

## **ASSIGNMENT NO. 4**

**TITLE:** Student Grade prediction

**AIM:** Implement an application for a selected problem statement using a suitable front-end language such as Python, HTML, CSS.

**OBJECTIVE:** To develop a mini project using suitable technology to provide a solution to the selected problem statement.

**OUTCOME:** Develop an application using suitable technology to provide a solution to the selected problem statement.

**A PRELIMINARY MINI PROJECT REPORT ON**  
**“Students Grade Predictor ”**

**SUBMITTED TOWARDS THE PARTIAL FULFILLMENT OF THE  
REQUIREMENTS OF**

**BACHELOR OF ENGINEERING (T.Y. B. Tech.)**

**Academic Year: 2022-23**

**By**

<b>PRN no.</b>	<b>Name</b>	<b>Roll no.</b>
<b>120B1B316</b>	<b>Naput Labde</b>	<b>TYCOD316</b>
<b>120B1B307</b>	<b>Siddhesh Choudhari</b>	<b>TYCOD307</b>
<b>120B1B312</b>	<b>Maithili Sude</b>	<b>TYCOD312</b>

**Under The Guidance of**  
**Prof. Shailesh Galande**



**DEPARTMENT OF COMPUTER ENGINEERING,**  
**PIMPRI CHINCHWAD COLLEGE OF ENGINEERING**  
**SECTOR 26, NIGDI, PRADHIKARAN**

**Abstract:**

Students have the freedom to learn anything they choose in the world of open educational systems because the learning materials are readily accessible. However, this resource may encourage complacency in students. As a result, forecasting the student's performance in advance becomes challenging. This project makes an effort to assist the learner in anticipating his performance. The Random Forest regression model is used for this. Based on expected grades, this would aid students in improving their performance and help teachers recognise those kids who require extra attention. The main goal of the "Student Grade Prediction Application" is to put into practise a straightforward computational model that forecasts each student's final grade for the academic year.

**Keywords:**

Students Grades, Grades Prediction, Examination, Random Forest.

## Index

<b>Chapter</b>	<b>Contents</b>		<b>Page no.</b>
1.	<b>Introduction</b>		<b>5</b>
	a.	Problem statement	<b>5</b>
	b.	Project Idea	<b>5</b>
	c.	Motivation	<b>5</b>
	d.	Scope	<b>6</b>
	e.	Literature Survey	<b>6</b>
2.	<b>Product Design</b>		<b>8</b>
	a.	Requirements	<b>8</b>
	b.	Dataset Design	<b>9</b>
3.	<b>Module Description</b>		<b>10</b>
	a.	Modeling	<b>10</b>
	b.	Model building with Flask API	<b>10</b>
4.	<b>Results</b>		<b>11</b>
	a.	Source code	<b>11</b>
	b.	Screenshots with GUI	<b>13</b>
5.	<b>Conclusion</b>		<b>15</b>
6.	<b>References</b>		<b>16</b>

## **1. Introduction**

### **a) Problem Statement:**

By utilizing a Random Forest regression model, the primary goal of student grade prediction is to assist the student in knowing their performance in advance. These methods would help the students perform better based on the expected grade and would allow the teachers to spot the pupils who might want extra aid.

### **b) Project Idea**

Student performance in the current educational systems is deteriorating day by day. Predicting student achievement beforehand can assist both students and their teachers in monitoring a student's development. Today, a lot of educational institutions use a continuous evaluation method. These programmes are helpful to kids in raising their performance. The continuous evaluation system's goal is to support ordinary students in their academic endeavors. Unit tests or class tests are done on a regular basis under the continuous assessment method. It is necessary to participate in every unit exam or class test in order to maintain consistency in performance for the final grade.

By utilizing a univariate linear regression model, the primary goal of student grade prediction is to assist the student in knowing their performance in advance. These methods would help the students perform better based on the expected grade and would allow the teachers to spot the pupils who might want extra aid.

### **c) Motivation**

Monitoring student growth has become crucial in today's classrooms and workplaces. This contributes to higher teacher and student performance evaluations. Thus, a mechanism for predicting student marks is created..

## **d) Scope**

Other machine learning techniques can be introduced to the model to increase accuracy, and data cleaning and analysis can be done more effectively. Predictions will be more accurate with a larger dataset. A dataset with enough features and an increase in size must be obtained in order to improve the findings. It is necessary to perform additional research to improve the current machine learning approaches so they can operate in real time and produce an effective model. In order to assess the effectiveness and scalability of the constructed models, they must also be tested on data of various volumes.

The outcome of regression on a balanced dataset can be investigated in subsequent work by altering the data distribution. To balance the types of data, this can be accomplished by choosing a sample of the dataset or eliminating specific entries.

