WEB PHISHING DETECTION

DONE BY:

Team leader: M.Nishanth
Team Member: R.Ruben
Team Member: S.Thangaraj
Team Member: M.Vijay Santhar

1.1 PROJECT OVERVIEW

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" e-mails to as many people as possible in an attempt to lure them into the scheme. When these e-mails 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.

Phishing has been accounted for many fraudulent incidents on the internet in the recent years, and it is showing no sign of stopping anytime soon. So, what is phishing? It is a term that is used to describe a malicious individual or a group of individuals who scam users. This is done by sending emails or creating web pages that are designed to collect an individual's online credentials, credit card details or other login information's. The concept of detecting phishing websites is usually done by looking through a huge database or a directory that contains all the malicious sites that has been logged by internet users or community members. An effective way for end users to benefit from phishing detection is by having the option to use an extension plugin that works on real time, as it gives them real time indication of what they are surfing and as well as if they are safe while browsing. With these issues in mind, and how it affects the security aspect of users on the internet and as well as giving concerns to privacy to the user, this research will be implemented as a google

chrome extension that can achieve the ability to do classification without needing a 3rd party server to do so.

1.2 PROJECT 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.

This research mainly will focus on implementing machine learning in JavaScript for it to run on a browser as an extension since JavaScript does not have much library support towards Machine Learning and also to keep in mind of the users' machines performance. This approach should be made with the intention of having it lite in order to achieve the capability to allow as much users as possible to use it. Random forest classifier for this project will be trained traditionally based on the phishing dataset 2 using Python scikit, and parameters of this model will then be exported in a JSON format to be used together with JavaScript.

2.1 Existing Problem

Cyber criminals use phishing emails because it's easy, cheap and effective. Email addresses are easy to obtain, and emails are virtually free to send. With little effort and cost, attackers can quickly gain access to valuable data. Those who fall for phishing scams may end up with malware infections (including ransomware), identity theft and data loss.

The data that cybercriminals go after includes personal identifiable information (PII)—like financial account data, credit card numbers and tax and medical records—as well as sensitive business data, such as customer names and contact information, proprietary product secrets and confidential communications.

Cybercriminals also use phishing attacks to gain direct access to email, social media and other accounts or to obtain permissions to modify and compromise connected systems, like point-of-sale terminals and order processing systems. Many of the biggest data breaches, like the headline-grabbing 2013 Target breach, start with a phishing email. By using a seemingly innocent email, cybercriminals can gain a small foothold and build on it.

in order to begin the development of a google chrome browser extension, it should emphasize on the ability to alert and warn the users if they accidentally visited a phishing webpage. This chrome extension will also be developed keeping in mind that, it should not have any 3rd party servers or API present to call services as this gives a narrow path for hackers to target users browsing pattern. Lastly, this extension plugin will also provide an instantaneous detection service that warns users as they view a phishing website, just so they avoid entering any confidential information before it is too late.

2.2 References

By : Ebubekir Bubar , Ajay , Christian charlotte.

Phishing is a form of fraud in which the attacker tries to learn sensitive information such as login credentials or account information by sending as a reputable entity or person in email or other communication channels.

Typically a victim receives a message that appears to have been sent by a known contact or organization. The message contains malicious software targeting the user's computer or has links to direct victims to malicious websites in order to trick them into divulging personal and financial information, such as passwords, account IDs or credit card details.

Phishing is popular among attackers, since it is easier to trick someone into clicking a malicious link which seems legitimate than trying to break through a computer's defense systems. The malicious links within the body of the message are designed to make it appear that they go to the spoofed organization using that organization's logos and other legitimate contents. In this article I explain: phishing domain (or Fraudulent Domain) characteristics, the features that distinguish them from legitimate domains, why it is important to detect these domains, and how they can be detected using machine learning and natural language processing techniques.

2.3 Problem Statement Definition

Problem Statement:

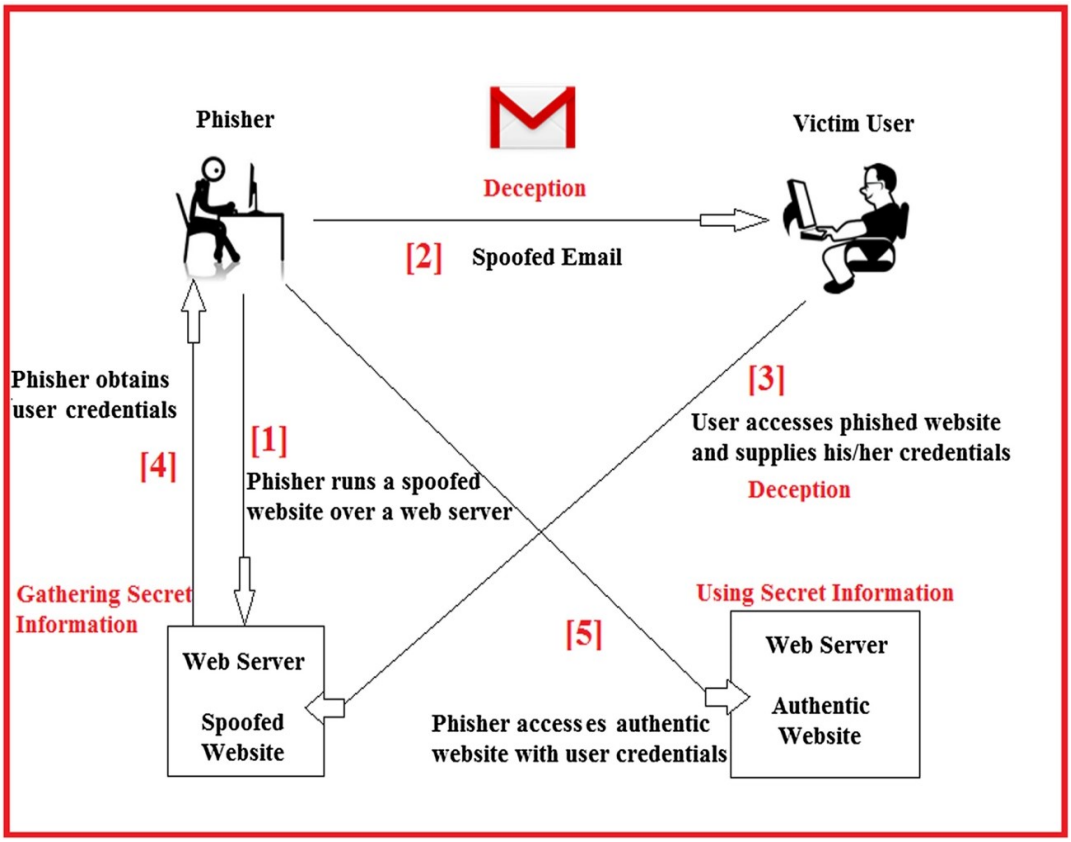
Internet is dominated as the world by dragging the half the population exponentially into the cyber world. with the booming of internet transactions, cybercrime rapidly increased and with anonymity presented by the internet, Hackers attempt to trap the end users through various forms like phishing.

Customers usually face problem with malicious link. These links will lead to a website that other steals login credential or financial information like credit card numbers. Scanning a URL may have expected on sequences, such as unsubscribing a user form a mailing list, so it is important that we limit these potential collateral damages.

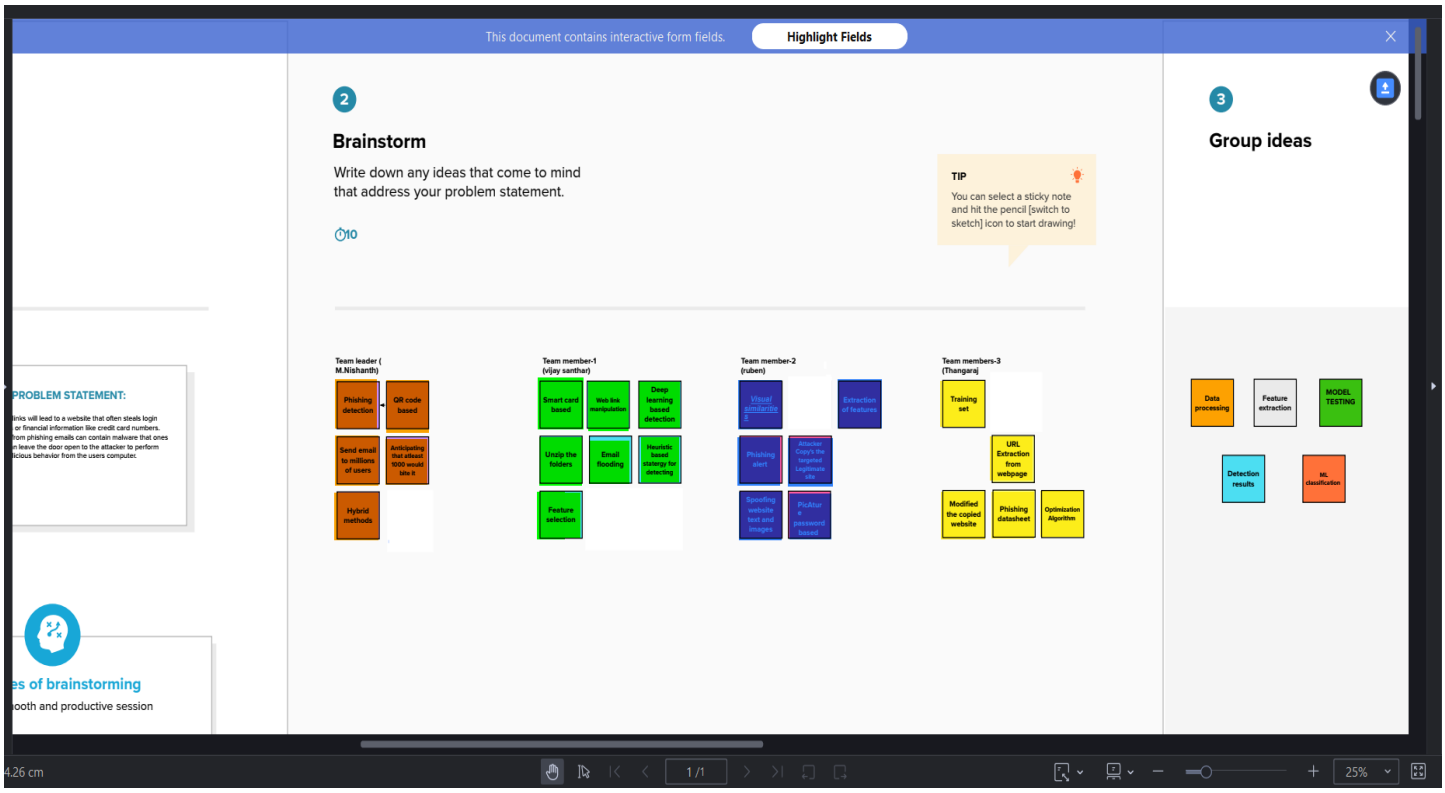
Phishing may lead to financial loss, emotional loss, stress, and information scam of customers.

