WEB PHISHING DETECTION

PROJECT REPORT

Submitted by

Team ID: PNT2022TMID36224

KEERTHANA. G	**110519205008**
GAYATHRI. K	**110519205004**
HARI PRIYA. M	**110519205005**
SWETHA. S	**110519205016**

In Partial Fulfillment for the Award of the Degree

OF

BACHELOR OF TECHNOLOGY

IN

INFORMATION TECHNOLOGY

GOJAN SCHOOL OF BUSINESS AND TECHNOLOGY

ANNA UNIVERSITY: CHENNAI – 600 025

NOVEMBER 2022

PROJECT REPORT
WEB PHISHING DETECTION

1. INTRODUCTION

1.1 Project Overview

Now days, As there are so many people are being aware of using internet to perform various activities like online shopping, online bill payment, online mobile recharge, banking transaction. There are many cybercrime that are widely performed for example fraud, spam, cyber terrorisms and phishing.

There are a number of users who purchase products online and make payments through e-banking. There are e-banking websites that ask users to provide sensitive data such as username, password & credit card details, etc., often for malicious reasons. This type of e-banking website is known as a phishing website. Web service is one of the key communications software services for the Internet. Web phishing is one of many security threats to web services on the Internet.

As the phishing website attacks mostly target online business, bank web user and government, so it as become a national security issue. It is necessary that these attacks are detected at an early stage. In this situation, it is preferred to develop guidelines to extract specific features from websites and then use them to predict the type of web page.

1.2 Purpose

The aim of this work was to develop model to safeguard user from phishing attacks. Phishing attacks are growing in similar manner as e-commerce industries are growing. Prediction and prevention of phishing attacks is a very critical towards safeguarding online transaction. Data mining tools can be applied in this regard as the technique is very easy and can mine millions of information within seconds and deliver accurate results.

2. LITERTURE SURVEY

2.1 Existing problem

Based Phishing Detection Systems These systems are based on visual similarity comparison of the web pages. Phishing and non-phishing sites are classified by taking a server-side view of them. These two data are compared with image processing techniques. Fake websites are often designed very close to the original ones. But visually, there are minor differences between them. It is easier to notice these differences, which users cannot easily notice, with image processing techniques. According to the similarity obtained, it is decided whether the website is phishing or not. In the literature, as in the study, there are studies, which detect the differences based on basic similarities.

The detection of the phishing website in Machine Learning Based Phishing Detection Systems is based on the classification of the specified features by using some artificial intelligence techniques. Features are created by collecting in different categories such as URL, domain name, website features or website content etc. Due to the dynamic structure, especially for the detection of the anomaly in the web sites, it has high popularity on the security of the users.

In these systems, features are obtained based on relational rule mining. The rules are estimated to emphasize features that are more common in phishing URLs. In studies using this type of system, it is aimed to use effective features more actively in the classification. In these systems, a set of rules are determined. Thus, the system gives a higher accuracy rate when trained with these rules. In this context, like CANTINA study, the Term Frequency - Inverse Document Frequency (TF-IDF) and rules were used to detect phishing attacks. In addition, in similar studies, models were created by using some features and rules.

2.2 Reference

1. Project Title : Detecting phishing websites using machine learning technique
   Author Name **:** Ashit Kumar Dutta
   Year: 2021

2. Project Title : Phishing Websites Detection using Machine Learning
   Author Name **:** Arun Kulkarni1, Leonard L. Brown, III2
   Year : 2019

3. Project Title : Detecting Phishing Websites Using Machine Learning
   Author Name : Aniket Garje1, Namrata Tanwani1 , Sammed Kandale1 , Twinkle Zope1 , Prof. Sandeep Gore2
   Year : 2021

4. Project Title : Phishing Detection using Machine Learning based URL Analysis: A Survey
   Author Name : Arathi Krishna V, Anusree A, Blessy Jose, Karthika Anilkumar, Ojus Thomas Lee
   Year : 2021

5. Project Name : Applications of deep learning for phishing detection: a systematic literature review
   Author Name : Cagatay Catal1 , Görkem Giray2 , Bedir Tekinerdogan3 , Sandeep Kumar4 , Suyash Shukla4
   Year : 2022

6. Project Name : A Systematic Literature Review on Phishing and Anti-Phishing Techniques
   Author Name : Ayesha Arshad1 , Attique Ur Rehman1 , Sabeen Javaid1, Tahir Muhammad Ali2 , Javed Anjum Sheikh1 , Muhammad Azeem
   Year : 2021

2.3 Problem Statement Definition

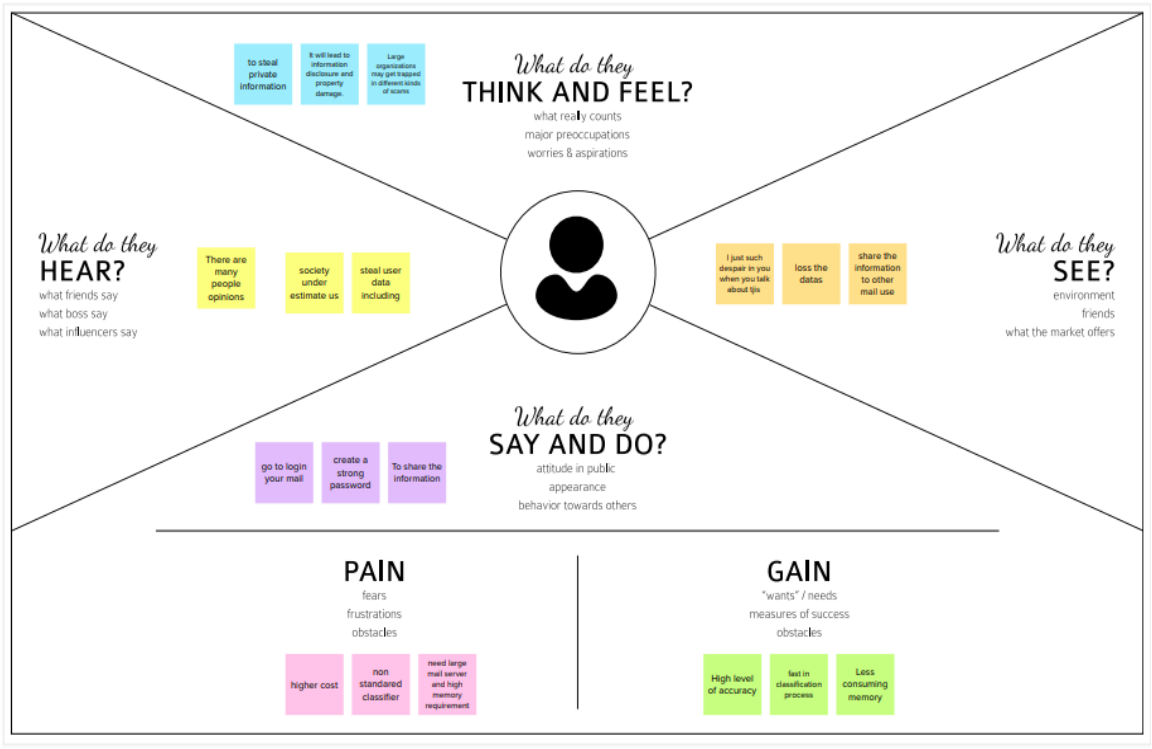
Internet has dominated the world by dragging half of the world's population exponentially into the cyber world. With the booming of internet transactions, cybercrimes rapidly increased and with anonymity presented by the internet, Hackers attempt to trap the end-users through various forms such as phishing, SQL injection, malware, man-in-the-middle, domain name system tunnelling, ransomware, web trojan, and so on.

Phishing attacks succeed when human users fail to detect phishing sites. Generally speaking, past work in anti-phishing falls into four categories: studies to understand why people fall for phishing attacks, methods for training people not to fall for phishing attacks, user interfaces for helping people make better decisions about rusting email and websites, and automated tools to detect phishing. Our work describes an automated approach to detect phishing.

Most of the end user normally takes decision only based on what he/she look and feel. When a user is accessing internet he/she only see the screen of a browser. He/she then work on the command of a web-page. The user doesn‘t concern about the back end process and most phishing attempts get this type of unintentional opportunity given by the user and make them fool.

3. IDEATION & PROPOSED SOLUTION

3.1 Empathy Map Canvas



An **EMPATHY MAP** is a collaborative visualization used to articulate what we know about a particular type of user. It externalizes knowledge about user in order to

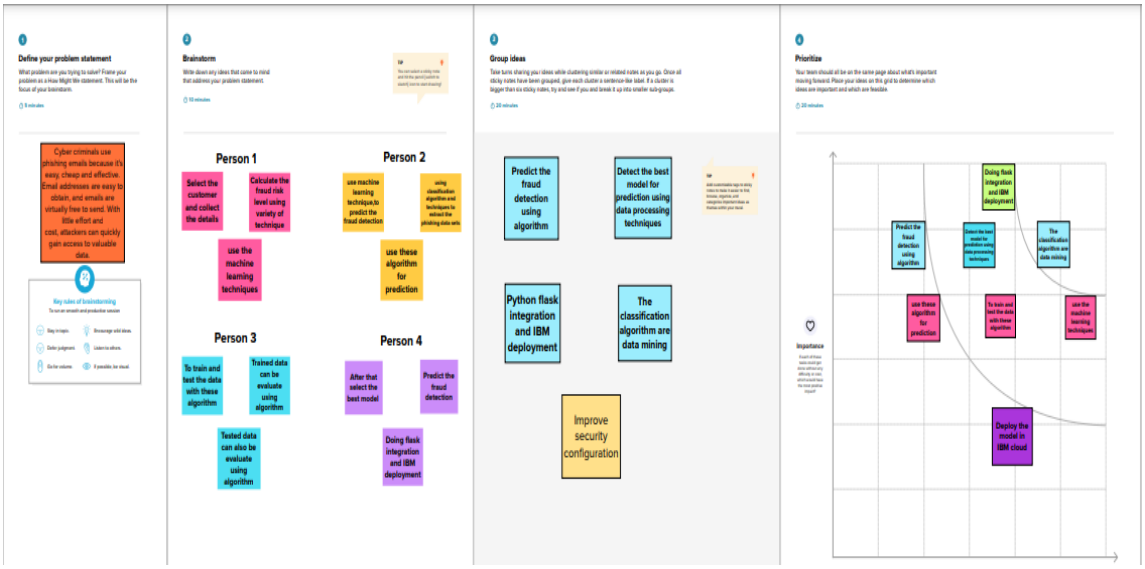
1. Create a shared understanding of user needs.
2. Aid in decision making.

Traditional empathy map is split into **four** quadrants. There are,

1. Says
2. Thinks
3. Does
4. Feels

Empathy map provides a glance into who a user is as a whole and are not chronological or sequential.

3.2 Ideation & Brainstorming



Brainstorming Audit plans have to be designed to find fraud. Here's help for your team on fraud brainstorming: delving into the details, thinking like a fraudster and using the knowledge of the processes to increase awareness of where frauds may be hiding.

For the most part, the audit team members will be the primary individuals involved in a fraud brainstorming session in advance of an audit so the objectives will remain relatively confidential. This also will minimize the possibility that the target group gets wind of the impending audit, especially steps designed to detect fraud. Therefore, carefully manage and safeguard the inclusion of others in this process.

Process Complexity:

Assess the complexity of the process' moving parts. The more complex a process, the greater the chance that fraud will slip through the cracks and crevices. Finding fraud is difficult. We all know that. We're constantly reminded at every audit, fraud and accounting conference we attend that fraud is inherently hidden.

