**IDENTIFYING PATTERNS AND TRENDS IN CAMPUS PLACEMENT DATA USING MACHINE LEARNING**

**TEAM LEADER:**

S.THAJITHA BANU

**TEAM MEMBER 1**

L.MONIKKA

**TEAM MEMBER 2**

M.MUTHU SOWMIYA

**TEAM MEMBER 3**

A.PRIYA

**MENDOR:**

Mrs E.ELAKIYA

**1. INTRODUCTION**

**1.1 Overview**

An approach for finding, attracting, and hiring young talent for internships and entry-level roles is campus recruitment. College recruiting is typically a strategy used by medium- to large-sized businesses with high volume recruiting needs, but it can range from modest efforts (such as collaborating with university career centers to find potential candidates) to extensive operations (such as visiting a variety of colleges and participating in recruiting events throughout the spring and fall semesters).Working with university career services offices and attending career fairs to meet in-person with college students and recent graduates are frequent aspects of campus recruitment. Our solution is centered on an Indian business school's placement season. Whereas there are numerous elements that go into hiring people, including work experience, exam %, etc., finally, it includes information about the recruitment status

**1.2 Purpose**

By examining the information gathered from students from past years, the goal is to anticipate where the kids will be placed for the upcoming academic year. This model is suggested along with an algorithm to make the same prediction.

It incorporates the whole machine learning model development life cycle, from data gathering through outcomes visualization.

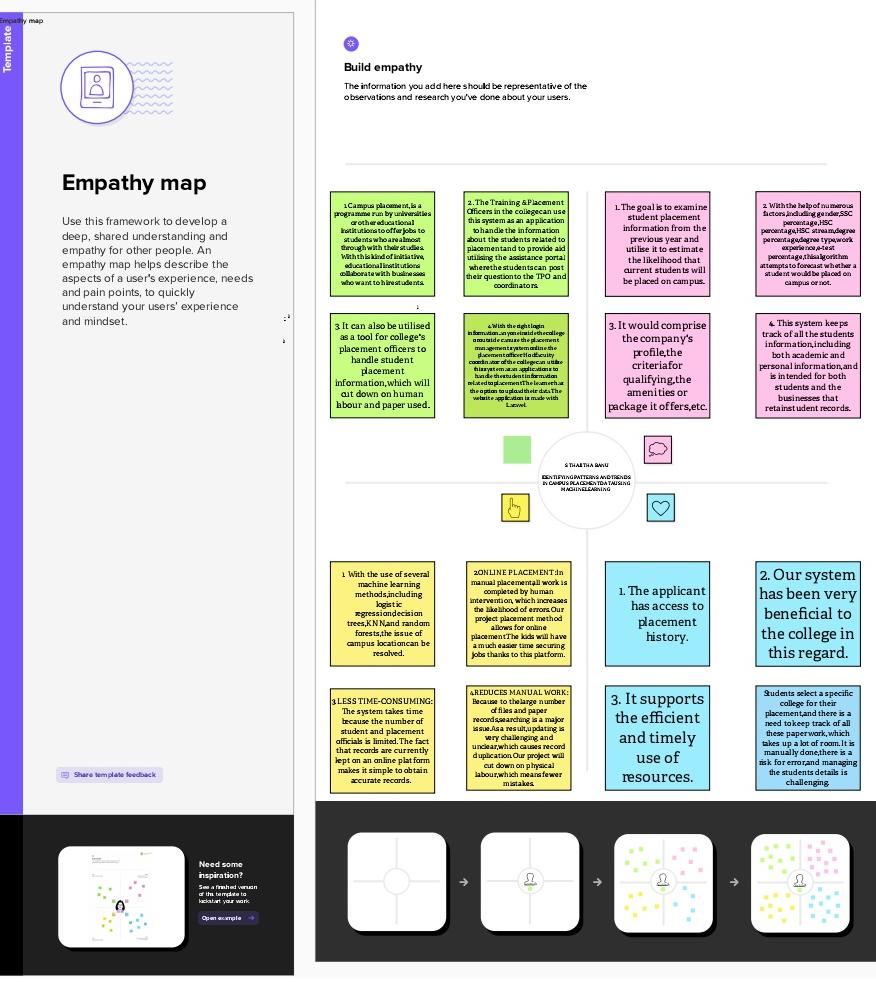
* Carries out data cleansing by addressing outliers and null values.
* Uses data scaling techniques to alter data so that it may be fitted into a machine learning model.

.

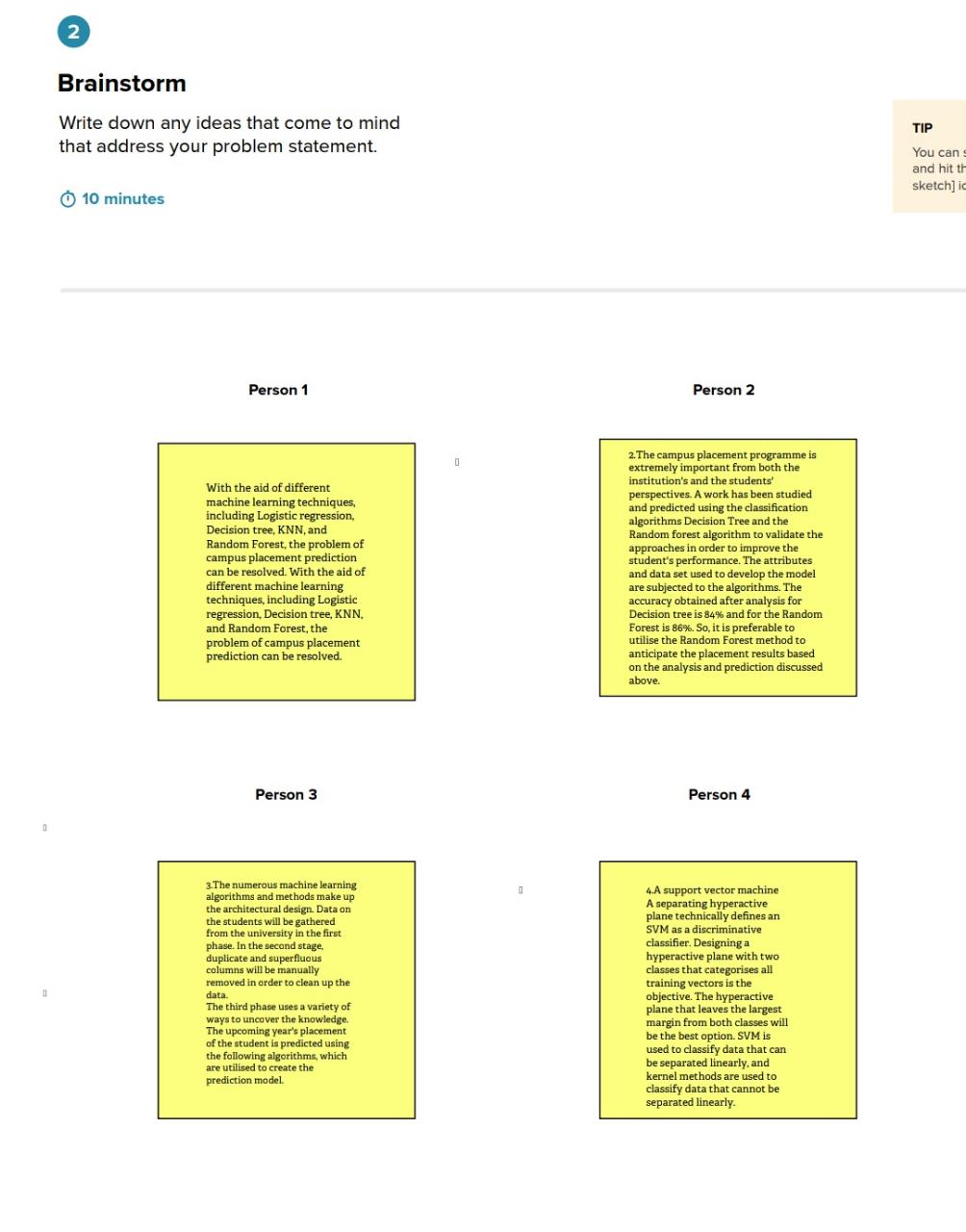
It performs k-fold cross validation on five different machine learning algorithms for data on student performance for job placement in universities AthreyaShetty B, Adithya R. Hande, Akram Pasha, Amith Singh S, and Shreyas N.

