

PHISHING URL DETECTION A REAL-CASE SCENARIO THROUGH LOGIN URLS

A. Durga Devi ¹, Kuthati.Rakesh,

¹**Assistant professor , PG DEPT, Dantuluri Narayana Raju College, Bhimavaram, Andhara pradesh**
Email:- adurgadevi760@gmail.com

²**PG Student of M.Sc, Dantuluri Narayana Raju College, Bhimavaram, Andhara pradesh**
Email:- kuthatirakesh67@gmail.com

ABSTRACT

Currently, numerous types of cybercrime are organized through the internet. Hence, this study mainly focuses on phishing attacks. Although phishing was first used in 1996, it has become the most severe and dangerous cybercrime on the internet. Phishing utilizes email distortion as its underlying mechanism for tricky correspondences, followed by mock sites, to obtain the required data from people in question. Different studies have presented their work on the precaution, identification, and knowledge of phishing attacks; however, there is currently no complete and proper solution for frustrating them. Therefore, machine learning plays a vital role in defending against cybercrimes involving phishing attacks. The proposed study is based on the phishing URL-based dataset extracted from the famous dataset repository, which consists of phishing and legitimate URL attributes collected from 11000+ website datasets in vector form. After preprocessing, many machine learning algorithms have been applied and designed to prevent phishing URLs and provide protection to the user.

1 INTRODUCTION

The internet plays a crucial role in various aspects of human life. The Internet is a collection of computers connected through telecommunication links such as phone lines, fiber optic lines, and wireless and satellite connections. It is a global computer network. The internet is used to obtain information stored on computers, which are known as hosts and servers. For communication purposes, they used a protocol called Internet protocol/transmission control protocol (IP-TCP). The government is not recognized as an owner of the Internet; many organizations, research agencies, and universities participate in managing the Internet. This has led to many convenient experiences in our lives regarding entertainment, education, banking, industry, online freelancing, social media, medicine, and many other fields in daily life. The internet provides many advantages in different fields of life. In the field of information search, the Internet has become a perfect opportunity to search for data for educational and research purposes. Email is a messaging source in fast way on the Internet through which we can send files, videos, pictures, and any applications, or write a letter to another person around the world. E-commerce is also used on the internet. People can conduct business and financial deals with customers worldwide through e-commerce.

Literature Survey

Rashmi Karnik et al., proposed a model of classification method, kernel-based approach. In this we categories phishing . This method produces estimated accuracy of 95% in detecting the phishing and malware sites. Andrei Butnaru et al., used a supervised Machine Learning algorithm to block phishing attacks, based on novel mixture phishing attacks and compare with Google Safe browsers. Vahid Shahrivari et al., proposed a one of the most successful techniques for identifying these malicious works is Machine Learning. It is because of most Phishing attacks have same features which can be noticed by Machine learning techniques. In this many machine learning-based classifiers are used for forecasting the phishing websites. The main advantage of machine learning is the ability to create

3 IMPLEMENTATION STUDY

EXISTING SYSTEM:

Phishing identification systems based on List use two different lists white lists and blacklists for the association and classification of authorized and phishing webpages. Whitelist based Phishing identification systems produce protected and reliable websites to produce the required data. A suspicious website just needs to match the website of the whitelists; if it is not in the whitelist, it means it is suspicious and threatened by the user. In [20]. To develop a whitelist-based system that generates a whitelist by monitoring and recording the IP address of every website that contains the login interface for the end-user used by the users to enter their details. When the user uses this login interface, the Windows 2008 system displays a warning for the incompatibility of registered information details.

Proposed System & algoritham

Phishing URL-based cyberattack detection is proposed in this study to prevent crime and protect people's privacy.

The dataset consists of 11000+ phishing URL attributes that help classify phishing URLs based on these attributes.

