Internship ID: TLS21A472

Name: Sunethra Ranganath

Semester: 7th

College: Don Bosco Institute of Technology

Assignment 2

1. Dataset Information:

Dataset chosen: Engineering Placements Prediction

Dataset URL: https://www.kaggle.com/tejashvi14/engineering-placements-prediction

Dataset Description: The dataset is Engineering College Placements Data of a University compiled over 2 years (2013 and 2014). We use this data to Predict And Analyse whether a student gets placed, based on his/her background.

Columns and Description:

Age: Age at the time of final year

Gender: Gender of candidate

Stream: Engineering stream that the candidate belongs to

Internships: Number of internships undertaken during the course of studies, not necessarily related to college studies or stream

CGPA: CGPA till 6th semester

Hostel: Whether student lives in college accomodation

HistoryOfBacklog: Whether student ever had any backlogs in any subjects

PlacedOrNot: Target Variable

2. Code:

```
import pandas as pd

import numpy as np

import matplotlib.pyplot as plt


dataset=pd.read_csv("collegePlace.csv")

#print(dataset)
```

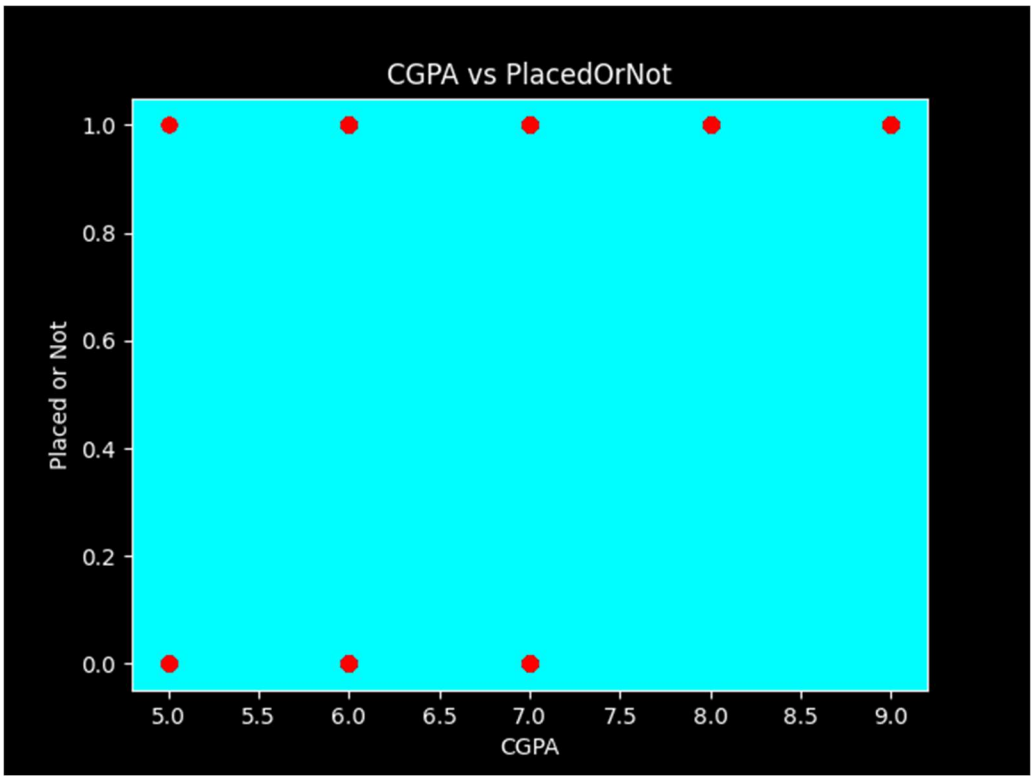
```
#To count % of IT students placed:
count_i=0
count_p=0
for i in dataset["Stream"]:
    if i=="Information Technology":
        count_i+=1
#print(count_i)
for i in range(0,len(dataset.index)):
    if dataset.iloc[i,2]=="Information Technology" and dataset.iloc[i,7]==1:
        count_p+=1
#print(count_p)
percent_p=(count_p/count_i)*100
print("\nPercentage of Information Technology students placed = ",percent_p)

#To plot graph of CGPA vs PlacedOrNot
x=dataset['CGPA'].values
#print(x)
y=dataset.PlacedOrNot.values
#print(y)
plt.style.use('dark_background')
ax=plt.axes()
ax.set_facecolor("cyan")
plt.scatter(x,y,c='r')
plt.title("CGPA vs PlacedOrNot")
plt.xlabel("CGPA")
plt.ylabel("Placed or Not")
plt.show()

#To predict in a student will get placed or not
print("\n\t\t\tPrediction of Placement:")
```

```
feature_cols=['Age','Internships','CGPA','Hostel','HistoryOfBacklogs']
x=dataset[feature_cols]
#print(x)
y=dataset.PlacedOrNot
#print(y)
from sklearn.model_selection import train_test_split
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.25,random_state=0)
from sklearn.linear_model import LogisticRegression
logreg=LogisticRegression()
logreg.fit(x_train,y_train)
y_pred=logreg.predict(x_test)
#stud_info=[22,1,8,1,1]
stud_info=[]
age=int(input("Enter Age: "))
stud_info.append(age)
intern=int(input("Enter no. of Internships: "))
stud_info.append(intern)
cgpa=int(input("Enter CGPA: "))
stud_info.append(cgpa)
hostel=int(input("Student lives in Hostel or not (1/0)? "))
stud_info.append(hostel)
back=int(input("Is there any History of Backlog (1/0)? "))
stud_info.append(back)
y_pred1=logreg.predict([stud_info])
print("\nGiven Info about Student: ",stud_info)
#print(y_pred1)
if y_pred1==[1]:
   print("Great! The Student will probably be placed.")
elif y_pred1==[0]:
   print("Sorry, there are very less chances of the student getting placed :(")
```

3. Graph and Explanation of the graph:



The graph plotted is student CGPA vs he/she is placed or not.

From this graph we can understand that among the students with lesser CGPA i.e. from 5 to 7, some of them were placed and some were not. Whereas among those who scored a CGPA greater than 7, all were placed. So, we can come to a possible conclusion that, the greater the CGPA, the more the chance of getting placed.

4. Output: