WEB PHISHING DETECTION

1. INTRODUCTION

1.1 PROJECT OVERVIEW

- Phishing attack is a simplest way to obtain sensitive information from innocent users.
- Aim of the phishing is to acquire critical information like username, password and bank account details.
- Cyber security persons are now looking for trustworthy and steady detection techniques for phishing websites detection.
- Machine learning technology for detection of phishing URLs by extracting and analyzing various features of legitimate and phishing URLs. Decision Tree, random forest and Support vector machine algorithms are used to detect phishing websites.
- Detect phishing URLs as well as narrow down to best machine learning algorithm by comparing accuracy rate, false positive and false negative rate of each algorithm.

1.2 PURPOSE

The main purpose of the project is to detect the fake or phishing websites who are trying to get access to thsensitive 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 andto detect the phishing websites who are trying to gain access on sensitive data.

2. LITERATURE SURVEY

2.1 EXISTING PROBLEM

1. Phishing detection techniques do suffer low detection accuracy and high false alarm especially when novel phishing approaches are introduced.

2. Furthermore, page content inspection has been used by some strategies to overcome the false negative problems and complement the vulnerabilities of the stale lists.

3. Moreover, page content inspection algorithms each have different approach to phishing website detection with varying degrees of accuracy.

2.2 REFERENCES

1. Gunter Ollmann,“The Phishing Guide Understanding & Preventing Phishing Attacks”, IBMInternet Security Systems, 2007.
2. https://resources.infosecinstitute.com/category/enterpris/phishing/the- phishing-landscape/phishing-data-attackstatistics/#gref
3. Mahmoud Khonji, Youssef Iraqi, "Phishing Detection: A Literature Survey IEEE, and Andrew Jones, 2013.
4. Mohammad R., Thabtah F. McCluskey L., (2015) Phishing websites dataset.

   Available:https://archive.ics.uci.edu/ml/datasets/Phishing+Websites Accessed January 2016
5. http://dataaspirant.com/2017/01/30/how-decision-tree-algorithm-works/ http://dataaspirant.com/2017/05/22/random-forest-algorithm-machine-learing/https://www.kdnuggets.com/2016/07/support-vector-machines- simple-explanation.

2.3 PROBLEM STATEMENT DEFINITION

Date	19 September 2022
Team ID	PNT2022TMID44414
Project Name	Web Phishing Detection
Maximum Marks	2mark

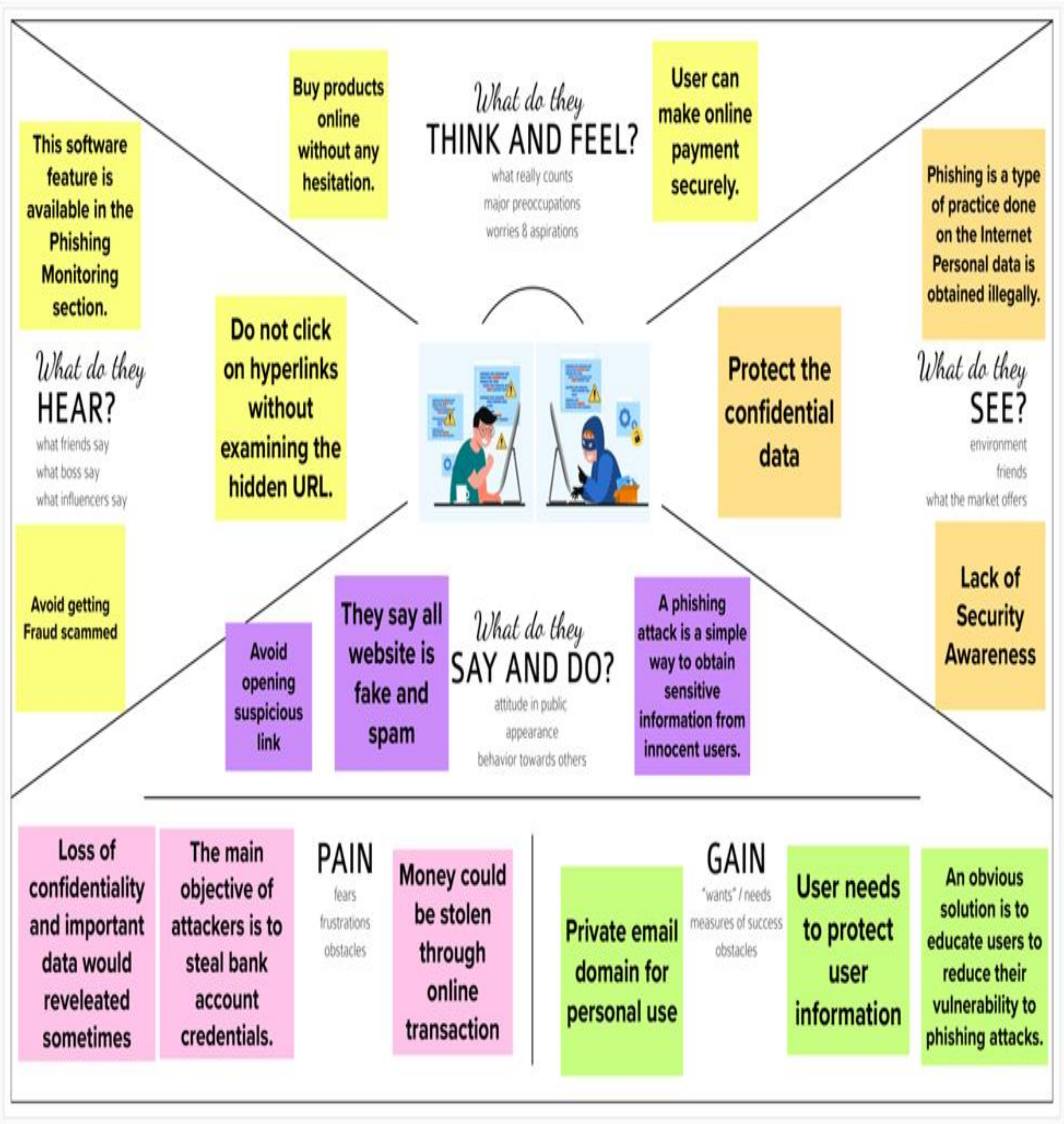
The main purpose of the project is to detect the fake or phishing websites who are trying to get access to the sensitive data or by creating the fake websites and trying to get access of the user personal credentials**.**

Whom does the problem affect?	Many users and organizations have fallen victim to phishing attacks, whereby their personally identifiable information, credentials and sensitive data have been stolen, resulting in identity theft, loss of money, loss of reputation, loss of intellectual property, as well as disruption of daily normaloperational activities.
What are the boundaries of the problem?	Phishing website looks very similar in appearance to its corresponding legitimate

	website to deceive users into believing that they are browsing the correct website. Visual similarity based phishing detection techniques utilize the feature set like text content, text format, HTML tags, Cascading Style Sheet (CSS), image, and so forth, to make the decision. These approaches compare the suspicious website with the corresponding legitimate website by using various features and if the similarity is greaterthan the predefined threshold value then it is declared phishing.
What is the issue?	The attacker easy to attack the bank account details and private data details And also
When does the issue occurs?	The issue occurs when an attacker, masquerading as a trusted entity, dupes a victim into opening an email, instant message, or text message
Where is the issue occurring?	96% of phishing attacks arrive by email. Another 3% are carried out through maliciouswebsites and just 1% via phone.
Why is it important that we fix the problem?	With sensitive information obtained from a successful phishing scam, these thieves can obtain loans or credit cards and even driver's licenses in your name. They can cause damage to your financial history and personalreputation that may take years to unravel.

3. IDEATION & PROPOSED SOLUTION

3.1 EMPATHY MAP CANVAS



What do they THINK AND FEEL?

what really counts
major preoccupations
worries & aspirations

Buy products online without any hesitation.

User can make online payment securely.

What do they HEAR?

what friends say
what boss say
what influencers say

This software feature is available in the Phishing Monitoring section.

Do not click on hyperlinks without examining the hidden URL.

Avoid getting Fraud scammed

What do they SEE?

environment
friends
what the market offers

Phishing is a type of practice done on the Internet Personal data is obtained illegally.

Protect the confidential data

Lack of Security Awareness

What do they SAY AND DO?

attitude in public
appearance
behavior towards others

Avoid opening suspicious link

They say all website is fake and spam

A phishing attack is a simple way to obtain sensitive information from innocent users.

PAIN

fears
frustrations
obstacles

Loss of confidentiality and important data would reveleated sometimes

The main objective of attackers is to steal bank account credentials.

