SECURING DATA WITH BLOCK CHAIN AND MACHINE LEARNING

Yaroju Madhu Kumar, Dr. Narsappa Reddy

[1]B.tech Student, Department Of Electronics and Computer Engineering, J.B Institute of Engineering and Technology

[2]Associate Professor, Department Of Electronics and Computer Engineering, J.B Institute of Engineering and Technology

Abstract: The aim of the phishers is to acquire critical information like username, password, and bank account details. Cyber security people are now looking for trustworthy and steady detection techniques for phishing websites detection and securing user's data. This project deals with machine learning technology for detection of phishing URLs by extracting and analyzing variousphishing URLs. The aim of the project is to detect phishing URLs as well as using light GBM andSVM algorithm. This project goal is to detect and present the phishing website and its phishing percentage in visual format and the technology and tool used are Block Chain, Machine Learning(SVM & LIGHTGBM), Artificial intelligence, Html and data visualization techniques

I. INTRODUCTION

Phishing affects individuals and companies worldwide. It is difficult to track the perpetrators sinceit is carried out across the borders. The phishers are using a method, "fast-flux," which uses a large pool of proxy servers and URLs to hide the actual location of the phishing site. Hence, it is more challenging to blacklist the site as the server used requires a lot of work.

Phishing has become a widespread issue impacting both individuals and companies on a global scale. Its effects transcend borders, making it challenging to track down perpetrators. One method gaining prominence among these cybercriminals is known as "fast-flux." This technique employs a vast network of proxy servers and URLs, effectively concealing the true location of the phishing site. Consequently, efforts to blacklist these sites face increased hurdles as the constantlyshifting servers demand extensive and persistent work to identify and take down. This tactic not only complicates the process of identifying the true origin of the fraudulent activities but also amplifies the difficulty in implementing preventive measures to curb these malicious activities. Asa result, combating phishing attacks employing fast-flux techniques requires innovative and adaptive strategies to mitigate their impact on individuals and organizations globally.

Phishing, a pervasive threat impacting both individuals and businesses on a global scale, presents a formidable challenge in terms of tracking down its perpetrators. This malicious activityoperates beyond borders, making it exceedingly difficult to pinpoint the origin of these attacks. One of the sophisticated techniques employed by these cybercriminals is the utilization of "fast- flux." This method involves the use of an extensive network of proxy servers and URLs, effectivelyconcealing the true location of the phishing site. This complex infrastructure complicates the task of blacklisting these fraudulent sites. Each server utilized necessitates a substantial amount of effort to identify and block, contributing to the heightened difficulty in combating these deceitful practices. As a result, the battle against phishing becomes even more arduous, requiring innovative strategies and collaborative efforts to safeguard against its detrimental impacts.

Phishing, a pervasive threat, casts its net over individuals and companies globally, leavinga wake of challenges in its path. Its cross-border nature makes tracking perpetrators an arduous task, complicating efforts to bring them to justice. Adding to the complexity is the emergence of the "fast-flux" technique employed by these cunning phishers. This method capitalizes on a vast array of proxy servers and URLs, effectively obscuring the true location of the phishing site. As aresult, the traditional approach of blacklisting such sites becomes an uphill battle, as the constantly shifting server infrastructure demands exhaustive efforts to identify and block them effectively. This dynamic landscape poses an immense challenge for cybersecurity professionals and organizations striving to safeguard against these sophisticated and elusive threats.