3.3 Proposed Solution

S.No.	Parameter	Description
1.	Problem Statement (Problem to be solved)	Phishing detection techniques do suffer low detection accuracy and high false alarm especially whennovel phishing approaches are introduced.
2.	Idea / Solution description	The sophisticated infrastructure of the Internet, people can do transactions such as shopping, banking etc. The Internet has many advantages,at the same time;it alsohas its own set of security and privacy problems so create a prominent platformfor cyber-attacks usingphishing.
3.	Novelty / Uniqueness	Web phishing detection is used to detect good domain.The dataset which will be used in the training phase is a very important point to build successful detectionmechanism.
4.	Social Impact / Customer Satisfaction	Phishing has a list of negative effects on a business,including loss of money,lossofintellectualproperty,damagetoreputation,anddisruptionofoperationalactivity. These effects work together to cause loss of company value,sometimes withirrreparable repercussions.
5.	Business Model (Revenue Model)	Phishing represents one aspect of the increasingly complex and convergingsecurity threats facing businesses today. The methods used by spammers havebecomemoresophisticated,andspamisnowincreasinglycombinedwithmalware and usedas a tool for online fraudor theft.
6.	Scalability of the Solution	Use anti-phishing protection and anti-spam software to protect yourself when malicious messages slip through to your computer.Anti-malware is includedto prevent other types of threats. Similar to anti-spam software, anti-malwaresoftware is programmed by security researchers to spot even the stealthiestmalware.

Proposed Solution means the technical solution to be provided by the Implementation agency in response to the requirements and the objectives of the Project. Proposed Solution means the combination of software, hardware, other products or equipment, and any and all services (including any installation, implementation, training, maintenance and support services) necessary to implement the solution described by Vendor in its Proposal.

Processing Time does not include the time when the incident is on status “Customer Action” or “SAP Proposed Solution”, whereas (a) the status Customer Action means the incident was handed over to Licensee; and (b) the status SAP Proposed Solution means SAP has provided a Corrective Action as outlined herein.

3.4 Problem Solution Fit

The problem-solution fit is when you-

Validate that the problem exists: When you validate your problem hypothesis using real-world data and feedback. That is, you gather information from real users to determine whether or not they care about the pain point you're trying to solve.

Validate that your solution solves the problem: When you validate that the target audience appreciates the value your solution delivers to them.

4. REQUIREMENT ANALYSIS

4.1 Functional requirement

FR No.	Functional Requirement (Epic)	Sub Requirement (Story / Sub-Task)
FR-1	User Registration	Registration through Form Registration through Gmail
FR-2	User Confirmation	Confirmation via Email Confirmation via OTP
FR-3	User Login	Login with User ID Login with Valid Email
FR-4	User Interface	Profile Details

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.

Functional system requirement: Extension plugin should provide a warning pop-up when they visit a website that is phished; therefore, it should strictly follow the following:

a. Extension plugin ability to present the pop-up to the users screen should be quick enough to the point, users will be aware before entering any confidential or sensitive details into a phishing website.

b. Extension plugin should not need the facilities and services from an 3rd party service or APIs, due the reason that those services will always the potential to leak users browsing data and pattern when it gets compromised by hackers.

c. Extension plugin will have the capability to also detect latest and new phishing websites.

4.2 Non-Functional requirements

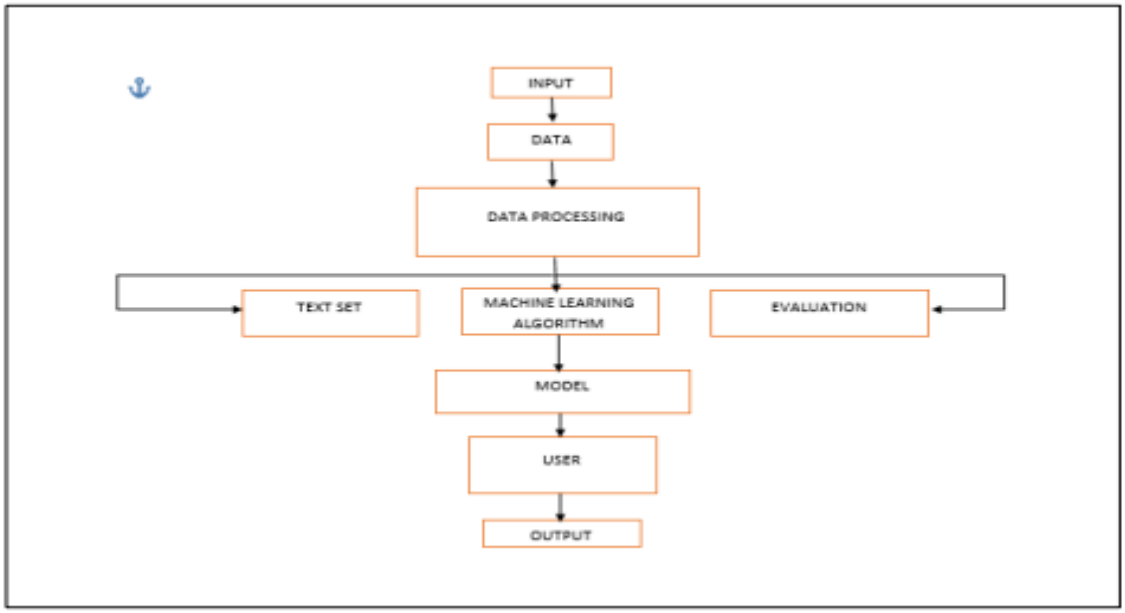
FR No.	Non-Functional Requirement	Description
NFR-1	**Usability**	The Ease at which an average person can use the Software or to achieve specific goals
NFR-2	**Security**	A secure website has a web application firewall activated to prevent attacks and hacks.
NFR-3	**Reliability**	Transparent about their sources and help the reader gain a deeper understanding of the topic.
NFR-4	**Performance**	The objective measurement and perceived user experience of a web site or application.
NFR-5	**Availability**	The ability of the users to access and use a website or web service
NFR-6	**Scalability**	The ability of an application to handle a growing number of users and load, without compromising on performance and causing disruptions to user experience.