I Am	Phishing can target the any sector and user, from a business executive to a home social network user or online banking consumer. Participants between the age of 18 and 25 are more susceptible to phishing then other age group.
I'M Trying to	• Never click any link or attachment in suspicious emails. • If the suspicious message appears to come from a person contact to the particular person to confirm the mail through the message or phone. • If it is suspicious message report the message. • And delete it
But	Phishing happens when a victim replies to a fraudulent email that demands urgent action. For example, the required action in a phishing website included: clicking an attachment. Enabling macro in Word document. Different types of phishing are Email Phishing, Spar Phishing, Whaling.
Because	Lack of security awareness among employees is also one of the major reasons for the success of phishing. Organizations should be aware of how the benefits and purpose of security awareness training can secure their employee from falling victim to phishing attacks
Which Make Me feel	Software-based phishing detection technique are preferred for fighting against the phishing attack. Mostly available method for detecting phishing attacks are blacklists/whitelists, natural language processing visual similarity, rules, machine learning techniques.

3.1 EMPATHY MAP CANVAS



3.2 Ideation and Brainstorm



3.3 Proposed Solution

Proposed Solution :

Project team shall fill the following information in proposed solution template.

S.No.	Parameter	Description
1.	Problem Statement (Problem to be solved)	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.
2.	Idea / Solution description	Anti-spyware and firewall settings should be used to prevent phishing attacks and users should Protect your mobile phone by setting software to update automatically. The website will be created with an opinion such that people are not only able to distinguish between legitimate and fraudulent websites, but also become aware of the mal-practices occurring in the current world.
3.	Novelty / Uniqueness	The website designed will be user friendly in means for any age. Easy to detect the fraudulent website and protect the sensitive credential information
4.	Social Impact / Customer Satisfaction	Feel protected by using the website as the business-related credentials will be safe. Parents can be relaxed when kids explore educational website as the fraudulent website will be detected by our website
5.	Business Model (Revenue Model)	This can be a efficient way to help banking sector as it secures the legitimate website from other malware that are set by hacker

6.	Scalability of the Solution	We would create an interactive and responsive website that will be used to detect whether a website is legitimate or phishing. This website is made using different web designing languages which include HTML, CSS, JavaScript and Python. This website is more useful to the user and it is user friendly also.

3.4 Proposed Solution Fit

Define CS, fit into CC

1. CUSTOMER SEGMENT(S) — CS

Who is your customer?

- Protect yourself and your family against malicious websites with the platform for free.
- With the platform, protecting your staff, data, brand, and your customer from malicious websites has never been easier.
- Proactively protect multiple customers against malicious websites at once with all-in-one platform.
- The platform can be used for government embeds to provide 100% security and privacy.

6. CUSTOMER CONSTRAINTS — CC

What constraints prevent your customers from taking action or limit their choices of solutions?

- The limitations of the web phishing detection approaches are explored by means of detection time, detection rate, and storage complexity to verify the level of robustness against the phishing attack.
- Thus most of the recent web phishing detection approaches lag in feature selection mechanism as they use handcrafted features to detect the attack.

5. AVAILABLE SOLUTIONS — AS

Which solutions are available to the customers when they face the problem or need to get the job done? What have they tried in the past? What pros & cons do these solutions have?

- Legitimate websites prevent web scraping by several techniques in respect to obfuscation using CSS sprites to display important data, replacing text with images.
- Spam filtering techniques are used to identify unsolicited emails before the user reads or clicks the link.
- When users visit a phishing web page that looks like a legitimate website, many people do not remember the legitimate website's domain name, particularly for some start-ups or unknown companies, so users cannot recognise the phishing website based on the URL. Some web browsers integrate a security component to detect phishing or malware sites, such as Chrome, which will display warning messages when one visits an unsafe web page.
- When the website detects that the IP address and device information of the user who is logging in does not match the commonly used information, it is necessary to verify the authenticity of the user

Explore AS, differentiate

Focus on J&P, tap into BE, understand RC

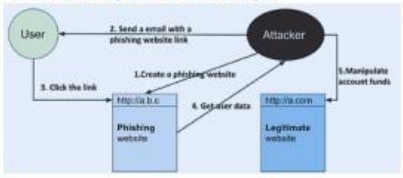
2. JOBS-TO-BE-DONE / PROBLEMS — J&P

Which jobs-to-be-done (or problems) do you address for your customers? There could be more than one; explore different sides.

- The objective of phishing website URLs is to purloin the personal information like user name, passwords and online banking transactions.
- Phishers use the websites which are visually and semantically similar to those real websites.
- As technology continues to grow, phishing techniques started to progress rapidly and this needs to be prevented by using anti-phishing mechanisms to detect phishing.

9. PROBLEM ROOT CAUSE — RC

What is the real reason that this problem exists?
What is the back story behind the need to do this job?



- A phishing attack is a type of cybersecurity threat that targets users directly through email, text or direct messages. During one of these scams, a cybercriminal will pose as a trusted contact to steal data from an unsuspecting user such as login information, account numbers and credit card information.
- While there are several types of phishing, the main purpose behind all of them is it to steal sensitive information or transfer malware.

7. BEHAVIOUR — BE

What does your customer do to address the problem and get the job done?
i.e. directly related: find the right solar panel installer, calculate usage and benefits; indirectly associated: customers spend free time on volunteering work

- Customers should take a "trust no one" approach when opening email.
- Check and verify the "From" address of the email.
- By carefully reading the email copy, users can typically spot something that seems "off" including:
- An email with an "urgent" request or An email offering the user something that's "too good to be true".
- Check grammar and spelling. Poor grammar and misspelled words in an email can be red flags.
- Be wary of generic salutations in an email. Legitimate companies, especially those with which you have accounts or have done business typically will address you by name versus by a generic greeting,
- Encourage your clients to look for any unusual or odd requests in their emails. Most fraudulent emails contain a request to respond to the email or click a link in it.
- Avoid clicking links or attachments in emails from unfamiliar sources.

Focus on J&P, tap into BE, understand RC

Identify strong TR & EM

3. TRIGGERS — TR

What triggers customers to act?

- Your users lack security awareness .
- Criminals are (unsurprisingly) following the money .
- You're not performing sufficient due diligence .
- Low-cost phishing and ransomware tools are easy to get hold of .
- Malware is becoming more sophisticated .

4. EMOTIONS: BEFORE / AFTER — EM

How do customers feel when they face a problem or a job and afterwards?

- Greed- Clicking on fake successful messages.
- Urgency-Hackers use fake security alerts with exclamation marks.
- Helpfulness-Hackers and cyber criminals use major tragedies to appeal for help but they are only helping themselves.
- Fear- Emails that spread fear and phishing links go hand in hand.

10. YOUR SOLUTION — SL

If you are working on an existing business, write down your current solution first, fill in the canvas, and check how much it fits reality.
If you are working on a new business proposition, then keep it blank until you fill in the canvas and come up with a solution that fits within customer limitations,
solves a problem and matches customer behaviour.

- We would create an interactive and responsive website that will be used to detect whether a website is legitimate or phishing.
- This website is made using different web designing languages which include HTML, CSS, JavaScript and Python.
- This website is more useful to the user and it is user friendly also.

8. CHANNELS of BEHAVIOUR — CH

8.1 ONLINE
What kind of actions do customers take online? Extract online channels from #3

- Nothing teaches like experience. When employees click on a link or an attachment in a simulated phishing email, it's important to communicate to them that they have potentially put both themselves and the organisation at risk.

8.2 OFFLINE
What kind of actions do customers take offline? Extract offline channels from #7 and use them for customer development.

- Phishing awareness training starts with educating your employees on why phishing is harmful, and empowering them to detect and report phishing attempts.
- Simulated phishing campaigns reinforce employee training, and to understand risk and improve workforce resiliency as these can take many forms, such as mass phishing, spear phishing, and whaling.

Extract online & offline CH of BE

4.1 Functional Requirement

Functional Requirements:

Following are the functional requirements of the proposed solution.

FR No.	Functional Requirement (Epic)	Sub Requirement (Story / Sub-Task)
FR-1	User Registration	Registration through Form Registration through Gmail Registration through LinkedIn
FR-2	User Confirmation	Confirmation via Email Confirmation via OTP
FR-3	Model Building	Building various Machine Learning model to detect web phishing and compare them.
FR-4	Check URL	Get the URL from the user and display if the website is malicious or not
FR-5	Integration	Integrate the developed Machine Learning Model using Flask
FR-6	Alert Message	Notify the user through email or phone regarding the malicious website.

4.2 NON- Functional Requirement

Non-functional Requirements:

Following are the non-functional requirements of the proposed solution.

FR No.	Non-Functional Requirement	Description
NFR-1	**Usability**	Any URL must be accepted for detection
NFR-2	**Security**	Alert message must be send to the users to enable secure browsing.
NFR-3	**Reliability**	The Phishing websites must detected accurately and the results must be reliable
NFR-4	**Performance**	The performance and interface must be user friendly
NFR-5	**Availability**	Anyone must be able to register and login
NFR-6	**Scalability**	It must be able to handle increase in the number of users.

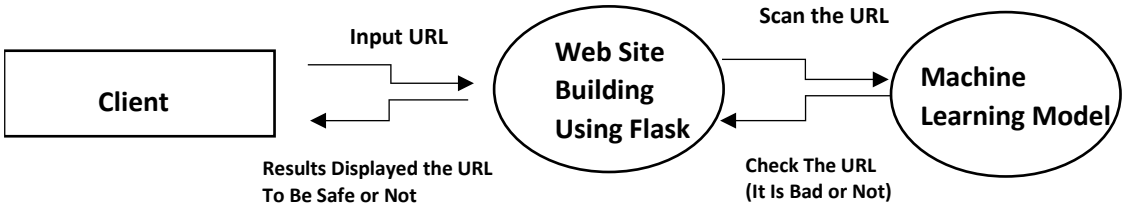
5.1 Data Flow Diagrams

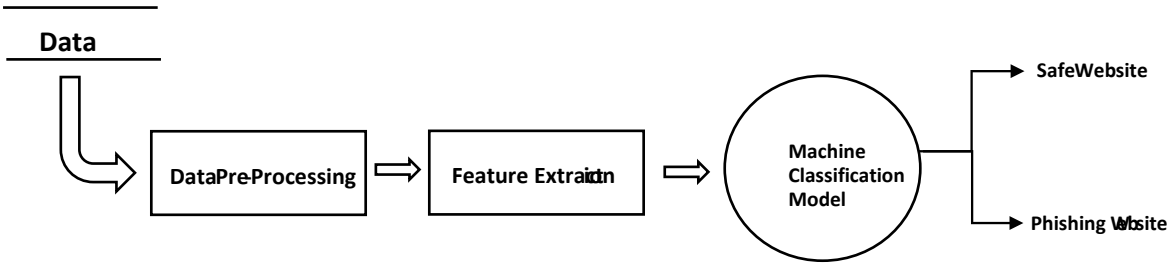
A Data Flow Diagram (DFD) is a traditional visual representation of the information flows within a system. A neat and clear DFD can depict the right amount of the system requirement graphically. It shows how data enters and leaves the system, what changes the information, and where data is stored.

