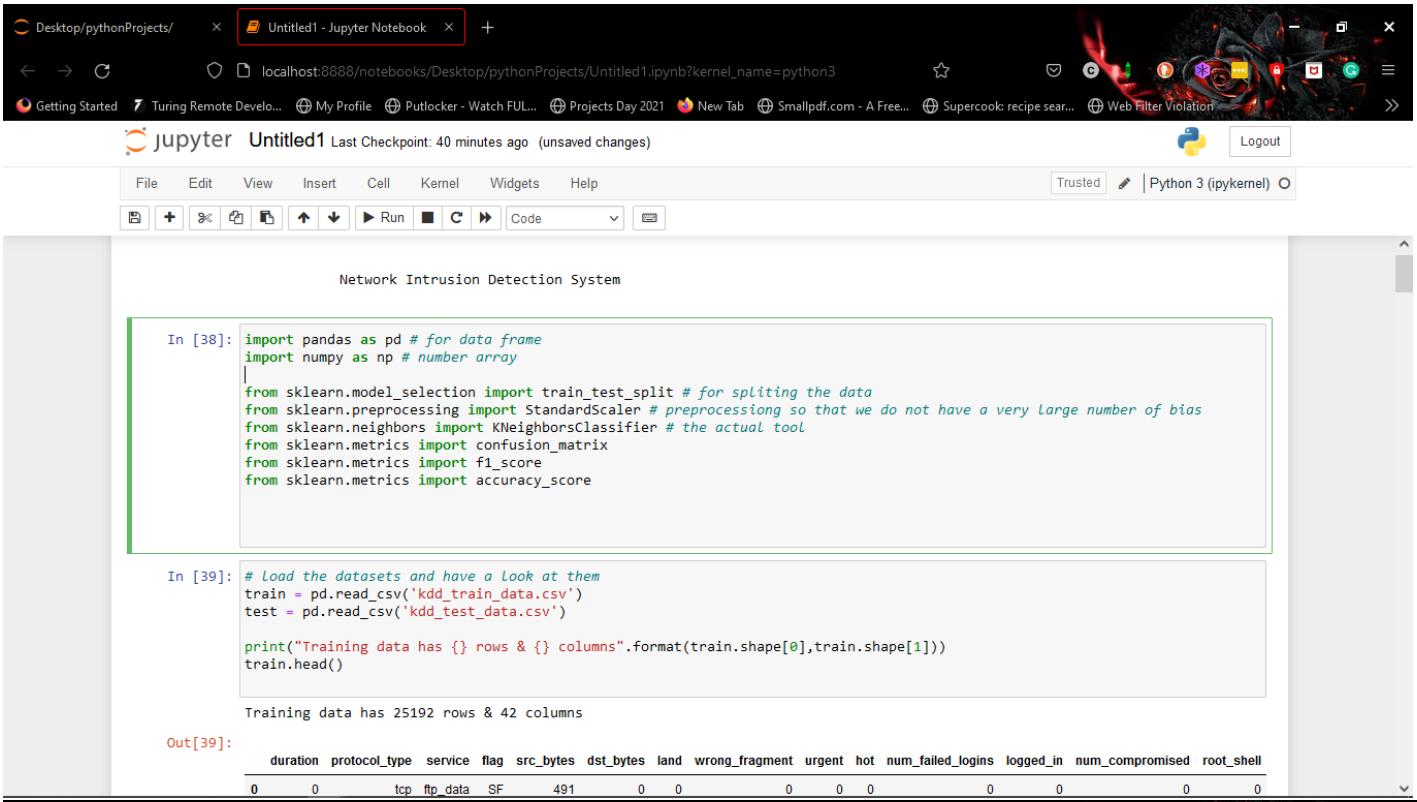
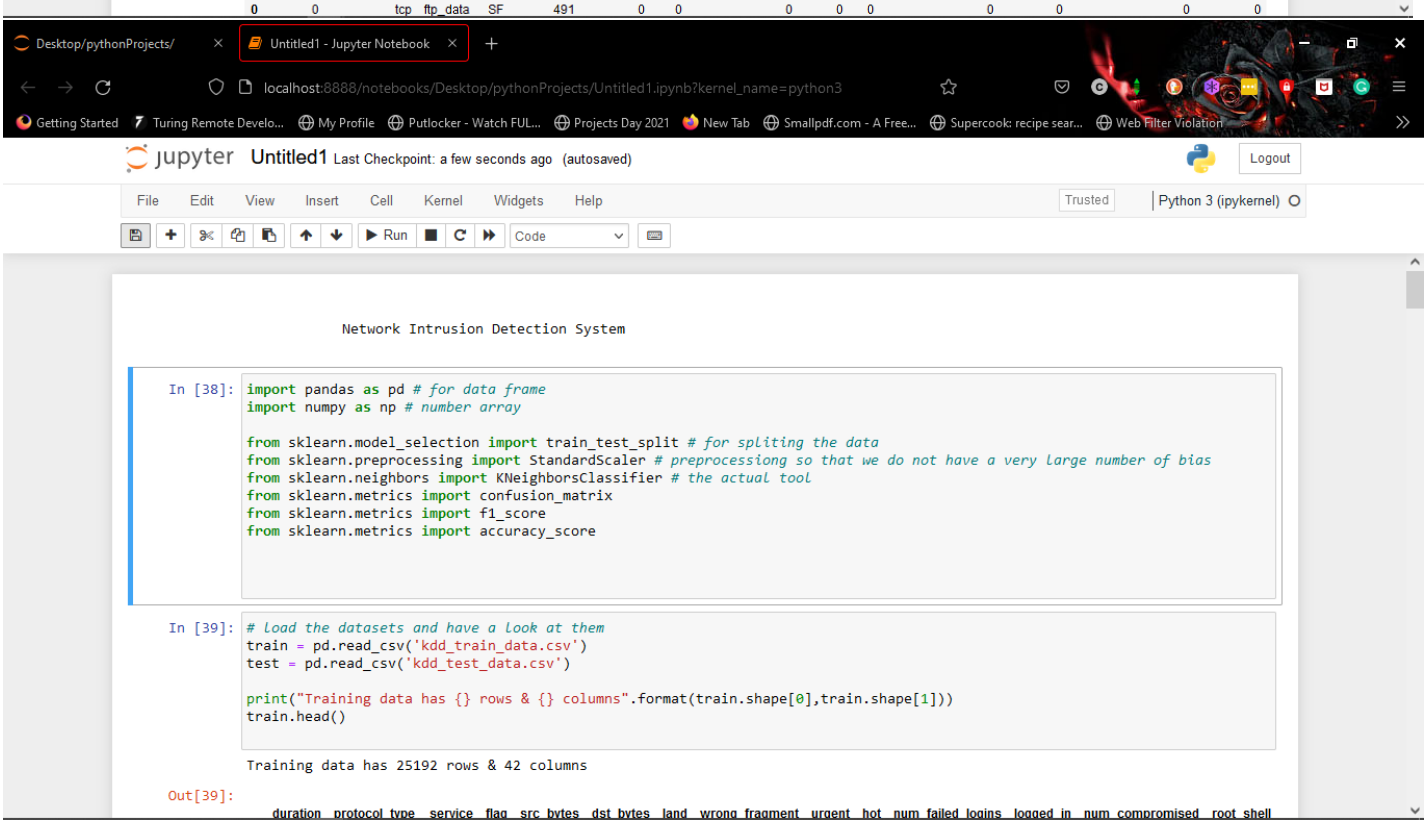
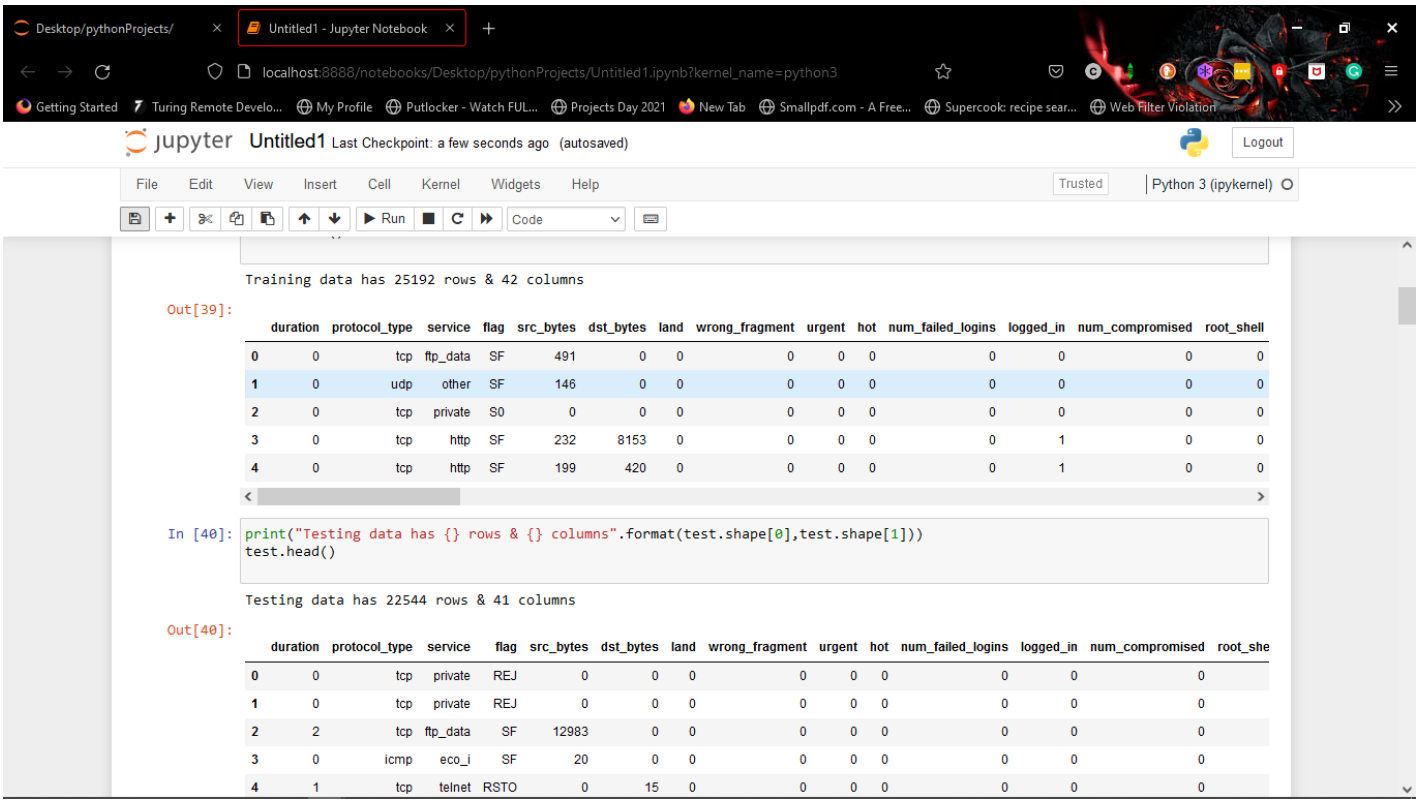

