ENG COLLEGE ADMISSION PREDICTION BASED ON STUDENT PERFORMANCE

SANDIP KUMAR THAKUR
REG NO:-1615500110116
ROLL NO:-15500216013

SK AMAN
REG NO:-1615500110122
ROLL NO:-15500216007

SANCHARI MAITY
REG NO:-1615500110115
ROLL NO:- 15500216014

DEPARTMENT-Information Technology(IT)

Durgapur Institute Of Advanced Technology and Management
Maulana Abul Kalam Azad University of Technology

ENG COLLEGE ADMISSION PREDICTION BASED ON STUDENT PERFORMANCE

A dissertation submitted in partial fulfillment of

Bachelor of Technology
In

Information Technology

By

Student's name
Sandip Kumar Thakur
Registration No. 1615500110116
Univ. Roll No. 15500216013
B.Tech. (IT), 8th Semester

Sk Aman
Registration No. 1615500110122
Univ. Roll No.15500216007
B.Tech. (IT), 8th Semester

Sanchari Maity
Registration No.1615500110115
Univ. Roll No.15500216014
B.Tech. (IT), 8th Semester

B.Tech. (INFORMATION TECHNOLOGY), 8th Semester

Under the Guidance
Of
Amit Kumar Dutta

Durgapur Institute of Advanced Technology and Management

Affiliated to

Maulana Abul Kalam Azad University of Technology
Kolkata
2020

To whom it may Concern

I hereby recommend that the Project (Major Project) prepared under my supervision, by **Sandip Kumar Thakur, Sk Aman, Sanchari Maity** (Roll. No. **15500216013, 15500216007, 15500216014** Regn. No. **161550110116, 161550110122, 161550110115** of **Durgapur Institute of Advanced Technology and Management**, B.Tech., Information Technology, 8th Semester) entitled ***"University Management System"*** be accepted in partial fulfillment of the requirement for the Degree of Bachelor of Technology in Information Technology.

Internal Guide
(Guide name)

Head Of The Department
(Name)

Submitted for the PROJECT PRESENTATION held on……. at College name.

Internal Examiner **External Examiner**

ACKNOWLEDGEMENT:

The feeling of acknowledge and expressing it in words are two things apart. It is weakness, but we honestly admit when we truly wish to express our warm gratitude and indebtedness towards somebody, we are always at loss of word.

The project could not have completed without support extends to us by **Mr. AMIT KUMAR DUTTA**, our **Mentor** as well as **Head of Department** of Information Technology, he guided us in preparation of the project and every time when we reached with difficulties, he welcomed them which helped us to successfully complete the project. We express our heartfelt gratitude with great pleasure and sense of obligation to him.

We would also like to thank **Dr. P.K. SINHA** **Honourable Principal** of our college who continued cooperation and support the department of Information Technology with pivotal force in making this project a success.

And finally a word of gratitude to our family and friends who were always there with their support and encouragement.

Sandip Kumar Thakur(15500216013)
Sk Aman (15500216007)
Sanchari Maity (15500216014)

CERTIFICATE OF ORIGINALITY:

This is to certify, that the Project submitted by us is an outcome of our independent and original work. We have duly acknowledged all the sources from which the ideas and extracts have been taken.

Student Names	Name of the Supervisor
Sandip Kumar Thakur Sk Aman Sanchari Maity	Amit Kumar Dutta

ABSTRACT:

When most of the people hear “Machine Learning,” they picture a robot: a dependable butler or a deadly terminator counting on who you ask. But Machine Learning isn't just a futuristic fantasy, it’s already here.

What Is Machine Learning?

Machine Learning is that the science (and art) of programming computers in order that they can learn from data.

Here may be a slightly more general definition:

[Machine Learning is the] field of study that provides computers with the power to find out without being

explicitly programmed. By-Arthur Samuel, 1959

And a more engineering-oriented one:

"A computer program is claimed to find out from experience E with reference to some task T and a few

performance measure P, if its performance on T, as measured by P, improves with experience E. By-Tom Mitchell, 1997."

CONTENTS:

Reference:

INTRODUCTION

OBJECTIVE & SCOPE OF THE PROJECT:

Education plays a vital role in today's era. While we talk about career - a person's degree, course, university and the knowledge that he possesses - is the key factor on which the firm hires a fresher. As soon as a student completes his/her Higher Secondary Schooling, the first goal of any student is to get into an appropriate College so that he can get a better education and guidance for his future. For that, students seek help from many sources like online sites or career experts to get the best options for their future. A good career counselor charges a huge amount for providing such solutions. Online sources are also not as reliable as the data from particular sources is not always accurate. Students also perform their analysis before applying to any institutions, but this method is slow and certainly not consistent for getting actual results and possibly includes human error. Since the number of applications in different universities for each year is way too high, there is a need to build up a system that is more accurate or precise to provide proper suggestions to students.

Our aim is to use machine learning concepts to predict the probability of a student to get admission into those preferred colleges and suggest a list of collages in a sequence of the probability of getting admission to that specific college. The following are the steps that include the work we have done in sequence of implementation.

THEORETICAL BACKGROUND

The core idea of machine learning, according to Arora, involves training a machine to search for patterns in data and improve from experience and interaction. This is very analogous to classic curve-fitting, a mathematical technique known for centuries. Training involves algorithms, the theoretical foundations of which are of great interest in mathematics (see related article). "Machine learning is a very important branch of the theory of computation and computational complexity," says Avi Wigderson, Herbert H. Maass Professor in the School of Mathematics, who heads the Theoretical Computer Science and Discrete Mathematics program. "It is something that needs to be understood and explained because it seems to have enormous power to do certain things—play games, recognize images, predict all sorts of behaviors. There is a really large array of things that these algorithms can do, and we don't understand why or how. Machine learning definitely suits our general attempts at IAS to understand algorithms, and the power and limits of computational devices."

the core idea of machine learning, according to Arora, involves training a machine to search for patterns in data and improve from experience and interaction. This is very analogous to classic curve-fitting, a mathematical technique known for centuries. Training involves algorithms, the theoretical foundations of which are of great interest in mathematics (see related article). "Machine learning is a very important branch of the theory of computation and computational complexity," says Avi Wigderson, Herbert H. Maass Professor in the School of Mathematics, who heads the Theoretical Computer Science and Discrete Mathematics program. "It is something that needs to be understood and explained because it seems to have enormous power to do certain things—play games, recognize images, predict all sorts of behaviors. There is a really large array of things that these algorithms can do, and we don't understand why or how. Machine learning definitely suits our general attempts at IAS to understand algorithms, and the power and limits of computational devices."

Why has machine learning become so pervasive in the past decade? According to Arora, this happened due to a symbiosis between three factors: data, hardware, and commercial reward. "Leading tech companies rely on such algorithms," says Arora. "This creates a self-reinforcing phenomenon: good algorithms bring them users, which in turn yields more user data for improving their algorithms, and the resulting rise in profits further lets them invest in better researchers, algorithms, and hardware."

With intense progress and momentum in the field coming from industry, the number of machine learning researchers who are trying to establish theoretical understanding is relatively small. But such study is essential—for reasons beyond its tantalizing connections to questions in mathematics and even physics. "Imagine if we didn't have a theory of aviation and could not predict how airplanes would behave under new conditions," says Cohen, current Member in the theoretical machine learning program. "Soon you will be putting your life in the hands of an

algorithm when you are sitting in a self-driving car or being treated in an operating room. We can't yet fully understand or predict the properties of today's machine learning algorithms."

DEFINITION OF THE PROBLEM

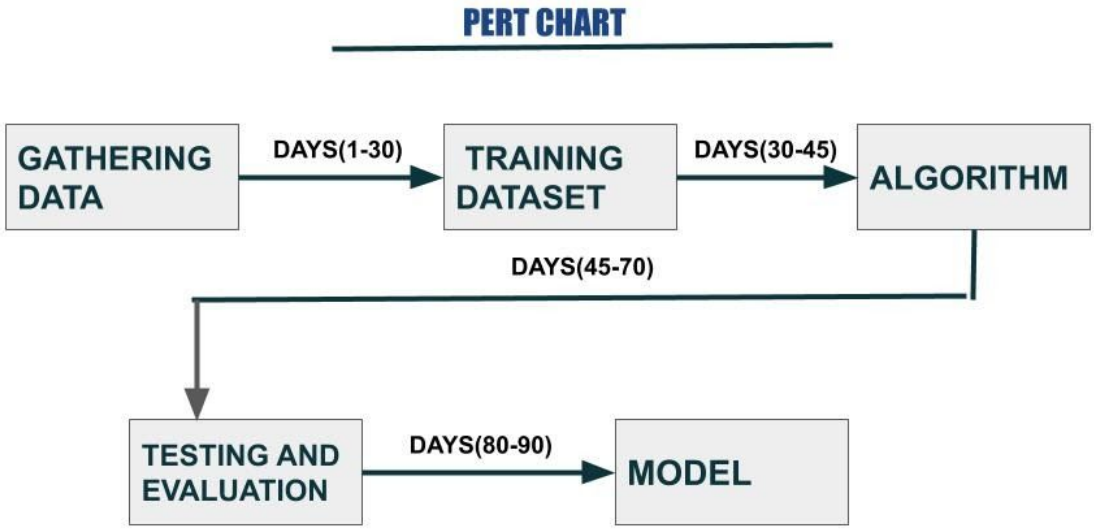
We know after 12th board results, the main problem of the students is to find an appropriate college for their future education. It is a tough decision to make for many students as to which college they should apply to. We have built a system that compares the student's data with past admission data and suggests colleges in a sequence of their preference. We have used XGBoost classifier, Decision tree classifier, etc. as our statistical model to predict the probability of getting admission to college. It was observed that the performance of XGBoost was achieved highest among all.

SYSTEM ANALYSIS AND DESIGN:-

User Requirements:-

NUMBER	DESCRIPTION
1	PC with 2GB hard-disk and 256MB RAM
2	Mobile with minimum 4GB RAM
3	Minimum 10MB space to download the space
4	Internet speed with minimum 200kbps
5	Intel core i3 9100F 9th Gen

System Planning (PERT Chart):-



Details of Hardware & Software used:-

HARDWARE USED:- HP laptop with RAM of 4GB , 1TB storage, 2GB graphics and processor CORE i3 intel.