Diagrams:

DFL – 0

DFL – 1





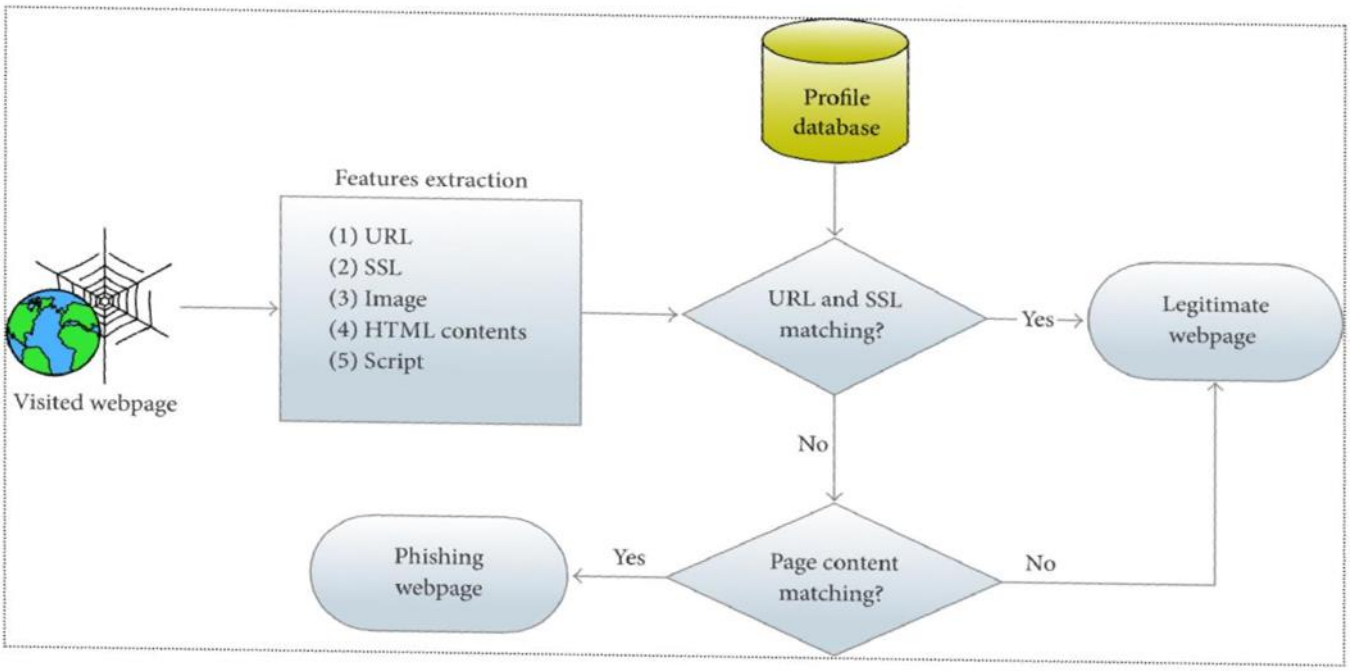
User Stories:

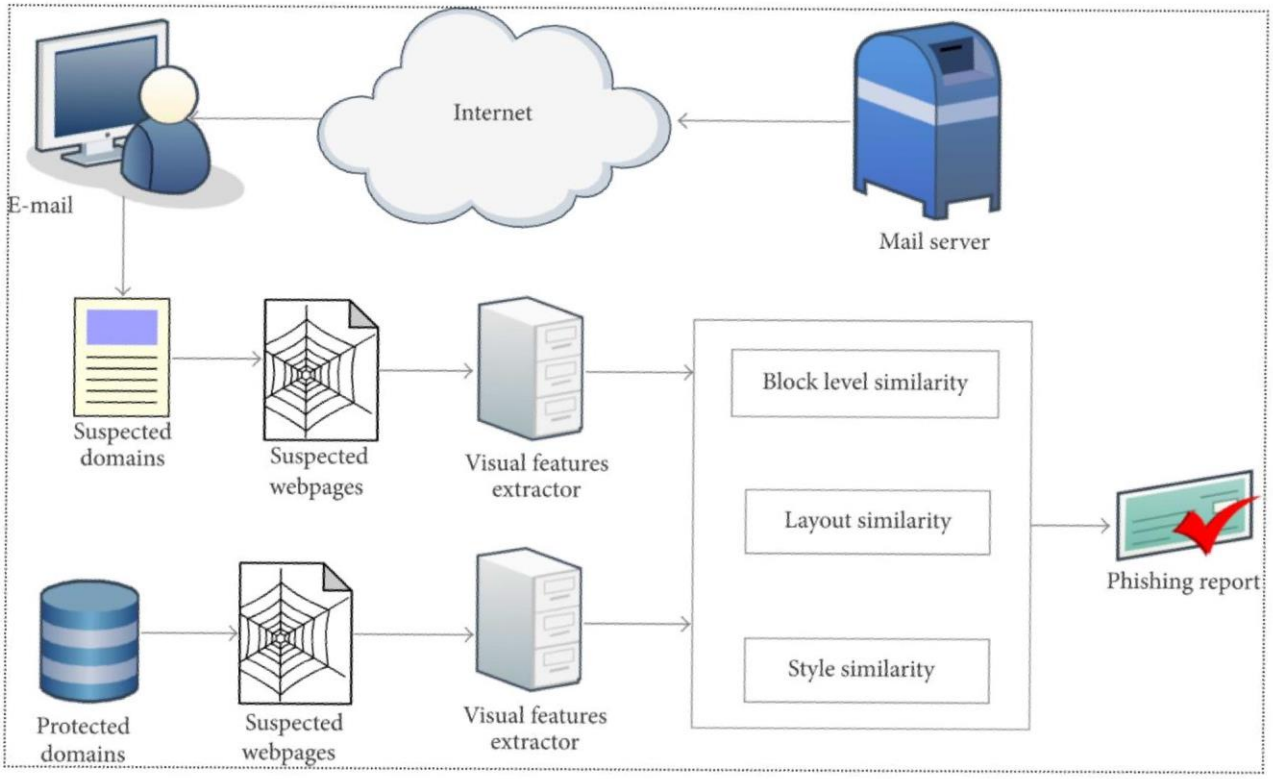
Use the below template to list all the user stories for the product.

User Type	Functional Requirement (Epic)	User Story Number	User Story / Task	Acceptation Criteria	Priority	Release
Customer (web user)	Website	USN – 1	Some, times the user need to check the website to check particular URL is safe or not	Website has easy to use and responsive.	High	Sprint-1
	Alter Notification	USN – 2	If enter into some malicious link, notification has to be sent to me	Received notification in mobile or to my mail ID	Low	Sprint-1
	Blocking	USN – 3	Whenever the link is not safe to enter, it should block to use to me.	I can register and access the dashboard with Facebook login.		Sprint-2
	Allowing	USN – 4	If I wish to use that website then it should also allow me to enter into that website			Sprint-1

	Login	USN – 5	As a user, I can log into the application by entering email & password	The phishing website has to be determined correctly.	High	Sprint-1
	DSHBOARD					
Customer (web view)	User input	USN - 1	As a user I can enter the required URL in the box while awaiting validation	I can access the website without any problem	High	
Customer care executive	Feature extraction	USN - 1	In the event that nothing is discovered during comparison, we can extract features	As a user I can have comparison between websites for security	High	
			using a heuristic and a visual similarity technique.			
administrator	Prediction	USN - 1	The model will use machine learning algorithms like a logistic regression and KNN model to forecast the URL of the websites.	I can accurately forecast the specific algorithms in this way	High	
	classifier	USN - 2	To create the final product. I will now feed all of the model output to classifier.	I will use this to identify the appropriate classifier for generating the outcome	Medium	Sprint-2

5.2 Solution And Technical Architecture





5.3 User Stories

User Type	Functional Requirement (Epic)	User Story Number	User Story / Task	Acceptation Criteria	Priority	Release
Customer (web user)	Website	USN – 1	Some, times the user need to check the website to check particular URL is safe or not	Website has easy to use and responsive.	High	Sprint-1
	Alter Notification	USN – 2	If enter into some malicious link, notification has to be sent to me	Received notification in mobile or to my mail ID	Low	Sprint-1
	Blocking	USN – 3	Whenever the link is not safe to enter, it should block to use to me.	I can register and access the dashboard with Facebook login.		Sprint-2
	Allowing	USN – 4	If I wish to use that website then it should also allow me to enter into that website			Sprint-1
	Login	USN – 5	As a user, I can log into the application by entering email & password	The phishing website has to be determined correctly.	High	Sprint-1
	DSHBOARD					
Customer (web view)	User input	USN - 1	As a user I can enter the required URL in the box while awaiting validation	I can access the website without any problem	High	

Customer care executive	Feature extraction	USN - 1	In the event that nothing is discovered during comparison, we can extract features	As a user I can have comparison between websites for security	High	
			using a heuristic and a visual similarity technique.			
administrator	Prediction	USN - 1	The model will use machine learning algorithms like a logistic regression and KNN model to forecast the URL of the websites.	I can accurately forecast the specific algorithms in this way	High	
	classifier	USN - 2	To create the final product. I will now feed all of the model output to classifier.	I will use this to identify the appropriate classifier for generating the outcome	Medium	Sprint-2

6.1 Sprint Planning And Estimtion

TITLE	DESCRIPTION	DATE
Literature Survey & Information Gathering	Literature survey on the selected project & gathering information by referring the, technical papers,research publications etc.	27 OCTOBER 2022
Prepare Empathy Map	Prepare Empathy Map Canvas to capture the user Pains & Gains, Prepare list of problem statements	27 OCTOBER 2022
Ideation	List the by organizing the brainstorming session and prioritize the top 3 ideas based on the feasibility & importance.	27 OCTOBER 2022
Proposed Solution	Prepare the proposed solution document, which includes the novelty, feasibility of idea, business model, social impact, scalability of solution, etc.	27 OCTOBER 2022
Problem Solution Fit	Prepare problem - solution fit document.	26 OCTOBER 2022
Solution Architecture	Prepare solution architecture document.	27 OCTOBER2022

Customer Journey	Prepare the customer journey maps to understand the user interactions & experiences with the application (entry to exit).	27 OCTOBER 2022
Functional Requirement	Prepare the functional requirement document.	8 OCTOBER 2022
Data Flow Diagrams	Draw the data flow diagrams and submit for review.	27 OCTOBER2022
Technology Architecture	Prepare the technology architecture diagram.	26 OCTOBER 2022
Prepare Milestone & Activity List	Prepare the milestones & activity list of the project.	not yet complete
Project Development - Delivery of Sprint-1, 2, 3 & 4	Develop & submit the developed code by testing it.	not yet completed

6.2 Sprint Delivery Schedule

Sprint schedule:

Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority	Team Members
Sprint-1	URL detector	USN-1	URL is the first thing to analyse a website to decide whether it is a phishing or not	10	High	M.Nishanth Ruben.R Vijay santhar.M Thangaraj.S
Sprint-1		USN-2	Some of URL-Based Features are • Digit count in the URL • Total length of URL • Checking whether the URL is typosquatted or not • Checking whether it includes a legitimate brand name or not • Number of subdomains in URL • TLD is one of the commonly used one	10	High	M.Nishanth Ruben.R Vijay santhar.M Thangaraj.S
Sprint-2	Domain detection	USN-3	The purpose of Phishing Domain Detection is detecting phishing domain names. Therefore, passive queries related to the domain name, which we want to classify as phishing or not, provide useful information to us.			

