Pattern & Anomaly Detection Lab

Experiment 13

Random Forest

Submitted By:

Dhruv Singhal

500075346

R177219074

AIML B3

Submitted To:

Dr. Gopal Phartiyal

Asst. Professor

SOCS

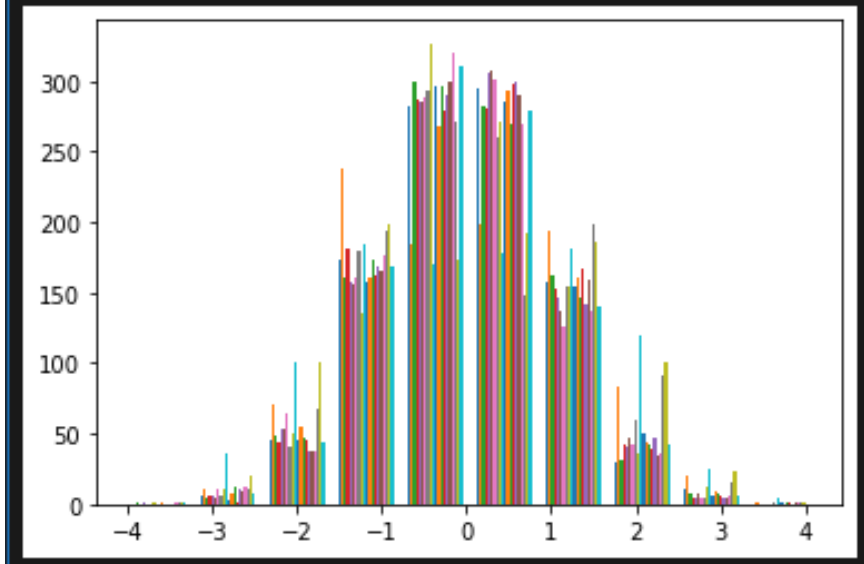
UPES

CODE:

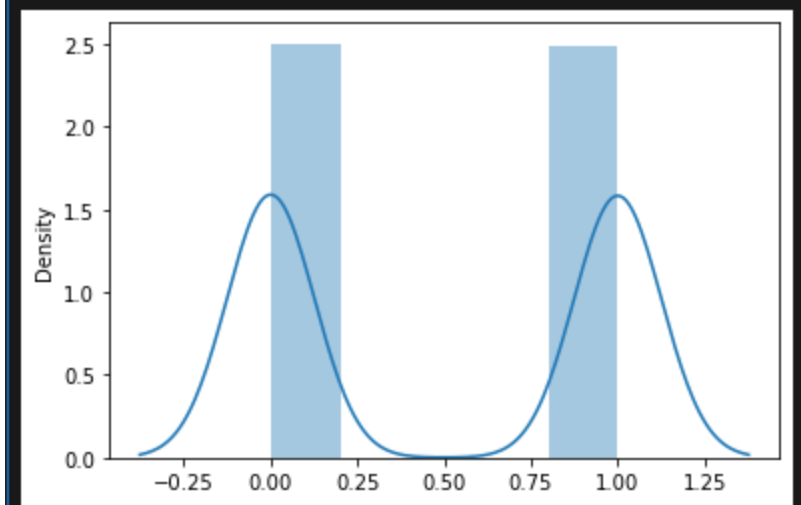
```python
from sklearn.datasets import make_classification
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestClassifier
import seaborn as sns
import pandas as pd
from sklearn.metrics import confusion_matrix
import matplotlib.pyplot as plt
#%%

#generate dataset
X,y = make_classification(n_samples=1000,n_classes=2)
X=pd.DataFrame(X)
y=pd.Series(y)
print("X values are:",X.head())
print("Y values are:",y.head())

#%% Visualization

plt.hist(X)
plt.show()

#%%
sns.distplot(y)
plt.show()

#%%
sns.distplot(X)
plt.show()

#%% feature extractionpreprocessing, correlation matrix
_,graph=plt.subplots(figsize=(15,10))
sns.heatmap(X.corr(),annot=True,ax=graph,square=True)
plt.show()


#%%
X_train,X_test,Y_train,Y_test=train_test_split(X,y,test_size=0.15,random_state=42)

model=RandomForestClassifier()
model.fit(X_train,Y_train)
print(model.classes_)

Y_pred=model.predict(X_test)
print(Y_pred)

print("train Accuracy:",model.score(X_train,Y_train))
print("test Accuracy:",model.score(X_test,Y_test))

print(confusion_matrix(Y_test,Y_pred))
```

