MachineLearning

December 14, 2023

[82]:
```
!pip3 install import_ipynb
```

```
Requirement already satisfied: import_ipynb in
/Users/debanilguha/opt/anaconda3/lib/python3.8/site-packages (0.1.4)
Requirement already satisfied: nbformat in
/Users/debanilguha/opt/anaconda3/lib/python3.8/site-packages (from import_ipynb)
(5.7.0)
Requirement already satisfied: IPython in
/Users/debanilguha/opt/anaconda3/lib/python3.8/site-packages (from import_ipynb)
(8.5.0)
Requirement already satisfied: pickleshare in
/Users/debanilguha/opt/anaconda3/lib/python3.8/site-packages (from
IPython->import_ipynb) (0.7.5)
Requirement already satisfied: prompt-toolkit<3.1.0,>3.0.1 in
/Users/debanilguha/opt/anaconda3/lib/python3.8/site-packages (from
IPython->import_ipynb) (3.0.36)
Requirement already satisfied: pexpect>4.3 in
/Users/debanilguha/opt/anaconda3/lib/python3.8/site-packages (from
IPython->import_ipynb) (4.8.0)
Requirement already satisfied: decorator in
/Users/debanilguha/opt/anaconda3/lib/python3.8/site-packages (from
IPython->import_ipynb) (5.1.1)
Requirement already satisfied: stack-data in
/Users/debanilguha/opt/anaconda3/lib/python3.8/site-packages (from
IPython->import_ipynb) (0.5.1)
Requirement already satisfied: appnope in
/Users/debanilguha/opt/anaconda3/lib/python3.8/site-packages (from
IPython->import_ipynb) (0.1.2)
Requirement already satisfied: traitlets>=5 in
/Users/debanilguha/opt/anaconda3/lib/python3.8/site-packages (from
IPython->import_ipynb) (5.7.1)
Requirement already satisfied: backcall in
/Users/debanilguha/opt/anaconda3/lib/python3.8/site-packages (from
IPython->import_ipynb) (0.2.0)
Requirement already satisfied: pygments>=2.4.0 in
/Users/debanilguha/opt/anaconda3/lib/python3.8/site-packages (from
IPython->import_ipynb) (2.11.2)
Requirement already satisfied: matplotlib-inline in
```

```
/Users/debanilguha/opt/anaconda3/lib/python3.8/site-packages (from
IPython->import_ipynb) (0.1.6)
Requirement already satisfied: jedi>=0.16 in
/Users/debanilguha/opt/anaconda3/lib/python3.8/site-packages (from
IPython->import_ipynb) (0.18.1)
Requirement already satisfied: jupyter-core in
/Users/debanilguha/opt/anaconda3/lib/python3.8/site-packages (from
nbformat->import_ipynb) (5.1.1)
Requirement already satisfied: jsonschema>=2.6 in
/Users/debanilguha/opt/anaconda3/lib/python3.8/site-packages (from
nbformat->import_ipynb) (4.16.0)
Requirement already satisfied: fastjsonschema in
/Users/debanilguha/opt/anaconda3/lib/python3.8/site-packages (from
nbformat->import_ipynb) (2.16.2)
Requirement already satisfied: parso<0.9.0,>=0.8.0 in
/Users/debanilguha/opt/anaconda3/lib/python3.8/site-packages (from
jedi>=0.16->IPython->import_ipynb) (0.8.3)
Requirement already satisfied: pkgutil-resolve-name>=1.3.10 in
/Users/debanilguha/opt/anaconda3/lib/python3.8/site-packages (from
jsonschema>=2.6->nbformat->import_ipynb) (1.3.10)
Requirement already satisfied: attrs>=17.4.0 in
/Users/debanilguha/opt/anaconda3/lib/python3.8/site-packages (from
jsonschema>=2.6->nbformat->import_ipynb) (22.1.0)
Requirement already satisfied: pyrsistent!=0.17.0,!=0.17.1,!=0.17.2,>=0.14.0 in
/Users/debanilguha/opt/anaconda3/lib/python3.8/site-packages (from
jsonschema>=2.6->nbformat->import_ipynb) (0.18.0)
Requirement already satisfied: importlib-resources>=1.4.0 in
/Users/debanilguha/opt/anaconda3/lib/python3.8/site-packages (from
jsonschema>=2.6->nbformat->import_ipynb) (5.2.0)
Requirement already satisfied: ptyprocess>=0.5 in
/Users/debanilguha/opt/anaconda3/lib/python3.8/site-packages (from
pexpect>4.3->IPython->import_ipynb) (0.7.0)
Requirement already satisfied: wcwidth in
/Users/debanilguha/opt/anaconda3/lib/python3.8/site-packages (from prompt-
toolkit<3.1.0,>3.0.1->IPython->import_ipynb) (0.2.5)
Requirement already satisfied: platformdirs>=2.5 in
/Users/debanilguha/opt/anaconda3/lib/python3.8/site-packages (from jupyter-
core->nbformat->import_ipynb) (2.5.2)
Requirement already satisfied: asttokens in
/Users/debanilguha/opt/anaconda3/lib/python3.8/site-packages (from stack-
data->IPython->import_ipynb) (2.0.8)
Requirement already satisfied: executing in
/Users/debanilguha/opt/anaconda3/lib/python3.8/site-packages (from stack-
data->IPython->import_ipynb) (1.1.1)
Requirement already satisfied: pure-eval in
/Users/debanilguha/opt/anaconda3/lib/python3.8/site-packages (from stack-
data->IPython->import_ipynb) (0.2.2)
Requirement already satisfied: zipp>=3.1.0 in
```

```
/Users/debanilguha/opt/anaconda3/lib/python3.8/site-packages (from importlib-
resources>=1.4.0->jsonschema>=2.6->nbformat->import_ipynb) (3.11.0)
Requirement already satisfied: six in
/Users/debanilguha/opt/anaconda3/lib/python3.8/site-packages (from
asttokens->stack-data->IPython->import_ipynb) (1.16.0)
```

[83]:
```
import import_ipynb
import numpy as np
import pandas as pd
import dataScience as ds
```

[84]:
```
def normalize(V,maxV,minV):
    if V==0:
        return V
    normalized_V = (V - minV) / (maxV - minV)

    return normalized_V
```

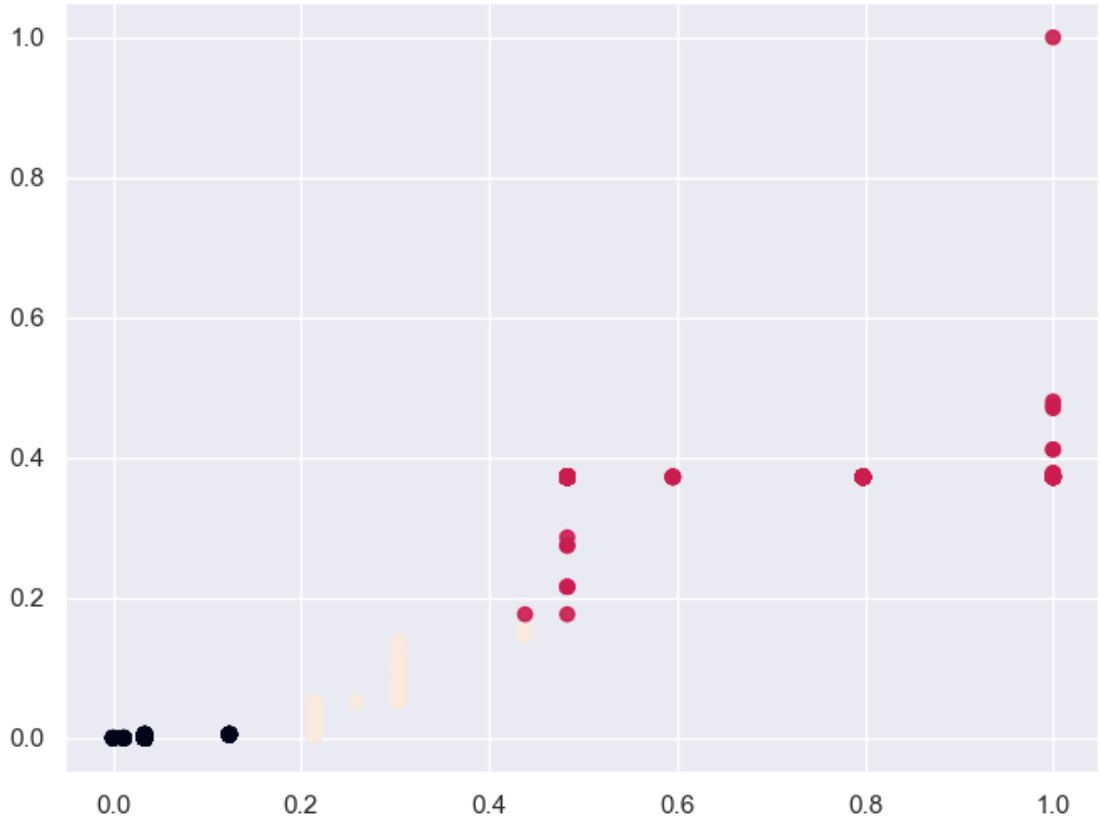
[85]:
```
def createClusteredData(X:list,Y:list):
    newArr = []
    for index,value in enumerate(X):
        newArr.append([normalize(X[index],np.max(X),np.
 ↪min(X)),normalize(Y[index],np.max(Y),np.min(Y))])
    return newArr
```

0.1 Using K Means Clustering Algorithm

[101]:
```
from sklearn.cluster import KMeans
import matplotlib.pyplot as plt
from sklearn.preprocessing import scale
dataForClustering = np.array(createClusteredData(sorted(ds.expectedPayment),␣
 ↪sorted(ds.paymentDoneByChampion)))
print(f"==>> dataForClustering.shape: ",dataForClustering[:
 ↪,0],dataForClustering)


model = KMeans(n_clusters=3)

model = model.fit(dataForClustering)

plt.figure(figsize=(8, 6))

plt.scatter(dataForClustering[:,0], dataForClustering[:,1], c=model.labels_.
 ↪astype(float),alpha=0.9)
plt.show()
```

```
==>> dataForClustering.shape:  [0. 0. 0. … 1. 1. 1.] [[0.         0.        ]
 [0.         0.        ]
```

```
[0.          0.        ]
…
[1.          0.4745098 ]
[1.          0.48039216]
[1.          1.        ]]
```



0.2 Real World Clustering on Active Project

Here this is an example of Clustering of Data in a real life scenario where Datas are clustered Together to give the Kmeans value . This is the scattered plot for Payment Done by the Champion verses Payment to be done on a daily basis