II. LITERATURE SURVEY

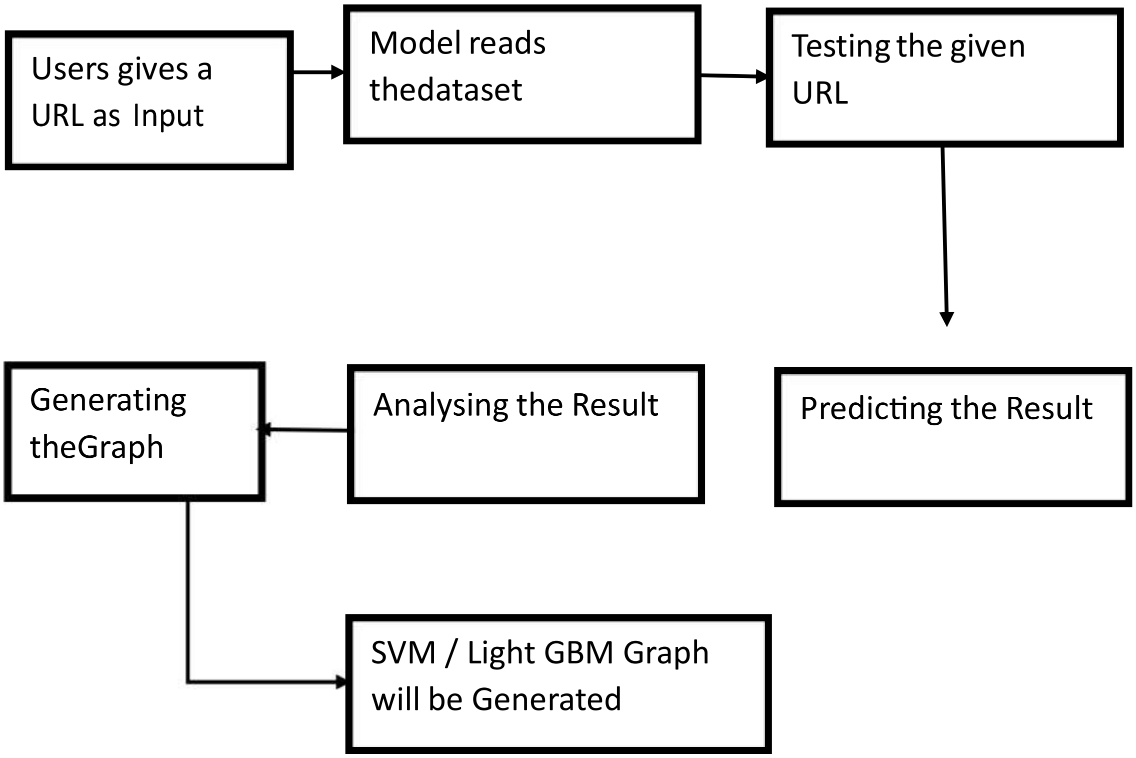
In the past decades, the operation of the internet has been increased extensively and makes our lives simple, easy and transforms our lives. It plays a major part in the areas of communication, education, business conditioning and commerce. A lot of useful data, information and data can be attained from the internet for organizational, profitable, and social development. The internet makes it easy to provide numerous services online and enables us to pierce colorful information atany time, from anywhere around the world. Phishing is the act of transferring an indistinguishable dispatch, dispatches, or vicious websites to trick the philanthropist / internet druggies into discovering delicate information similar as personal identification number (PIN) and word of bankaccount, credit card information, date of birth or social security figures. Phishing assaults affect hundreds of thousands of internet druggies across the globe. Individualizes and associations have lost a huge sum of plutocrat and confidential information through Phishing attacks. Detecting thephishing attack proves to be a challenging task. Tis attack may take a sophisticated form and fooleven the savviest users: such as substituting a few characters of the URL with alike Unicode characters. By cons, it can come in sloppy forms, as the use of an IP address instead of the domainname. Nonetheless, in the literature, several works tackled the phishing attack detection challenge while using artificial intelligence and data mining techniques [5–9] achieving some satisfying recognition rate peaking at 99.62%. However, those systems are not optimal to smartphones and other embed devices because of their complex computing and their high battery usage, since theyrequire as entry complete HTML pages or at least HTML links, tags, and webpage JavaScript elements some of those systems uses image processing to achieve the recognition. Opposite to ourrecognition system since it is a less greedy in terms of CPU and memory unlike other proposed systems as it needs only six features completely extracted from the URL as input. In this paper, after a summary of this Feld key research, we will detail the characteristics of the URL that our system uses to do the recognition. Otherwise, we will describe our recognition system, next in thepractical part we will test the proposed system while presenting the results obtained. Finally, we will enumerate the implications and advantages that our system brings as a solution to the phishing

attack**.** The purpose of phishing website detection is detecting phishing website names. Therefore,passive queries related to website names, which we want to classify as phishing or not, provide useful information to us. To prevent phishing from stealing confidential information, such as usernames, passwords, and credit card details, by way of impersonating a legitimate entity, a detection mechanism is needed. Hence, it will lead to prevention of information disclosure and property damage. The paper aims to detect phishing URLs and narrow down to the best machine learning algorithm by comparing accuracy rate, false positive and false negative rate of each algorithm.

ANALYSIS

The aim of the phishers is to acquire critical information like username, password, and bank account details. Cyber security people are now looking for trustworthy and steady detection techniques for phishing websites detection and securing user's data. This project deals with machine learning technology for detection of phishing URLs by extracting and analyzing variousphishing URLs. The aim of the project is to detect phishing URLs as well as using light GBM andSVM algorithm. This project goal is to detect and present the phishing website and its phishing percentage in visual format and the technology and tool used are Block Chain, Machine Learning (SVM & LIGHTGBM), Artificial intelligence, Html and data visualization techniques.

CONTENT DIAGRAM OF PROJECT



III. DESIGN

Designing a phishing website detection system involves a multi-layered approach that integrates various components to effectively identify and mitigate potential threats. Here's an overview of a system design for detecting phishing website. A web crawler to gather a vast array of web pages, extracting content and metadata. Extract relevant features from web pages, including URL structures, content similarity, SSL certificates, and webpage elements. Collect and maintain a database of known phishing websites and their characteristics. Train models using label data to classify websites as phishing or legitimate based on extracted features. Employ clustering algorithms to identify anomalies and patterns in website features, detecting previously unseen phishing attempts. Combine multiple models to enhance accuracy and robustness.