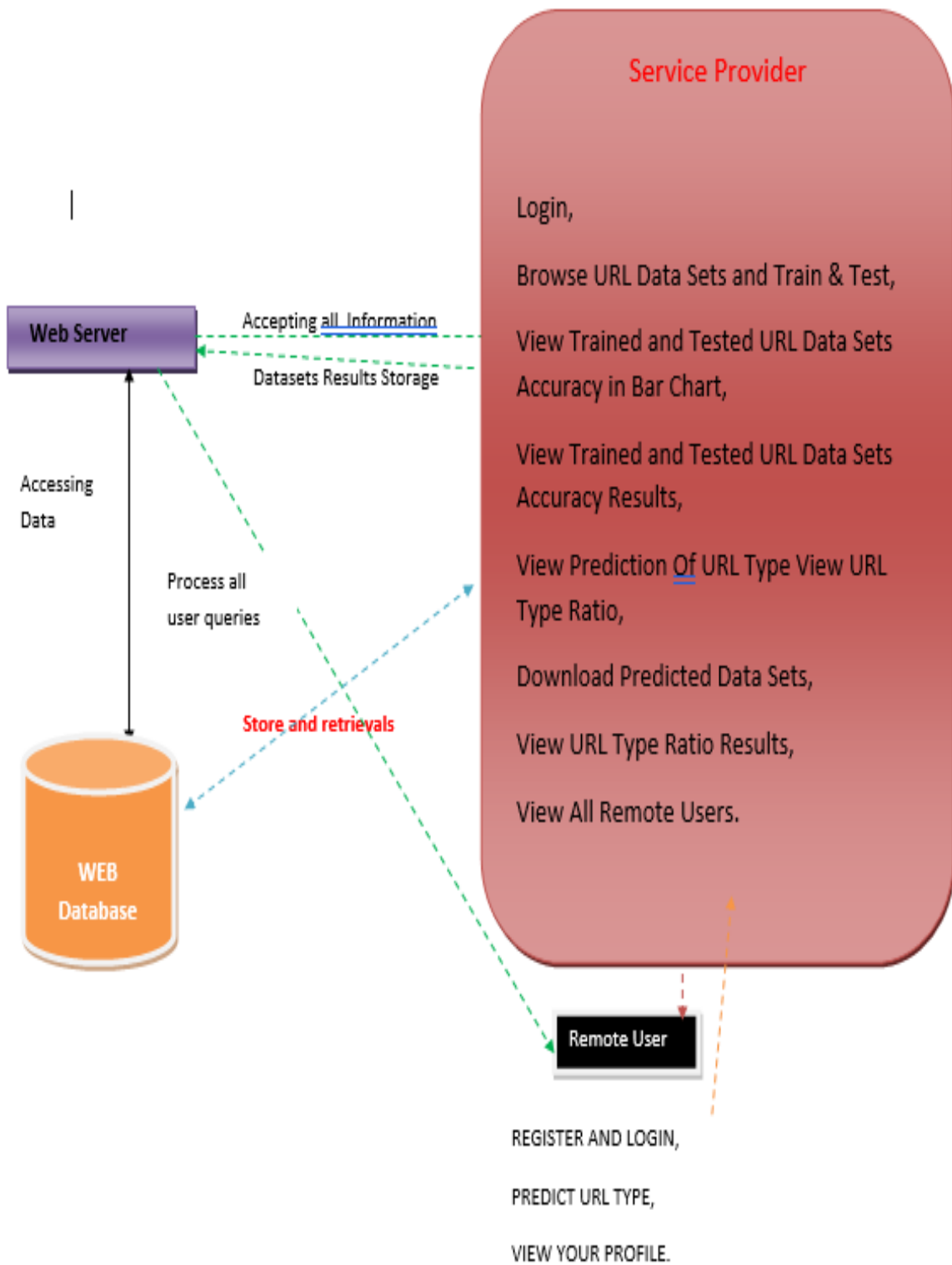


Fig:3.1 System Architecture

IMPLEMENTATION

MODULES:

Service Provider:

In this module, the Service Provider has to login by using valid user name and password. After login successful he can do some operations such as Browse URL Data Sets and Train & Test, View Trained and Tested URL Data Sets Accuracy in Bar Chart, View Trained and Tested URL Data Sets Accuracy Results, View Prediction Of URL Type View URL Type Ratio, Download Predicted Data Sets, View URL Type Ratio Results, View All Remote Users.

View and Authorize Users:

In this module, the admin can view the list of users who all registered. In this, the admin can view the user's details such as, user name, email, address and admin authorizes the users.

5 RESULTS AND DISCUSSION

Screen Shorts:

Web URL Page:

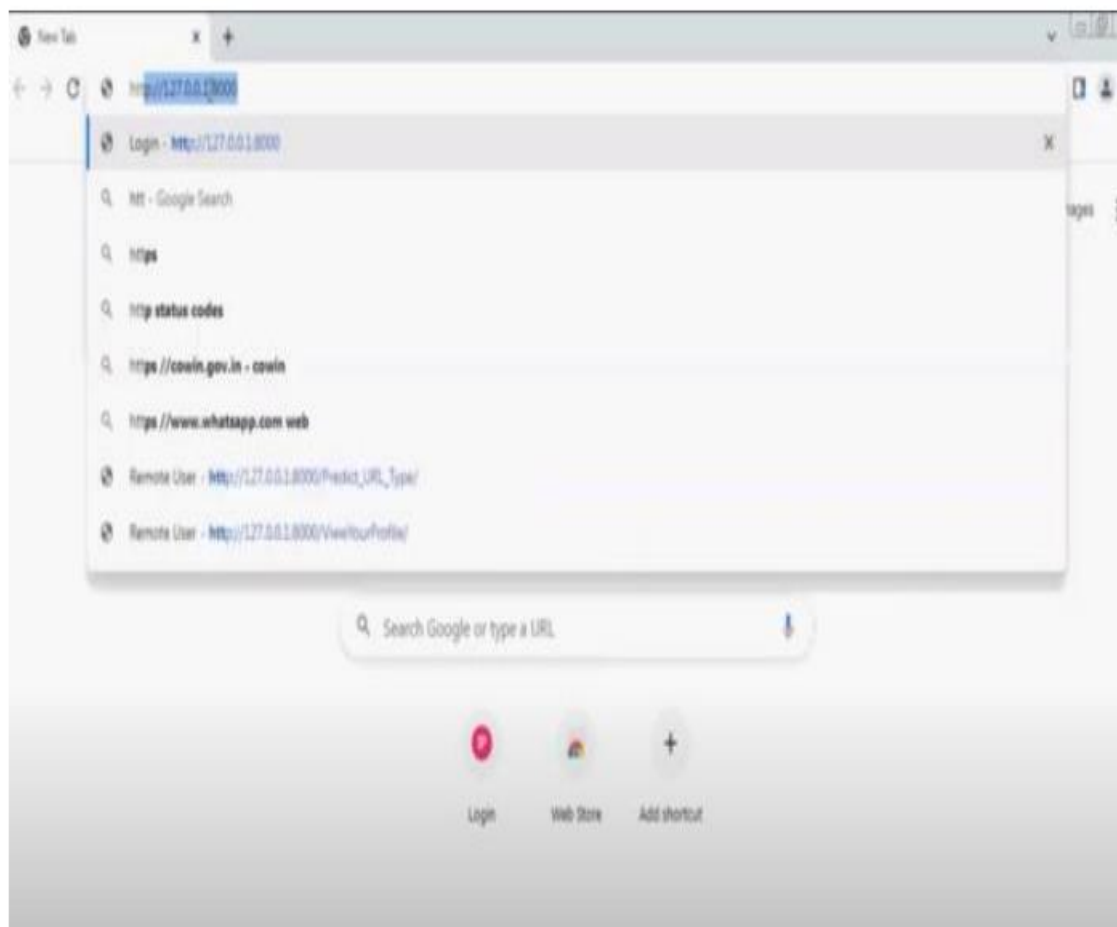


Fig.1

Admin login page:

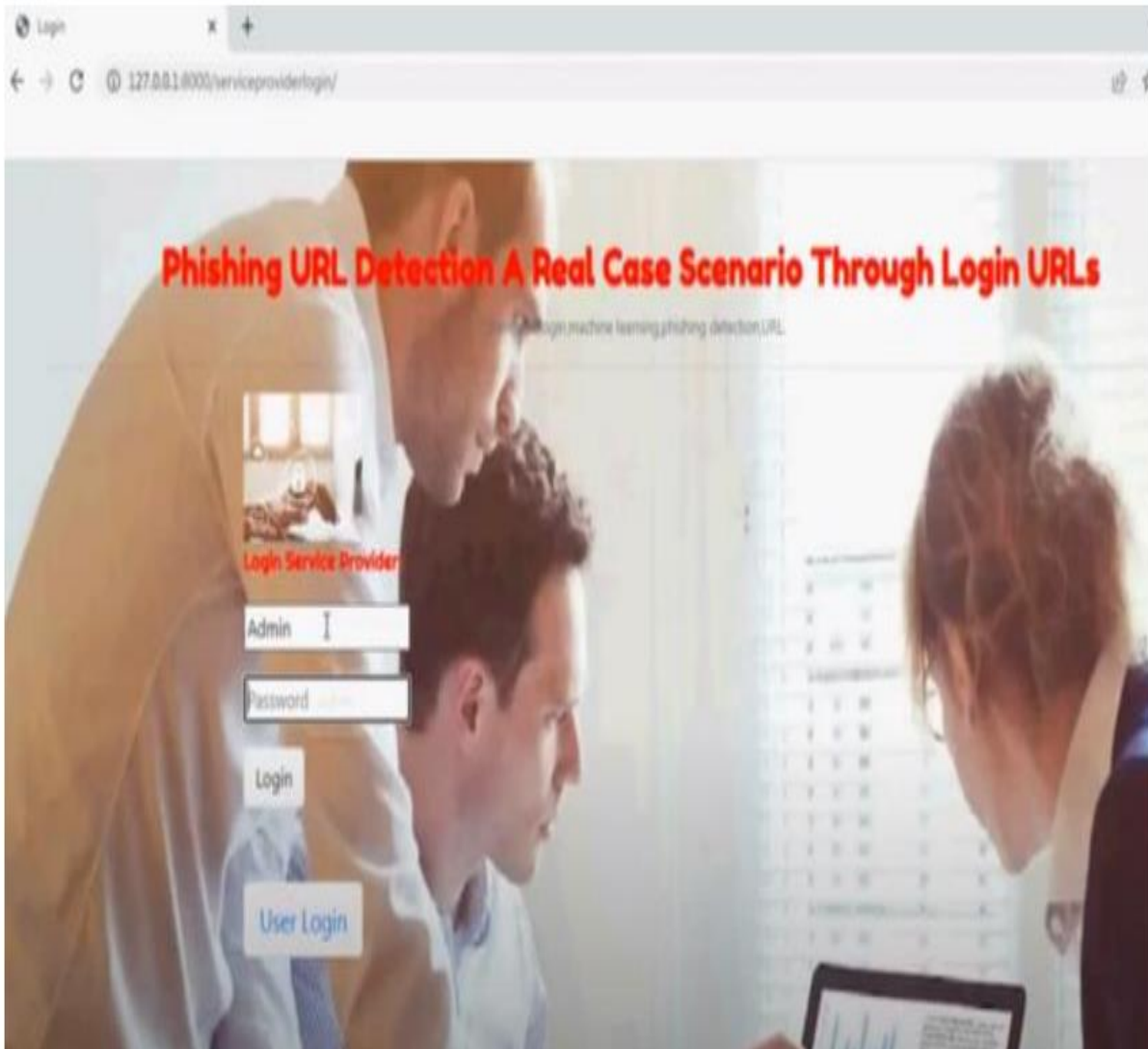
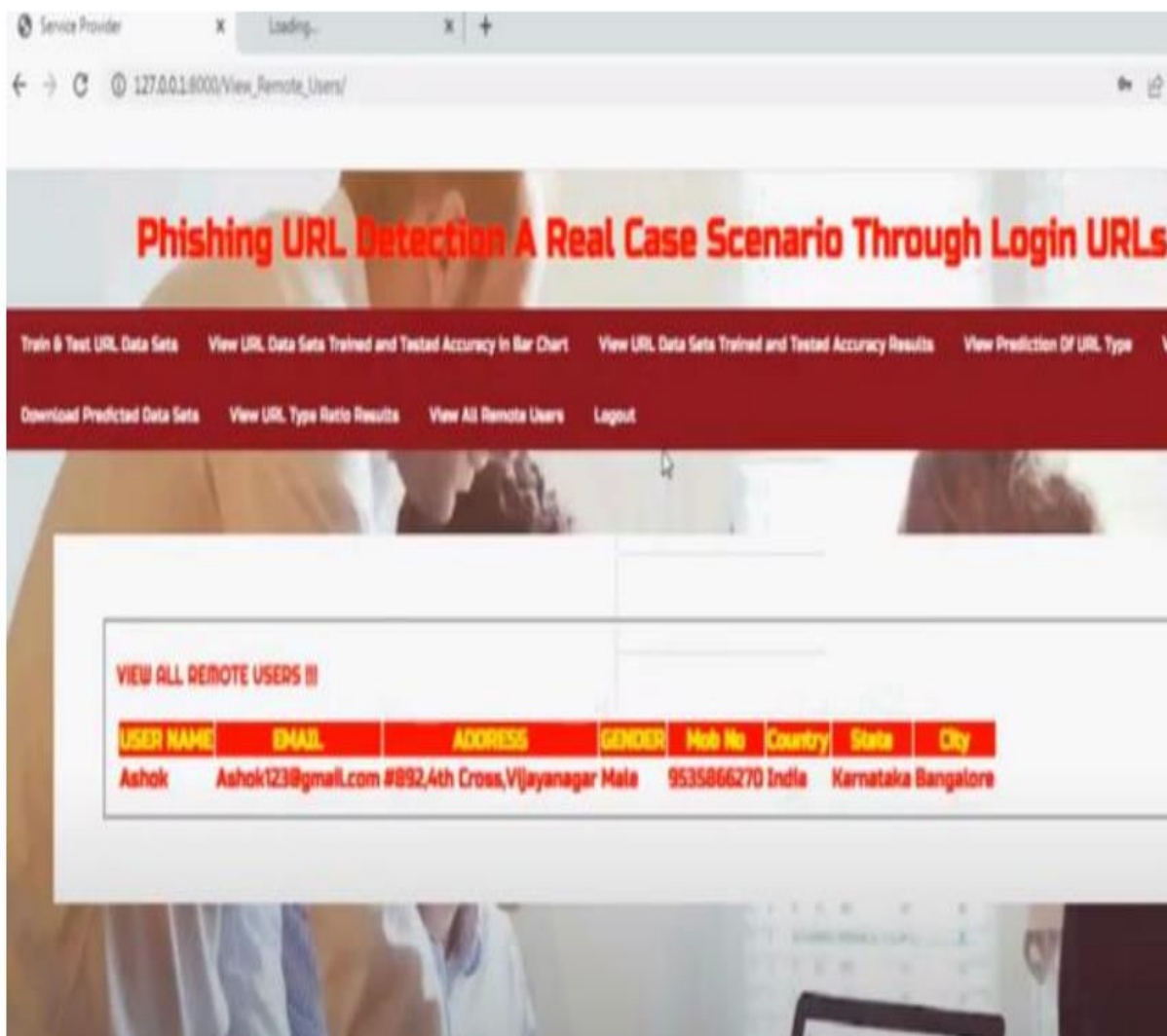


