



ARTIFICIAL INTELLIGENCE LAB (CSL5402)

Name: Lakhan Kumawat

Roll: 1906055

Program: B.Tech CSE
(5th Sem JUL-DEC 2021)

Assignment - 11

Ques . 1: Read wines.csv dataset (shared with the named “wines.csv” in the file section inside the dataset folder) using PCA.

Ques . 2: Read and analyze the wines.csv dataset (shared with the named “wines.csv” in the file section inside the dataset folder) using PCA calculate the accuracy and F1 score.

```

1 from sklearn.metrics import confusion_matrix
2 import numpy as np
3 import matplotlib.pyplot as plt
4 import pandas as pd
5 import seaborn as sn
6 from sklearn.metrics import accuracy_score
7 from sklearn.metrics import f1_score
8 from sklearn.preprocessing import StandardScaler

```

```

1 wine=pd.read_csv("wines.csv")
2 wine.head(10)

```



	Wine	Alcohol	Malic.acid	Ash	Ac1	Mg	Phenols	Flavanoids	Nonflavanoid.phenols	Proanth	Color.int
0	1	14.23	1.71	2.43	15.6	127	2.80	3.06	0.28	2.29	15.1
1	1	13.20	1.78	2.14	11.2	100	2.65	2.76	0.26	1.28	12.8
2	1	13.16	2.36	2.67	18.6	101	2.80	3.24	0.30	2.81	15.3
3	1	14.37	1.95	2.50	16.8	113	3.85	3.49	0.24	2.18	14.3
4	1	13.24	2.59	2.87	21.0	118	2.80	2.69	0.39	1.82	13.2
5	1	14.20	1.76	2.45	15.2	112	3.27	3.39	0.34	1.97	14.5
6	1	14.39	1.87	2.45	14.6	96	2.50	2.52	0.30	1.98	14.9
7	1	14.06	2.15	2.61	17.6	121	2.60	2.51	0.31	1.25	13.5
8	1	14.83	1.64	2.17	14.0	97	2.80	2.98	0.29	1.98	14.8
9	1	13.86	1.35	2.27	16.0	98	2.98	3.15	0.22	1.85	13.8

```

1 wine.isnull().sum()

```

```

Wine          0
Alcohol       0
Malic.acid    0
Ash           0
Ac1           0
Mg            0
Phenols       0
Flavanoids    0
Nonflavanoid.phenols  0
Proanth       0
Color.int     0
Hue           0
OD            0
Proline       0
dtype: int64

```

```

1 wine.Wine.value_counts()

```

```
1      59
3      48
Name: Wine, dtype: int64
```

```
1 wx=wine.iloc[:,1:13]
2 wy=wine.iloc[:,0]
3 wx.shape,wy.shape
```

```
((178, 12), (178,))
```

```
1 scaler = StandardScaler()
2 wx=scaler.fit_transform(wx)
```

```
1 from sklearn.model_selection import train_test_split
2 x_train,x_test,y_train,y_test = train_test_split(wx,wy,test_size=0.2,random_st
```

```
1 from sklearn.decomposition import PCA
2 pca=PCA(n_components=2)
3 x_train=pca.fit_transform(x_train)
4 x_test=pca.fit_transform(x_test)
```

```
1 from sklearn.neighbors import KNeighborsClassifier
```

```
1 knn=KNeighborsClassifier(n_neighbors=20)
2 knn.fit(x_train,y_train)
```

```
KNeighborsClassifier(n_neighbors=20)
```

```
1 pred=knn.predict(x_test)
2 confusion_matrix(pred,y_test)
```

```
array([[12,  1,  0],
       [ 3, 11,  0],
       [ 0,  2,  7]], dtype=int64)
```

```
1
```

```
1 print("f1_score",f1_score(pred,y_test,average = 'macro'))
```

```
f1_score 0.8392857142857143
```

```
1
```

```
1
```

```

1 import numpy as np # linear algebra
2 import pandas as pd # data processing, CSV file I/O (e.g. pd.read_csv)
3 import matplotlib.pyplot as plt
4 from sklearn.metrics import confusion_matrix
5 from sklearn.metrics import accuracy_score
6 from sklearn.metrics import f1_score

```

```
1 df=pd.read_csv(r"C:\Users\Lakhan Kumawat\Downloads\wines.csv")
```

```
1 df.head()
```