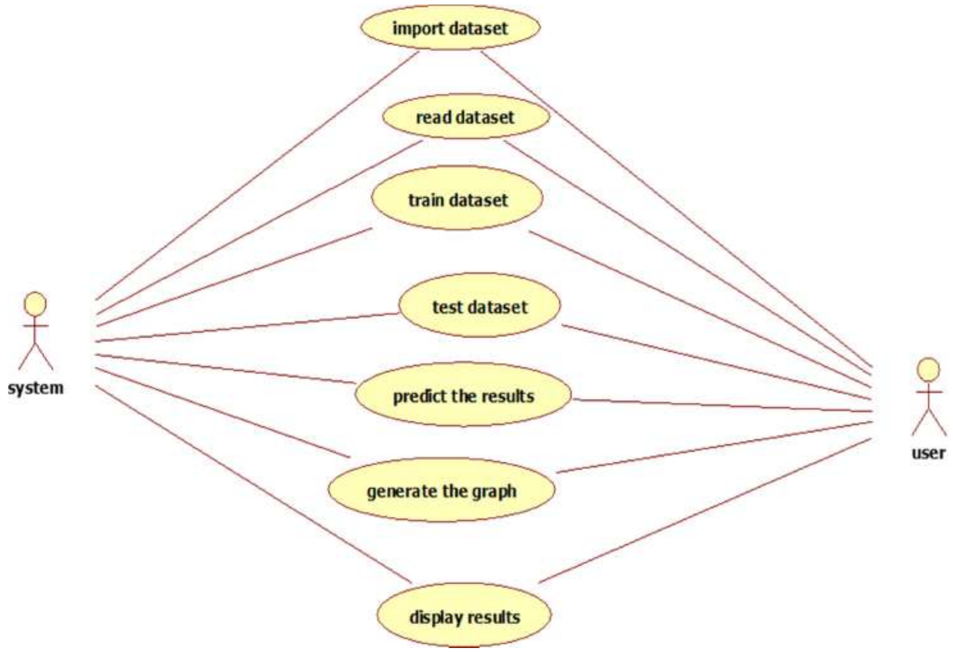
Utilize a blockchain ledger to store verified website identities, SSL certificates, and user- reported suspicious activities. Implement smart contracts for secure and automated verification processes, ensuring the integrity of data and actions taken. Enable a decentralized consensus mechanism to validate and update the ledger securely. Monitor web traffic and user interactions in real-time to identify suspicious behaviour. Analyse user behaviour patterns (clicks, form submissions) to detect deviations from normal behaviour. Use natural language processing (NLP)and image recognition to identify phishing content or mimicked interfaces.

Allow users to report suspicious websites or activities, feeding into the system for analysis and verification. Incorporate user feedback to enhance the system's learning and adaptability overtime. Trigger alerts or warnings for users visiting suspected phishing websites. Implement mechanisms to block access or redirect users away from identified phishing sites. Develop protocols for incident response, including reporting to authorities and updating security measures.Continuously update machine learning models based on new data and emerging phishing tactics. Integrate feedback loops to incorporate user-reported incidents and improve detection accuracy. In today's digital landscape, the proliferation of phishing websites poses a significant threat to online security.

DFD OR UML DIAGRAMS

Use Case Diagram

A use case diagram represents the interactions between users (actors) and the system. For a phishing website detection system.

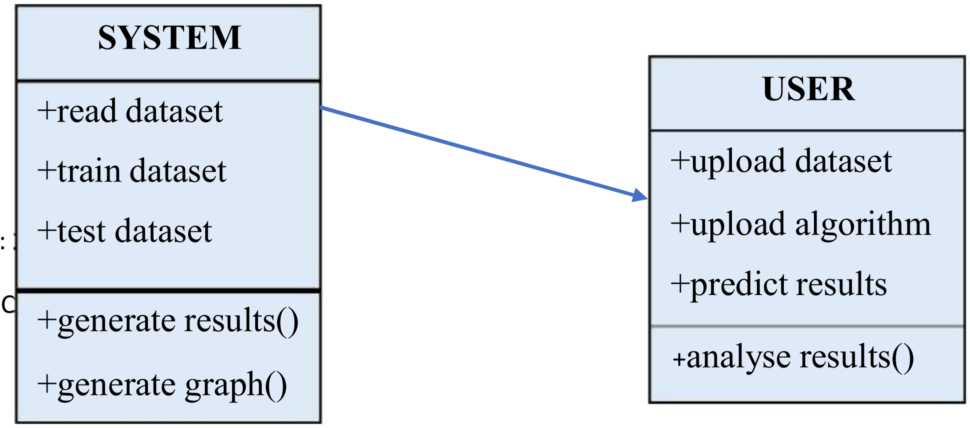


use case diagram

This diagram provides an overview of the interactions between users and the system components involved in detecting phishing websites. It illustrates how users interact with the system functionalities to report suspicious sites

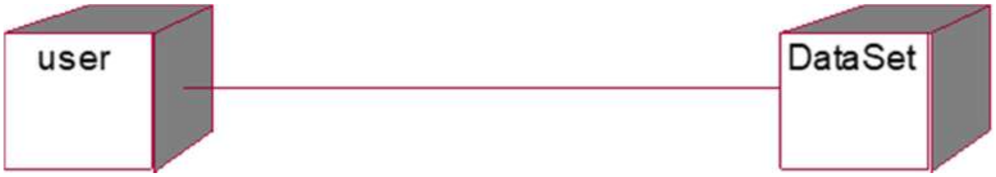
Class Diagram

Creating a class diagram for detecting phishing websites involves identifying the key classes,their attributes, and their relationships.