Fig.2

User details:

The screenshot shows a web browser window with the address bar displaying "127.0.0.1:8000/View_Remote_Users/". The page title is "Phishing URL Detection A Real Case Scenario Through Login URLs". The navigation menu includes links for "Train & Test URL Data Sets", "View URL Data Sets Trained and Tested Accuracy in Bar Chart", "View URL Data Sets Trained and Tested Accuracy Results", "View Prediction Of URL Type", "Download Predicted Data Sets", "View URL Type Ratio Results", "View All Remote Users", and "Logout". The main content area displays a table titled "VIEW ALL REMOTE USERS III" with the following data:

USER NAME	EMAIL	ADDRESS	GENDER	Mobile No	Country	State	City
Ashok	Ashok123@gmail.com	#892,4th Cross,Vijayanagar Male		9535866270	India	Karnataka	Bangalore

Fig.3

Register Page:

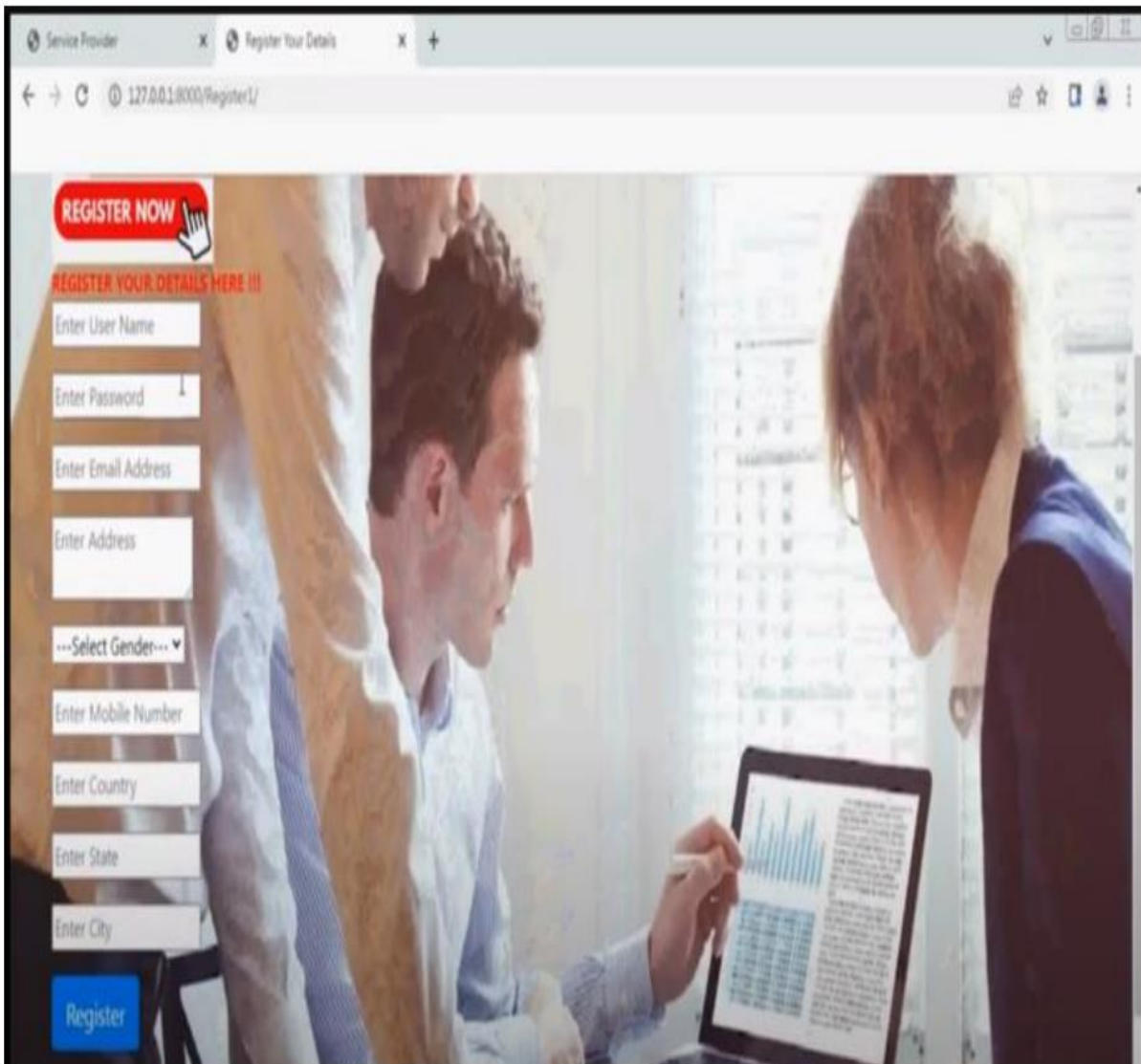


Fig.4

URL upload page:

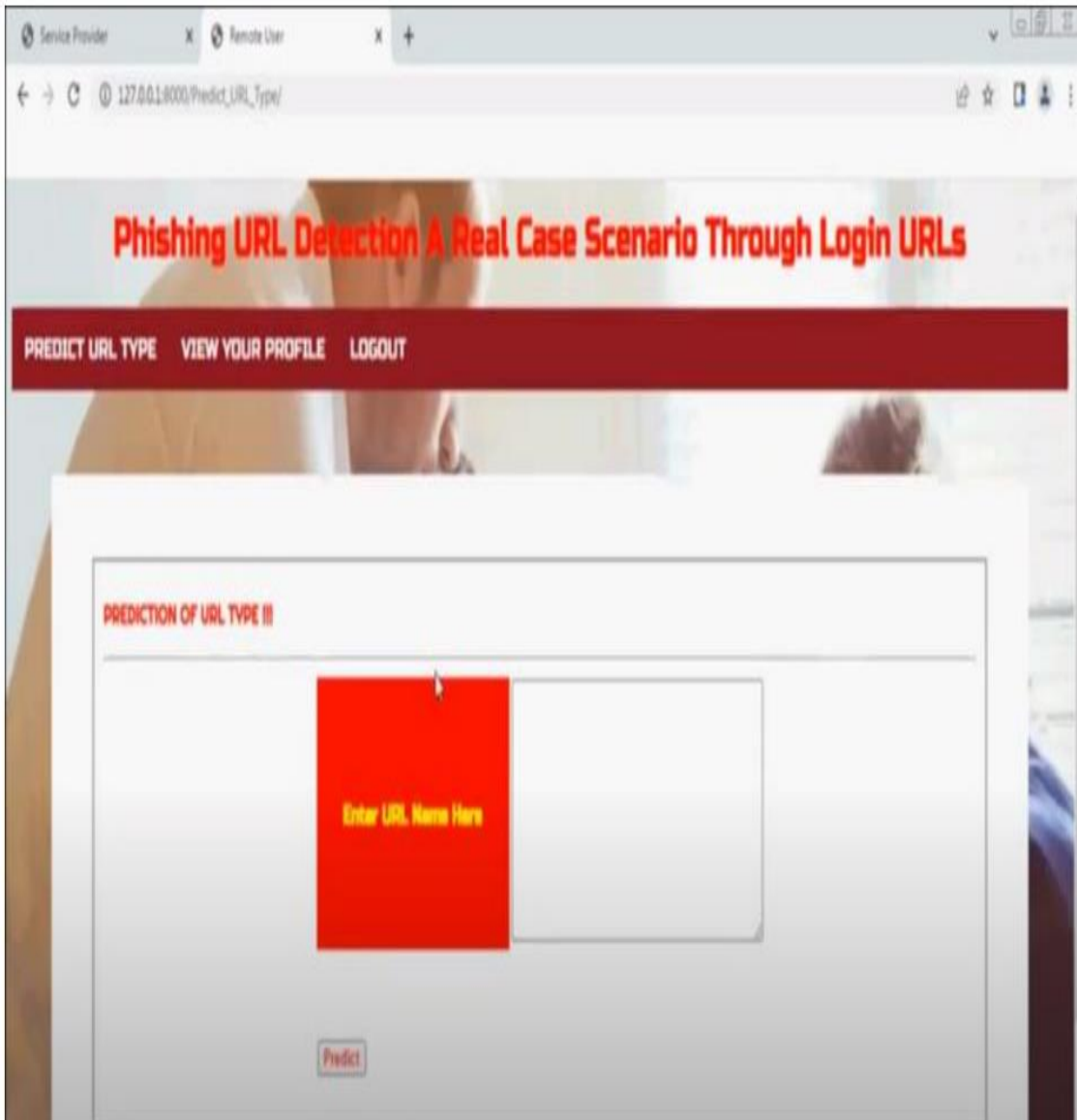


Fig.5

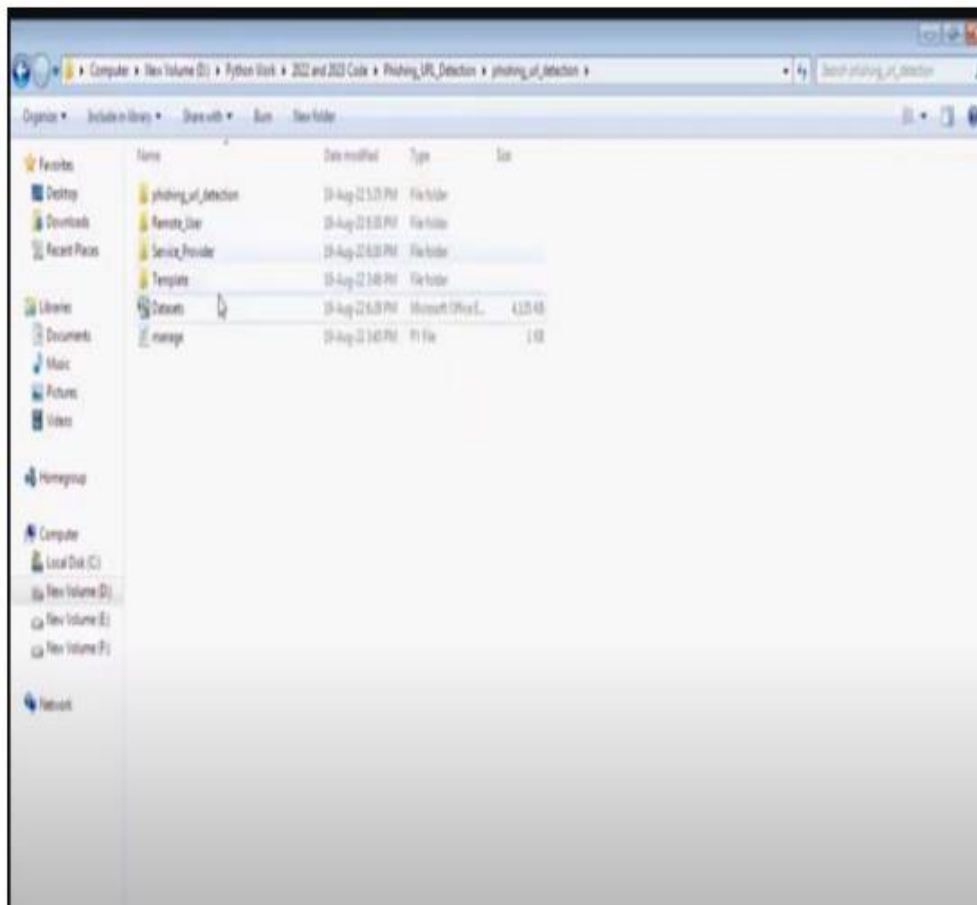
Upload dataset details:

Fig.6

URL dataset with accuracy

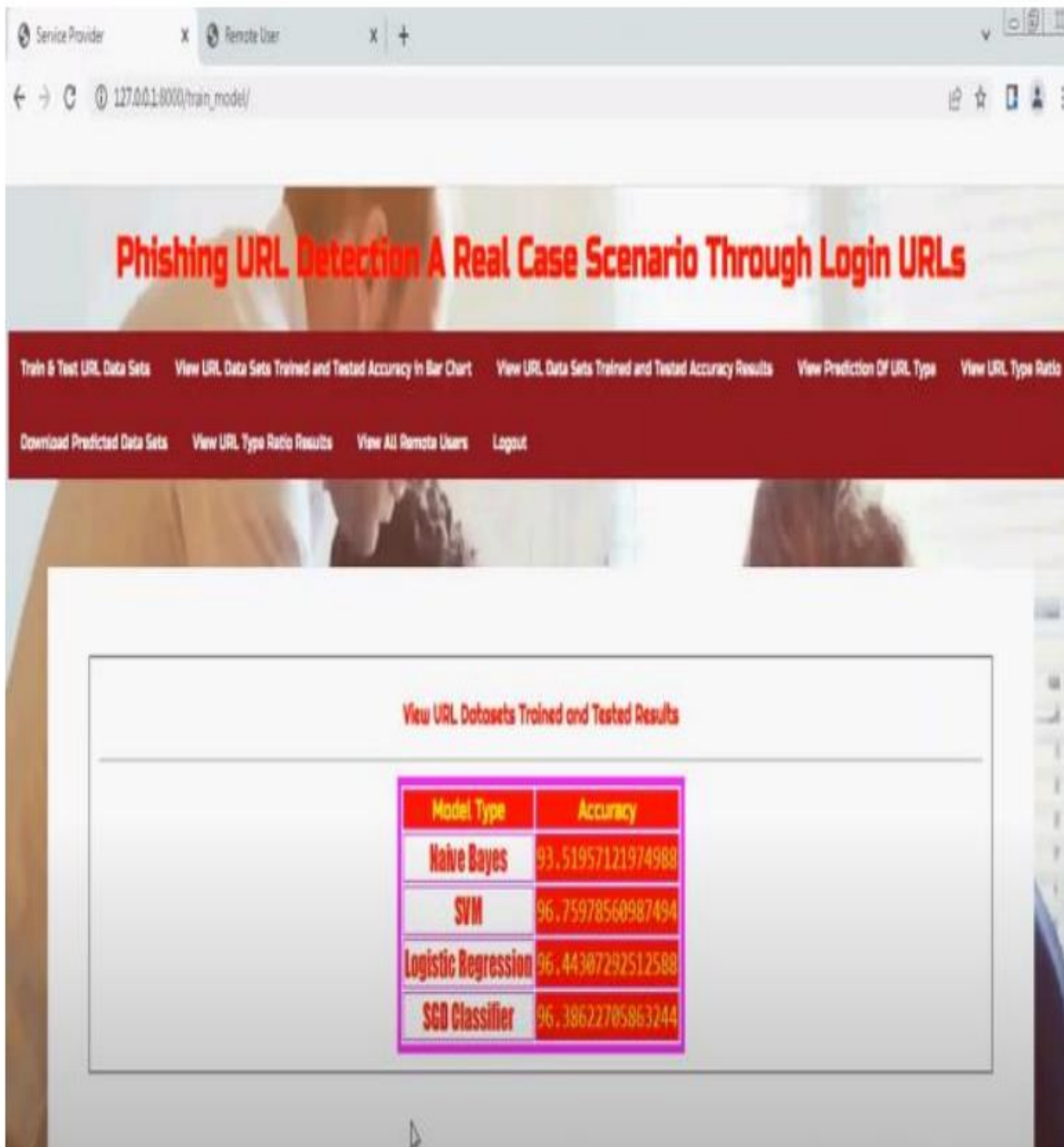


Fig.7

Output Graphs



Fig.8

Accuracy levels

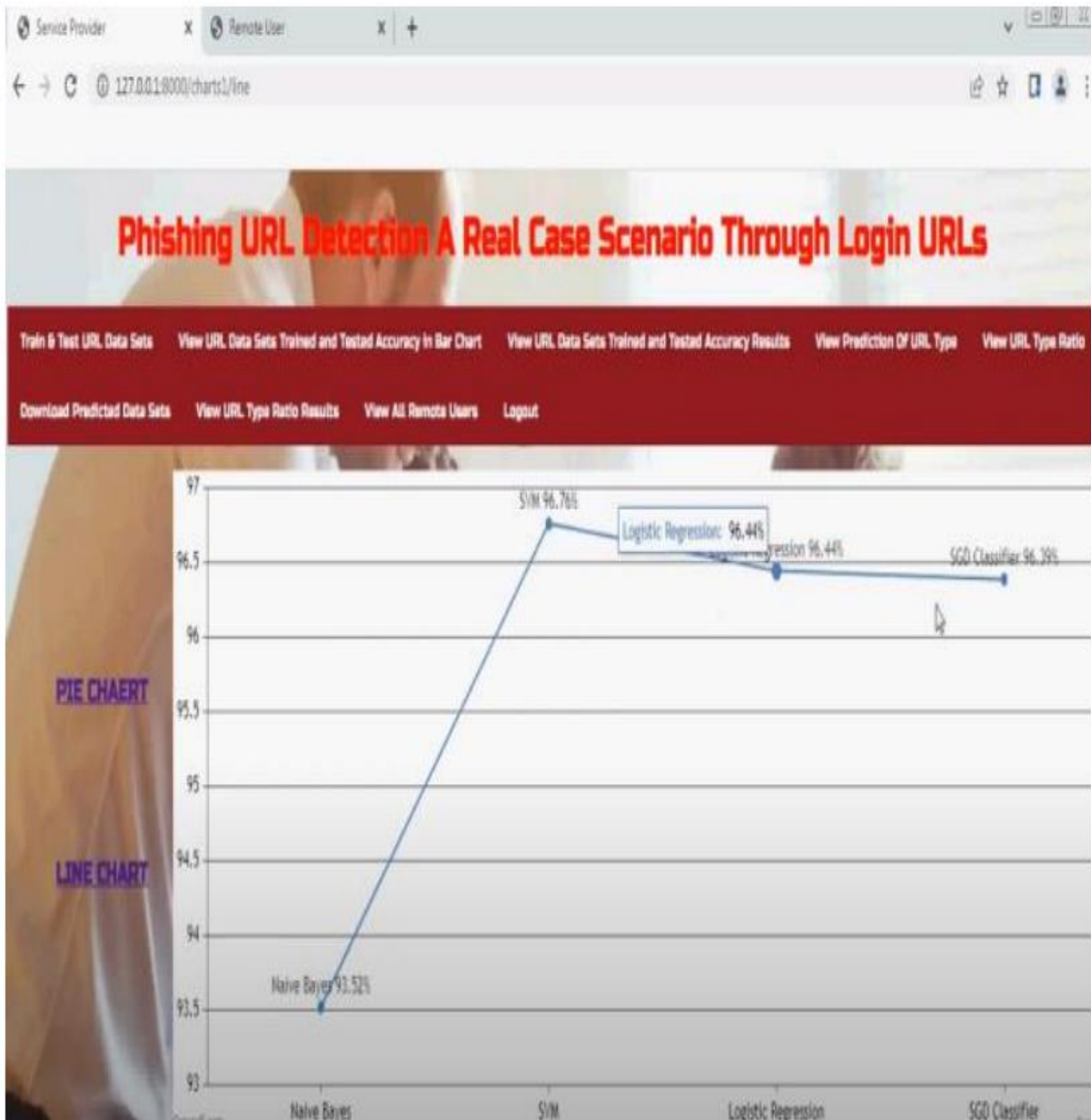
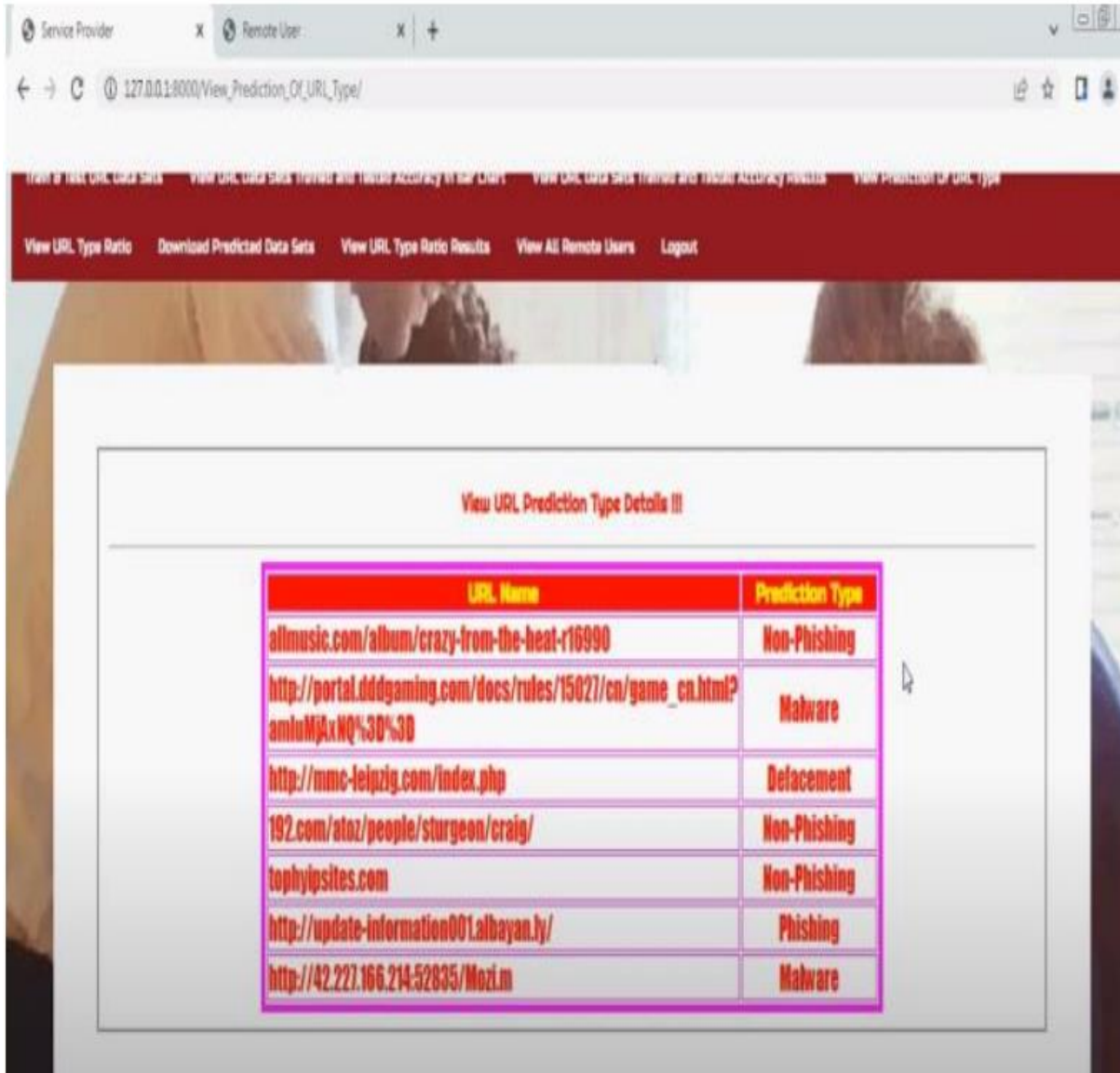