OUTPUT:

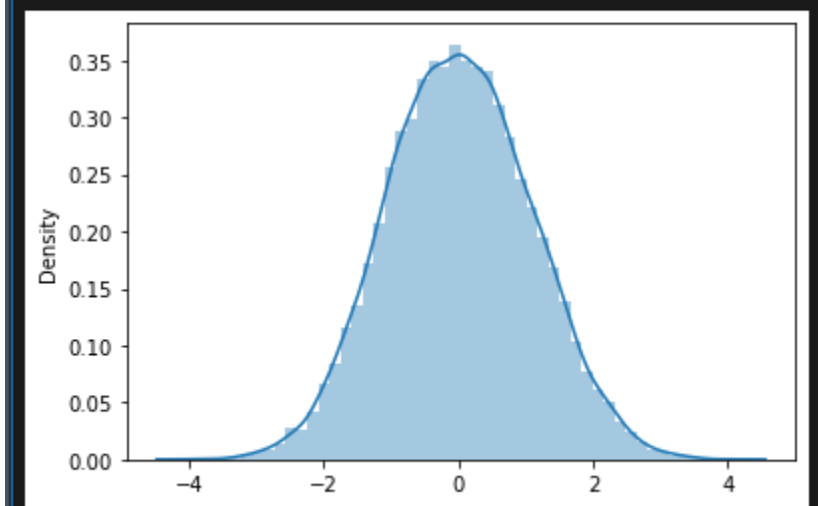
```
In [2]: runcell(2, 'B:/3rd year/5th sem/P&AD/random_normal.py')
X values are:          0         1         2   ...        17        18        19
0  1.370177  0.062732  0.236024  ...  0.109538 -0.201451  2.372252
1  1.454009  1.366763 -0.714880  ...  1.536780 -0.850639  0.993668
2  0.261344 -1.142819 -0.622376  ... -0.733287 -1.585985 -0.511575
3  0.215769 -2.041450 -2.239620  ... -2.029578  0.163686  1.164900
4 -0.654511  0.740525  0.538260  ...  0.251433  1.959272  1.413851

[5 rows x 20 columns]
Y values are: 0    1
1    1
2    0
3    0
4    1
dtype: int32

In [3]: runcell('Visualization', 'B:/3rd year/5th sem/P&AD/random_normal.py')
```