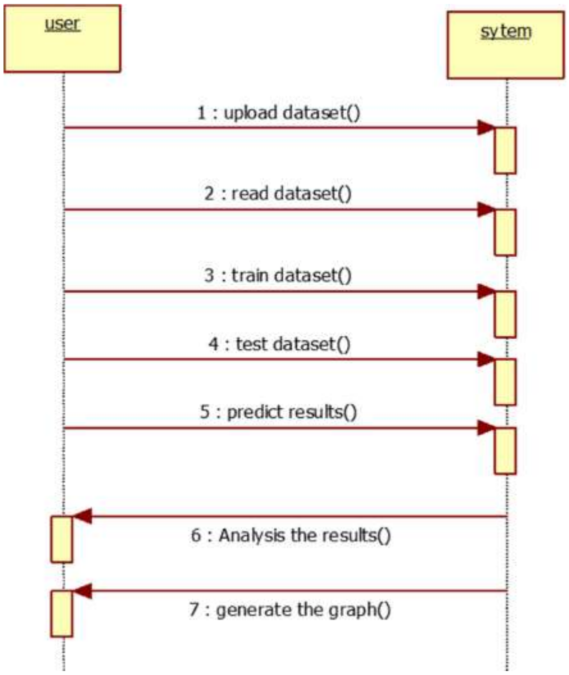
Deployment Diagram:

A deployment diagram in UML showcases the physical deployment of artifacts on nodes(hardware or software elements).

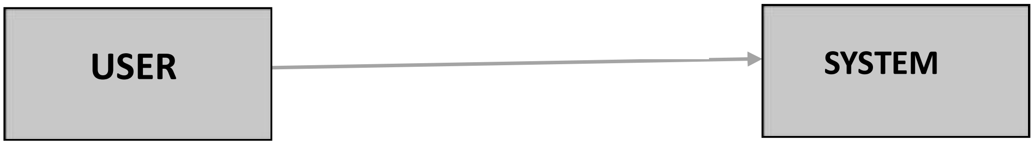


Sequence Diagram:

A sequence diagram in UML illustrates the interactions between different components or objectsin a particular sequence.

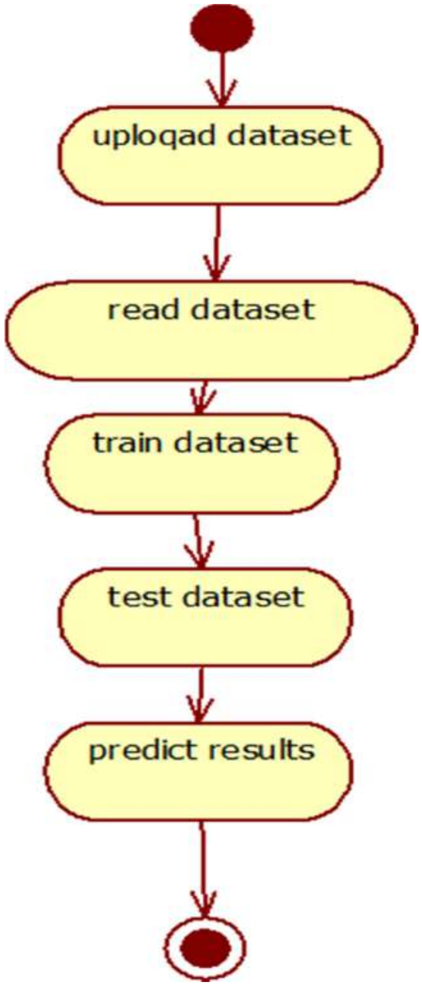


Collaboration Diagram



Activity Diagram

This simplified activity diagram outlines the sequential flow of actions involved in checking a URL for phishing by validating its format, checking against a blacklist, and taking appropriate actions based on the result. Depending on the complexity of the detection process, this diagram can be expanded to include more detailed steps or decision points.

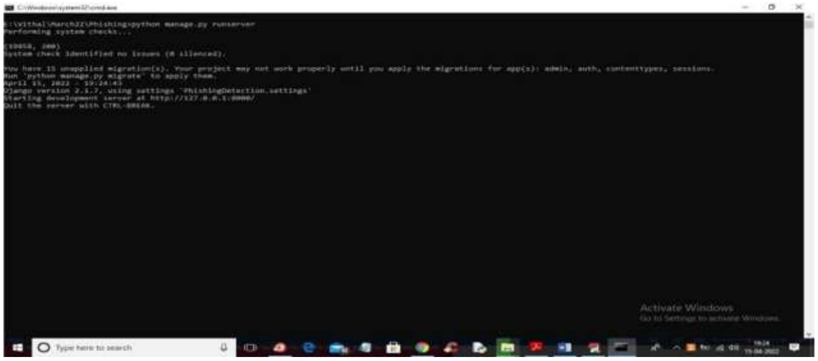


IV. IMPLEMENTATION AND RESULTS

This chapter tells us about the implementation part of the website. This section deals with the briefintroduction about the important functions used to create the securing data with blockchain and machine learning model. It consists of various source codes used in building this web page and user dashboards. Also lists out the outputs of each section which makes it clear about the differentoptions available to complete the quiz successful