**2. PROBLEM DEFINITION & DESIGN THINKING**

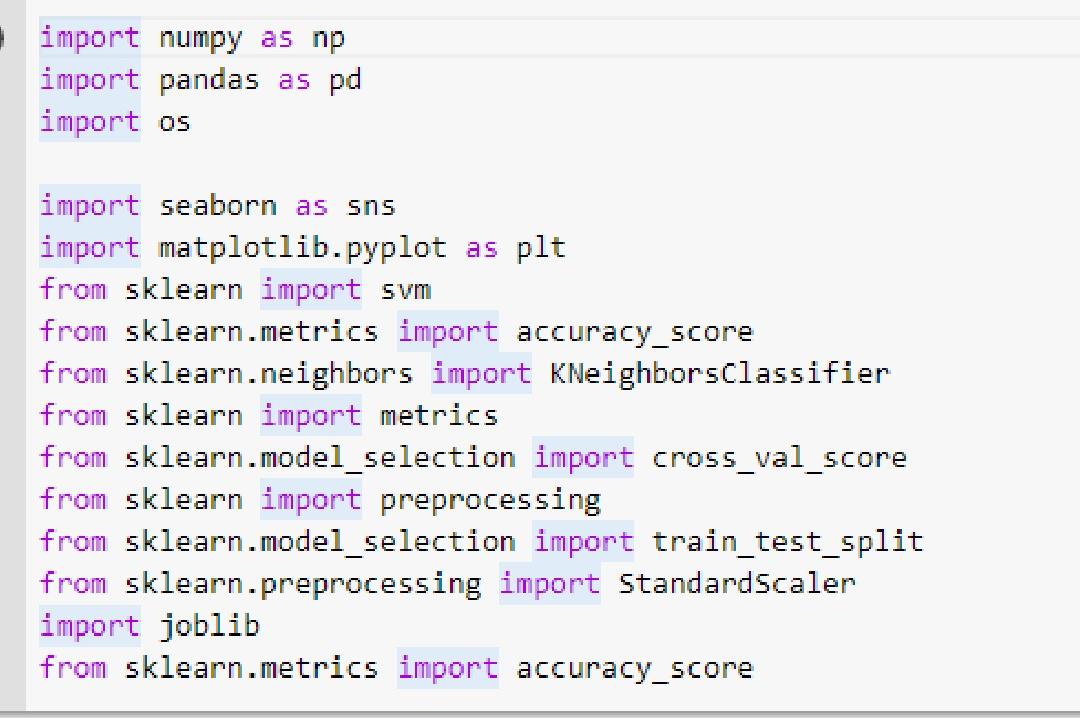
**2.1Empathy Map**

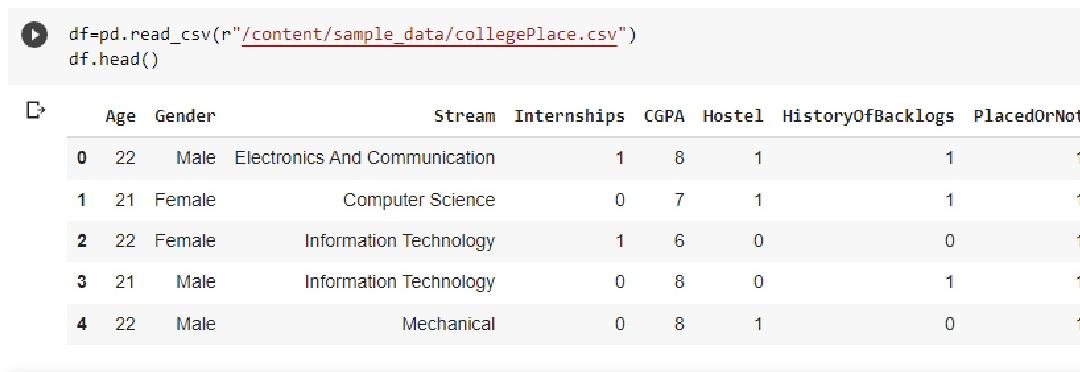


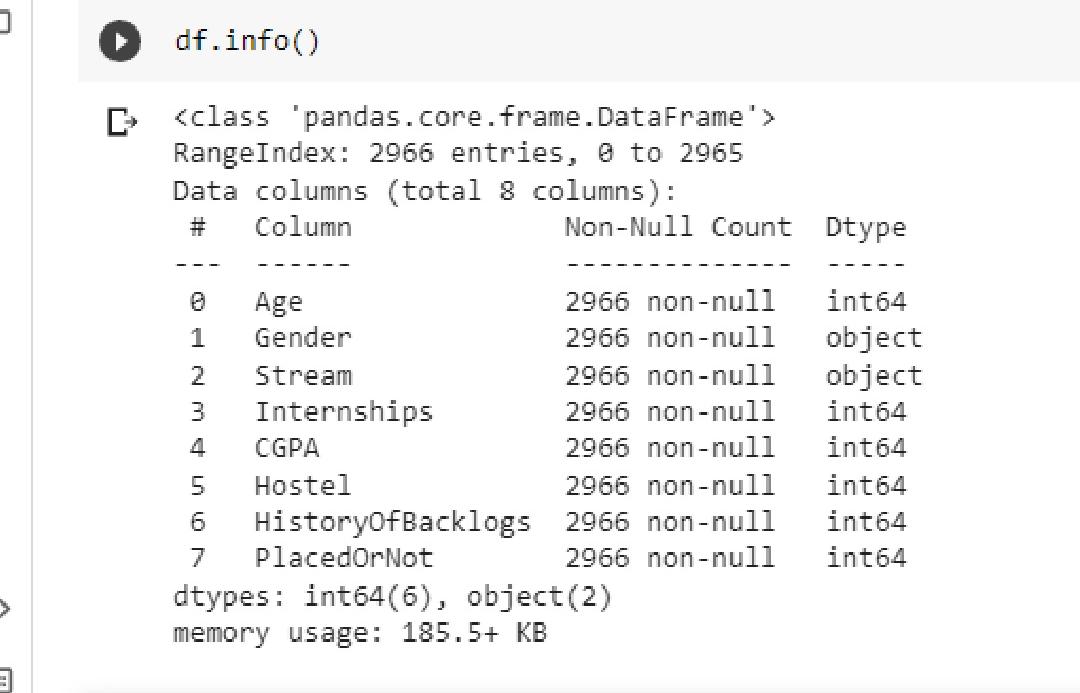
**2.2 Ideation & Brainstorming map**

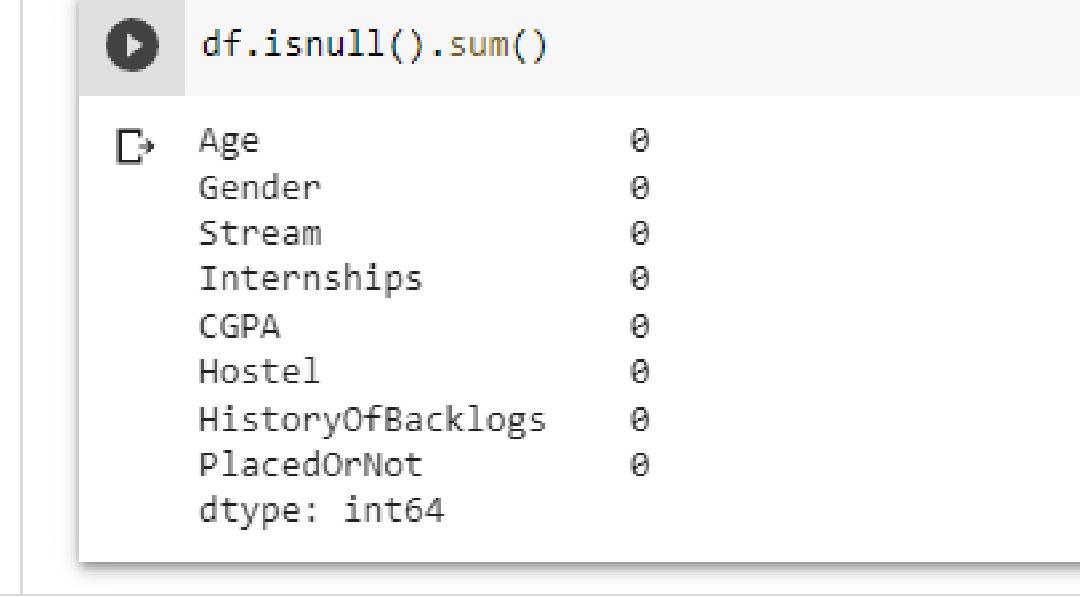


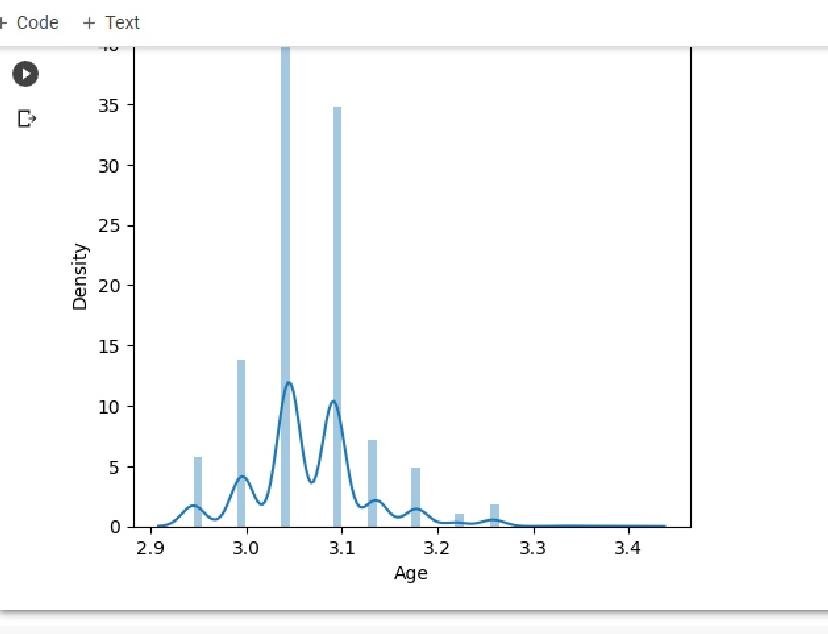
**3. RESULT**

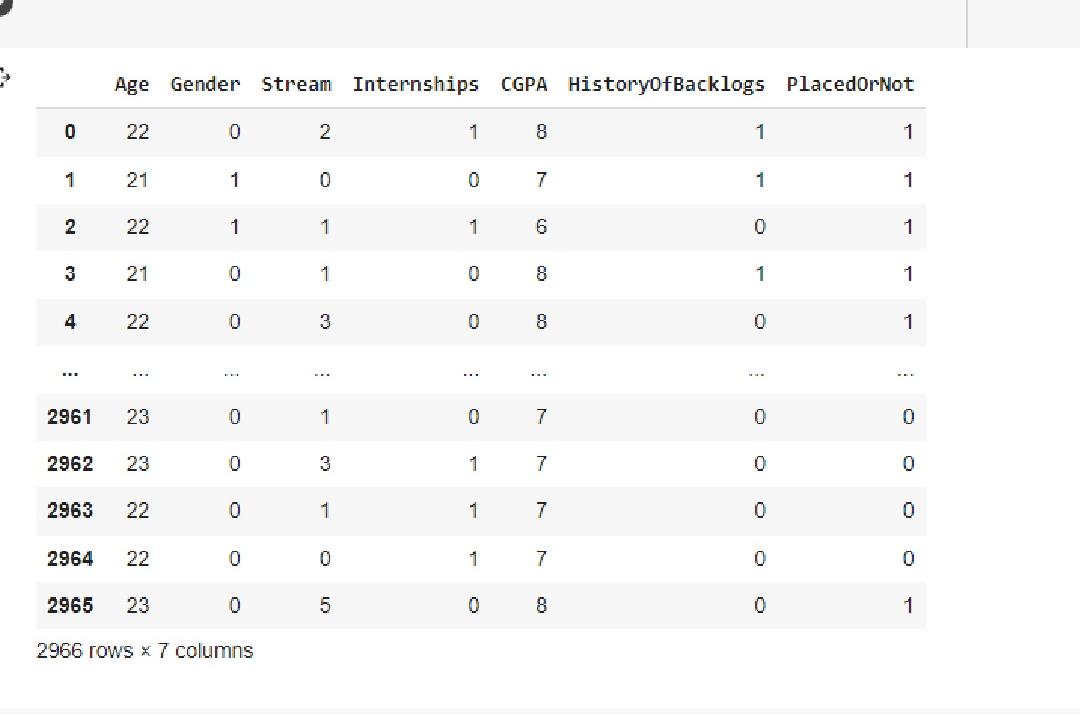


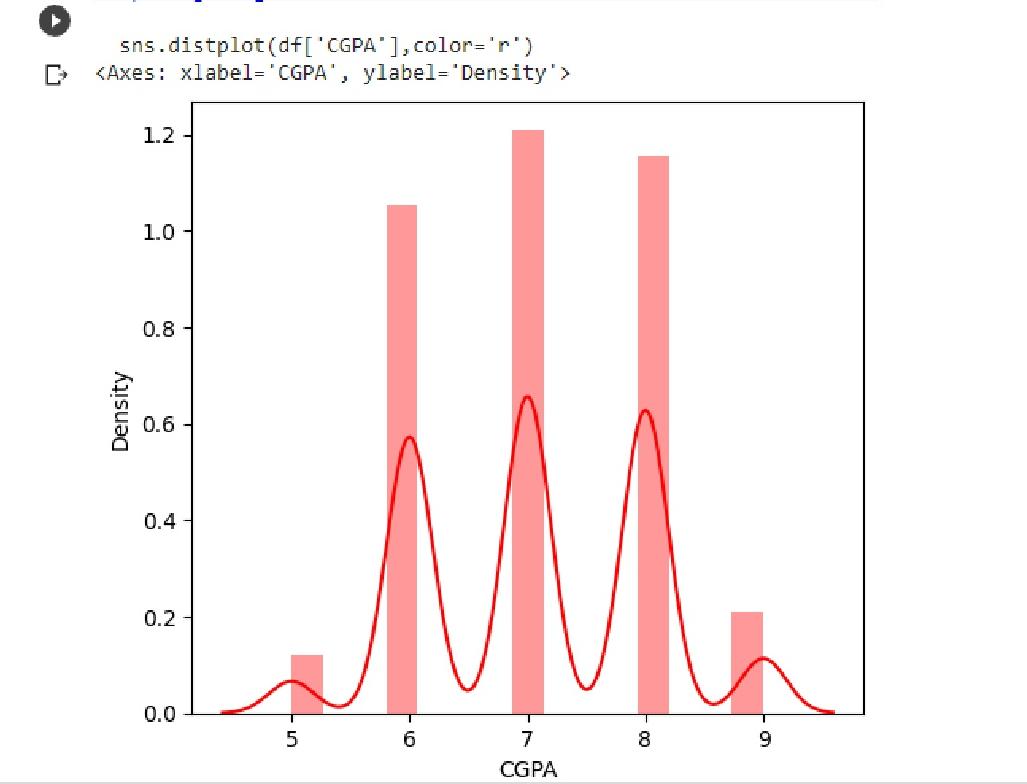


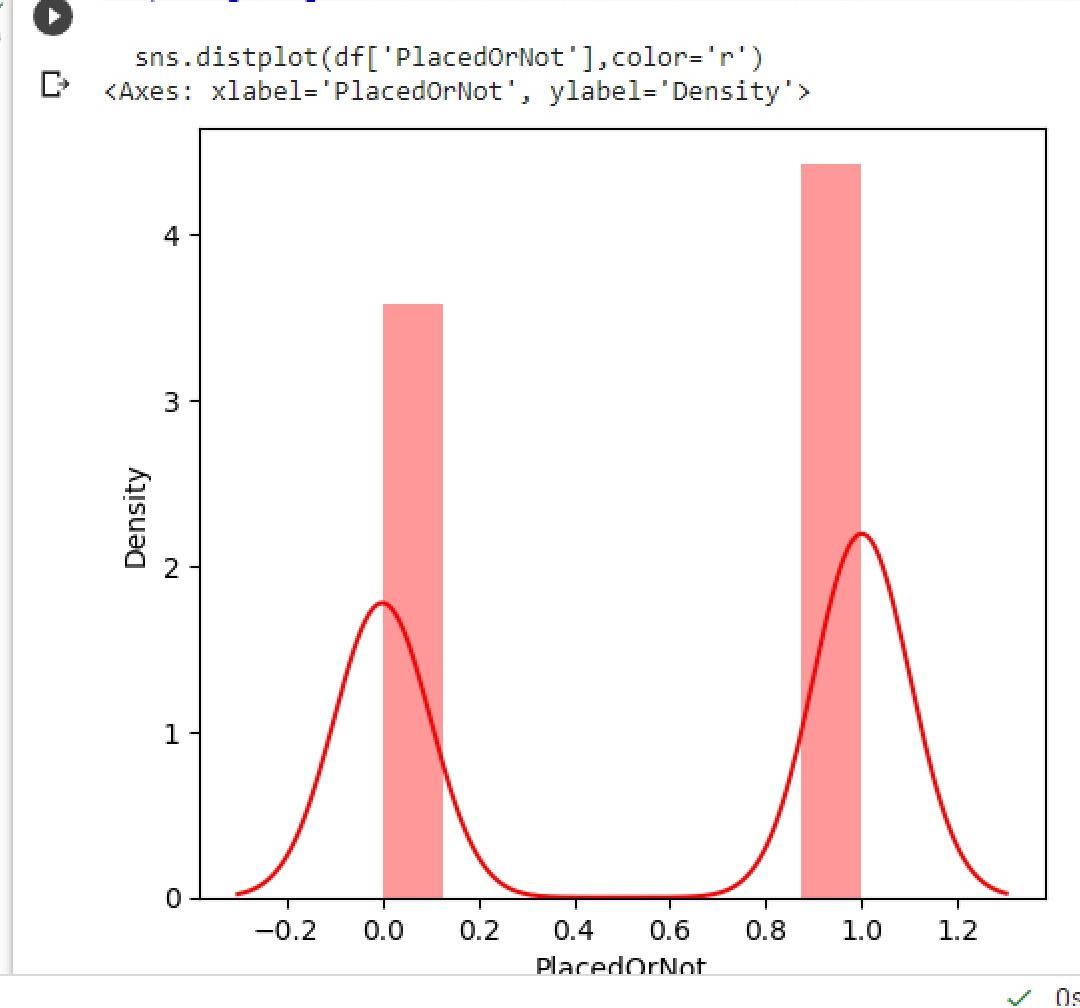


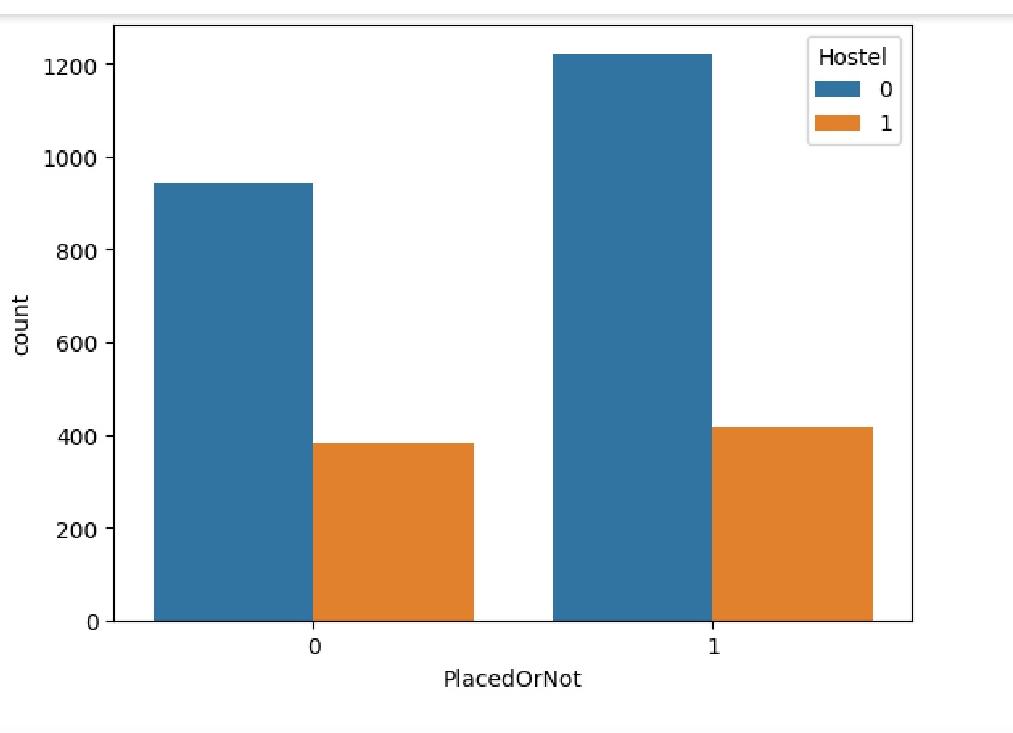


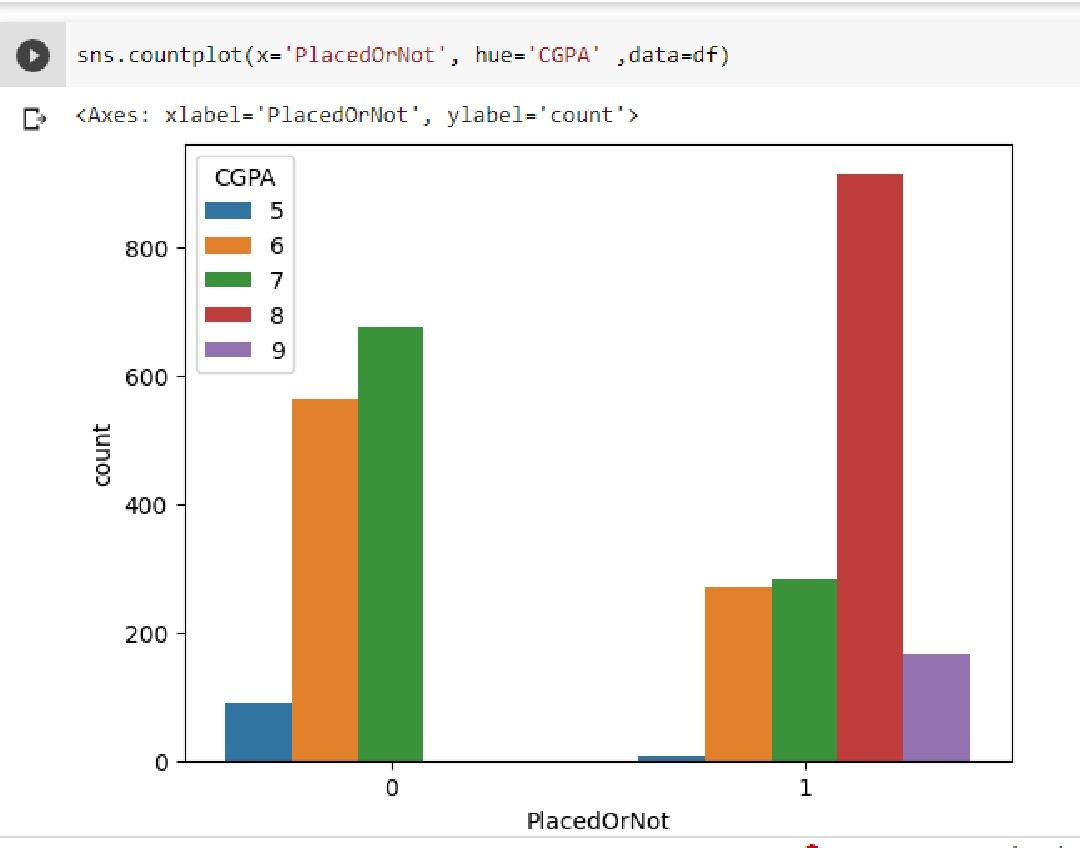


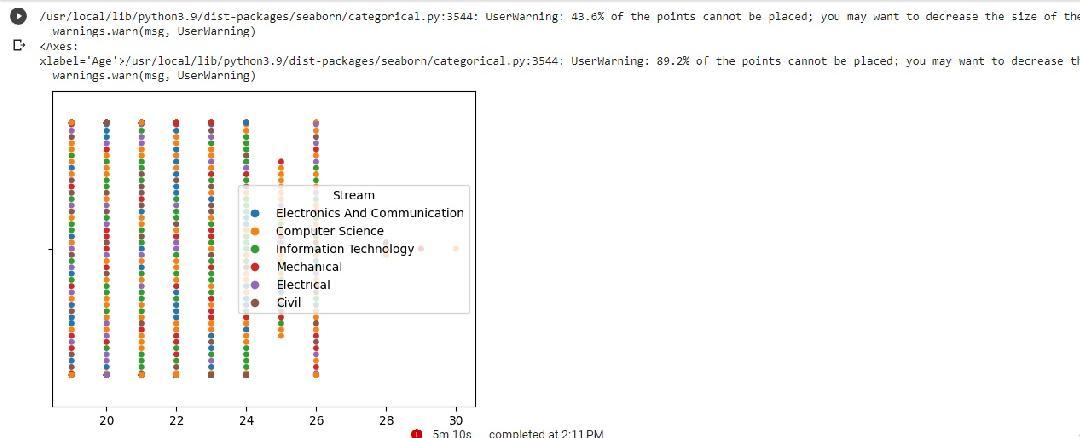


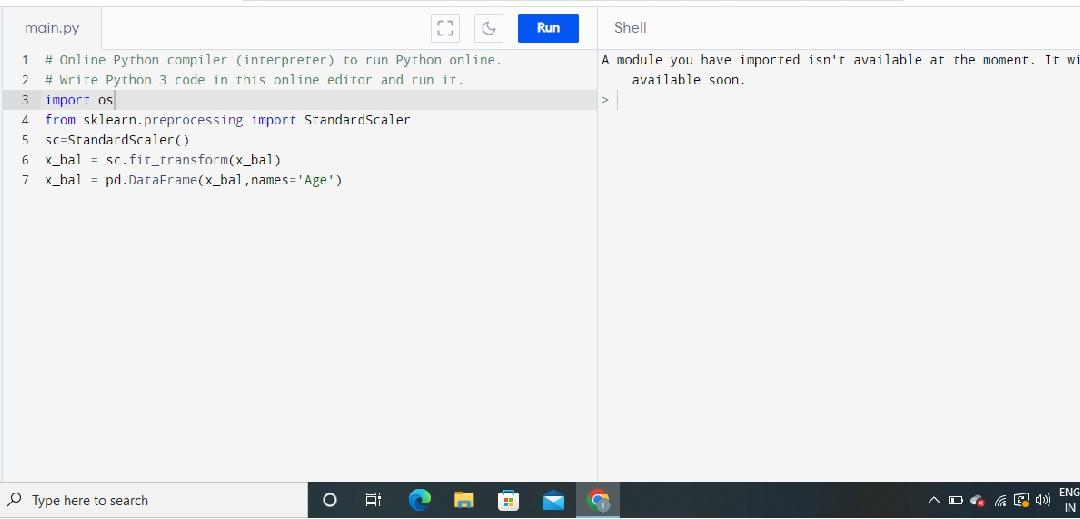


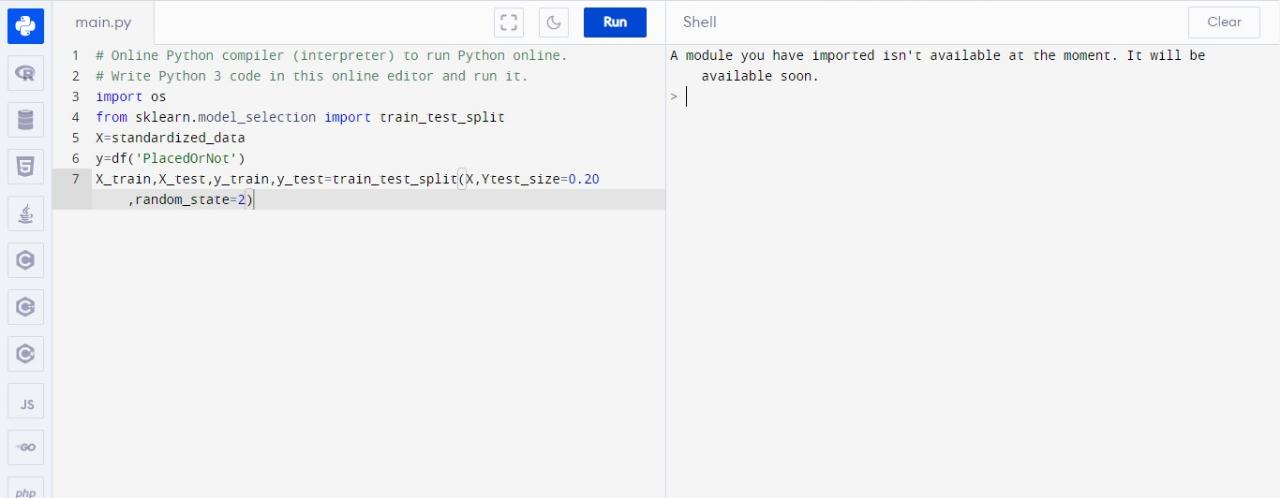


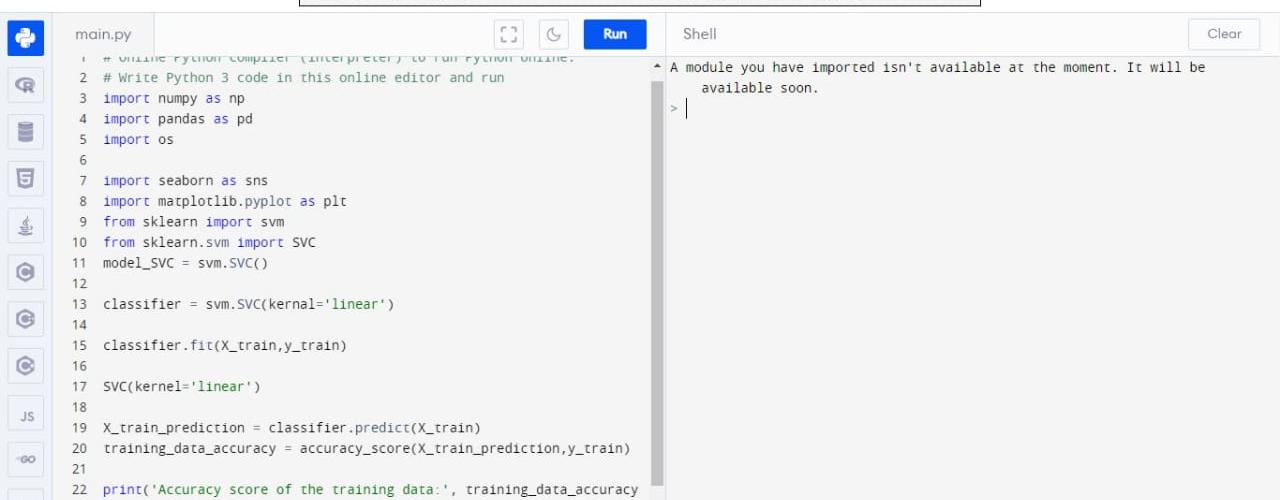


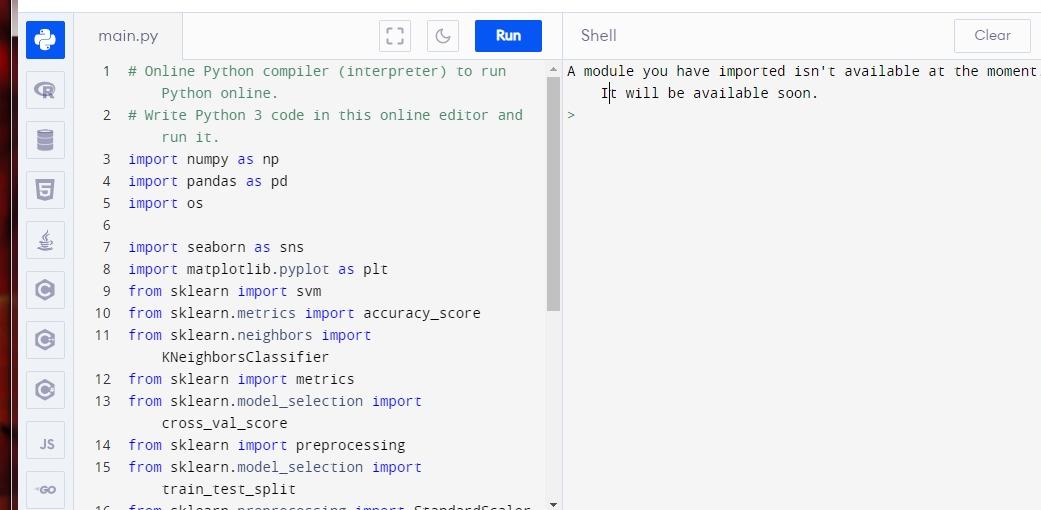


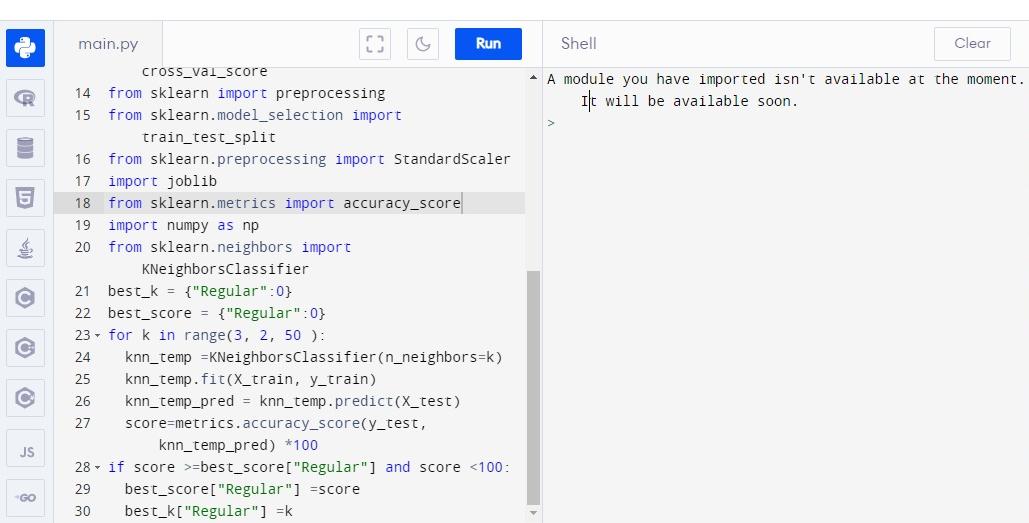


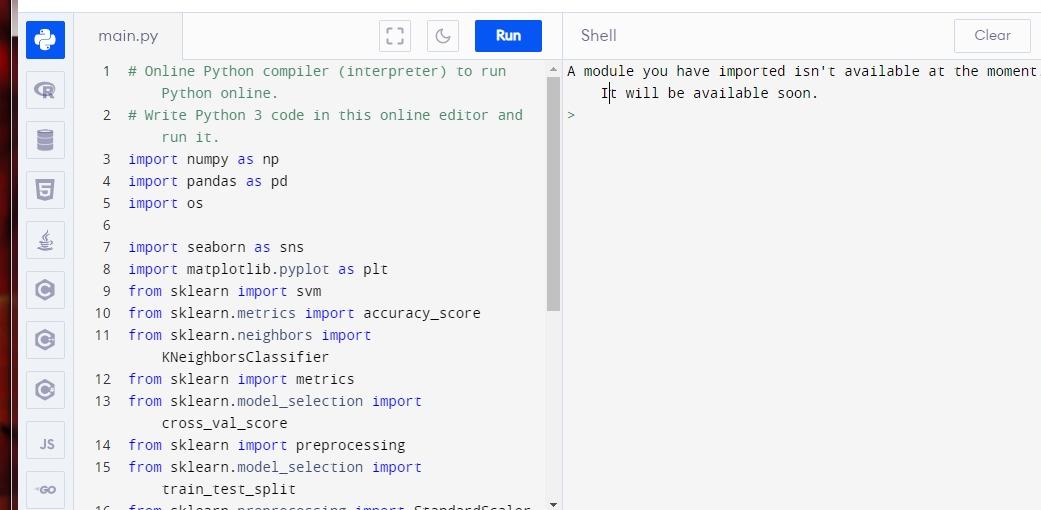


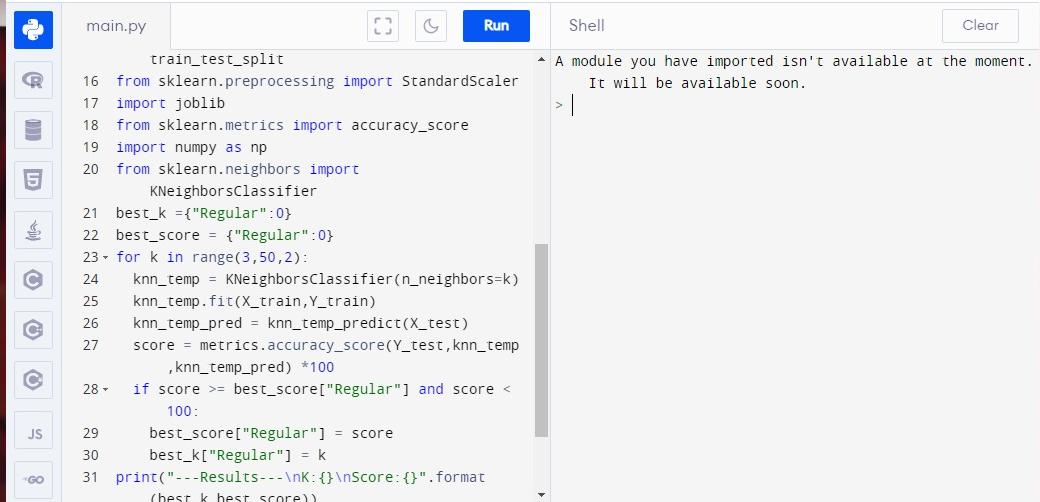


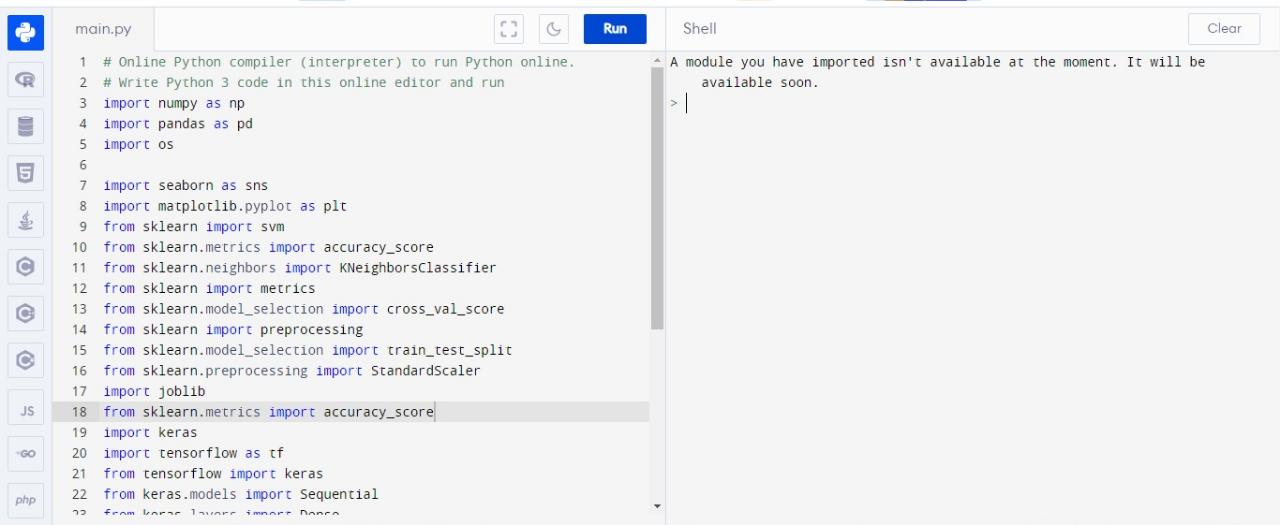


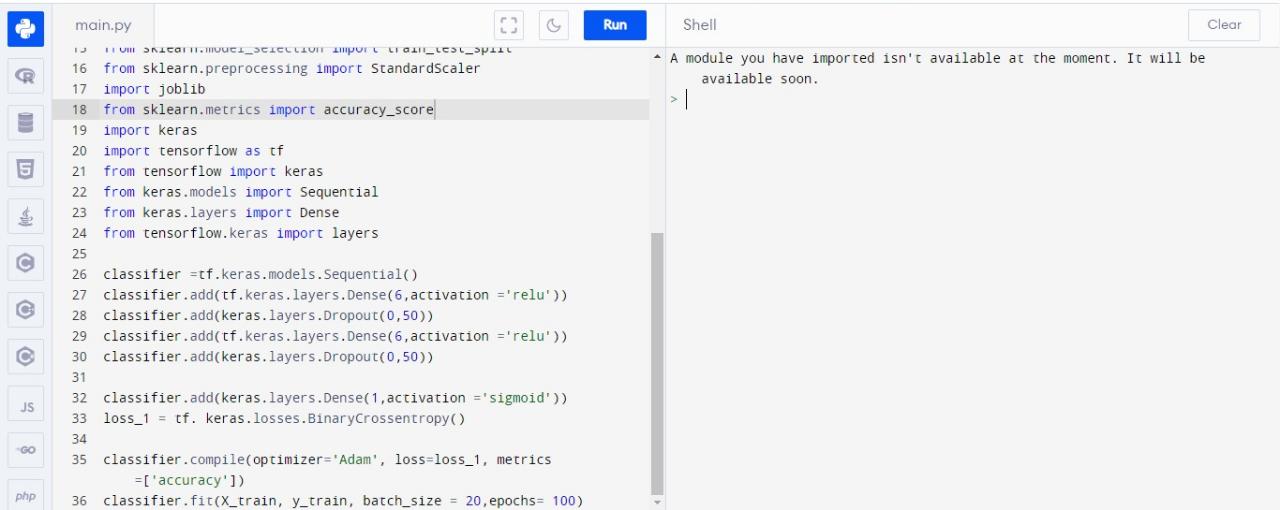


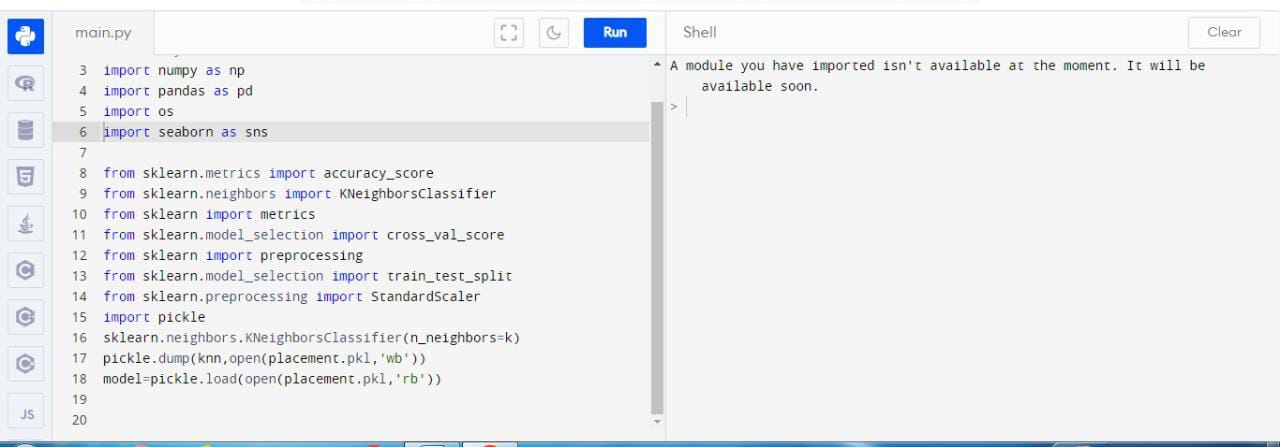