Sprint-2		USN-4	Some useful Domain-Based Features are • Its domain name or its IP address in blacklists of well-known reputation services? • How many days passed sincethe domain was registered? • Is the registrant name hidden?
Sprint-3	Page based features and Content based features	USN-5	Page-Based Features are using information about pages which are calculated reputation ranking services. Obtaining these types of features requires active scan to target domain. Page contents are processed for us to detect whether target domain is used for phishing or not
Sprint-3			• Global pagerank • Country pagerank • Position at the Alexa top 1 million site Some processed information about pages are • Page titles • Meta tags • Hidden text • Text in the body Images etc.
Sprint-4	Detection process	USN-6	Detecting Phishing Domains is a classification problem, so it means we need labeled data which has samples as phish domains and legitimate domains in the training phase

Sprint-2	Domain detection	USN-3	The purpose of Phishing Domain Detection is detecting phishing domain names. Therefore, passive queries related to the domain name, which we want to classify as phishing or not, provide useful information to us.	10	High	M.Nishanth Ruben.R Vijay santhar.M Thangaraj.S
Sprint-2		USN-4	Some useful Domain-Based Features are • Its domain name or its IP address in blacklists of well-known reputation services? • How many days passed sincethe domain was registered? • Is the registrant name hidden?	10	High	M.Nishanth Ruben.R Vijay santhar.M Thangaraj.S
Sprint-3	Page based features and Content based features	USN-5	Page-Based Features are using information about pages which are calculated reputation ranking services. Obtaining these types of features requires active scan to target domain. Page contents are processed for us to detect whether target domain is used for phishing or not	10	High	M.Nishanth Ruben.R Vijay santhar.M Thangaraj.S
Sprint-3			• Global pagerank • Country pagerank • Position at the Alexa top 1 million site Some processed information about pages are • Page titles • Meta tags • Hidden text • Text in the body Images etc.	10	High	M.Nishanth Ruben.R Vijay santhar.M Thangaraj.S

Sprint-4	Detection process	USN-6	Detecting Phishing Domains is a classification problem, so it means we need labeled data which has samples as phish domains and legitimate domains in the training phase	20	High	M.Nishanth Ruben.R Vijay santhar.M Thangaraj.S

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	10	29 Oct 2022
Sprint-2	20	6 Days	31 Oct 2022	05 Nov 2022	10	05 Nov 2022
Sprint-3	20	6 Days	07 Nov 2022	12 Nov 2022	10	12 Nov 2022
Sprint-4	20	6 Days	14 Nov 2022	19 Nov 2022	20	19 Nov 2022

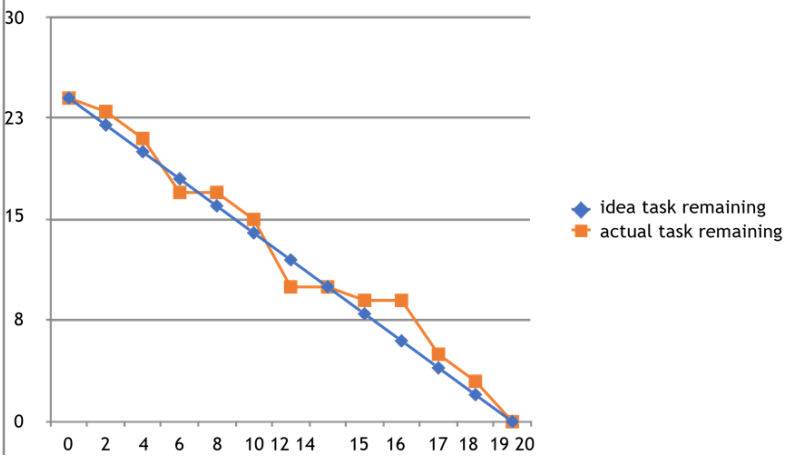
Velocity:

Imagine we have a 10-day sprint duration, and the velocity of the team is 20 (points per sprint). Let's calculate the team's average velocity (AV) per iteration unit (story points per day)

$$AV = \frac{sprint\ duration}{velocity} = \frac{20}{10} = 2$$

Burndown Chart:

A burn down chart is a graphical representation of work left to do versus time. It is often used in agile software development methodologies such as Scrum. However, burn down charts can be applied to any project containing measurable progress over time.



7.1 Feature 1

Detection:

Obtaining these types of features requires active scan to target domain. Page contents are processed for us to detect whether target domain is used for phishing or not. Some processed information about pages are given below.

- Page Titles
- Meta Tags
- Hidden Text
- Text in the Body
- Images etc.

By analysing these information, we can gather information such as;

- Is it required to login to website
- Website category
- Information about audience profile etc.

All of features explained above are useful for phishing domain detection. In some cases, it may not be useful to use some of these, so there are some limitations for using these features. For example, it may not be logical to use some of the features such as Content-Based Features for the developing fast detection mechanism which is able to analyze the number of domains between 100.000 and

200.000. Another example would be, if we want to analyze new registered domains Page-Based Features is not very useful. Therefore, the features that will be used by the detection mechanism depends on the purpose of the detection mechanism. Which features to use in the detection mechanism should be selected carefully.

CODE:

```
<!DOCTYPE html>
<html lang="en">

<head>
    <meta charset="UTF-8">
    <meta name="viewport" content="width=device-width, initial-scale=1.0">
    <title>Legitimate site Detector</title>
    <link rel="stylesheet" href="https://cdn.jsdelivr.net/npm/shorthandcss@1.1.1/dist/shorthand.min.css" />
    <link rel="stylesheet" href="https://fonts.googleapis.com/css?family=Muli:200,300,400,500,60 0,700,800,900&display=swap" />
    <link rel="stylesheet" type="text/css"
       href="https://cdnjs.cloudflare.com/ajax/libs/slick-carousel/1.9.0/slick.min.css" />
    <link rel="stylesheet" type="text/css" href="//cdn.jsdelivr.net/npm/slick-carousel@1.8.1/slick/slick-theme.css" />
</head>
```

```
<body class="bg-black muli">
   <nav class="w-100pc flex flex-column md-flex-row md-px-10 py-5 bg-black">
      <div class="flex justify-between">
         <a href="#" class="flex items-center p-2 mr-4 no-underline">
            <!-- <img class="max-h-l2 w-auto" src="assets/img/Logo-NVRT-50.jpg" /> -->
         </a>
         <a data-toggle="toggle-nav" data-target="#nav-items" href="#"
            class="flex items-center ml-auto md-hidden indigo-lighter opacity-50 hover-opacity-100 ease-300 p-1 m-3">
            <i data-feather="menu"></i>
         </a>
      </div>

   </nav>
```

This Above Code creates a web page with a url search box which is used to enter the url and check whether the given url is legit or not. Thus making the user aware of any illegal activites.

CODE:

```
<section id="home" class="min-h-100vh flex justify-start items-center">
    <div class="mx-5 md-mx-l5">
        <h1 class="white fs-l3 lh-2 md-fs-xl1 md-lh-1 fw-900 ">Check and Be safe from <br />Phishing Sites</h1>

        <div class="br-8 mt-10 inline-flex">
            <form action="/" method ="post">
            <input type="text"
                class="input-lg half bw-0 fw-200 bg-indigo-lightest-10 white ph-indigo-lightest focus-white opacity-80 fs-s3 py-5 min-w-25vw br-r-0"
                class="form__input" name ='url' id="url"
                placeholder="Enter URL" required="" />
            <button
                class="button-lg bg-indigo-lightest-20 indigo-lightest focus-white fw-300 fs-s3 mr-0 br-l-0" role="button">Detect</button>
            </form>
        </div>
        <div class="white opacity-20 fs-s3 mt-3">eg:https://amazon.global.com</span>
```

```
            </div>
            <iframe src='https://my.spline.design/untitled-81091ba6f7db7fba14e41a82633b20ca/' frameborder='0' width='100%' height='100%'></iframe></center>
            <div class="col-md" id="form2">


                <br>
                <h6 class = "right "><a href= {{ url }} target="_blank">{{ url }}</a></h6>


                <br>
                <h3 id="prediction"></h3>
            </div>
        </div>
    </section>


    <script>


        let x = '{{xx}}';
        let num = x*100;
        if (0<=x && x<0.50){
            num = 100-num;
        }
        let txtx = num.toString();
        if(x<=1 && x>=0.50){
```

```
        var label = "Website is "+txtx +"% safe to use...";
        document.getElementById("prediction").innerHTML
= label;

document.getElementById("button1").style.display="block";
    }
    else if (0<=x && x<0.50){
        var label = "Website is "+txtx +"% unsafe to use..."
        document.getElementById("prediction").innerHTML
= label ;

document.getElementById("button2").style.display="block";
    }

</script>
  </section>
```

7.2 Phishing for The phishing Sites:

By using our algorithm we try to verify for the authenticity of the websites by different means disscussed above.

Detecting Phishing Domains is a classification problem, so it means we need labeled data which has samples as phish domains and legitimate domains in the training phase. The dataset which will be used in the training phase is a very important point to build successful detection mechanism. We

have to use samples whose classes are precisely known. So it means, the samples which are labeled as phishing must be absolutely detected as phishing. Likewise the samples which are labeled as legitimate must be absolutely detected as legitimate. Otherwise, the system will not work correctly if we use samples that we are not sure about.

For this purpose, some public datasets are created for phishing. Some of the well-known one is PhishTank. These data sources are used commonly in academic studies.

Collecting legitimate domains is another problem. For this purpose, site reputation services are commonly used. These services analyse and rank available websites. This ranking may be global or may be country-based. Ranking mechanism depends on a wide variety of features. The websites which have high rank scores are legitimate sites which are used very frequently. One of the well-known reputation ranking service is Alexa. Researchers are using top lists of Alexa for legitimate sites.

When we have raw data for phishing and legitimate sites, the next step should be processing these data and extract meaningful information from it to detect fraudulent domains. The dataset to be used for machine learning must actually consist these features. So, we must process the raw data which is collected from Alexa, Phishtank or other data resources, and create a new dataset to train our system with machine learning algorithms. The feature values should be selected according to our needs and purposes and should be calculated for every one of them.

There so many machine learning algorithms and each algorithm has its own working mechanism. In this article, we have explained ***Decision Tree Algorithm***, because I think, this algorithm is a simple and powerful one.

Initially, as we mentioned above, phishing domain is one of the classification problem. So, this means we need labeled instances to build detection mechanism. In this problem we have two classes: **(1)** phishing and **(2)** legitimate.

When we calculate the features that we've selected our needs and purposes, our dataset looks like in figure below. In our examples, we selected 12 features, and we calculated them. Thus we generated a dataset which will be used in training phase of machine learning algorithm.

No.	1: domain	2: tld	3: brandName	4: editDbrandName	5: digitCount	6: length	7: isKnownTld	8: www	9: keywords	10: punnyCode	11: randomDomain	12:
	String	String	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric	
...	ayanasalon	com	0.0	1.0	0.0	10.0	0.0	0.0	0.0	0.0	0.0	
...	esteticabrasilbeauty	com	0.0	1.0	0.0	20.0	0.0	0.0	0.0	0.0	0.0	
...	erate365	com	0.0	1.0	3.0	8.0	0.0	0.0	0.0	0.0	0.0	
...	upstatescbusiness	com	1.0	1.0	0.0	17.0	0.0	0.0	1.0	0.0	0.0	
...	6-4c	com	0.0	0.0	2.0	4.0	0.0	0.0	0.0	0.0	0.0	
...	services-confirmatio...	com	1.0	1.0	0.0	23.0	0.0	0.0	1.0	0.0	0.0	
...	hmsinformatica	com	1.0	1.0	0.0	14.0	0.0	0.0	1.0	0.0	0.0	

A Decision Tree can be considered as an improved nested-if-else structure. Each features will be checked one by one. An example tree model is given below.

Generating a tree is the main structure of detection mechanism. Yellow and elliptical shaped ones represent features and these are called nodes. Green and angular ones represent classes and these are called leaves. The *length* is checked when an example arrives and then the other features are checked according to the result. When the journey of the samples is completed, the class that a sample belongs to will become clear.

NOTE: the above mentioned points a reference but slimly similar to vision in this project

CODE:

```
<section class="relative bg-indigo-lightest-10">
    <div id="slider-1">
```

<h2 class="white fs-l3 fw-900">About</h2>

 measurement for phishing detection is the

number of suspicious e-mails reported to

the security team. This measurement is designed to evaluate the number of employees who

followed the proper procedure for reporting suspicious messages.