Non-functional 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.

Graphical User Interface design Interface developed should be done with the understanding that it must meet the simplicity of what users would like to see when they need an extension for detecting things, and also it needs to adhere to non IT literate users as well. It must also provide the exact information on what the user wants like identifying a phishing website quickly without needing to click on many options.

5. PROJECT DESIGN

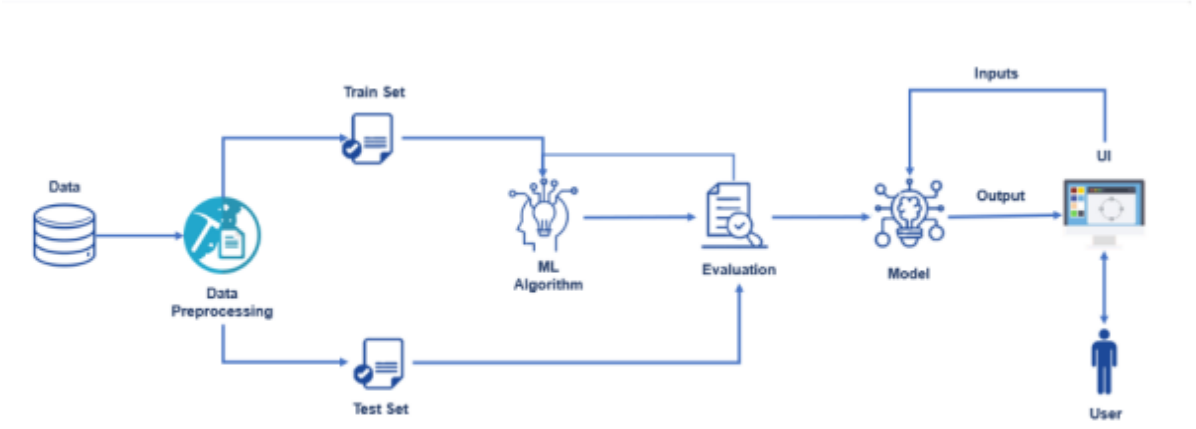
5.1 Data Flow Diagram



A data-flow diagram is a way of representing a flow of data through a process or a system. The DFD also provides information about the outputs and inputs of each entity and the process itself. A data-flow diagram has no control flow- there are no decision rules and no loops.

A data-flow diagram is a way of representing a flow of data through a process or a system (usually an information system). The DFD also provides information about the outputs and inputs of each entity and the process itself. A data-flow diagram has no control flow — there are no decision rules and no loops. Specific operations based on the data can be represented by a flowchart.

5.2 Solution & Technical Architecture



Solution architecture is a complex process – with many sub-processes – that bridges the gap between business problems and technology solutions. Its goals are to:

• Find the best tech solution to solve existing business problems.
• Describe the structure, characteristics, behaviour, and other aspects of the software to project stakeholders.
• Define features, development phases, and solution requirements.
• Provide specifications according to which the solution is defined, managed, and delivered.

Solutions architect, is a professional who helps clients by creating personalized solutions for their IT requirements. They assist companies' management teams by assessing business requirements and presenting them with strategic plans. These professionals may work alongside other IT architects, engineers and analysts.

5.3 User stories

User Type	Functional Requirement (Epic)	User Story Number	User Story / Task	Acceptance criteria	Priority	Release
Customer (Mobile user)	Registration	USN-1	As a user, I can register for the application by entering my email, password, and confirming my password.	I can access my account / dashboard	High	Sprint-1
		USN-2	As a user, I will receive confirmation email once I have registered for the application	I can receive confirmation email & click confirm	High	Sprint-1
		USN-3	As a user, I can register for the application through Facebook	I can register & access the dashboard with Facebook Login	Low	Sprint-2
		USN-4	As a user, I can register for the application through Gmail	I can receive confirmation Gmail	Medium	Sprint-1
	Login	USN-5	As a user, I can log into the application by entering email & password	I can receive confirmation email & Password	High	Sprint-1
		USN-6	As a user, I can log into the application by entering user id	I can register by user id information	High	Sprint-1
Customer (Web user)	User Interface	USN-7	As a user, I can log into the application by profile details	I can access my account / dashboard	High	Sprint-1
Customer Care Executive	User Expectation	USN-8	As a user, I can customize the output	I can receive expect output	High	Sprint-1
Administrator	Registration	USN-9	As an administrator, accept the user application via email / dashboard	I can access User registration form	Medium	Sprint-1
		USN-10	As an administrator, send a confirmation link to user	I can accept the user request	High	Sprint-1
	Executive	USN-11	As an administrator, understand the user expected output	I can complete expect output	High	Sprint-1

6. PROJECT PLANNING & SCHEDULES

6.1 Sprint Planning & Estimation

Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority	Team Members
Sprint-1	Home page	USN-1	As a user, I can survey the resource of the home page functioning.	5	High	Haripriya.M Keerthana.G
Sprint-1		USN-2	As a user, I can learn about the various resource sites of web phishing	6	High	Gayathri.K Swetha.S
Sprint-2	Final page	USN-3	As a user, I can survey the resource of the Final page functioning.	4	Low	Haripriya.M Gayathri.K
Sprint-3	Prediction	USN-4	As a user, I can predict the URL easly for detecting	10	High	Keerthana.G Haripriya.M Gayathri.K Swetha.S
Sprint-4	Dashboard	USN-5	As a user, I can check more information through dashboard	4	High	Keerthana.G Gayathri.K
Sprint-5	Chat	UNS-6	As a user, I can share the experience or contact the admin for the support	3	Low	Haripriya.M Swetha.S

Sprint planning is an event in scrum that kicks off the sprint. The purpose of sprint planning is to define what can be delivered in the sprint and how that work will be achieved. Sprint planning is done in collaboration with the whole scrum team. In scrum, the sprint is a set period of time where all the work is done. However, before you can leap into action you have to set up the sprint. You need to decide on how long the time box is going to be, the sprint goal, and where you're going to start. The sprint planning session kicks off the sprint by setting the agenda and focus. If done correctly, it also creates an environment where the team is motivated, challenged, and can be successful. Bad sprint plans can derail the team by setting unrealistic expectations.