SOFTWARE USED:-

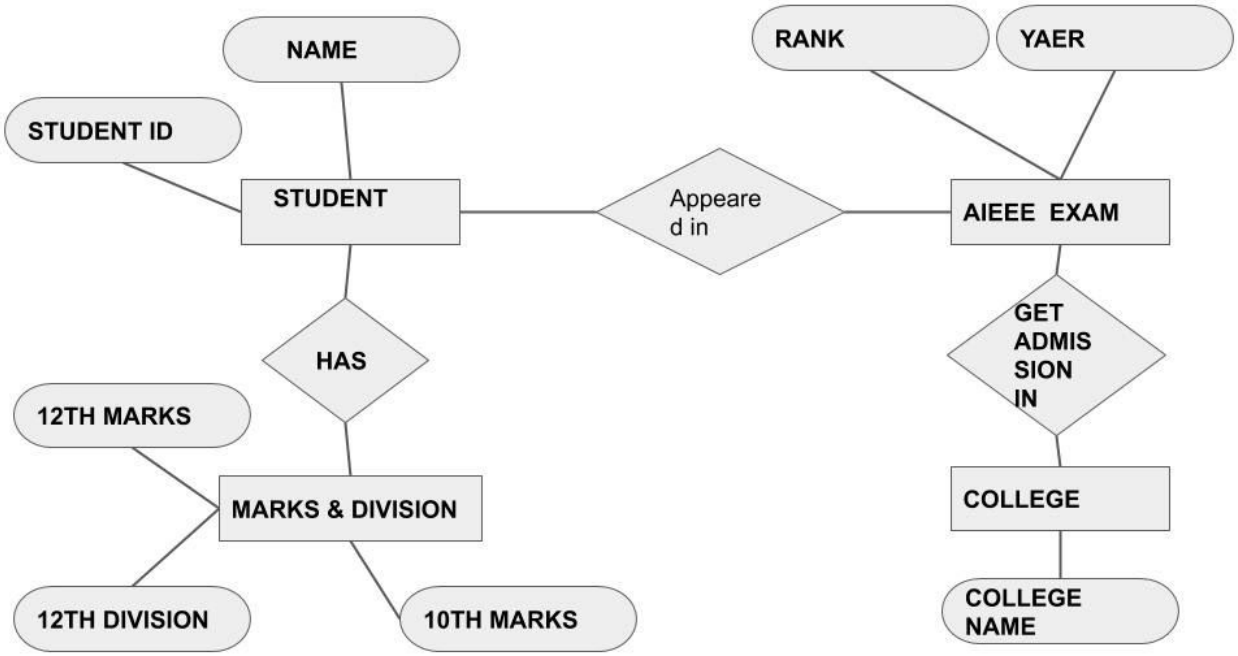
- **ANACONDA-** Anaconda is a free and open-source distribution of the Python and R programming languages for scientific computing (data science, machine learning applications, large-scale data processing, predictive analytics, etc.), that aims to simplify package management and deployment.

- **JUPYTER NOTEBOOK-**The Jupyter Notebook is an open-source web application that allows you to create and share documents that contain live code, equations, visualizations and narrative text. Uses include: data cleaning and transformation, numerical simulation, statistical modeling, data visualization, machine learning, and much more.

- **PYTHON VERSION>3.7-** Python is an interpreted, high-level, general-purpose programming language. Created by Guido van Rossum and first released in 1991, Python's design philosophy emphasizes code readability with its notable use of significant whitespace.