<h2 class="white fs-l3 fw-900">Who are the Victims??</h2>

Those aged 25 to 44 years are most likely to be targeted, according to results from the Telephone-operated Crime Survey of England and Wales (TCSEW). Traditionally sent via email, phishing involves messages from fraudsters posing as legitimate organisations to extract personal information, or money, from the victim.

<h2 class="white fs-l3 fw-900">Why should you use this site</h2>

GPage-Based Features are using information about

pages which are calculated reputation ranking services. Some of these features give information about how much reliable a web site is.</p>

```
            </div>
        </div>
        <ul class="absolute list-none w-100pc flex justify-between top-50pc">
            <li><button
                    class="prev ml-10 br-round border-indigo-lightest indigo-lightest  bg-transparent flex justify-center items-center p-2 focus-indigo-lighter outline-none"><i
                        data-feather="chevron-left"></i></button></li>
            <li><button
                    class="next mr-10 br-round border-indigo-lightest indigo-lightest  bg-transparent flex justify-center items-center p-2  focus-indigo-lighter outline-none"><i
                        data-feather="chevron-right"></i></button></li>
        </ul>
    </section>


    <!-- big text -->
    <section class="p-10 md-py-10">
        <div class="w-100pc md-w-70pc mx-auto py-10">
            <h2 class="white fs-12 md-fs-xl1 fw-900 lh-2 ">We'll Detect the
```

```
            <span class="border-b bc-indigo bw-4"> Website
</span>
            for you.</h2>
        </div>
    </section>

    <!-- product options -->
    <section id="team" class="min-h-100vh flex justify-start
items-center">
    <section class="py-l10">
        <div class="flex flex-column md-flex-row md-w-80pc
mx-auto">
            <div class="w-100pc md-w-50pc">
                <div class="br-8 p-5 m-5 bg-indigo-lightest-10
pointer hover-scale-up-1 ease-300">
                    <div class="inline-block bg-indigo indigo-lightest
br-3 px-4 py-1 mb-10 fs-s4 uppercase">
                        Team Leader</div>
                    <div class="indigo-lightest fw-600 fs-
m1">Nishanth M<br><span class="opacity-30">
Reg.No:961219104035
                        <br><br>Loyola Institute of Technology and
Science.
                    </link></span> </div>

            </div>
        </div>
```

```
<div class="w-100pc md-w-50pc">
    <div class="br-8 p-5 m-5 bg-indigo-lightest-10 pointer hover-scale-up-1 ease-300">
        <div class="inline-block bg-indigo indigo-lightest br-3 px-4 py-1 mb-10 fs-s4 uppercase">
            Team Member</div>
        <div class="indigo-lightest fw-600 fs-m1">Vijay Santhar M<br><span class="opacity-30">
Reg.No:961219104055<br><br>Loyola Institute of Technology and Science</link></span> </div>

    </div>
</div>
<div class="w-100pc md-w-50pc">
    <div class="br-8 p-5 m-5 bg-indigo-lightest-10 pointer hover-scale-up-1 ease-300">
        <div class="inline-block bg-indigo indigo-lightest br-3 px-4 py-1 mb-10 fs-s4 uppercase">
            Team Member</div>
        <div class="indigo-lightest fw-600 fs-m1">Thangaraj S<br><span class="opacity-30">
Reg.No:9612191053<br><br>Loyola Institute of Technology and Science</link></span> </div>

    </div>
</div>
```

```
            <div class="w-100pc md-w-50pc">
                <div class="br-8 p-5 m-5 bg-indigo-lightest-10
pointer hover-scale-up-1 ease-300">
                    <div class="inline-block bg-indigo indigo-lightest
br-3 px-4 py-1 mb-10 fs-s4 uppercase">
                        Team Member</div>
                    <div class="indigo-lightest fw-600 fs-
m1">Ruban R<br><span class="opacity-30">
Reg.No:961219104042<br><br>Loyola Institute of
Technology and Science</link></span> </div>


                </div>
            </div>


        </div>
    </section>
</section>




    <!-- footer -->
    <footer class="p-5 md-p-l5 bg-indigo-lightest-10">
        <div class="flex flex-wrap">
            <div class="md-w-25pc mb-10">
```

```
            <!--<img src="assets/images/logo.png" class="w-
l5" alt="">-->
            <div class="white opacity-70 fs-s2 mt-4 md-pr-
10">
                <p>Copyright ©LSD 2022. <br>All Rights
Rseserved.<br>Team ID: PNT2022TMID51354</p>


            </div>
        </div>



    </footer>
        <i class="w-4" data-feather="download"></i>
    </div>
    <script src="https://code.jquery.com/jquery-
3.4.1.min.js"></script>
    <script src="https://unpkg.com/feather-icons"></script>
    <script src="https://cdnjs.cloudflare.com/ajax/libs/slick-
carousel/1.9.0/slick.min.js"></script>
    <script src="https://cdn.jsdelivr.net/gh/cferdinandi/smooth-
scroll@15.0.0/dist/smooth-scroll.polyfills.min.js"></script>
    <!--<script src="assets/js/script.js"></script>-->
</body>


</html>
```

8.1 Test Cases

The experimental system is structured to empirically test and verify the efficacy of the proposed methods for phishing website detection.

For training and evaluating the proposed techniques, three phishing datasets from the UCI repositories are used and the K-fold (where k = 10) cross-validation (CV) approach is used for the creation and evaluation of the phishing models.

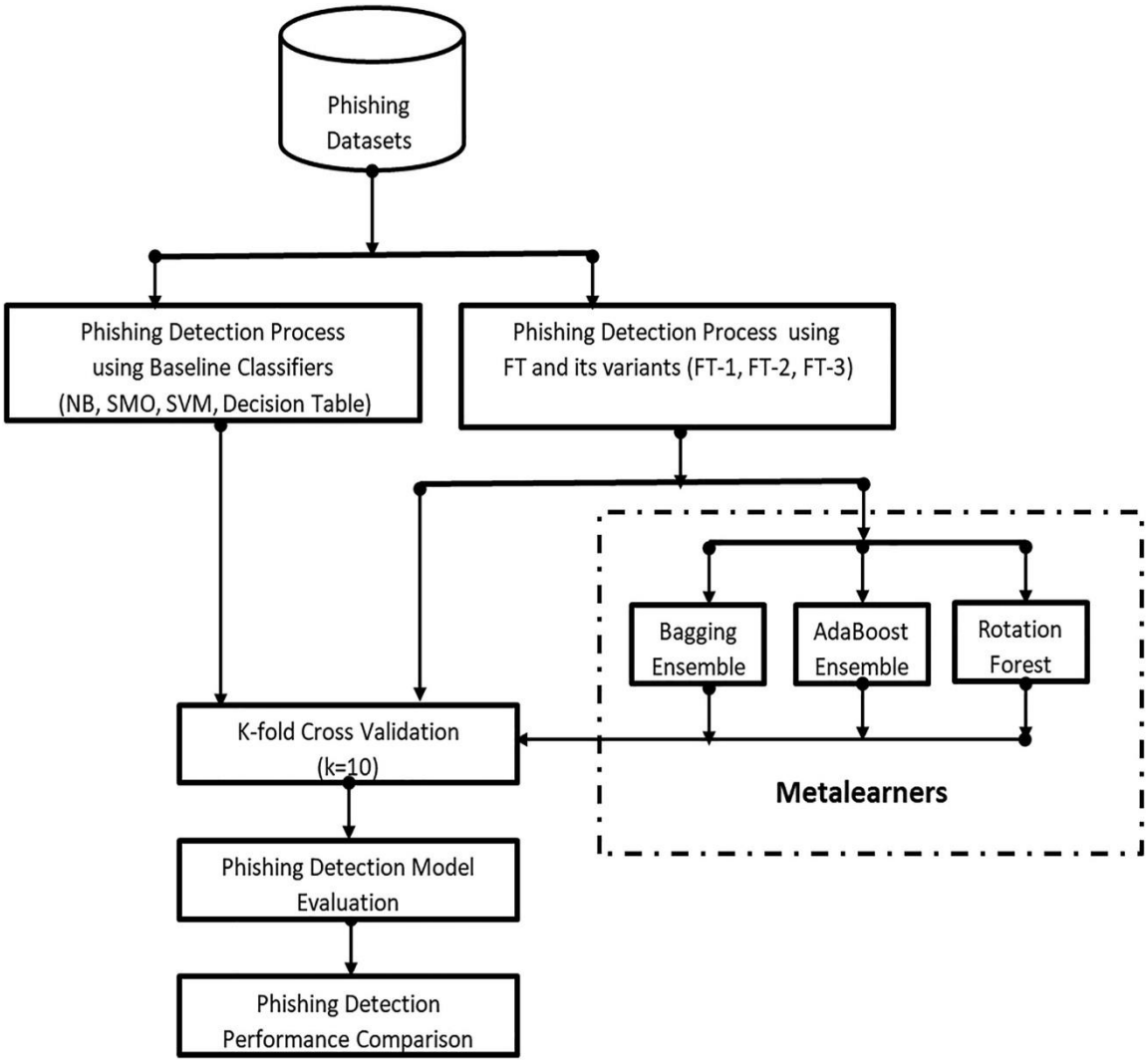
The 10-fold CV option is based on its ability to create phishing models with the low impact of the issue of class imbalance [49, 50, 51, 52, 53].

Moreover, the K-fold CV approach ensures that each instance can be used iteratively for both training and testing [54, 55, 56].

On phishing datasets, based on 10-fold CV, the proposed methods and the chosen baseline classifiers (NB, SMO, SVM, and Decision Table (Dec Table)) are then implemented.

The phishing detection efficiency of the developed phishing models is then tested and contrasted with other experimented methods of phishing detection.

All experiments were performed using the WEKA machine learning tool in the same environment [57].


Phishing
Datasets
Phishing Detection Process
using Baseline Classifiers
(NB, SMO, SVM, Decision Table)
Phishing Detection Process using
FT and its variants (FT-1, FT-2, FT-3)
Bagging
Ensemble
AdaBoost
Ensemble
Rotation
Forest
Metalearners
K-fold Cross Validation
(k=10)
Phishing Detection Model
Evaluation
Phishing Detection
Performance Comparison

8.2 User Acceptance Testing

.. Phishing attacks around the world cost billions of dollars in loss every year (Mc Combie et.al, 2009). Phishing has a huge negative impact on organizations' client relationships revenues, marketing pains and general corporate appearance (Dhanalakshmi et.al, 2011). Statistics report that 35.9% of financial sector is the target of Phishing frauds (APWG, 2010). ...

However these approaches are cannot detect and identify fresh phishes because of lists, where maintenance and human resources required and the scalability and run time are not suitable. This is the reason the list based approaches combine with other approaches [2, 3, 10, [12][13][14][15][16][17][18][19][20]. The Heuristics based approaches are predicted through one or more websites features like URL, source code and visual features.