Desktop/pythonProjects/
Untitled1 - Jupyter Notebook
localhost:8888/notebooks/Desktop/pythonProjects/Untitled1.ipynb?kernel_name=python3
Getting Started
Turing Remote Develo...
My Profile
Putlocker - Watch FUL...
Projects Day 2021
New Tab
Smallpdf.com - A Free...
Supercook: recipe sear...
Web Filter Violation
jupyter
Untitled1 Last Checkpoint: 40 minutes ago (unsaved changes)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3 (ipykernel)
Run
Code
Network Intrusion Detection System
In [38]:
import pandas as pd # for data frame
import numpy as np # number array
from sklearn.model_selection import train_test_split # for spliting the data
from sklearn.preprocessing import StandardScaler # preprocessiong so that we do not have a very large number of bias
from sklearn.neighbors import KNeighborsClassifier # the actual tool
from sklearn.metrics import confusion_matrix
from sklearn.metrics import f1_score
from sklearn.metrics import accuracy_score
In [39]:
# load the datasets and have a look at them
train = pd.read_csv('kdd_train_data.csv')
test = pd.read_csv('kdd_test_data.csv')
print("Training data has {} rows & {} columns".format(train.shape[0],train.shape[1]))
train.head()
Training data has 25192 rows & 42 columns
Out[39]:
duration protocol_type service flag src_bytes dst_bytes land wrong_fragment urgent hot num_failed_logins logged_in num_compromised root_shell
0 0 tcp ftp_data SF 491 0 0 0 0 0 0 0 0 0