Money could be stolen through online transaction

GAIN

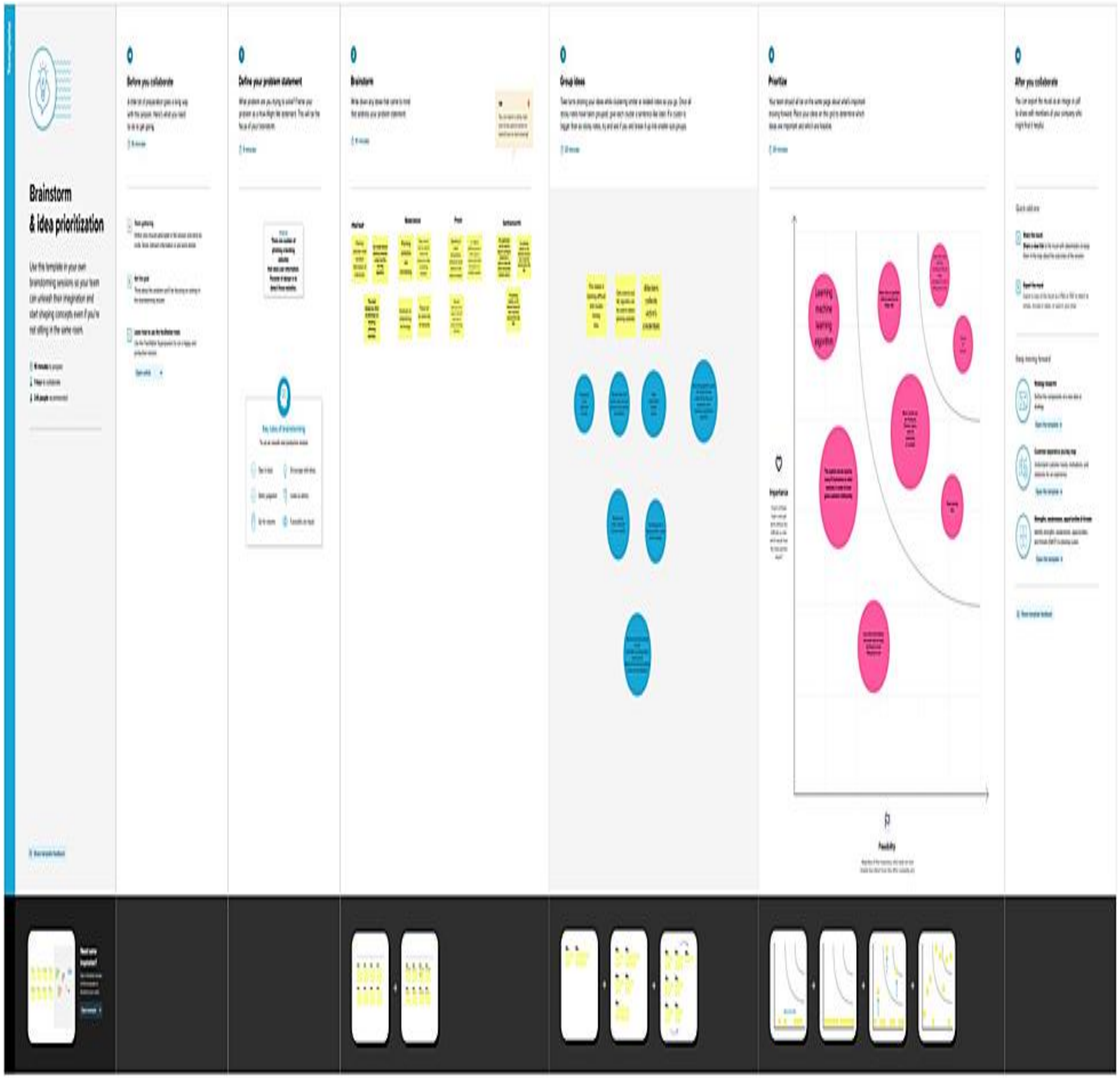
"wants" / needs
measures of success
obstacles

Private email domain for personal use

User needs to protect user information

An obvious solution is to educate users to reduce their vulnerability to phishing attacks.

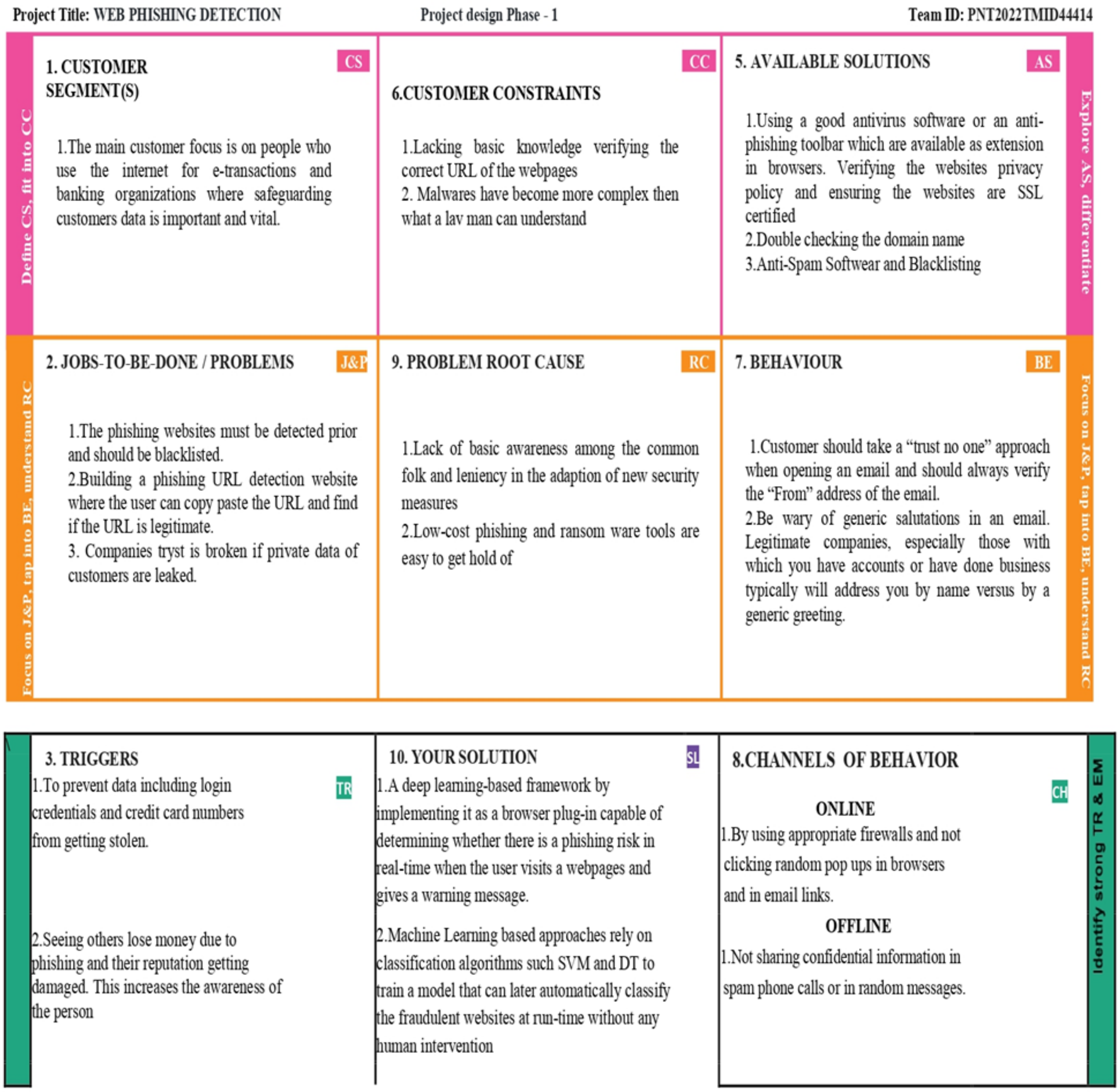
3.2 IDEATION & BRAINSTORMING



Brainstorm & idea prioritization

Use this template in your own brainstorming sessions so your team can unleash their imagination and start shaping concepts even if you're not sitting in the same room.

3.3 PROBLEM SOLUTION FIT



Project Title: WEB PHISHING DETECTION

Project design Phase - 1

Team ID: PNT2022TMID44414

Define CS, fit into CC

1. CUSTOMER SEGMENT(S) CS

1.The main customer focus is on people who use the internet for e-transactions and banking organizations where safeguarding customers data is important and vital.

6.CUSTOMER CONSTRAINTS CC

1.Lacking basic knowledge verifying the correct URL of the webpages
2. Malwares have become more complex then what a lav man can understand

5. AVAILABLE SOLUTIONS AS

1.Using a good antivirus software or an anti-phishing toolbar which are available as extension in browsers. Verifying the websites privacy policy and ensuring the websites are SSL certified
2.Double checking the domain name
3.Anti-Spam Softwear and Blacklisting

Explore AS, differentiate

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

2. JOBS-TO-BE-DONE / PROBLEMS J&P

1.The phishing websites must be detected prior and should be blacklisted.
2.Building a phishing URL detection website where the user can copy paste the URL and find if the URL is legitimate.
3. Companies tryst is broken if private data of customers are leaked.

9. PROBLEM ROOT CAUSE RC

1.Lack of basic awareness among the common folk and leniency in the adaption of new security measures
2.Low-cost phishing and ransom ware tools are easy to get hold of

7. BEHAVIOUR BE

1.Customer should take a "trust no one" approach when opening an email and should always verify the "From" address of the email.
2.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.

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

3. TRIGGERS TR

1.To prevent data including login credentials and credit card numbers from getting stolen.

2.Seeing others lose money due to phishing and their reputation getting damaged. This increases the awareness of the person

10. YOUR SOLUTION SL

1.A deep learning-based framework by implementing it as a browser plug-in capable of determining whether there is a phishing risk in real-time when the user visits a webpages and gives a warning message.

2.Machine Learning based approaches rely on classification algorithms such SVM and DT to train a model that can later automatically classify the fraudulent websites at run-time without any human intervention

8.CHANNELS OF BEHAVIOR CH

ONLINE

1.By using appropriate firewalls and not clicking random pop ups in browsers and in email links.

OFFLINE

1.Not sharing confidential information in spam phone calls or in random messages.