9.1 Performance Matrix

Coding For Metrics:

```
import pandas as pd
import numpy as np
from sklearn.preprocessing import MinMaxScaler
from sklearn.metrics import confusion_matrix,accuracy_score
```

```
#Import Dataset
ds= pd.read_csv("dataset_website.csv")
ds.head()
```

```
from sklearn.linear_model import LogisticRegression
lr=LogisticRegression()
lr.fit(x_train,y_train)
```

```
y_pred1=lr.predict(x_test)
from sklearn.metrics import accuracy_score
log_reg=accuracy_score(y_test,y_pred1)
log_reg
```

```
0.9167797376752601
```

	index	having_IPhaving_IP_Address	URLURL_Length	Shortining_Service	having_At_Symbol	double_slash_redirecting	Prefix_Suffix	having_Sub_Domain
0	1	-1	1	1	1	-1	-1	-1
1	2	1	1	1	1	1	-1	0
2	3	1	0	1	1	1	-1	-1
3	4	1	0	1	1	1	-1	-1
4	5	1	0	-1	1	1	-1	1

5 rows × 32 columns

```
from sklearn.linear_model import LogisticRegression
lr=LogisticRegression()
lr.fit(x_train,y_train)
```

```
import pickle
pickle.dump(lr,open('Phishing_Website.pkl','wb'))
```

```
import numpy as np
from flask import Flask, request, jsonify, render_template
import pickle
#importing the inputScript file used to analyze the URL
import inputScript
```

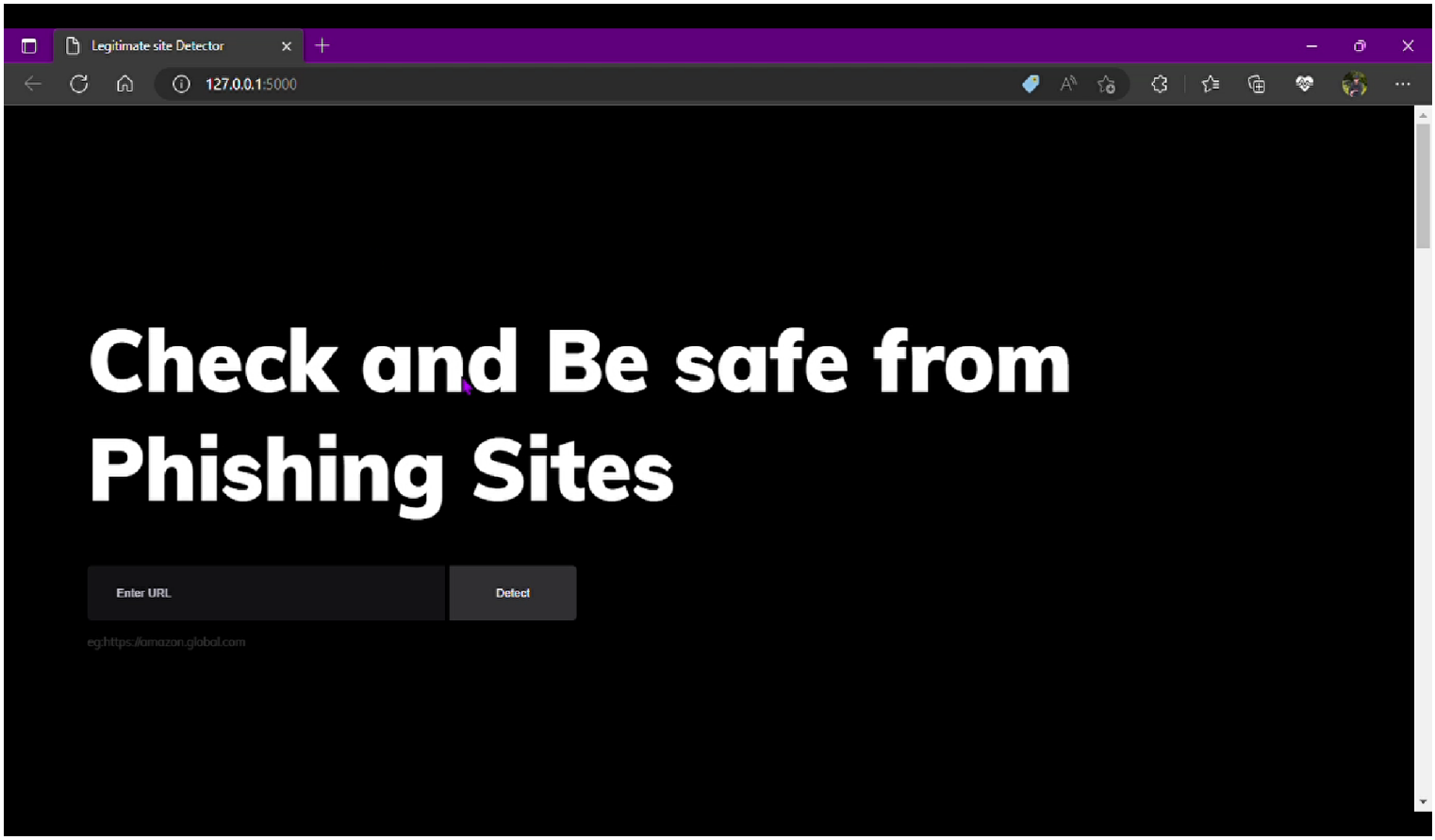
```
#load model
app = Flask(__name__)
model = pickle.load(open('Phishing_Website.pkl', 'rb'))

```

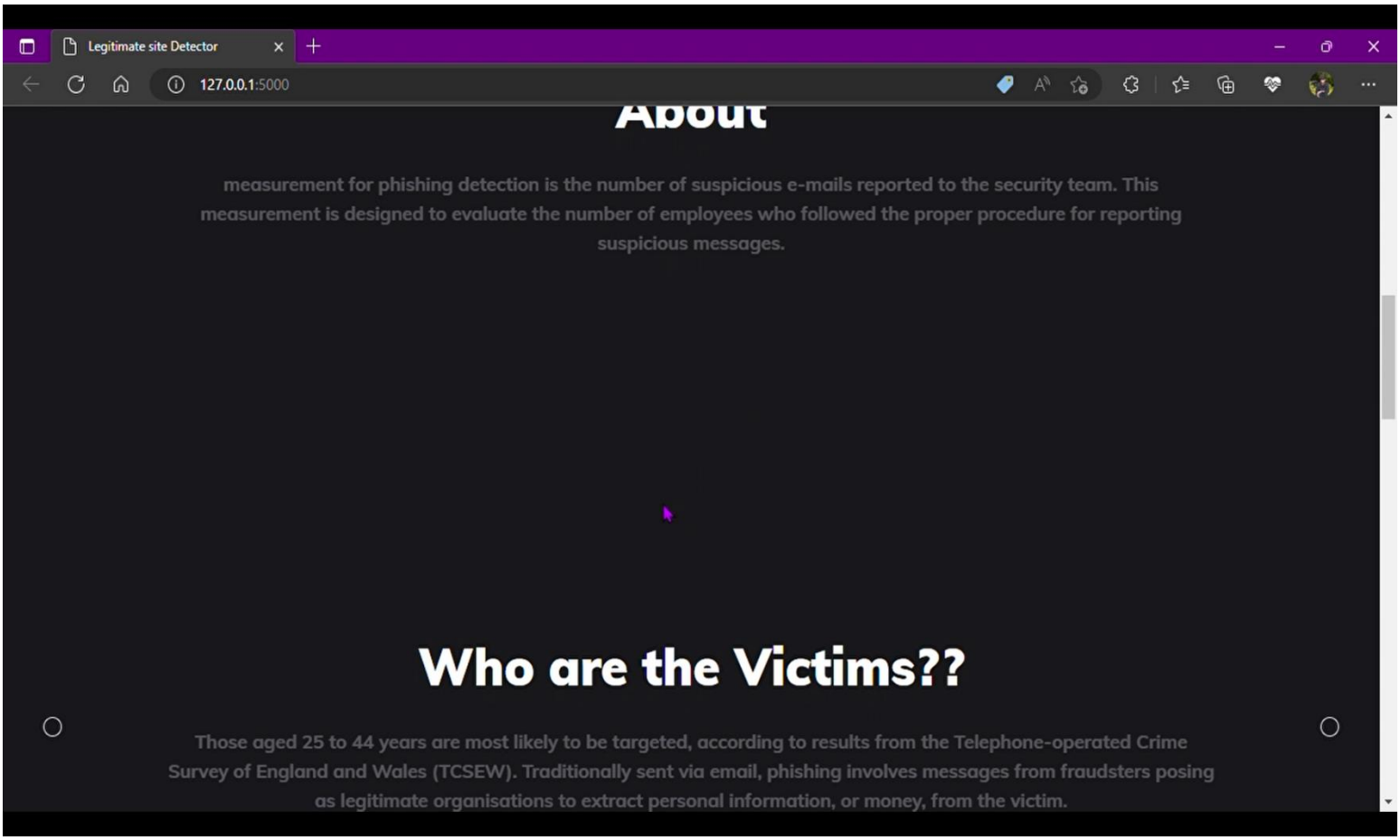
```python
#Redirects to the page to give the user iput URL.
@app.route('/predict')
def predict():
    return render_template('final.html')

#Fetches the URL given by the URL and passes to inputScript
@app.route('/y_predict',methods=['POST'])
def y_predict():
    '''
    For rendering results on HTML GUI
    '''
    url = request.form['URL']
    checkprediction = inputScript.main(url)
    prediction = model.predict(checkprediction)
    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 cautious!"
    return render_template('final.html', prediction_text='{}'.format(pred),url=url)

#Takes the input parameters fetched from the URL by inputScript and returns the predictions
@app.route('/predict_api',methods=['POST'])
def predict_api():
    '''
    For direct API calls trought request
    '''
    data = request.get_json(force=True)
    prediction = model.y_predict([np.array(list(data.values()))])

    output = prediction[0]
    return jsonify(output)

```

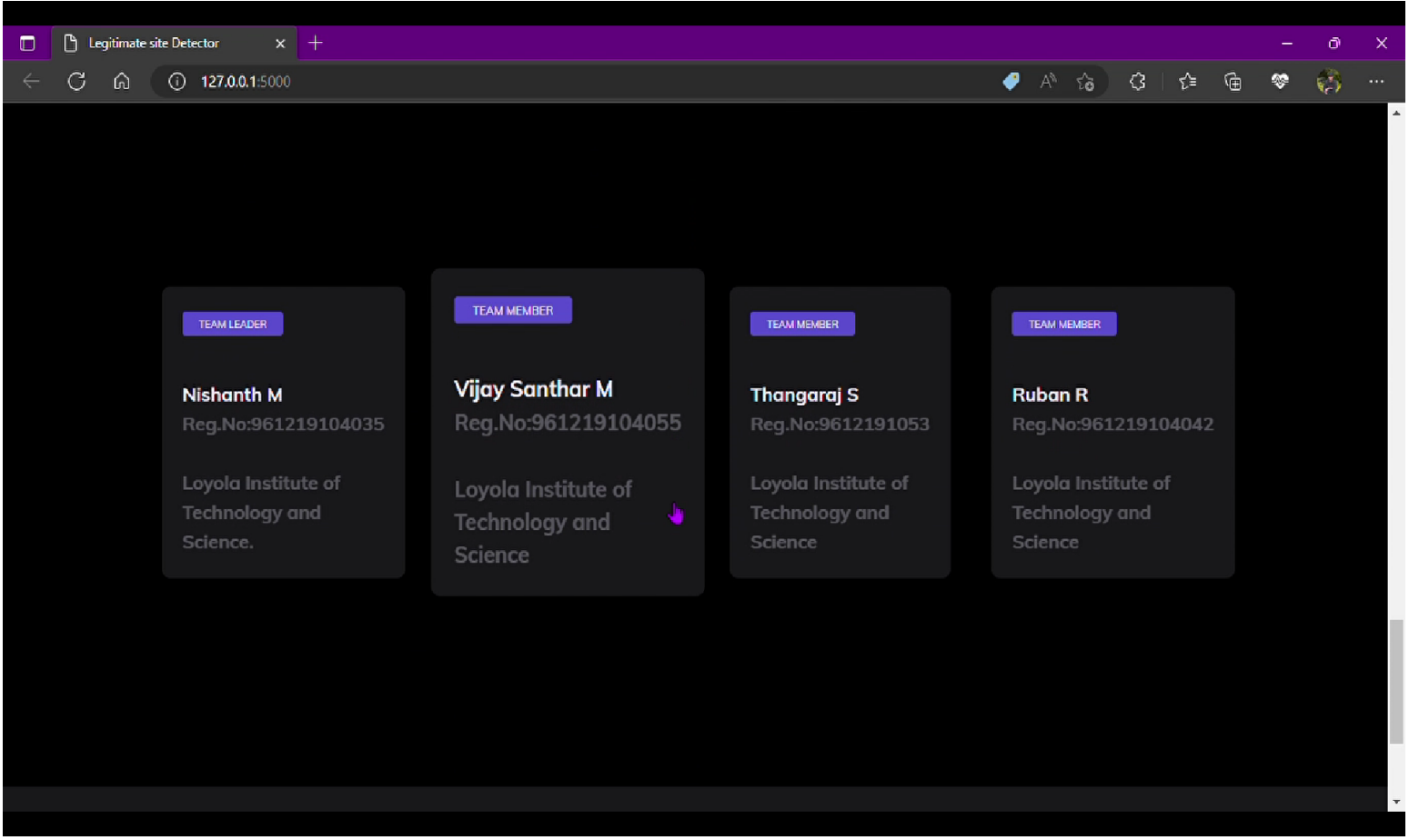
```python
if __name__ == '__main__':
    app.run(host='0.0.0.0', debug=True)
```

