

Project Development Phase

Delivery of Sprint-1

Date	03 November 2022
Team ID	PNT2022TMID07498
Project Name	Web Phishing Detection
Maximum Marks	4 Marks

The screenshot shows a Jupyter Notebook titled "Sprint 1" with a last checkpoint 7 minutes ago. The interface includes a menu bar (File, Edit, View, Insert, Cell, Kernel, Widgets, Help) and a toolbar with icons for file operations, running, and code execution. The notebook content is divided into three sections:

Import Required Libraries

```
In [1]: import pandas as pd
import numpy as np
from sklearn.preprocessing import MinMaxScaler
from sklearn.metrics import confusion_matrix, accuracy_score
```

Read the Dataset

```
In [6]: data=pd.read_csv("dataset_website.csv")
data.head()
```

The output of the code is a preview of the dataset, showing 5 rows and 32 columns. The first five rows are displayed in a table format:

	index	having_IPhaving_IP_Address	URLURL_Length	Shortining_Service	having_At_Symbol	double_slash_redirecting	Prefix_Suffix	having_Sub_Domain	SSI
0	1	-1	1	1	1	-1	-1	-1	
1	2	1	1	1	1	1	-1	0	
2	3	1	0	1	1	1	-1	-1	
3	4	1	0	1	1	1	-1	-1	
4	5	1	0	-1	1	1	-1	1	

Below the table, it indicates "5 rows x 32 columns".

Handling Null Values

jupyter Sprint 1 Last Checkpoint: 7 minutes ago (autosaved) Logout

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3 (pykernel)

Handling Null Values

```
In [7]: data.info()
data.isnull().any()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 11055 entries, 0 to 11054
Data columns (total 32 columns):
#   Column                                Non-Null Count  Dtype
---  --
0   index                                11055 non-null  int64
1   having_IPhaving_IP_Address           11055 non-null  int64
2   URLURL_Length                        11055 non-null  int64
3   Shortining_Service                  11055 non-null  int64
4   having_At_Symbol                     11055 non-null  int64
5   double_slash_redirecting             11055 non-null  int64
6   Prefix_Suffix                        11055 non-null  int64
7   having_Sub_Domain                    11055 non-null  int64
8   Sslfinal_State                       11055 non-null  int64
9   Domain_registration_length           11055 non-null  int64
10  Favicon                               11055 non-null  int64
11  port                                 11055 non-null  int64
12  HTTPS_token                          11055 non-null  int64
13  Request_URL                           11055 non-null  int64
14  URL_of_Anchor                         11055 non-null  int64
15  Links_in_tags                         11055 non-null  int64
16  SFH                                   11055 non-null  int64
17  Submitting_to_email                  11055 non-null  int64
18  Abnormal_URL                         11055 non-null  int64
19  Redirect                              11055 non-null  int64
20  on_mouseover                         11055 non-null  int64
21  Rightclick                           11055 non-null  int64
22  popupwindow                          11055 non-null  int64
23  IFrame                               11055 non-null  int64
```

jupyter Sprint 1 Last Checkpoint: 7 minutes ago (autosaved) Logout

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3 (pykernel)

```
23  IFrame                               11055 non-null  int64
24  age_of_domain                       11055 non-null  int64
25  DNSRecord                           11055 non-null  int64
26  web_traffic                          11055 non-null  int64
27  Page_Rank                           11055 non-null  int64
28  Google_Index                        11055 non-null  int64
29  Links_pointing_to_page               11055 non-null  int64
30  Statistical_report                   11055 non-null  int64
31  Result                              11055 non-null  int64
dtypes: int64(32)
memory usage: 2.7 MB

Out[7]: index                                False
having_IPhaving_IP_Address               False
URLURL_Length                           False
Shortining_Service                       False
having_At_Symbol                         False
double_slash_redirecting                  False
Prefix_Suffix                            False
having_Sub_Domain                        False
Sslfinal_State                           False
Domain_registration_length                False
Favicon                                  False
port                                     False
HTTPS_token                              False
Request_URL                              False
URL_of_Anchor                            False
Links_in_tags                            False
SFH                                       False
Submitting_to_email                       False
Abnormal_URL                             False
Redirect                                 False
on_mouseover                             False
Rightclick                               False
popupwindow                              False
IFrame                                   False
```

Jupyter Sprint 1 Last Checkpoint: 7 minutes ago (autosaved) Logout

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3 (pykernel)

iframe
age_of_domain False
DNSRecord False
web_traffic False
Page_Rank False
Google_Index False
Links_pointing_to_page False
Statistical_report False
Result False
dtype: bool

Splitting The Data

```
In [8]: 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]
```

```
In [9]: from scipy.sparse.construct import random  
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)
```

```
In [ ]:
```

02:39 PM
03-11-2022