Identify strong TR & EM

4. REQUIREMENT ANALYSIS

4.1 FUNCTIONAL REQUIREMENT

Following are the functional requirements of the proposed solution.

FR No	Functional Requirement (Epic)	Requirement (Story / Sub-Task)
FR-1	User Registration	Register by entering details suchas name, email, password, phonenumber, etc. (“A visitor can register himself to the website to access it”).
FR-2	User Login	Login using the registered email idand password. (“ After a successful registration, user/adminmay input his credentials to login into the system”)
FR-3	Model Building	Build various machine learning model to detect web phishing andcompare them.
FR-4	Check URL	Here, the user checks for the blacklisted website by inputtingthe URL.

FR-5	Integration	Integrate the frontend and the developed ML model using flask
FR-6	Alert Message	Notify the user through email orphone regarding the malicious website.

4.2 NON-FUNCTIONAL REQUIREMENT

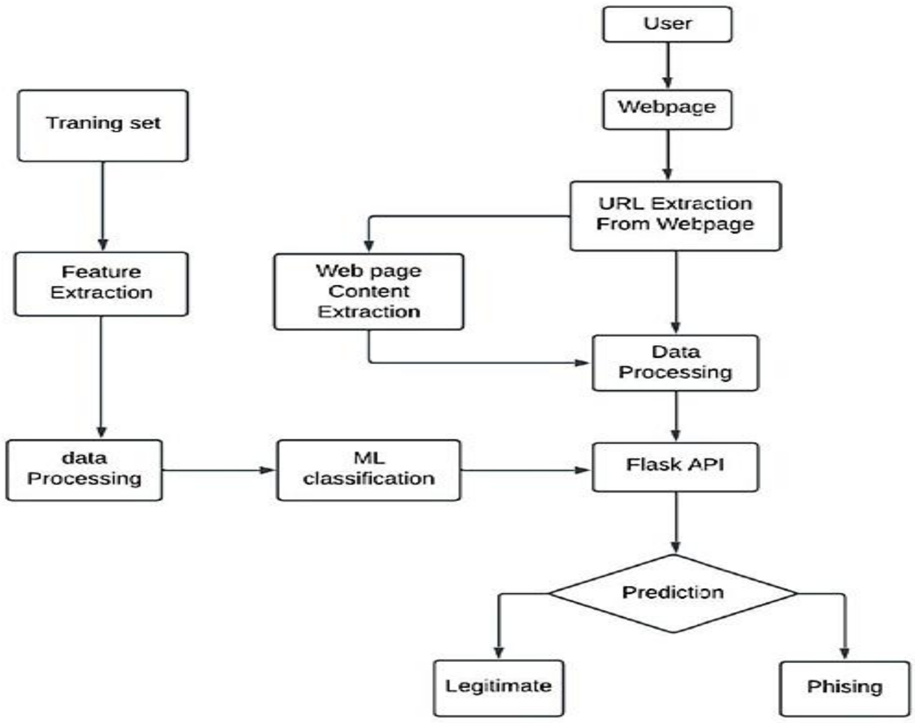
Following are the non-functional requirements of the proposed solution.

FR NO	**Non-Functional Requirement**	**Description**
FR-1	Usability	Any URL must be accepted fordetection.
FR-2	Security	Alert message must be sent tothe users to enable secure browsing.
FR-3	Reliability	The web phishing websites mustdetected accurately and the result must be reliable.
FR-4	Performance	The performance and interfacemust be user friendly

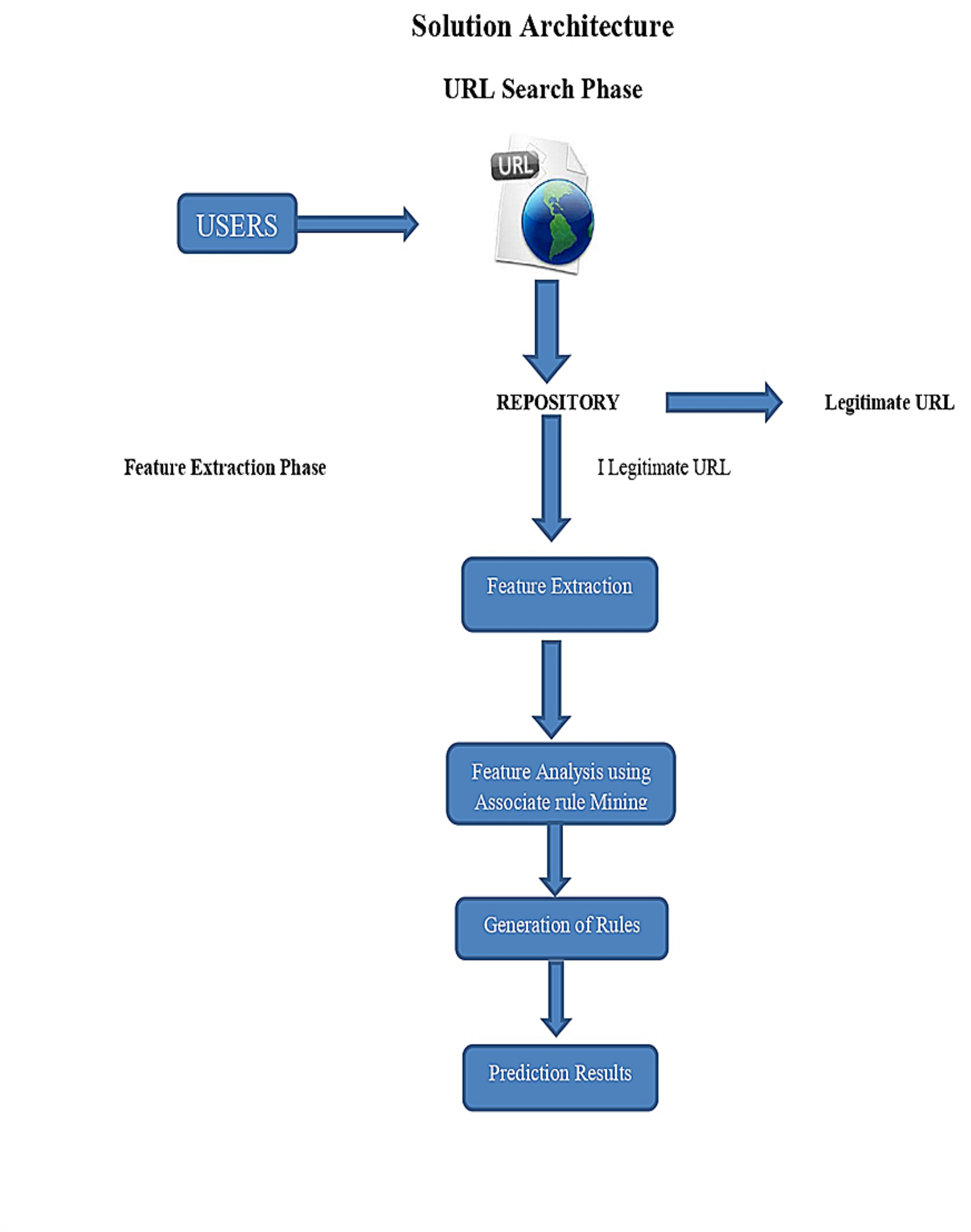
FR-5	Availability	Anyone must be able to registerand login.
FR-6	Scalability	It must be able to handle increase in the number of users.

5. PROJECT DESIGN

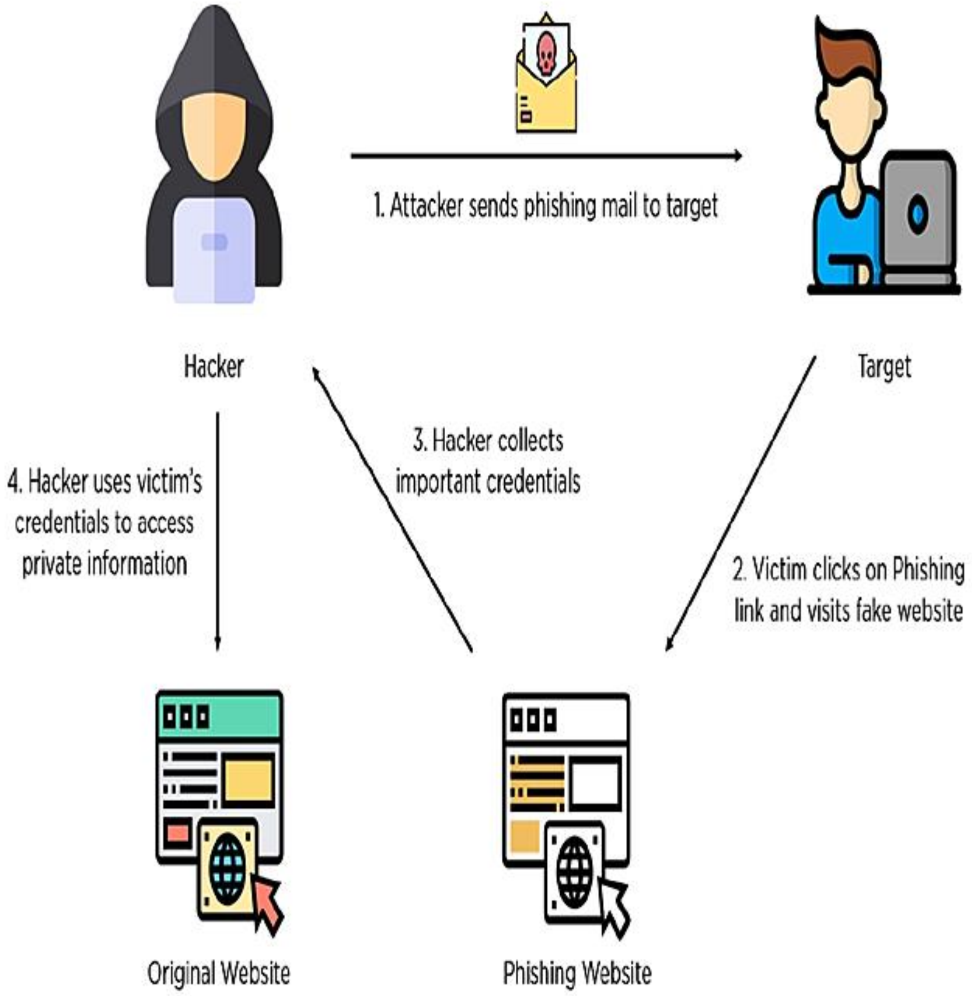
5.1 DATA FLOW DIAGRAMS



5.2 SOLUTION & TECHNICAL ARCHITECTURE



Technical Architecture for the model:


