**TEAM ID : PNT2022TMID51470**

**TEAM MEMBERS : PAUL GALVIN P (960619104301)**

**SUGADEV M (960619104040)**

**STEEN ROHAN J (960619104037)**

**BOOBATHY RAM A (960619104012)**

**VIGNESH S (960619104302)**

**DEPARTMENT : CSE**

**COLLEGE NAME : DMI ENGINEERING COLLEGE,**

**ARALVAIMOZHI**.

**WEB PHISHING DETECTION**

**USING APPLIED DATA SCIENCE**

**ABSTRACT**

A web service is one of the most important Internet communications software services. Using fraudulent methods to get personal information is becoming increasingly widespread these days. However, it makes our lives easier, it leads to numerous security vulnerabilities to the Internet's private structure. Web phishing is just one of the many security risks that web services face. Phishing assaults are usually detected by experienced users however, security is a primary concern for system users who are unaware of such situations. Phishing is the act of portraying malicious web runners as genuine web runners to obtain sensitive information from the end-user. Phishing is currently regarded as one of the most dangerous threats to web security. Vicious Web sites significantly encourage Internet criminal activity and inhibit the growth of Web services. As a result, there has been a tremendous push to build a comprehensive solution to prevent users from accessing such websites. We suggest a literacy-based strategy to categorize Web sites into three categories: benign, spam, and malicious. Our technology merely examines the Uniform Resource Locator (URL) itself, not the content of Web pages. As a result, it removes run-time stillness and the risk of drug users being exposed to cyber surfer-based vulnerabilities. When compared to a blacklisting service, our approach performs better on generality and content since it uses learning techniques. The criminals, who want to obtain sensitive data, first create unauthorized replicas of a real website and e-mail. The e-mail will be created using logos and slogans of a legitimate company. The nature of website creation is one of the reasons that the Internet has grown so rapidly as a communication medium. Phisher then send the "spoofed" emails to as many people as possible in an attempt to lure them into the scheme. When these emails are opened or when a link in the mail is clicked, the consumers are redirected to a spoofed website, appearing to be from the legitimate entity. We discuss the methods used for detection of phishing Web sites based on url importance properties.

**TABLE OF CONTENTS**

**CHAPTER NO TITLE PAGE NO**

**1 ABSTRACT 1**

1.1 Project Overview

1.2 Purpose

**2 LITERATURE SURVEY 8**

2.1 Existing problem

2.2 References

2.3 Problem Statement Definition

**3 IDEATION & PROPOSED SOLUTION 10**

3.1 Empathy Map Canvas

3.2 Ideation & Brainstorming

3.3 Proposed Solution

3.4 Problem Solution fit

**4 REQUIREMENT ANALYSIS 13**

4.1 Functional requirement

4.2 Non-Functional requirements

**5 PROJECT DESIGN 14**

5.1 Data Flow Diagrams

5.2 Solution & Technical Architecture

5.3 User Stories

**6 PROJECT PLANNING & SCHEDULING 18**

6.1 Sprint Planning & Estimation

6.2 Sprint Delivery Schedule

6.3 Reports from JIRA

**7 CODING & SOLUTIONING 20**

7.1 Feature 1

7.2 Feature 2

**8 TESTING 23**

8.1 Test Cases

8.2 User Acceptance Testing

**9 RESULTS 28**

9.1 Performance Metrics

9.2 Outputs

**10.ADVANTAGES & DISADVANTAGES 30**

**11.CONCLUSION 31**

**12.FUTURE SCOPE 40**

**13.FUTURE SCOPE 76**

(i) Source code

(ii) Github link

(II)Project demo link

**1. INTRODUCTION**

**1.1 PROJECT OVERVIEW**

Internet constitutes a tremendous change in today's world because of its versatility. By utilizing the sophisticated infrastructure of the internet, people can do transactions such as shopping, banking etc. whenever and wherever they want. The internet has many advantages, at the same time; it also has its own set of security and privacy problems. By using the anonymous and independent infrastructure of the internet, attackers create a prominent platform for cyberattacks, such as phishing, malware distribution, privacy disclosure etc., which causes severe threats to the end-users of the internet.

Phishing is the most widespread and pernicious cyber attack , which mostly targets the human rather than the computer by exploiting their vulnerabilities. According to Anti-Phishing Working Group (APWG), phishing is a criminal mechanism employing both social engineering and technical tricks to steal user's identity data and financial account credentials by disguising as a trusted one. To lure the end-users, attackers use malicious websites and e-mails by posing themselves as a trusted one. The supreme goal of phishing is to abduct confidential data such as user name, password, bank details, credit card details etc. Attackers perform phishing for many reasons – to gain benefits financially, to steal personal information, to ruin the reputation of the organizations and sometimes just to get fame .

The term ‘phishing’ is coined in the mid-1990s and is from the term ‘fishing’ because it involves trying to outwit someone into a trap. The history of phishing scams can be traced back to the beginning of the 1990s via America Online (AOL), which was carried out by generating random credit card numbers to make fake AOL accounts . Later, attackers turned it into a million-dollar growth business, by impersonating many organizations such as banks, credit card companies, online payment service providers (e.g. InstaReM), and social media websites (e.g. Face book). Even Internet giants such as Google and Face book were scammed out of more than $100 million between 2013 and 2015 through email phishing . Recently in 2020, Texas school district lost $2.3 million in phishing raid and the Federal Bureau of Investigation is currently investigating on it.

**1.2 PURPOSE**

The main purpose of the project is to detect the fake or phishing websites who are trying to get access to the sensitive data or by creating the fake websites and trying to get access of the user personal credentials. We are using machine learning algorithms to safeguard the sensitive data and to detect the phishing websites who are trying to gain access on sensitive data.

**2. LITERATURE SURVEY**

The purpose or goal behind phishing is data, money or personal information stealing through the fake website. The best strategy for avoiding the contact with the phishing web site is to detect real time malicious URL. Phishing websites can be determined on the basis of their domains. They usually are related to URL which needs to be registered (low-level domain and upper-level domain, path, query). Recently acquired status of intra-URL relationship is used to evaluate it using distinctive properties extracted from words that compose a URL based on query data from various search engines such as Google and Yahoo. These properties are further led to the machine-learning based classification for the identification of phishing URLs from a real dataset. This paper focus on real time URL phishing against phishing content by using phish - STORM. For this a few relationship between the register domain rest of the URL are consider also intra URL relentless is consider which help to dusting wish between phishing or non phishing URL. For detecting a phishing website certain typical blacklisted url’s are used, but this technique is unproductive as the duration of phishing websites is very short. Phishing is the name of avenue. It can be defined as the manner of deception of an organization's customer to communicate with their confidential information in an unacceptable behavior. It can also be defined as intentionally using harsh weapons such as Spasm to automatically target the victims and targeting their private information. As many of the failures being occurred in the SMTP are exploiting vectors for the phishing websites, there is a greater availability of communication for malicious message deliveries.

**2.1 EXISTING PROBLEM**

The existing system uses the Classifiers, Fusion Algorithm, and Bayesian Model to detect the phishing sites. The classifiers can classify the text content and image content. Text classifier is to classify the text content and Image classifier is to classify the image content. Bayesian model estimates the threshold value. Fusion Algorithm combines the both classifier results and decides whether the site is phishing or not. The performance of different classifiers based on correct classification ratio, F-score, Matthews’s correlation coefficient, False negative ratio, and False alarm ratio. The threshold value will be decided by the developer only. This leads to the problems like false positive and false negative. False positive means, the probability of being a phishing webpage is greater than the threshold value but that webpage is not a phishing webpage. False negative means, the probability of being a phishing webpage is less than the threshold value but that webpage is a phishing webpage. This results the reduction in security levels. The existing system handles the only one kind of phishing attacks.

**2.2 REFERENCES**

**AKANBI, O.A., AMIRI, I.S., FAZELDEHKORDI,E: ‘A MACHINE LEARNIN APPROACH TO PHISHING DETECTION AND DEFENSE’ (SYNGRESS, 2014, 1ST**

**EDITION)**. Phishing is the most widespread and pernicious cyber attack , which mostly targets the human rather than the computer by exploiting their vulnerabilities. According to AntiPhishing Working Group (APWG), phishing is a criminal mechanism employing both social engineering and technical tricks to steal user's identity data and financial account credentials by disguising as a trusted one.

**GOEL, D., JAIN, A.K.: ‘MOBILE PHISHING ATTACKS AND DEFENCE MECHANISMS: STATEOF ART AND OPEN RESEARCH CHALLENGES’, COMPUT.**

**SEC., 2018**. To lure the end-users, attackers use malicious websites and e-mails by posing themselves as a trusted one. The supreme goal of phishing is to abduct confidential data such as username, password, bank details, credit card details etc. Attackers perform phishing for many reasons – to gain benefits financially, to steal personal information, to ruin the reputation of the organizations and sometimes just to get fame.

**BASE PAPER**

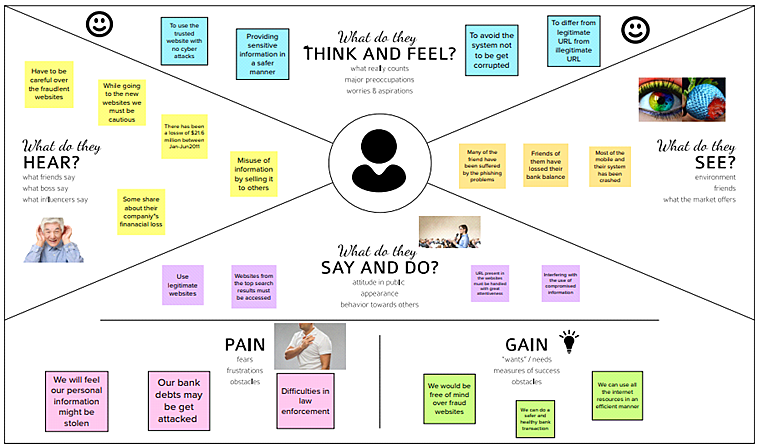
**JAIN, A.K., GUPTA, B.B.: ‘A MACHINE LEARNING BASED APPROACH FOR PHISHINGDETECTION USING HYPERLINKS INFORMATION’, J. AMBIENT**

**INTELL. HUMANIZ. COMPUT,2019**. Jain and Gupta proposed a novel web phishing detection approach by extracting hyperlinks of the web pages. The proposed approach has extracted 12 specific hyperlink features such as total hyperlink feature, no hyperlink feature, internal hyperlinks, external hyperlinks, null hyperlink, internal CSS, external CSS, internal redirection, external redirection, internal error, external error, login form link, internal favicon, and external favicon. The extracted features are then fed into ML algorithms such as naïve Bayes, random forest, SVM, Ada boost, neural network, C4.5, and logistic regression. The performance of all the ML algorithms was measured and reported**.**

**2.3 PROBLEM STATEMENT DEFINITION**

Phishing is one of the techniques which are used by the intruders to get access to the user credentials or to gain access to the sensitive data. This type of accessing the is done by creating the replica of the websites which looks same as the original websites which we use on our daily basis but when a user click on the link he will see the website and think its original and try to provide his credentials . To overcome this problem we are using some of the machine learning algorithms in which it will help us to identify the phishing websites based on the features present in the algorithm. By using these algorithm we cam be able to keep the user personal credentials or the sensitive data safe from the intruders.

**3. IDEATION & PROPOSED SOLUTION**



**3.3 PROPOSED SOLUTION**

Fields such as domain, sub domain, Top Level Domain (TLD), protocol, directory, file name, path and query allow creating different URL addresses. These related fields in the phishing URLs are generally different from the legitimate ones on websites. Therefore, URLs have an important place in detecting phishing attacks especially for classifying the web page quickly. It was observed in the literature review that effective features obtained from the URL increase the accuracy of the classification. Additionally, third-party service usage, site layout, CSS, content, meta information, etc. features can also improve accuracy. However, these features will cause an increase in the classification time of the new websites which needed to be classified. The proposed model, trained only with the features obtained from the URL, is expected to classify in a shorter time than other models. Considering this information, only URL analysis is planned in the study. Thus, the classification results of the obtained features in different algorithms in machine learning are compared. In addition, the results from another study with the same dataset are compared with those of the current study.

**NOVELTY**

Attackers take advantage from published phishing counter measure to modify their websites and bypass detection systems. We propose 27 novel features with relevant information to achieve high performance for today’s phishing detection tasks. We include a novel type of feature, web technology analysis.

**SOCIAL IMPACTS**

Stolen data can have many uses. Master card information will be accustomed purchase goods and services, ATM card information may well be accustomed duplicate ATM card sand use them for withdrawal of money. Account be accustomed steal information or to be ready to act as another user online**.**

**FEASIBILITY**

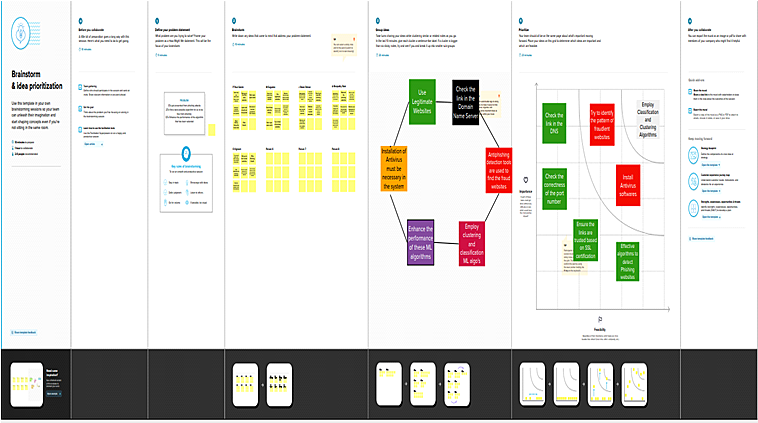
This is the first attempt to systematically understand the threat posed by the ease of correlating user information across caller ID lookup application (True caller), and social networking application (Facebook). This was executed using phone num-as unique identifiers. We show the attack is feasible with easily available computational resources, and poses a significant security and privacy threat. An attacker can use these cross-application features to launch highly targeted attacks on multiple channels like OTT applications, voice, e-mail, or SMS

**SCALABILITY**

Indian phone numbers that we enumerated, it is possible to launch social and spear phishing attacks against 51,409 and 180,000 users respectively. Phishing attacks can be launched against 722,696 users. We also found 91,487 highly influential victims who can be attacked by crafting whaling attacks against them.