	Wine	Alcohol	Malic.acid	Ash	Ac1	Mg	Phenols	Flavanoids	Nonflavanoid.phenols
0	1	14.23	1.71	2.43	15.6	127	2.80	3.06	0.28
1	1	13.20	1.78	2.14	11.2	100	2.65	2.76	0.26
2	1	13.16	2.36	2.67	18.6	101	2.80	3.24	0.30
3	1	14.37	1.95	2.50	16.8	113	3.85	3.49	0.24
4	1	13.24	2.59	2.87	21.0	118	2.80	2.69	0.39

```
1 df.isnull().sum()
```

```

Wine                0
Alcohol             0
Malic.acid          0
Ash                 0
Ac1                 0
Mg                  0
Phenols             0
Flavanoids          0
Nonflavanoid.phenols 0
Proanth             0
Color.int           0
Hue                 0
OD                  0
Proline             0
dtype: int64

```

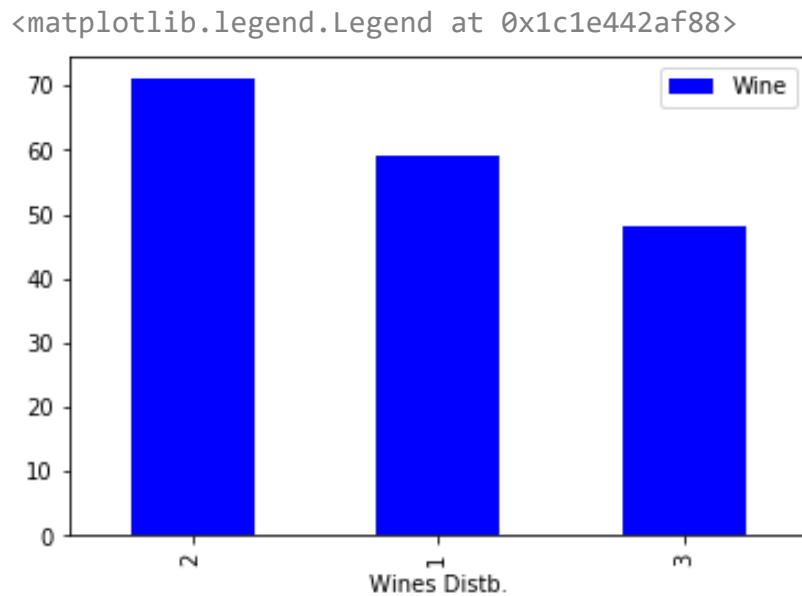
```
1 df["Wine"].value_counts()
```

```

2    71
1    59

```

```
1 df["Wine"].value_counts().plot.bar(color='Blue')
2 plt.xlabel("Wines Distb.")
3 plt.legend()
```



```
1 X = df.iloc[:, 1:13].values
2 y = df.iloc[:, 0].values
```

```
[ [14.23  1.71  2.43 ...  5.64  1.04  3.92]
  [13.2   1.78  2.14 ...  4.38  1.05  3.4 ]
  [13.16  2.36  2.67 ...  5.68  1.03  3.17]
  ...
  [13.27  4.28  2.26 ... 10.2   0.59  1.56]
  [13.17  2.59  2.37 ...  9.3    0.6   1.62]
  [14.13  4.1   2.74 ...  9.2    0.61  1.6  ]]
```

```
1 from sklearn.model_selection import train_test_split
2 X_train, X_test, y_train, y_test = train_test_split(X, y, test_size =
```

```
1 from sklearn.decomposition import PCA
2 pca = PCA(n_components = 3)
3 X_train = pca.fit_transform(X_train)
4 X_test = pca.transform(X_test)
```

```
LogisticRegression(C=1.0, class_weight=None, dual=False, fit_intercept=True,
                    intercept_scaling=1, l1_ratio=None, max_iter=100,
                    multi_class='auto', n_jobs=None, penalty='l2',
                    random_state=10, solver='lbfgs', tol=0.0001, verbose=0,
                    warm_start=False)
```

```
1 y_pred = classifier.predict(X_test)

1 from sklearn.metrics import confusion_matrix, accuracy_score
2 cm = confusion_matrix(y_test, y_pred)
3 print(cm)
4 accuracy_score(y_test,y_pred)
```

```
[[10  0  0]
 [ 1 15  1]
 [ 0  0  9]]
0.9444444444444444
```

```
1 f1_score(y_test, y_pred,average='weighted')
```

```
0.9441007031467559
```

```
1 from sklearn.ensemble import RandomForestClassifier
2 rdf_c=RandomForestClassifier(n_estimators=10,criterion='entropy',random
3 rdf_c.fit(X_train,y_train)
4 rdf_pred=rdf_c.predict(X_test)
5 rdf_cm=confusion_matrix(y_test,rdf_pred)
6 print("The accuracy of RandomForestClassifier is:",accuracy_score(rdf_
```

```
The accuracy of RandomForestClassifier is: 0.9166666666666666
```

```
1 print("f1_score of RandomForestClassifier is:")
2 f1_score(y_test, rdf_pred,average='weighted')
```

```
f1_score of RandomForestClassifier is:
0.9174558080808082
```

```
1
```