OUTPUT SCREENS

Login Page



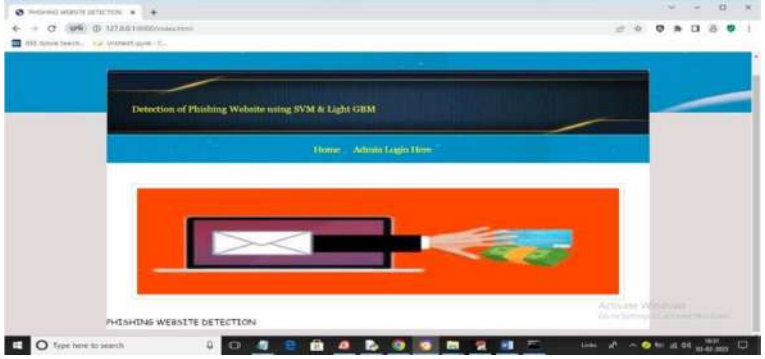
```
E:\vithal\March22\Phishing>python manage.py runserver
Performing system checks...

(10858, 200)
System check identified no issues (0 silenced).

You have 15 unapplied migration(s). Your project may not work properly until you apply the migrations for app(s): admin, auth, contenttypes, sessions.
Run 'python manage.py migrate' to apply them.
April 15, 2022 - 19:24:43
Django version 2.1.7, using settings 'PhishingDetection.settings'
Starting development server at http://127.0.0.1:8000/
Quit the server with CTRL-BREAK.
```

Fig 5.4.1

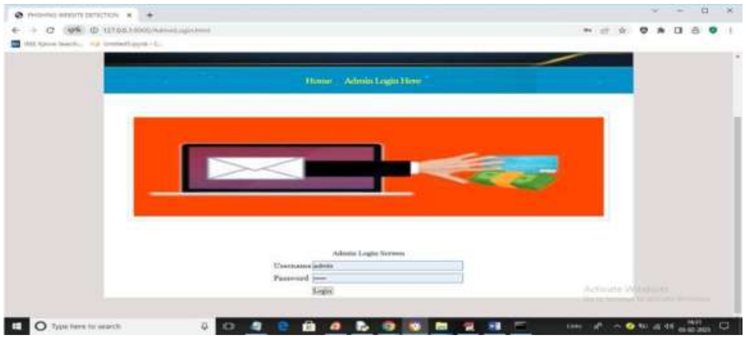
In above screen DJANGO webserver started and now open browser and enter URL http://127.0.0.1:8000/index.html and press enter key to get below output



PHISHING WEBSITE DETECTION
127.0.0.1:8000/index.html
Detection of Phishing Website using SVM & Light GBM
Home
Admin Login Here
PHISHING WEBSITE DETECTION

Fig 5.4.2

In above screen click on ‘Admin Login Here’ link to get below login screen



Admin Login Screen

Username admin

Password •••••

Fig 5.4.3

In above screen enter username and password as ‘admin’ and ‘admin’ and then press button to getbelow output

SVM Algorithm Screen

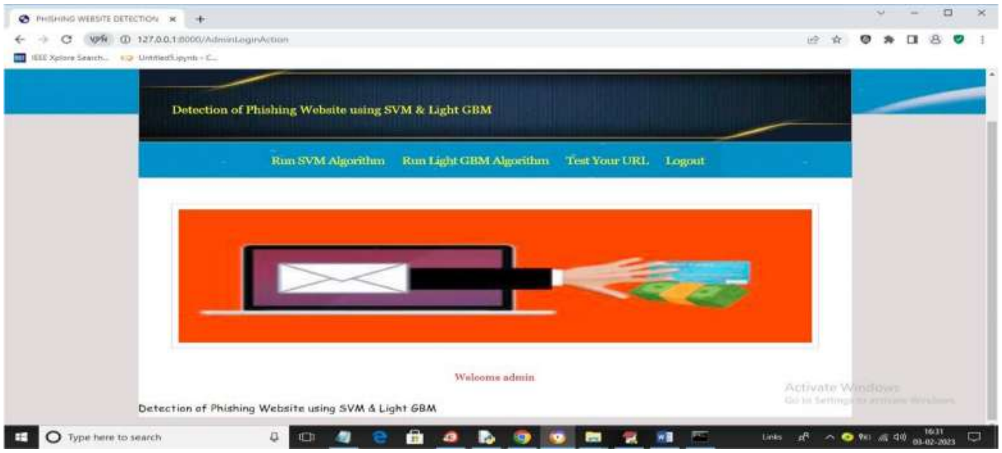

PHISHING WEBSITE DETECTION
127.0.0.1:8000/AdminLoginAction
Detection of Phishing Website using SVM & Light GBM
Run SVM Algorithm
Run Light GBM Algorithm
Test Your URL
Logout
Welcome admin
Detection of Phishing Website using SVM & Light GBM
Type here to search

Fig 5.4.4

In above screen click on ‘Run SVM Algorithm’ link to train SVM algorithm and get below output