ERD

ERD DIAGRAM

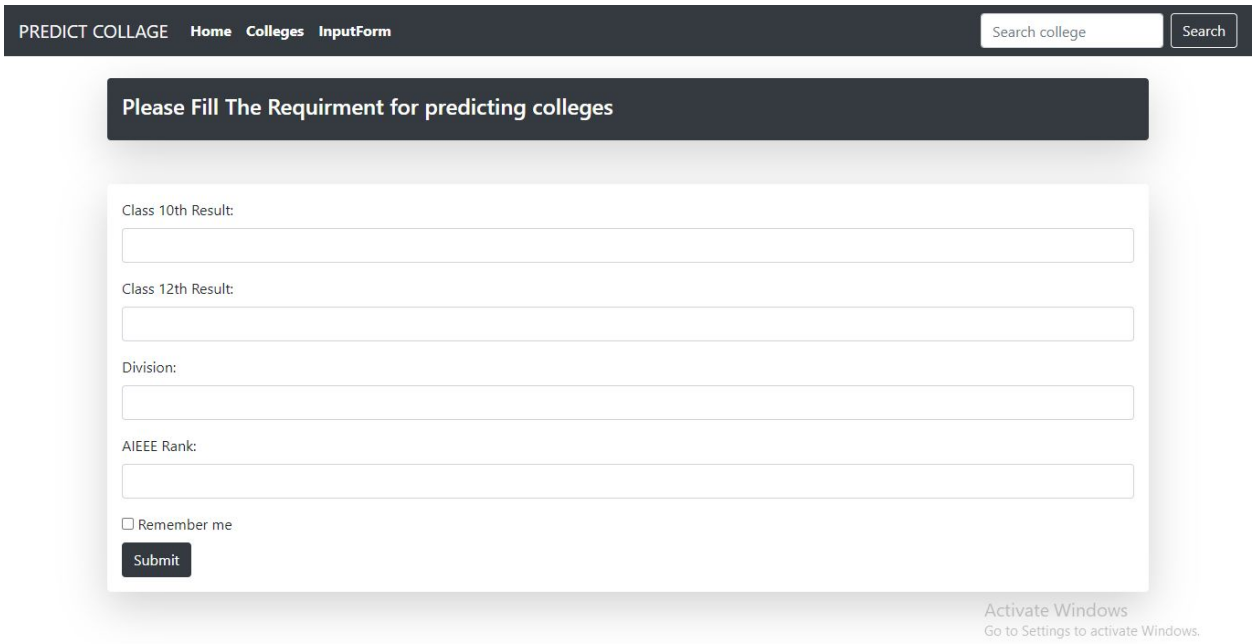


Database Tables:-

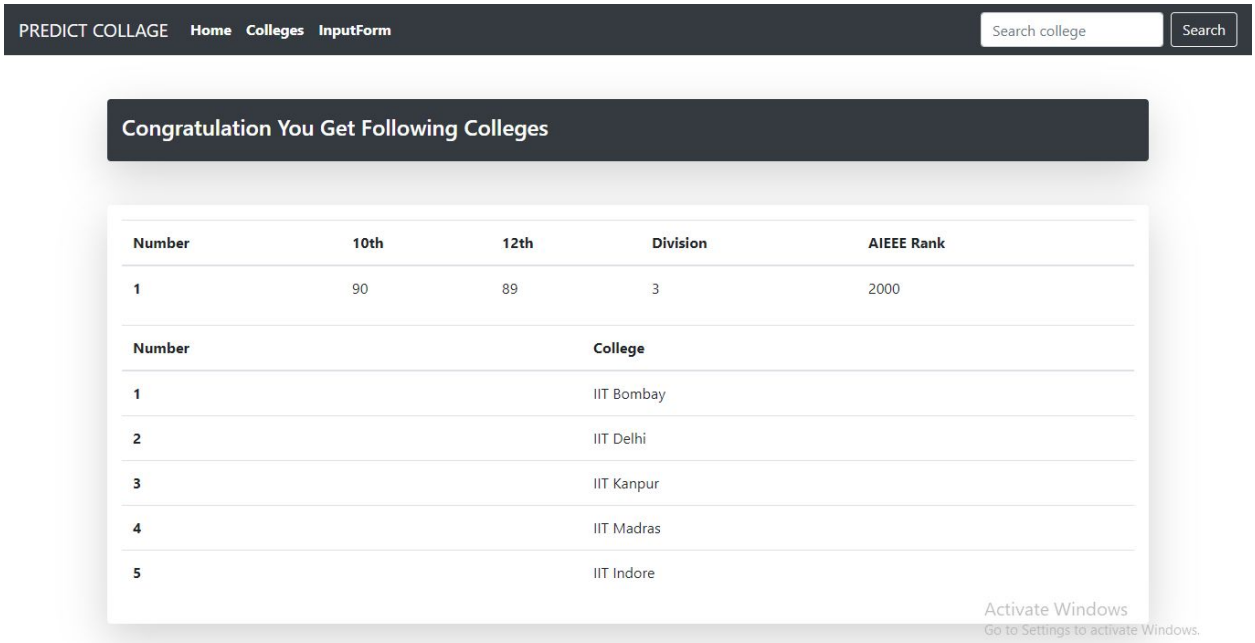
SL NUMBER	ATTRIBUTE	DESCRIPTION
1	Name	Name of a student
2	Year	Year of completing exam
3	10TH Marks	10th marks obtained by student in class 10
4	12TH Marks	12th marks obtained by student in class 12
5	12TH Division	12th division obtained by student in class 12
6	AIEEE Rank	AIEEE Rank obtained by student
7	Colleges	Colleges according to the students performance

INPUT AND OUTPUT SCREEN DESIGN

Input Form Design:-



Output Form Design:-



System Implementation:

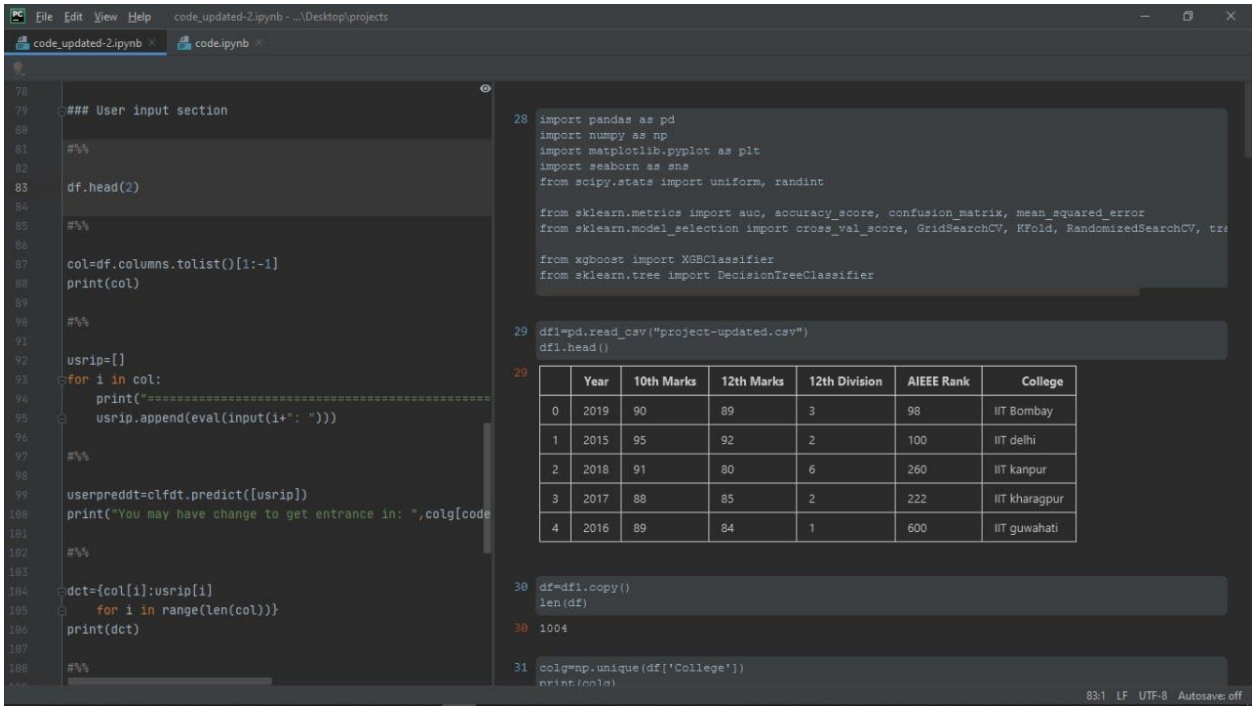
Sample code:-

Libraries we have imported

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from scipy.stats import uniform, randint