Desktop/pythonProjects/
Untitled1 - Jupyter Notebook
localhost:8888/notebooks/Desktop/pythonProjects/Untitled1.ipynb?kernel_name=python3
Getting Started
Turing Remote Develo...
My Profile
Putlocker - Watch FUL...
Projects Day 2021
New Tab
Smallpdf.com - A Free...
Supercook: recipe sear...
Web Filter Violation
jupyter
Untitled1 Last Checkpoint: a few seconds ago (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3 (ipykernel)
Run
Code
Network Intrusion Detection System
In [38]:
import pandas as pd # for data frame
import numpy as np # number array
from sklearn.model_selection import train_test_split # for spliting the data
from sklearn.preprocessing import StandardScaler # preprocessiong so that we do not have a very large number of bias
from sklearn.neighbors import KNeighborsClassifier # the actual tool
from sklearn.metrics import confusion_matrix
from sklearn.metrics import f1_score
from sklearn.metrics import accuracy_score
In [39]:
# load the datasets and have a look at them
train = pd.read_csv('kdd_train_data.csv')
test = pd.read_csv('kdd_test_data.csv')
print("Training data has {} rows & {} columns".format(train.shape[0],train.shape[1]))
train.head()
Training data has 25192 rows & 42 columns
Out[39]:
duration protocol_type service flag src_bytes dst_bytes land wrong_fragment urgent hot num_failed_logins logged_in num_compromised root_shell