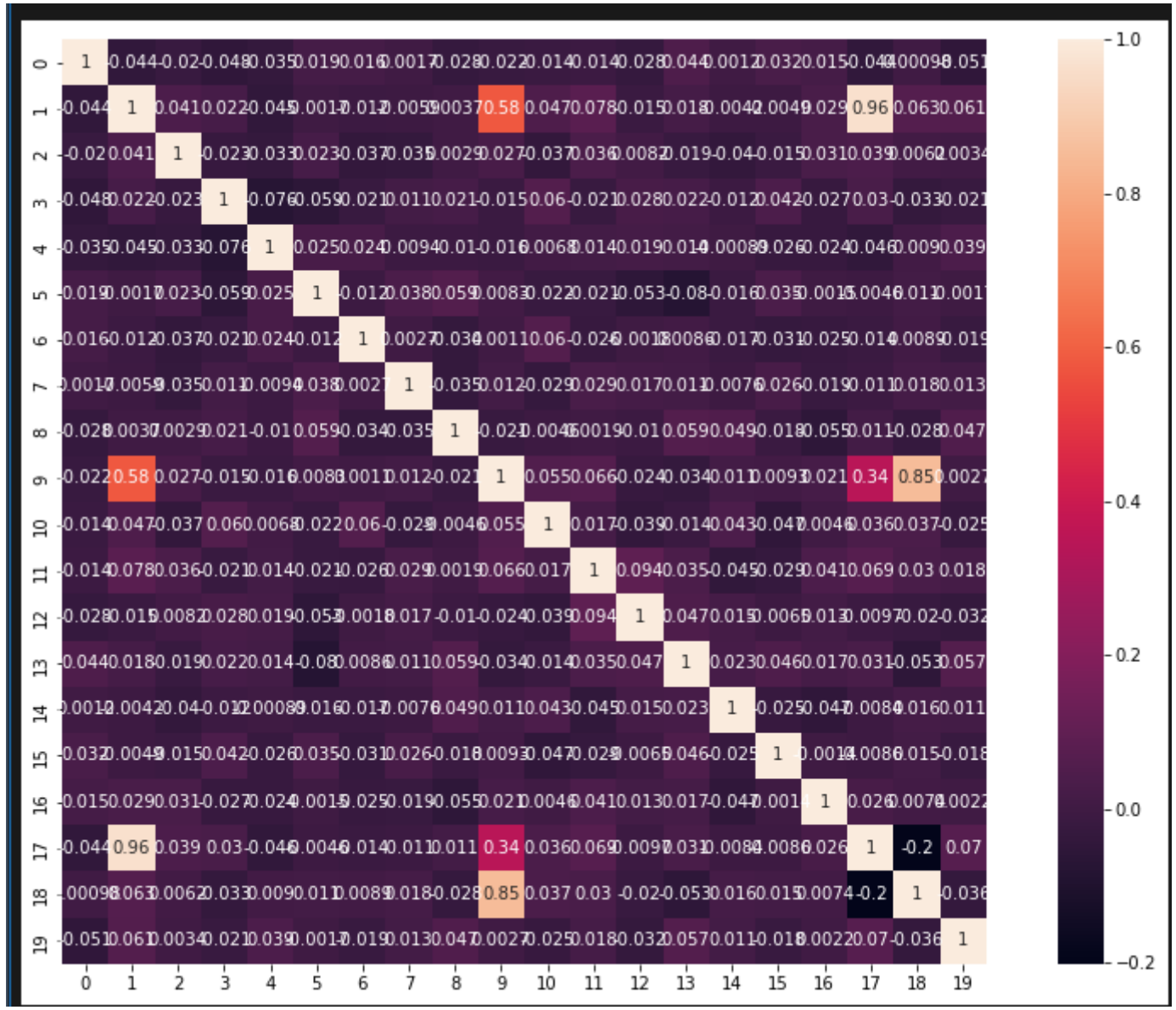


```
In [4]: runcell(4, 'B:/3rd year/5th sem/P&AD/random_normal.py')
C:\Users\Dhruv Singhal\anaconda3\lib\site-packages\seaborn\distributions.py:2619: FutureWarning:
`distplot` is a deprecated function and will be removed in a future version. Please adapt your code to
use either `displot` (a figure-level function with similar flexibility) or `histplot` (an axes-level
function for histograms).
  warnings.warn(msg, FutureWarning)
```



```
In [5]: runcell(5, 'B:/3rd year/5th sem/P&AD/random_normal.py')
C:\Users\Dhruv Singhal\anaconda3\lib\site-packages\seaborn\distributions.py:2619: FutureWarning:
`distplot` is a deprecated function and will be removed in a future version. Please adapt your code to
use either `displot` (a figure-level function with similar flexibility) or `histplot` (an axes-level
function for histograms).
  warnings.warn(msg, FutureWarning)
```




0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19
1.0 0.8 0.6 0.4 0.2 0.0 −0.2

```
In [7]: runcell(7, 'B:/3rd year/5th sem/P&AD/random_normal.py')
[0 1]
[0 0 0 1 0 0 0 0 1 1 0 0 0 1 1 0 0 1 0 0 0 0 0 1 0 0 0 0 1 1 0 0 0 0 1 0 0
 1 1 0 0 0 0 1 0 1 1 1 1 0 1 0 1 0 1 1 0 1 1 0 0 1 1 0 1 0 0 0 0 0 0 0 0 0
 1 1 1 0 1 0 0 1 1 1 1 1 1 0 0 0 0 1 1 0 1 0 1 0 0 0 1 0 1 0 1 1 1 1 1 0 0
 1 1 0 1 1 1 1 0 0 0 0 1 1 1 0 1 1 0 1 0 1 0 1 0 0 1 1 1 0 1 0 0 1 0 1 1 1
 0 1]
train Accuracy: 1.0
test Accuracy: 0.9133333333333333
[[75  8]
 [ 5 62]]
```