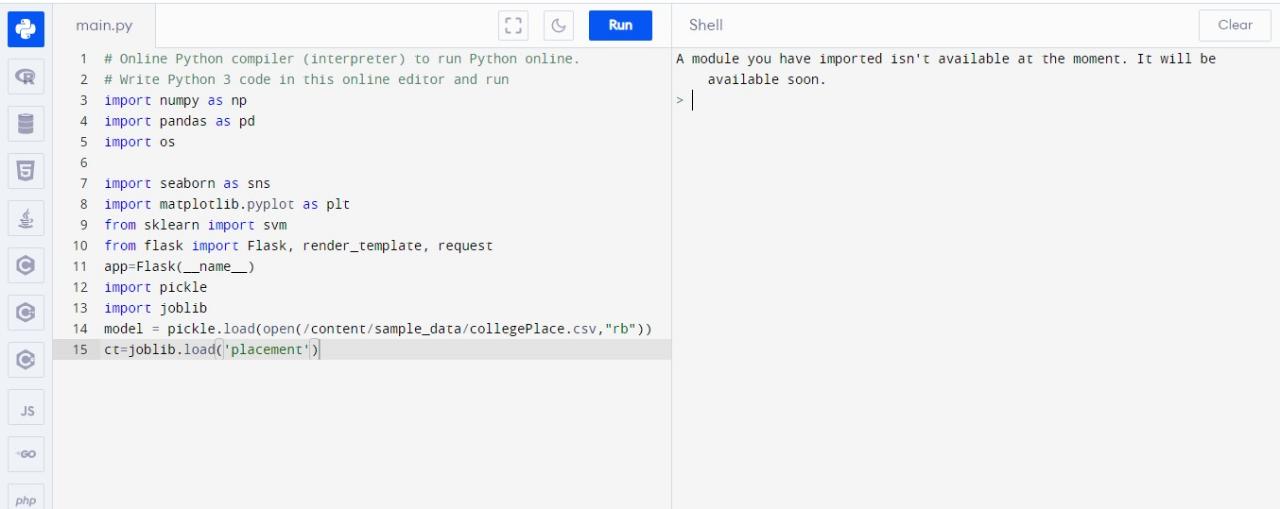




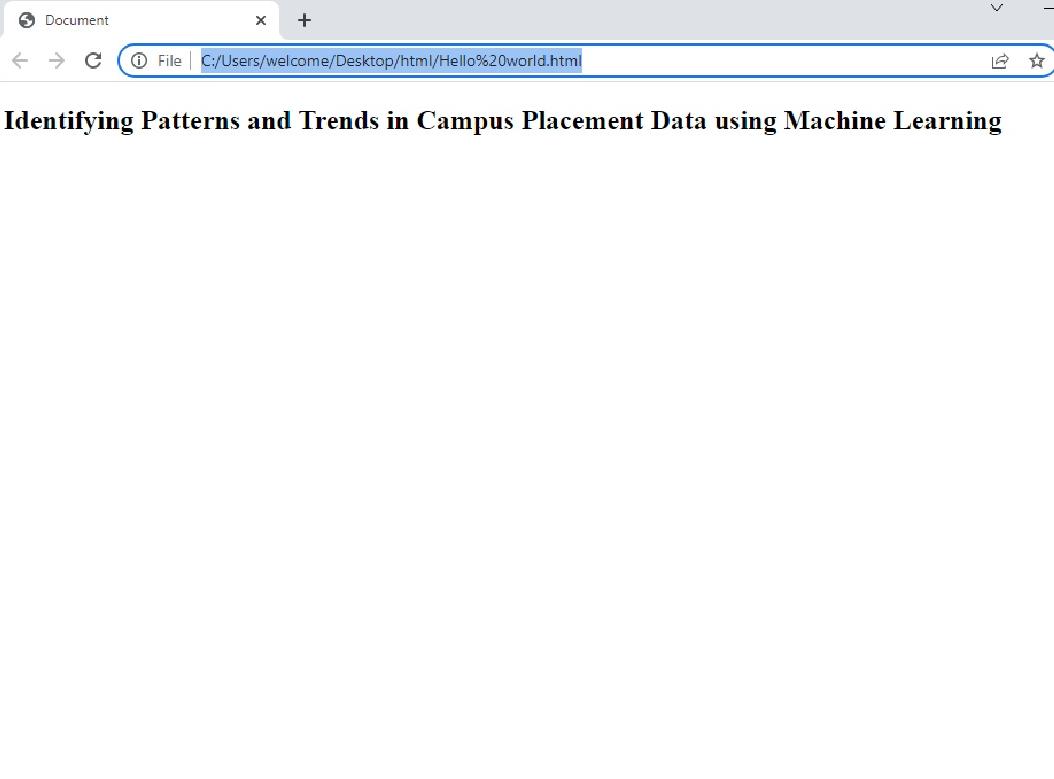


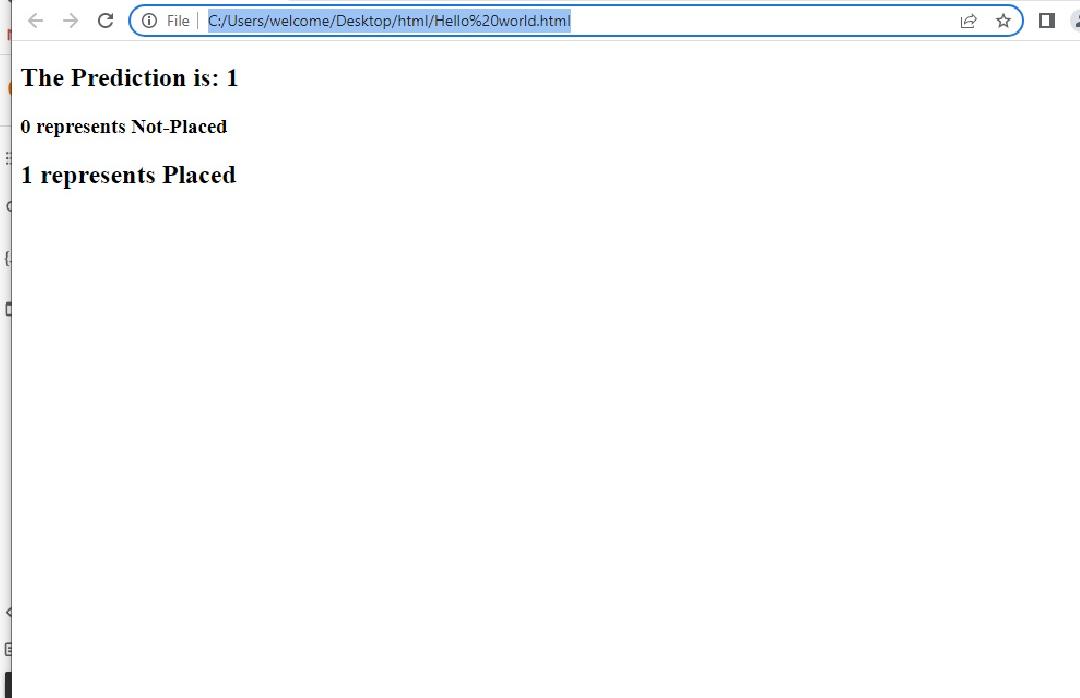


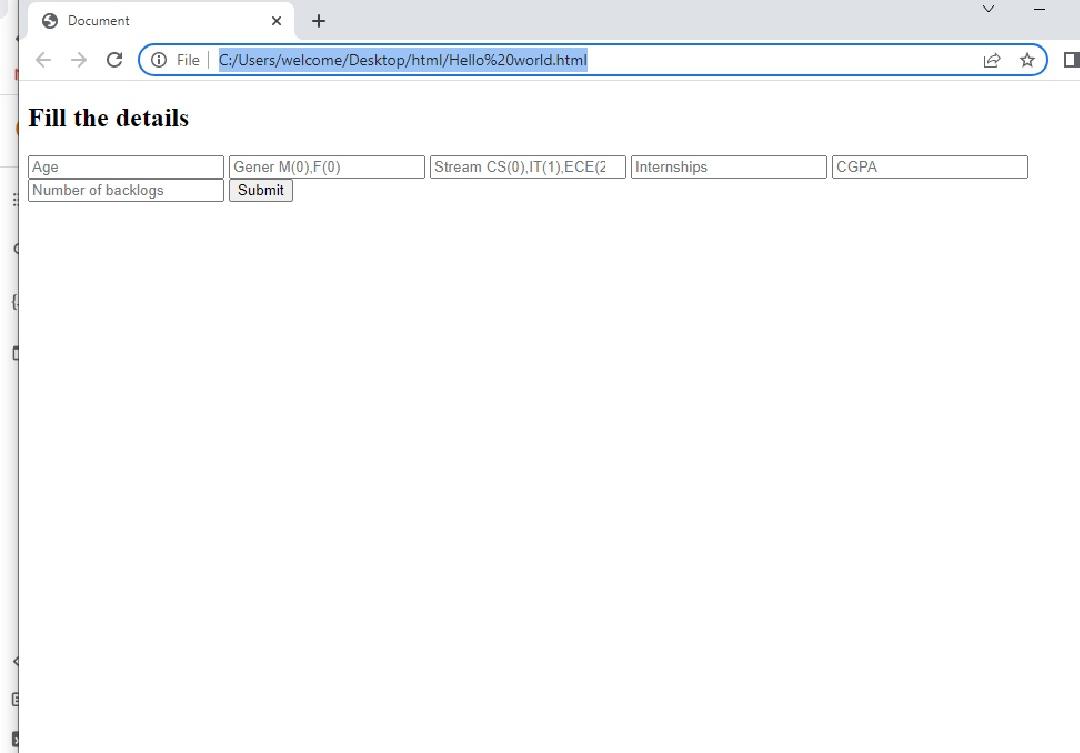














**4. ADVANTAGES & DISADVANTAGES**

**4.1 Advantages:**

A person can profit much by applying pattern recognition techniques. It aids in trend analysis and forecast making, among other things.

1. It aids in the recognition of objects at various angles and distances. simple and fully automated.
2. It is not complex science and doesn't call for original thought.

3. Very helpful for making accurate sales estimate in the financial sector.

4. Effective solutions to current issues.

5. Useful in the medical areas for DNA (Deoxyribonucleic Acid) sequencing and forensic analysis.

**4.2 Disadvantages:**

**1. A Small Staff:**

The majority of campus recruitment initiatives follow a set structure. Companies send out a small hiring team to visit institutions. Now that the personnel is outnumbered by the large number of candidates, coordinating the entire