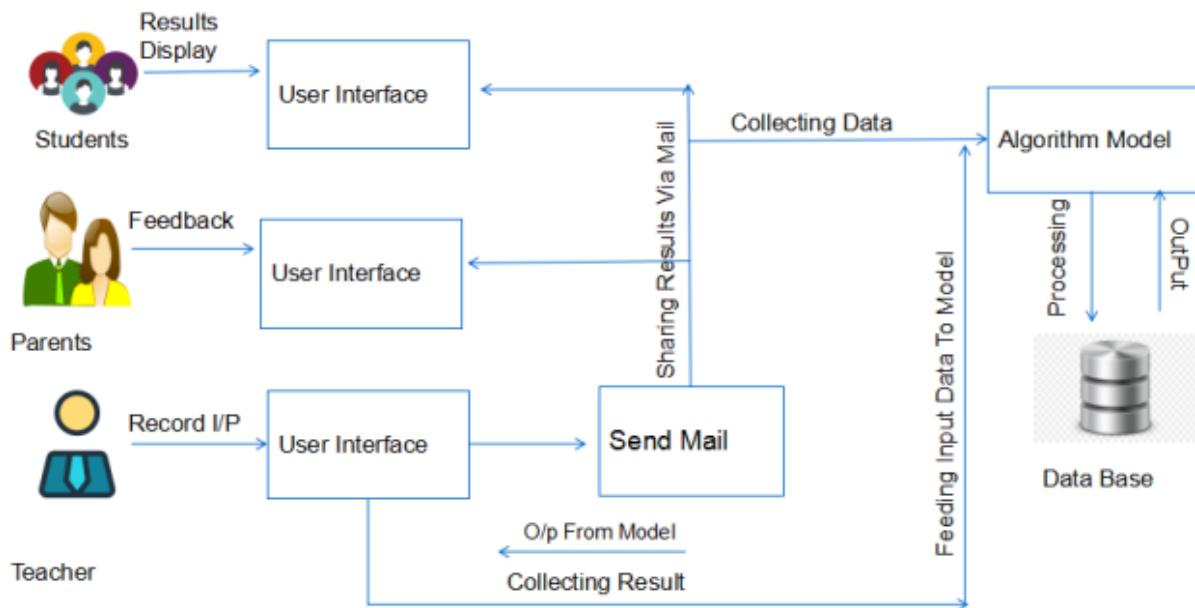


Fig. 1 System Architecture

### 3.2 Algorithms

Naive Bayes algorithm is the algorithm that learns the probability of an object with certain features belonging to a particular group/class. In short, it is a probabilistic classifier. The Naive Bayes algorithm is called 'naïve' because it makes the assumption that the occurrence of a certain feature is independent of the occurrence of other features. Here we classify the heart disease based on heart checkup attributes. Naive Bayes or Bayes' Rule is the basis for many machine learning and data mining methods. The rule (algorithm) is used to create models with predictive capabilities. It provides new ways of exploring and understanding data.

Why to prefer naive Bayes implementation:

- When the data is high.
- When the attributes are independent of each other.
- When we expect more efficient output, as compared to other methods output. Based on all these information and steps we classify to predict the heart disease depending on heart checkup attributes.

Steps:

1. Given training dataset D which consists of documents belonging to different class say Class A and Class B
2. Calculate the prior probability of class A = number of objects of class A / total number of objects Calculate the prior probability of class B = number of objects of class B / total number of objects
3. Find NI, the total no of frequency of each class Na = the total no of frequency of class A Nb = the total no of frequency of class B
4. Find conditional probability of keyword occurrence given a class:

$$P(\text{value 1/Class A}) = \text{count}/n_i(A) \quad P(\text{value 1/Class B}) = \text{count}/n_i(B) \quad P(\text{value 2/Class A}) = \text{count}/n_i(A)$$

$P(\text{value } 2/\text{Class B}) = \text{count}/n_i(B)$  .....