1. Attacker sends phishing mail to target
Hacker
Target
3. Hacker collects important credentials
4. Hacker uses victim's credentials to access private information
2. Victim clicks on Phishing link and visits fake website
Original Website
Phishing Website

5.3 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	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 confirm in Any password	I can access my account / dashboar	High	Sprint-1
		USN-2	As a user, I will receive confirmation email once I have registered for the application	I can receive confirmation on email & click confirm	High	Sprint-1
		USN-3	As a user, I can register for the application through	I can register & access the	low	Sprint-2

			Facebook	dashboard with Facebook Login		
		USN-4	As a user, I can register for the application through Gmail		Medium	Sprint-1
		USN-5	As a user, I can log into the application by entering email & password		High	Sprint-1
						Sprint-1
		USN-1	As a user i can input the particular URL in the requiredfield and waiting for validation	I can go access the website without any problem	High	Sprint-1

		USN-1	After i compare in case if none found on comparison thenwe can extract feature using heuristic and visualsimilarity approach	As a Useri can have comparis on between websites for security	High	Sprint-1
		USN-1	Here the Model will predict the URL websites using Machine Learning algorithms such as Logistic Regression, KNN	In this i can have correct prediction on the particular algorithms	High	Sprint-1

6. PROJECT PLANNING & SCHEDULING

6.1 SPRINT PLANNING & ESTIMATION

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 analyses a websiteto decide whether it is a phishing or not	10	High	
Sprint-1		USN-2	Some of URL-BasedFeatures are Digit count inthe URL Total length of URL Checking whether the URL is typosquatted or not Checking whether it includes a	10	High	

			legitimate brand name or not Number of subdomains in URL TLD is one of the commonlyused one			

Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority	Team Members
			● Number of subdomains inURL ● TLD is one of the commonlyused one			
Sprint-2	Domain detection	USN-3	The purpose of Phishing Domain Detection is detectingphishing domain	10		

			names. Therefore, passive queries related to the domain name, which we wantto classify as phishing or not, provide useful information to us.		High	
Sprint-2	Domain Detection	USN-4	● Some useful Domain-Based Features are ● Its domain name or its IPaddress in ● blacklists of well-known reputation ● services? ● How many days passed sincethe ● domain was registered? ● Is the registrant	10	High	

			name hidden?			
Sprint-3	Page based features and Content based features		Page-Based Features are using information about pages whichare calculated reputation ranking services.	10	High	

Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority	Team Members
	Features	USN-5	Obtaining these typesof 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		

			● Global page rank ● Country page rank ● 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.		High	
sprint-3	Detection process	USN-6	Detecting Phishing Domains is a classification problem, so it meanswe need labeled data which has samples as	20	High	

			phish domains and legitimate domains inthe training phase			

6.2 PROJECT TRACKKER,VELOCITY & BURNDOWN CHAR

Sprint	**Total Story Points**	**Duration**	**Sprint Start Date**	**Sprint End Date (Planned)**	**Story Points Completed (as on Planned End Date)**	**Sprint ReleaseDate (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)

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

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.

6.3 SPRINT DELIVERY SCHEDULE

Sprint	Milestone
Sprint1	URL detector URL is the first thing to analyses a website to decide whether it is a phishing or not Some of URL-BasedFeatures are Digit count in the URL Total lengthof URL Checking whether the URL is typo squatted or not Checking whether it includes a legitimate brand name or not Number of subdomains in URL TLD is one of the commonlyused one
Sprint 2	Domain Detection 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. Some useful Domain-Based Features are Its domain name or its IP address in blacklists of well-known reputation services? How many days passed since the domain was registered? Is the registrant name hidden?
Sprint 3	Page Based Features and Content Based Features Page-Based Features are using information aboutpages which are calculated reputation rankingservices. Obtaining these types of features requiresactive scan to target domain. Page contents areprocessed for us to detect whether target domain isused for phishing or not Global page rank Country page rank Position at the Alex a top 1 million site Some processed information aboutpages are Page titles Meta tags Hidden text in the body Images etc.
Sprint 4	Detection process 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.

7. CODING & SOLUTIONING

7.1 FEATURE 1

```
<!DOCTYPE html>

<html lang="en">

 <head>

 <meta charset="UTF-8">

 <meta http-equiv="X-UA-Compatible" content="IE=edge">

 <meta name="viewport" content="width=device-width, initial-scale=1.0">

 <meta name="description" content="This website is develop for identify the
safety ofurl.">

<meta name="keywords" content="phishing url,phishing,cyber
security,machinelearning,classifier,python">

<meta name="author" content="NSP">

<!-- BootStrap -->

<link rel="stylesheet"
href="https://stackpath.bootstrapcdn.com/bootstrap/4.5.0/css/bootstrap.min.css
"

 integrity="sha384-
9aIt2nRpC12Uk9gS9baDl411NQApFmC26EwAOH8WgZl5MYYxFfc+NcPb1d
KGj7Sk" crossorigin="anonymous">


 <link href="static/styles.css" rel="stylesheet">

 <title>Web Phishing detection</title>

 <style>
```

```
body {

  margin: 0;

  padding:0;

  font-family: Arial, Helvetica, sans-serif;

}


.topnav {

  overflow: hidden;

  background-color: #659999,#f4791f;

}


.topnav a {

  float: right;

  display:flex;

  color: #f2f2f2;

  text-align: center;

  padding: 14px 16px;

  text-decoration: none;

  font-size: 17px;

  justify-content:center;

}
```

```
.topnav a:hover {
  background-color: #ddd;
  color: black;
}

.topnav a.active {
  background-color: #04AA6D;
  color: white;
}
  nav{
                position:relative;
                top: 0;
                left: 0;
                width: 100%;
                height: 70px;
                padding: 10px 100px;
                box-sizing:border-box;
                background:#161616;
            }
             nav .logo{
                padding: 15px;
                height: 30px;
```

```
    float: left;

    font-size: 15px;

    font-weight: bold;

    color: #fff;

}

nav ul {

    list-style:none;

    float: right;

    margin: 0;

    padding: 0;

    display: flex;

}

nav ul li a{

    float: right;

    display: block;

    color: #f2f2f2;

    text-align: center;

    padding: 15px;

    text-decoration: none;

    font-size: 17px;


}
```

```
            nav ul li a:hover{
                background: rgb(200, 212, 200);
                border-radius: 6px;
                color: rgb(70, 27, 13);
            }
            nav ul li a.active{
                background: #e2472f;
                border-radius: 6px;

            }
</style>
</head>

<body>
 <div class="wrap">
        <nav>
        <div class="logo" ><h2>Web Phishing Detection</h2> </div>
        <ul>
            <li class="active">
             <a href="{{ url_for ('index') }}">Home</a></li>
            <li><a href="{{ url_for ('about') }}">About</a></li>
        </ul>
 </div>

 </nav><br><br><br><br><br>


<center><br><br>

<div class=" container">

    <div class="row">

        <div class="form col-md" id="form1">

           <center>

                 <h1 style="font-family:'Franklin Gothic Medium', 'Arial Narrow', Arial Black,
          sans-serif;color: rgb(39, 41, 40);">PHISHING WEBSITE DETECTION USING
          MACHINE LEARNING</h1>

           </center>


            <br>


            <form action="/" method ="post">

                 <center> <input type="text" class="form__input" name ='url' id="url"
          placeholder="Enter Your URL" required="" />

                <label for="url" class="form__label">URL</label>

                <button class="button" role="button" href="index.html" >Predict
          here</button> </center>

            </form>

</center>
```

```
    </div>

 <center>

    <div class="col-md" id="form2">



        <br>

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



        <br>

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

        <button class="button2" id="button2" role="button"
          onclick="window.open('{{url}}')" target="_blank" >Still want to
          Continue</button>

        <button class="button1" id="button1" role="button"
          onclick="window.open('{{url}}')" target="_blank">Continue</button>

    </div>

</div>

<br>

</div>



    <!-- JavaScript -->

    <script src="https://code.jquery.com/jquery-3.5.1.slim.min.js"

        integrity="sha384-
          DfXdz2htPH0lsSSs5nCTpuj/zy4C+OGpamoFVy38MVBnE+IbbVYUew+OrCXa
```

```
      Rkfj"
    crossorigin="anonymous"></script>
<script src="https://cdn.jsdelivr.net/npm/popper.js@1.16.0/dist/umd/popper.min.js"
    integrity="sha384-
      Q6E9RHvbIyZFJoft+2mJbHaEWldlvI9IOYy5n3zV9zzTtmI3UksdQRVvoxMfoo
      Ao"
    crossorigin="anonymous"></script>
<script src="https://stackpath.bootstrapcdn.com/bootstrap/4.5.0/js/bootstrap.min.js"
    integrity="sha384-
      OgVRvuATP1z7JjHLkuOU7Xw704+h835Lr+6QL9UvYjZE3Ipu6Tp75j7Bh/kR0J
      KI"
    crossorigin="anonymous"></script>
<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>
</body>
</html>
```