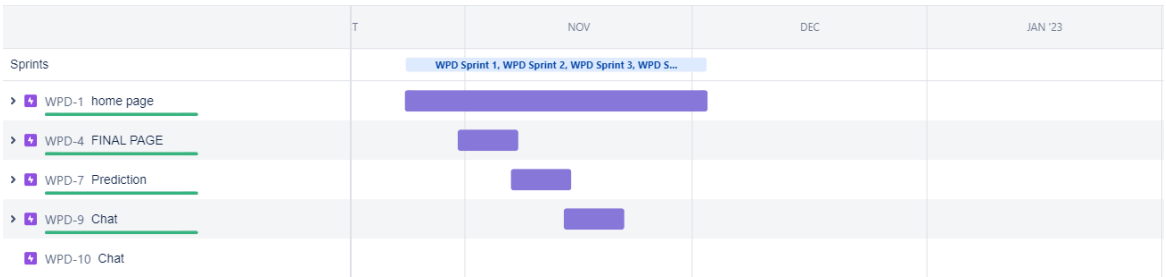
6.2 Sprint Delivery Schedule

Sprint	Total Story Points	Duration	Sprint Start Date	Sprint End Date (Planned)	Story Points Completed (as on Planned End Date)	Sprint Release Date (Actual)
Sprint-1	20	6 Days	24 Oct 2022	29 Oct 2022	20	30 Oct 2022
Sprint-2	20	6 Days	31 Oct 2022	05 Nov 2022	20	05 Nov 2022
Sprint-3	20	6 Days	07 Nov 2022	12 Nov 2022	20	12 Nov 2022
Sprint-4	20	6 Days	14 Nov 2022	19 Nov 2022	20	12 Nov 2022

The objective of sprint planning is to work out the key details regarding the team's planned work during the next sprint. With that in mind, the sprint team should plan to address at least the following issues during this meeting.

Objectives are the desired benefits, outcomes, or performance improvements that you expect from the project. The deliverables, also called "outputs", are the tangible things that the project will produce to enable the objectives to be achieved.

6.3 Reports from JIRA



7. CODING & SOLUTIONING

7.1 Feature 1

HOME PAGE:

A home page is the primary web page that a visitor will view when they navigate to a website via a search engine, and it may also function as a landing page to attract visitors.

Home page is also a web page but is being considered as the starting page of your website with the navigation bar that provides links to different sections with in the particular website.

Home page can have lots of information, links, and sources and the purpose it is provide plenty of information and resources to potential customers.

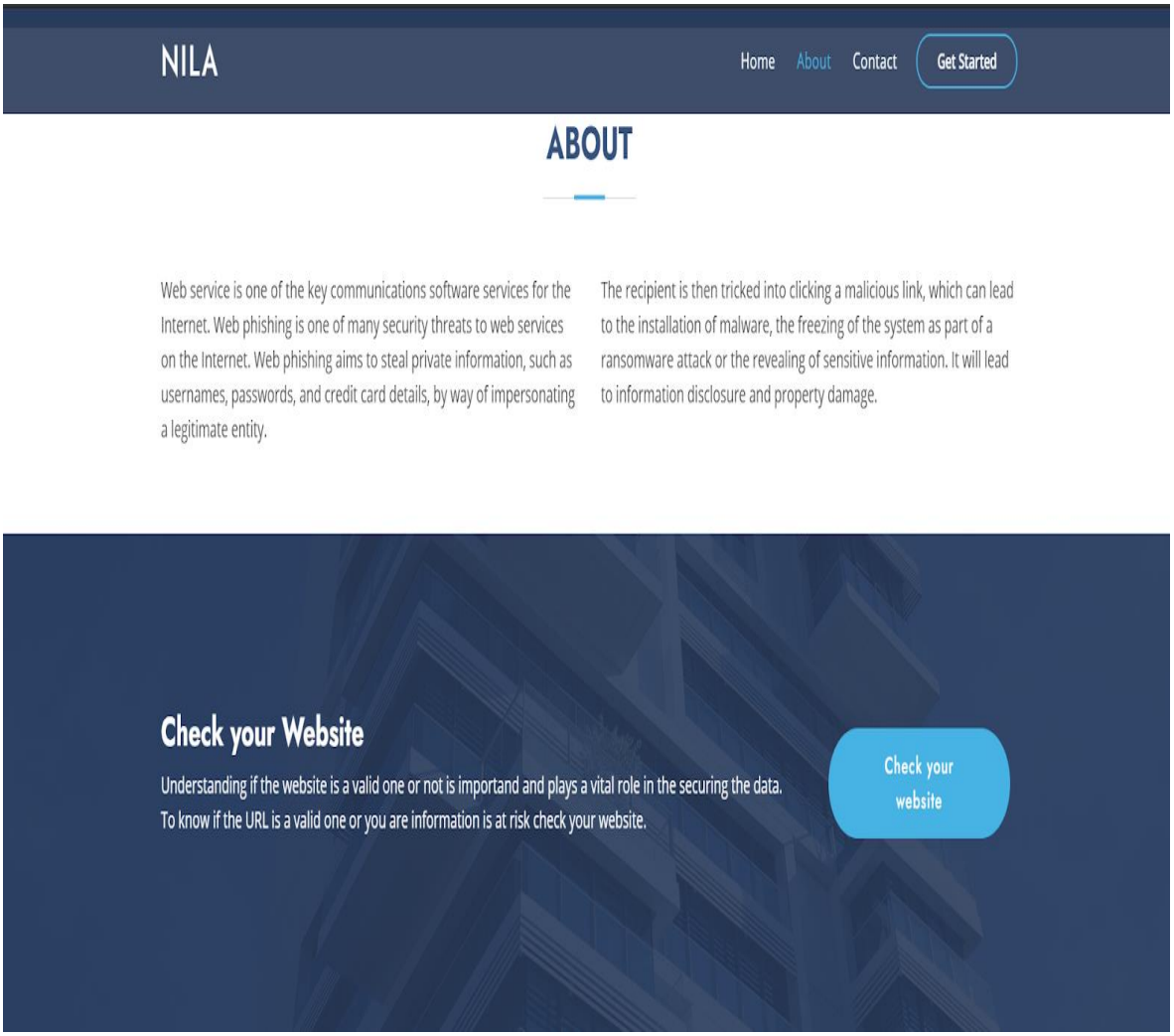
For Example: A Website home page often has standard links at the top of the page and often at the side of the page for items like,