Fig.9

Phishing detected



The screenshot displays a web browser window with the address bar showing "127.0.0.1:8000/View_Prediction_Of_URL_Type/". The page features a dark red navigation bar with several menu items: "View URL Type Ratio", "Download Predicted Data Sets", "View URL Type Ratio Results", "View All Remote Users", and "Logout". Below the navigation bar, a white box titled "View URL Prediction Type Details III" contains a table with two columns: "URL Name" and "Prediction Type". The table lists seven URLs with their corresponding prediction types.

URL Name	Prediction Type
allmusic.com/album/crazy-from-the-heat-r16990	Non-Phishing
http://portal.dddgaming.com/docs/rules/15027/cn/game_cn.html?amluMjAxNQ%3D%3D	Malware
http://mmc-leipzig.com/index.php	Defacement
192.com/atoz/people/sturgeon/craig/	Non-Phishing
tophypsites.com	Non-Phishing
http://update-information001.albayan.ly/	Phishing
http://42.227.166.214:52835/Mozi.m	Malware

Fig.10

6. CONCLUSION AND FUTURE WORK

CONCLUSION

The Internet consumes almost the whole world in the upcoming age, but it is still growing rapidly. With the growth of the Internet, cybercrimes are also increasing daily using suspicious and malicious URLs, which have a significant impact on the quality of services provided by the Internet and industrial companies. Currently, privacy and confidentiality are essential issues on the internet. To breach the security phases and interrupt strong networks, attackers use phishing emails or URLs that are very easy and effective for intrusion into private or confidential networks. Phishing URLs simply act as legitimate URLs. A machine-learning-based phishing system is proposed in this study. A dataset consisting of 32 URL attributes and more than 11054 URLs was extracted from 11000+websites. This dataset was extracted from the Kaggle repository and used as a benchmark for research. This dataset has already been presented in the form of vectors used in machine learning models. Decision tree, linear regression, random forest, support vector machine, gradient boosting machine, K-Neighbor classifier, naive Bayes, and hybrid (LR+SVC+DT) with soft and hard voting were applied to perform the experiments and achieve the highest performance results. The canopy feature selection with cross fold validation and Grid search hyper parameter optimization techniques are used with LSD Ensemble model.

7. REFERENCES

1 REFERENCES

- N. Z. Harun, N. Jaffar, and P. S. J. Kassim, “Physical attributes significant in preserving the social sustainability of the traditional malay settlement,” in *Reframing the Vernacular: Politics, Semiotics, and Representation*. Springer, 2020, pp. 225–238.
 - D. M. Divakaran and A. Oest, “Phishing detection leveraging machine learning and deep learning: A review,” 2022, arXiv:2205.07411.
 - A. Akanchha, “Exploring a robust machine learning classifier for detecting phishing domains using SSL certificates,” *Fac. Comput. Sci., Dalhousie Univ., Halifax, NS, Canada, Tech. Rep. 10222/78875*, 2020.
 - H. Shahriar and S. Nimmagadda, “Network intrusion detection for TCP/IP packets with machine learning techniques,” in *Machine Intelligence and Big Data Analytics for Cybersecurity Applications*. Cham, Switzerland: Springer, 2020, pp. 231–247.
 - J. Kline, E. Oakes, and P. Barford, “A URL-based analysis of WWW structure and dynamics,” in *Proc. Netw. Traffic Meas. Anal. Conf. (TMA)*, Jun. 2019, p. 800.
-

- A. K. Murthy and Suresha, “XML URL classification based on their semantic structure orientation for web mining applications,” *Proc. Comput. Sci.*, vol. 46, pp. 143–150, Jan. 2015.
 - A. A. Ubing, S. Kamilia, A. Abdullah, N. Jhanjhi, and M. Supramaniam, “Phishing website detection: An improved accuracy through feature selection and ensemble learning,” *Int. J. Adv. Comput. Sci. Appl.*, vol. 10, no. 1, pp. 252–257, 2019.
 - A. Aggarwal, A. Rajadesingan, and P. Kumaraguru, “PhishAri: Automatic realtime phishing detection on Twitter,” in *Proc. eCrime Res. Summit*, Oct. 2012, pp. 1–12.
 - S. N. Foley, D. Gollmann, and E. Sneekenes, *Computer Security—ESORICS 2017*, vol. 10492. Oslo, Norway: Springer, Sep. 2017. P. George and P. Vinod, “Composite email features for spam identification,” in *Cyber Security*. Singapore: Springer, 2018, pp. 281–289.
-