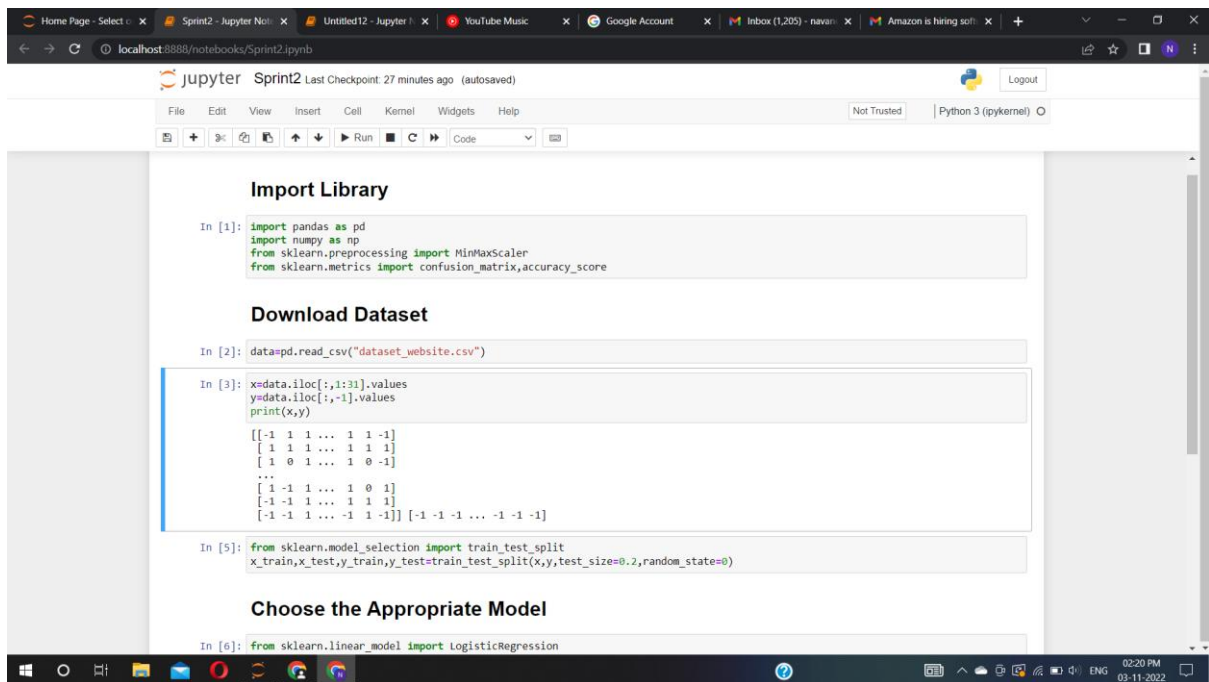


PROJECT DEVELOPMENT PHASE

Delivery of Sprint-2

Date	3 November 2022
Team ID	PNT2022TMID07498
Project Name	Web Phishing Detection
Maximum Marks	4 Marks



The screenshot shows a Jupyter Notebook interface with the following content:

```
import pandas as pd
import numpy as np
from sklearn.preprocessing import MinMaxScaler
from sklearn.metrics import confusion_matrix, accuracy_score
```

Download Dataset

```
data = pd.read_csv("dataset_website.csv")

x = data.iloc[:, 1:31].values
y = data.iloc[:, -1].values
print(x, y)
```

```
[[[-1  1  1 ...  1  1 -1]
 [ 1  1  1 ...  1  1  1]
 [ 1  0  1 ...  1  0 -1]
 ...
 [ 1 -1  1 ...  1  0  1]
 [-1 -1  1 ...  1  1  1]
 [-1 -1  1 ... -1  1 -1]] [-1 -1 -1 ... -1 -1 -1]
```

```
from sklearn.model_selection import train_test_split
x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.2, random_state=0)
```

Choose the Appropriate Model

```
from sklearn.linear_model import LogisticRegression
```

Home Page - Select x Sprint2 - Jupyter Notebook x Untitled12 - Jupyter Notebook x YouTube Music x Google Account x Inbox (1,205) - navan x Amazon is hiring soft x

localhost:8888/notebooks/Sprint2.ipynb

jupyter Sprint2 Last Checkpoint: 27 minutes ago (autosaved)

File Edit View Insert Cell Kernel Widgets Help Not Trusted Python 3 (ipykernel)

[-1 -1 1 ... 1 1 1]
[-1 -1 1 ... -1 1 -1]] [-1 -1 -1 ... -1 -1 -1]

```
In [5]: from sklearn.model_selection import train_test_split
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.2,random_state=0)
```

Choose the Appropriate Model

```
In [6]: from sklearn.linear_model import LogisticRegression
lr=LogisticRegression()
lr.fit(x_train,y_train)

Out[6]: LogisticRegression()

In [7]: y_pred=lr.predict(x_test)
from sklearn.metrics import accuracy_score
log_reg=accuracy_score(y_test,y_pred)
log_reg

Out[7]: 0.9167797376752601

In [8]: import pickle
pickle.dump(lr,open('phishing_website.pkl','wb'))

In [ ]:
```

02:21 PM 03-11-2022