drive presents several challenges. Engaging candidates is a difficult task, and occasionally certain candidates are not properly informed by the organization about the next steps.

The result is a process that is incredibly cumbersome, slow, and ineffective.

**2. Insufficient branding:**

Diversity and inclusivity are essential components of every company's culture, and they can only exist when employees can juggle their work and contribute effectively to the business at the same time. Freshmen are the first to be hired because they are untrained and may be trained from scratch.

However, because to time constraints, most companies are unable to visit most institutions or, if they can, they try to finish the drive quickly, which creates a lot of issues. Keep in mind that a company's branding is largely influenced by how it conducts campus drives. A antiquated campus hiring strategy can quickly put your business behind rivals.

1. **Numerous Candidate:**

Let's face it: thousands of applicants will be vetted during any Campus Drive. Typically, 2-4 representatives from companies visit colleges. This implies that there are more staff members than available to screen applicants. Any company that values its employees won't advocate placing an unreasonable strain on any of them. Still, it's a difficult task to evaluate thousands of applicants in only a few days. Additionally, it's important to conduct the screening effectively to avoid passing on eligible applicants.

Another difficult issue is to establish effective communication with thousands of prospects at once.

1. **A bad hiring experience:**

Due to the candidate-driven nature of the employment market, Campus Drives' first impressions count for a lot. Any candidate will pick a company that employs contemporary technologies for hiring over one that uses time-consuming, old approaches. Additionally, if a corporation doesn't perform campus drives properly, they run the risk of passing up on top talent.

1. **The significance of a resume:**