from sklearn.metrics import auc, accuracy_score, confusion_matrix,
mean_squared_error
from sklearn.model_selection import cross_val_score, GridSearchCV,
KFold, RandomizedSearchCV, train_test_split

from xgboost import XGBClassifier
from sklearn.tree import DecisionTreeClassifier
```

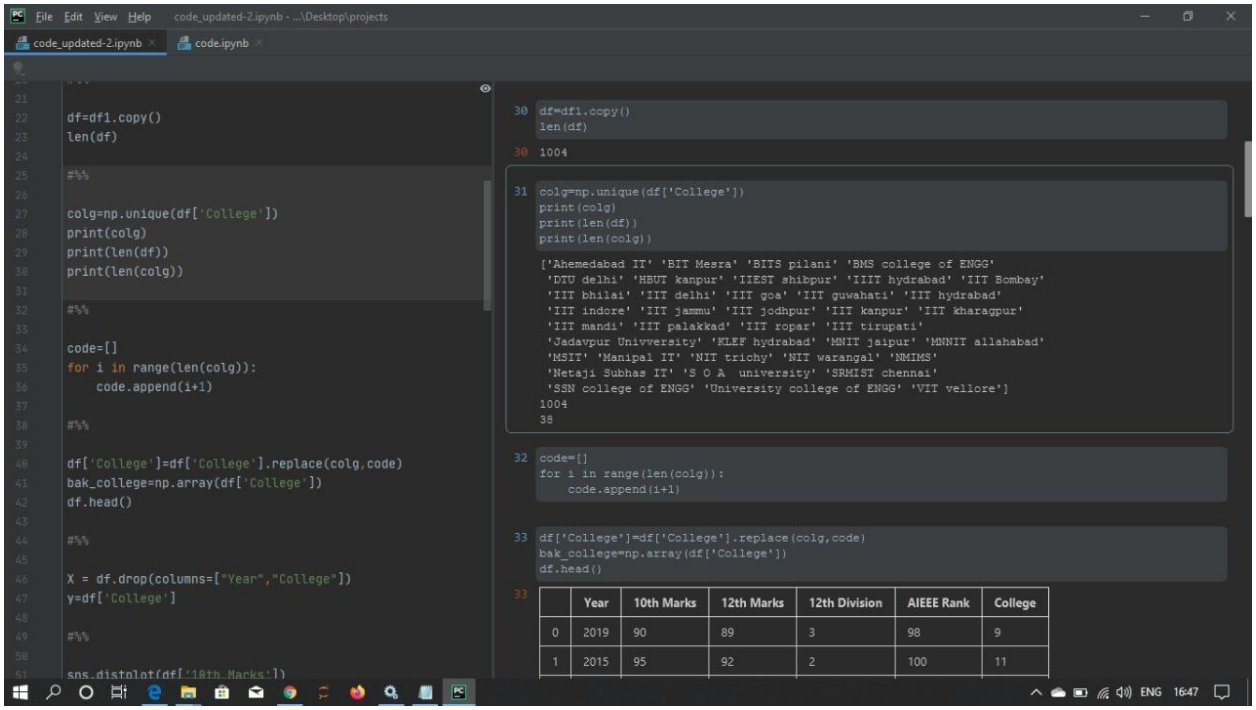
Reading of .CSV file

