

MACHINE LEARNING

SPECIAL ASSIGNMENT

PHISHING EMAIL DETECTION

Presented by

Tanisha (19bec132)

Yashasvi Patel (19bec148)



A brief outline

Pre processing data

Exploratory Data Analysis

Model Training and Prediction

Summary

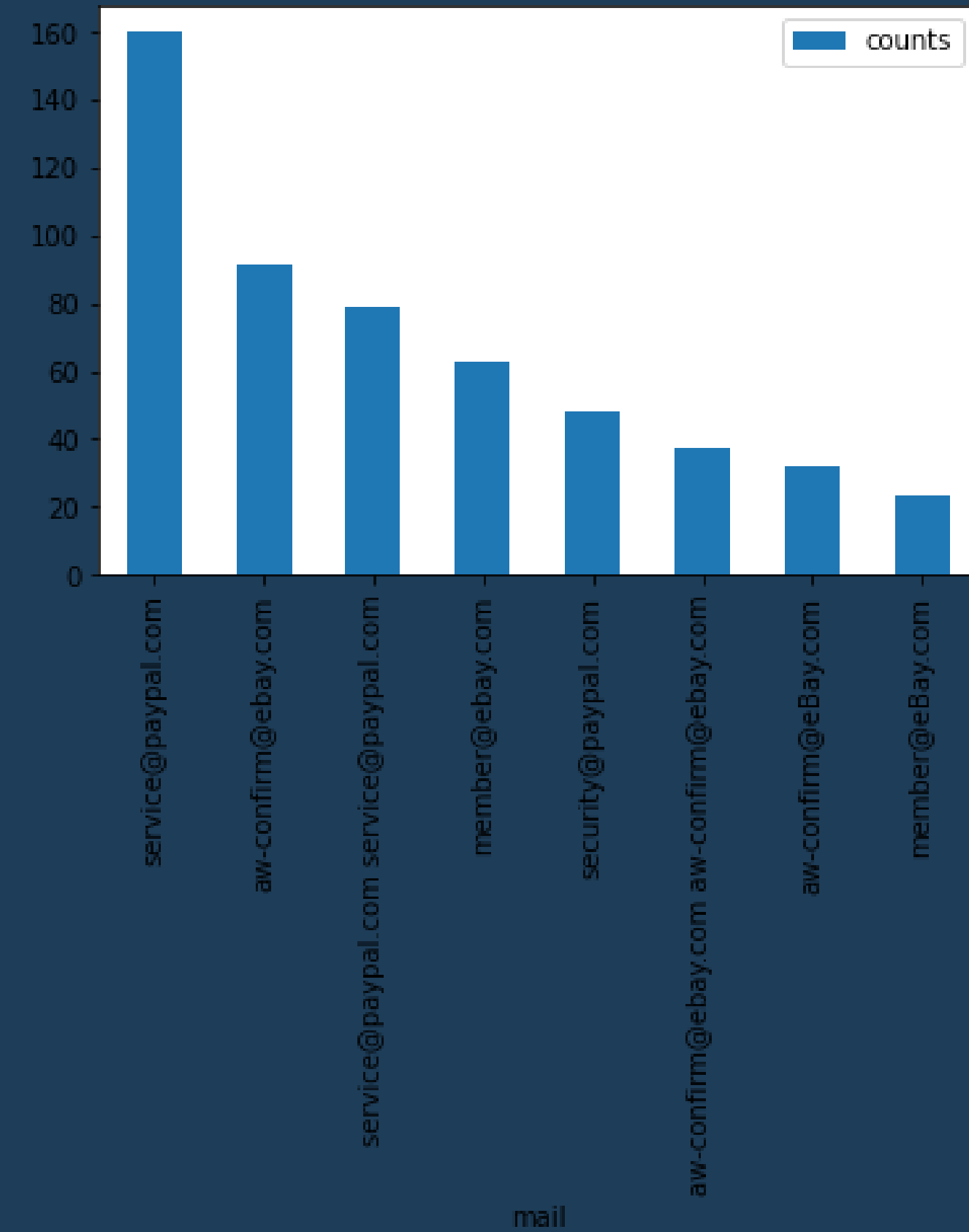
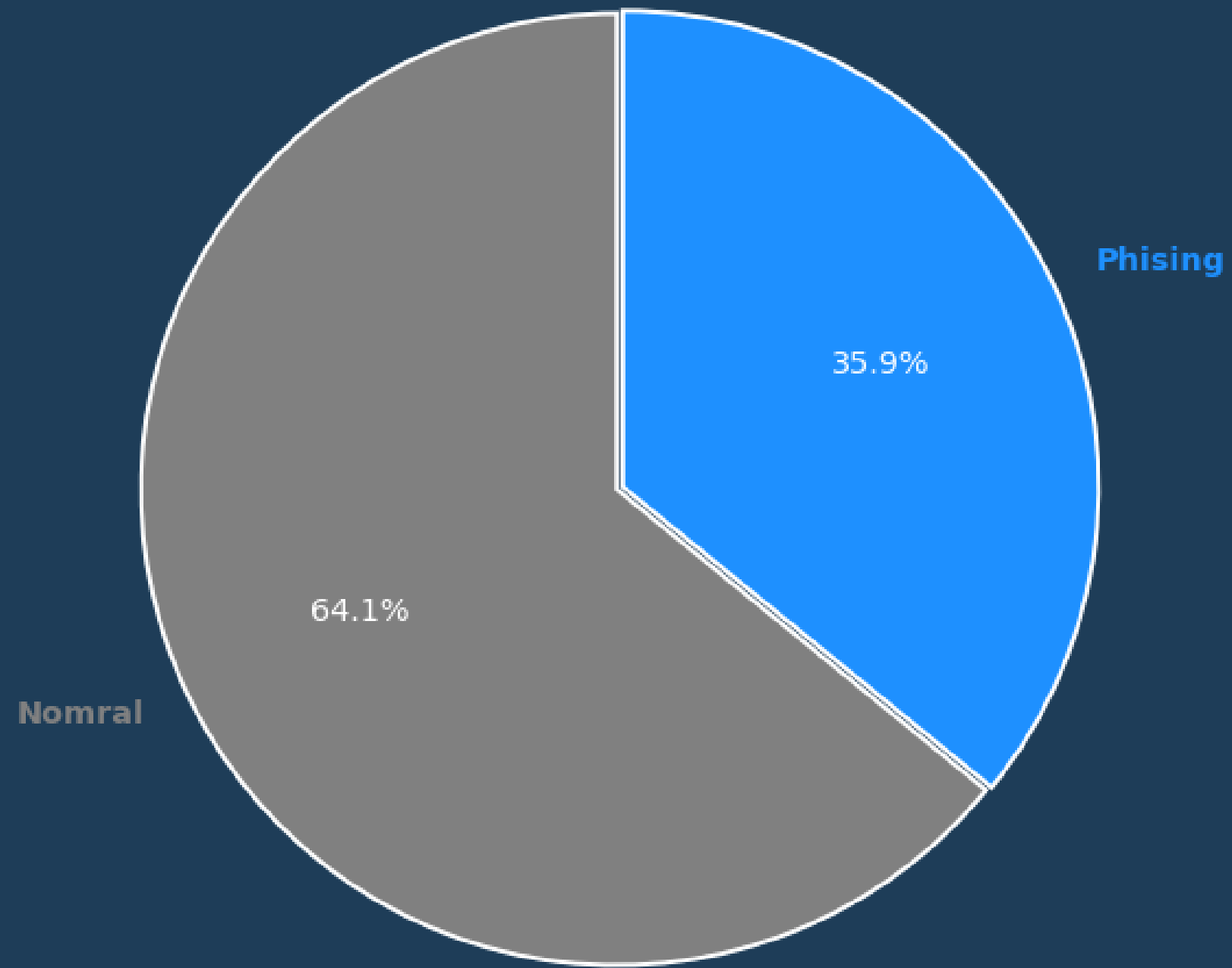


Data Preprocessing

- 1) All the mail bodies with only float values are converted to string by replacing with NAN
- 2) All the stop words like is, the, i, you etc are removed
- 3) Concaving phishing data and normal data



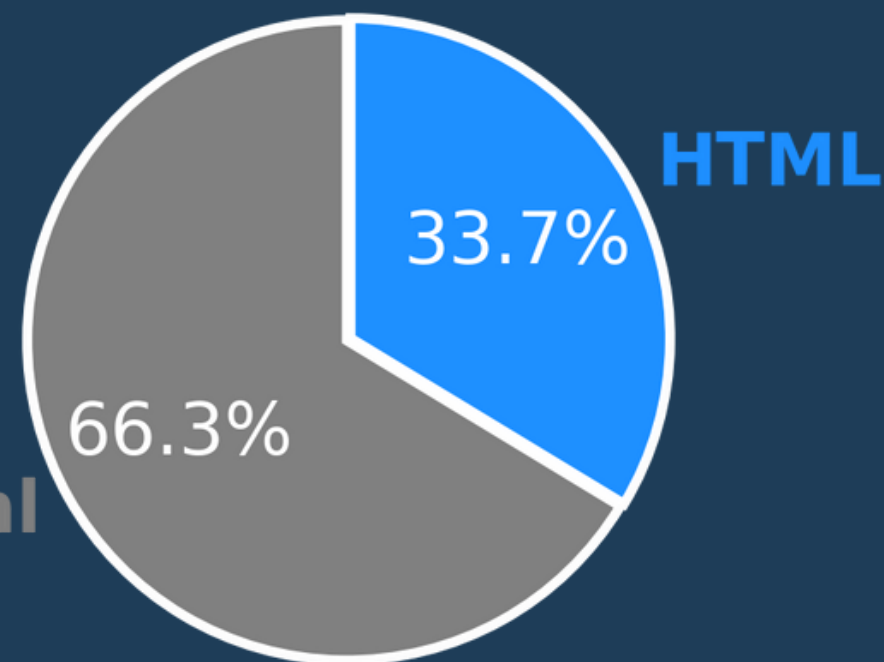
Email



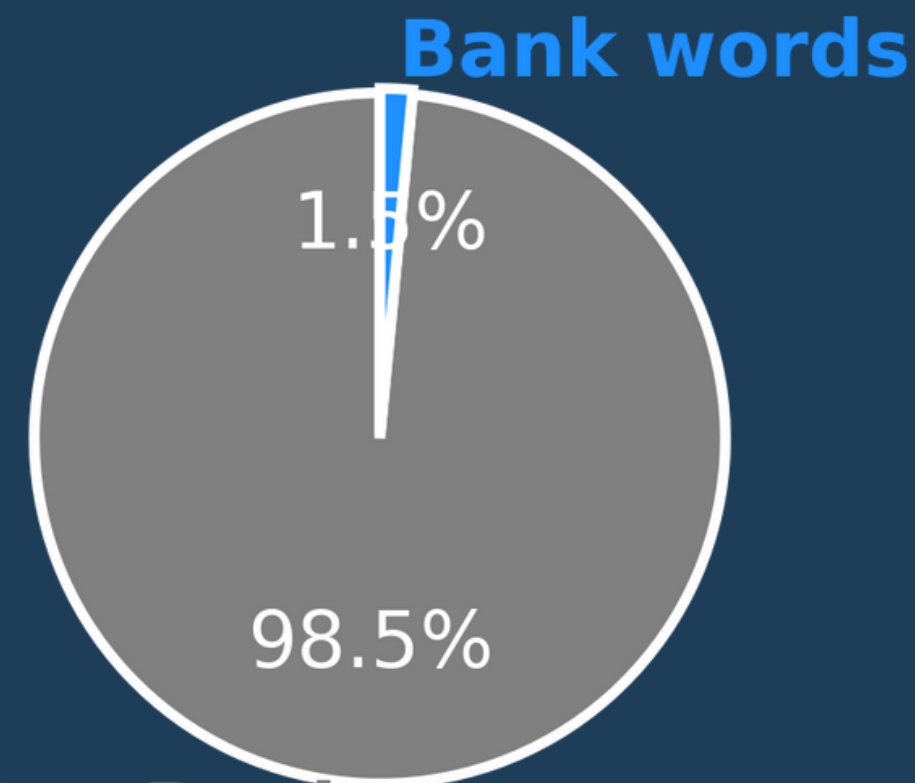
TOOK 50 PHISHING SAMPLES

Common words like HTML Tags, words related to online banking and redirecting URLs were analysed and removed

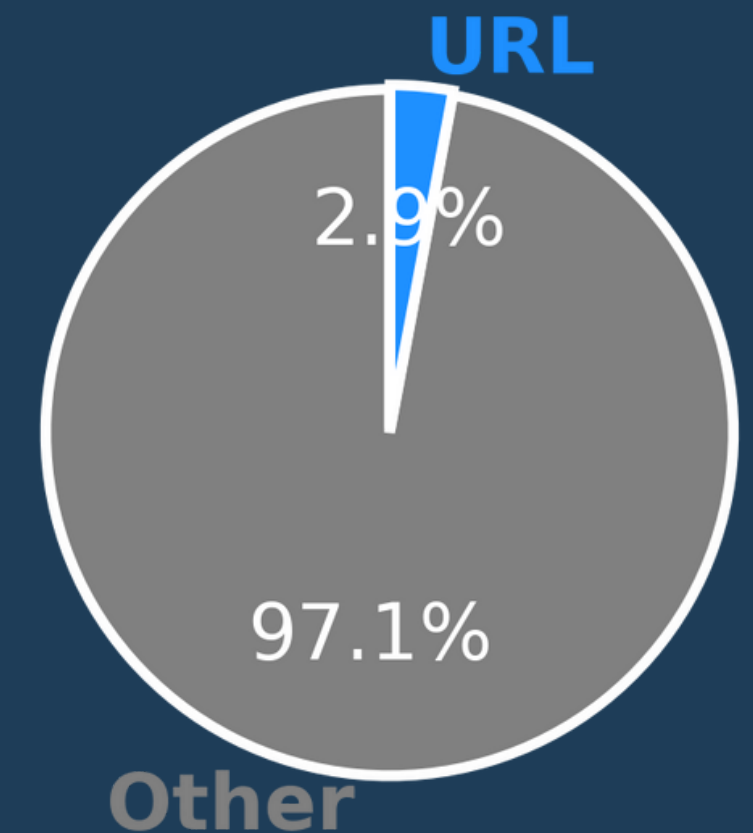
HTML



Bank



URLs





Model Training and Prediction



DECISION TREE



SUPPORT VECTOR MACHINE

SVM-Classification

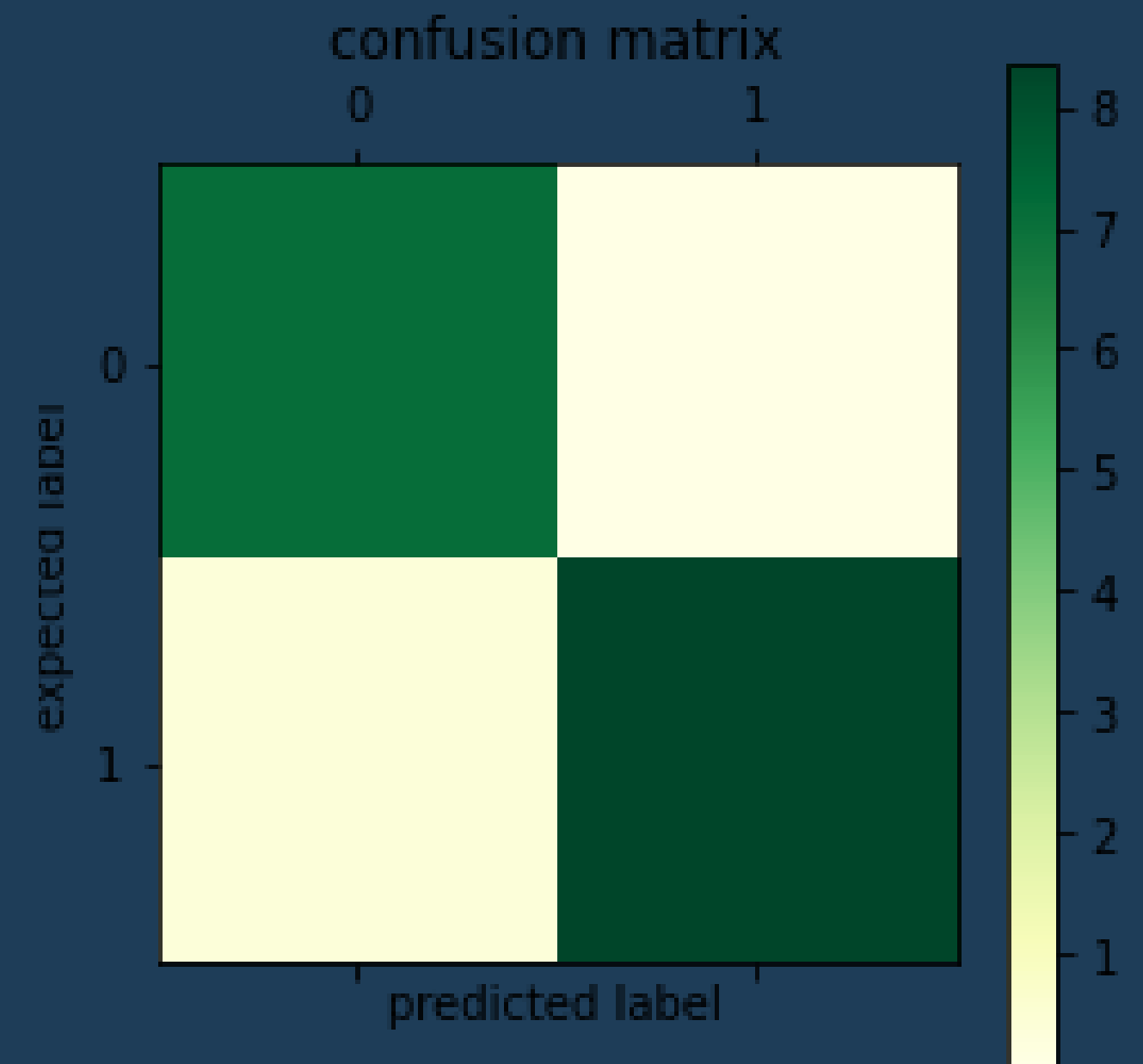
```
confusion matrix::
```

```
[[711   9]
 [ 37 837]]
```

```
accuracy :: 0.9711417816813049
```

```
Classification Report::
```

	precision	recall	f1-score	support
0	0.95	0.99	0.97	720
1	0.99	0.96	0.97	874
accuracy			0.97	1594
macro avg	0.97	0.97	0.97	1594
weighted avg	0.97	0.97	0.97	1594