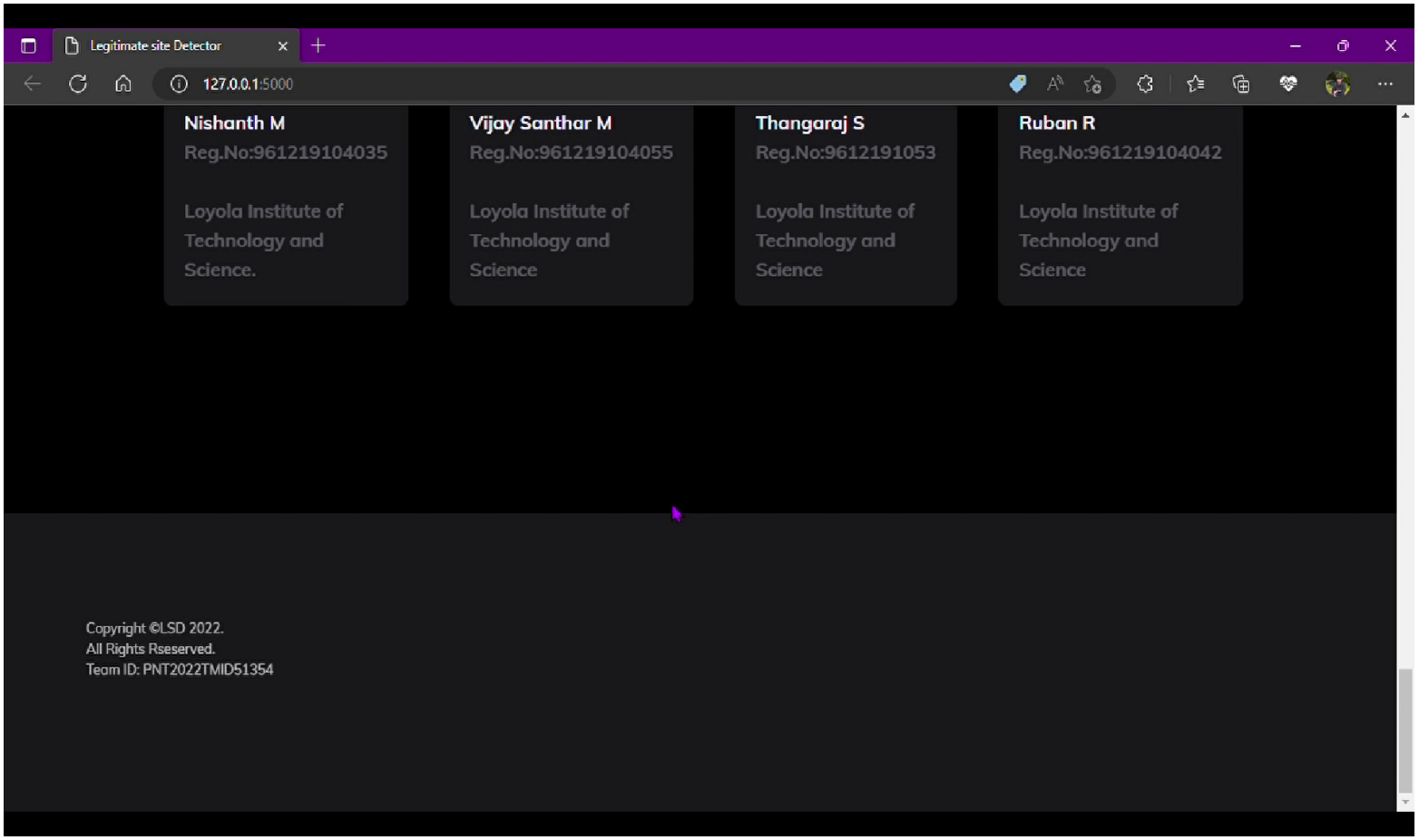
NOTE :The above mentioned code models are given by IBM.

Output:


Legitimate site Detector
127.0.0.1:5000
Check and Be safe from Phishing Sites
Enter URL
Detect
eg:https://amazon.global.com


Legitimate site Detector
127.0.0.1:5000
About
measurement for phishing detection is the number of suspicious e-mails reported to the security team. This measurement is designed to evaluate the number of employees who followed the proper procedure for reporting suspicious messages.
Who are the Victims??
Those aged 25 to 44 years are most likely to be targeted, according to results from the Telephone-operated Crime Survey of England and Wales (TCSEW). Traditionally sent via email, phishing involves messages from fraudsters posing as legitimate organisations to extract personal information, or money, from the victim.


Legitimate site Detector
127.0.0.1:5000
TEAM LEADER
Nishanth M
Reg.No:961219104035
Loyola Institute of Technology and Science.
TEAM MEMBER
Vijay Santhar M
Reg.No:961219104055
Loyola Institute of Technology and Science
TEAM MEMBER
Thangaraj S
Reg.No:9612191053
Loyola Institute of Technology and Science
TEAM MEMBER
Ruban R
Reg.No:961219104042
Loyola Institute of Technology and Science


Legitimate site Detector
127.0.0.1:5000
Nishanth M
Reg.No:961219104035
Loyola Institute of Technology and Science.
Vijay Santhar M
Reg.No:961219104055
Loyola Institute of Technology and Science
Thangaraj S
Reg.No:9612191053
Loyola Institute of Technology and Science
Ruban R
Reg.No:961219104042
Loyola Institute of Technology and Science
Copyright ©LSD 2022.
All Rights Rseserved.
Team ID: PNT2022TMID51354

The Above Output is From :

```
<!DOCTYPE html>
<html lang="en">

<head>
    <meta charset="UTF-8">
    <meta name="viewport" content="width=device-width,
initial-scale=1.0">
    <title>Legitimate site Detector</title>
    <link rel="stylesheet"
href="https://cdn.jsdelivr.net/npm/shorthandcss@1.1.1/dist/sh
orthand.min.css" />
    <link rel="stylesheet"
href="https://fonts.googleapis.com/css?family=Muli:200,300,
400,500,60 0,700,800,900&display=swap" />
    <link rel="stylesheet" type="text/css"
        href="https://cdnjs.cloudflare.com/ajax/libs/slick-
carousel/1.9.0/slick.min.css" />
    <link rel="stylesheet" type="text/css"
href="//cdn.jsdelivr.net/npm/slick-carousel@1.8.1/slick/slick-
theme.css" />
</head>

<body class="bg-black muli">
    <nav class="w-100pc flex flex-column md-flex-row md-
px-10 py-5 bg-black">
```

```
        <div class="flex justify-between">
            <a href="#" class="flex items-center p-2 mr-4 no-underline">
                <!-- <img class="max-h-l2 w-auto" src="assets/img/Logo-NVRT-50.jpg" /> -->
            </a>
            <a data-toggle="toggle-nav" data-target="#nav-items" href="#"
                class="flex items-center ml-auto md-hidden indigo-lighter opacity-50 hover-opacity-100 ease-300 p-1 m-3">
                <i data-feather="menu"></i>
            </a>
        </div>

    </nav>

    <!-- hero section -->
    <section id="home" class="min-h-100vh flex justify-start items-center">
        <div class="mx-5 md-mx-l5">
            <h1 class="white fs-l3 lh-2 md-fs-xl1 md-lh-1 fw-900 ">Check and Be safe from <br />Phishing Sites</h1>


            <div class="br-8 mt-10 inline-flex">
```

```
<form action="/" method ="post">
<input type="text"
class="input-lg half bw-0 fw-200 bg-indigo-lightest-10 white ph-indigo-lightest focus-white opacity-80 fs-s3 py-5 min-w-25vw br-r-0"
class="form__input" name ='url' id="url"
placeholder="Enter URL" required="" />
<button
class="button-lg bg-indigo-lightest-20 indigo-lightest focus-white fw-300 fs-s3 mr-0 br-l-0" role="button">Detect</button>
</form>
</div>
<div class="white opacity-20 fs-s3 mt-3">eg:https://amazon.global.com</span>
</div>
<iframe src='https://my.spline.design/untitled-81091ba6f7db7fba14e41a82633b20ca/' frameborder='0' width='100%' height='100%'></iframe></center>
<div class="col-md" id="form2">

<br>
<h6 class = "right "><a href= {{ url }} target="_blank">{{ url }}</a></h6>

<br>
<h3 id="prediction"></h3>
```

```
            </div>
        </div>
    </section>

    <script>

        let x = '{{xx}}';
        let num = x*100;
        if (0<=x && x<0.50){
            num = 100-num;
        }
        let txtx = num.toString();
        if(x<=1 && x>=0.50){
            var label = "Website is "+txtx +"% safe to use...";
            document.getElementById("prediction").innerHTML
= label;

document.getElementById("button1").style.display="block";
        }
        else if (0<=x && x<0.50){
            var label = "Website is "+txtx +"% unsafe to use..."
            document.getElementById("prediction").innerHTML
= label ;

document.getElementById("button2").style.display="block";
        }
```

```
</script>
    </section>


    <!-- slider -->
    <section class="relative bg-indigo-lightest-10">
        <div id="slider-1">
            <div class="p-10 md-p-l10 flex justify-center items-center flex-column text-center">
                <h2 class="white fs-l3 fw-900">About</h2>
                <p class="indigo-lightest fw-600 fs-m1 opacity-30 my-5"> measurement for phishing detection is the
                    number of suspicious e-mails reported to
                the security team. This measurement is designed to evaluate the number of employees who
                    followed the proper procedure for reporting suspicious messages.</p>
            </div>
            <div class="p-10 md-p-l10 flex justify-center items-center flex-column text-center">
                <h2 class="white fs-l3 fw-900">Who are the Victims??</h2>
                <p class="indigo-lightest fw-600 fs-m1 opacity-30 my-5">Those aged 25 to 44 years are most likely to be targeted, according to results from the Telephone-operated Crime Survey of England and Wales (TCSEW). Traditionally sent via email, phishing involves messages from fraudsters
```

posing as legitimate organisations to extract personal information, or money, from the victim.