```
df1=pd.read_csv("project-updated.csv")
df1.head()
```



```
df=df1.copy()
len(df)
```

```
colg=np.unique(df['College'])
print(colg)
print(len(df))
print(len(colg))
```



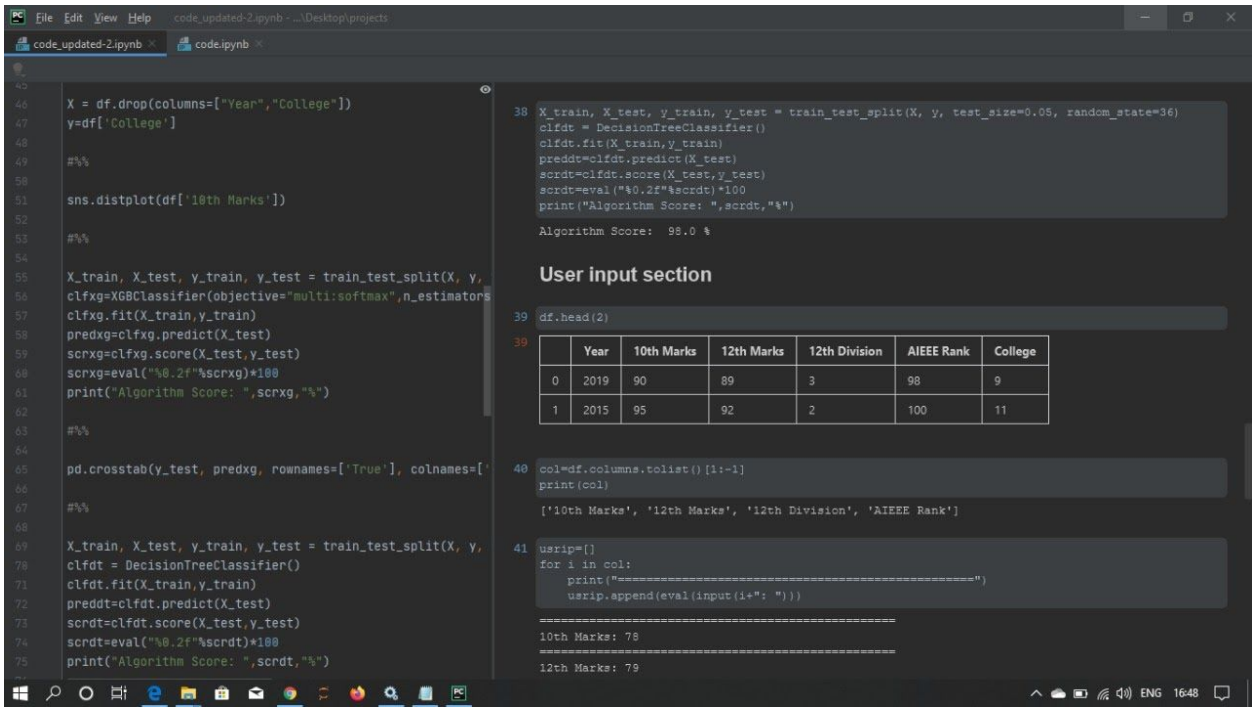
```
code=[]
for i in range(len(colg)):
    code.append(i+1)

df['College']=df['College'].replace(colg,code)
bak_college=np.array(df['College'])
df.head()
```

```
X = df.drop(columns=["Year","College"])
y=df['College']
```

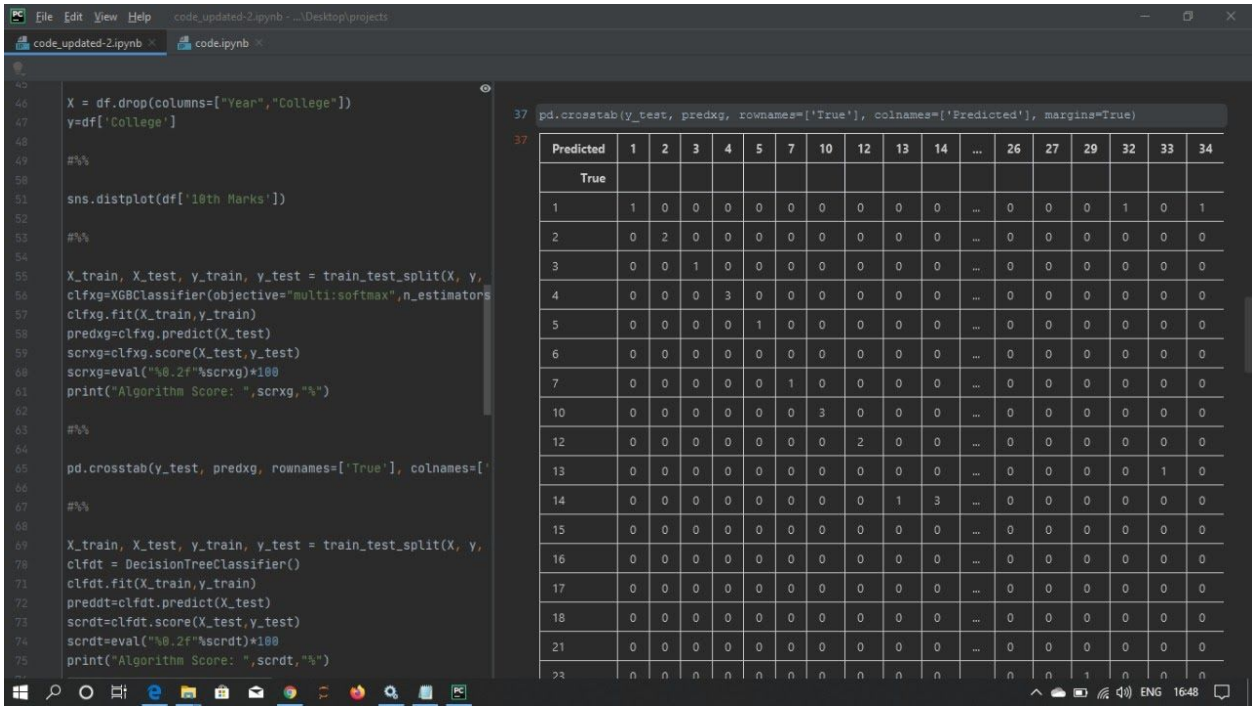
Training of dataset

