

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 11, November 2019

Predicting Student's Performance using Data Mining Techniques

Akanksha Gujar¹, Pranay Katariya², Rida Kazi³, Krupali Mahale⁴, Mahendra Nivangune⁵

Department of Computer Engineering, Sinhgad Academy of Engineering, Pune, India^{1,2,3,4,5}

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. RELATED WORK

N.V. Krishna Rao, Dr. N Mangathayaru Professor and Dr. M. Sreenivasa Rao conducted the student performance evaluation which is very important for performance enhancement which plays vital role in socioeconomic of the country. The presented work gives an idea about student performance evaluation prediction algorithms. A Learning algorithm is implemented by allowing student data to store in clustered memory for repeated querying. A Clustering technique boosts the searching process speed and analysis. Clustering is done to access the vital student data from big data and process it with amazing speed. Classification technique with decision tree algorithm is implemented to find the student performance. Various existing classification algorithms are discussed in this paper. The accuracy levels of various techniques are compared in this work.[1]

FebriantiWidyahastuti, ViannyUtamiTjhin studied the prediction of student's performance in final examination by applying linear regression and multilayer perceptron in WEKA- in terms of accuracy, performance and error rate- to compare their feasibility. The basis of data was derived from extraction and analysis of e-learning logged-post in discussion forum and attendance. Final examination has been predicted using two classification algorithms namely linear regression and multilayer perceptron and, then, are compared the greatness between the two based on the value of mean absolute error difference. The result showed that multilayer perceptron gives good prediction results than linear regression. This study contributed to the students to improve their learning experiences through online discussion forums.[2]

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 11, November 2019

Badr HSSINA, Abdelkarim MERBOUHA, Hanane EZZIKOURI, Mohammed ERRITALI studied Decision tree method which is generally used for the Classification, because it has simple hierarchical structure for the user understanding & decision making. ID3 and C4.5 algorithms have been introduced by J.R Quinlan which produce reasonable decision trees. The objective of this paper was to present these algorithms. In this paper they presented the classical algorithm that is ID3, and details of C4.5, which is natural extension of the ID3 algorithm and comparison between these two algorithms and others algorithms such as C5.0 and CART. They have focused on the key elements of the construction from a set of data, and presented the ID3 and C4.5 algorithms that responded to these specifications. Comparison of ID3/C4.5, C4.5/C5.0 and C5.0/CART is done, which led to confirm that the most powerful and preferred method in machine learning is certainly C4.5.[3]

Dr. Anjali B Raut, Ms. Ankita A Nichat analyzed data mining methods and techniques students' data to construct a predictive model for students' performance prediction. They focused on measuring student performance using classification technique such as decision tree. They also focused on classification techniques which are used to analyze performance by the scope of knowledge. This paper focused on C4.5 decision making algorithm for predicting students performance. Given the details about the results, and the specific needs of studies for improvement, such as the accompaniment of students along their learning process, and the taking of timely decisions in order to prevent academic risk and desertion. Lastly, some recommendations and thoughts are laid out for the future development of performance are discussed.[4]

Prabhjot Kaur, Williamjeet Singh developed the Student SGPA Prediction System (SSPS) which uses rules extracted from the best algorithm among J48, LMT, Random Tree and REP Tree algorithms to predict SGPA of students in first six semesters. These four classification algorithms are compared by building student performance prediction model based on student's social conditions and previous academic performance using WEKA. The records of 236 computer engineering students at Punjabi University are used to build these models. REP Tree algorithm with average accuracy and minimum average error rate is found to be better than the J48, Random Tree and LMT algorithms. This resulted in predicting performance of students with amazing speed.[5]

Abhishek Kumar Jha, Manasa U Hegde, Animesh Giri endeavored to augment the efficiency and reliability of the student performance prediction system and other analogous intelligence systems in Educational Data mining, by merging of multiple algorithms used for the purpose. Meta Learning and Ensemble methods are utilized to improve the combined outcome in a multi class system. This paper illustrates the potential of a meta-learning model in educational data mining with exemplification of multiple algorithms, for manifold classes, and their accuracy over different test set of data with cross validation. The Main motivation for combining multiple base models is to reduce the chances of misclassification based on a single algorithm, by increasing systems area of expertise, which reduces the bias and variation caused by any of the poor base algorithm. This approach proposed the enhancement of the performance by adding a additional layer to conventional Meta-Learning Technique like Bagging, by using a Model to learn and then predict the outcome, instead of using a max-argument voting as there are cases where dirty data, empty values or a small set of base learners can cause most of the base prediction to be wrong. Here the efficiency of performance prediction model is increased but the computational cost of the model is high.[6]

International Journal of Innovative Research in Science, Engineering and Technology

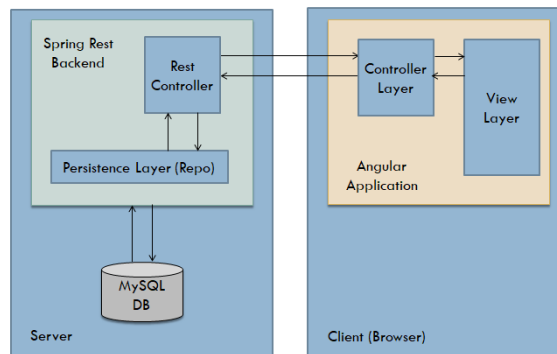
(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 11, November 2019

III. 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 interchange 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.

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 11, November 2019

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.

IV. MATHEMATICAL MODEL

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.

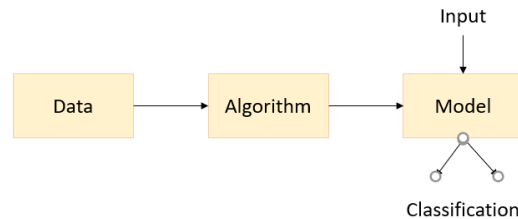
Attributes for Naive Bayes algorithm to develop model will be major affecting factors in aptitude and logical thinking like 10th result, mathematics marks in 10th, 12th result, mathematics mark in 12th, CGPA in UG. Based on above attributed model will be developed.

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 11, November 2019



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.

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 11, November 2019

V. 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. It is expected that the Naïve Bayes algorithm outperforms the Random Forest 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.

VI. 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