The resume is a crucial document that can completely change the game for any candidate. It is always the first document that students and recruiters refer to in any Campus Drive. Yes, factors like a strong resume and the location of a student's internship contribute a lot, but candidates are more complicated than that. Employers frequently make hiring judgments based only on a candidate's resume, missing out on top talent even after physically visiting the campus.

They ought to develop a system that places emphasis on the candidate’s completes personality rather than just one document.

**5. APPLICATIONS**

**Applications of Pattern Recognition:**

Trend analysis: Pattern recognition aids in spotting trends in the provided data so that the proper analysis may be performed. Future sales, for instance, might be forecast by looking at recent trends in the sales generated by a specific firm or organization.

Help - Pattern is an essential component of our daily lives. It greatly facilitates our daily activities. Today's market is flooded with software and applications that utilize machine learning algorithms to detect the presence of obstacles and warn the user of potential void misses.

E-commerce-Visual search engines for e-commerce identify the requested item based on its specifications and return relevant results. Recommender systems are used by the majority of websites that are solely focused on online buying. These programmers gather information about each purchase made by a customer and provide recommendations. Making accurate predictions requires analyzing historical trends, which is how all these activities are completed.

Computer vision-The user interacts with the computer vision system by providing an image or video as input. To look for patterns, the machine compares it to thousands, or even millions, of other photos that are saved in its database. An algorithm that is primarily intended for grouping similar-looking items and patterns is used to draw the vital features. Computer vision is the phrase used for this. For instance, cancer detection.

