

Readme

Team : 32 {Reetesh Deshmukh (201IT248) , M Sai Srinivas (201IT233), V Balaji (201IT267)}

About code :

Given source code is generating 15 features from given url. Dataset is collected from various sources to cover large variety for phishing website detection. Model is still under process...

Instruction to run code :

- Place required dataset name which u want for feature extraction in data_extractor.py -> getData() function.

```
def getData(self):  
    df = pd.read_csv('legitimate-test.csv')  
    # df = pd.read_sql('SELECT rec_id, url FRO
```

- Run 'python const_data.py' on terminal
- Enter required details and It will generate .csv file with extracted features.

(for eg. If u want to extract feature of legitimate train dataset)

```
please enter phishing dataset or legitimate dataset  
legitimate  
please enter feature extraction is for training or testing  
train
```

- Feature extraction may take a lot of time to compute since it contains

Phish_test : 518 urls ,**Phish_train :** 4800 urls, **Legitimate_test :** 4800 urls, **Legitimate_train :** 518 urls