

# Predicting Student's Performance For Early Prevention using Supervised Machine Learning Algorithm

Shubha Pawar<sup>1</sup>, Omkar Jadhav<sup>2</sup>, Shubham Kadam<sup>3</sup>, Pratik Jadhav<sup>4</sup>

Prof. Sneha Deshmukh<sup>5</sup>

<sup>1-4</sup>Student, Department of Computer Engineering, DPCOE - Dhole Patil College of Engineering, Pune, Maharashtra, India

<sup>5</sup>Assistant Professor, Department of Computer Engineering, DPCOE - Dhole Patil College of Engineering, Pune, Maharashtra, India

\*\*\*

**Abstract** - Nowadays educational institutions are trying to improve quality of education and also the approach of the student towards the examinations. In this Research, we are trying to find out student's current status and predict his/her future results. By predicting the score before the exams, it will benefit for both students as well as the teachers. A teacher can guide a student to cope up with their weakness and enhancement in their academic to scores good marks. Even when student knows the marks before exams, he himself can also analyze in which area he needs to work. The aim is helping the students for the all-round performance using AI. Techniques used in this are Decision tree, SVM, Naïve bayes. It is noted that from all of these techniques SVM produces better results. The highest accuracy with SVM which occurred was about 87.26%.

**Key Words:** Predicting Score, AI, Decision tree, SVM, Academic, Examinations, Analyse.

## 1. INTRODUCTION

For any country/nation students are their future. Students play a vital role in grooming a nation to the next level. Every student aims that his role should also be there in developing his nation. A nation's development is totally based on quality of education the student poses. That's the reason Education plays a vital role in a student's life. Education works on shaping the life of student. In the entire life of a student, they undergo through various examination at every point of their life. The exams are taken to test student's mentality, his approach towards the studies, thinking capability, problem facing approach. But because there is no linking between practical study and theoretical studies students face issues in understanding the concepts and they either take

burden on themselves or don't study at that time. But when exams approach, they get frustrated and don't know what to study and how to pass the exams. Due to which many students' scoreless marks or even can't pass the exam, and because of this many students think that everything is over and they give up their life and out of which some of them goes under depression. So to overcome all this negativity our module comes into the picture. A teacher should know before a student lags behind. So, we decided how to predict the student performance using previous academic data by AI. If teacher contain knowledge of students result before the exams, teacher can take the necessary steps to improve the students' performance. So primarily we will predict student performance in this research according to past report.

## 2. PROPOSED METHODOLOGY

Our proposed methodology started with gathering data set. So, we tried to collect student's class test marks, assignment marks, presentation marks, attendance marks, midterm marks, and final examination marks.

### 2.1 Dataset

The data used for prediction of student's performance is used from UCI Machine Learning repository. The data mainly comprises of performance repository. The data mainly comprises of performance of the student's achievement in secondary education of Portuguese school. Mainly there are two datasets provided which consist of performance in Math's and Portuguese subjects respectively. There are overall 649 records in Math's and Portuguese subjects combined.

## 2.2 Modifying and Preparing Dataset

We are using the Dataset that is the csv file which consist of student’s performance. So, there are two dataset we first merge this dataset to single dataset. The next step is cleaning the dataset i.e., handling the missing values, removing duplicate records, handling of the categorical variables& so on. But in our case the dataset is already cleaned. For Predicting the performance, we have added one more column final grade which is consist of the performance of student like poor, average and good by converting G3 grade. Then the unnecessary columns are dropped from the dataset while identifying through data visualizations and then finally converting the categorical variables to numerical variable and the dataset are finally ready for training.

## 2.3 Model Used

The main purpose is to evaluate student’s performance and see how the different features has impact on it. In this paper we have used some traditional classification methods. In this paper we have used different classification algorithms like SVM, Naïve Bayes, Logistic Regression, Decision Tree, KNN and Random Forest Classifier. SVM (Support Machines) classifier is one of the best methods in classification. One of the major advantages of SVM’s are its effective in high dimensions spaces and it selects the best plane which is suitable for classification. Gaussian Naïve Bayes Classifier is classification which is Bayesian theorem in probability and majorly works for independent features. Logistics Regression are easier for implementation, interpret and very efficient in training. Decision tree classifiers are one of the best is used to classify the data in different classes it is also supervised learning algorithm. So, in this paper KNN algorithm will take certain input features and classify it among 3 classes i.e., poor, average and good performances. Random Forest Classifier is also one of the supervised machine learning which is also an ensemble learning method for classification. The forest ensemble decision tree using a bagging method It also tries to increase the overall performance.