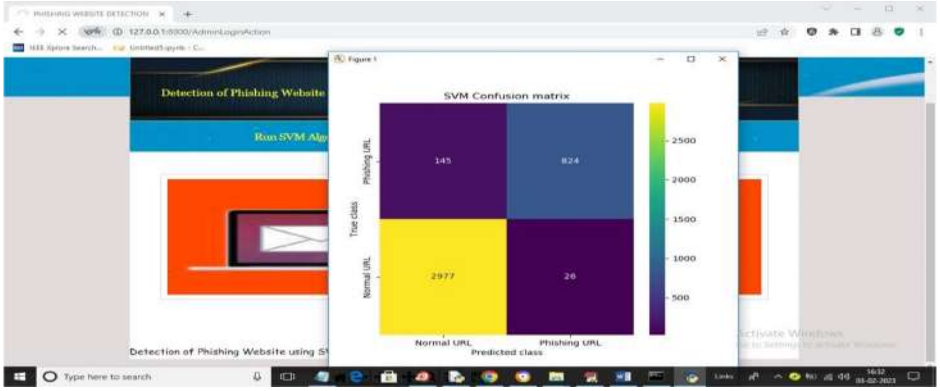
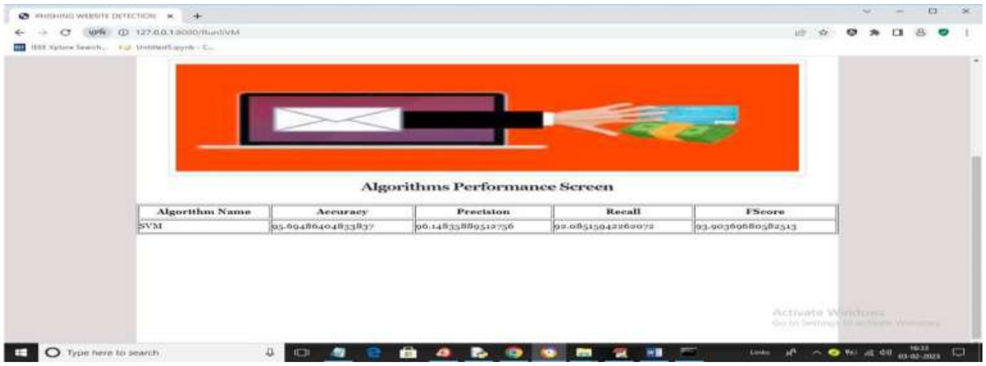


Fig 5.4.5

In above screen we can see SVM confusion matrix where x-axis represents predicted class and y-axis represents TRUE class and we can see SVM predict 2977 records correctly



Algorithms Performance Screen

Algorithm Name	Accuracy	Precision	Recall	FScore
SVM	95.69486404833837	96.14835889512756	92.08515942260072	93.90369680582513

Fig 5.4.6

Testing Screen

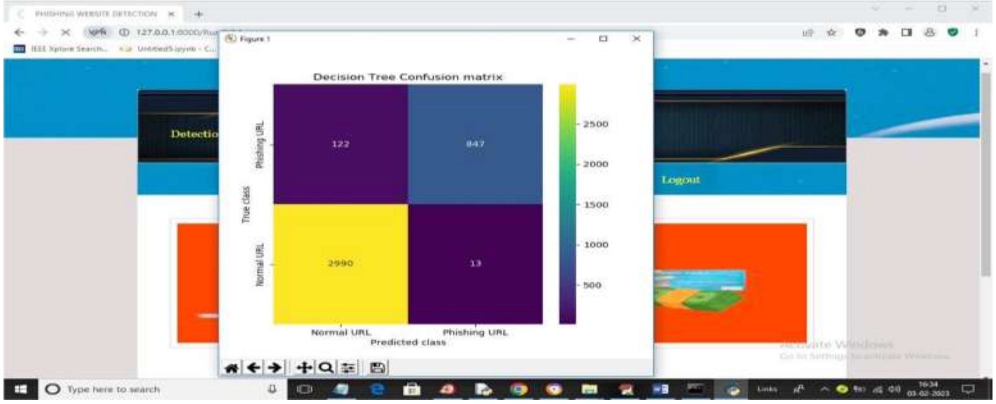
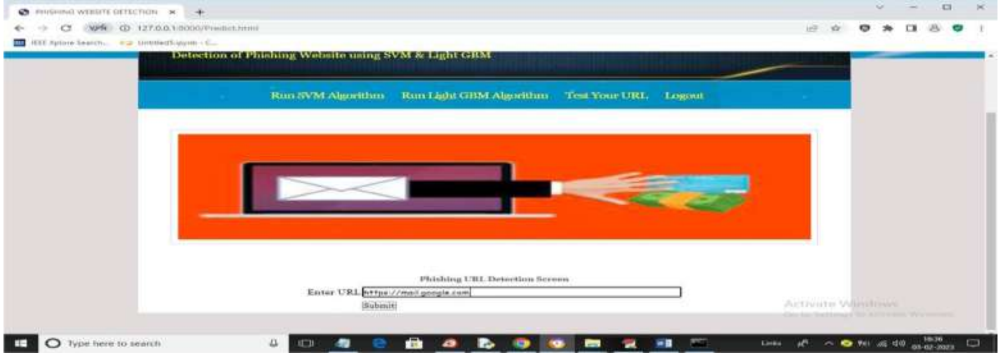


Fig 5.4.7

In above screen we can see Decision Tree confusion matrix graph and now close above graph toget below output