7.2FEATURE 2

```
 <!DOCTYPE html>
<html lang="en">
    <head>
        <title> Web Phishing Detection</title>
        <meta charset="utf-8">
        <meta name="viewport" content="width=device-width, initial-scale=1">
        <link rel="stylesheet"
          href="https://maxcdn.bootstrapcdn.com/bootstrap/3.4.1/css/bootstrap.min.css">
        <link rel="stylesheet" href="https://cdnjs.cloudflare.com/ajax/libs/font-
          awesome/4.7.0/css/font-awesome.min.css">
        <script
          src="https://ajax.googleapis.com/ajax/libs/jquery/3.6.0/jquery.min.js"></script>
        <script
          src="https://maxcdn.bootstrapcdn.com/bootstrap/3.4.1/js/bootstrap.min.js"></scrip
 t>
<style>
    body{
        margin: 0;
        padding: 0;
        font-family:Arial, Helvetica, sans-serif
    }
    nav{
        position:relative;
        top: 0;
        left: 0;
        width: 100%;
        height: 70px;
        padding: 10px 100px;
        box-sizing:border-box;
        background:#161616;
   }
    nav .logo{
        padding: 15px;
        height: 30px;
        float: left;
        font-size: 25px;
    font-weight: bold;

    color: #fff;

}

nav ul {

    list-style:none;

    float: right;

    margin: 0;

    padding: 0;

    display: flex;

    font-size: 25px;

}

nav ul li a{

    float: right;

    display: block;

    color: #f2f2f2;

    text-align: center;

    padding: 15px;

    text-decoration: none;

    font-size: 22px;


}

nav ul li a:hover{
```

```
            background: rgb(200, 212, 200);
            border-radius: 6px;
            color: rgb(70, 27, 13);
        }
        nav ul li a.active{
            background: #e2472f;
            border-radius: 6px;

        }
        .end {
            overflow: hidden;
            background-color: rgb(63, 63, 63);
            position: fixed;
            bottom: 0;
            height: 55px;
            width: 100%;
        }
        .continer {
            align-self:auto;
        }
        .button1{
appearance: button;
```

```
background-color: transparent;

background-image: linear-gradient(to bottom, rgb(160, 245, 174), #37ee65);

border: 0 solid #e5e7eb;

border-radius: .5rem;

box-sizing: border-box;

color: #482307;

column-gap: 1rem;

cursor:  pointer;

display: flex;

font-family: ui-sans-serif,system-ui,-apple-system,system-ui,"Segoe UI",Roboto,"Helvetica Neue",Arial,"Noto Sans",sans-serif,"Apple Color Emoji","Segoe UI Emoji","Segoe UI Symbol","Noto Color Emoji";

font-size: 100%;

font-weight:  700;

line-height: 24px;

margin: 0;

outline: 2px solid transparent;

padding: 1rem 1.5rem;

text-align: center;

text-transform: none;

transition: all .1s cubic-bezier(.4, 0, .2, 1);

user-select: none;

-webkit-user-select: none;
```

```
  touch-action: manipulation;
  box-shadow: -6px 8px 10px rgba(81,41,10,0.1),0px 2px 2px rgba(81,41,10,0.2);
  display: none;
}
.button2{
  appearance: button;
  background-color: transparent;
  background-image: linear-gradient(to bottom, rgb(252, 162, 162), #ee3737);
  border: 0 solid #e5e7eb;
  border-radius: .5rem;
  box-sizing: border-box;
  color: #482307;
  column-gap: 1rem;
  cursor:  pointer;
  display: flex;
  font-family: ui-sans-serif,system-ui,-apple-system,system-ui,"Segoe UI",Roboto,"Helvetica Neue",Arial,"Noto Sans",sans-serif,"Apple Color Emoji","Segoe UI Emoji","Segoe UI Symbol","Noto Color Emoji";
  font-size: 100%;
  font-weight:  700;
  line-height: 24px;
  margin: 0;
  outline: 2px solid transparent;
  padding: 1rem 1.5rem;

  text-align: center;

  text-transform: none;

  transition: all .1s cubic-bezier(.4, 0, .2, 1);

  user-select: none;

  -webkit-user-select: none;

  touch-action: manipulation;

  box-shadow: -6px 8px 10px rgba(81,41,10,0.1),0px 2px 2px rgba(81,41,10,0.2);

  display: none;

}


  </style>

    </head>

    <body style="background-image: linear-gradient(to right,#c6ffdd, #fbd786,
          #f7797d);">

        <div class="wrap">

        <nav>

        <div class="logo" >Web Phishing Detection</div>

        <ul>

            <li class="active">

             <a href="{{ url_for ('index') }}">Home</a></li>

            <li><a href="index.html">About</a></li>
```

```
</ul>
 </div>
</nav><br><br>
<div class="container">
<center>
  <h2 style="font-family:'Franklin Gothic Medium', 'Arial Narrow', Arial, sans-serif;color: rgb(39, 41, 40);">ABOUT PROJECT </h2>
  <br>
   <p style="font-size:20px;font-family: 'Times New Roman', Times, serif;">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="font-size:20px;font-family: 'Times New Roman', Times, serif;">This Guided Project mainly focuses on applying a machine-learning algorithm to detect Phishing websites.
```

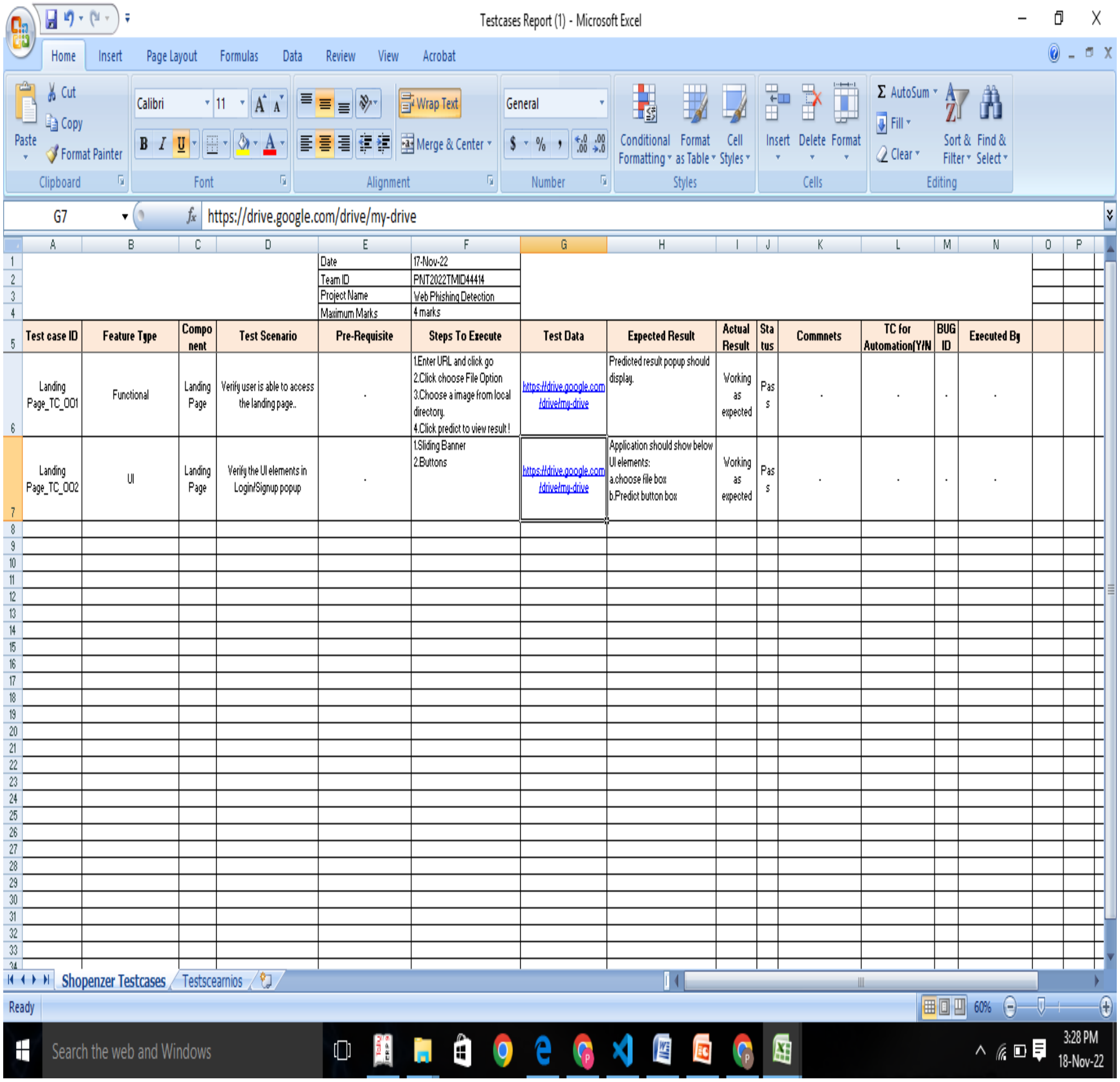
In order to detect and predict e-banking phishing websites, we proposed an intelligent, flexible and effective system that is based on using classification algorithms. We implemented classification algorithms and techniques to extract the phishing datasets criteria to classify their legitimacy.