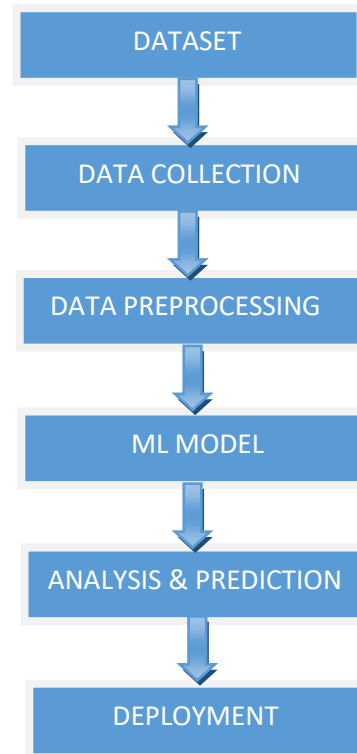


Fig.1. System Architecture

## 3. MODEL PERFORMANCE

Our model got 87% using SVM Classifier and 85% accuracy using Random Forest Classifier accuracy on predicting student performance. By analyzing the .

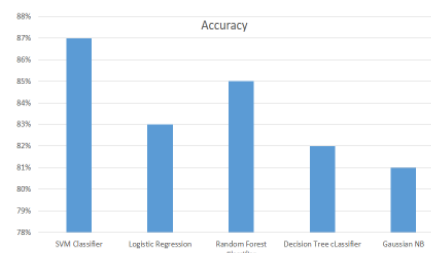


Fig.2. Chart of Accuracy in different algorithms.

In the above figure, we can see the accuracy rate of different machine learning algorithms when we have implemented the model. It gives the accuracy rate of more than 80% in different algorithms like

SVM Classifier, Logistic Regression, Random Forest Classifier, Decision Tree Classifier & Gaussian NB.

subject. So ultimately more the students absent less is the score.

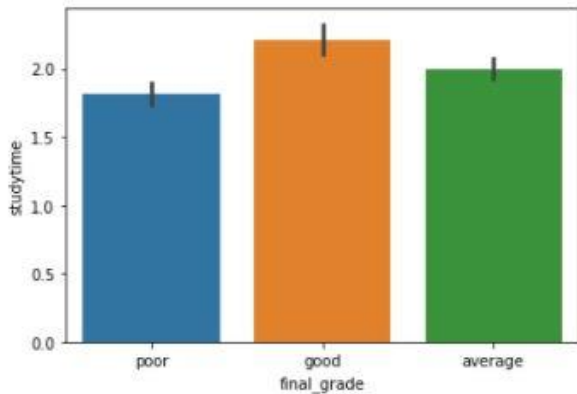


Fig.3. Categorized study time according to their grades.

In figure-3, we have classified the study-time according to classes i.e., poor, average, good. Those who study more are getting good grades while those study less are having poor score so this impact the overall performance of the students.

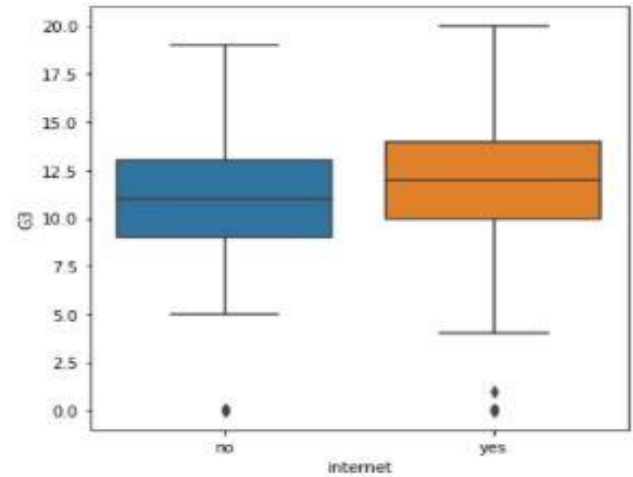


Fig.5. Categorized Internet time.

As we can see in fig-5 that those students who have internet has been having slightly greater grades than those who are not having. Internet also provides online resources to learn new things. So, the students are utilizing the internet properly.

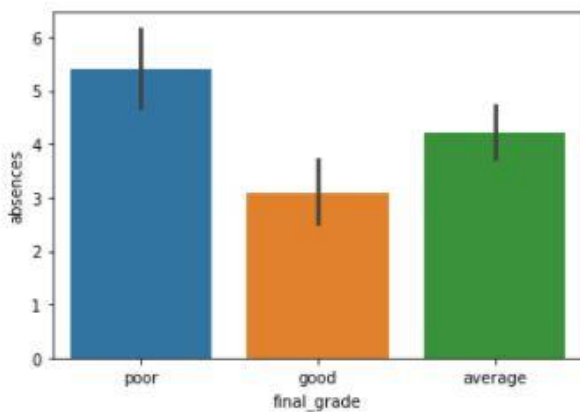


Fig.4. Categorized absences according to their grade

In the fig-4 we have classified absences according to classes i.e., poor, average, good. Attending classes plays an important role in enhancing the student's performance. So more you attend the classes, the more you learn about the respective

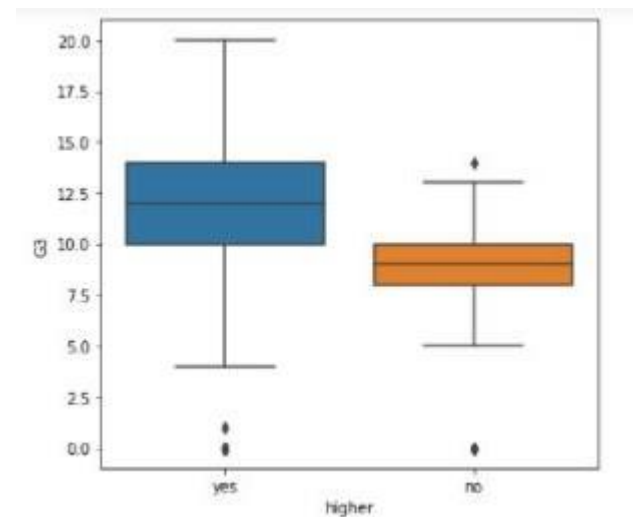


Fig.6. Categorized according to higher studies

As we can see in fig-6 that those students who have done their higher studies has been having slightly greater grades than those who are not having their higher studies. Higher studies will allow you to understand the concepts in well manner.

#### 4. CONCLUSIONS

Predicting student's academic performance is of great concern to the education institutes. It helps us to identify abilities of students, interest and their weaknesses. Students' performance can be influenced by many different types. Here we have used different classification algorithms like SVM, Naïve Bayes, Logistic Regression, Decision Tree, KNN and Random Forest Classifier. Out of all the SVM gave the highest accuracy of 87.26%. Previous years data plays an important role in predicting the final score.

Moreover, we want to increase our data sets in the future and also, we will focus on the accuracy and the result. We have the plan to revise our Model in the future and want to include some more extra features in some academic and industry purposes. We can implement the model in some company to predict or analyze employees' performance annually or monthly and by this the company can warn the employees about their drawbacks to improve company's productivity. Also, this research will help us in many fields. This could be alert for the students, teachers and also for the parents who are very concern about their daughter's/son's academic establishment.

#### 5. REFERENCES

- 1) Exawatt Tanura and Yaya Hayride, "Using Machine Learning Techniques to Earlier Predict Student's Performance", IEEE, 2018.
- 2) Rajitha Ravikumar, "Early Alert System for Detection of At-Risk Student's", IEEE, 2018.
- 3) H.M Rafi Hasan and Mohammed Islam, "Machine Learning Algorithm for Student's Performance Prediction", IEEE, 2019.
- 4) Has bur Rahman, "Predict Student's Academic Performance and Evaluate the Impact of Different Attributes on the Performance Using Data Mining Techniques", ICEEE, 2017.
- 5) Saga deep Roy and Panchal Garg, "Predict Academic Performance of Student Using Classification Techniques", IEEE, 2017.
- 6) Raza Hasan and Sella pan, "Student Academic Performance Prediction by using Decision Tree Algorithm", ICCOINS, 2018.
- 7) It Burman and Subanal Some, "Predicting Student's Academic Performance Using Support Vector Machine", IEEE, 2019.