Experiment – 2

Student Name: Vivek Kumar **UID: 21BCS8129**
Branch: BE-CSE(LEET) **Section/Group: WM-20BCS-616/A**
Semester: 5th **Date of Performance: 16/08/2022**
Subject Name: Machine Learning Lab **Subject Code: 20CSP-317**

1. Aim/Overview of the practical:

For a given set of training data examples stored in a .CSV file, implement and demonstrate the Candidate-Elimination algorithm to output a description of the set of all hypotheses consistent with the training examples.

2. Task to be done/ Which logistics used:

Data Visualization using matplotlib

3. Algorithm/Flowchart (For programming-based labs):

4. Steps for experiment/practical/Code:

```
import pandas as pd

data = pd.read_csv("/content/drive/MyDrive/Data/Students_data.csv")

data.head(10)

data.tail()

pip install matplotlib

import matplotlib.pyplot as plt

plt.scatter(data['race'],data['GPA'])
plt.title('Scatter Plot')
plt.xlabel('Race')
plt.ylabel('GPA')
plt.show()

plt.scatter(data['race'],data['GPA'],c=data['Probability'],s=data['Statistics'])
plt.title('Scatter Plot')
plt.xlabel('Race')
plt.ylabel('GPA')
plt.colorbar()
```

```
plt.show()

plt.bar(data['race'],data['GPA'])
plt.title('Bar Plot')
plt.xlabel('Race')
plt.ylabel('GPA')
plt.show()

plt.hist(data['race'])
plt.title('Histogram Plot')
plt.show()

import seaborn as sb

sb.scatterplot(x='race',y='GPA',data=data)

sb.scatterplot(x='race',y='GPA',data=data,hue='gender')

sb.lineplot(x='race',y='GPA',data=data)

sb.lineplot(x='race',y='GPA',data=data,hue='gender')

sb.barplot(x='race',y='GPA',data=data,hue='gender')

sb.histplot(x='GPA', data=data, kde=True, hue='gender')
```

5. Observations/Discussions/ Complexity Analysis:

In this have done Data visualization with matplotlib and used various function such as scatter, scatter with colorbar, bar with x-y label and hist. Another library which I have used seaborn and plotted various graph such as scatterplot, lineplot, barplot and histplot.

6. Result/Output/Writing Summary:

Machine Learning Lab - 2

To Implement data Visualization

```
[1] from google.colab import drive
    drive.mount('/content/drive')
```

```
Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.mount("/content/drive", force_remount=True).
```

```
[2] import pandas as pd
```

```
data = pd.read_csv("/content/drive/MyDrive/Data/Students_data.csv")
```

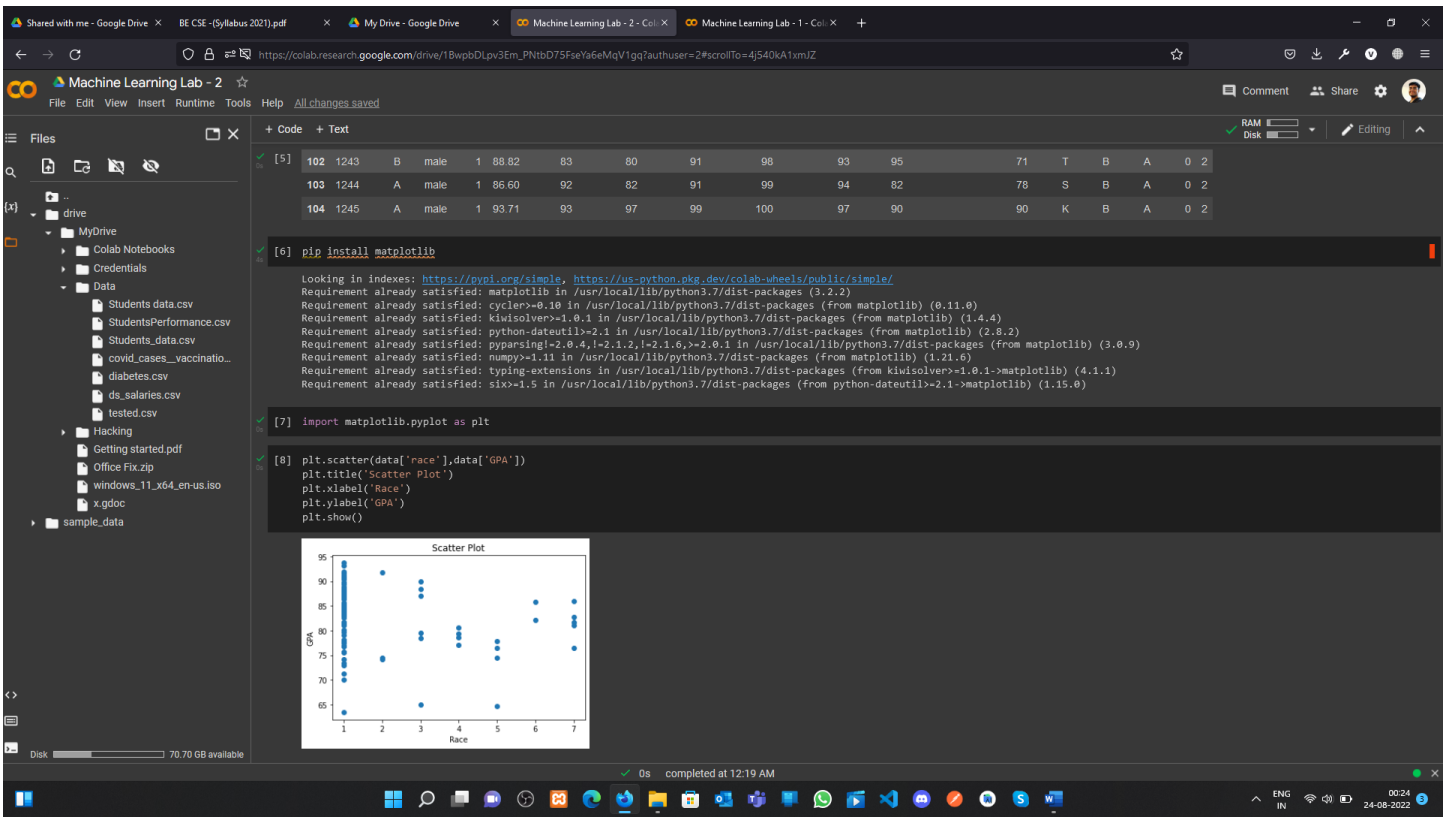
```
[4] data.head(10)
```

	ID	class	gender	race	GPA	Algebra	Calculus1	Calculus2	Statistics	Probability	Measure	Functional_analysis	from1	from2	from3	from4	y
0	1141	A	male	1	73.47	64	81	87	60	74	71	60	A	A	A	3	0
1	1142	A	female	1	71.22	57	50	51	51	55	62	61	B	A	A	2	0
2	1143	A	female	2	74.56	47	48	71	60	61	68	64	C	A	A	0	1
3	1144	A	female	1	72.89	46	72	38	60	29	54	51	D	A	A	0	0
4	1145	A	female	1	70.11	49	45	63	60	66	66	61	E	A	A	0	0
5	1146	A	male	3	65.04	60	60	39	61	65	74	60	F	B	C	0	0
6	1147	A	male	4	77.11	60	43	52	63	71	72	75	G	A	A	0	1
7	1148	A	female	5	64.75	60	38	60	63	70	68	51	H	B	C	0	0
8	1149	B	female	5	77.92	61	60	66	68	80	78	71	I	B	A	0	0
9	1150	A	female	5	76.50	60	61	60	69	73	60	62	H	B	A	0	0

```
[5] data.tail()
```