<h2 class="white fs-l3 fw-900">Why should you use this site</h2>

GPage-Based Features are using information about pages which are calculated reputation ranking services. Some of these features give information about how much reliable a web site is.

<ul class="absolute list-none w-100pc flex justify-between top-50pc">

<button

class="prev ml-10 br-round border-indigo-lightest indigo-lightest bg-transparent flex justify-center items-center p-2 focus-indigo-lighter outline-none"><i

data-feather="chevron-left"></i></button>

<button

class="next mr-10 br-round border-indigo-lightest indigo-lightest bg-transparent flex justify-center items-center p-2 focus-indigo-lighter outline-none"><i

data-feather="chevron-right"></i></button>

```
</ul>
```

10 Advantages And Disadvantages

Advantages:

- Could Easily find the phishing websites before clicking on them
- Can Identify Which Sites Have The Most Target
- Could save your data before it is too late
- DETECTION is Better Than Cure.

Disadvantages:

- Hackers Find New Way To Attack,
- May not be able to detect all the websites.
- Could possibly bypass the detection.

11.Conclusion

The importance to safeguard online users from becoming victims of online fraud, divulging confidential information to an attacker among other effective uses of phishing as an attacker's tool, phishing detection tools play a vital role in ensuring a secure online experience for users. Unfortunately, many of the existing phishing-detection tools, especially those that depend on an existing blacklist, suffer limitations such as low detection accuracy and high false alarm that is often caused by either a delay in blacklist update as a result of human verification process involved in classification or perhaps, it can be attributed to human error in classification which may lead to improper classification of the classes. These critical issues have drawn many researchers to work on various approaches to improve detection accuracy of phishing attacks and to minimize false alarm rate. The inconsistent nature of attacks behaviors and continuously changing URL phish patterns require timely updating of the reference model. Therefore, it requires an effective technique to regulate retraining as to enable machine learning algorithm to actively adapt to the changes in phish patterns.

12.Future Scope

Despite there are several ways to carry out these attacks, unfortunately the current phishing detection techniques cover some attack vectors like email and fake websites. Therefore, building a specific limited scope detection system will not provide complete protection from the wide phishing attack vectors.

13.**Source Code**

HTML

```
<!DOCTYPE html>
<html lang="en">


<head>
    <meta charset="UTF-8">
    <meta name="viewport" content="width=device-width, initial-scale=1.0">
    <title>Legitimate site Detector</title>
    <link rel="stylesheet" href="https://cdn.jsdelivr.net/npm/shorthandcss@1.1.1/dist/shorthand.min.css" />
    <link rel="stylesheet" href="https://fonts.googleapis.com/css?family=Muli:200,300,400,500,60 0,700,800,900&display=swap" />
    <link rel="stylesheet" type="text/css"
        href="https://cdnjs.cloudflare.com/ajax/libs/slick-carousel/1.9.0/slick.min.css" />
    <link rel="stylesheet" type="text/css" href="//cdn.jsdelivr.net/npm/slick-carousel@1.8.1/slick/slick-theme.css" />
```

```
</head>


<body class="bg-black muli">
   <nav class="w-100pc flex flex-column md-flex-row md-px-10 py-5 bg-black">
      <div class="flex justify-between">
         <a href="#" class="flex items-center p-2 mr-4 no-underline">
            <!-- <img class="max-h-l2 w-auto" src="assets/img/Logo-NVRT-50.jpg" /> -->
         </a>
         <a data-toggle="toggle-nav" data-target="#nav-items" href="#"
            class="flex items-center ml-auto md-hidden indigo-lighter opacity-50 hover-opacity-100 ease-300 p-1 m-3">
            <i data-feather="menu"></i>
         </a>
      </div>

   </nav>

   <!-- hero section -->
   <section id="home" class="min-h-100vh flex justify-start items-center">
      <div class="mx-5 md-mx-l5">
         <h1 class="white fs-l3 lh-2 md-fs-xl1 md-lh-1 fw-900 ">Check and Be safe from <br />Phishing Sites</h1>
```

```
            <div class="br-8 mt-10 inline-flex">
                <form action="/" method ="post">
                <input type="text"
                    class="input-lg half bw-0 fw-200 bg-indigo-lightest-10 white ph-indigo-lightest focus-white opacity-80 fs-s3 py-5 min-w-25vw br-r-0"
                    class="form__input" name ='url' id="url"
                    placeholder="Enter URL" required="" />
                <button
                    class="button-lg bg-indigo-lightest-20 indigo-lightest focus-white fw-300 fs-s3 mr-0 br-l-0" role="button">Detect</button>
                </form>
            </div>
            <div class="white opacity-20 fs-s3 mt-3">eg:https://amazon.global.com</span>
        </div>
        <iframe src='https://my.spline.design/untitled-81091ba6f7db7fba14e41a82633b20ca/' frameborder='0' width='100%' height='100%'></iframe></center>
        <div class="col-md" id="form2">

            <br>
            <h6 class = "right "><a href= {{ url }} target="_blank">{{ url }}</a></h6>

            <br>
            <h3 id="prediction"></h3>
        </div>
    </div>
```

```
    </section>

    <script>

        let x = '{{xx}}';
        let num = x*100;
        if (0<=x && x<0.50){
            num = 100-num;
        }
        let txtx = num.toString();
        if(x<=1 && x>=0.50){
            var label = "Website is "+txtx +"% safe to use...";
            document.getElementById("prediction").innerHTML = label;
            document.getElementById("button1").style.display="block";
        }
        else if (0<=x && x<0.50){
            var label = "Website is "+txtx +"% unsafe to use..."
            document.getElementById("prediction").innerHTML = label ;
            document.getElementById("button2").style.display="block";
        }

</script>
    </section>

    <!-- slider -->
    <section class="relative bg-indigo-lightest-10">
        <div id="slider-1">
```

<h2 class="white fs-l3 fw-900">About</h2>

 measurement for phishing detection is the

number of suspicious e-mails reported to

the security team. This measurement is designed to evaluate the number of employees who

followed the proper procedure for reporting suspicious messages.

<h2 class="white fs-l3 fw-900">Who are the Victims??</h2>

Those aged 25 to 44 years are most likely to be targeted, according to results from the Telephone-operated Crime Survey of England and Wales (TCSEW). Traditionally sent via email, phishing involves messages from fraudsters posing as legitimate organisations to extract personal information, or money, from the victim.

<h2 class="white fs-l3 fw-900">Why should you use this site</h2>

GPage-Based Features are using information about pages which are calculated reputation ranking services. Some of these features give information about how much reliable a web site is.

```
        </div>
        <ul class="absolute list-none w-100pc flex justify-between top-50pc">
            <li><button
                    class="prev ml-10 br-round border-indigo-lightest indigo-lightest  bg-transparent flex justify-center items-center p-2 focus-indigo-lighter outline-none"><i
                        data-feather="chevron-left"></i></button></li>
            <li><button
                    class="next mr-10 br-round border-indigo-lightest indigo-lightest  bg-transparent flex justify-center items-center p-2 focus-indigo-lighter outline-none"><i
                        data-feather="chevron-right"></i></button></li>
        </ul>
    </section>


    <!-- big text -->
    <section class="p-10 md-py-10">
        <div class="w-100pc md-w-70pc mx-auto py-10">
            <h2 class="white fs-12 md-fs-xl1 fw-900 lh-2 ">We'll Detect the
                <span class="border-b bc-indigo bw-4"> Website </span>
                for you.</h2>
        </div>
    </section>


    <!-- product options -->
    <section id="team" class="min-h-100vh flex justify-start items-center">
```

```
<section class="py-l10">
    <div class="flex flex-column md-flex-row md-w-80pc mx-auto">
        <div class="w-100pc md-w-50pc">
            <div class="br-8 p-5 m-5 bg-indigo-lightest-10 pointer hover-scale-up-1 ease-300">
                <div class="inline-block bg-indigo indigo-lightest br-3 px-4 py-1 mb-10 fs-s4 uppercase">
                    Team Leader</div>
                <div class="indigo-lightest fw-600 fs-m1">Nishanth M<br><span class="opacity-30"> Reg.No:961219104035
                    <br><br>Loyola Institute of Technology and Science.
                </link></span> </div>

            </div>
        </div>
        <div class="w-100pc md-w-50pc">
            <div class="br-8 p-5 m-5 bg-indigo-lightest-10 pointer hover-scale-up-1 ease-300">
                <div class="inline-block bg-indigo indigo-lightest br-3 px-4 py-1 mb-10 fs-s4 uppercase">
                    Team Member</div>
                <div class="indigo-lightest fw-600 fs-m1">Vijay Santhar M<br><span class="opacity-30">
Reg.No:961219104055<br><br>Loyola Institute of Technology and Science</link></span> </div>

            </div>
        </div>
```

```
        <div class="w-100pc md-w-50pc">
            <div class="br-8 p-5 m-5 bg-indigo-lightest-10 pointer
hover-scale-up-1 ease-300">
                <div class="inline-block bg-indigo indigo-lightest br-3
px-4 py-1 mb-10 fs-s4 uppercase">
                    Team Member</div>
                <div class="indigo-lightest fw-600 fs-m1">Thangaraj
S<br><span class="opacity-30">
Reg.No:9612191053<br><br>Loyola Institute of Technology and
Science</link></span> </div>

            </div>
        </div>

        <div class="w-100pc md-w-50pc">
            <div class="br-8 p-5 m-5 bg-indigo-lightest-10 pointer
hover-scale-up-1 ease-300">
                <div class="inline-block bg-indigo indigo-lightest br-3
px-4 py-1 mb-10 fs-s4 uppercase">
                    Team Member</div>
                <div class="indigo-lightest fw-600 fs-m1">Ruban
R<br><span class="opacity-30">
Reg.No:961219104042<br><br>Loyola Institute of Technology and
Science</link></span> </div>

            </div>
        </div>

    </div>
</section>
```

```
</section>

    <!-- footer -->
    <footer class="p-5 md-p-l5 bg-indigo-lightest-10">
        <div class="flex flex-wrap">
            <div class="md-w-25pc mb-10">
                <!--<img src="assets/images/logo.png" class="w-l5"
alt="">-->
                <div class="white opacity-70 fs-s2 mt-4 md-pr-10">
                    <p>Copyright ©LSD 2022. <br>All Rights
Rseserved.<br>Team ID: PNT2022TMID51354</p>

                </div>
            </div>

    </footer>
        <i class="w-4" data-feather="download"></i>
    </div>
    <script src="https://code.jquery.com/jquery-3.4.1.min.js"></script>
    <script src="https://unpkg.com/feather-icons"></script>
    <script src="https://cdnjs.cloudflare.com/ajax/libs/slick-
carousel/1.9.0/slick.min.js"></script>
    <script src="https://cdn.jsdelivr.net/gh/cferdinandi/smooth-
scroll@15.0.0/dist/smooth-scroll.polyfills.min.js"></script>
```

```
    <!--<script src="assets/js/script.js"></script>-->
</body>


</html>
```

TEAM GITHUB LINK: https://github.com/IBM-EPBL/IBM-Project-9393-1658999833

PROJECT DEMO LINK: