



e-ISSN: 2319-8753 | p-ISSN: 2320-6710

# IJRSET

International Journal of Innovative Research in  
**SCIENCE | ENGINEERING | TECHNOLOGY**



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

**Volume 9, Issue 9, September 2020**

**ISSN** INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA

**Impact Factor: 7.512**

 9940 572 462

 6381 907 438

 [ijirset@gmail.com](mailto:ijirset@gmail.com)

 [www.ijirset.com](http://www.ijirset.com)

# Predicting Student’s Performance using Data Mining Techniques

Akanksha Gujar<sup>1</sup>, Pranay Katariya<sup>2</sup>, Rida Kazi<sup>3</sup>, Krupali Mahale<sup>4</sup>, Mahendra Nivangune<sup>5</sup>

Department of Computer Engineering, Sinhgad Academy of Engineering, Pune, India<sup>1,2,3,4,5</sup>

**ABSTRACT:** A country’s growth is largely dependent on the quality of education system. To improve this system and to measure the capability or success rate of students, the prediction of student performance plays a vital role. An early detection of the weaker students will help the college management to look after these students and provide proper guidance and counselling to them. In the proposed system we will build a performance predicting model using data mining technique. The students take a series of test and on the basis of their results their performance is generated. Naïve Bayes and Random forest Algorithms are two data mining techniques used in this paper. Naïve Bayes is an algorithm based on the Bayes theorem, which is used for the classification tasks. Random Forest is also used for classification and regression. Lastly, some e-books and practice papers are given to the students for the future development of the student.

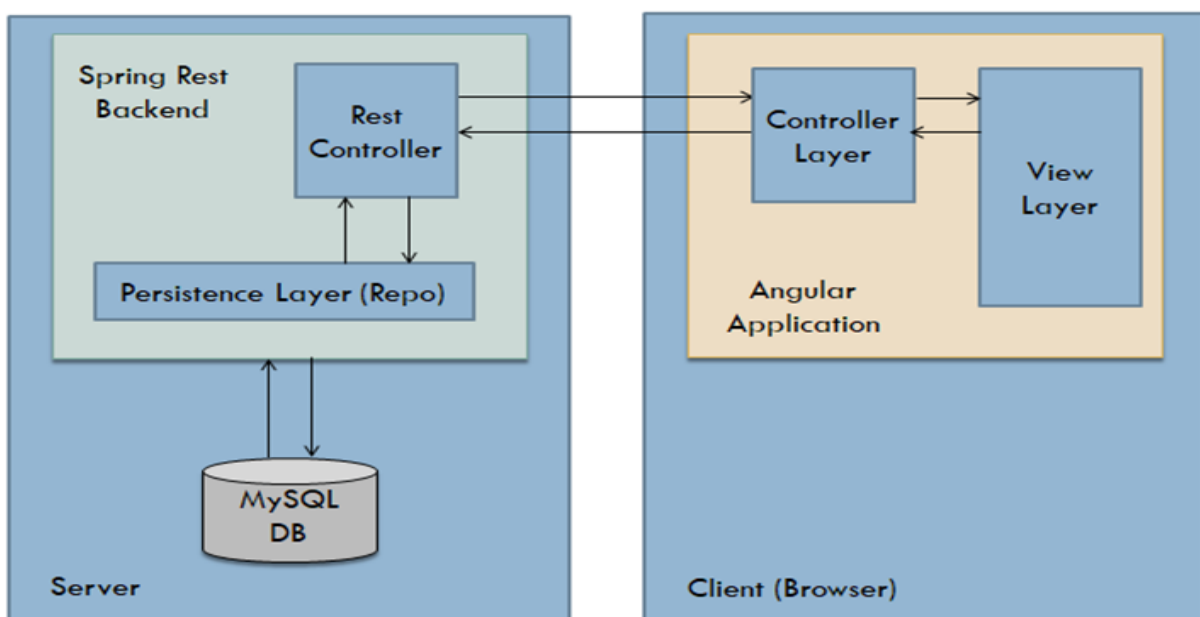
**KEYWORDS:** PREDICTION, PERFORMANCE, NAÏVE BAYES, RANDOM FOREST.