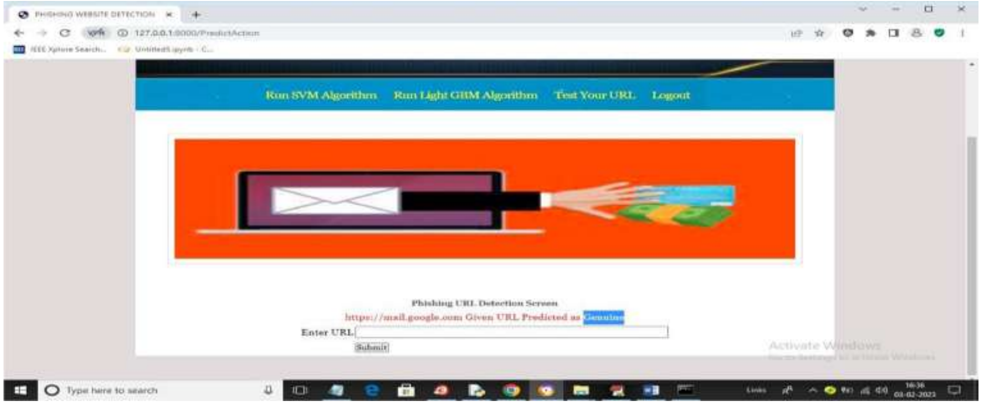

PHISHING WEBSITE DETECTION
127.0.0.1:8000/Predict.html
Detection of Phishing Website using SVM & Light GBM
Run SVM Algorithm
Run Light GBM Algorithm
Test Your URL
Logout
Submit
Type here to search
16:36
03-02-2023

Phishing URL Detection Screen

Enter URL https://mail.google.com

Fig 5.4.8

In above screen I entered URL as https://mail.google.com and then press button to get below output



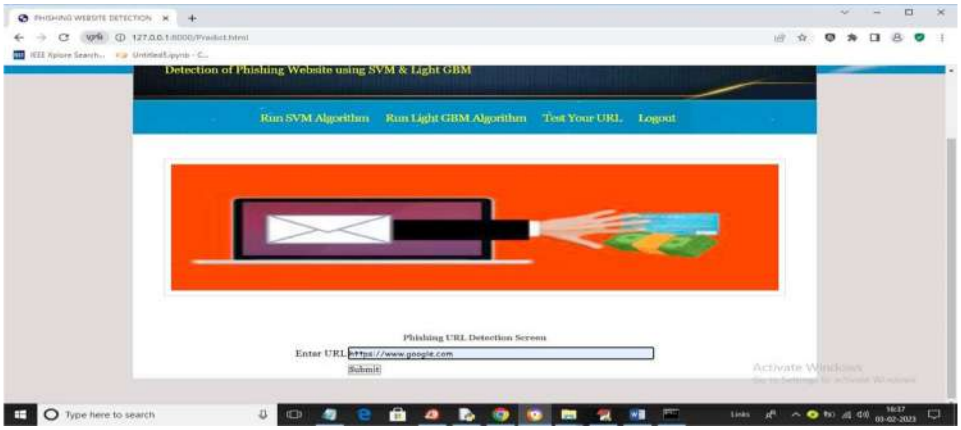
Phishing URL Detection Screen

https://mail.google.com Given URL Predicted as Genuine

Enter URL

Fig 5.4.9

In above screen in blue colour text we can see given URL predicted as GENUINE (normal) andnow test other URL. Similarly now I will enter Google.com in below screen