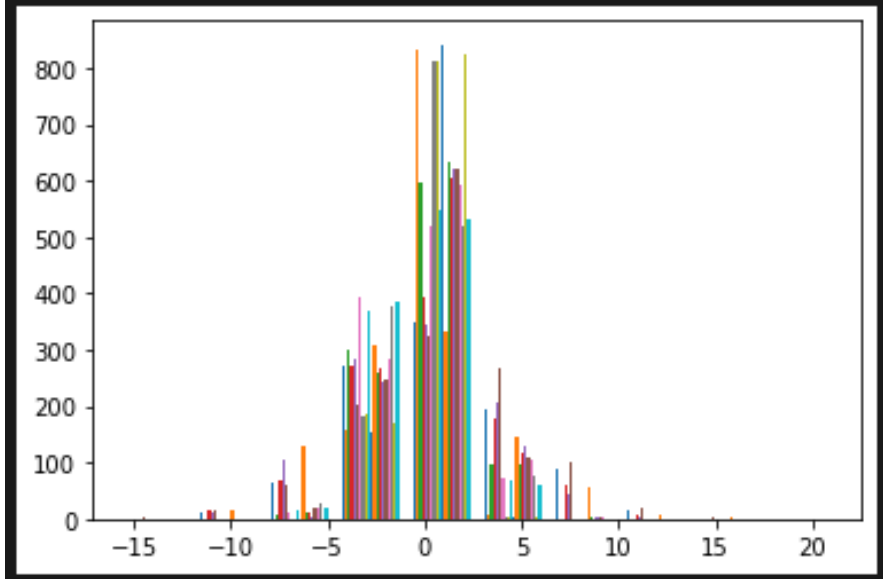
CODE: Multiple Class

```
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestClassifier
import seaborn as sns
import pandas as pd
import matplotlib.pyplot as plt
import warnings
warnings.filterwarnings('ignore')
#%%

#generate dataset
X,y = make_classification(n_samples=1000,n_classes=5,n_features=20,
n_informative=10,n_redundant=5,shuffle=True,random_state=42)
X=pd.DataFrame(X)
y=pd.DataFrame(y)
print("X values are:",X.head())
print("Y values are:",y.head())

#%% Visualization

plt.hist(X)
plt.show()

#%%
sns.distplot(y)
plt.show()

#%%
sns.distplot(X)
plt.show()

#%% feature extractionpreprocessing, correlation matrix
_,graph=plt.subplots(figsize=(15,10))
sns.heatmap(X.corr(),annot=True,ax=graph,square=True)
plt.show()
#%%
X_train,X_test,Y_train,Y_test=train_test_split(X,y,test_size=0.15,random_state=42)

model=RandomForestClassifier()
model.fit(X_train,Y_train)
print(model.classes_)

Y_pred=model.predict(X_test)
print(Y_pred)
```

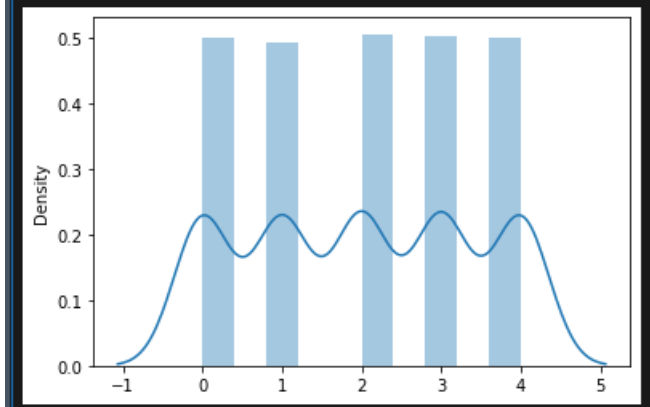
```
45
46  print("train Accuracy:",model.score(X_train,Y_train))
47  print("test Accuracy:",model.score(X_test,Y_test))
48  #%%
49  from sklearn.model_selection import KFold
50  from sklearn.model_selection import cross_val_score
51  clf = RandomForestClassifier(max_depth=5, random_state=0)
52  kf=KFold(n_splits=7)
53  score=cross_val_score(clf, X, y, cv=kf)
54  print("Cross Validation Scores are {}".format(score))
55  print("Average Cross Validation score :{}".format(score.mean()))
56
57
58  #%%
59  from sklearn.model_selection import GridSearchCV
60  tuned_parameters = [{'n_estimators':[10,20,40,100],'criterion':['gini', 'entropy'],
61                       'max_features':['auto', 'sqrt', 'log2'],'bootstrap':[True,False]}]
62  clf=GridSearchCV(RandomForestClassifier(),tuned_parameters,scoring=('accuracy'),verbose=3)
63  clf.fit(X,y)
64  print("Best parameters set found on development set:")
65  print()
66  print(clf.best_params_)
67  print()
68  print("Best Score:",clf.best_score_)
69  z=clf.cv_results_
70
```

OUTPUT:

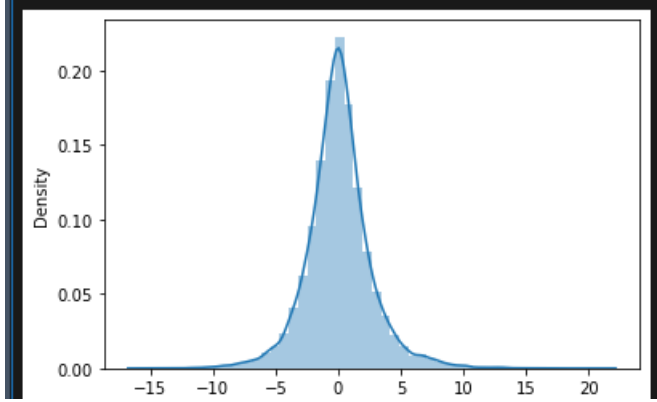
```
In [1]: runcell(0, 'B:/3rd year/5th sem/P&AD/exp13_multi.py')

In [2]: runcell(1, 'B:/3rd year/5th sem/P&AD/exp13_multi.py')
X values are:          0         1         2   ...        17        18        19
0  8.777957 -1.390182  2.866176  ... -2.372538  0.321516 -3.421634
1 -0.296949 -1.350663 -2.596690  ...  0.577353  0.005837 -1.315612
2  3.198753  0.020124  2.434817  ... -1.054403  1.699146 -3.275074
3  4.725085 -0.114722 -0.956705  ...  0.660762 -2.022246 -3.007724
4 -3.280166 -0.763541  0.201406  ... -3.444687 -0.181709 -0.402945

[5 rows x 20 columns]
Y values are:    0
0  2
1  3
2  2
3  0
4  3

In [3]: runcell('Visualization', 'B:/3rd year/5th sem/P&AD/exp13_multi.py')
```

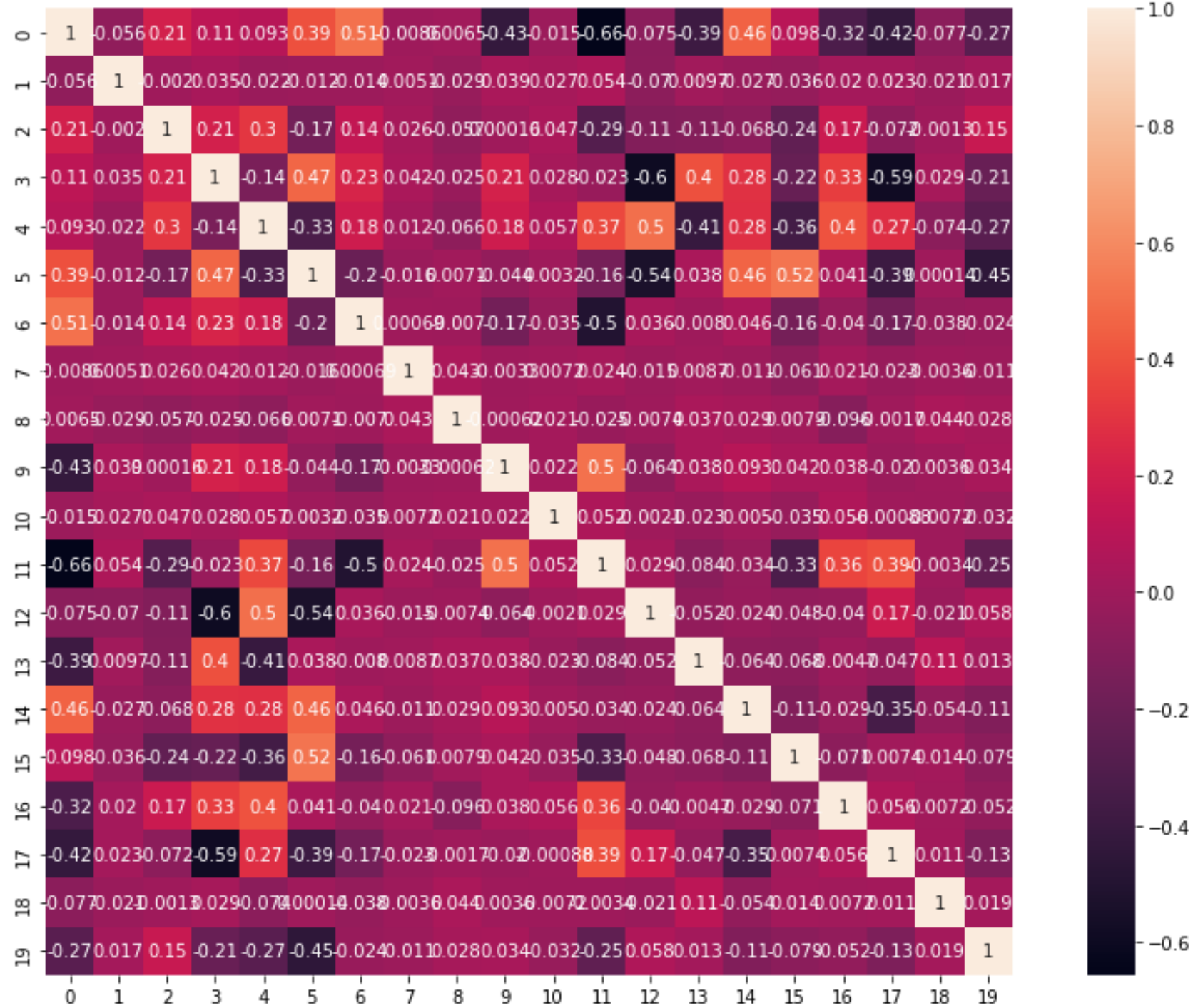


```
In [4]: runcell(3, 'B:/3rd year/5th sem/P&AD/exp13_multi.py')
```




```
In [5]: runcell(4, 'B:/3rd year/5th sem/P&AD/exp13_multi.py')
```