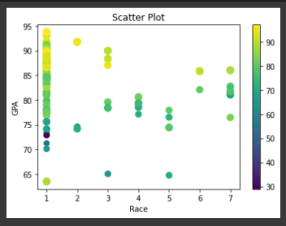
	ID	class	gender	race	GPA	Algebra	Calculus1	Calculus2	Statistics	Probability	Measure	Functional_analysis	from1	from2	from3	from4	y
100	1241	A	female	1	88.34	87	83	92	98	93	86	90	M	B	A	0	1
101	1242	B	male	1	89.84	98	77	95	98	96	88	100	A	B	A	0	1

0s completed at 12:19 AM

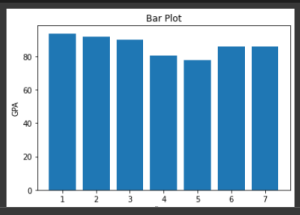


Machine Learning Lab - 2

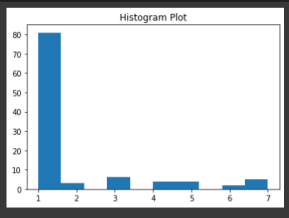
```
[9] plt.scatter(data['race'],data['GPA'],c=data['Probability'],s=data['Statistics'])
    plt.title('Scatter Plot')
    plt.xlabel('Race')
    plt.ylabel('GPA')
    plt.colorbar()
    plt.show()
```




```
[10] plt.bar(data['race'],data['GPA'])
     plt.title('Bar Plot')
     plt.xlabel('Race')
     plt.ylabel('GPA')
     plt.show()
```



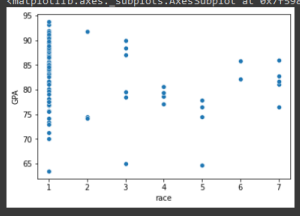

```
[11] plt.hist(data['race'])
     plt.title('Histogram Plot')
     plt.show()
```