- ✓ HOME
- ✓ ABOUT
- ✓ CONTACT
- ✓ GET STARTED

The home page as the largest job of all pages. It must let the viewers know they have arrived at the correct place. It needs to provide current customers with quick contact information. It must also help build upon and start the research stage.

The page that shows up by default when someone accesses your URL from their browser. Usually index.html, index.php etc….

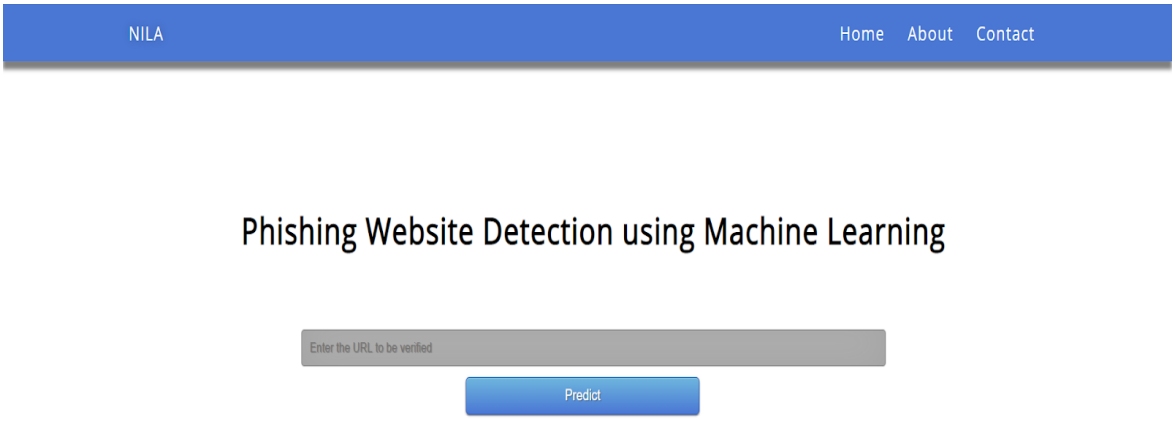
7.2 Feature 2



ABOUT PAGE:

An About page is a special web page on a site where your readers/visitors learn more about you and what you do. This is not a contact us page. Writing this page isn't the easiest thing to master, but it's possible once you understand the essential elements that must be included. The primary purpose of an about us page is to inform the reader about the company and its operations. This is a straightforward goal that nearly all businesses have to fulfill in some fashion or another. However, there are other reasons why about us pages are common fixtures on business websites

7.3 Database Schema

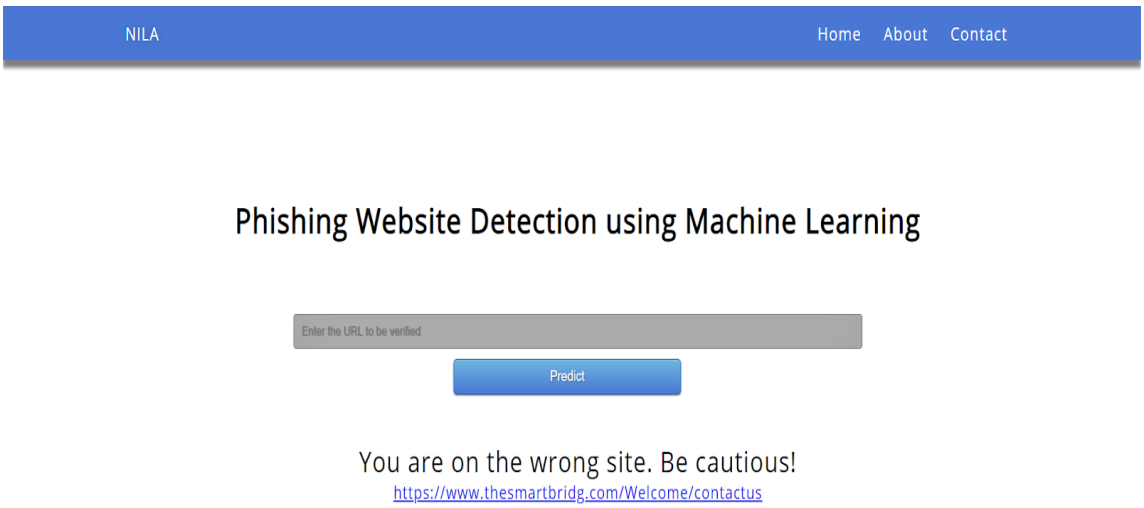


Predictive research is chiefly concerned with forecasting (predicting) outcomes, consequences, costs, or effects. This type of research tries to extrapolate from the analysis of existing phenomena, policies, or other entities in order to predict something that has not been tried, tested, or proposed before.

Machine learning model predictions allow businesses to make highly accurate guesses as to the likely outcomes of a question based on historical data, which can be about all kinds of things – customer churn likelihood, possible fraudulent activity, and more.

