

Detection of Phishing Urls Using Machine Learning

Recap

- I already made dataset using Free Hosted Domains(FHDs) Urls and trained on Random Forest Classifier and got the results :-

Model name	train_accuracy	test_accuracy
Multilayer Perceptrons(MLP)	97.2	96.5
Random Forest Classifier + MLP (stacked)		95.837
Random Forest Classifier	98.3	95.8
Random Forest Classifier + XGB(stackd)		94.66
Support Vector Machine	93.2	93.6
Logistic Regression	92.4	92.8
XG Boost Classifier	98.1	94.4
Decision Tree Classifier	92.9	90.6

Continuing From Previous Meeting

- So as per your suggestion
- Now I have collected **2512** number of Non Free Hosted Domain (NFHD) Urls which is phishing checked by Virus Total and
- Collected **2499** number of Non Free Hosted Domain (NFHD) Urls which is legitimate checked by Virus Total

Now I have performed experiment in following manner :-

Trained on **FHD urls** and tested on **Non-FHD urls** and vice-versa.

Result

Training on **Free hosted domain urls** on 3937 number of urls

Tested on following urls:-

These testing were conducted on **FHD urls** only and number in bracket indicate total number of urls testing were conducted on.

Accuracy on train Data(3937) : 0.980

Accuracy on test Data(985) : 0.953

Now Testing were conducted on **Non-FHD urls** which is Legitimate and got the accuracy :-

test on general legitimate urls(2499) : 0.76

Now Testing were conducted on **Non-FHD urls** which is Phishing and got the accuracy :-

test on general phishing urls(2512) : 0.06

Cont...

Training on **Non Free hosted domain** urls on 4008 number of urls

Tested on following urls:-

These testing were conducted on **non-fhd urls** only and number in bracket indicate total number of urls testing were conducted on.

Accuracy on train Data(4008): 0.998

Accuracy on test Data(1002): 0.963

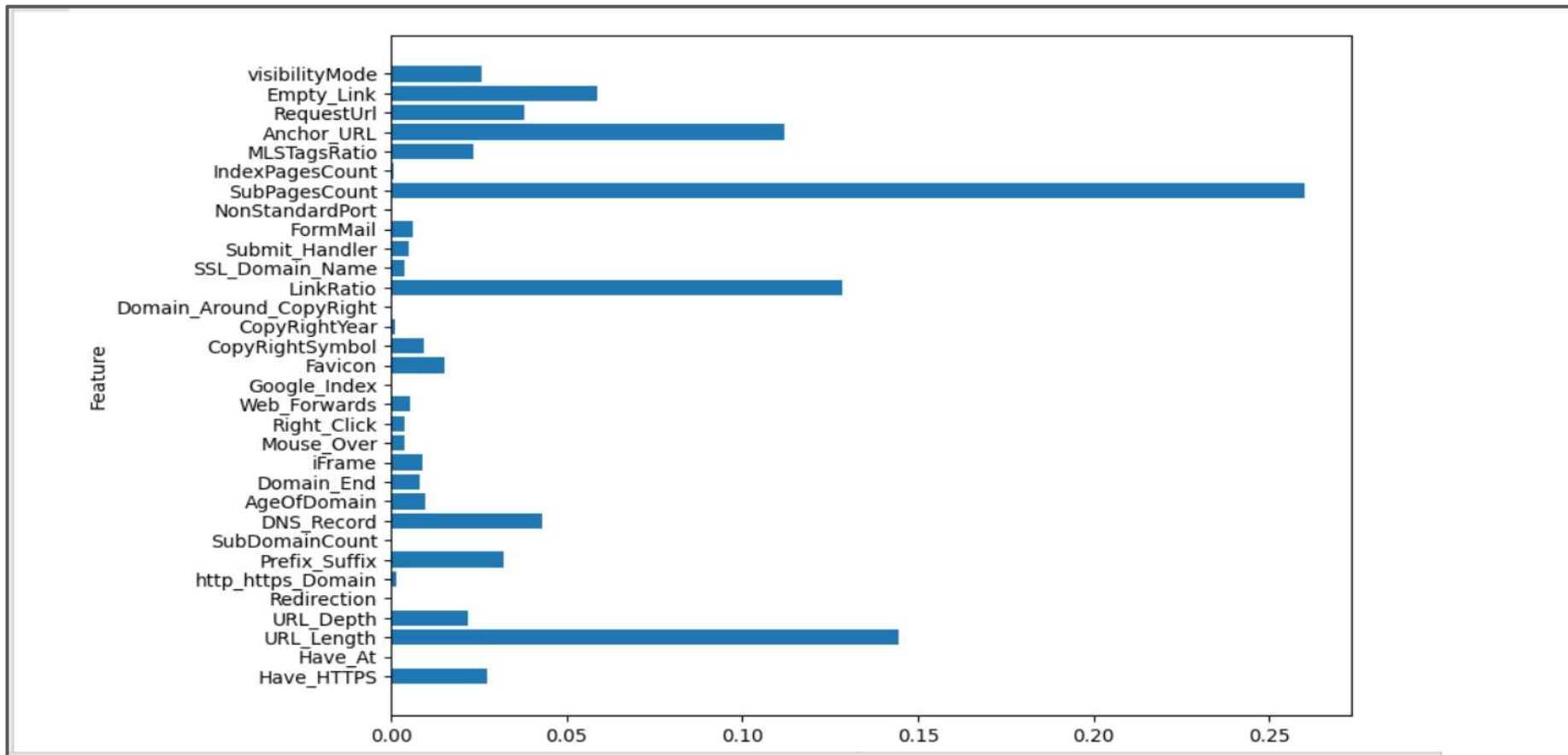
Now Testing were conducted on **FHD urls** which is Legitimate and got the accuracy :-

