Phishing Detection

Advanced Cyber Security PROJECT REPORT

BACHELOR OF TECHNOLOGY
IN
COMPUTER SCIENCE AND ENGINEERING

by

Paduchuru Abhishek (18BCN7057)

Under the Guidance of

DR. SOMYA RANJAN SAHOO

SCHOOL OF COMPUTER SCIENCE ENGINEERING
VIT-AP UNIVERSITY
AMARAVATI- 522237

JANUARY 2020

ABSTRACT

Phishing is one of the main security threats on the internet today. Web phishing helps the attacker to steal the private information like user name,passwords,credit and debit card detials by entering into our world through back. It will lead to the damage of our software property and most importantly the secure information.

Phishing is one of the social engineering attacks that helps in stealing the main information from the normal and business personalities .Phishers oftenly exploit the trust of users on the appearance of the site by using the web pages which look the same as the real one and through that they steal the data.

In this project a phishing detection approach is used to find the site is legal or illegal. In this we used machine learning technique to blacklist the phishing sites .the technique which used in the approach gives the 96 percent the accurate solution whether the site is phishing or not

TABLE OF CONTENTS

Introduction

The first phishing was done in 1995 on AOL solutions . one phisher successfully got the users personal information like name, credit card details etc.and also conducted an attack on an online payment system. This attack created a main impact to the attackers who attacked the the company

The url phishing websites looks the same like real websites with different ip addresses
When we click on phishing link it will redirect to the attackers server if we provide details it will catch the details and start blackmailing and earns money

At the technical point of view the phishing detection is also possible with some machine learning techniques and codes .which helps in protection from crime links .in this way we can blacklist the phishing sites and be safe

Objective

In this Project, we will focus on the detection model using machine learning techniques. The main contributions are as follows:

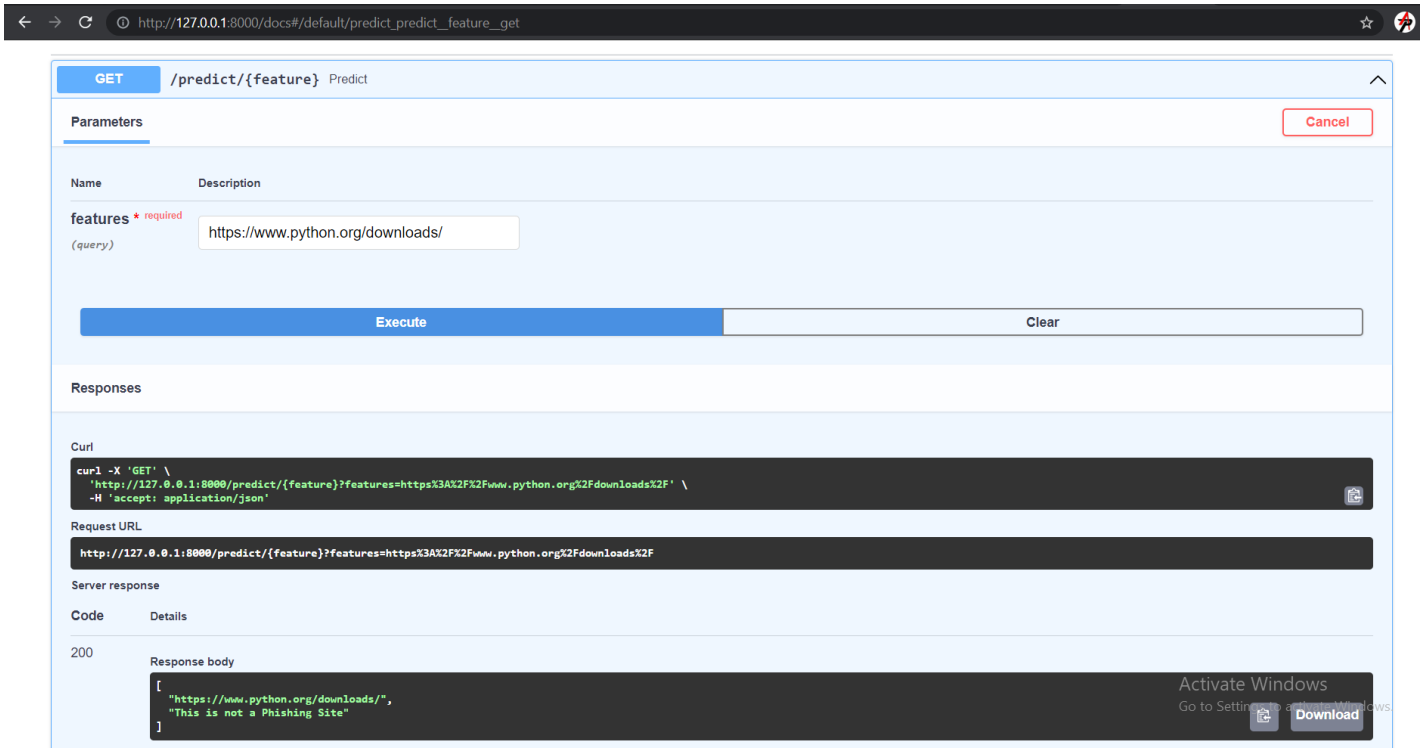
- Firstly collecting dataset links and separating both of them by good and bad links using the main keywords .
- Build The Model
- Train the Model
- Import Links To Prediction
- Deploy The Model using LogisticRegression which gives 96percent accurate solution