## **e) Literature Survey**

### **A. A Comparative Analysis of Techniques for Predicting Student Performance:**

The problem of student final grade prediction in a particular course has recently been addressed using data mining techniques. In this paper, we present two different approaches solving this task. Both approaches are validated on 138 courses which were offered to students of the Faculty of Informatics of Masaryk University between the years of 2010 and 2013. The first approach is based on classification and regression algorithms that search for patterns in study-related data and also data about students' social behavior. We prove that students' social behavior characteristics improve prediction for a quarter of courses. The second approach is based on collaborative filtering techniques. We predict the final grades based on previous achievements of similar students. The results show that both approaches reached similar average results and can be beneficially utilized for student final grade prediction. The first approach reaches significantly better results for courses with a small number of students. In contrary, the second approach achieves significantly better results for mathematical courses. We also identified groups of courses for which we are not able to predict the grades reliably. Finally, we are able to correctly identify half of all failures (that constitute less than a quarter of all grades) and predict the final grades only with the error of one degree in the grade scale.

## **B. Student Grade Prediction Using Data Analytics:**

In the world of open education systems, students have flexibility to learn anything with ease as the learning content is easily available. But this facility can make student complacent. Therefore, it becomes difficult to predict the student's performance in advance. In this project, an attempt is made to help the student to know his performance in advance. This is done by using univariate linear regression model. This would help students to improve their performance based on predicted grades and would enable teachers to identify those individuals who need assistance.

The Main Objective of "Student Grade Prediction Application" is to implement a simple algorithmic model that predicts the score of an individual student at the end of the year. "G3" or the final grade is our label (output) and the rest of the columns will be our features (inputs).

## **C. Prediction of Student's Performance Using Random Forest Classifier :**

Measuring student performance based on both qualitative and quantitative factors is essential because many undergraduate students could not be able to complete their degree in recent pasts. At present, students' dropout rate in university is gradually increasing and many bright students sometimes just cannot cope with the universities. The first-year result of a student is very important because in the majority of cases this drives the students to be either motivated or demotivated. So, the first-year student performance of a renowned university in Bangladesh is investigated in this paper. This research is mainly based on finding the reasons for students' different types of results and then predicting students' performance based on those eleven significant factors. For this purpose, a popular supervised machine learning algorithm, random forests (RF) have been used for classifying students' different levels of results and predicting students' performances. The input dataset for both training and testing were taken by merging the values obtained from two surveys done on students and experts using fuzzy ANFIS analysis. The result exhibits that RF can perform the classification of multiple classes based on many distinguishing features with 96.88 percent accuracy. This proposed model can also be applied to predict course-wise students' performances and its precision can also be greatly improved by adding new factors.

## 2. Product Design

### a) Requirements

#### Hardware requirements

- Operating system- Windows 7,8,10
- Processor- dual core 2.4 GHz (i3 or i5 or i7 series Intel processor or equivalent AMD)
- RAM-4GB

#### Software Requirements

- Python Pycharm
- PIP 2.7
- Jupyter Notebook
- Chrome

#### Technologies used:

-Frontend

- Html, CSS, JavaScript
- Bootstrap
- Flask

-Backend

- Machine learning model



## b.) Dataset Design

This data set consists of the marks secured by the students in various subjects which is available at [www.kaggle.com](http://www.kaggle.com).

Below are the dataset (before cleaning):

	gender	race/ethnicity	parental level of education	lunch	test preparation course	math score	reading score	writing score
0	female	group B	bachelor's degree	standard	none	72	72	74
1	female	group C	some college	standard	completed	69	90	88
2	female	group B	master's degree	standard	none	90	95	93
3	male	group A	associate's degree	free/reduced	none	47	57	44
4	male	group C	some college	standard	none	76	78	75
...	...	...	...	...	...	...	...	...
995	female	group E	master's degree	standard	completed	88	99	95
996	male	group C	high school	free/reduced	none	62	55	55
997	female	group C	high school	free/reduced	completed	59	71	65
998	female	group D	some college	standard	completed	68	78	77
999	female	group D	some college	free/reduced	none	77	86	86

1000 rows × 8 columns

About the dataset:

1. gender (String)
2. race/ethnicity(string)
3. parental level of education(string)
4. lunch(string)
5. test preparation course(string)
6. math score(integer)
7. reading score(integer)
8. writing score(integer)

## Modul description

## Modeling

### 1.) Random forest Regressor

Random Forest is already revealing that it creates a forest and then somehow randomizes it. It builds the forest through the ensemble of Decision Trees and most of the time trains it using a method called the Bagging Method. Since it uses the ensemble method, the result is improved. Decision tree and bagging classifier hyperparameters are the same. Each feature in the tree can be made random simply by adding thresholds.

```
from sklearn.ensemble import RandomForestRegressor
```

```
regressor=RandomForestRegressor()
```