These techniques can provide managers and executives with decision-making tools to influence upselling, sales and revenue forecasting, manufacturing optimization, and even new product development.

FINAL PAGE:



Good readers make predictions as they read, to help them deepen their thinking and better comprehend what they read. Predicting is when readers use text clues and their own personal experiences, to anticipate what is going to happen next in the story.

The key notable points of our initial work embed: Phishing sites and their domains reveal the features that are different from other sites and domains. (For example, Google; www.google.com and some random phishing website be like; www.googlee.com). Phishing Uniform Resource Locators and 'domain names' typically have a different length when compared to other websites and domain names.

8. TESTING

8.1 Test Cases

The goal of a phishing test is to educate users of phishing dangers and to reduce the risk of hackers gaining access to sensitive information. Phish testing is a program that lets organizations send a realistic but fake phishing email to employees in order to see how they respond. Phish testing is used to gauge the effectiveness of phishing training programs that are designed to help employees spot phishing emails and to handle them appropriately.

For example, a test scenario might be, “Verify login functionality.” Test scenarios typically have their own ID numbers for tracking. QA teams often derive test cases (low-level actions) from test scenarios (high-level actions); and test scenarios typically come from software and business requirements documentation.

The purpose of a test case is to determine if different features within a system are performing as expected and to confirm that the system satisfies all related standards, guidelines and customer requirements. The process of writing a test case can also help reveal errors or defects within the system.

A Test case is to ensure if different features within an application are working as expected. It helps the tester to validate if the application is free of defects and if it is working as per the expectations of the end-users. A tester or QA professional typically writes test cases, which are run after the completion of a feature or the group of features that make up the release. Test cases also confirm whether the product meets its software requirement

8.2 User Acceptance Testing

User Acceptance Testing (UAT), which is performed on most UIT projects, sometimes called beta testing or end-user testing, is a phase of software development in which the software is tested in the "real world" by the intended audience or business representative.

The User methodology where the developed software is tested by the business user to validate if the software is working as per the specifications defined.

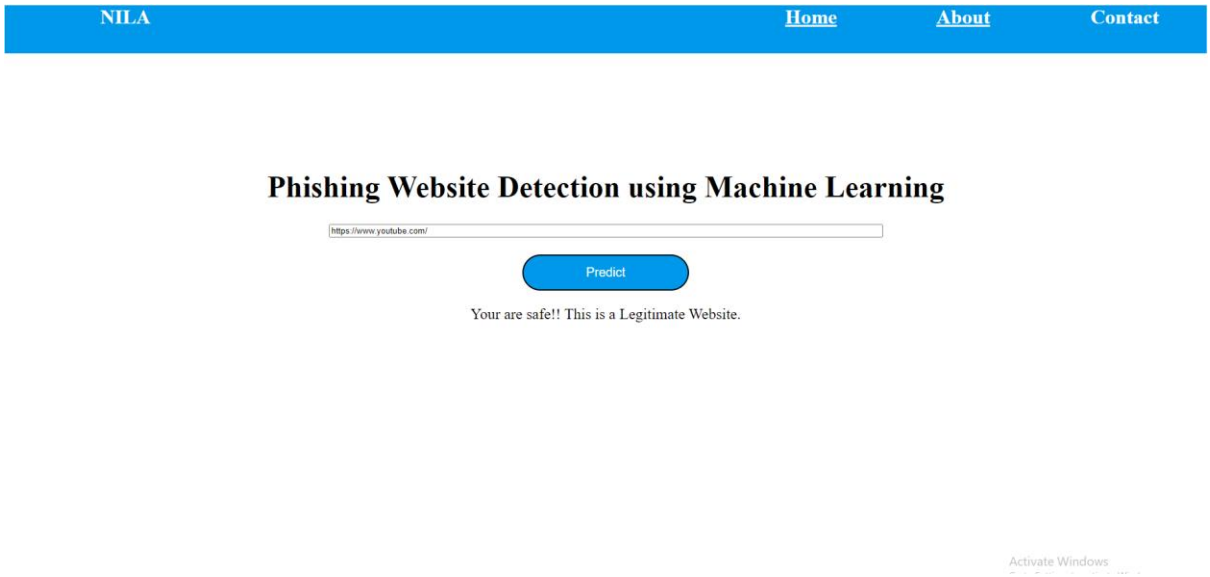
Need of User Acceptance Testing arises once software has undergone Unit, Integration and System testing because developers might have built software based on requirements document by their own understanding and further required changes during development may not be effectively communicated to them, so for testing whether the final product is accepted by client/end-user, user acceptance testing is needed.

User acceptance testing, a testing methodology where the client's /end users involved in testing the product to validate the product against their requirements. It is performed at client location at developer's site.

UAT Tester should possess good knowledge of the business. He should be independent and think as an unknown user to the system. Tester should be Analytical and Lateral thinker and combine all sort of data to make the UAT successful. Software based on requirements document which is their "own" understanding of the requirements and may not actually be what the client needs from the software.

9. RESULTS

9.1 Performance Metrics



Performance metrics are used to measure the behavior, activities, and performance of a business. This should be in the form of data that measures required data within a range, allowing a basis to be formed supporting the achievement of overall business goals.

By measuring and analyzing the KPIs, it becomes easier to determine what areas of your website are functioning better than others. Make sure you are aware of some common KPI warning signs. With the right website performance tools and/or team of professionals, tracking and analyzing your data is simple.

Web performance is important **for accessibility and also for other website metrics that serve the goals of an organization or business**. Good or bad website performance correlates powerfully to user experience, as well as the overall effectiveness of most sites. This is why you should care about web performance.

10. ADVANTAGES

- This system can be used by many E-commerce or other websites in order to have good customer relationship.
- User can make online payment securely.
- Data mining algorithm used in this system provides better performance as compared to other traditional classifications algorithms.
- With the help of this system user can also purchase products online without any hesitation.
- Measure the degrees of corporate and employee vulnerability.
- Eliminate the cyber threat risk level.
- Increase user alertness to phishing risks.
- Instill a cyber security culture and create cyber security heroes.