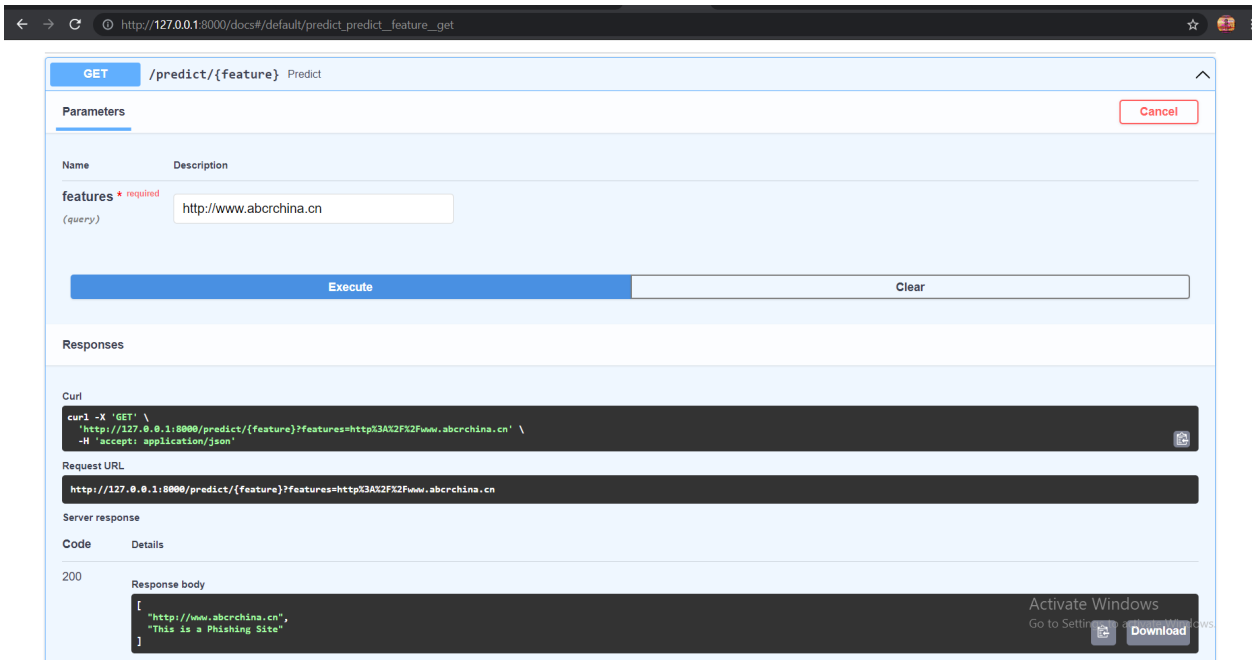
CODE

```
import uvicorn
from fastapi import FastAPI
import joblib,os

app = FastAPI()

#pkl
phish_model = open('phishing.pkl','rb')
phish_model_ls = joblib.load(phish_model)

# ML Aspect
@app.get('/predict/{feature}')
async def predict(features):
        X_predict = []
        X_predict.append(str(features))
        y_Predict = phish_model_ls.predict(X_predict)
        if y_Predict == 'bad':
                result = "This is a Phishing Site"
        else:
                result = "This is not a Phishing Site"

        return (features, result)
if __name__ == '__main__':
        uvicorn.run(app,host="127.0.0.1",port=8000)
```