test on fhd legitimate urls(2499) : 0.000 (quite surprise to see that non-FHD trained model not able to classify any FHD legitimate urls)

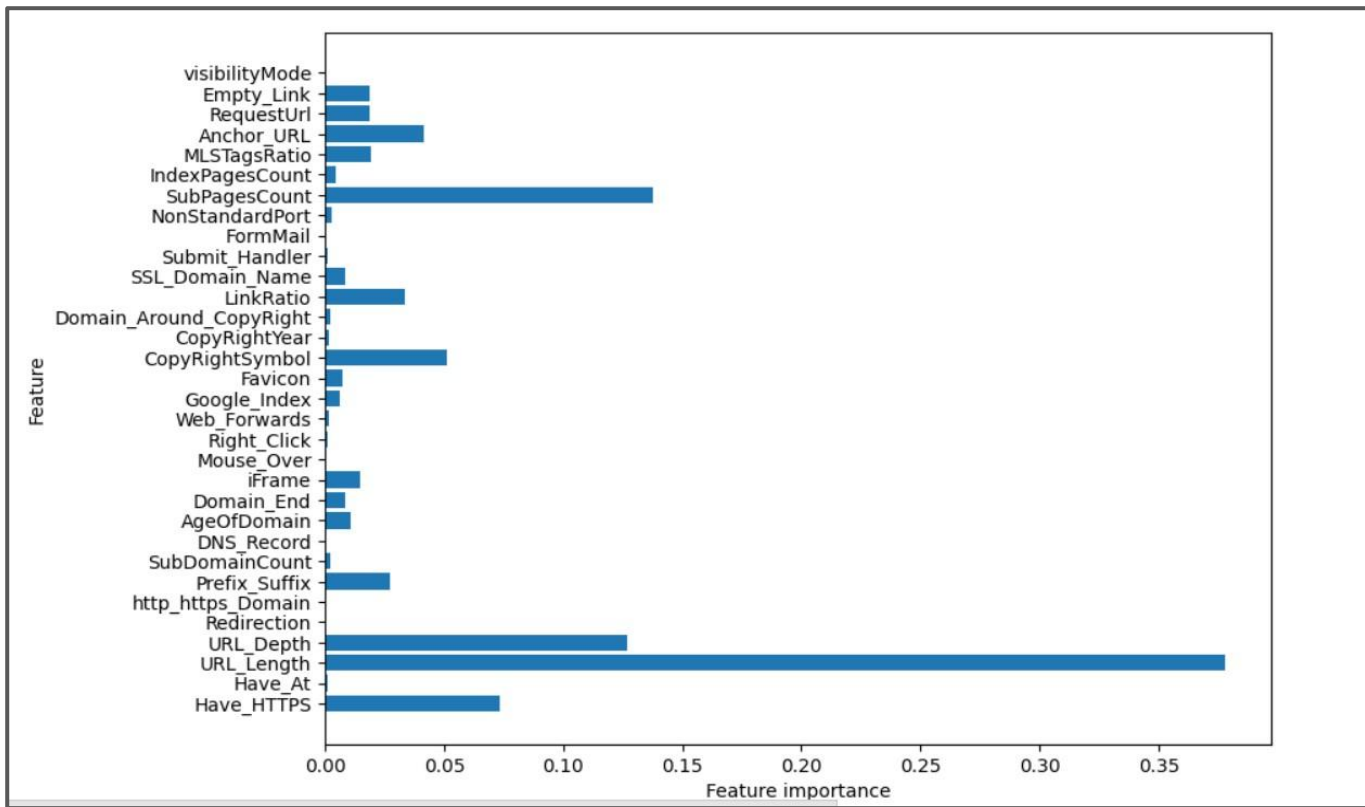
Now Testing were conducted on **FHD urls** which is Phishing and got the accuracy :-

test on fhd phishing urls(2423) : 0.495

Feature importance graph for training on FHD type Urls



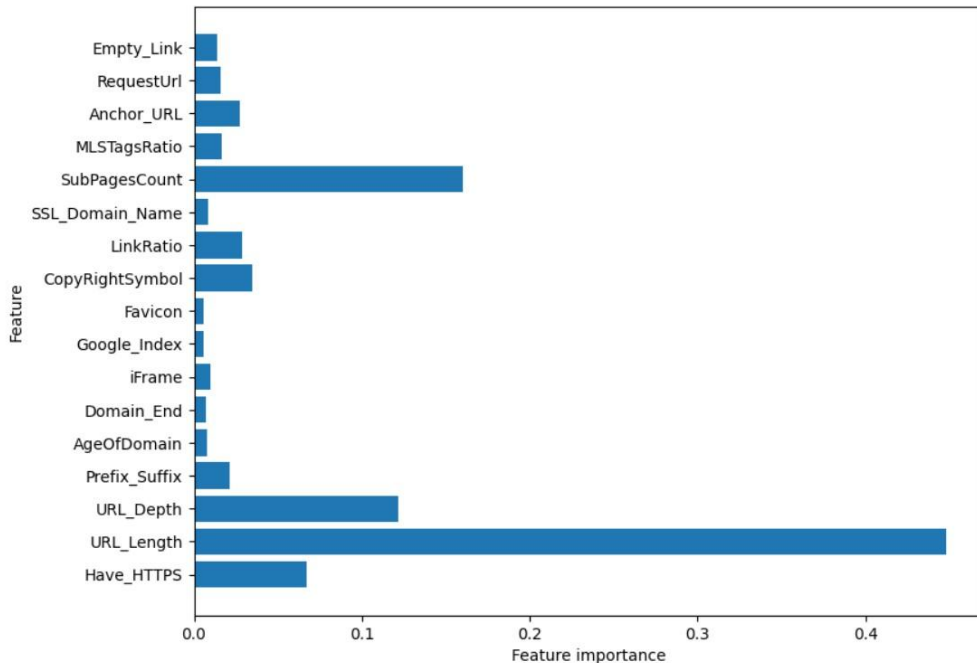
Feature importance graph for training on Non-FHD type Urls(general urls)



Now performing Feature selection

On Non-FHD Urls Training(General Urls)

Choosing TOP 17 features based on their importance value



Result after feature selection on Non-FHD urls

Training on **Non Free hosted domain** urls on 4008 number of urls

Tested on following urls:-

These testing were conducted on **non-fhd urls** only and number in bracket indicate total number of urls testing were conducted on.

Accuracy on train Data(4008): 0.998

Accuracy on test Data(1002): 0.963

Now Testing were conducted on **FHD urls** which is Legitimate and got the accuracy :-

test on fhd legitimate urls(2499) : 0.000 (quite surprise to see that non-FHD trained model not able to classify any FHD legitimate urls)

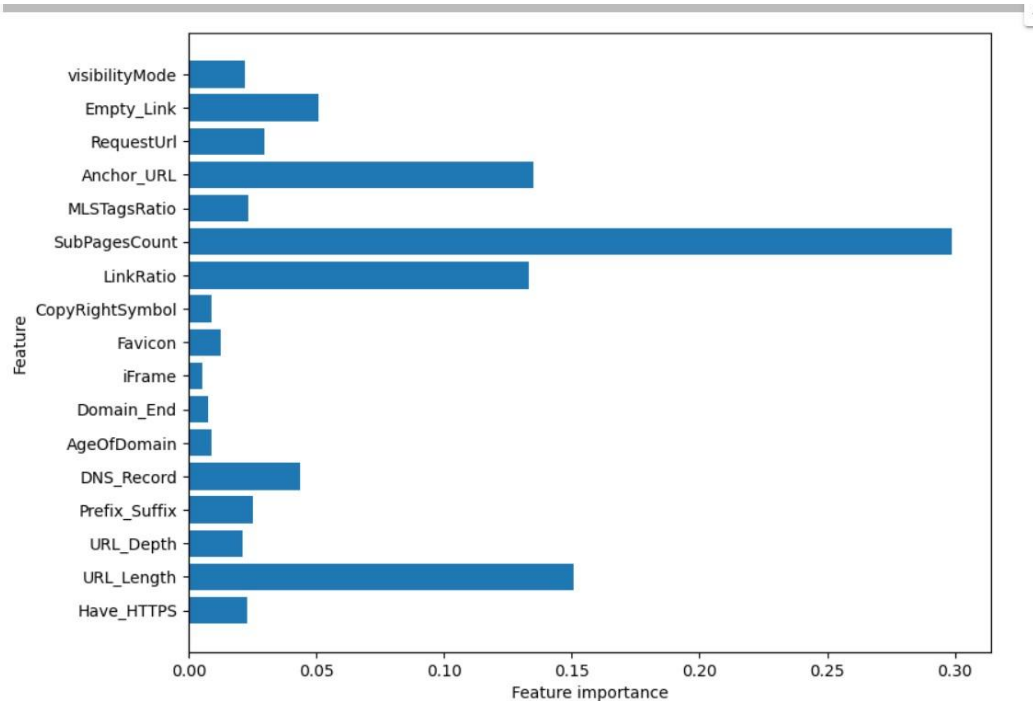
Now Testing were conducted on **FHD urls** which is Phishing and got the accuracy :-

test on fhd phishing urls(2423) : 0.549

Now performing Feature selection

On FHD Urls Training

Choosing TOP 17 features based on their importance value



Result after feature selection on FHD urls

Training on **Free hosted domain urls** on 3937 number of urls

Tested on following urls:-

These testing were conducted on **FHD urls** only and number in bracket indicate total number of urls testing were conducted on.

Accuracy on train Data(3937) : 0.980

Accuracy on test Data(985) : 0.954

Now Testing were conducted on **Non-FHD urls** which is Legitimate and got the accuracy :-

test on general legitimate urls(2499) : 0.681

Now Testing were conducted on **Non-FHD urls** which is Phishing and got the accuracy :-

test on general phishing urls(2512) : 0.041

Thank You Sir...