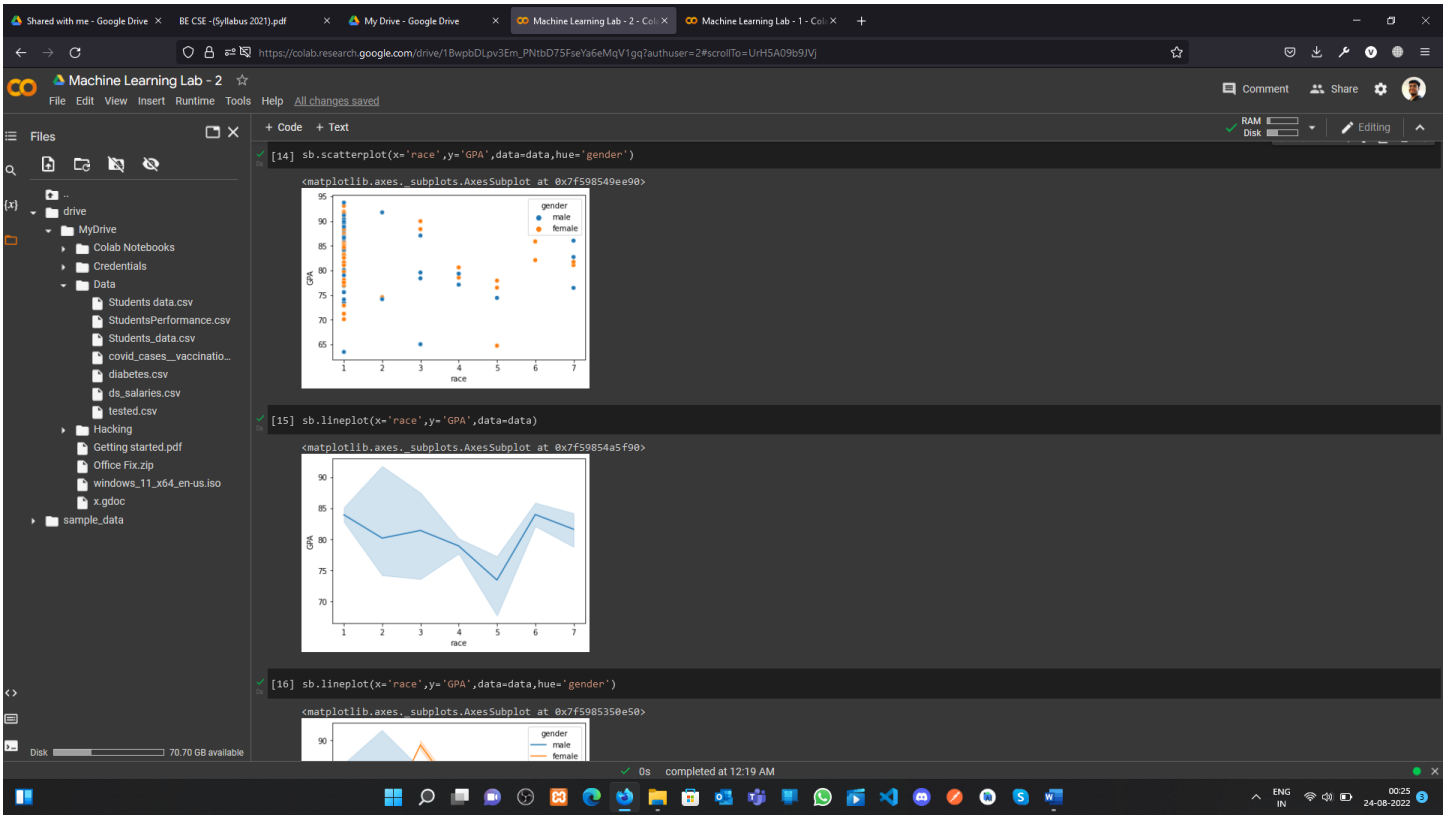


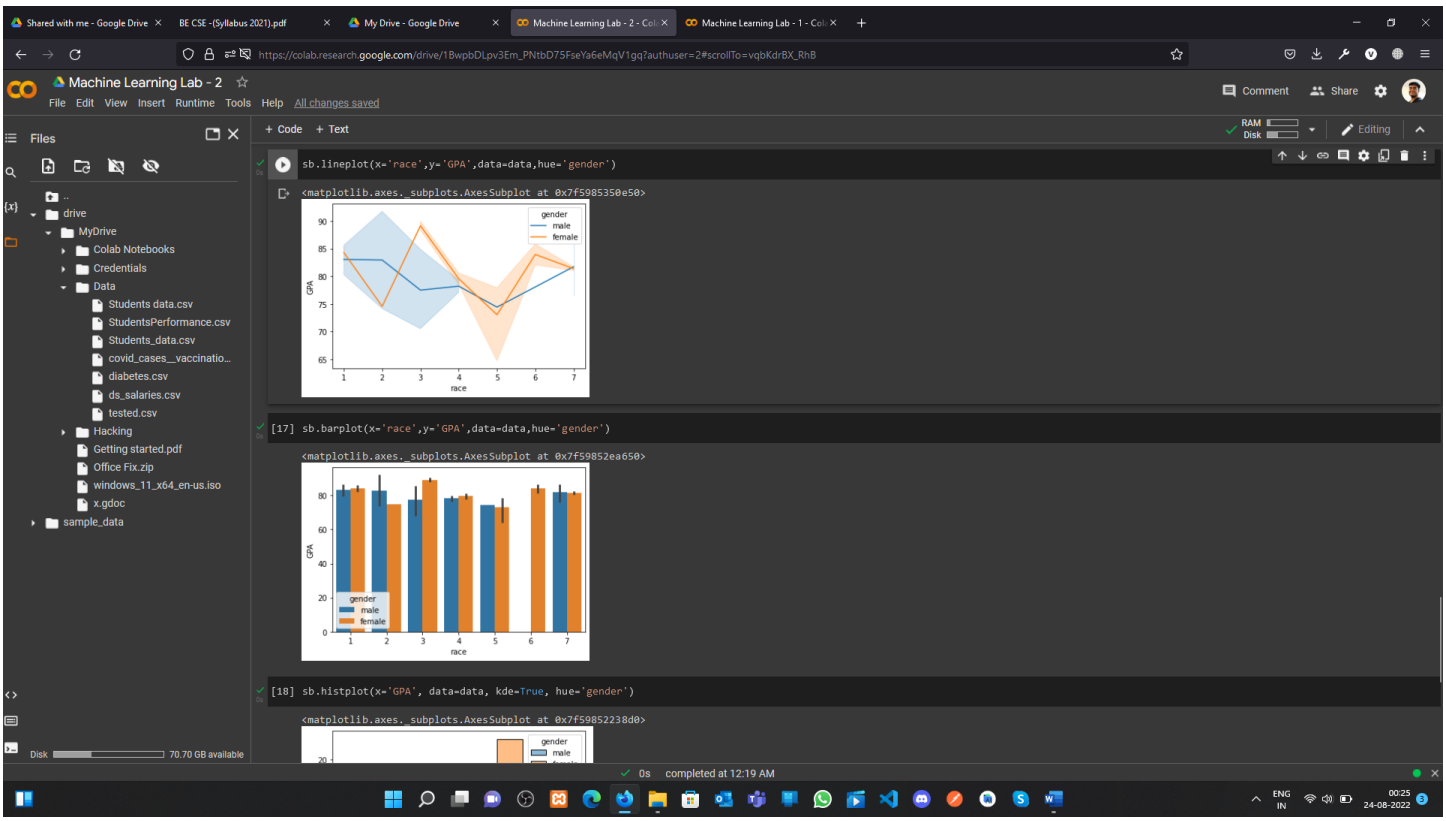

```
[12] import seaborn as sb
```

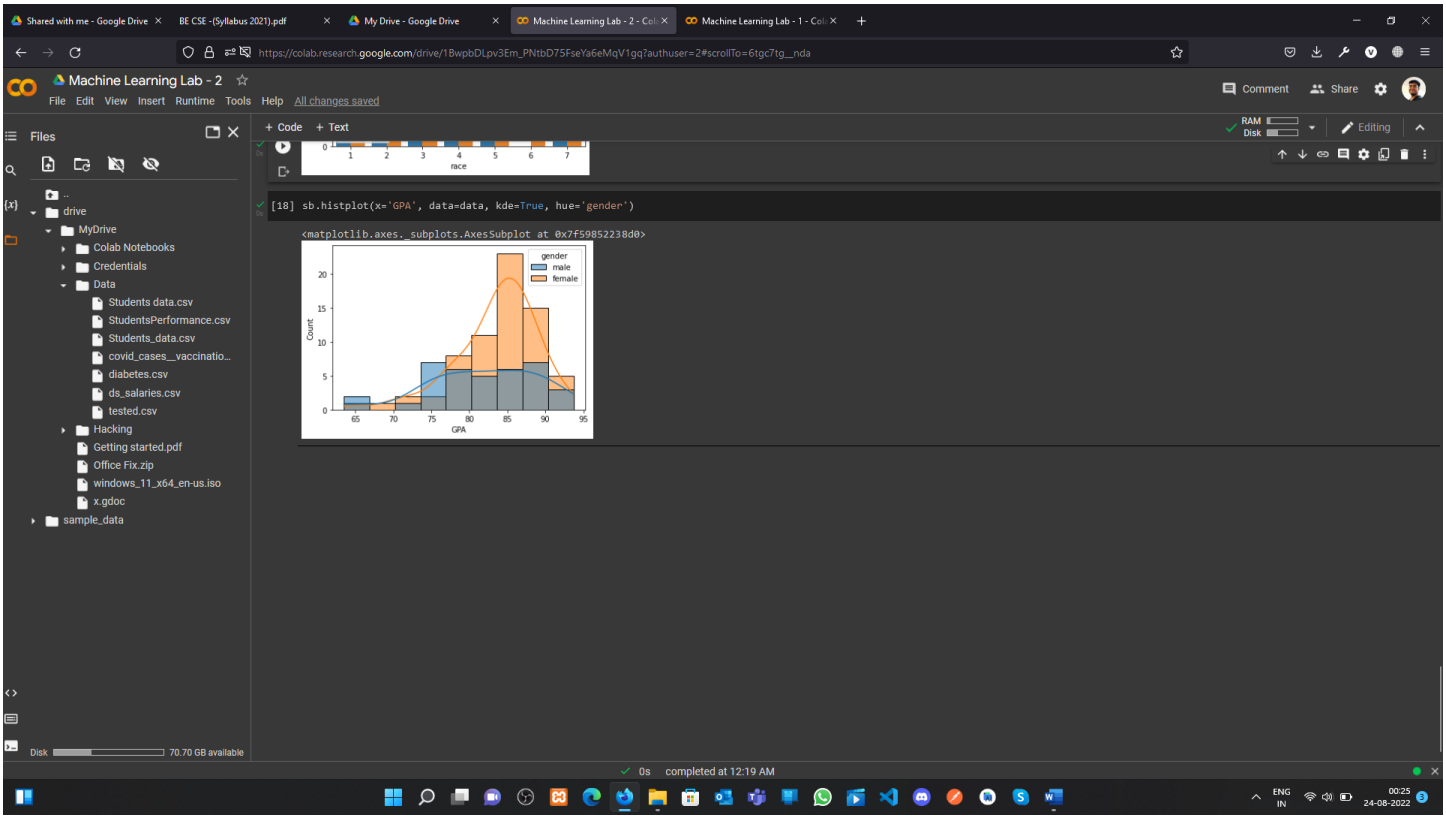
```
[13] sb.scatterplot(x='race',y='GPA',data=data)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f598511210>
```




Machine Learning Lab - 2
File Edit View Insert Runtime Tools Help All changes saved
+ Code + Text
Files
drive
MyDrive
Colab Notebooks
Credentials
Data
Students data.csv
StudentsPerformance.csv
Students_data.csv
covid_cases__vaccinatio...
diabetes.csv
ds_salaries.csv
tested.csv
Hacking
Getting started.pdf
Office Fix.zip
windows_11_x64_en-us.iso
x.gdoc
sample_data
[14] sb.scatterplot(x='race',y='GPA',data=data,hue='gender')
<matplotlib.axes._subplots.AxesSubplot at 0x7f598549ee90>
[15] sb.lineplot(x='race',y='GPA',data=data)
<matplotlib.axes._subplots.AxesSubplot at 0x7f59854a5f90>
[16] sb.lineplot(x='race',y='GPA',data=data,hue='gender')
<matplotlib.axes._subplots.AxesSubplot at 0x7f5985350e50>
gender
male
female
GPA
race
Disk 70.70 GB available
0s completed at 12:19 AM


Machine Learning Lab - 2
File Edit View Insert Runtime Tools Help All changes saved
+ Code + Text
Files
drive
MyDrive
Colab Notebooks
Credentials
Data
Students data.csv
StudentsPerformance.csv
Students_data.csv
covid_cases__vaccinatio...
diabetes.csv
ds_salaries.csv
tested.csv
Hacking
Getting started.pdf
Office Fix.zip
windows_11_x64_en-us.iso
x.gdoc
sample_data
sb.lineplot(x='race',y='GPA',data=data,hue='gender')
<matplotlib.axes._subplots.AxesSubplot at 0x7f5985350e50>
[17] sb.barplot(x='race',y='GPA',data=data,hue='gender')
<matplotlib.axes._subplots.AxesSubplot at 0x7f59852ea650>
[18] sb.histplot(x='GPA', data=data, kde=True, hue='gender')
<matplotlib.axes._subplots.AxesSubplot at 0x7f59852238d0>
gender
male
female
GPA
race
Disk 70.70 GB available
0s completed at 12:19 AM



Learning outcomes (What I have learnt):

1. Data Visualization using matplotlib
2. Data visualization using seaborn

Evaluation Grid (To be created as per the SOP and Assessment guidelines by the faculty):

Sr. No.	Parameters	Marks Obtained	Maximum Marks
1.			
2.			
3.			