Proposed model/Framework/ Implementation Details

```
:\Users\HP>cd desktop

:\Users\HP\Desktop>cd new folder (6)

:\Users\HP\Desktop\New folder (6)>cd Phishing Site URLs Prediction

:\Users\HP\Desktop\New folder (6)\Phishing Site URLs Prediction>pyhon prediction_app.py
pyhon' is not recognized as an internal or external command,
perable program or batch file.

:\Users\HP\Desktop\New folder (6)\Phishing Site URLs Prediction>python prediction_app.py
:\Users\HP\AppData\Local\Programs\Python\Python39\lib\site-packages\sklearn\base.py:310: UserWarning: Trying to unpickl
 estimator CountVectorizer from version 0.23.1 when using version 0.24.2. This might lead to breaking code or invalid r
sults. Use at your own risk.
 warnings.warn(
:\Users\HP\AppData\Local\Programs\Python\Python39\lib\site-packages\sklearn\base.py:310: UserWarning: Trying to unpickl
 estimator LogisticRegression from version 0.23.1 when using version 0.24.2. This might lead to breaking code or invali
 results. Use at your own risk.
 warnings.warn(
:\Users\HP\AppData\Local\Programs\Python\Python39\lib\site-packages\sklearn\base.py:310: UserWarning: Trying to unpickl
 estimator Pipeline from version 0.23.1 when using version 0.24.2. This might lead to breaking code or invalid results.
Use at your own risk.
 warnings.warn(
[32mINFO[0m:     Started server process [[36m1144[0m]
[32mINFO[0m:     Waiting for application startup.
[32mINFO[0m:     Application startup complete.
[32mINFO[0m:     Uvicorn running on [1mhttp://127.0.0.1:8000[0m (Press CTRL+C to quit)
```


http://127.0.0.1:8000/docs#/default/predict_predict__feature__get
GET
/predict/{feature} Predict
Parameters
Cancel
Name
Description
features * required
(query)
https://www.python.org/downloads/
Execute
Clear
Responses
Curl
curl -X 'GET' \
'http://127.0.0.1:8000/predict/{feature}?features=https%3A%2F%2Fwww.python.org%2Fdownloads%2F' \
-H 'accept: application/json'
Request URL
http://127.0.0.1:8000/predict/{feature}?features=https%3A%2F%2Fwww.python.org%2Fdownloads%2F
Server response
Code
Details
200
Response body
[
"https://www.python.org/downloads/",
"This is not a Phishing Site"
]
Download
Activate Windows


http://127.0.0.1:8000/docs#/default/predict_predict__feature__get
GET
/predict/{feature}
Predict
Parameters
Cancel
Name
Description
features * required
(query)
http://www.abcrchina.cn
Execute
Clear
Responses
Curl
curl -X 'GET' \
'http://127.0.0.1:8000/predict/{feature}?features=http%3A%2F%2Fwww.abcrchina.cn' \
-H 'accept: application/json'
Request URL
http://127.0.0.1:8000/predict/{feature}?features=http%3A%2F%2Fwww.abcrchina.cn
Server response
Code
Details
200
Response body
[
"http://www.abcrchina.cn",
"This is a Phishing Site"
]
Download
Activate Windows