

ML Offline - 1

1605063

Prerequisite

- OS: Ubuntu/Debian (Linux)
- Python-3
- pip

Create Virtual Environment

```
python3 -m pip install --user virtualenv
```

```
python3 -m venv ./venv
```

```
source venv/bin/activate
```

Install Libraries

```
pip install -r requirements.txt
```

Required Libraries

```
cycler==0.11.0  
fonttools==4.28.3  
joblib==1.1.0  
kiwisolver==1.3.2  
matplotlib==3.5.0  
numpy==1.21.4  
packaging==21.3  
pandas==1.3.4  
Pillow==8.4.0  
pyparsing==3.0.6  
python-dateutil==2.8.2  
pytz==2021.3  
scikit-learn==1.0.1  
scipy==1.7.3  
setuptools-scm==6.3.2  
six==1.16.0  
sklearn==0.0  
threadpoolctl==3.0.0  
tomli==1.2.2  
tqdm==4.62.3
```

Run Python File

```
python 1605063.py
```

Selecting Dataset & Round No

- Pass file paths accordingly on lines 485 (Dataset 1), 486 (Dataset 2) and 487 (Dataset 3)
- We have 3 datasets in “dataset_1”, “dataset_2” and “dataset_3” variables, to use the desired dataset change values at lines 489 & 490 accordingly.
- No of rounds can be updated at line 526 by changing the “n_round” parameter

Results for [Dataset-1](#) (Telco Customer Churn)

Performance Measure	Training	Test
Accuracy	0.7965921192758253	0.7955997161107168
True Positive Rate	0.47558528428093644	0.47558528428093644
True Negative Rate	0.9125392606909881	0.9125392606909881
Positive Predictivity Value	0.6626281453867661	0.6626281453867661
False discovery rate	0.33737185461323393	0.33737185461323393
F1 Score	0.5537383177570093	0.5537383177570093

Number of Boosting Rounds	Training	Test
05	0.788604898828541	0.7906316536550745
10	0.7884274050408235	0.7877927608232789
15	0.7887823926162584	0.7892122072391767
20	0.788604898828541	0.78708303761533

Results for [Dataset-2](#) (Adult's Income)

Performance Measure	Training	Test
Accuracy	0.8275851478762938	0.8300472943922363
True Positive Rate	0.45950771585256983	0.45950771585256983
True Negative Rate	0.944336569579288	0.944336569579288
Positive Predictivity Value	0.7236392849969874	0.7236392849969874
False discovery rate	0.2763607150030127	0.2763607150030127
F1 Score	0.5620904836193448	0.5620904836193448

Number of Boosting Rounds	Training	Test
05	0.8316390774239121	0.8304158221239482
10	0.8349559288719633	0.8337939929979731
15	0.8353244679217469	0.8337939929979731
20	0.8355087374466387	0.8338554142865917

Results for [Dataset-3](#) (Credit Card Fraud)

Performance Measure	Training	Test
Accuracy	0.9815615752333258	0.9909008189262967
True Positive Rate	0.7969543147208121	0.7969543147208121
True Negative Rate	0.9997499374843711	0.9997499374843711
Positive Predictivity Value	0.9968253968253968	0.9968253968253968
False discovery rate	0.0031746031746031746	0.0031746031746031746
F1 Score	0.8857545839210155	0.8857545839210155

Number of Boosting Rounds	Training	Test
05	0.9699521966765309	0.9772520473157416
10	0.9710903710448441	0.978161965423112
15	0.9708627361711815	0.978161965423112
20	0.9708627361711815	0.9772520473157416