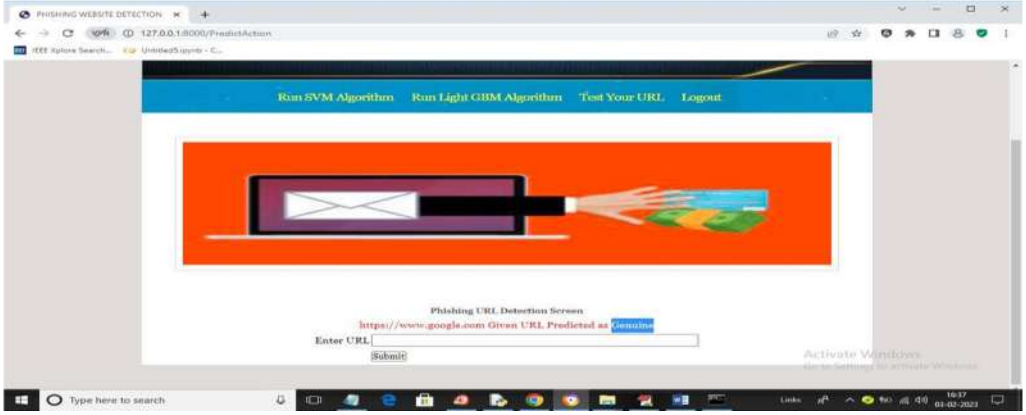


Phishing URL Detection Screen

Enter URL https://www.google.com

Fig 5.4.10

In above screen I gave URL as Google.com and below is the output



Phishing URL Detection Screen

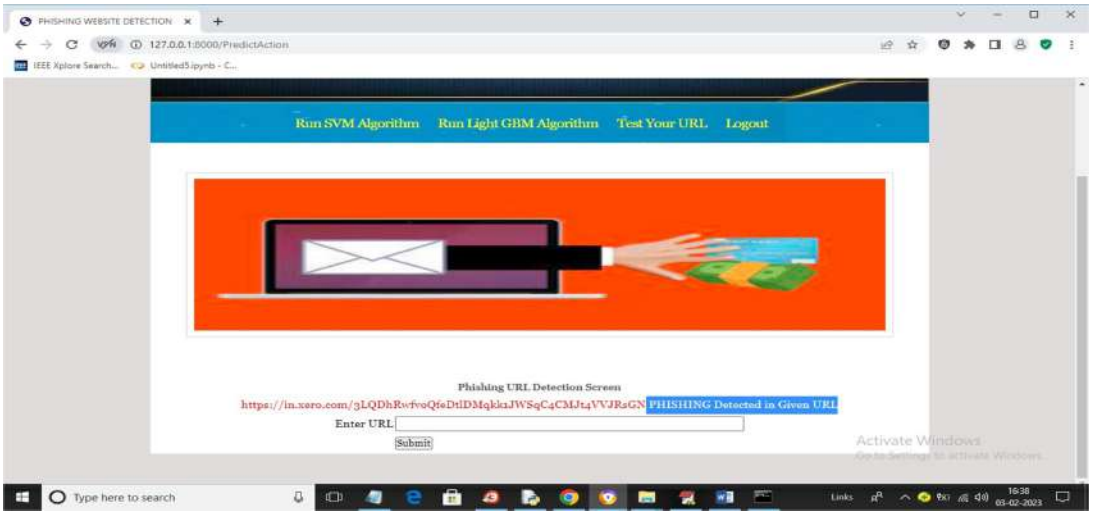
https://www.google.com Given URL Predicted as Genuine

Enter URL

Fig 5.4.11

In above screen Google.com also predicted as Genuine. Now in below screen from internet I am taking one phishing URL and then input to my application to get prediction. In above screen bluecolour URL is the phishing URL and I will input that to my application in below screen and belowis the phishing URL from internet
‘https://in.xero.com/3LQDhRwfvoQfeDtlDMqkk1JWSqC4CMJt4VVJRsGN’

In above screen I entered same URL and press button to get below output



Phishing URL Detection Screen

https://in.xero.com/3LQDhRwfvoQfeDtlDMqkk1JWSqC4CMJt4VVJRsGN PHISHING Detected in Given URL

Enter URL

Submit

Fig 5.4.11

In above screen in blue colour text we can see application detected PHISHING in given URL andsimilarly you can enter any URL and detect it as NORMAL or phishing.

Result Analysis

This chapter provides information about the website's implementation phase. This section provides a succinct overview of the key features that were used to develop the URL checking application. It is made up of numerous source codes that were used to create this website. Additionally providesthe results of each area, which clarifies the various possibilities available to correctly finish project.

V. Conclusion:

This project aims to enhance detection methods to detect phishing websites using machine learningtechnology. The proposed study emphasized the phishing technique in the context of classification, where phishing website is considered to involve automatic categorization of websites into a predetermined set of class values based on several features and the class variable The ML based phishing techniques depend on website functionalities to gather information that can help classifywebsites for detecting phishing sites. The problem of phishing cannot be eradicated nonetheless can be reduced by combating it in two ways, improving targeted anti-phishing procedures and techniques and informing the public on how fraudulent phishing websites can be detected and identified. To combat the ever evolving and complexity of phishing attacks and tactics, ML anti- phishing techniques are essential. We achieved 97.14% detection accuracy using SVM and Light GBM algorithms with lowest false positive rate.

REFERENCES

[1] Sophiya Shikalgar, Mrs. Swati Narwane (2019), Detecting of URL based Phishing Attackusing Machine Learning. (vol. 8 Issue 11, November – 2019)

[2] Rashmi Karnik, Dr. Gayathri M Bhandari, Support Vector Machine Based Malware andPhishing Website Detection.

[3] Arun Kulkarni, Leonard L. Brown , III2, Phishing Websites Detection using Machine Learning (vol. 10, No. 7,2019)

[4] R. Kiruthiga, D. Akila, Phishing Websites Detection using Machine Learning.

[5] Ademola Philip Abidoye, Boniface Kabaso, Hybrid Machine Learning: A Tool to detect Phishing Attacks in Communication Networks. (vol. 11 No. 6,2020)

[6] Andrei Butnaru, Alexios Mylonas and Nikolaos Pitropakis, Article Towards Lightweight URL-Based Phishing Detection.13 June 2021

[7] Ashit Kumar Dutta (2021), Detecting phishing websites using machine learning technique.Oct 11, 2021

[8] Nguyet Quang Do, Ali Selamat, Ondrej Krejcar, Takeru Yokoi and Hamido Fujita (2021)Phishing Webpage Classification via Deep Learning-Based Algorithms: An Empirical study.

[9] Ammara Zamir, Hikmat Ullah Khan and Tassawar Iqbal, Phishing website detection usingdiverse machine learning algorithms.

[10] Valid Shahrivari, Mohammad Mahdi Darabi and Mohammad Izadi (2020), Phishing Detection Using Machine Learning Techniques.

[11] Orunsolu, A. S. Sodiya and A.T. Akinwale (2019), A predictive model for phishing detection.

[12] Wong, R. K. K. (2019). An Empirical Study on Performance Server Analysis and URL Phishing Prevention to Improve System Management Through Machine Learning. In Economics of Grids, Clouds, Systems, and Services: 15th International Conference, GECON 2018, Pisa, Italy, September 18-20, 2018, Proceedings (Vol. 11113, p. 199). Springer.