PRACTICAL NO: 2A

Aim: Perform Data Loading, Feature selection (Principal Component analysis) and Feature Scoring and Ranking.

Code:

```
from pandas import read_csv

from sklearn.decomposition import PCA

from sklearn.ensemble import ExtraTreesClassifier


#We will use PCA to select best 3 Principal components from Pima Indians Diabetes dataset.

path = 'pima-indians-diabetes.csv'

names = ['preg', 'plas', 'pres', 'skin', 'test', 'mass', 'pedi', 'age', 'class']

dataframe = read_csv(path, names=names)

array = dataframe.values


#Next, we will separate array into input and output components −

X = array[:,0:8]

Y = array[:,8]


#The following lines of code will extract features from dataset −

pca = PCA(n_components=3)

fit = pca.fit(X)

print("Explained Variance: %s",fit.explained_variance_ratio_)

print(fit.components_)


#From the output, we can observe that there are scores for each attribute.

#The higher the score, higher is the importance of that attribute.

model = ExtraTreesClassifier()

model.fit(X, Y)

print("Scores for each attribute")

print(model.feature_importances_)
```

```
#Rank

dataframe["BMIRanking"] = dataframe["mass"].rank(ascending=False).astype("int")

print(dataframe.head()) #prints the first 5 rows.
```

Output:

```
Python 3.7.9 Shell
File  Edit  Shell  Debug  Options  Window  Help
Python 3.7.9 (tags/v3.7.9:13c94747c7, Aug 17 2020, 18:58:18) [MSC v.1900 64 bit (AMD64)] on win32
Type "help", "copyright", "credits" or "license()" for more information.
>>>
================== RESTART: D:/MSc IT/Sem3/ML Pracs/Prac2A.py ==================
Explained Variance: %s [0.88854663 0.06159078 0.02579012]
[[-2.02176587e-03  9.78115765e-02  1.60930503e-02  6.07566861e-02
   9.93110844e-01  1.40108085e-02  5.37167919e-04 -3.56474430e-03]
 [-2.26488861e-02 -9.72210040e-01 -1.41909330e-01  5.78614699e-02
   9.46266913e-02 -4.69729766e-02 -8.16804621e-04 -1.40168181e-01]
 [-2.24649003e-02  1.43428710e-01 -9.22467192e-01 -3.07013055e-01
   2.09773019e-02 -1.32444542e-01 -6.39983017e-04 -1.25454310e-01]]
Scores for each attribute
[0.10966063 0.23830732 0.101225   0.07923757 0.07325185 0.14131782
 0.11869353 0.13830628]
Ranking based on BMI
   preg  plas  pres  skin  test  mass   pedi  age  class  BMIRanking
0     6   148    72    35     0  33.6  0.627   50      1         314
1     1    85    66    29     0  26.6  0.351   31      0         593
2     8   183    64     0     0  23.3  0.672   32      1         697
3     1    89    66    23    94  28.1  0.167   21      0         539
4     0   137    40    35   168  43.1  2.288   33      1          57
>>>
```

1. **Describe the use of diagram using principle of PCA**
2. **2. Explain procedure for computation of principle of datas**

PRACTICAL NO: 2B

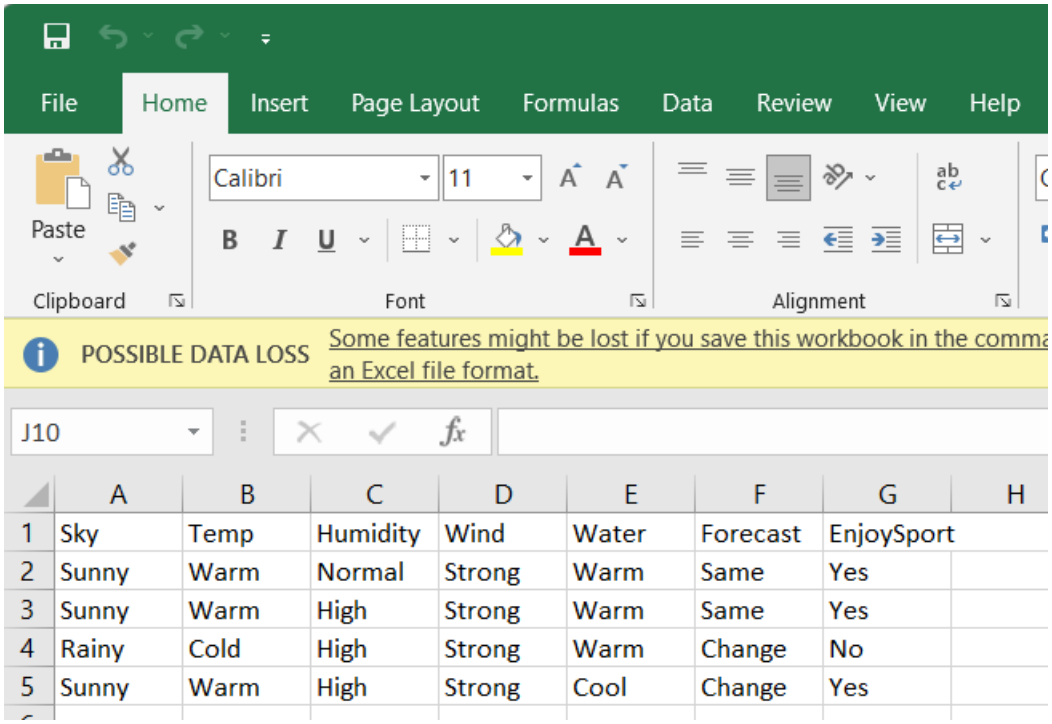
Aim: For a given set of training data examples stored in a .CSV file, implement and demonstrate the Candidate-Elimination algorithm to output a description of the set of all hypotheses consistent with the training examples.