**BUSINESS MODEL**

The main reason is the lack of awareness of users. But security defenders must take precautions to prevent users from confronting these harmful sites. A business is aware that a spoof site has launched, the next step is to alert customers to ensure they don’t visit the fake website and enter credentials. But organizations still need to provide a countermeasure in case customers aren't notified in time. As deception technology matures, defenders have new ways to foil phishing, even if hackers have managed to gather victim credentials. One approach involves injecting the spoof site with decoy credentials. Decoys are highly convincing fake credentials that lessen the value of stolen credentials to the point where the attacker is unsure if they’ve taken anything they can use. An email-focused strategy fails to extend protection to an organization’s customers. All one needs to do is look at phishing attacks against British Airways or American Express to see that consumers are an extremely vulnerable group. An email-centric strategy can't protect customers because they’re all using different email providers that a company cannot monitor, much less control. It’s also not possible for a large enterprise to train every one of its customers to recognize a phishing attack.



**3.4 PROPOSED SOLUTION FIT**

The overview of the proposed solution to detect phishing attacks is the data source contained URLs and HTML codes of web pages. The URLs are directly used as inputs to the model with a minimum pre-processing, and that is separately discussed in a below subsection. However, HTML features need to be extracted from the web pages. Therefore, a feature extraction model is used for the extraction before finalizing the model input features. After extracting the relevant features from the web pages, HTML features, and URLs concatenate to have input feature vectors for the detection model. Finally, the detection model will use the input feature vector and produce an output as legitimate or phishing. However, the detection model is a combination of two deep networks. It can analyze URLs and HTML features separately and combine both decisions in making the final output of the model. The major components included in the solution, namely, a feature extraction model and detection model, are introduced in the following

**4. REQUIREMENT ANALYSIS**

**4.1 FUNCTIONAL REQUIREMENT**

A function of software system is defined in functional requirement and the behavior of the system is evaluated when presented with specific inputs or conditions which may include calculations, data manipulation and processing and other specific functionality.

**HARDWARE REQUIREMENTS:**

1. Processor : Any Processor above 500 MHz
2. RAM : 8 GB
3. Hard Disk : 1 TB
4. Input device : Standard keyboard and mouse
5. Internet Connectivity

**4.2 NON-FUNCTIONAL REQUIREMENTS**

Nonfunctional requirements describe how a system must behave and establish constraints of its functionality. This type of requirements is also known as the system’s quality attributes. Attributes such as performance, security, usability, compatibility are not the feature of the system, they are a required characteristic. They are "developing" properties that emerge from the whole arrangement and hence we can't compose a particular line of code to execute them. Any attributes required by the customer are described by the specification. We must include only those requirements that are appropriate for our project.

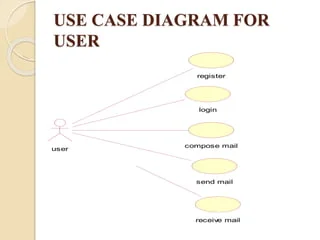
**SOFTWARE REQUIREMENTS** :r

1. OS : Windows 11
2. Platform : Jupyter Notebook,Microsoft Visual Studio
3. Language : Python ,HTML,flask
4. IDE/tool : Anaconda

