# Phishing URL Detection

## PHISHING URL DETECTION

Enter URL

Check here

p4code

https://www.youtube.com/

## Website is safe to use...

Continue

## PHISHING URL DETECTION

Enter URL

Check here

p4code

http://games.javascriptsource.com

## Website is unsafe to use...

Still want to Continue

# Table of Content

## Introduction

The Internet has become an indispensable part of our life, However, It also has provided opportunities to anonymously perform malicious activities like Phishing. Phishers try to deceive their victims by social engineering or creating mockup websites to steal information such as account ID, username, password from individuals and organizations. Although many methods have been proposed to detect phishing websites, Phishers have evolved their methods to escape from these detection methods. One of the most successful methods for detecting these malicious activities is Machine Learning. This is because most Phishing attacks have some common characteristics which can be identified by machine learning methods.

## Installation

The Code is written in Python 3.6.10. If you don't have Python installed you can find it here. If you are using a lower version of Python you can upgrade using the pip package, ensuring you have the latest version of pip. To install the required packages and libraries, run this command in the project directory after cloning the repository:

```
pip install -r requirements.txt
```

## Directory Tree

```
├── RandomForestClassifier.pkl
├── static
│   ├── styles.css
├── templates
│   ├── index.html
├── Phishing URL Detection.ipynb
├── Procfile
├── README.md
├── app.py
├── feature.py
├── phishing.csv
├── requirements.txt
```

## Technologies Used
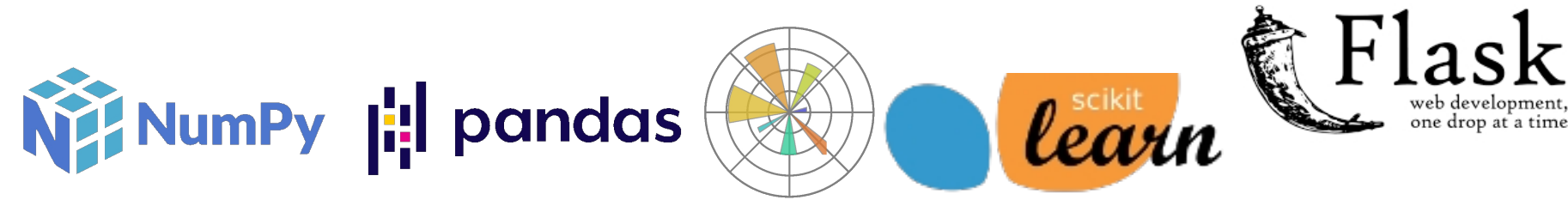




## Result

Accuracy of various model used for URL detection

| | ML Model | Accuracy | f1_score | Recall | Precision |
|---|---|---|---|---|---|
| 1 | Random Forest | 0.959 | 0.963 | 0.983 | 0.975 |

| ML Model | Accuracy | f1_score | Recall | Precision |
|---|---|---|---|---|
| 2 Decision Tree | 0.954 | 0.959 | 0.981 | 0.977 |
| 3 K-Nearest Neighbors | 0.940 | 0.946 | 0.964 | 0.956 |

Feature importance for Phishing URL Detection



Feature importances using permutation on full model
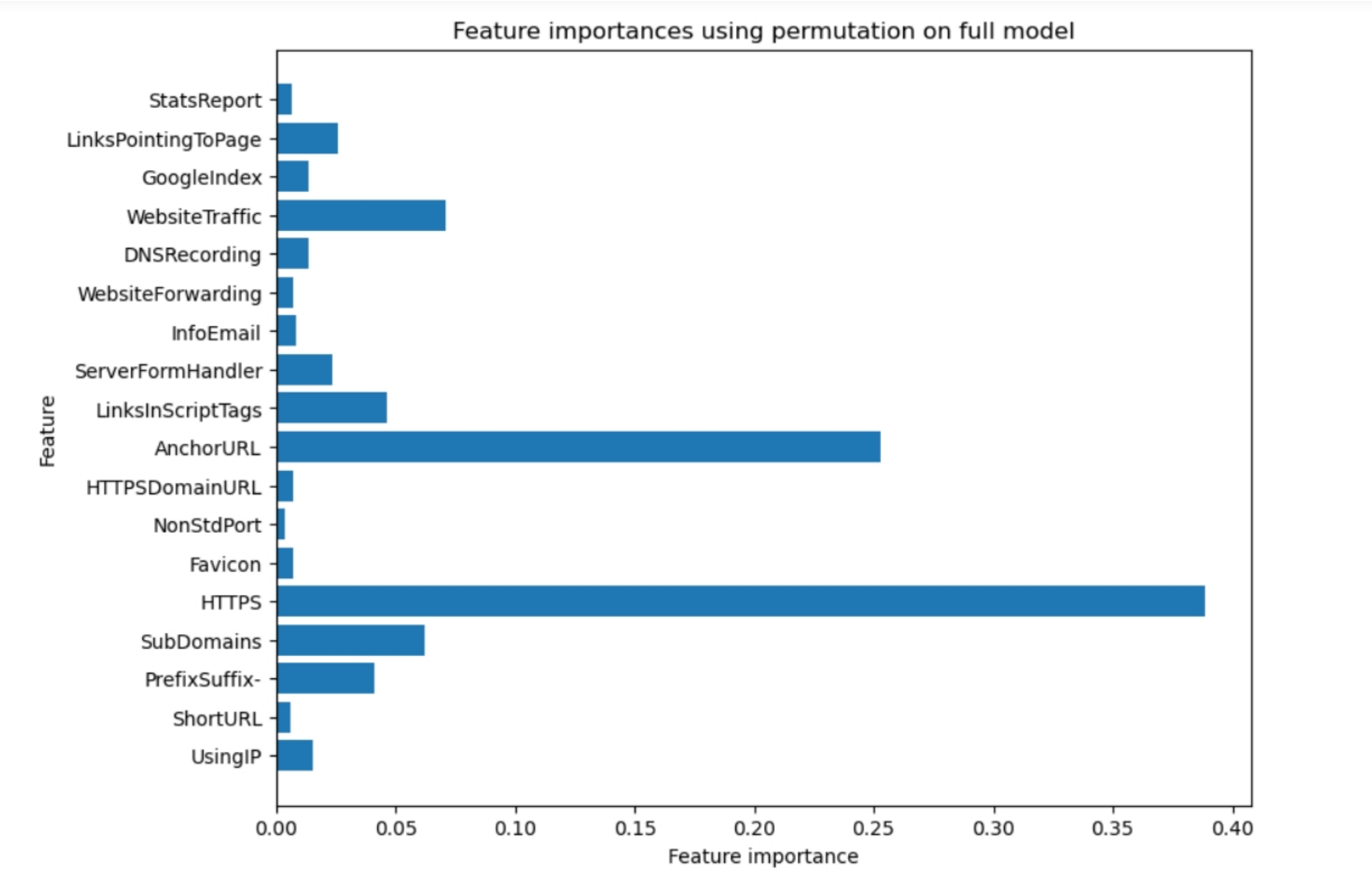
# Conclusion

1. The final take away form this project is to explore various machine learning models, perform Exploratory Data Analysis on phishing dataset and understanding their features.
2. Creating this notebook helped us to learn a lot about the features affecting the models to detect whether URL is safe or not, also we came to know how to tuned model and how they affect the model performance.
3. The final conclusion on the Phishing dataset is that the some feature like "HTTTPS", "AnchorURL", "WebsiteTraffic" have more importance to classify URL is phishing URL or not.
4. Random Forest Classifier currently classify URL upto 95.9% respective classes and hence reduces the chance of malicious attachments.