0 1 -0.056 0.21 0.11 0.093 0.39 0.51 -0.008 0.0065 -0.43 -0.015 -0.66 -0.075 -0.39 0.46 0.098 -0.32 -0.42 -0.077 -0.27
1 -0.056 1 0.002 0.035 -0.022 0.012 0.014 -0.0051 0.029 0.039 0.027 0.054 -0.07 0.0097 0.027 0.036 0.02 0.023 -0.021 0.017
2 0.21 -0.002 1 0.21 0.3 -0.17 0.14 0.026 -0.057 0.00016 0.047 -0.29 -0.11 -0.11 -0.068 -0.24 0.17 -0.074 0.0013 0.15
3 0.11 0.035 0.21 1 -0.14 0.47 0.23 0.042 -0.025 0.21 0.028 0.023 -0.6 0.4 0.28 -0.22 0.33 -0.59 0.029 -0.21
4 0.093 -0.022 0.3 -0.14 1 -0.33 0.18 0.012 -0.066 0.18 0.057 0.37 0.5 -0.41 0.28 -0.36 0.4 0.27 -0.074 -0.27
5 0.39 -0.012 -0.17 0.47 -0.33 1 -0.2 -0.016 0.0071 0.044 0.0032 0.16 -0.54 0.038 0.46 0.52 0.041 -0.39 0.00014 0.45
6 0.51 -0.014 0.14 0.23 0.18 -0.2 1 0.00069 -0.007 -0.17 -0.035 -0.5 0.036 0.008 0.046 -0.16 -0.04 -0.17 -0.038 0.024
7 -0.0086 0.0051 0.026 0.042 0.012 -0.016 0.00069 1 0.043 0.0033 0.0072 0.024 0.015 0.0087 0.011 0.061 0.021 -0.023 0.0036 0.011
8 0.0065 0.029 0.057 -0.025 0.066 0.0071 0.007 0.043 1 0.00062 0.021 0.025 0.0074 0.037 0.029 0.0079 0.096 0.0017 0.044 0.028
9 -0.43 0.039 0.00016 0.21 0.18 -0.044 -0.17 0.0033 0.00062 1 0.022 0.5 -0.064 0.038 0.093 0.042 0.038 -0.02 0.0036 0.034
10 -0.015 0.027 0.047 0.028 0.057 0.0032 0.035 0.0072 0.021 0.022 1 0.052 0.0024 0.023 0.005 0.035 0.056 0.00088 0.0072 0.032
11 -0.66 0.054 -0.29 -0.023 0.37 -0.16 -0.5 0.024 -0.025 0.5 0.052 1 0.029 0.084 0.034 -0.33 0.36 0.39 0.0034 0.25
12 -0.075 -0.07 -0.11 -0.6 0.5 -0.54 0.036 0.015 0.0074 0.064 0.0021 0.029 1 0.052 0.024 0.048 -0.04 0.17 -0.021 0.058
13 -0.39 0.0097 -0.11 0.4 -0.41 0.038 0.008 0.0087 0.037 0.038 0.023 0.084 0.052 1 0.064 0.068 0.0047 0.047 0.11 0.013
14 0.46 -0.027 0.068 0.28 0.28 0.46 0.046 0.011 0.029 0.093 0.005 0.034 0.024 0.064 1 0.11 0.029 0.35 0.054 0.11
15 0.098 0.036 0.24 0.22 0.36 0.52 0.16 0.061 0.0079 0.042 0.035 0.33 0.048 0.068 0.11 1 0.071 0.0074 0.014 0.079
16 0.32 0.02 0.17 0.33 0.4 0.041 0.04 0.021 0.096 0.038 0.056 0.36 0.04 0.0047 0.029 0.071 1 0.056 0.0072 0.052
17 0.42 0.023 0.072 0.59 0.27 0.39 0.17 0.023 0.0017 0.02 0.00088 0.39 0.17 0.047 0.35 0.0074 0.056 1 0.011 0.13
18 0.077 0.021 0.0013 0.029 0.074 0.00014 0.038 0.0036 0.044 0.0036 0.0072 0.0034 0.021 0.11 0.054 0.014 0.0072 0.011 1 0.019
19 0.27 0.017 0.15 0.21 0.27 0.45 0.024 0.011 0.028 0.034 0.032 0.25 0.058 0.013 0.11 0.079 0.052 0.13 0.019 1
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19
1.0 0.8 0.6 0.4 0.2 0.0 −0.2 −0.4 −0.6

```
In [7]: runcell(6, 'B:/3rd year/5th sem/P&AD/exp13_multi.py')
[0 1 2 3 4]
[2 1 2 3 2 2 0 2 0 4 0 1 2 1 1 0 2 4 3 4 2 2 0 1 0 0 3 4 0 0 2 3 3 0 4 1 2
 0 0 1 4 0 1 0 1 1 4 4 4 2 3 3 2 3 1 2 3 3 1 1 1 2 3 0 0 4 0 4 1 3 0 3 1 3
 2 1 0 1 4 0 0 1 1 2 3 2 1 2 0 3 2 3 1 3 4 3 2 4 3 4 1 3 1 4 2 3 0 4 0 0 2
 3 2 2 4 1 1 2 1 3 4 4 2 4 2 1 1 2 2 2 0 2 2 1 3 2 3 0 0 3 0 0 2 4 4 2 4 4
 1 3]
train Accuracy: 1.0
test Accuracy: 0.7333333333333333

In [8]: runcell(7, 'B:/3rd year/5th sem/P&AD/exp13_multi.py')
Cross Validation Scores are [0.6013986  0.62237762 0.62937063 0.64335664 0.6993007  0.6013986
 0.62676056]
Average Cross Validation score :0.6319947657975826
```

```
Best parameters set found on development set:

{'bootstrap': False, 'criterion': 'gini', 'max_features': 'auto', 'n_estimators': 100}

Best Score: 0.723
```