Training data has 25192 rows & 42 columns

Out[39]:

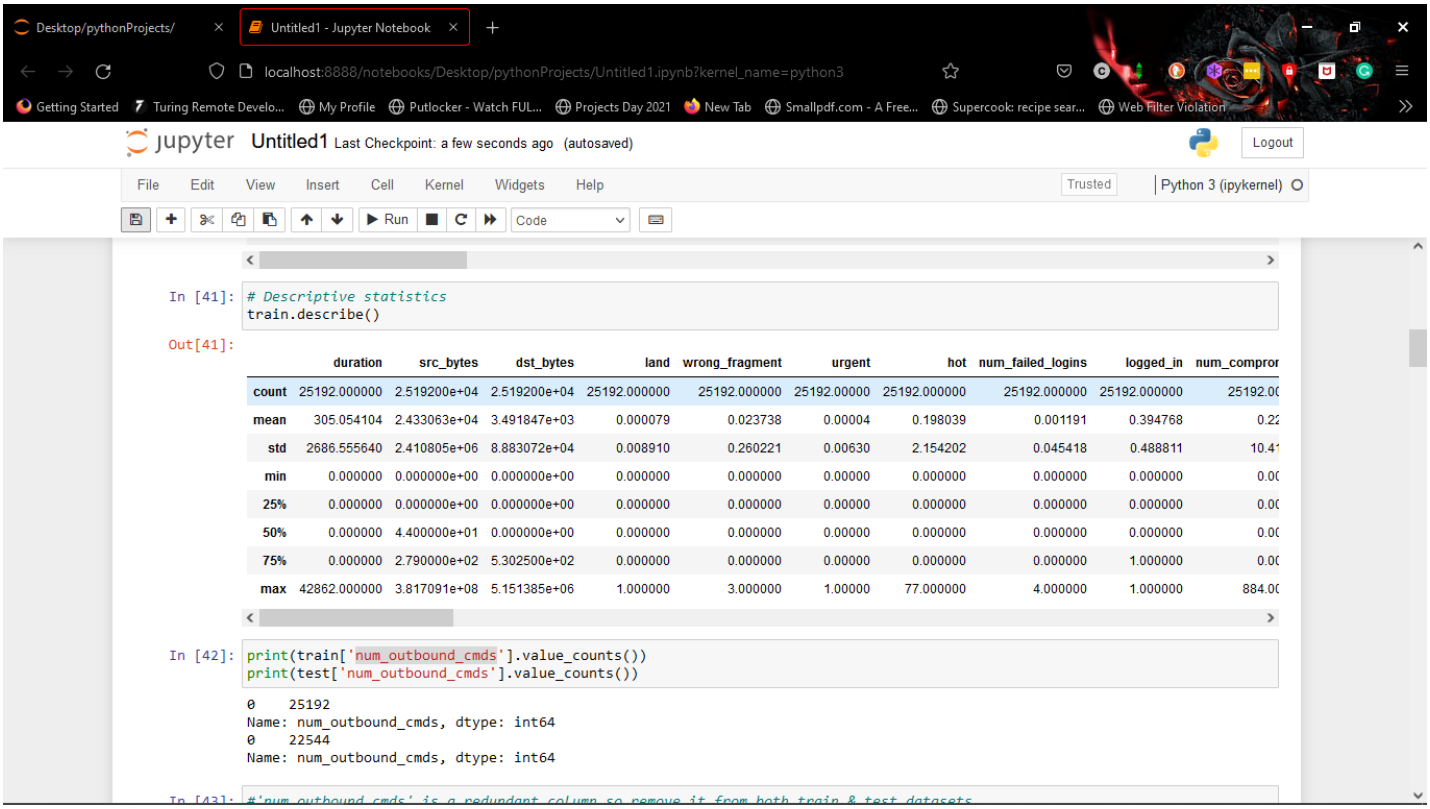
	duration	protocol_type	service	flag	src_bytes	dst_bytes	land	wrong_fragment	urgent	hot	num_failed_logins	logged_in	num_compromised	root_shell
0	0	tcp	ftp_data	SF	491	0	0	0	0	0	0	0	0	0
1	0	udp	other	SF	146	0	0	0	0	0	0	0	0	0
2	0	tcp	private	S0	0	0	0	0	0	0	0	0	0	0
3	0	tcp	http	SF	232	8153	0	0	0	0	0	1	0	0
4	0	tcp	http	SF	199	420	0	0	0	0	0	1	0	0