.....

.....

$P(\text{value } n/\text{Class B}) = \text{count}/n_i(B)$

5. Avoid zero frequency problems by applying uniform distribution

6. Classify Document C based on the probability  $p(C/W)$  a. Find  $P(A/W) = P(A) * P(\text{value } 1/\text{Class A}) * P(\text{value } 2/\text{Class A}) \dots P(\text{value } n/\text{Class A})$  b. Find  $P(B/W) = P(B) * P(\text{value } 1/\text{Class B}) * P(\text{value } 2/\text{Class B}) \dots P(\text{value } n/\text{Class B})$

7. Assign document to class that has higher probability.

### 3.3 Mathematical Model

Mathematical equation:

The algorithm implemented in this project is describe as: Algorithm  $P(\text{class/features}) = P(\text{class}) * P(\text{features/class})P(\text{features})$

- $P(\text{class/features})$  : Posterior Probability
- $P(\text{class})$  : Class Prior Probability
- $P(\text{features/class})$  : Likelihood
- $P(\text{features})$  : Predictor Prior Probability A. Normal distribution Normal distribution The probability density of the normal distribution is:  $f(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$  (1) Where
- 'μ' is the mean or expectation of the distribution,
- 'σ' is the standard deviation, and
- 'σ<sup>2</sup>' is the variance.

### IV. CONCLUSION

Machine learning techniques can be useful in the field of students performance prediction considering that they helps to identify from the beginning of academic year. The aim of this paper is to apply machine learning algorithms for prediction of student performance. An early analysis of student having poor performance helps the management take timely action to improve their performance through predicting their academic details. Accurately predicting student performance based on their ongoing academic records is predicted. Also we conclude that proposed system is helping us to make the student performance better. In this paper machine learning can prove to be powerful tool and all algorithms we used increases with increase in dataset size.

### REFERENCES

- [1] Castro, F., Vellido, T., Àngela Nebot, and Mugica F., Applying Data Mining Techniques to e-Learning Problems.
- [2] Romero, C. and Ventura, S. 2007, Educational data mining: A survey from 1995 to 2005, Expert Systems with Applications, 135-146.
- [3] Kotsiantis, S., Piarrekeas, C., Pintelas, P., 2007. Predicting Students' performance in Distance Learning using Machine Learning Techniques, Applied Artificial Intelligence, 18:411-426.
- [4] Nguyen, N. N., Janeck, P, Haddawy, P., 2007, A Comparative Analysis of Techniques for Predicting Academic Performance, 37th ASEE/IEEE Frontiers in Education Conference.



- [5] Ramaswami, M., Bhaskaran, R.,2010, A CHAID Based Performance Prediction Model in Educational Data Mining, International Journal of Computer Science Issues, Vol. 7, Issue 1, No. 1.
- [6] Minei-Bidgoli, B., Kashy, D., Kortemeyer G., Punch W F, 2003. Predicting Student Performance: An Application of Data Mining Methods with the Educational Web-Based System LON-CAPA, 33rd ASEE/IEEE Frontiers in Education Conference,.
- [7] Kovacic, Z. J.,2007, Early Prediction of Student Success: Mining Students Enrolment Data, Informing Science &IT Education Conference .online and no one sees it, is it still an ad?" Journal of AdvertisingResearch, 2013.
- [8] Livieris, E., Drakopoulou, E., Pintelas, P., Predicting students' performance using artificial neural networks.
- [9] Stefanowski, J., An Experimental Study of Methods Combining Multiple Classifiers - Diversified both by Feature Selection and Bootstrap Sampling.
- [10] Zhao, Y., and Zhang, Y.,Comparison of decision tree methods for finding active objects.
- [11] Acharya, A., Sinha, D., 2014, Application of Feature Selection Methods in Educational Data Mining , International Journal of Computer Applications 103(2):34-38.