

Model Development Phase Template

Date	10 june 2024
Team ID	739879
Project Title	Detection of phishing websites from URLs
Maximum Marks	4 Marks

Initial Model Training Code, Model Validation and Evaluation Report

The initial model training code will be showcased in the future through a screenshot. The model validation and evaluation report will include classification reports, accuracy, and confusion matrices for multiple models, presented through respective screenshots.

Initial Model Training Code:

```
[15]: x = ds.iloc[:,1:-1].values
      y = ds.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]] [0 0 0 ... 0 0 0]

[16]: from sklearn.model_selection import train_test_split
[17]: x_train,x_test,y_train,y_test = train_test_split(x,y,test_size=0.2,random_state=0)
[18]: x_train.shape
[18]: (8844, 30)
[19]: x_test.shape
[19]: (2211, 30)
[20]: y_train.shape
[20]: (8844,)

[21]: y_test.shape
[21]: (2211,)

[22]: from sklearn.linear_model import LogisticRegression
[23]: lr=LogisticRegression()
[24]: lr.fit(x_train,y_train)
```

```
[22]: from sklearn.linear_model import LogisticRegression
```

```
[23]: lr=LogisticRegression()
```

```
[24]: lr.fit(x_train,y_train)
```

```
[24]: ▼ LogisticRegression  
LogisticRegression()
```

```
[25]: y_pred1=lr.predict(x_test)
```

```
[26]: y_pred1
```

```
[26]: array([0, 0, 1, ..., 0, 0, 1], dtype=int64)
```

```
[27]: y_test
```

```
[27]: array([0, 0, 0, ..., 1, 1, 1], dtype=int64)
```

```
[28]: from sklearn.metrics import accuracy_score
```

```
[29]: log_reg=accuracy_score(y_test,y_pred1)
```

```
[30]: log_reg
```

```
[30]: 0.9167797376752601
```

```
[31]: from sklearn.ensemble import RandomForestClassifier  
from sklearn.metrics import accuracy_score
```

```
[32]: model=RandomForestClassifier()  
model.fit(x_train,y_train)
```

```
[32]: ▼ RandomForestClassifier  
RandomForestClassifier()
```

```
[33]: y_pred = model.predict(x_test)  
print("accuracy",accuracy_score(y_test,y_pred))  
accuracy 0.9678878335594754
```

Model Validation and Evaluation Report:

Model	Classification Report	F1 Score	Confusion Matrix
Logistic regression	<pre> 223: from sklearn.linear_model import LogisticRegression 224: lr=LogisticRegression() 225: lr.fit(x_train,y_train) 226: y_pred=lr.predict(x_test) 227: y_pred1 228: array([0, 0, 1, ..., 0, 0, 1], dtype=int64) 229: y_test 230: array([0, 0, 0, ..., 1, 1, 1], dtype=int64) 231: from sklearn.metrics import accuracy_score 232: log_reg_accuracy_score(y_test,y_pred1) 233: log_reg 234: 0.9167777777777778 </pre>	91%	-

Random forest	<pre>[93]: y_pred = model.predict(x_test) print("accuracy", accuracy_score(y_test, y_pred)) accuracy 0.9678878335594754</pre>	96%	-
KNN	-		
Gradient Boosting	-		