```
X_train, X_test, y_train, y_test = train_test_split(X, y,
```

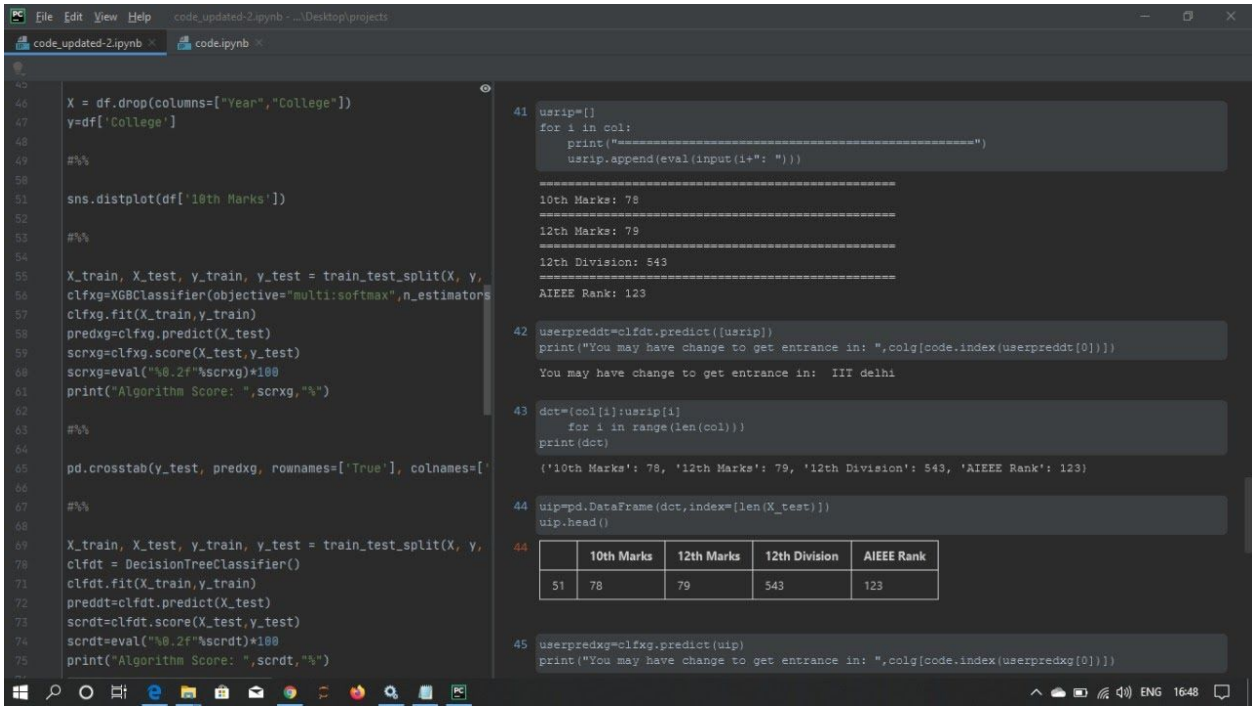
```
test_size=0.05, random_state=22)
clfxg=XGBClassifier(objective="multi:softmax",n_estimators=50,learning_r
ate=0.0001)
clfxg.fit(X_train,y_train)
predxg=clfxg.predict(X_test)
scrxg=clfxg.score(X_test,y_test)
scrxg=eval("%0.2f"%scrxg)*100
print("Algorithm Score: ",scrxg,"%")
```



```
pd.crosstab(y_test, predxg, rownames=['True'], colnames=['Predicted'],
margins=True)
```

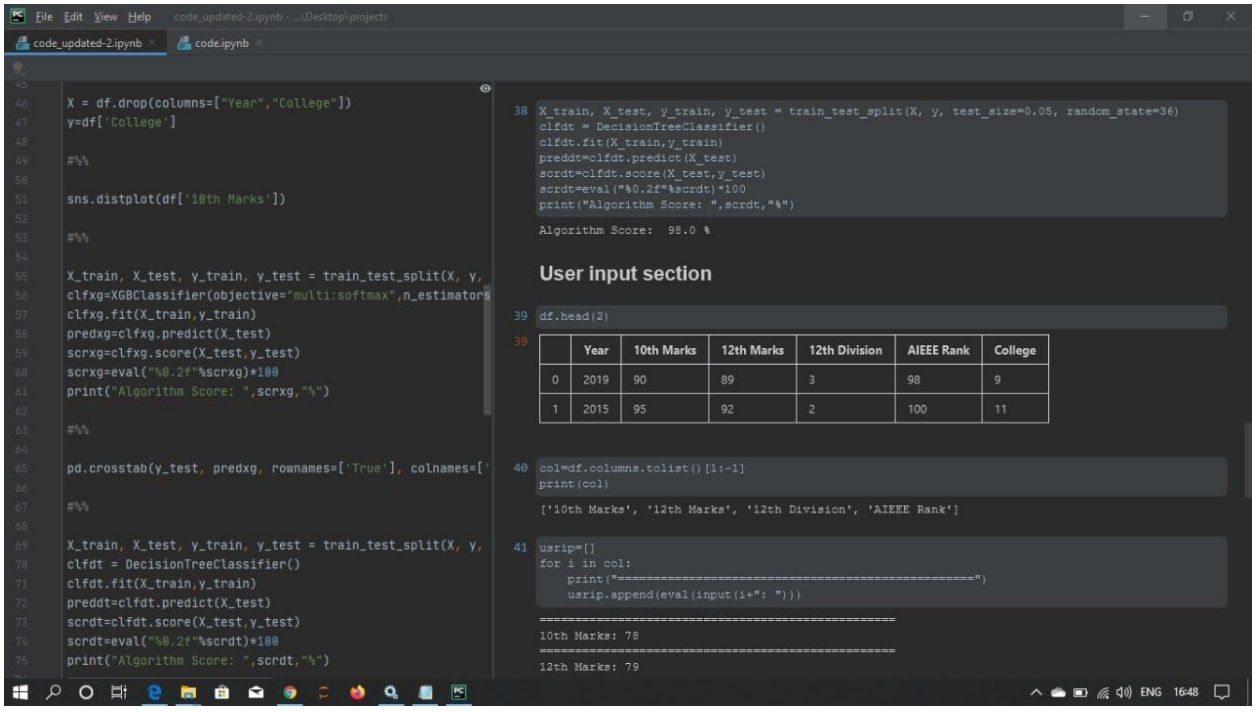


