Functional Features

Date	17 November 2022
Team ID	PNT2022TMID15890
Project Name	Project – Web Phishing Detection

Functional Features

Functional features are manufactured features which perform a particular function. These functional features are important part of the application which comes in the later portion of the Application Designing Process. They contain the functions that the application uses to perform the said task. The functional features do not affect the model of the application that is being developed.

Functional Features of the application

The application developed uses 30 different features of the URL to predict the fraudulence of the website. The application thus developed has around 30 functions or methods to obtain the features from the URL and checks if the site is phishing website or an authentic website.

Some of the function/methods that checks the fraudulence of the website are as follows

```
def having_ip(url):
    try:
        domain = urlparse(url).netloc
        ip = socket.gethostbyname(domain)
        if domain == ip:
            return 1
        else:
            return -1
    except socket.gaierror:
        return 0
    except:
        return 1
```

```
def url_length(url):
    length = len(url)
    if(length < 54):
        return -1
    elif(54 <= length <= 75):
        return 0
    else:
        return 1
```

```
def sslfinal_state(url):
    try:
        https = urlparse(url).scheme
        if 'https' in https:
            return -1
        else:
            return 1
    except:
        return 1
```

```
def domain_registration_length(url):
    try:
        domain = urlparse(url).netloc
        domain_info = whois.whois(domain)

        domain_creation_date = domain_info.creation_date[0]
        domain_expiration_date = domain_info.expiration_date[0]

        age = (domain_expiration_date - domain_creation_date).days

        if age <= 365:
            return 1
        else:
            return -1
    except:
        return 1
```

The model used in the application is developed using **‘Random Forest Classifier’** which is then trained and tested using 11,000 data samples consisting of 30 features each that is obtained from the dataset send to the model. The prediction feature of the model uses these 30 features’ functional output as input to predict the result as a **‘Phishing’** or **‘Legitimate’** website.