Decision Tree

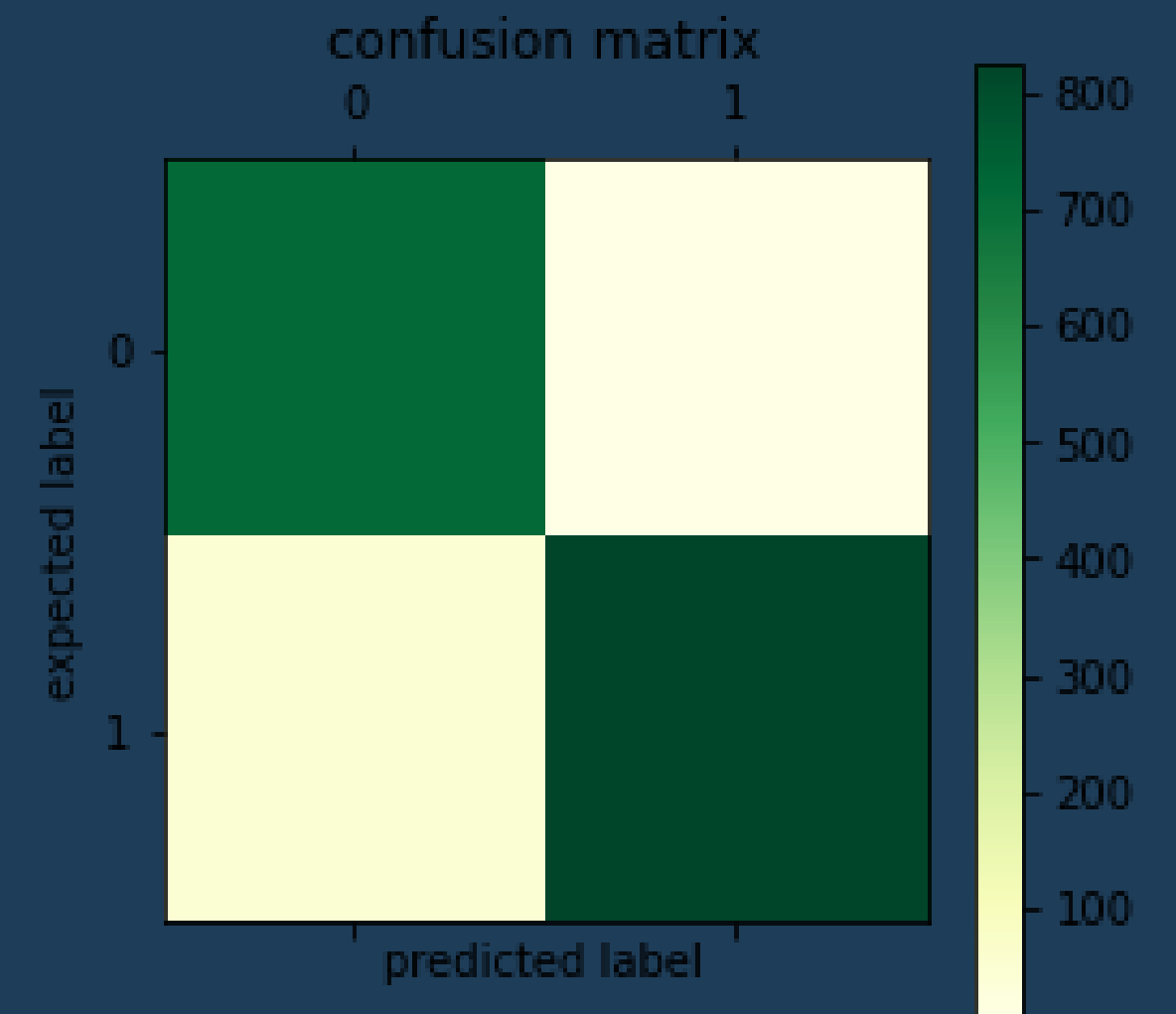
```
confusion matrix::
```

```
[[716   4]
 [ 47 827]]
```

```
accuracy :: 0.9680050188205772
```

```
Classification Report::
```

	precision	recall	f1-score	support
0	0.94	0.99	0.97	720
1	1.00	0.95	0.97	874
accuracy			0.97	1594
macro avg	0.97	0.97	0.97	1594
weighted avg	0.97	0.97	0.97	1594



Decision-Tree