### 2.)Model Building with Flask API

For building the model first pickle the model which you have created in colab.

```
import pickle  
pickle.dump(model,open('df.pkl','wb'))
```

Create a new app.py file. Now, import every important module and library to deploy the model. Also, load the model in the app.py file. For the web application, you need to create an HTML file for the structure of a website and you can add CSS and JS for styling and other thing

## 4. Results and Description

### 1. Source Code

```
In [1]: import pandas as pd
import matplotlib.pyplot as plt
```

```
In [2]: df = pd.read_csv("StudentsPerformance.csv")
```

```
In [7]: df.isnull().sum()
```

```
In [16]: gender = {
    'male':1,
    'female':0
}
```

```
In [17]: race = {
    'group A':0,
    'group B':1,
    'group C':2,
    'group D':3,
    'group E':4
}
```

```
In [18]: df['gender']=df['gender'].map(gender)
df['race/ethnicity']=df['race/ethnicity'].map(race)
```

```
In [20]: level = {
    "bachelor's degree":0,
    'some college':1,
    "master's degree":2,
    "associate's degree":3,
    "high school":4,
    "some high school":5
}
```

```
In [21]: df['parental level of education']=df['parental level of education'].map(level)
```

```
In [23]: df = pd.get_dummies(df,drop_first=True)
```

---

```
In [42]: from sklearn.model_selection import train_test_split
```

```
In [48]: x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.25,random_state=0)
```

```
In [49]: from sklearn.ensemble import RandomForestRegressor
```

```
In [50]: model=RandomForestRegressor()
```

```
In [51]: model.fit(x_train,y_train)
```

```
Out[51]: RandomForestRegressor()
```

```
In [52]: predictions=model.predict(x_test)
```

```
In [53]: predictions
```