```
df.head(2)
col=df.columns.tolist()[1:-1]
print(col)
usrip=[]
for i in col:
    print("==========================================================")
    usrip.append(eval(input(i+": ")))
```




```
dct={col[i]:usrip[i]
    for i in range(len(col))}
print(dct)
uip=pd.DataFrame(dct,index=[len(X_test)])
uip.head()
userpredxg=clfxg.predict(uip)
print("You may have change to get entrance in:
",colg[code.index(userpredxg[0])])
df1.head(2)
df1['College Code']=bak_college
```

```
df1.head(2)
```


File Edit View Help
code_updated-2.ipynb
code.ipynb
X = df.drop(columns=["Year","College"])
y=df['College']
#%%
sns.distplot(df['10th Marks'])
#%%
X_train, X_test, y_train, y_test = train_test_split(X, y,
clfxg=XGBClassifier(objective="multi:softmax",n_estimators
clfxg.fit(X_train,y_train)
predxg=clfxg.predict(X_test)
scrxg=clfxg.score(X_test,y_test)
scrxg=eval("%0.2f"%scrxg)*100
print("Algorithm Score: ",scrxg,"%")
#%%
pd.crosstab(y_test, predxg, rownames=['True'], colnames=['
#%%
X_train, X_test, y_train, y_test = train_test_split(X, y,
clfdt = DecisionTreeClassifier()
clfdt.fit(X_train,y_train)
preddt=clfdt.predict(X_test)
scrdt=clfdt.score(X_test,y_test)
scrdt=eval("%0.2f"%scrdt)*100
print("Algorithm Score: ",scrdt,"%")
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.05, random_state=36)
clfdt = DecisionTreeClassifier()
clfdt.fit(X_train,y_train)
preddt=clfdt.predict(X_test)
scrdt=clfdt.score(X_test,y_test)
scrdt=eval("%0.2f"%scrdt)*100
print("Algorithm Score: ",scrdt,"%")
Algorithm Score: 98.0 %
User input section
df.head(2)
Year 10th Marks 12th Marks 12th Division AIEEE Rank College
0 2019 90 89 3 98 9
1 2015 95 92 2 100 11
col=df.columns.tolist()[1:-1]
print(col)
['10th Marks', '12th Marks', '12th Division', 'AIEEE Rank']
usrip=[]
for i in col:
print("==")
usrip.append(eval(input(i+": ")))
==
10th Marks: 78
==
12th Marks: 79
ENG 16:48

Methodology used for testing

Software quality in Machine learning and Deep learning systems are different. In this, accuracy, robustness, learning efficiency and adaptation and performance of the system checked. There are two types of technique used in training and testing the Machine Learning model. Which are Train and Test on the same data and Train and Test on different dataset. In our model we have used the train test split method which gave us more accurate results.

Annexure:

Brief background of the organization where the student has developed the project

Durgapur Institute of Advanced Technology and Management, better known as DIATM, saw the light in the year 2002. A brainchild of Shri R. N. Majumder, an industrialist of great repute, the institute is committed to excellence in technical education. in its short span since inception in 2002, the institute has established itself as one of the best of its kind in engineering education. DIATM is affiliated to the West Bengal University of Technology (WBUT) and approved by All India Council for Technical Education (AICTE). DIATM has created a holistic campus with an infrastructure conducive to academic and co-curricular pursuits and a support system catering to the needs of the students.

DIATM, though quite recently established, is now known as one of the most reputed engineering colleges of its kind. The institute not only nurtures talents but aims to turn out only highly trained and knowledgeable techies who are responsible citizens as well. DIATM strives to bring the latest to its students in curriculum, technology and applications to help keep them abreast of the happenings in the real world.

Data Dictionary.

Prediction:- This term is used for predicting the value by machine or model.

Gathering Data:- For performing the train and test method we need data.

Model:- Once one gathers data it's time to create a model.

Anaconda:-Anaconda is a free and open-source distribution of the Python and R programming languages for scientific computing (data science, machine learning applications, large-scale data processing, predictive analytics, etc.), that aims to simplify package management and deployment.

Python:- Python is an interpreted, high-level, general-purpose programming language. Created by Guido van Rossum and first released in 1991, Python's design philosophy emphasizes code readability with its notable use of significant whitespace

Student:- In our project we first gather the student's data and show the available colleges according to their rank.

Importing:- is an important term in ML and also in python. Python is a library rich programming language. Where most of the work is done by importing the libraries. Hence, for importing the libraries we use the import keyword.

BIBLIOGRAPHY:

- Introduction to Machine Learning with Python

  By Andreas C. Müller

- Python Machine Learning

  By Sebastian Raschka

- Hands-On Machine Learning with Scikit-Learn and TensorFlow

  By Aurélien Géron

WEBSITES VISITED:

- www.kaggle.com
- www.mrmachine.com/
- www.reddit.com
- www.youtube.com
- www.google.com
- www.tutorialpoint.com
- www.wikipedia.com
- www.msdn.microsoft.com
- www.webopedia.com