Code:

```
import csv


with open("tennis.csv") as f:
    csv_file=csv.reader(f)
    data=list(csv_file)


    s=data[1][:-1]
    g=[['?' for i in range(len(s))] for j in range(len(s))]


    for i in data:
        if i[-1]=="Yes":
            for j in range(len(s)):
                if i[j]!=s[j]:
                    s[j]='?'
                    g[j][j]='?'


        elif i[-1]=="No":
            for j in range(len(s)):
                if i[j]!=s[j]:
                    g[j][j]=s[j]
                else:
```

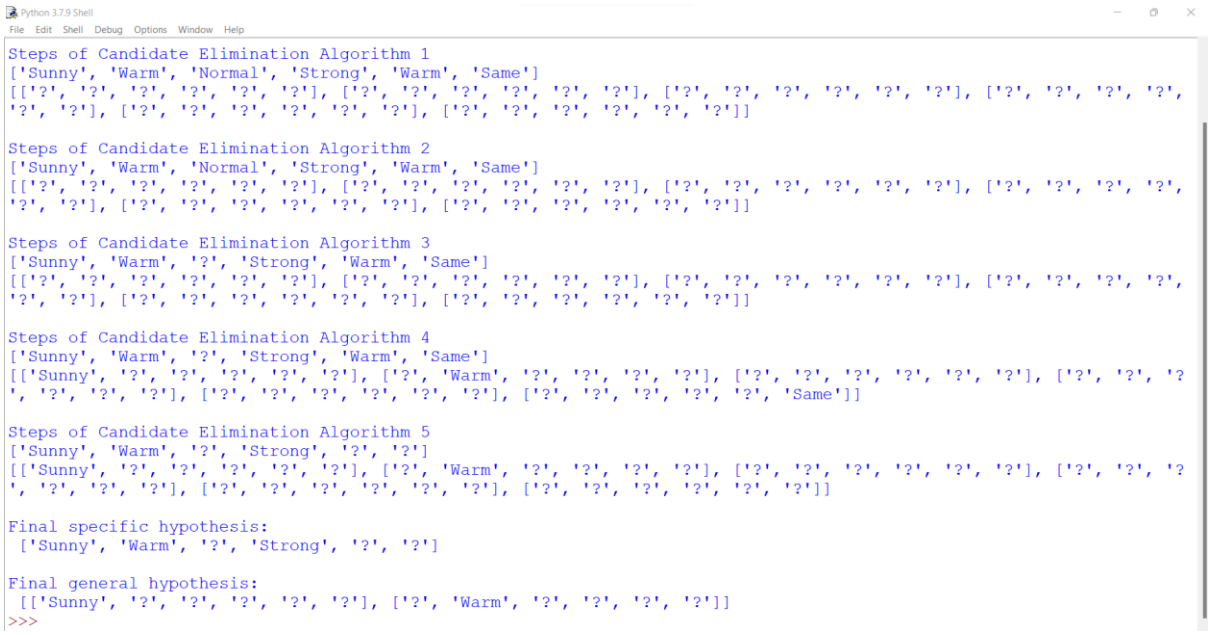
```
                g[j][j]="?"
        print("\nSteps of Candidate Elimination Algorithm",data.index(i)+1)
        print(s)
        print(g)
    gh=[]
    for i in g:
        for j in i:
            if j!='?':
                gh.append(i)
                break
    print("\nFinal specific hypothesis:\n",s)
    print("\nFinal general hypothesis:\n",gh)
```

tennis.csv file



	A	B	C	D	E	F	G	H
1	Sky	Temp	Humidity	Wind	Water	Forecast	EnjoySport	
2	Sunny	Warm	Normal	Strong	Warm	Same	Yes	
3	Sunny	Warm	High	Strong	Warm	Same	Yes	
4	Rainy	Cold	High	Strong	Warm	Change	No	
5	Sunny	Warm	High	Strong	Cool	Change	Yes	

Output:



1. Given following data , compute principle component vectors and the first principle component

x	**2**	**3**	**7**
y	**11**	**14**	**26**

2. Given following data , compute principle component vectors and the first principle component

x	**-3**	**1**	**-2**
y	**2**	**-1**	**3**