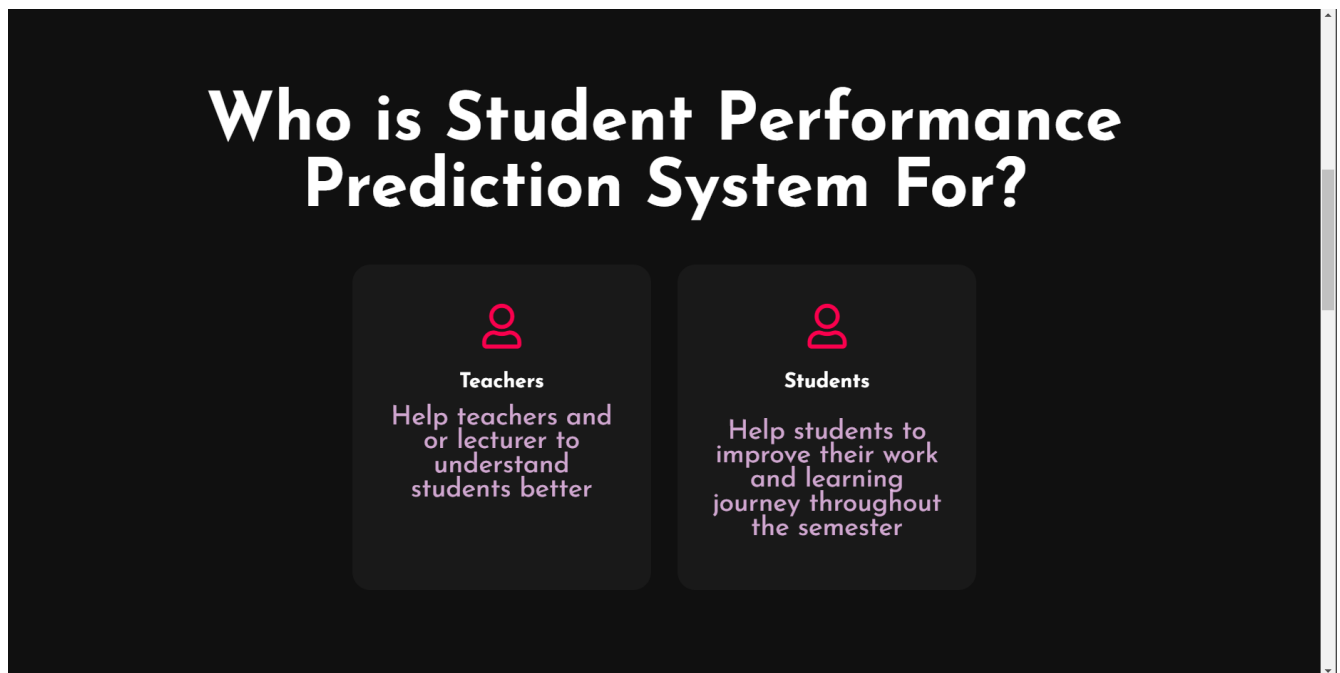
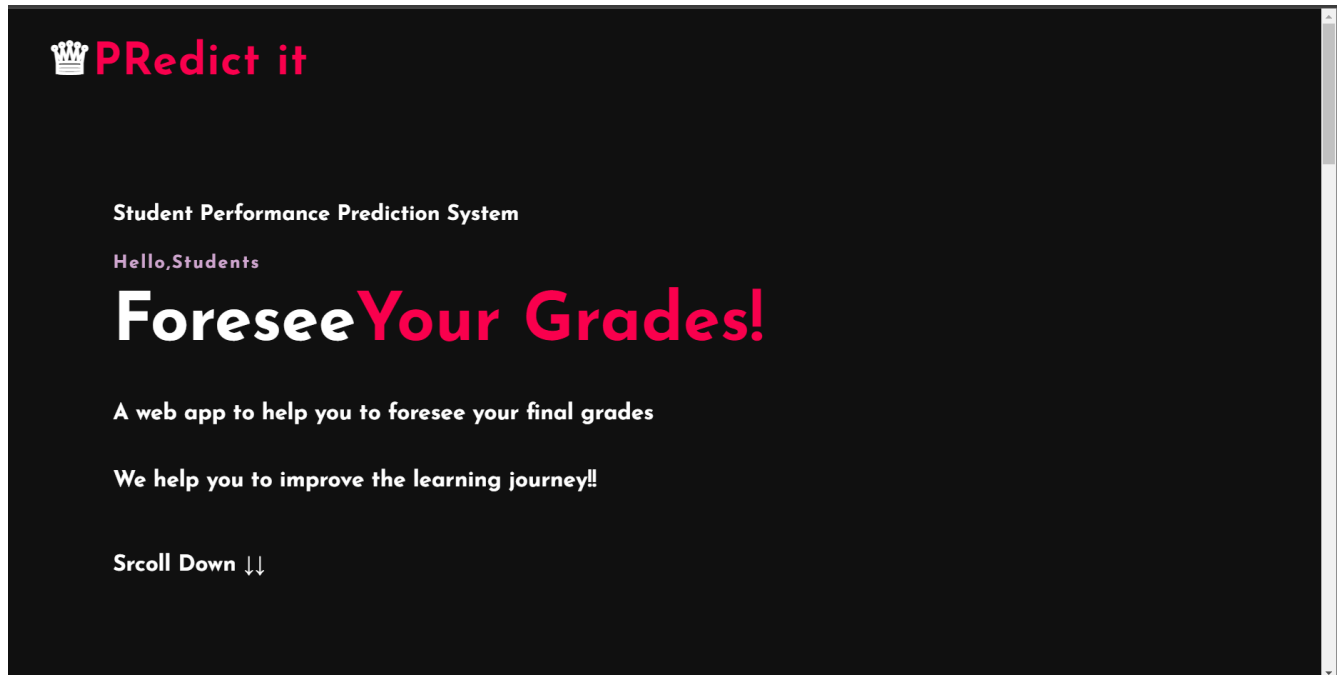
```
In [54]: from sklearn.metrics import r2_score
```

```
In [55]: print(r2_score(predictions,y_test))
```

```
0.9972615591014802
```

```
In [58]: import pickle  
pickle.dump(model,open('df.pkl','wb'))
```

## 2. Screenshot of GUI



Average Score: 66.02

GENDER

RACE/ETHNICITY

PARENTAL LEVEL OF EDUCATION

Female

Group A

Some High School

MATH SCORE

READING SCORE

WRITING SCORE

LUNCH

Standard

TEST PREPARATION COURSE

None

SUBMIT

## **5. Conclusion**

The main goal of this paper is to analyze the student data and predict the students performance. Here, the algorithm are used and compared to find the best accurate algorithm. The implementation tools used are pycharm, Jupyter and atom. Hence using this prediction, the teacher can be focus on weak stuends. This project opens scope for future work in predict data for all semester.

## 6. References

1. Hana Bydžovská, “A Comparative Analysis of Techniques for Predicting Student Performance”, 9th International Conference on Educational Data Mining , oct ,2017
2. Sourav Kumar Ghosh and Farhatul Janan,” Prediction of Student’s Performance Using Random Forest Classifier”, IEOM Society International, aug , 2018
3. Lahari Pampati, B. Varshini, V. Tanmaya and P. Archana ,”Student Grade Prediction Using Data Analytics”, THINK INDIA JOURNAL ,jul, 2019