## I. INTRODUCTION

All students have different learning power and motivation. Student’s performance is an important parameter in deciding their future. It is also essential for the success of the institution. Data mining applied in educational area is called as Educational Data mining (EDM). This paper uses the Naïve Bayes algorithm and Random Forest algorithm for the prediction of the students’ performance. The proposed system will help in increasing the quality of students, their personality and will prepare them for various placement activities as well. At the same time the weaker students can be given timely-help for improvement.

## II. PROPOSED WORK

### A. ARCHITECTURE



Angular applications are ideally suited for being built as single page web apps. The main idea compared to other more traditional server-side architectures is to build the server as a set of stateless reusable REST services, and from an MVC perspective to take the controller out of the backend and move it into the browser.

The client is MVC (Model-View-Controller) capable and contains all the presentation logic which is separated in a view layer, a controller layer, and a frontend services layer. When angular application is running only JSON (Javascript Object Notation) data interchanged between client and server

## B. DESIGN

### 1. How is the backend built?

Backend System is one complete application developed in spring boot which has REST (Representational State Transfer) endpoints. Our Backend System is connected to the database through JPA (Java Persistence API).

Spring boot uses Spring MVC in the background to build REST API. MVC stands for Model View Controller, though we don't use View when building REST API. We can't neglect Model and Controller.

Our Backend service provide specific API service on specific URL which uses famous HTTP methods like GET, POST, etc. Our front-end application request backend in JSON data interchange format and backend API delivers response in JSON only.

Backend System is totally developed in spring framework which follows MVC architecture. Here in our REST endpoints, first request goes to Rest Controller (class which is annotated with @RestController) then specific method is mapped according the request and then for data controller invokes the repository which is responsible to give data from the database according to model class and then data is returned to the controller class and served as JSON response over the web then it is the responsibility of front end in our case angular application to consume that data.

### 2. How is the frontend built?

Front end of PSP application is developed in angular. Angular follows MVC architecture. Angular application interacts with API developed in spring boot and get the required data by doing authentication to the REST API on the web through HTTP. Then it is responsibility of angular to take that data and create view to the end user accordingly. Angular uses component-based development process. Components and Routing are the main feature of angular which is used to achieve the concept of single page application.

#### 2.1 The View Layer

The view layer is composed of HTML component, CSS component and any Angular directives representing the different UI components.

#### 2.2 The Controller Layer

Component.ts file is used in controller layer which is totally written in TypeScript language which is superset of JavaScript.

Controller layer is responsible to interact with backend system and fetch the JSON data using HTTP request and popular methods like GET, POST, etc.

One of the main responsibilities of the controller is to perform frontend validations. Any validations done on the frontend are for user convenience only - for example they are useful to immediately inform the user that a field is required.

#### 2.3 Web Application overall design

Web application named as PSP - Predicting Student Performance developed using Spring framework and Angular.

Main purpose of developing PSP is to provide student with a portal to help them understand their aptitude and overall soft skills and to provide study material to them to improve their skills. PSP has different modules integrated in single page applications and can be reached to these different modules using angular routing. Main modules of application are Communication test, Presentation test, Aptitude test and Technical test. Final result of logged in user will get calculated after user gives all the test. When all users gives the test and test result is stored in the database then Naive bayes algorithm will be trained on that data set which is stored in database and after training of algorithm it will predict the next test result of the user and based on prediction proper study material will be provided to the user to study and improve again to appear in next test and perform well.

After all this report for all students will be generated in the system and communicated to the students from their login id and through emails.

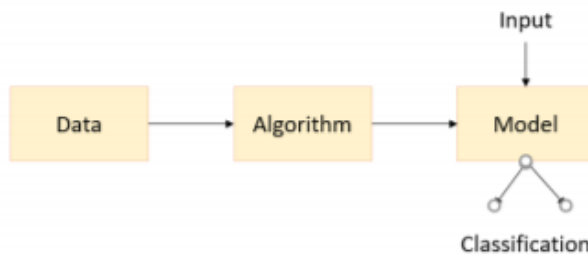
### III. CLASSIFICATION ALGORITHMS

Classification is a data mining technique mostly used for predictive data mining tasks. Classification technique is used to classify data into predefined classes. Classification technique has Naive Bayes, Random Forest Tree, Decision Trees etc different classifiers to classify data are compared and the best one is used to extract rules for the prediction of students' performance.