DISADVANTAGES

In spite of its advantages, there will always be cases where old-fashioned manual reviews will be preferable

Less control: This is especially true of black box machine learning engines, which can make mistakes without anyone noticing them

False positives: If a legitimate action is marked as fraud and you don't realize it, it will influence the whole system negatively. In that sense, a badly calibrated machine learning engine can create a negative loop where the more false positives aren't flagged, the less precise your results will be in the future.

No human understanding: If you're trying to get to the bottom of understanding why a user action is suspicious, it's hard to beat good old psychology.

11. CONCLUSION

Phishing is growing continuously irrespective of intelligence security development, there is definitely need of special care toward safeguarding of people being cheated. In this work different machine learning algorithms have been compared on phishing dataset and found that random forest works better in terms of accuracy, error rate and other parameters. The proposed algorithm has also been compared with existing similar works as and it has been found that the purposed model achieves considerably higher accuracy as compared to works reported by different author's. The proposed algorithm has also been compared with work on different dataset with similar algorithm and the results show that the proposed model achieves considerably better accuracy as compared to works reported by different authors.

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.

12. FUTURE SCOPE

Further enhancement of this work could be use of more advanced algorithms with 10-fold cross validation. Feature selection method can also be varied to see the effect on the varies parameters.

A combination or hybrid machine learning algorithm can also be implemented to improve success rate and minimize false rate. One of the challenges faced by our research was the unavailability of reliable training datasets.

In fact, this challenge faces any researcher in the field. However, although plenty of articles about predicting phishing websites using data mining techniques have been disseminated these days, no reliable training dataset has been published publically, maybe because there is no agreement in literature on the definitive features that characterize phishing websites, hence it is difficult to shape a dataset that covers all possible features.

In this article, we shed light on the important features that have proved to be sound and effective in predicting phishing websites. In addition, we proposed some new features, experimentally assign new rules to some well-known features and update some other features.

We have implemented python program to extract features from URL. Below are the features that we have extracted for detection of phishing URLs.

13. APPENDIX

Source code:

Index.html:

```
<html>
<head>
<title>Phishing Website Detection</title>
<link rel="stylesheet"
 href="{{ url_for('static', filename = 'style.css') }}">
</head>
 <body >
 <div id="blue">
 <h1 id="title">NILA<span style="float:right">
<a href="\">Home</a>
<a href="About">About</a>
<a>Contact</a>
<a href="\predict">Get Started</a></span></h1>
<br>
<img src="{{ url_for('static', filename = 'phishing image.jpg') }}" style="float:
right" width="700" height="500">
</img>
<h1 id="sub-t">Solution to Detect <br>Phishing Websites</h1>
<p>Be aware of what's happening with you <br>confidential data</p>
<span> <button style="font-size:20px" id="start">
<a href="\predict">Get Started</a>
</button>  <label for="video">Watch Video</label>
<button style="font-size:20px" id="Video"></button></span></div>
<div id="white">
<h1><center>About</center></h1>
 </div>
 <div>
 <div style="display: flex; justify-content: space-between;">
 <p style="background-color: white;">Web service is one of the key
communications software services for the Internet. Web phishing is one of many
security threats to web services on the Internet. Web phishing aims to steal private
information, such as usernames,passwords,and credit card details,by way of
impersonating a legitimate entity.</p>
 <p style="background-color: white;">The recipient is then tricked into clicking
a malicious link, which can lead to the installation of malware,the freezing of the
system as part of a ransomware attack or the revealing of sensitive information.It
will lead to information disclosure and property damage.</p>
 </div>
```

```
 </div>
 <div id="dark">
 <div id="div1">
 <p id="s1"><b>Check your Website</b><br>Understanding if the website is valid one or not is important and plays a vital role in the security of data.<br>To know if the URL is a valid one or you are information is at risk check your website.</p>
</div>
<div id="div2"><a id="stop" href="\predict">Check your website</a>
</div>
</div>
</div>
</body>
</html>
```

Final.html:

```
<html>
<head>
<link rel="stylesheet" href="{{ url_for('static', filename = 'style1.css') }}">
 </head>
 <body>
 <div id="blue">
 <h1 id="title">NILA<span style="float:right">
 <a href="\">Home</a>
 <a href="\">About</a>
 <a>Contact</a></span></h1>
 </div>
 <div id="white">
 <center>
 <form action="{{ url_for('y_predict') }}" method="post">
 <h1><b><center>Phishing Website Dete ction using Machine Learning</center></b></h1>
 <input type="text" id="url" name="url" value="{{ url_path }}" size="120"><br><br>
 <input type="submit" style="font-size:20px" id="start" value="Predict">
 </form>
 {{ url }}</center></div></body>
</html>
```

PYTHON CODE:

```
import numpy as np
import pickle
import inputScript
from flask import Flask, render_template, request
app = Flask(__name__)

model = pickle.load(open('phishing_website.pkl','rb'))

@app.route("/")
def index():
  return render_template("index.html")

@app.route("/predict")
def predict():
   return render_template("final.html")


@app.route("/y_predict", methods = ['GET', 'POST'])
def y_predict():
   geturl = request.form['url']
   check_prediction = inputScript.main(geturl)
   prediction = model.predict(check_prediction)
   print(prediction)
   output = prediction[0]
   if(output==-1):
      pred ='Your are safe!! This is a Legitimate Website.'
   else:
      pred = 'You are on the wrong site. Be cautions!'
   return render_template('final.html',url_path = geturl,url = pred)



if __name__ == '__main__':
  app.run(debug = True)
```

Github:

https://github.com/IBM-EPBL/IBM-Project-39600-1660463800

Project Demo Link:

https://github.com/IBM-EPBL/IBM-Project-39600-1660463800/blob/main/Final%20Deliverables/Project%20Demo.mp4