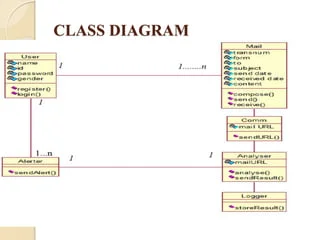
**5. PROJECT DESIGN**

**5.1 DATA FLOW DIAGRAMS**

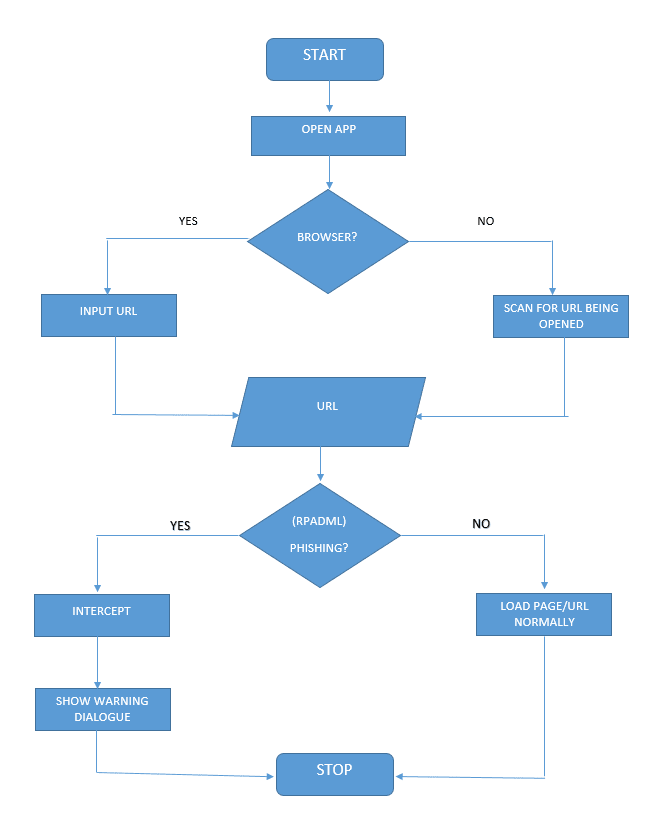
5.1.1 USECASE DIAGRAMS



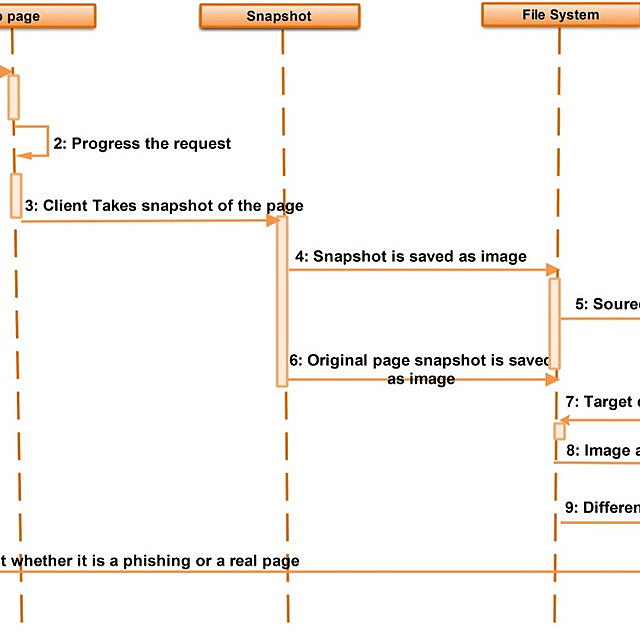
5.1.2 CLASS DIAGRAM



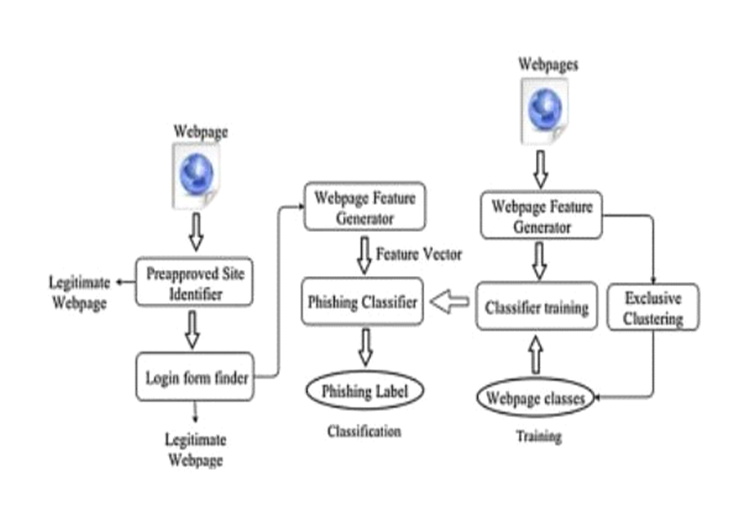
5.1.3 ACTIVITY DIAGRAM



5.1.4 SEQUENCE DIAGRAM



**5.2 SOLUTION & TECHNICAL ARCHITECTURE**



**5.3 USER STORIES**

Designing an effective anti-phishing strategy involves considering multiple factors, such as how, when and at what frequency users should undergo training. In this work, the term training is used

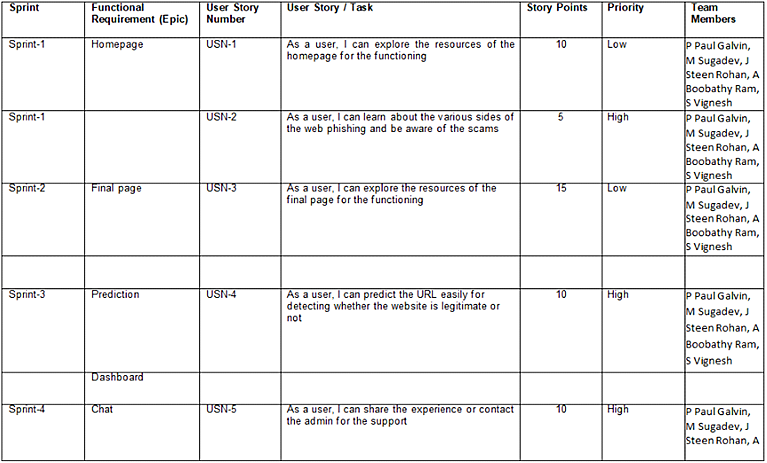
to refer to a process (e.g., a course), intended to improve a person’s awareness and knowledge of phishing, which in turn has a potential impact on his or her ability to detect and respond to phishing attempts. Such training can involve different instruments or media, such as computer based simulations, videos, and leaflets or other printed materials. To address the challenges associated with training employees to avoid such attacks, we identify relevant factors that should be considered in a company’s anti-phishing training program, then provide a comprehensive survey of relevant research results and, based on these findings, present a proposal for an ideal anti-phishing training program.

1. What are the implications of current research findings for designing effective anti phishing training programs? is effort is crucial, as insights into anti-phishing training and into how an effective training program can be developed are instrumental in improving defense against phishing attacks.

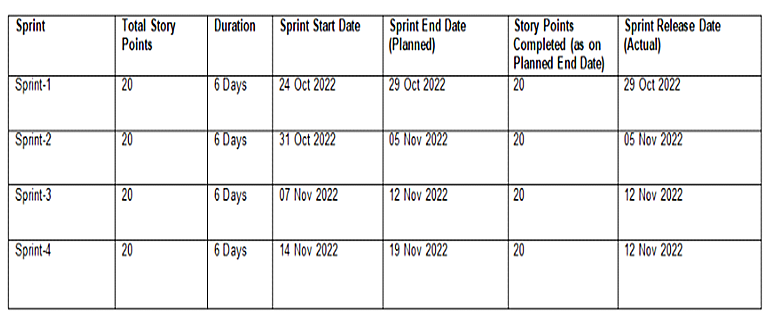
Moreover, a training program serves to reduce potential damage and increase the overall security of organizations. Current research indicates that factors such as the selected training method, how feedback should be provided to users, how training materials should be designed and how retraining intervals should be organized are relevant and thus have direct impact on the success of an anti-phishing program . Considering these findings, this paper makes the following contributions:

1. It identifies relevant academic works on anti-phishing training (“Methodology” section) • It defines multiple categories, each covering one or several of the identified core areas by examining and categorizing the surveyed works (“Categories” section)
2. It concisely presents the most important findings of each study and their implications for an envisaged training program (“Literature analysis” section)
3. It proposes an effective anti-phishing training program based on the performed analysis (“Discussion” section).