Biometric devices—Brut]alizing technology for face recognition and fingerprint detection, these gadgets secure authentication and security. Machine learning algorithms serve as the underpinning for the application of technologies like face and fingerprint recognition.

**6. CONCLUSION**

* Both both the organization’s and the student's perspectives, the campus placement project is incredibly important. In order to validate the techniques, an effort to improve student performance has been researched and forecast utilizing the classification algorithms Decision Tree, Naive Bays, and Random Forest. To create the model, features are chosen and the algorithms are applied to the data set. Following examination, the accuracy for the Decision Tree is 83%, the Naive Bayes is 84.65%, and the Random Forest is 86%. These findings suggest that the Random Forest classifier, among the machine learning algorithms validated, has the ability to greatly advance the traditional classification techniques for use in placement.
* The machine learning algorithms that we have covered can be used to determine placement trends, which will help universities attract more students in the future. We compared the algorithms and determined the accuracy by taking some student characteristics into account. Here, we employed a deep neural network classifier with 71%, 77%, and 91% accuracy over 1000, 2000, and 5000 iterations.

**Patterns:**

One of the words of the century is machine learning. Due to its features and widely used machine learning applications, it is in great demand. With its incredible powers, it has revolutionized all the industries. Pattern recognition, data mining, analysis, and other areas are some of the many topics and applications of machine learning.

The application of pattern recognition in machine learning is widespread in both technical and non-technical industries nowadays. The research and visualization of numerous trends have benefited from it. Not only has it made analysis and prediction more effective and simple, but it has also improved career opportunities in the industry.

For the purpose of creating accurate forecasts, prestigious corporations like Microsoft, Google, and Amazon are seeking candidates with expertise in pattern recognition and data analysis. Thus, we may state that one of the most cutting-edge areas of machine learning is pattern recognition.

College placement managers face numerous challenges when it comes to handling student information. It is more challenging because complete information management is required, allowing for the creation of a web-based solution to address the issue. The input from the system is more precise. Accessible precise data is available. It increases the dependability of systems and software by assuring security. We believe our project will help a lot of students in the future because admin can check which businesses' students have applied to and been accepted into. We draw the conclusion that the proposed strategy will address the system's flaws as a result.

**7. FUTURE SCOPE**

1. **The future performance of a student will notify the institute and the student whether they will pass the course or not. Machine learning algorithms are highly useful when forecasting something.**

**Scope of Pattern Recognition in Machine Learning:**

**Data Mining:**

Data mining is the process of obtaining valuable information from vast amounts of data originating from many sources. For creating predictions and data analysis, data mining techniques produce relevant data.

**Recommender System:**

Recommender systems are used by the majority of websites for online purchasing. These systems gather information about each client purchase and then use machine learning algorithms to find trends in the purchasing patterns of customers.

**Image processing:**

Image processing-Digital image processing and Analog image processing are the two main categories of image processing. Intelligent machine learning techniques are used in digital image processing to improve the quality of images obtained from remote sources, such satellites.

**Bioinformatics:**

It is a branch of science that employs software and computing tools to generate predictions about biological data. Consider the scenario where a new protein was identified in the lab but its sequence is unknown. In order to anticipate a sequence based on comparable patterns, the unknown protein is compared with a vast number of proteins kept in the database using bioinformatics methods.

**Analysis:**

For spotting significant data trends, pattern recognition is employed. Future forecasts can be made using these trends. In practically every field, whether technical or not, an analysis is necessary. By employing natural language processing to find trends in the posts, tweets sent by a user on Twitter, for instance, aid in sentiment analysis.

**SCOPE:**

This project includes a wide range of student data that may be recorded in CV format and is categorized into different streams. Numerous businesses have access to student data, which they can update and manage. Students receive notifications from the companies. Students have access to prior information

1. View Student Profile Page for Company.

2. View the hiring schedule

3. View Selected Student List

**8. APPENDIX**

**A. Source Code**

import numpy as np

import pandas as pd

import os

import seaborn as sns

import matplotlib.pyplot as plt

from sklearn import svm

from sklearn.metrics import accuracy\_score

from sklearn.neighbors import KNeighborsClassifier

from sklearn import metrics

from sklearn.model\_selection import cross\_val\_score

from sklearn import preprocessing

from sklearn.model\_selection import train\_test\_split

from sklearn.preprocessing import StandardScaler

import  joblib

from sklearn.metrics import accuracy\_score

df=pd.read\_csv(r"/content/sample\_data/collegePlace.csv")

df.head()

df.info()

df.isnull().sum()

def transformationplot(feature):

    plt.figure(figsize=(12,5))

    plt.subplot(1,2,1)

    sns.distplot(feature)

transformationplot(np.log(df['Age']))

df=df.replace(['Male'],[0])

df=df.replace(['Female'],[1])

df=df.replace(['Computer Science','Information Technology','Electronics And Communication','Mechanical','Electrical','Civil'],[0,1,2,3,4,5])

df=df.drop(['Hostel'],axis=1)

plt.figure(figsize=(12,5))

plt.subplot(121)

sns.distplot(df['CGPA'],color='r')

plt.figure(figsize=(12,5))

plt.subplot(121)

sns.distplot(df['PlacedOrNot'],color='r')

import seaborn as sns

sns.countplot(x='PlacedOrNot', hue='Hostel' ,data=df)

sns.countplot(x='PlacedOrNot', hue='CGPA' ,data=df)

import os

from sklearn.preprocessing import StandardScaler

sc= StandardScaler()

x\_bal=sc.fit\_transform(x\_bal)

x\_bal= pd.DataFrame(x\_bal,name='Age')

import seaborn as sns

import pandas as pd

from pandas import read\_csv

pd=read\_csv("/content/sample\_data/collegePlace.csv")

sns.swarmplot(x='Age', hue ='Stream',data=df)

import os

from sklearn.model\_selection import train\_test\_split

X=standardized\_data

y=df('PlacedOrNot')

X\_train,X\_test,y\_train,y\_test=train\_test\_split(X,y,test\_size=0.20,random\_state=2)

import numpy as np

import pandas as pd

import os

import seaborn as sns

import matplotlib.pyplot as plt

from sklearn import svm

from sklearn.svm import SVC

model\_SVC = svm.SVC()

classifier = svm.SVC(kernal='linear')

classifier.fit(X\_train,y\_train)

SVC(kernel='linear')

X\_train\_prediction = classifier.predict(X\_train)

training\_data\_accuracy = accuracy\_score(X\_train\_prediction,y\_train)

print('Accuracy score of the training data:', training\_data\_accuracy)

import keras

import tensorflow as tf

from tensorflow import keras

from keras.models import Sequential

from keras.layers import Dense

from tensorflow.keras import layers

classifier = tf.keras.models.Sequential()

classifier.add(tf.keras.layers.Dense(6,activation ='relu'))

classifier.add(keras.layers.Dropout(0,50))

classifier.add(tf.keras.layers.Dense(6,activation ='relu'))

classifier.add(keras.layers.Dropout(0,50))

classifier.add(keras.layers.Dense(1,activation ='sigmoid'))

loss\_1 = tf. keras.losses.BinaryCrossentropy()

classifier.compile(optimizer='Adam', loss=loss\_1, metrics=['accuracy'])

<section id="hero"class="d-flex flex-column justify-content-center">

 <div class="container">

  <div class="row justify-content-center">

    <div class="col-xl-8">

      <h1>Identifying Patterns and Trends in Campus Placement Data using Machine Learning</h1>

    </div>

  </div>

 </div>

</section>

<section id="about" class="about">

  <div class="container">

    <div class="section-title">

    <h2>Fill the details</h2>

    </div>

     <div class="row content">

      <div class="first">

       <form action="{{url\_for('y\_predict')}}" method="POST">

           <input type="number" id="sen1" name="sen1" placeholder="Age">

           <input type="number" id="sen2" name="sen2" placeholder="Gener M(0),F(0)">

           <input type="number" id="sen3" name="sen3" placeholder="Stream CS(0),IT(1),ECE(2),Mech(3),EEE(4)Civil(5)">

           <input type="number" id="sen4" name="sen4" placeholder="Internships">

           <input type="number" id="sen5" name="sen5" placeholder="CGPA">

           <input type="number" id="sen6" name="sen6" placeholder="Number of backlogs">

           <input type="submit" value="Submit">

       </form>

      </div>

     </div>

</div>

  </section>

<section id="hero" class="d-flex flex-column justify-content-center">

     <div class="container">

      <div class="row justify-content-center">

       <div class="col-xl-8">

           <h1>The Prediction is: {{y}} </h1>

           <h3> 0 represents Not-Placed </h3>

           <h2> 1 represents Placed </h2>

       </div>

      </div>

     </div>

</section>

import numpy as np

import pandas as pd

import os

import seaborn as sns

import matplotlib.pyplot as plt

from sklearn import svm

from flask import Flask, render\_template, request

app=Flask(\_name\_)

import pickle

import joblib

model = pickle.load(open(/content/sample\_data/collegePlace.csv,"rb"))

ct=joblib.load('placement')

@app.route('/')

def hello():

   return render\_template("index.html")

@app.route('/guest',methods=["POST"])

def Guest():

  sen1=request.form["sen1"]

  sen2=request.form["sen2"]

  sen3=request.form["sen3"]

  sen4=request.form["sen4"]

  sen5=request.form["sen5"]

  sen6=request.form["sen6"]

@app.route('/y\_predicct',methods=["POST"])

def y\_predict():

  x\_test=[[(yo) for yo in request.fom.values()]]

  prediction =model.predict(x\_test)

  prediction = prediction[0]

  return render\_template("secondpage.html",y=prediction)

app.run(debug=True)