The e-banking phishing website can be detected based on some important characteristics like URL and domain identity, and security and encryption criteria in the final phishing detection rate. Once a user makes a transaction online when he makes payment through an e-banking website our system will use a data mining algorithm to detect whether the e-banking website is a phishing website or not.

```
    <br>
</div>
</center>
<br>
 <!--  <div class="end">
      <p style="color:rgb(255, 246, 246);
                margin-top: 20px;
                text-align: center;">
      </p>

    </div>-->
  </body>
  </html>
```

8. TESTING

8.1 TEST CASE



https://drive.google.com/drive/my-drive

Date	17-Nov-22
Team ID	PNT2022TMID44414
Project Name	Web Phishing Detection
Maximum Marks	4 marks

Test case ID	Feature Type	Component	Test Scenario	Pre-Requisite	Steps To Execute	Test Data	Expected Result	Actual Result	Status	Commnets	TC for Automation(Y/N	BUG ID	Executed By
Landing Page_TC_OO1	Functional	Landing Page	Verify user is able to access the landing page..	.	1.Enter URL and click go 2.Click choose File Option 3.Choose a image from local directory. 4.Click predict to view result !	https://drive.google.com/drive/my-drive	Predicted result popup should display.	Working as expected	Pass	.	.	.	.
Landing Page_TC_OO2	UI	Landing Page	Verify the UI elements in Login/Signup popup	.	1.Sliding Banner 2.Buttons	https://drive.google.com/drive/my-drive	Application should show below UI elements: a.choose file box b.Predict button box	Working as expected	Pass	.	.	.	.

USER ACCEPTANCE TESTING

Purpose of Document

The purpose of this document is to briefly explain the test coverage and open issues of the [ProductName] project at the time of the release to User Acceptance Testing (UAT).

1. Defect Analysis

This report shows the number of resolved or closed bugs at each severity level, and how they were resolved

Resolution	Severity 1	Severity 2	Severity 3	Severity 4	Subtotal
By Design	10	4	2	3	20
Duplicate	1	0	3	0	4
External	2	3	0	1	6
Fixed	11	2	4	20	37
Not Reproduced	0	0	1	0	1
Skipped	0	0	1	1	2
Won't Fix	0	5	2	1	8
Totals	24	14	13	26	77

1.Test Case Analysis

This report shows the number of test cases that have passed, failed, and untested

Section	Total **Cases**	Not Tested	Fai l	Pass
Print Engine	7	0	0	7
Client Application	51	0	0	51
Security	2	0	0	2
Outsource Shipping	3	0	0	3
Exception Reporting	9	0	0	9
Final Report Output	4	0	0	4
Version Control	2	0	0	2

9. RESULTS

Scikit-learn tool has been used to import Machine learning algorithms.Dataset is divided into training set and testing set in 50:50, 70:30 and 90:10

ratios respectively. Each classifier is trained using training set and testing set is used to evaluate performance of classifiers. Performance of classifiers has been evaluated by calculating classifier's accuracy score, false negative rate and false positive rate.

Result shows that Random forest algorith1m gives better detection accuracy which is 97.14 with lowest false negative rate than decision tree and support vector machine algorithms . Result also shows that detection accuracy of phishing websites increases as more dataset used as training dataset. All classifiers perform well when 90% of data used as training dataset. the detection accuracy of all classifiers when 50%, 70% and 90% of data used as training dataset and graph clearly shows that detection accuracy increases when 90% of data used as training dataset and random forest detection accuracy is maximum than other two classifiers.

10. ADVANTAGES

- The 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.

10. DISADVANTAGES

- If internet connection fails, this system won’t work.
- All websites related data will be stored in one place.

11.CONCLUSION

After reviewing and researching for appropriate monitoring tools, proposed system has been identified and chosen to address the complexity of monitoring requirement for current situation.This software is designed to show awareness of the extensive level of its functionality, features that can be displayed in the monitoring era. The system fosters many features in comparison of other software. Its unique features such as capturing blacklisted URL's from the browser directly to verify the validity of the website, notifying user on blacklisted websites while they are trying to access through pop-up, andalso notifying through email.

This system will assist user to be alert when they are trying to access a blacklisted website.In conclusion, this system is designed for resources are used as intended, prevents from valuable information from leaks out, produce better control mechanism and alerts the user to keep their private information safe. Like any other programs, there are improvements which could be made into this system. Based on the capabilities which the current system processes, text message integration would a great recommendation that could be made to improve the program in the future.

The future version of the application could also implement an option to directly notify the blacklisted website with a text message. The program could be made to access the list as an attachment. This text message integration function would further the usability of the application.

12.FUTURE SCOPE

one of the challenges faced by our research was the unavalibility reliable training datasets. in fact, this challengs faces any research in the field. however, although plenty of articles aboud predicting phishing website using data mining techniques have been disseminated these days, no reliable 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 effectivein predicting phishing websites. in addtion, we proposed some new features, experimentally assign new rules to some well-known features and updates some other features.

13. APPENDIX

Source code

app.py

```
#importing required libraries

from flask import Flask, request, render_templateimport numpy as
np
import pandas as pd

from sklearn import metrics
```

```
import warningsimport
pickle
warnings.filterwarnings('ignore') from feature import
FeatureExtractionfile = open("pickle/model.pkl","rb")gbc =
pickle.load(file)
file.close()

app = Flask(__name__)

@app.route("/", methods=["GET", "POST"])def index():
    if request.method == "POST":url =
        request.form["url"]
        obj = FeatureExtraction(url)

        x = np.array(obj.getFeaturesList()).reshape(1,30)y_pred =
        gbc.predict(x)[0]
        #1 is safe

        #-1 is unsafe

        y_pro_phishing = gbc.predict_proba(x)[0,0]
        y_pro_non_phishing =  gbc.predict_proba(x)[0,1] # if(y_pred
        ==1 ):
        pred = "It is {0:.2f} % safe to go ".format(y_pro_phishing*100)return
       render_template('index.html',xx
=round(y_pro_non_phishing,2),url=url  )

    return  render_template("index.html",  xx  =-1)@app.route('/about.html')
```

```
def about():
    return render_template("about.html") if__name
______________== "_main_":
    app.run(debug=True)
```

Index

```
<!DOCTYPE html>

<html lang="en">

<head>

    <meta charset="UTF-8">

    <meta http-equiv="X-UA-Compatible" content="IE=edge">

    <meta name="viewport" content="width=device-width, initial-scale=1.0">

    <meta name="description" content="This website is develop for identifythe safety of url.">
    <meta name="keywords" content="phishing url,phishing,cybersecurity,machine
learning,classifier,python">
    <meta name="author" content="NSP">


    <!-- BootStrap -->

    <link rel="stylesheet" href="https://stackpath.bootstrapcdn.com/bootstrap/4.5.0/css/bootstrap.min
```

```
.css"
        integrity="sha384-
9aIt2nRpC12Uk9gS9baDl411NQApFmC26EwAOH8WgZl5MYYxFfc+NcPb1dKGj7Sk"
crossorigin="anonymous">

    <link href="static/styles.css" rel="stylesheet">

    <title>Web Phishing detection</title>

    <style> body
{ margin: 0;
padding:0;
  font-family: Arial, Helvetica, sans-serif;

}


.topnav  { overflow:
  hidden;
  background-color: #659999,#f4791f;

}


.topnav a { float:
  right;
  display:flex;
  color: #f2f2f2;
  text-align: center;
  padding: 14px 16px;
  text-decoration: none;font-
  size: 17px;
  justify-content:center;

}



.topnav a:hover { background-
  color: #ddd;color: black;
}


.topnav a.active {

  background-color: #04AA6D;color:
  white;
}

  nav{

              position:relative;top:
              0;
              left: 0;

              width: 100%;
              height: 70px;
              padding: 10px 100px; box-
              sizing:border-box;
              background:#161616;



          }

           nav .logo{
```

```
     padding: 15px;
     height: 30px; float:
     left;
     font-size: 15px; font-
     weight: bold;color:
     #fff;
 }

 nav ul {

     list-style:none;
     float: right;
     margin: 0;
     padding: 0;
     display: flex;
 }

 nav ul li a{
     float: right;
     display: block; color:
     #f2f2f2; text-align:
     center;padding: 15px;
     text-decoration: none;font-
size: 17px;


 }

 nav ul li a:hover{
```

```
                background: rgb(200, 212, 200);border-
                radius: 6px;
                color: rgb(70, 27, 13);

            }

            nav ul li a.active{ background:
                #e2472f;border-radius: 6px;


            }

  </style>

   </head>


  <body>

 <div class="wrap">

        <nav>

        <div class="logo" ><h2>Web Phishing Detection</h2> </div>

        <ul>

            <li class="active">

             <a href="{{ url_for ('index') }}">Home</a></li>

            <li><a href="{{ url_for ('about') }}">About</a></li>

        </ul>

 </div>

 </nav><br><br><br><br><br>


<center><br><br>
```

```
<div class=" container">

    <div class="row">

        <div class="form col-md" id="form1">

           <center>

                 <h1 style="font-family:'Franklin Gothic Medium', 'Arial Narrow',Arial Black,
sans-serif;color: rgb(39, 41, 40);">PHISHING WEBSITE DETECTION USING MACHINE
LEARNING</h1>
           </center>



            <br>



            <form action="/" method ="post">

            <center> <input type="text" class="form____________input" name ='url' id="url"
placeholder="Enter Your URL" required="" />
                <label for="url" class="form________label">URL</label>

                <button class="button" role="button" href="index.html" >Predicthere</button>
</center>
            </form>

</center>



    </div>

 <center>

    <div class="col-md" id="form2">



        <br>
```