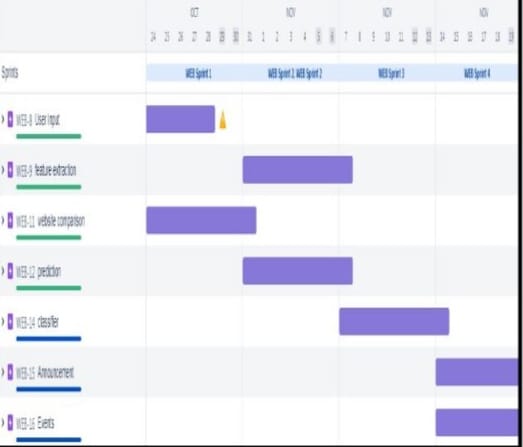
**6. PROJECT PLANNING & SCHEDULING**



**6.2 SPRINT DELIVERY SCHEDULE**



**6.3 REPORTS FROM JIRA**

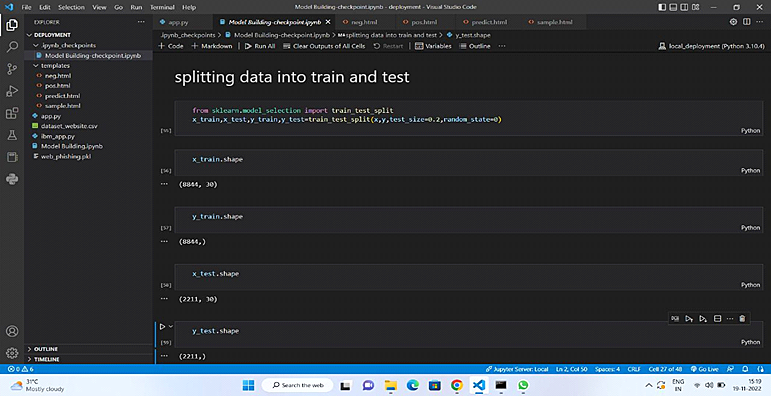


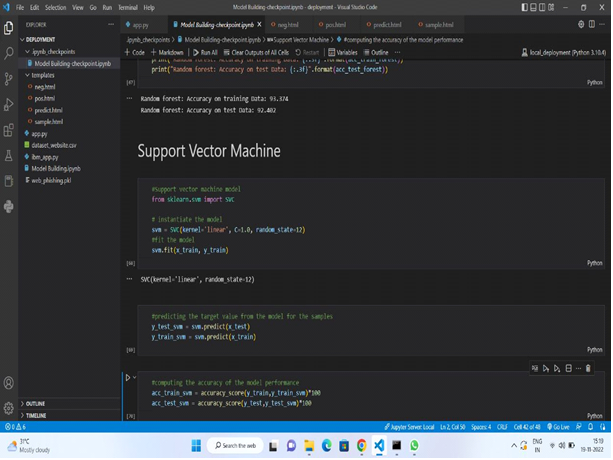
**7. CODING & SOLUTIONING**

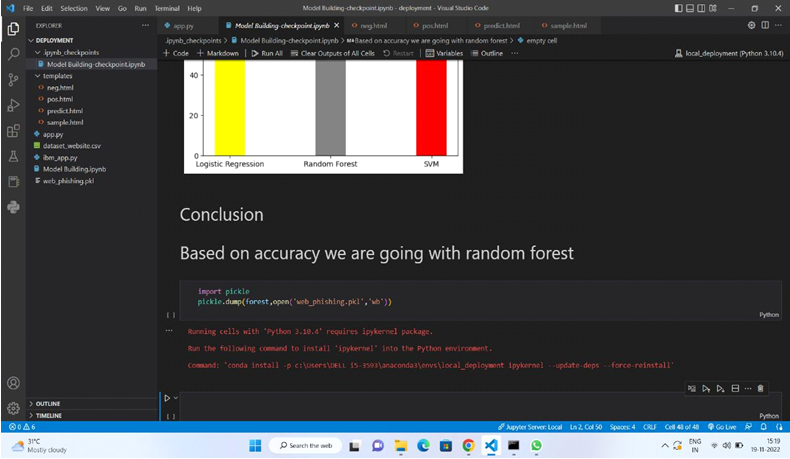
**7.1 FEATURES**

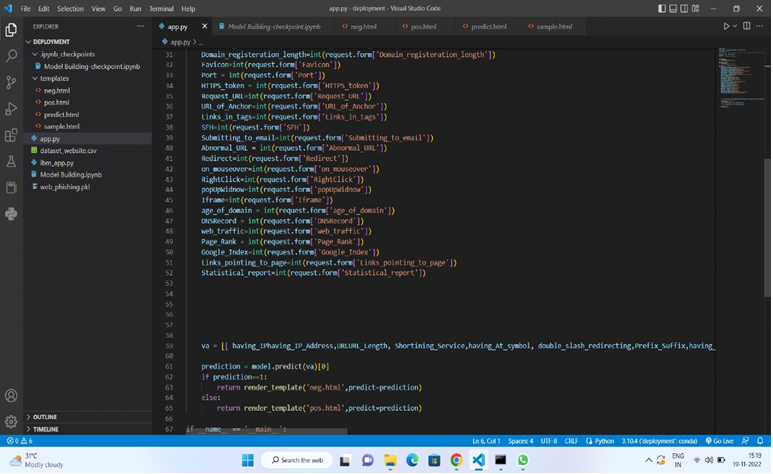
Mainly, the feature selection method aims at reducing the feature space dimensionality and enhancing the compactness of features by exploring the most contributing features in order to eliminate the less contributing ones. In the hybrid phishing detection, feature selection has been an active field of research owing to the curse of high dimensional web data (emails or websites), the existence of many irrelevant features and redundant in the examined web data, and less comprehensive and less effective machine learning classifiers against phishing evolution (Fahmy and Ghoneim, 2011; Gowtham and Krishna murthi, 2014a, 2014b; Islam and Abawajy, 2013; Xiang et al., 2011). For such key challenges, different methods of feature selection have been employed in hybrid phishing detection approaches (Basnet and Sung, 2012; Basnet et al., 2012; Hamid and Abawajy, 2011; Olivo et al., 2013; Toolan and Carthy, 2010). In Table 1, examples of the most salient feature selection methods that frequently used in the domain of phishing detection, are briefly described with respect to their search procedure, selection concept, specifics and evaluation criteria. It is noteworthy to mention that feature selection methods currently in use have shared the same process of selection involving search procedure and evaluation criterion (Chen et al., 2006; Molina et al., 2002). This means that the search procedure often discards or adds one feature against the evaluation criterion.

7.2 FEATURES









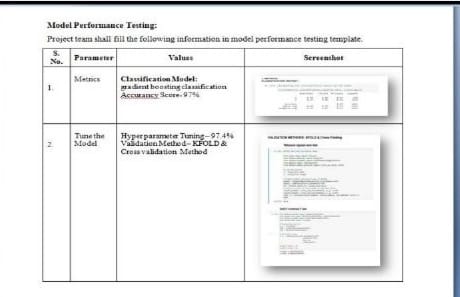
**8. TESTING**

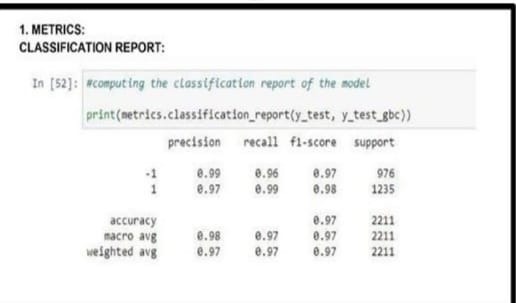
**8.1 TEST CASES**

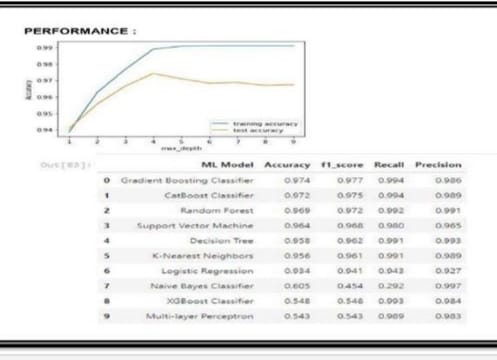
1. Performance testing
2. User Acceptance testing

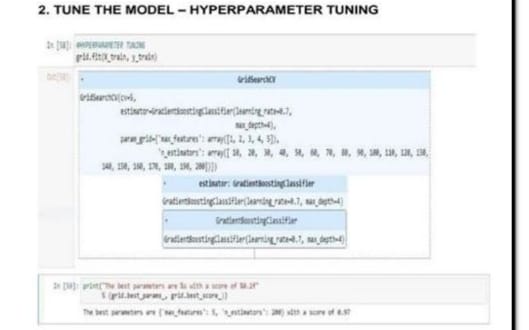
**8.2 PERFORMANCE TESTING**

Performance Testing is a software testing process used for testing the speed, response time, stability, reliability, scalability, and resource usage of a software application under a particular workload. The main purpose of performance testing is to identify and eliminate the performance bottlenecks in the software application.



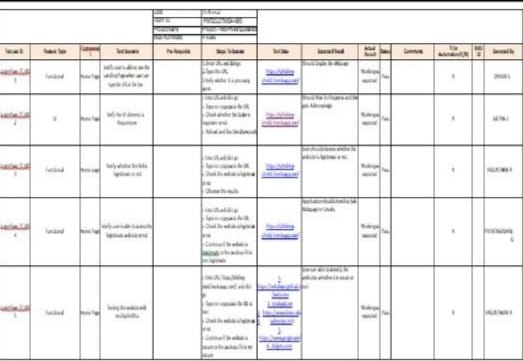


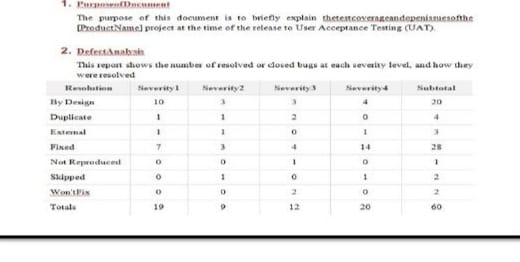






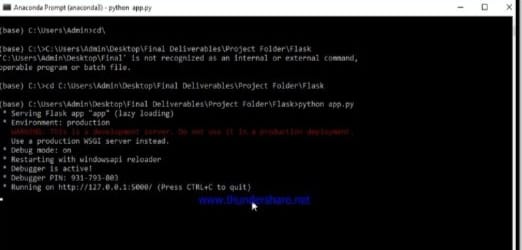
**8.2 USER ACCEPTANCE TESTING**



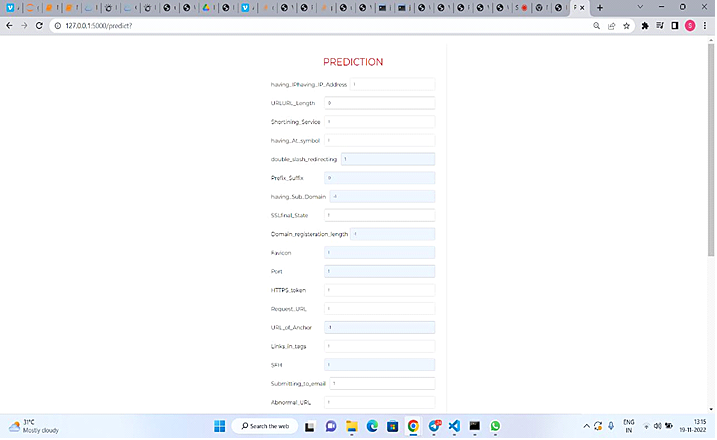


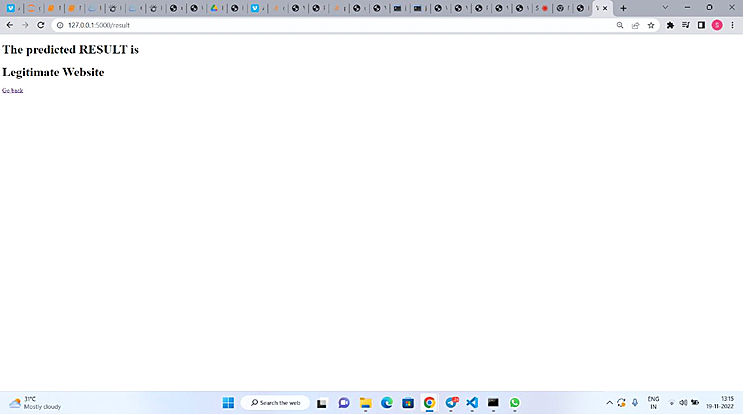


**9. RESULTS**

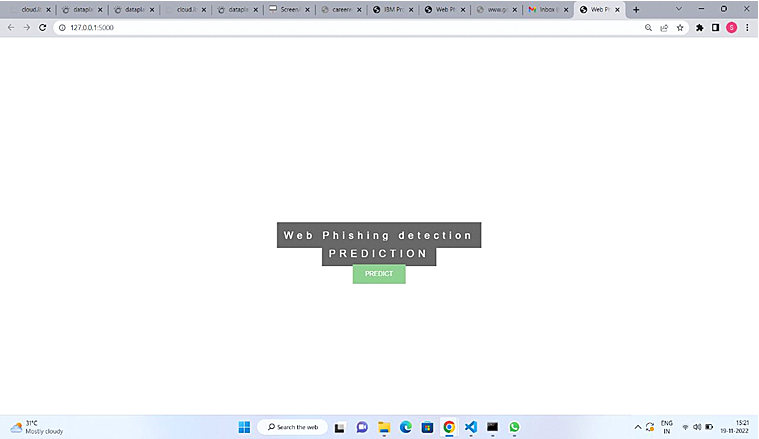
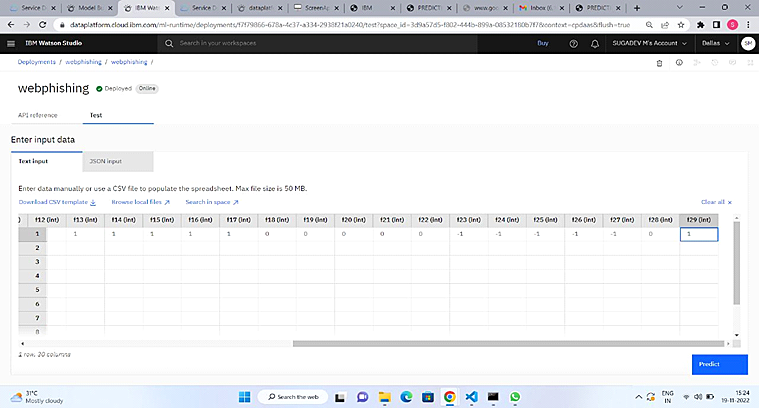


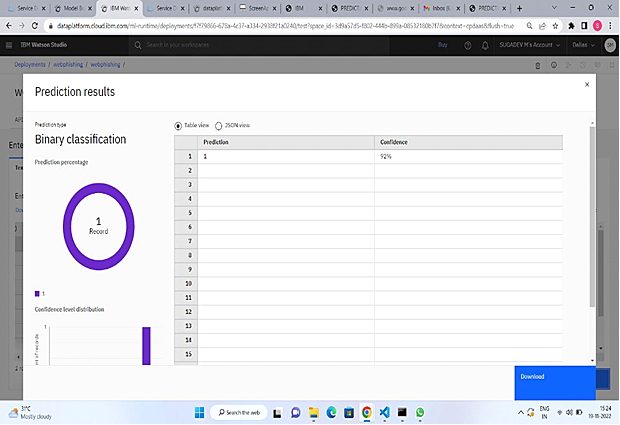
**9.2 OUTPUTS**





**9.2.2 THE FINAL STEP**





**10.** **ADVANTAGES & DISADVANTAGES**

**10.1 ADVANTAGES**

1. This system can be used by many E-commerce or other websites in order to have good customer relationship.
2. User can make online payment securely.
3. Data mining algorithm used in this system provides better performance as compared to other traditional classifications algorithms.
4. With the help of this system user can also purchase products online without any hesitation**.**

**10.2 DISADVANTAGES**

1. Time consuming .
2. huge number of features Agent (MTA) and mail user.
3. consuming memory.
4. Non standard classifier.
5. Time consuming because this technique has many layers to make the final result. huge number of features many algorithm for classification which mean time consuming  higher cost .
6. need large mail server and high memory requirement.
7. Less accuracy because it depend on unsupervised learning, need feed continuously.
8. Need feed continuously.

**11.** **CONCLUSION**

It is outstanding that a decent enemy of phishing apparatus ought to anticipate the phishing assaults in a decent timescale. We accept that the accessibility of a decent enemy of phishing device at a decent time scale is additionally imperative to build the extent of anticipating phishing sites. This apparatus ought to be improved continually through consistent retraining. As a matter of fact, the accessibility of crisp and cutting-edge preparing dataset which may gained utilizing our very own device will help us to retrain our model consistently and handle any adjustments in the highlights, which are influential in deciding the site class. Albeit neural system demonstrates its capacity to tackle a wide assortment of classification issues, the procedure of finding the ideal structure is very difficult, and much of the time, this structure is controlled by experimentation. Our model takes care of this issue via computerizing the way toward organizing a neural system conspire; hence, on the off chance that we construct an enemy of phishing model and for any reasons we have to refresh it, at that point our model will encourage this procedure, that is, since our model will mechanize the organizing procedure and will request scarcely any client defined parameters**.**

**12. FUTURE SCOPE**

The means by which hackers access user information have quickly evolved beyond traditional phishing emails. Phishing has always had the aim of baiting users to take an action or share a piece of sensitive information by appearing as a non-threat but awareness has since grown. Unprompted password reset emails, while once effective, no longer drive the same volume of user action and are often detected by spam filters.

Today, phishing attacks are targeted, can be difficult to detect, and grant malicious individuals broad permissions over user data, user devices, and online services. The days of basic phishing schemes have more or less passed. Attacks now rely on advanced forms of infiltration that better disguise malicious intent.

**13. APPENDIX**

**13.1 SOURCE CODE**

**SAMPLE.HTML:**

|  |  |
| --- | --- |
|  |  |
|  | <!DOCTYPE html> |
|  | <html lang="en"> |
|  | <head> |
|  | <meta charset="UTF-8"> |
|  | <meta http-equiv="X-UA-Compatible" content="IE=edge"> |
|  | <meta name="viewport" content="width=device-width, initial-scale=1.0"> |
|  | <title>Web Phishing Detection</title> |
|  | </head> |
|  | <body> |
|  | <h1>The predicted RESULT is</h1> |
|  | <h1>Illegitimate website</h1> |
|  | <a href="/">Go back</a> |
|  | </body> |
|  | </html> |

**PREDICT.HTML**

|  |  |
| --- | --- |
|  |  |
|  |  |
|  | <html> |
|  | <head> |
|  | <title>PREDICTION</title> |
|  | <meta charset="UTF-8"> |
|  | <meta name="viewport" content="width=device-width, initial-scale=1.0"> |
|  | <meta http-equiv="X-UA-Compatible" content="ie=edge"> |
|  | <style> |
|  | @import url('https://fonts.googleapis.com/css2?family=Montserrat&display=swap'); |
|  | \*{ |
|  | margin: 0; |
|  | padding: 0; |
|  | box-sizing: border-box; |
|  | font-family: 'montserrat',sans-serif ; |
|  | } |
|  | body{ |
|  |  |
|  | padding: 0px 10px; |
|  | background-size: cover; |
|  | } |
|  | .container{ |
|  | max-width: 500px; |
|  | width: 100%; |
|  | background: transparent; |
|  | margin: 20px auto; |
|  | padding: 30px; |
|  | box-shadow: 1px 1px 2px rgba(0, 0, 0, 0.125); |
|  | } |
|  | .container .title{ |
|  | font-size: 25px; |
|  | font-weight: 1000px; |
|  | margin-bottom: 25px; |
|  | color: red; |
|  | text-transform: uppercase; |
|  | text-align: center; |
|  | } |
|  | .container .form{ |
|  | width: 100%; |
|  | } |
|  |  |
|  | .container .form .input\_field{ |
|  | margin-bottom: 15px; |
|  | display: flex; |
|  | align-items: center; |
|  | } |
|  | .container .form .input\_field label{ |
|  | width: 200px; |
|  | margin-right: 10px; |
|  | font-size: 14px; |
|  |  |
|  | } |
|  | .container .form .input\_field .input, |
|  | .container .form .input\_field .textarea{ |
|  | width: 100%; |
|  | outline: none; |
|  | border: 1px solid #d5d5d5; |
|  | font-size: 10px; |
|  | padding: 8px 10px; |
|  | border-radius: 3px; |
|  | transition: all 0.3s ease; |
|  | } |
|  |  |
|  | .container .form .input\_field .textarea{ |
|  | resize: none; |
|  | height: 90px; |
|  |  |
|  | } |
|  | .container .form .input\_field .custom\_radio{ |
|  | position: relative; |
|  |  |
|  | width: 100%; |
|  | } |
|  |  |
|  | .container .form .input\_field .input:focus, |
|  | .container .form .input\_field .textarea:focus, |
|  | .container .form .input\_field select:focus{ |
|  | border: 1px solid #fec107; |
|  |  |
|  | } |
|  | #next{ |
|  | border-radius : 100px; |
|  | padding: 10px 15px; |
|  | background: rgb(40, 24, 252); |
|  | border: none; |
|  | outline: none; |
|  | color: white; |
|  | margin-top: 15px; |
|  |  |
|  | } |
|  | #next:hover{ |
|  | background-color: rgb(12, 221, 221); |
|  | cursor: pointer; |
|  | color: black; |
|  | } |
|  |  |
|  |  |
|  | #reset{ |
|  | border-radius : 60px; |
|  | padding: 10px 15px; |
|  | background: rgb(40, 24, 252); |
|  | width: 70%; |
|  | border: none; |
|  | outline: none; |
|  | color: white; |
|  | margin-top: 25px; |
|  | margin-left: 100px; |
|  |  |
|  | } |
|  | #reset:hover{ |
|  | background-color: rgb(16, 247, 47); |
|  | cursor: pointer; |
|  | color: black; |
|  | } |
|  |  |
|  |  |
|  | @media (max-width:420px){ |
|  | .container .form .input\_field{ |
|  | flex-direction: column; |
|  | align-items: flex-start; |
|  | } |
|  | .container .form .input\_field{ |
|  | margin-bottom: 5px; |
|  | } |
|  | } |
|  |  |
|  | </style> |
|  | <div class="container"> |
|  | <div class ="title"> |
|  | PREDICTION |
|  | </div> |
|  | <form method="POST" action="{{ url\_for('result')}}"> |
|  | <div class="form"> |
|  | <div class="input\_field"> |
|  | <label>having\_IPhaving\_IP\_Address</label> |
|  | <input type="text" placeholder="Enter the value" class="input" name="having\_IPhaving\_IP\_Address"required> |
|  | </div> |
|  |  |
|  | <div class="input\_field"> |
|  | <label>URLURL\_Length </label> |
|  | <input type="text" placeholder="Enter the value" class="input" name="URLURL\_Length"required> |
|  | </div> |
|  |  |
|  | <div class="input\_field"> |
|  | <label>Shortining\_Service</label> |
|  | <input type="text" placeholder="Enter the value" class="input" name="Shortining\_Service"required> |
|  | </div> |
|  |  |
|  | <div class="input\_field"> |
|  | <label>having\_At\_symbol</label> |
|  | <input type="text" placeholder="Enter the value" class="input" name="having\_At\_symbol"required> |
|  | </div> |
|  |  |
|  | <div class="input\_field"> |
|  | <label>double\_slash\_redirecting</label> |
|  | <input type="text" placeholder="Enter the value" class="input" name="double\_slash\_redirecting"required> |
|  | </div> |
|  |  |
|  | <div class="input\_field"> |
|  | <label>Prefix\_Suffix </label> |
|  | <input type="text" placeholder="Enter the value" class="input" name="Prefix\_Suffix"required> |
|  | </div> |
|  |  |
|  | <div class="input\_field"> |
|  | <label>having\_Sub\_Domain </label> |
|  | <input type="text" placeholder="Enter the value" class="input" name="having\_Sub\_Domain"required> |
|  | </div> |
|  |  |
|  | <div class="input\_field"> |
|  | <label>SSLfinal\_State</label> |
|  | <input type="text" placeholder="Enter the value" class="input" name="SSLfinal\_State"required> |
|  | </div> |
|  |  |
|  | <div class="input\_field"> |
|  | <label>Domain\_registeration\_length </label> |
|  | <input type="text" placeholder="Enter the value" class="input" name="Domain\_registeration\_length"required> |
|  | </div> |
|  |  |
|  | <div class="input\_field"> |
|  | <label>Favicon </label> |
|  | <input type="text" placeholder="Enter the value" class="input" name="Favicon"required> |
|  | </div> |
|  | <div class="input\_field"> |
|  | <label>Port </label> |
|  | <input type="text" placeholder="Enter the value" class="input" name="Port"required> |
|  | </div> |
|  | <div class="input\_field"> |
|  | <label>HTTPS\_token </label> |
|  | <input type="text" placeholder="Enter the value" class="input" name="HTTPS\_token"required> |
|  | </div> |
|  | <div class="input\_field"> |
|  | <label>Request\_URL</label> |
|  | <input type="text" placeholder="Enter the value" class="input" name="Request\_URL"required> |
|  | </div> |
|  | <div class="input\_field"> |
|  | <label>URL\_of\_Anchor</label> |
|  | <input type="text" placeholder="Enter the value" class="input" name="URL\_of\_Anchor"required> |
|  | </div> |
|  | <div class="input\_field"> |
|  | <label>Links\_in\_tags </label> |
|  | <input type="text" placeholder="Enter the value" class="input" name="Links\_in\_tags"required> |
|  | </div> |
|  | <div class="input\_field"> |
|  | <label>SFH </label> |
|  | <input type="text" placeholder="Enter the value" class="input" name="SFH"required> |
|  | </div> |
|  | <div class="input\_field"> |
|  | <label>Submitting\_to\_email </label> |
|  | <input type="text" placeholder="Enter the value" class="input" name="Submitting\_to\_email"required> |
|  | </div> |
|  | <div class="input\_field"> |
|  | <label>Abnormal\_URL</label> |
|  | <input type="text" placeholder="Enter the value" class="input" name="Abnormal\_URL"required> |
|  | </div> |
|  | <div class="input\_field"> |
|  | <label>Redirect</label> |
|  | <input type="text" placeholder="Enter the value" class="input" name="Redirect"required> |
|  | </div> |
|  | <div class="input\_field"> |
|  | <label>on\_mouseover</label> |
|  | <input type="text" placeholder="Enter the value" class="input" name="on\_mouseover"required> |
|  | </div> |
|  | <div class="input\_field"> |
|  | <label>RightClick </label> |
|  | <input type="text" placeholder="Enter the value" class="input" name="RightClick"required> |
|  | </div> |
|  | <div class="input\_field"> |
|  | <label>popUpWidnow</label> |
|  | <input type="text" placeholder="Enter the value" class="input" name="popUpWidnow"required> |
|  | </div> |
|  | <div class="input\_field"> |
|  | <label>Iframe</label> |
|  | <input type="text" placeholder="Enter the value" class="input" name="Iframe"required> |
|  | </div> |
|  | <div class="input\_field"> |
|  | <label>age\_of\_domain </label> |
|  | <input type="text" placeholder="Enter the value" class="input" name="age\_of\_domain"required> |
|  | </div> |
|  | <div class="input\_field"> |
|  | <label>DNSRecord</label> |
|  | <input type="text" placeholder="Enter the value" class="input" name="DNSRecord"required> |
|  | </div> |
|  | <div class="input\_field"> |
|  | <label>web\_traffic</label> |
|  | <input type="text" placeholder="Enter the value" class="input" name="web\_traffic"required> |
|  | </div> |
|  | <div class="input\_field"> |
|  | <label>Page\_Rank</label> |
|  | <input type="text" placeholder="Enter the value" class="input" name="Page\_Rank"required> |
|  | </div> |
|  | <div class="input\_field"> |
|  | <label>Google\_Index </label> |
|  | <input type="text" placeholder="Enter the value" class="input" name="Google\_Index"required> |
|  | </div> |
|  | <div class="input\_field"> |
|  | <label>Links\_pointing\_to\_page</label> |
|  | <input type="text" placeholder="Enter the value" class="input" name="Links\_pointing\_to\_page"required> |
|  | </div> |
|  | <div class="input\_field"> |
|  | <label>Statistical\_report</label> |
|  | <input type="text" placeholder="Enter the value" class="input" name="Statistical\_report"required> |
|  | </div> |
|  |  |
|  | <center><button type="submit" id="next">SUBMIT</button></center> |
|  | </div> |
|  | </form> |
|  | </div> |
|  |  |
|  |  |
|  | </head> |
|  | </html> |

**POS.HTML**

|  |  |
| --- | --- |
|  | <!DOCTYPE html> |
|  | <html lang="en"> |
|  | <head> |
|  | <meta charset="UTF-8"> |
|  | <meta http-equiv="X-UA-Compatible" content="IE=edge"> |
|  | <meta name="viewport" content="width=device-width, initial-scale=1.0"> |
|  | <title>Web Phishing Detection</title> |
|  | </head> |
|  | <body> |
|  | <h1>The predicted RESULT is</h1> |
|  | <h1>Legitimate Website</h1> |
|  | <a href="/">Go back</a> |
|  | </body> |
|  | </html> |

**NEG.HTML**

|  |  |
| --- | --- |
|  |  |
|  | <!DOCTYPE html> |
|  | <html lang="en"> |
|  | <head> |
|  | <meta charset="UTF-8"> |
|  | <meta http-equiv="X-UA-Compatible" content="IE=edge"> |
|  | <meta name="viewport" content="width=device-width, initial-scale=1.0"> |
|  | <title>Web Phishing Detection</title> |
|  | </head> |
|  | <body> |
|  | <h1>The predicted RESULT is</h1> |
|  | <h1>Illegitimate website</h1> |
|  | <a href="/">Go back</a> |
|  | </body> |
|  | </html> |

**GITHUB LINK**

[GitHub - IBM-EPBL/IBM-Project-40830-1660636265: Web Phishing Detection](https://github.com/IBM-EPBL/IBM-Project-40830-1660636265)

**PROJECT DEMO LINK**

<https://drive.google.com/file/d/1BFyynpQuQj9aCeMztE6dK5zx3hfMEtNb/view?usp=drivesdk>