In [40]:
```
print("Testing data has {} rows & {} columns".format(test.shape[0],test.shape[1]))
test.head()
```

Testing data has 22544 rows & 41 columns

Out[40]:

	duration	protocol_type	service	flag	src_bytes	dst_bytes	land	wrong_fragment	urgent	hot	num_failed_logins	logged_in	num_compromised	root_she
0	0	tcp	private	REJ	0	0	0	0	0	0	0	0	0	
1	0	tcp	private	REJ	0	0	0	0	0	0	0	0	0	
2	2	tcp	ftp_data	SF	12983	0	0	0	0	0	0	0	0	
3	0	icmp	eco_i	SF	20	0	0	0	0	0	0	0	0	
4	1	tcp	telnet	RSTO	0	15	0	0	0	0	0	0	0	



In [41]:
```
# Descriptive statistics
train.describe()
```

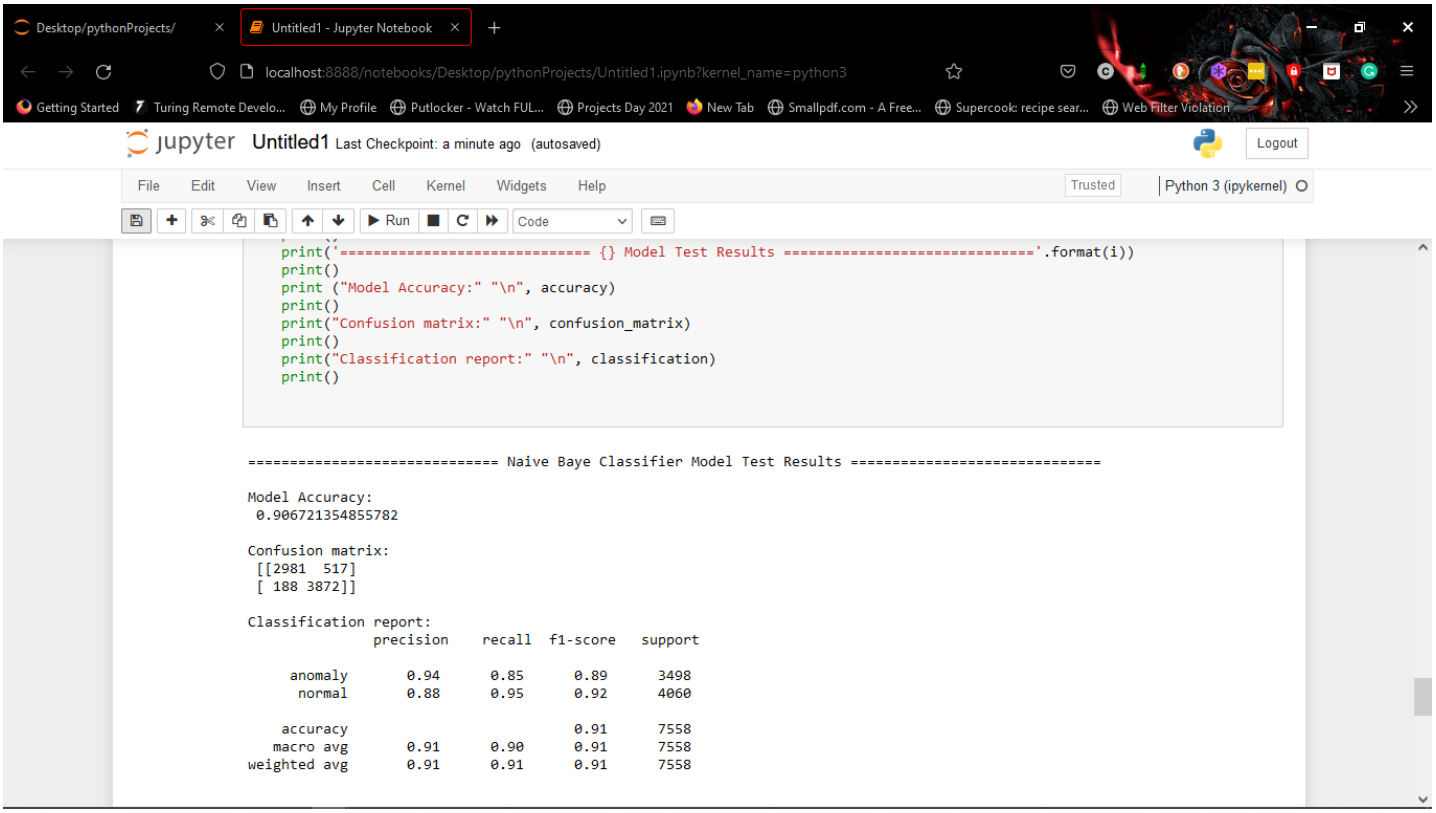
Out[41]:

	duration	src_bytes	dst_bytes	land	wrong_fragment	urgent	hot	num_failed_logins	logged_in	num_compror
count	25192.000000	2.519200e+04	2.519200e+04	25192.000000	25192.000000	25192.00000	25192.000000	25192.000000	25192.000000	25192.00
mean	305.054104	2.433063e+04	3.491847e+03	0.000079	0.023738	0.00004	0.198039	0.001191	0.394768	0.22
std	2686.555640	2.410805e+06	8.883072e+04	0.008910	0.260221	0.00630	2.154202	0.045418	0.488811	10.41
min	0.000000	0.000000e+00	0.000000e+00	0.000000	0.000000	0.00000	0.000000	0.000000	0.000000	0.00
25%	0.000000	0.000000e+00	0.000000e+00	0.000000	0.000000	0.00000	0.000000	0.000000	0.000000	0.00
50%	0.000000	4.400000e+01	0.000000e+00	0.000000	0.000000	0.00000	0.000000	0.000000	0.000000	0.00
75%	0.000000	2.790000e+02	5.302500e+02	0.000000	0.000000	0.00000	0.000000	0.000000	1.000000	0.00
max	42862.000000	3.817091e+08	5.151385e+06	1.000000	3.000000	1.00000	77.000000	4.000000	1.000000	884.00

In [42]:
```
print(train['num_outbound_cmds'].value_counts())
print(test['num_outbound_cmds'].value_counts())
```

```
0    25192
Name: num_outbound_cmds, dtype: int64
0    22544
Name: num_outbound_cmds, dtype: int64
```

In [43]:
```
#'num_outbound_cmds' is a redundant column so remove it from both train & test datasets
```


Desktop/pythonProjects/
Untitled1 - Jupyter Notebook
localhost:8888/notebooks/Desktop/pythonProjects/Untitled1.ipynb?kernel_name=python3
Getting Started
Turing Remote Develo...
My Profile
Putlocker - Watch FUL...
Projects Day 2021
New Tab
Smallpdf.com - A Free...
Supercook: recipe sear...
Web Filter Violation
jupyter
Untitled1
Last Checkpoint: a minute ago (autosaved)
Logout
File
Edit
View
Insert
Cell
Kernel
Widgets
Help
Trusted
Python 3 (ipykernel)
Run
Code
print('============================== {} Model Test Results ==============================='.format(i))
print()
print ("Model Accuracy:" "\n", accuracy)
print()
print("Confusion matrix:" "\n", confusion_matrix)
print()
print("Classification report:" "\n", classification)
print()
============================== Naive Baye Classifier Model Test Results ==============================
Model Accuracy:
0.906721354855782
Confusion matrix:
[[2981 517]
[188 3872]]
Classification report:
precision recall f1-score support
anomaly 0.94 0.85 0.89 3498
normal 0.88 0.95 0.92 4060
accuracy 0.91 7558
macro avg 0.91 0.90 0.91 7558
weighted avg 0.91 0.91 0.91 7558