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


      <br>

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

      <button class="button2" id="button2" role="button" onclick="window.open('{{url}}')"
target="_blank" >Still want toContinue</button>
      <button class="button1" id="button1" role="button" onclick="window.open('{{url}}')"
target="_blank">Continue</button>
   </div>

</div>

<br>

</div>


  <!-- JavaScript -->

   <script src="https://code.jquery.com/jquery-3.5.1.slim.min.js"
      integrity="sha384-
DfXdz2htPH0lsSSs5nCTpuj/zy4C+OGpamoFVy38MVBnE+IbbVYUew+OrCXaRkfj"
      crossorigin="anonymous"></script>

   <script src="https://cdn.jsdelivr.net/npm/popper.js@1.16.0/dist/umd/popper.min.js"
      integrity="sha384-
Q6E9RHvbIyZFJoft+2mJbHaEWldlvI9IOYy5n3zV9zzTtmI3UksdQRVvoxMfooAo"
```

```
        crossorigin="anonymous"></script>

    <script src="https://stackpath.bootstrapcdn.com/bootstrap/4.5.0/js/bootstrap.min.js"
        integrity="sha384-
OgVRvuATP1z7JjHLkuOU7Xw704+h835Lr+6QL9UvYjZE3Ipu6Tp75j7Bh/kR0JKI"
        crossorigin="anonymous"></script>




    <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 ;
```

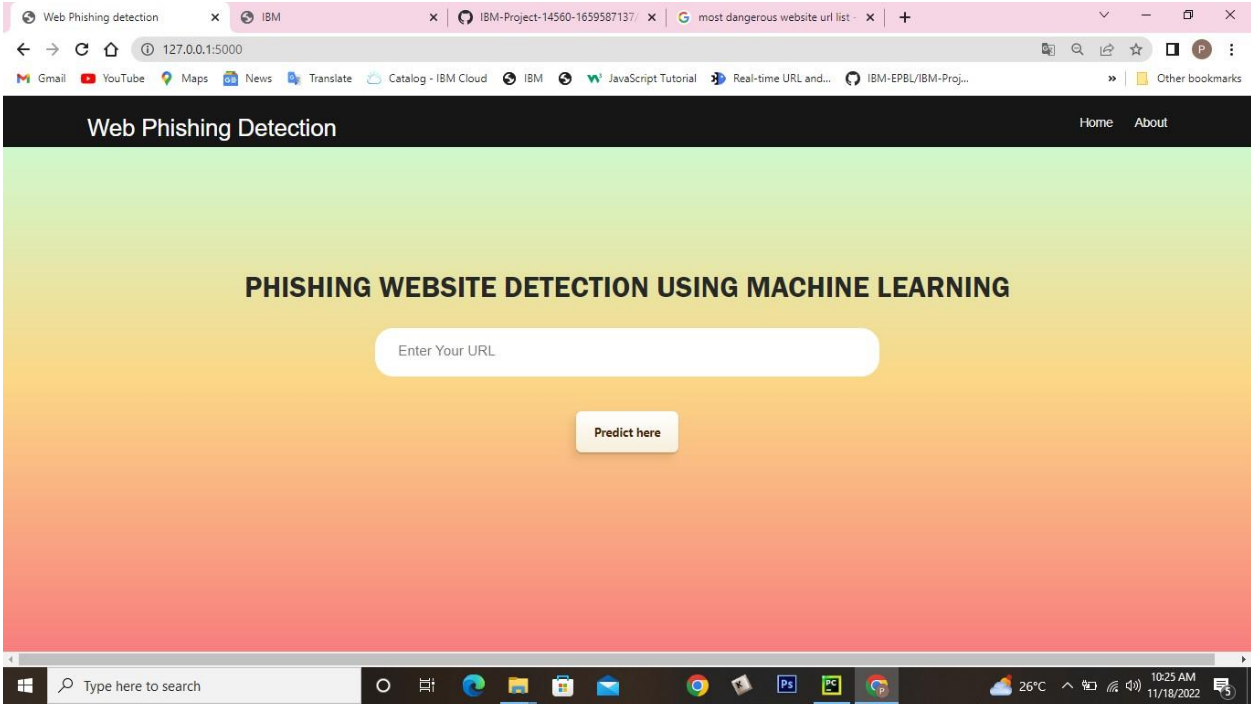
```
                        document.getElementById("button2").style.display="block";

                    }



            </script>


        </body>

        </html>
```

Aobut

```
            <!DOCTYPE html>

            <html lang="en">

                <head>

                    <title> Web Phishing Detection</title>

                    <meta charset="utf-8">

                    <meta name="viewport" content="width=device-width, initial-scale=1">
                    <link rel="stylesheet"
            href="https://maxcdn.bootstrapcdn.com/bootstrap/3.4.1/css/bootstrap.min.css">
                    <link rel="stylesheet"
            href="https://cdnjs.cloudflare.com/ajax/libs/font-
            awesome/4.7.0/css/font-awesome.min.css">
                    <script src="https://ajax.googleapis.com/ajax/libs/jquery/3.6.0/jquery.min.js">
            </script>
                    <script src="https://maxcdn.bootstrapcdn.com/bootstrap/3.4.1/js/bootstrap.min.js"></script>
```

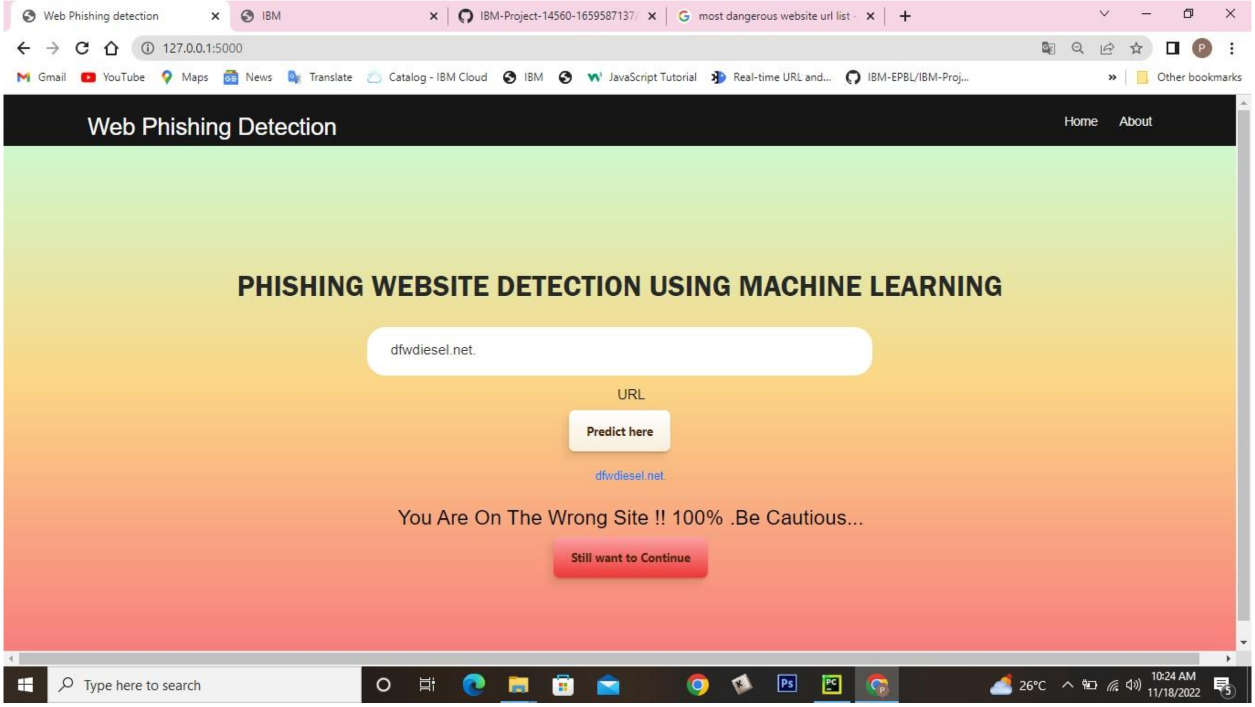
```
<style>
    body{
        margin: 0;
        padding: 0;
        font-family:Arial, Helvetica, sans-serif
    }
    nav{
        position:relative;top:
        0;
        left: 0;
        width: 100%;
        height: 70px;
        padding: 10px 100px; box-
        sizing:border-box;
        background:#161616;
   }
    nav .logo{ padding:
        15px;height: 30px;
        float: left;
        font-size: 25px; font-
        weight: bold;color:
        #fff;
    }
    nav ul {
        list-style:none;
        float: right;
        margin: 0;
        padding: 0;
        display: flex; font-
        size: 25px;
    }
```

```
nav ul li a{
    float: right;
    display: block; color:
    #f2f2f2; text-align:
    center;padding: 15px;
    text-decoration: none;font-
    size: 22px;

}
nav ul li a:hover{
    background: rgb(200, 212, 200);border-
    radius: 6px;
    color: rgb(70, 27, 13);
}
nav ul li a.active{ background:
    #e2472f;border-radius: 6px;

}
.end {
    overflow: hidden;
    background-color: rgb(63, 63, 63);position:
    fixed;
    bottom: 0;
    height: 55px;
    width: 100%;
}
.continer {
    align-self:auto;
}
.button1{
  appearance: button; background-color:
  transparent;
  background-image: linear-gradient(to bottom, rgb(160, 245, 174),#37ee65);
  border: 0 solid #e5e7eb;border-
  radius: .5rem; box-sizing:
  border-box; color: #482307;
  column-gap: 1rem;
  cursor: pointer; display:
  flex;
  font-family: ui-sans-serif,system-ui,-apple-system,system-ui,"Segoe
UI",Roboto,"Helvetica Neue",Arial,"Noto Sans",sans-serif,"Apple Color Emoji","Segoe
UI Emoji","Segoe UI Symbol","Noto Color Emoji";
  font-size: 100%;
  font-weight: 700;line-
  height:  24px; margin:
  0;
  outline: 2px solid transparent;padding: 1rem
  1.5rem;
  text-align: center; text-
  transform: none;
  transition: all .1s cubic-bezier(.4, 0, .2, 1);user-select:
  none;
  -webkit-user-select: none; touch-
  action: manipulation;
  box-shadow: -6px 8px 10px rgba(81,41,10,0.1),0px 2px 2pxrgba(81,41,10,0.2);
  display: none;
}
.button2{
  appearance: button; background-color:
  transparent;
  background-image: linear-gradient(to bottom, rgb(252, 162, 162),#ee3737);
  border: 0 solid #e5e7eb;border-
  radius: .5rem; box-sizing:
  border-box; color: #482307;
  column-gap: 1rem;
  cursor: pointer; display:
  flex;
  font-family: ui-sans-serif,system-ui,-apple-system,system-ui,"Segoe
UI",Roboto,"Helvetica Neue",Arial,"Noto Sans",sans-serif,"Apple Color Emoji","Segoe
UI Emoji","Segoe UI Symbol","Noto Color Emoji";
  font-size: 100%;
  font-weight: 700;line-
  height:  24px; margin:
  0;
  outline: 2px solid transparent;padding: 1rem
  1.5rem;
  text-align: center; text-
  transform: none;
  transition: all .1s cubic-bezier(.4, 0, .2, 1);user-select:
  none;
 -
```



PHISHING WEBSITE DETECTION USING MACHINE LEARNING

Enter Your URL



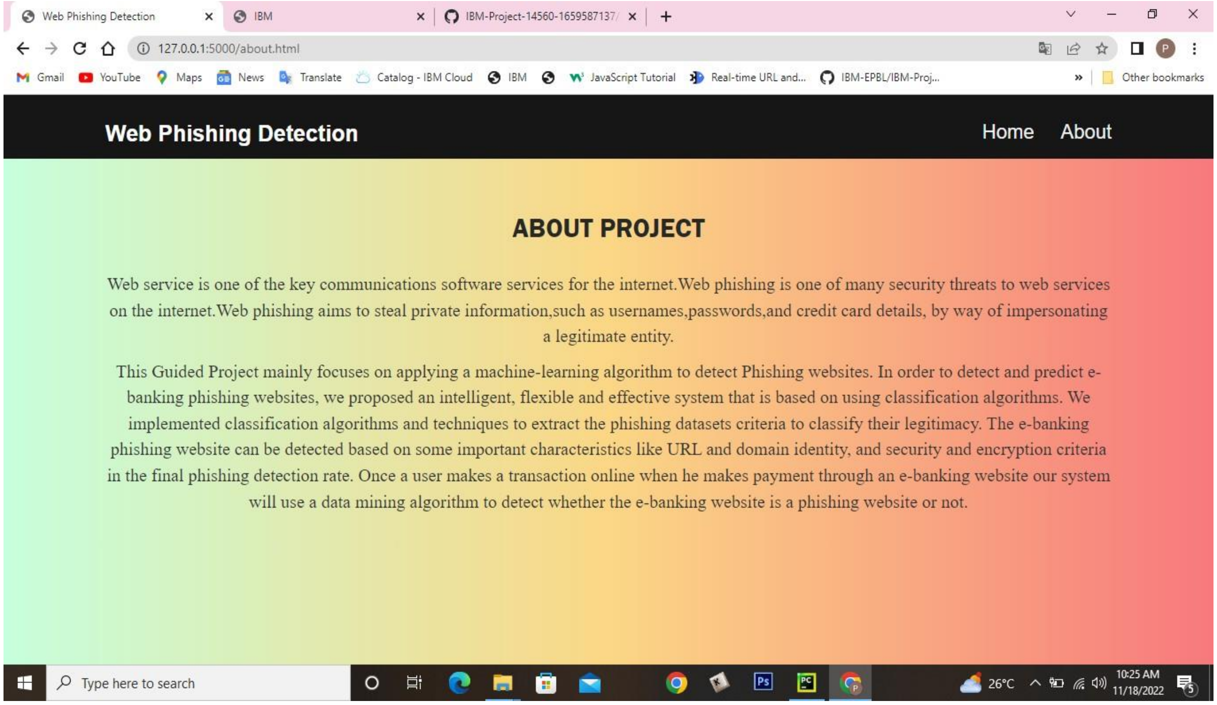
PHISHING WEBSITE DETECTION USING MACHINE LEARNING

dfwdiesel.net.

URL

dfwdiesel.net.

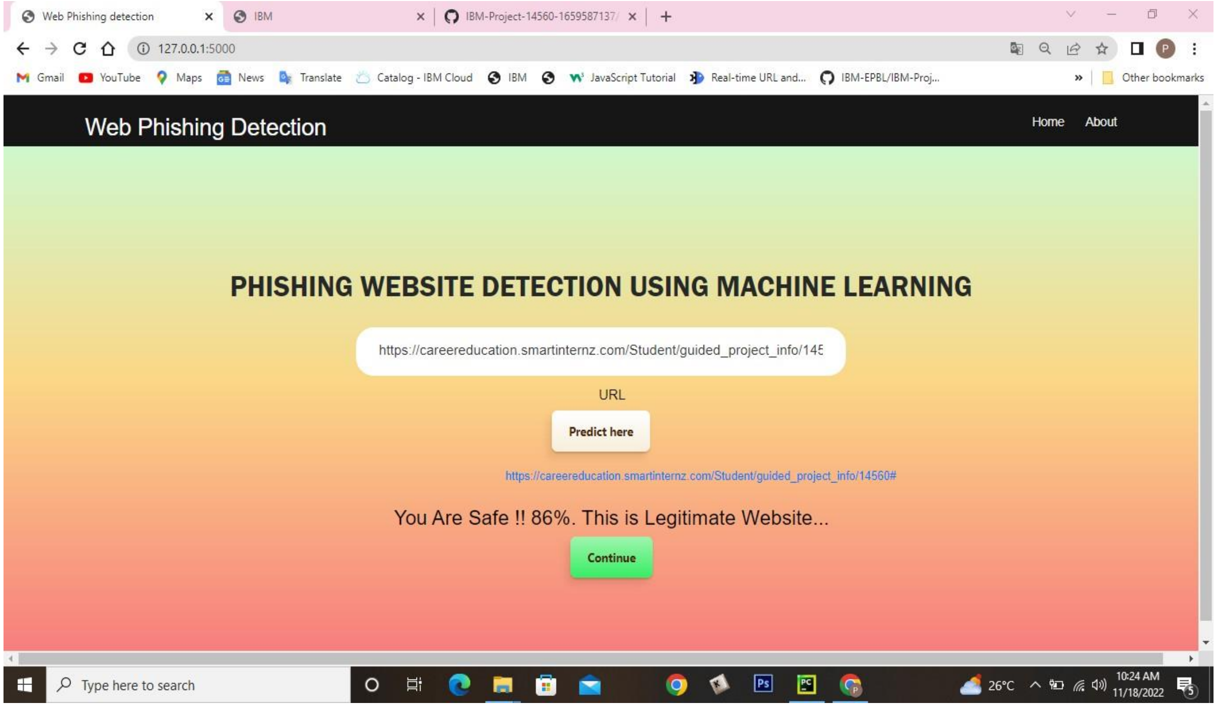
You Are On The Wrong Site !! 100% .Be Cautious...



ABOUT PROJECT

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.

This Guided Project mainly focuses on applying a machine-learning algorithm to detect Phishing websites. In order to detect and predict e-banking phishing websites, we proposed an intelligent, flexible and effective system that is based on using classification algorithms. We implemented classification algorithms and techniques to extract the phishing datasets criteria to classify their legitimacy. The e-banking phishing website can be detected based on some important characteristics like URL and domain identity, and security and encryption criteria in the final phishing detection rate. Once a user makes a transaction online when he makes payment through an e-banking website our system will use a data mining algorithm to detect whether the e-banking website is a phishing website or not.



PHISHING WEBSITE DETECTION USING MACHINE LEARNING

https://careereducation.smartinternz.com/Student/guided_project_info/145

URL

https://careereducation.smartinternz.com/Student/guided_project_info/14560#

You Are Safe !! 86%. This is Legitimate Website...

GITHUB & PROJECT DEMO LINK

https://github.com/IBM-EPBL/IBM-Project-145601659587137