#### A. Naive Bayes Classifier

Naive Bayes is probabilistic machine learning algorithm based on the Bayes theorem used in a wide variety of classification task. Since it is a probabilistic model the algorithm can be easily implemented and the prediction is also made really quick. It is really scalable and is algorithm of choice for the real-world applications that are required to respond to users request instantaneously.

Attributes for Naive Bayes algorithm to develop model will be major affecting factors in aptitude and logical thinking like 10th result, mathematics marks in 10th, 12th result, mathematics mark in 12th, CGPA in UG. Based on above attributed model will be developed.



Mathematics Behind Naive Bayes Classifier:

Bayes theorem provides a way of calculating Posterior probability  $P(c/x)$  from  $P(c)$ ,  $P(x)$  and  $P(x/c)$ . Consider the following equation:

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

Likelihood
Class Prior Probability  
Posterior Probability
Predictor Prior Probability

---


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

Here,  $P(c/x)$ : posterior probability of class(c,target) given predictor(x,attributes). This represents the probability of c being true, provided x is true.

$P(c)$ : is the prior probability of class. This is the observed probability of class out of all the observations.

$P(x/c)$ : is the likelihood which is the probability of predictor-given class. This represents the probability of x being true, provided x is true.

$P(x)$ : is the prior probability of predictor. This is the observed probability of predictor out of all the observations.

#### B. Random Forest Algorithm

Random forest algorithm is supervised learning technique classification algorithm. Algorithm internally creates the forest of decision trees.

If there are more trees in the forest the more reliable the forest looks like. In the same way in the random forest classifier algorithm, maximum the number of decision trees in the forest gives the high accuracy results for classification.

Why Random Forest Algorithm:

1. Random forest classifier will handle the missing values of attributes.
2. When we have highest number of decision trees in the forest, random forest classifier won't overfit the model.

3. Can model the random forest classifier for categorical values.

**IV. COMPARISON BETWEEN RANDOM FOREST TREE AND NAIVE BAYES ALGORITHMS**

To compare the performance of Naïve Bayes and Random Forest Tree algorithms, a sample Student marks data set is considered.

Comparison of the Naïve Bayes and Random Forest Tree algorithm shows that the Random Forest Tree algorithm generates more optimized solution than the Naive Bayes algorithm. Better classification rules are provided by the Random Forest Tree algorithm.

In the proposed system to predict the performance of students, Naïve Bayes and Random Forest Tree algorithm is applied on the student database. The database consists of one thousand ten student’s information based on which classification is done. Classification accuracy given by the algorithms is shown in table below. Random Forest Tree algorithm has given more classification accuracy.

Algorithm	Total Students	Accuracy based on Training Data	Accuracy based on Testing Data
Naïve Bayes	1010	0.8342776203966006	0.8514851485148515
Random Forest	1010	1.0	0.858085808580858

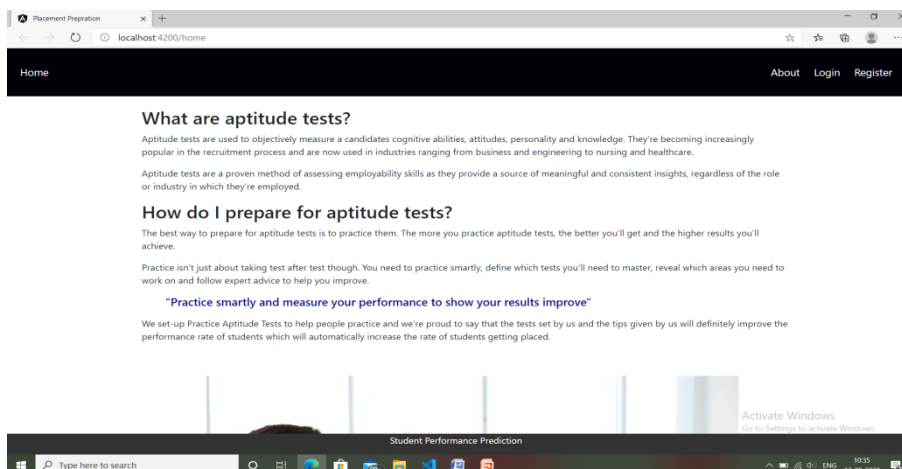
**V. EXPERIMENTAL RESULTS**

Home page is an introductory page for the user. Fig(1) shows the home page of the SPPS system. There are two forms used in this system one is login form that stores the login information of the user and other is student registration form that stores the information entered by the student.

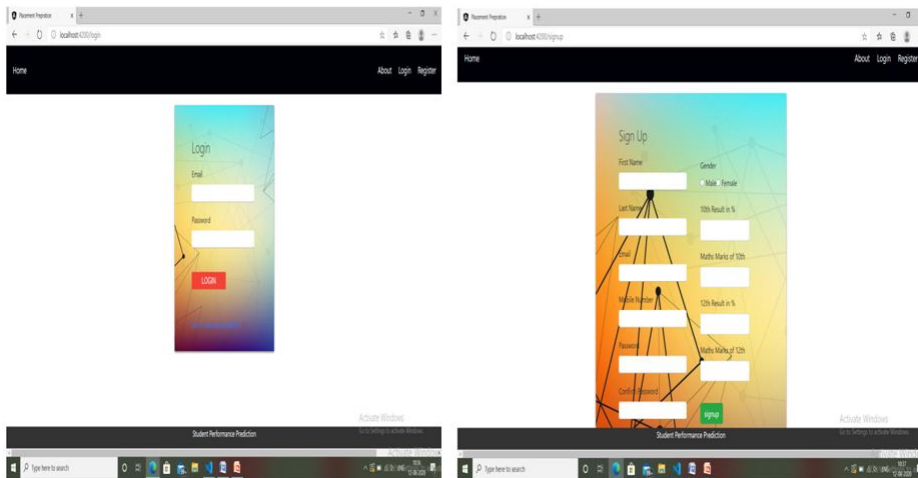
Username and Password are the security parameters to grant access to the SPPS system as shown in Fig(2). After login System redirects user to the next form to enter personal information shown in Fig(3). Student fills his first name, last name, email address, mobile number, password, confirm password, gender and previous academic marks. On submission of this entire information SPPS system redirects to instruction page as shown in Fig(4).

Figure(2)Login Page to access SPPS system, (3) Form to enter personal information of the student, (4) Instructions for taking test (5) Test consists of the communication test form (6) Shows the writing test form (7) Shows the aptitude test form (8) Shows the technical test form (9) After giving the test shows the output Result of test given by student based on the rules extracted from the Random Forest Tree and Naïve Bayes algorithms. (10) Shows tips for improving performance of student.

Figure (a),(b),(c) shows the admin login page, admin page and student result to the admin respectively.

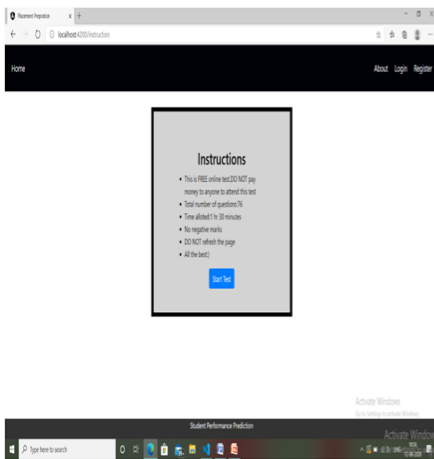


Fig(1)

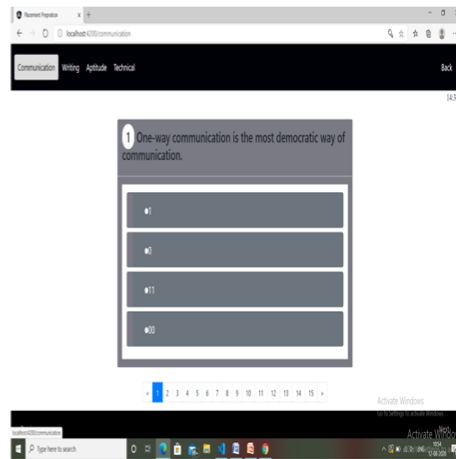


Fig(2)

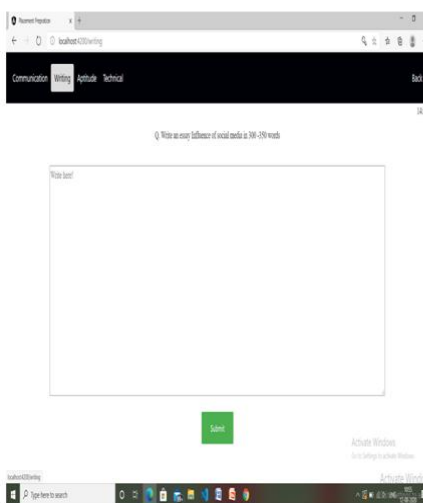
Fig(3)



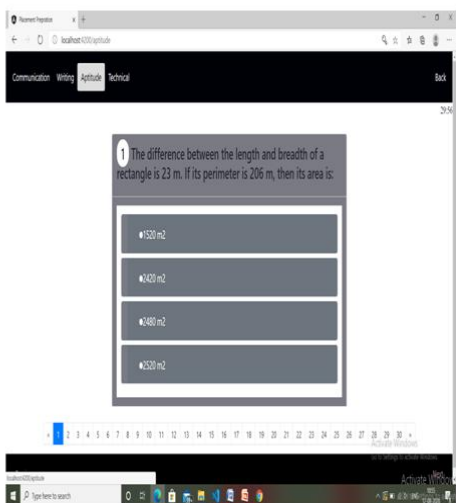
Fig(4)



Fig(5)



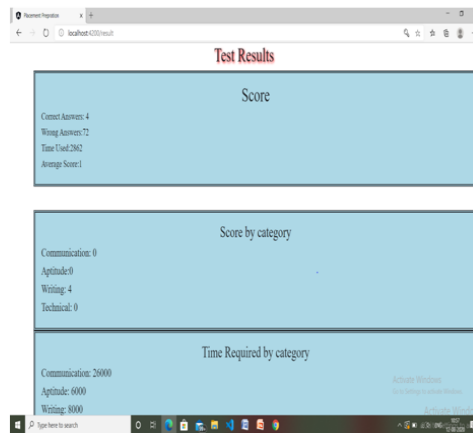
Fig(6)



Fig(7)



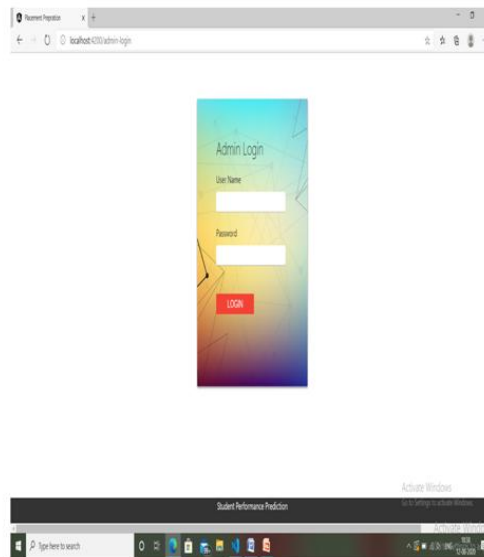
Fig(8)



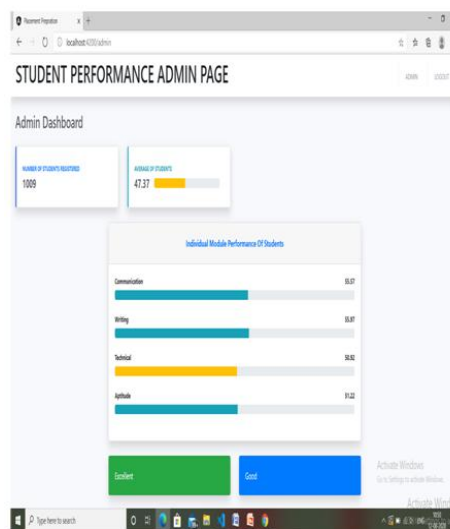
Fig(9)



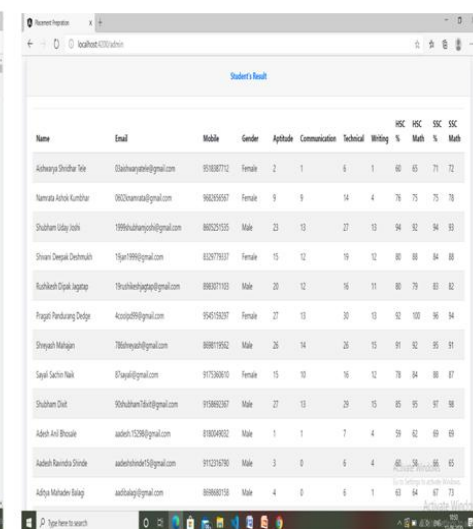
Fig(10)



Fig(a)



Fig(b)



Fig(c)



## VI. CONCLUSION

In this paper, classification algorithms such as Naïve Bayes and Random Forest algorithm are employed on the student test results and their performance is evaluated. Comparison of Naïve Bayes and Random Forest Tree algorithms is carried out. Random Forest Tree algorithm gives better performance as compared to Naïve Bayes algorithm. This study will help both the college and the students to boost the performance. Special attention can then be given to the weaker students.

## VII. FUTURE SCOPE

Various different factors such as academic performance, extracurricular activities can be considered for the evaluation of the students' performance. Physiological thinking, social network interaction and Family background can also be the additional factors for deciding their performance. People with disability should have different factors to predict their performance. Demographics such as Gender can also be considered as the learning pattern and styles is different between males and females.

## REFERENCES

- [1] N.V. Krishna Rao, Dr. N Mangathayaru, Dr. M. Sreenivasa Rao, "Evolution and Prediction of Radical MultiDimensional E-Learning System with Cluster based Data Mining Techniques", International Conference on Trends in Electronics and Informatics ICEI 2017, 978-1-5090-4257-9/17/\$31.00 ©2017 IEEE
- [2] Febrianti Widyahastuti, Viany Utami Tjhin, "Predicting Students Performance in Final Examination using Linear Regression and Multilayer Perceptron", 978-1-5090-4688-1/17/\$31.00 ©2017 IEEE
- [3] Badr HSSINA, Abdelkarim MERBOUHA, Hanane EZZIKOURI, Mohammed ERRITALI TIAD, "A comparative study of decision tree ID3 and C4.5", (IJACSA) International Journal of Advanced Computer Science and Applications, Special Issue on Advances in Vehicular Ad Hoc Networking and Applications
- [4] Dr. Anjali B Raut, Ms. Ankita A Nichat, "Students Performance Prediction Using Decision Tree Technique", International Journal of Computational Intelligence Research ISSN 0973-1873 Volume 13, Number 7 (2017), pp. 1735-1741 <http://www.ripublication.com>
- [5] Prabhjot Kaur, Williamjeet Singh, "Implementation of Student SGPA Prediction System (SSPS) Using Optimal Selection of Classification Algorithm"
- [6] Abhishek Kumar Jha, Manasa U Hegde, Animesh Giri, "A Composer System based on Meta-Learning for Student Performance Prediction", 3rd IEEE International Conference on Computational Systems and Information Technology for Sustainable Solutions 2018, ISBN: 978-1-5386-6078-2 © 2018 IEEE





**INNO**  **SPACE**  
SJIF Scientific Journal Impact Factor

**Impact Factor:**  
**7.512**

**ISSN** INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
**INDIA**



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 **9940 572 462**  **6381 907 438**  **ijirset@gmail.com**



[www.ijirset.com](http://www.ijirset.com)

Scan to save the contact details