WEB PHISHING DETECTION

IBM-Project-42143-1660652065

PROFESSIONAL READLINES FOR INNOVATION, EMPLOYABILITY AND ENTERPERNEURSHIP

PROJECT REPORT

TEAM ID : PNT2022TMID40190

TEAM MEMBERS

1. ARVINSARATH B [512719104003]
2. LOKESH V [512719104013]
3. MUTHARASAN P [512719104018]
4. SETHURAMAN S R [512719104019]
5. YUVARAJ E [512719104027]

BACHELOUR OF ENGINEERING IN COMPUTER SCIENCE AND ENGINEERING

CHENNAI INSTITUTE OF TECHNOLOGY
CHENNAI-600060

1. INTRODUCTION

1.1 Project Overview

Phishing is one of the most severe cyber-attacks where researchers are interested to find a solution. In phishing, attackers lure end-users and steal their personal in-formation. To minimize the damage caused by phishing must be detected as early as possible. There are various phishing attacks like spear phishing, whaling, vishing, smishing, pharming and so on. There are various phishing detection techniques based on white-list, black-list, content-based, URL-based, visual-similarity and machine-learning. In this paper, we discuss various kinds of phishing attacks, attack vectors and detection techniques for detecting the phishing sites.

1.2 Purpose

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.

2. LITERATURE SURVEY

2.1 Existing Problem

2.1.1 Protecting user against phishing using Antiphishing: -

Anti Phish is used to avoid users from using fraudulent web sites which in turn may lead to phishing attack. Here, AntiPhish traces the sensitive information to be filled by the user and alerts the user whenever he/she is attempting to share his/her information to a untrusted web site. The much effective elucidation for this is cultivating the users to approach only for trusted websites. However, this approach is unrealistic. Anyhow, the user may get tricked. Hence, it becomes mandatory for the associates to present such explanations to overcome the problem of phishing. Widely accepted alternatives are based on the creepy websites for the identification of “clones” and maintenance of records of phishing websites which are in hit list.

2.1.2 Learning to Detect Phishing Emails:

An alternative for detecting these attacks is a relevant process of reliability of machine on a trait intended for the reflection of the besieged deception of user by means of electronic communication. This approach can be used in the detection of phishing websites, or the text messages sent through emails that are used for trapping the victims. Approximately, 800 phishing mails and 7,000 nonphishing mails are traced till date and are detected accurately over 95% of them along with the categorization on the basis of 0.09% of the genuine emails. We can just wrap up with the methods for identifying the deception, along with the progressing nature of attacks .

2.1.3 Phishing detection system for e-banking using fuzzy data mining: -

Phishing websites, mainly used for e-banking services, are very complex and dynamic to be identified and classified. Due to the involvement of various ambiguities in the detection, certain crucial data mining techniques may prove an effective means in keeping the e-commerce websites safe since it deals with considering various quality factors rather than exact values. In this paper, an effective approach to overcome the "fuzziness" in the e-banking phishing website assessment is used an intelligent resilient and effective model for detecting e-banking phishing websites is put forth. The applied model is based on fuzzy logics along with data mining algorithms to consider various effective factors of the e-banking phishing website.

2.1.4 Collaborative Detection of Fast Flux Phishing Domains:-

Here, two approaches are defined to find correlation of evidences from multiple servers of DNS and multiple suspects of FF domain. Real life examples can be used to prove that our correlation approaches expedite the detection of the FF domain, which are based on an analytical model which can quantify various DNS queries that are required to verify a FF domain. It also shows implementation of correlation schemes on a huge level by using a distributed model, that is more scalable as compared to a centralized one, is publish N subscribe correlation model known as LARSID. In deduction, it is quite difficult to detect the FF domains in a accurate and timely manner, as the screen of proxies is used to shield the FF Mother ship. A theoretical approach is used to analyze the problem of FF detection by calculating the number of DNS queries required to get back a certain amount of unique IP addresses

2.1.5 A Prior-based Transfer Learning Method for the Phishing Detection: -

A logistic regression is the root of a priority based transferrable learning method, which is presented here for our classifier of statistical machine learning. It is used for the detection of the phishing websites depending on our selected characteristics of the URLs. Due to the divergence in the allocation of the features in the distinct phishing areas, multiple models are proposed for different regions. It is almost impractical to gather enough data from a new area to restore the detection model and use the transfer learning algorithm for adjusting the existing model. An appropriate way for phishing detection is to use our URLbased method. To cope with all the prerequisites of failure of detecting characteristics, we have to adopt the transferring method to generate a more effective model

2.2 Reference

[1]. "Protecting Users Against Phishing Attacks with AntiPhish" Engin Kirda and Christopher Kruegel Technical University of Vienna

[2]. "Learning to Detect Phishing Emails" Ian Fette School of Computer Science Carnegie Mellon University Pittsburgh, PA, 15213, USA icf@cs.cmu.edu Norman Sadeh School of Computer Science Carnegie Mellon University Pittsburgh, PA, 15213, USA Anthony Tomasic School of Computer Science Carnegie Mellon University Pittsburgh, PA, 15213, USA

[3]. Modeling and Preventing Phishing Attacks by Markus Jakobsson, Phishing detection system for e -banking using fuzzy data mining by Aburrous, M. ; Dept. of Comput., Univ. of Bradford, Bradford, UK ; Hossain, M.A. ; Dahal, K. ; Thabatah, F.

[4] M. Chandrasekaran, et al., "Phishing email detection based on structural properties", in New York State Cyber Security Conference (NYS) , Albany, NY ," 2006

[5] P. R. a. D. L. Ganger, ”Gone phishing: Evaluating anti-phishing tools for windows. Technical report, ,” September 2006

[6] M. Bazarganigilani, ”Phishing E-Mail Detection Using Ontology Concept and Nave Bayes Algorithm,” International Journal of Research and Reviews in Computer Science, vol. 2,no.2, 2011.

[7] M. Chandrasekaran, et al., ”Phoney: Mimicking user response to detect phishing attacks,” in In: Symposium on World of Wireless, Mobile and Multimedia Networks, IEEE Computer Society, 2006, pp. 668-672

[8] I. Fette, et al., ”Learning to detect phishing emails,” in Proc. 16th International World Wide Web Conference (WWW 2007), ACM Press, New York, NY, USA, May 2007, pp. 649-656

[9] A. Bergholz, et al., ”Improved phishing detection using model-based features,” in Proc. Conference on Email and Anti-Spam (CEAS). Mountain View Conf, CA, aug 2008

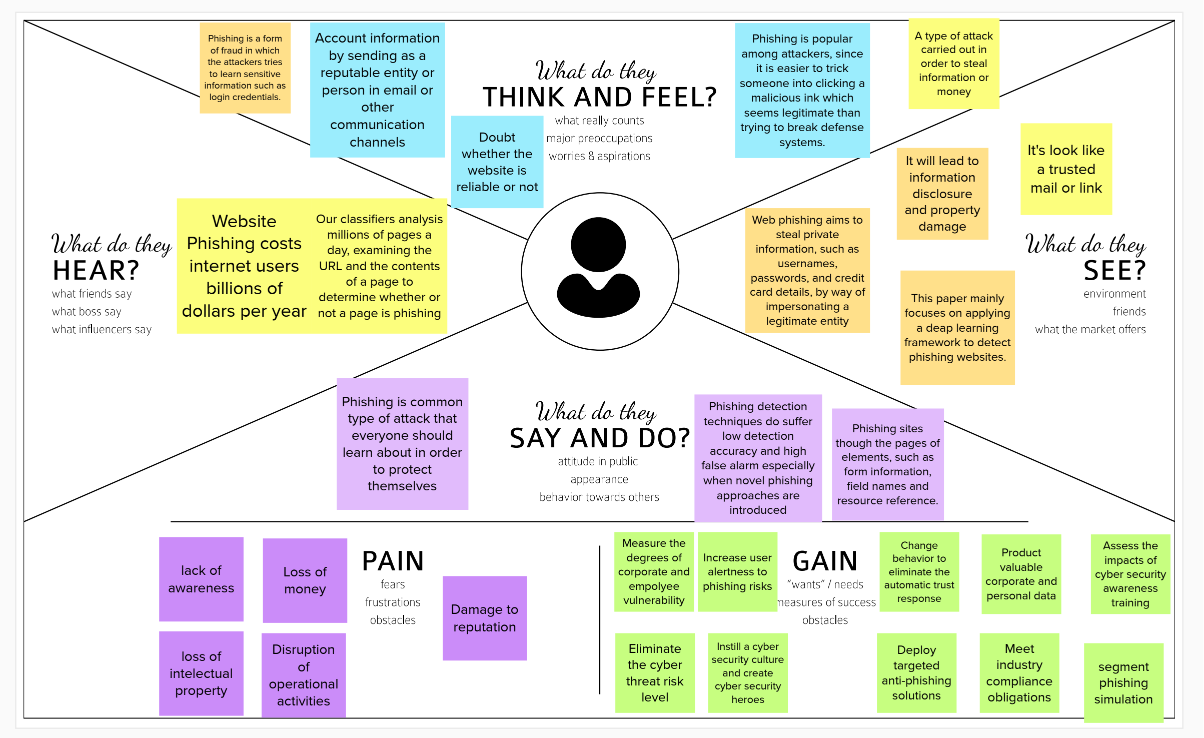
[10] L. Ma, et al.,”Detecting phishing emails using hybrid features,”IEEE Conf, 2009, pp. 493-497

2.3 Problem Statement Definition

Phishing is a fraudulent technique that is used over the internet to manipulate user to extract their personal information such as Username, Passwords, Credit Cards, Bank Account information etc. Phishing use multiple methods, including E-mail, Uniform Resource Locators(URL’s), Instant messages, Form posting, Telephone calls and Text messages to steal user information. Many cypher infiltrations are accomplished through phishing attacks where user are tricked into interacting with web pages that appear to be legitimate. This project aim tto develop these methods of defense utilizing various approaches to categorising Websites and narrow them down to the best Machine Learning algorithm by comparing the accuracy rate, false positive and false negative rate of each algorithm.

3. IDEATION & PROPOSED SOLUTION

3.1 Empathy Map Canvas



What do they THINK AND FEEL?
what really counts
major preoccupations
worries & aspirations

Phishing is a form of fraud in which the attackers tries to learn sensitive information such as login credentials.

Account information by sending as a reputable entity or person in email or other communication channels

Doubt whether the website is reliable or not

Phishing is popular among attackers, since it is easier to trick someone into clicking a malicious ink which seems legitimate than trying to break defense systems.

A type of attack carried out in order to steal information or money

What do they HEAR?
what friends say
what boss say
what influencers say

Website Phishing costs internet users billions of dollars per year

Our classifiers analysis millions of pages a day, examining the URL and the contents of a page to determine whether or not a page is phishing

What do they SEE?
environment
friends
what the market offers

It's look like a trusted mail or link

It will lead to information disclosure and property damage

Web phishing aims to steal private information, such as usernames, passwords, and credit card details, by way of impersonating a legitimate entity

This paper mainly focuses on applying a deap learning framework to detect phishing websites.

What do they SAY AND DO?
attitude in public
appearance
behavior towards others

Phishing is common type of attack that everyone should learn about in order to protect themselves

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

Phishing sites though the pages of elements, such as form information, field names and resource reference.

PAIN
fears
frustrations
obstacles

lack of awareness

Loss of money

Damage to reputation

loss of intelectual property

Disruption of operational activities

GAIN
“wants” / needs
measures of success
obstacles

Measure the degrees of corporate and empolyee vulnerability

Increase user alertness to phishing risks

Change behavior to eliminate the automatic trust response

Product valuable corporate and personal data

Assess the impacts of cyber security awareness training

Eliminate the cyber threat risk level

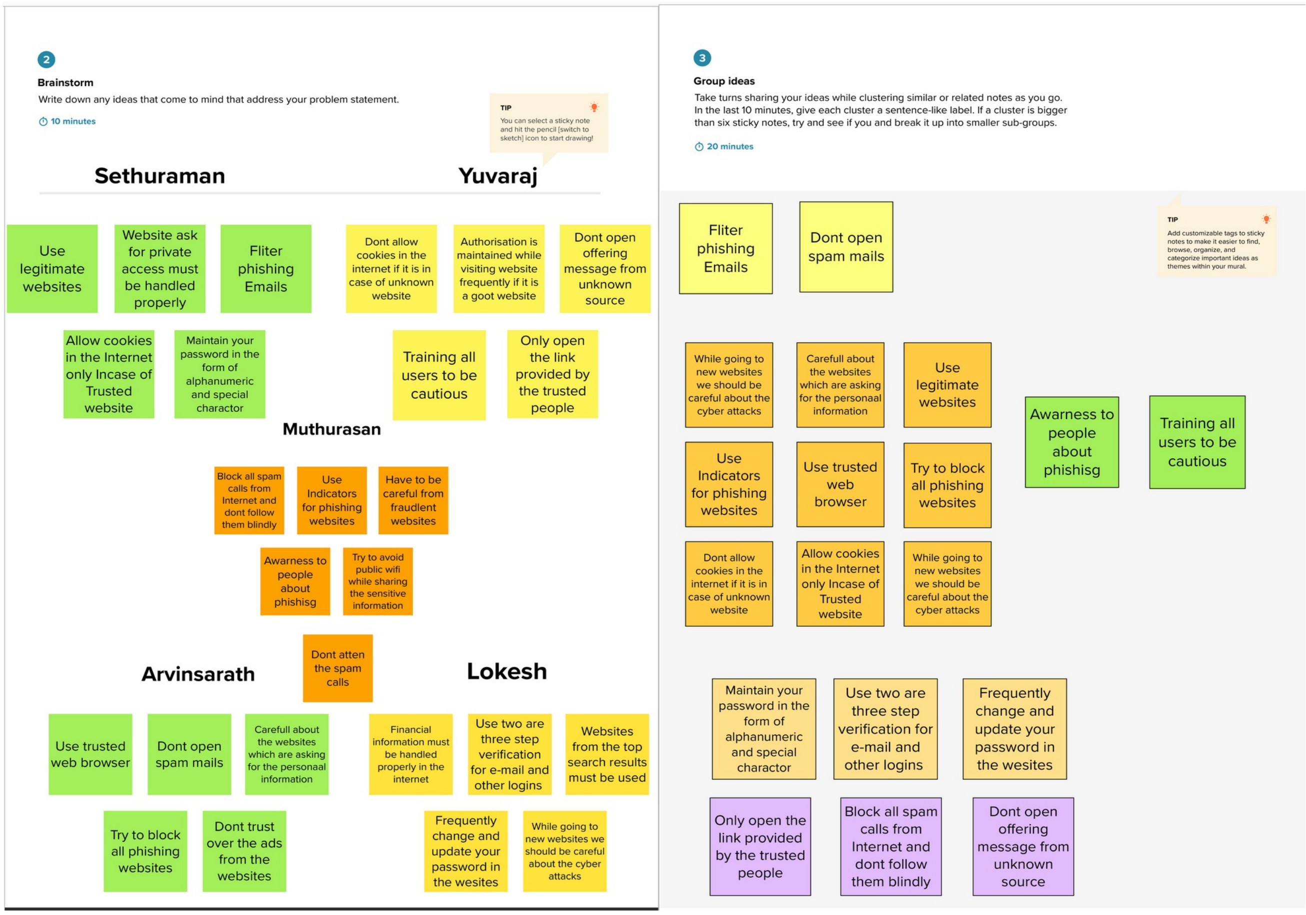
Instill a cyber security culture and create cyber security heroes

Deploy targeted anti-phishing solutions

Meet industry compliance obligations

segment phishing simulation

3.2 Ideation & Brainstorming



2

Brainstorm

Write down any ideas that come to mind that address your problem statement.

10 minutes

TIP

You can select a sticky note and hit the pencil [switch to sketch] icon to start drawing!

Sethuraman

Use legitimate websites

Website ask for private access must be handled properly

Fliter phishing Emails

Allow cookies in the Internet only Incase of Trusted website

Maintain your password in the form of alphanumeric and special charactor

Yuvaraj

Dont allow cookies in the internet if it is in case of unknown website

Authorisation is maintained while visiting website frequently if it is a goot website

Dont open offering message from unknown source

Training all users to be cautious

Only open the link provided by the trusted people

Muthurasan

Block all spam calls from Internet and dont follow them blindly

Use Indicators for phishing websites

Have to be careful from fraudlent websites

Awarness to people about phishisg

Try to avoid public wifi while sharing the sensitive information

Dont atten the spam calls

Arvinsarath

Use trusted web browser

Dont open spam mails

Carefull about the websites which are asking for the personaal information

Try to block all phishing websites

Dont trust over the ads from the websites

Lokesh

Financial information must be handled properly in the internet

Use two are three step verification for e-mail and other logins

Websites from the top search results must be used

Frequently change and update your password in the wesites

While going to new websites we should be careful about the cyber attacks

3

Group ideas

Take turns sharing your ideas while clustering similar or related notes as you go. In the last 10 minutes, give each cluster a sentence-like label. If a cluster is bigger than six sticky notes, try and see if you and break it up into smaller sub-groups.

20 minutes

TIP

Add customizable tags to sticky notes to make it easier to find, browse, organize, and categorize important ideas as themes within your mural.

Fliter phishing Emails

Dont open spam mails

While going to new websites we should be careful about the cyber attacks

Carefull about the websites which are asking for the personaal information

Use legitimate websites

Use Indicators for phishing websites

Use trusted web browser

Try to block all phishing websites

Dont allow cookies in the internet if it is in case of unknown website

Allow cookies in the Internet only Incase of Trusted website

While going to new websites we should be careful about the cyber attacks

Awarness to people about phishisg

Training all users to be cautious

Maintain your password in the form of alphanumeric and special charactor

Use two are three step verification for e-mail and other logins

Frequently change and update your password in the wesites

Only open the link provided by the trusted people

Block all spam calls from Internet and dont follow them blindly

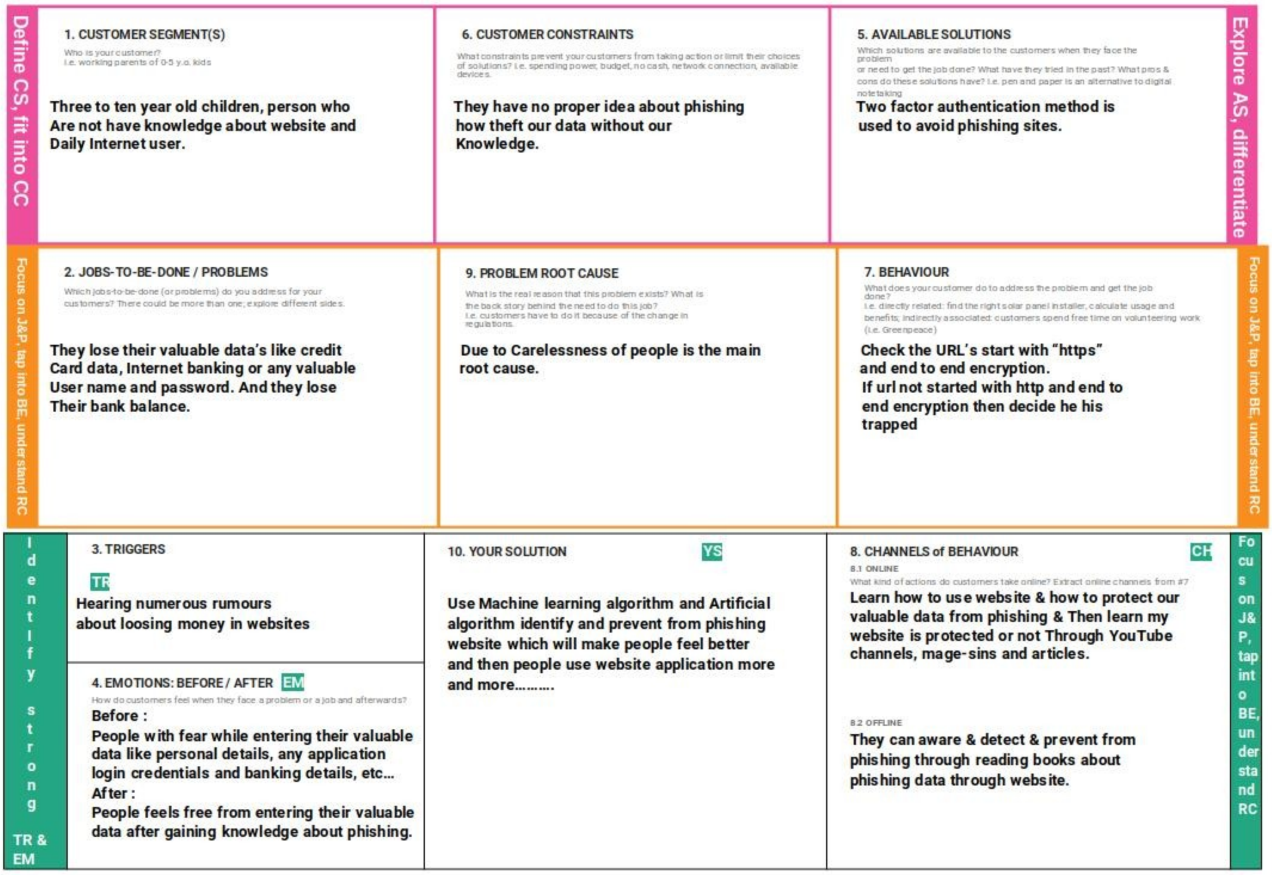
Dont open offering message from unknown source

3.3 Proposed Solution

S. No	Parameter	Description
1.	Problem Statement (Problem to be solved)	Phishing is a fraudulent technique that is used over the internet to manipulate user to extract their personal information such as Username, Passwords, Credit Cards, Bank Account information etc.Phishing use multiple methods, including E-mail, Uniform Resource Locators(URL's), Instant messages, Form posting, Telephone calls and Text messages to steal user information. Many cypher infiltrations are accomplished through phishing attacks where user are tricked into interacting with web pages that appear to be legitimate. This project aim tto develop these methods of defense utilizing various approaches to categorising Websites and narrow them down to the best Machine Learning algorithm by comparing the accuracy rate, false positive and false negative rate of each algorithm.
2.	Idea / Solution description	This project aim to develop these methods of defense utilizing various approaches to categorising Websites and narrow them down to the best Machine Learning algorithm by comparing the accuracy rate, false positive and false negative rate of each algorithm.To find unknown malicious urls compared to the

		blacklist approach.
3.	Novelty / Uniqueness	Our model uses the power of Machine learning to detect phishing sites. Python serves as a powerful tool to execute the application with Low false positives, High accuracy. Uses the latest techniques that gives an efficient and great performance. It can easily differentiate the fake and safe URL’s. If it’s fake means, a warning message will be intimate to the users.
4.	Social Impact / Customer Satisfaction	According to recent research by Google, these was a 4505 increase in phishing websites from January to March 2021. Phishing has a list of negative effects on a business, including loss of money, loss of intellectual property, damage to reputation, and disruption of operational activities. As an impact of this model, people can able to find fraudulent websites of fake ones. So that, they can avoid sharing sensitive data with unrecognized websites.
5.	Business Model (Revenue Model)	Our model can be used by all user’s to secure their data from malicious websites. It’s an open source tool.
6.	Scalability of the Solution	A-part from E-Banking sector the idea proposed can be developed into platfrom independent model. Adapts to all sort of web application and ease of preventing users from scam.

3.4 Problem Solution fit



Define CS, fit into CC

1. CUSTOMER SEGMENT(S)

Who is your customer?
i.e. working parents of 0-5 y.o. kids

Three to ten year old children, person who
Are not have knowledge about website and
Daily Internet user.

6. CUSTOMER CONSTRAINTS

What constraints prevent your customers from taking action or limit their choices of solutions? i.e. spending power, budget, no cash, network connection, available devices.

They have no proper idea about phishing
how theft our data without our
Knowledge.

5. AVAILABLE SOLUTIONS

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? i.e. pen and paper is an alternative to digital notetaking

Two factor authentication method is
used to avoid phishing sites.

Explore AS, differentiate

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

2. JOBS-TO-BE-DONE / PROBLEMS

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

They lose their valuable data’s like credit
Card data, Internet banking or any valuable
User name and password. And they lose
Their bank balance.

9. PROBLEM ROOT CAUSE

What is the real reason that this problem exists? What is the back story behind the need to do this job?
i.e. customers have to do it because of the change in regulations.

Due to Carelessness of people is the main
root cause.

7. BEHAVIOUR

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 (i.e. Greenpeace)

Check the URL’s start with “https”
and end to end encryption.
If url not started with http and end to
end encryption then decide he his
trapped

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

Identify strong TR & EM

3. TRIGGERS

TR

Hearing numerous rumours
about loosing money in websites

4. EMOTIONS: BEFORE / AFTER EM

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

Before :
People with fear while entering their valuable data like personal details, any application login credentials and banking details, etc…
After :
People feels free from entering their valuable data after gaining knowledge about phishing.

10. YOUR SOLUTION YS

Use Machine learning algorithm and Artificial
algorithm identify and prevent from phishing
website which will make people feel better
and then people use website application more
and more………

8. CHANNELS of BEHAVIOUR CH

8.1 ONLINE

What kind of actions do customers take online? Extract online channels from #7

Learn how to use website & how to protect our valuable data from phishing & Then learn my website is protected or not Through YouTube channels, mage-sins and articles.

8.2 OFFLINE

They can aware & detect & prevent from phishing through reading books about phishing data through website.

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

4. REQUIREMENT ANALYSIS

4.1 Functional requirement

Following are the functional requirements of the proposed solution.

FR No.	Functional Requirement (Epic)	Sub Requirement (Story / Sub-Task)
FR-1	User Input	User input an URL to chech it is legal or phishing site.
FR-2	Website Comparison	Model comparing the entered URL with the help of Blacklist and Whitelist.
FR-3	Feature extraction	After comparing, if none found on comparison the it extracts feature using heuristic and visual similarity approach.
FR-4	Prediction	Model Predicts the URL using Machine Learning algorithm such as Logistic Regression, KNN.
FR-5	Classifier	Model sends output to classifier and it produce final result.
FR-6	Announcement	Model the displays whether the website is a legal or phishing site.
FR-7	Events	Model needs the capability of reetrieving and displaying accurate result for a website.

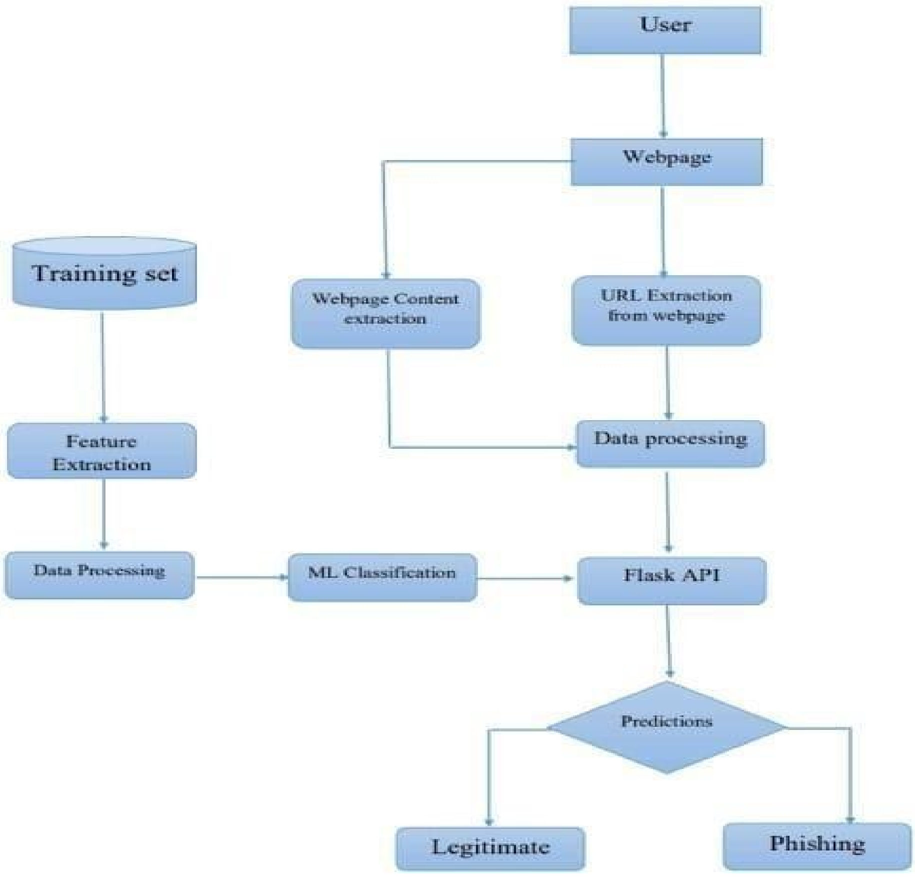
4.2 Non-Functional requirements

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

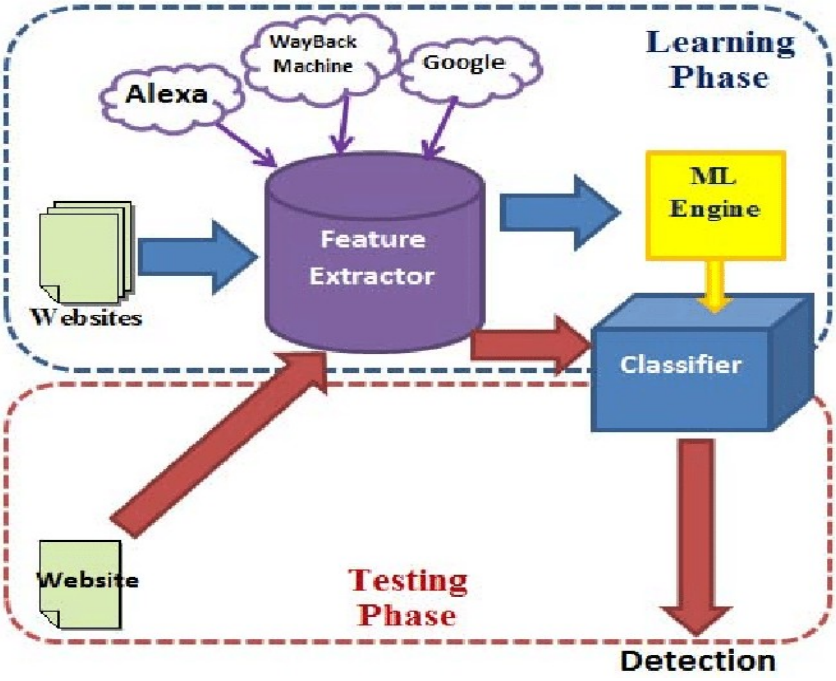
FR No.	Non-Functional Requirement	Description
NFR-1	**Usability**	A set of specifications that describe the system’s operation capabilities and constraints and attempt to improve its functionality.
NFR-2	**Security**	Assuring all data inside the system or its part will be protected against malware attacks or unauthorized access.
NFR-3	**Reliability**	This approach gives more accuracy then existing system.
NFR-4	**Performance**	Parameters for the proposed system gives accurate predicted value which is compared to the existing system.
NFR-5	**Availability**	The system is accessible by user at any time using web browser.
NFR-6	**Scalability**	The design will be suitable and performs with full efficieny according to rising demands.

5. PROJECT DESIGN

5.1 Data Flow Diagrams



5.2 Solution & Technical Architecture



5.3 User Stories

User Type	Functional Requirement (Epic)	User Story Number	User Story / Task	Acceptance criteria	Priority	Release
Customer (Web user)	Dashboard	USN-1	As a user, I can easily navigate through dashboard and I can use the dashboard to get details about app and instruction to use the app.	Using dashboard i can easily access the application.	High	Sprint-1
	Url prediction and Result page	CCE-2	As a user, i can able to enter the URL to predict and View the corresponding result to that entered URL.	I can enter the URL and able to view the result	High	Sprint-2
	Add URL and Experience page and About page	USN-3	As a user, i can share my perviously experienced Phishing site and View about page of the website	I can add or enter experience and submit it	High	Sprint-3
Model Buliding	Prediction of Phishing sites	M-1	As an User, I can enter the url and Predict it as a Phishing site or not.	I can predict the URL is bad or good	High	Sprint-4
Model Testing	Testing of Model is worked as properly	MT-1	If the model Predict the URL as Phishing site or not with accuracy rate above 95%.		High	Sprint-4

6. PROJECT PLANNING & SCHEDULING

6.1 Sprint Planning & Estimation

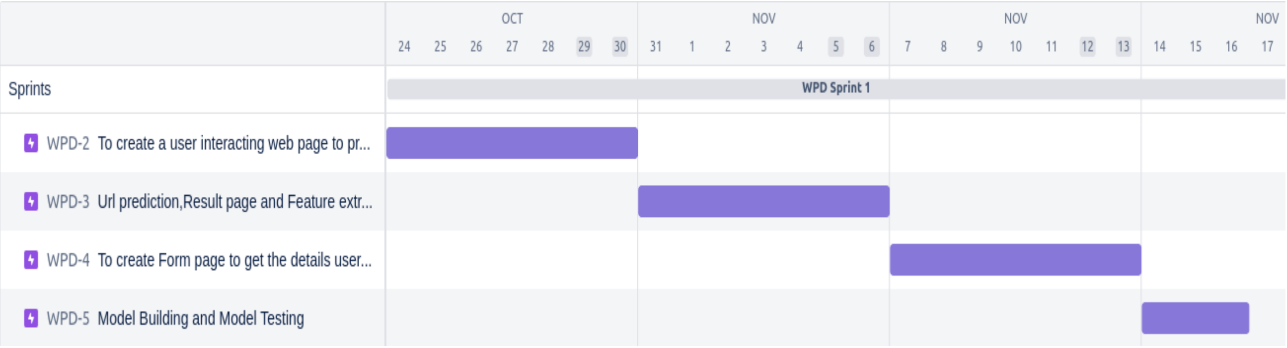
Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority	Team Members
Sprint-1	Home Page	USN-1	As a user, i can access the home page content	2	Medium	Lokesh V
Sprint-2	Url prediction and Result page	USN-2	As an User, I can enter the url and Predict it as a Phishing site or not.		High	Arvinsarath
Sprint-3	Add URL and Experience page and About page	USN-3	As an administrator, I can login and access the Experience form submitted by user.	1	High	Yuvaraj & mutharasan
Sprint-4	Model Buliding and Model Testing	MB/MT-1	If the model Predict the URL as Phishing site or not with accuracy rate above 95%.	2	High	Lokesh v & sethuraman

6.2 Sprint Delivery Schedule

Sprint	Total Story Points	Duration	Sprint Start Date	Sprint End Date (Planned)	Story Points Completed (as on Planned End Date)	Sprint Release Date (Actual)
Sprint-1	20	6 Days	24 Oct 2022	29 Oct 2022	20	29 Oct 2022

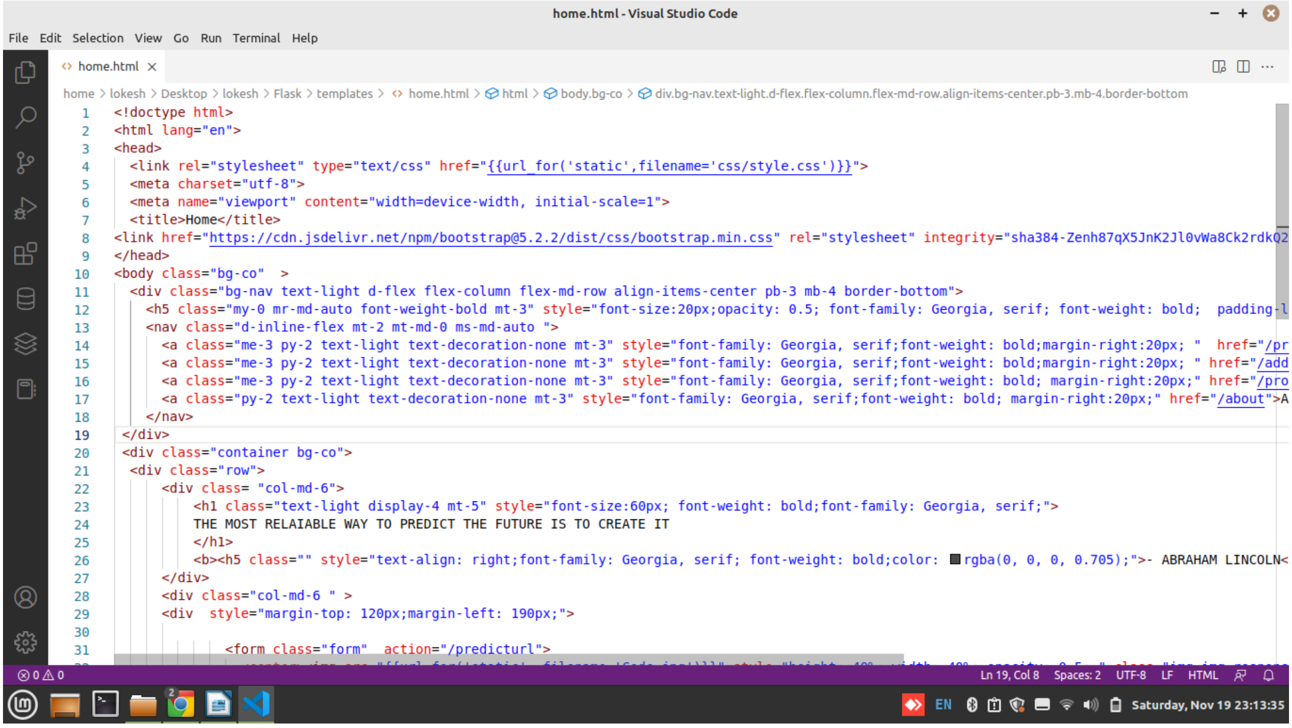
Sprint	Total Story Points	Duration	Sprint Start Date	Sprint End Date (Planned)	Story Points Completed (as on Planned End Date)	Sprint Release Date (Actual)
Sprint-2	20	6 Days	31 Oct 2022	05 Nov 2022	20	06 Nov 2022
Sprint-3	20	6 Days	07 Nov 2022	12 Nov 2022	20	12 Nov 2022
Sprint-4	20	6 Days	14 Nov 2022	19 Nov 2022	20	18 nov 2022

6.3 Reports from JIRA



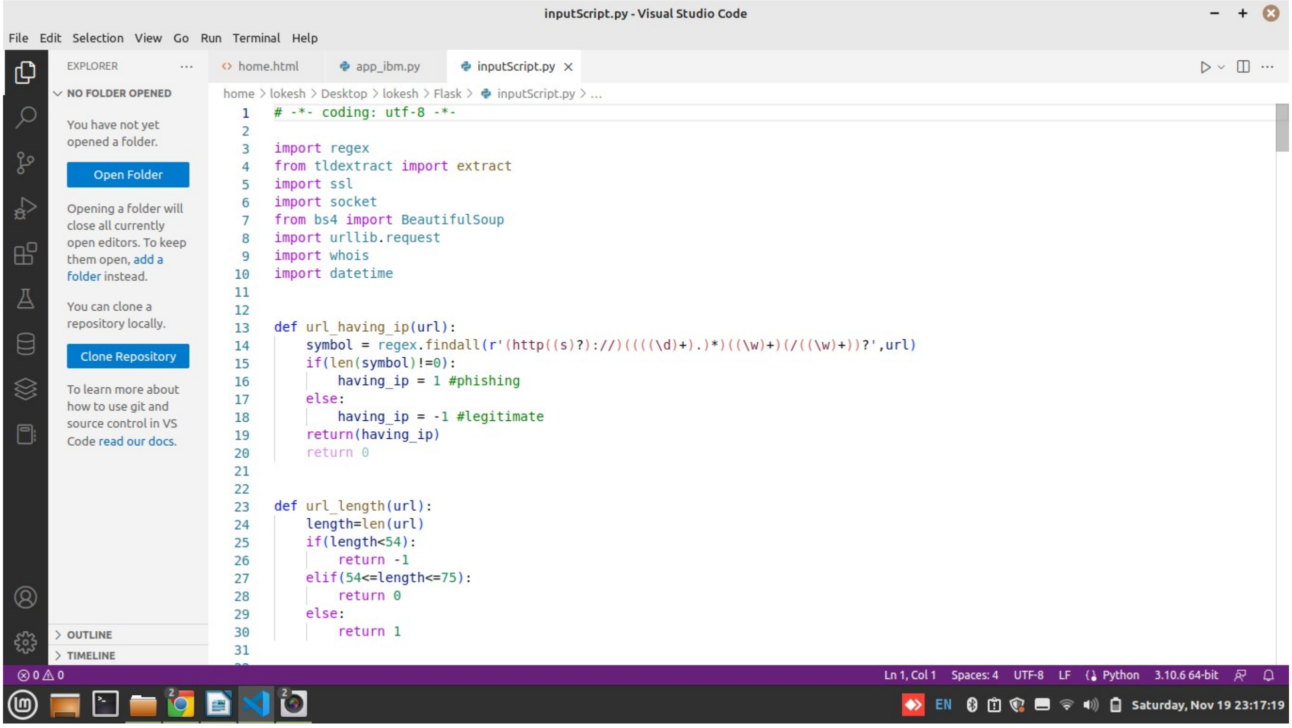
7. CODING & SOLUTIONING

7.1 Feature 1


home.html - Visual Studio Code
File Edit Selection View Go Run Terminal Help
home.html
home > lokesh > Desktop > lokesh > Flask > templates > home.html > html > body.bg-co > div.bg-nav.text-light.d-flex.flex-column.flex-md-row.align-items-center.pb-3.mb-4.border-bottom
⊗ 0 ⚠ 0
Ln 19, Col 8 Spaces: 2 UTF-8 LF HTML
EN
Saturday, Nov 19 23:13:35

```
<!doctype html>
<html lang="en">
<head>
  <link rel="stylesheet" type="text/css" href="{{url_for('static',filename='css/style.css')}}">
  <meta charset="utf-8">
  <meta name="viewport" content="width=device-width, initial-scale=1">
  <title>Home</title>
<link href="https://cdn.jsdelivr.net/npm/bootstrap@5.2.2/dist/css/bootstrap.min.css" rel="stylesheet" integrity="sha384-Zenh87qX5JnK2Jl0vWa8Ck2rdkQ2
</head>
<body class="bg-co"  >
  <div class="bg-nav text-light d-flex flex-column flex-md-row align-items-center pb-3 mb-4 border-bottom">
    <h5 class="my-0 mr-md-auto font-weight-bold mt-3" style="font-size:20px;opacity: 0.5; font-family: Georgia, serif; font-weight: bold;  padding-l
    <nav class="d-inline-flex mt-2 mt-md-0 ms-md-auto ">
      <a class="me-3 py-2 text-light text-decoration-none mt-3" style="font-family: Georgia, serif;font-weight: bold;margin-right:20px; "  href="/pr
      <a class="me-3 py-2 text-light text-decoration-none mt-3" style="font-family: Georgia, serif;font-weight: bold;margin-right:20px; " href="/add
      <a class="me-3 py-2 text-light text-decoration-none mt-3" style="font-family: Georgia, serif;font-weight: bold; margin-right:20px;" href="/pro
      <a class="py-2 text-light text-decoration-none mt-3" style="font-family: Georgia, serif;font-weight: bold; margin-right:20px;" href="/about">A
    </nav>
 </div>
 <div class="container bg-co">
  <div class="row">
      <div class= "col-md-6">
          <h1 class="text-light display-4 mt-5" style="font-size:60px; font-weight: bold;font-family: Georgia, serif;">
          THE MOST RELAIABLE WAY TO PREDICT THE FUTURE IS TO CREATE IT
          </h1>
          <b><h5 class="" style="text-align: right;font-family: Georgia, serif; font-weight: bold;color: ■rgba(0, 0, 0, 0.705);">- ABRAHAM LINCOLN<
      </div>
      <div class="col-md-6 " >
      <div  style="margin-top: 120px;margin-left: 190px;">

              <form class="form"  action="/predicturl">
```

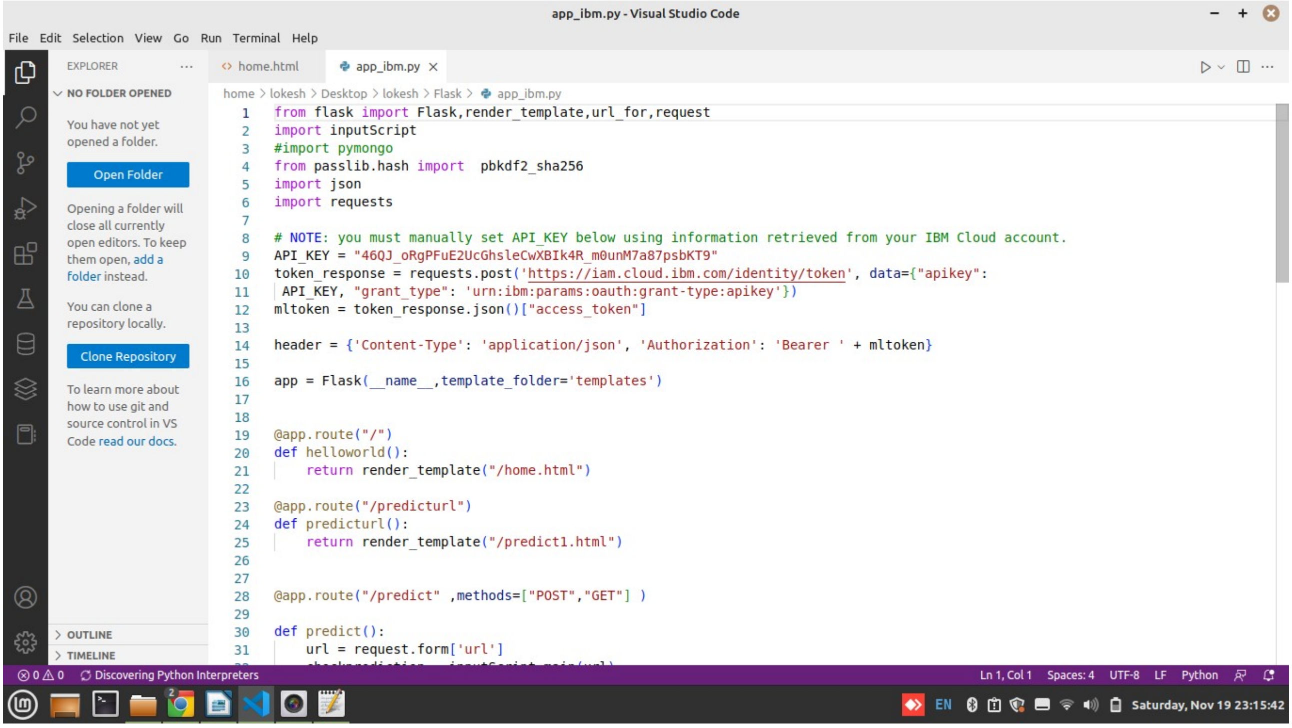
Figure 1: Home page


inputScript.py - Visual Studio Code
File Edit Selection View Go Run Terminal Help
EXPLORER
NO FOLDER OPENED
You have not yet opened a folder.
Open Folder
Opening a folder will close all currently open editors. To keep them open, add a folder instead.
You can clone a repository locally.
Clone Repository
To learn more about how to use git and source control in VS Code read our docs.
OUTLINE
TIMELINE
home.html
app_ibm.py
inputScript.py
home > lokesh > Desktop > lokesh > Flask > inputScript.py > ...
Ln 1, Col 1 Spaces: 4 UTF-8 LF Python 3.10.6 64-bit
EN
Saturday, Nov 19 23:17:19

```python
# -*- coding: utf-8 -*-

import regex
from tldextract import extract
import ssl
import socket
from bs4 import BeautifulSoup
import urllib.request
import whois
import datetime


def url_having_ip(url):
    symbol = regex.findall(r'(http((s)?)://)((((\d)+).)*)((\w)+)(/((\w)+))?',url)
    if(len(symbol)!=0):
        having_ip = 1 #phishing
    else:
        having_ip = -1 #legitimate
    return(having_ip)
    return 0


def url_length(url):
    length=len(url)
    if(length<54):
        return -1
    elif(54<=length<=75):
        return 0
    else:
        return 1
```

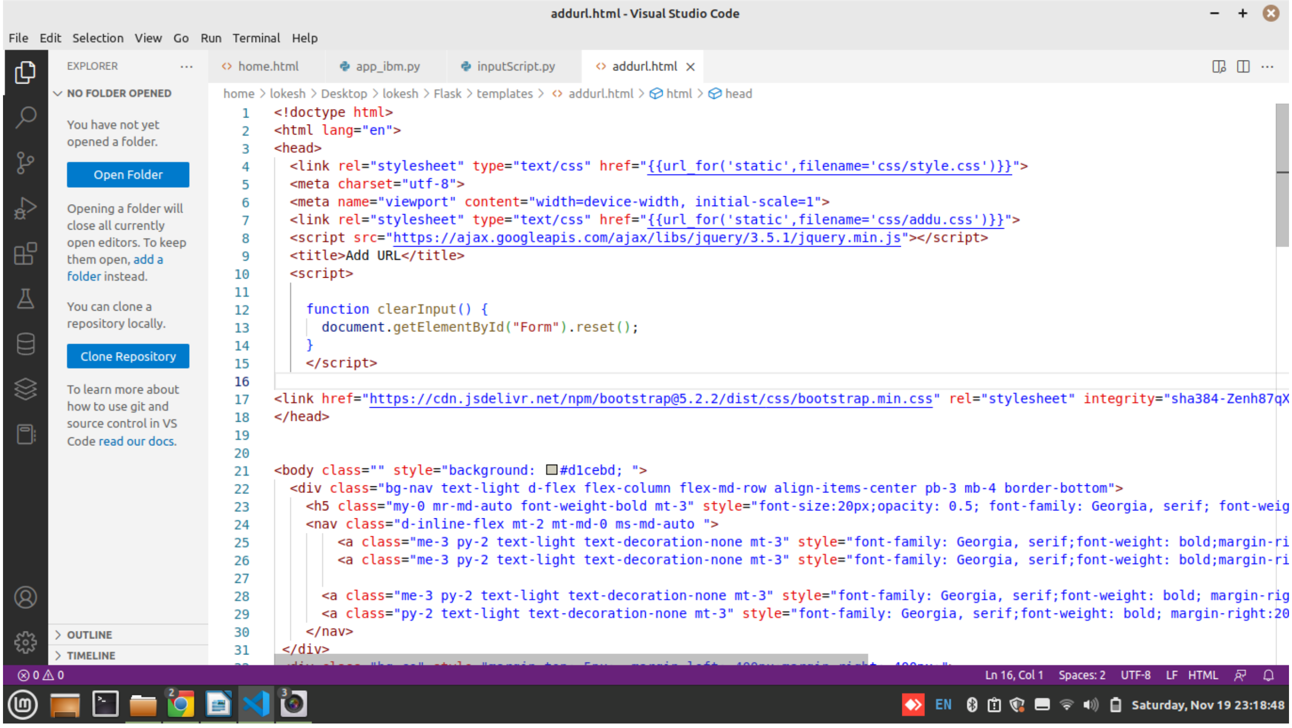
Figure 2: Input Script

7.2 Feature 2


app_ibm.py - Visual Studio Code
File Edit Selection View Go Run Terminal Help
EXPLORER
NO FOLDER OPENED
You have not yet opened a folder.
Open Folder
Opening a folder will close all currently open editors. To keep them open, add a folder instead.
You can clone a repository locally.
Clone Repository
To learn more about how to use git and source control in VS Code read our docs.
OUTLINE
TIMELINE
home.html
app_ibm.py
home > lokesh > Desktop > lokesh > Flask > app_ibm.py
Discovering Python Interpreters
Ln 1, Col 1 Spaces: 4 UTF-8 LF Python
EN
Saturday, Nov 19 23:15:42

```python
from flask import Flask,render_template,url_for,request
import inputScript
#import pymongo
from passlib.hash import  pbkdf2_sha256
import json
import requests

# NOTE: you must manually set API_KEY below using information retrieved from your IBM Cloud account.
API_KEY = "46QJ_oRgPFuE2UcGhsleCwXBIk4R_m0unM7a87psbKT9"
token_response = requests.post('https://iam.cloud.ibm.com/identity/token', data={"apikey":
 API_KEY, "grant_type": 'urn:ibm:params:oauth:grant-type:apikey'})
mltoken = token_response.json()["access_token"]

header = {'Content-Type': 'application/json', 'Authorization': 'Bearer ' + mltoken}

app = Flask(__name__,template_folder='templates')


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

@app.route("/predicturl")
def predicturl():
    return render_template("/predict1.html")


@app.route("/predict" ,methods=["POST","GET"] )

def predict():
    url = request.form['url']
```

Figure 3: Flask file

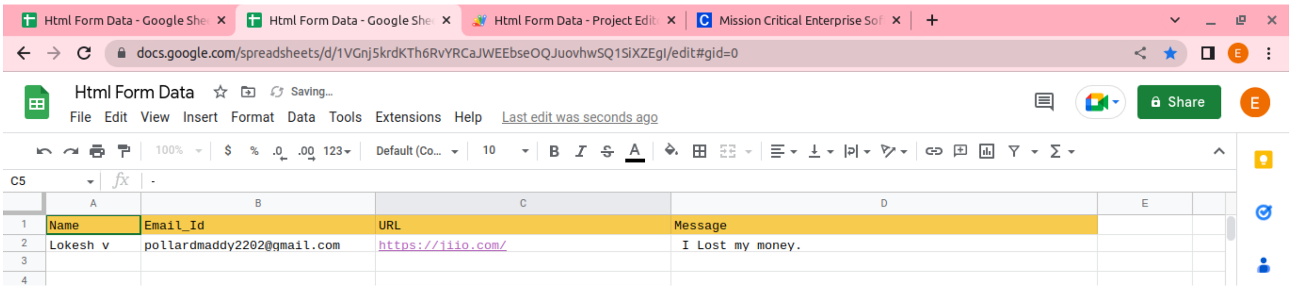


```
<!doctype html>
<html lang="en">
<head>
  <link rel="stylesheet" type="text/css" href="{{url_for('static',filename='css/style.css')}}">
  <meta charset="utf-8">
  <meta name="viewport" content="width=device-width, initial-scale=1">
  <link rel="stylesheet" type="text/css" href="{{url_for('static',filename='css/addu.css')}}">
  <script src="https://ajax.googleapis.com/ajax/libs/jquery/3.5.1/jquery.min.js"></script>
  <title>Add URL</title>
  <script>

    function clearInput() {
      document.getElementById("Form").reset();
    }
    </script>

<link href="https://cdn.jsdelivr.net/npm/bootstrap@5.2.2/dist/css/bootstrap.min.css" rel="stylesheet" integrity="sha384-Zenh87qX
</head>


<body class="" style="background: #d1cebd; ">
  <div class="bg-nav text-light d-flex flex-column flex-md-row align-items-center pb-3 mb-4 border-bottom">
    <h5 class="my-0 mr-md-auto font-weight-bold mt-3" style="font-size:20px;opacity: 0.5; font-family: Georgia, serif; font-weig
    <nav class="d-inline-flex mt-2 mt-md-0 ms-md-auto ">
        <a class="me-3 py-2 text-light text-decoration-none mt-3" style="font-family: Georgia, serif;font-weight: bold;margin-ri
        <a class="me-3 py-2 text-light text-decoration-none mt-3" style="font-family: Georgia, serif;font-weight: bold;margin-ri

      <a class="me-3 py-2 text-light text-decoration-none mt-3" style="font-family: Georgia, serif;font-weight: bold; margin-rig
      <a class="py-2 text-light text-decoration-none mt-3" style="font-family: Georgia, serif;font-weight: bold; margin-right:20
    </nav>
 </div>
```

addurl.html - Visual Studio Code
File Edit Selection View Go Run Terminal Help
home.html
app_ibm.py
inputScript.py
addurl.html
home > lokesh > Desktop > lokesh > Flask > templates > addurl.html > html > head
Ln 16, Col 1
Spaces: 2
UTF-8
LF
HTML
Saturday, Nov 19 23:18:48

Figure 4: Add Experience Page

7.3 Database Schema



	A	B	C	D
1	Name	Email_Id	URL	Message
2	Lokesh v	pollardmaddy2202@gmail.com	https://iiio.com/	I Lost my money.

Html Form Data - Google She
Html Form Data - Project Edit
Mission Critical Enterprise Sof
docs.google.com/spreadsheets/d/1VGnj5krdKTh6RvYRCaJWEEbseOQJuovhwSQ1SiXZEgI/edit#gid=0
Html Form Data
Saving...
File Edit View Insert Format Data Tools Extensions Help
Last edit was seconds ago
Share
C5

8. TESTING

8.1 Test Cases

Test Case ID	Test Case Description	Test Steps
TC01	Check Predict button is rooted to Prediction page	In home page, Click Prediction URL button.
TC02	In Prediction Page, Check prediction of url is done or not.	In prediction page, 1. Enter Url 2. Then press Prediction Button to predict URL
TC03	In Prediction output page, check the “Predict another URL” button.	In result page, press Predict another URL button.
TC04	In, Prediction Page,	In prediction page,

	Check Prediction is done in positive and negative.	1.Enter URL for good site and bad site. 2.then press Predict button.
TC05	Check User experience form is submitted in google form or not.	In add URL page, 1.Enter the Rrequired fields. 2.press submit button.
TC06	Check About button root to About page.	Press about button.
TC07	Check project Details button root's to Project details button.	Press Project details button
TC08	Check all buttons are working properly or not	Press all button and check it root's to corresponding page or not.

8.2 User Acceptance Testing

8.2.1 Purpose of Document

This document is to briefly explain the test coverage and open issues of the Web Phishing Detection project at the time of the release to User Acceptance Testing (UAT).

8.2.2 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	8	4	2	3	17
Duplicate	1	0	3	0	4
External	0	3	0	1	4
Fixed	9	2	4	15	30
Not Reproduced	0	0	1	0	1
Skipped	0	0	1	1	2
Won't Fix	0	5	2	1	8
Totals	18	14	13	20	65

8.2.3 Test Case Analysis

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

Section	Total Cases	Not Tested	Fail	Pass
Print Engine	2	0	0	2
Client Application	2	0	0	4
Security	1	0	0	1
Outsource Shipping	1	0	0	1
Exception Reporting	1	0	0	1
Final Report Output	1	0	0	1
Version Control	1	0	0	1

9. RESULTS

9.1 Performance Metrics

Our execution confirms that we had successfully implemented our project work and we had also tested them is different cases in the given timeline. Our project is distributes the work of design, implementation, testing and documentation in different levels so that we can complete our project on time. As the result, Our project Machine learning model predict Url is good or bad with 96% accuracy.

10. ADVANTAGES & DISADVANTAGES

Measure the degrees of corporate and employee vulnerability.Eliminate the cyber threat risk level. Increase user alertness to phishing risks. Instill a cyber security culture and create cyber security heroes.

11. CONCLUSION

Our execution confirms that we had successfully implemented our project work and we had also tested them in different cases in the given timeline. Our project is distributes the work of design, implementation, testing and documentation in different levels so that we can complete our project on time. The results generated are up to the expected marks from which we concluded that our project is accomplised effectively, As a proof of completion we had produce the Demo video link and Coding of the project in our Documentation.

12. FUTURE SCOPE

We were planning to create a Google extension to predict whether a URL is Trusted or not.

13. APPENDIX

Source Code

app_ibm.py

```
from flask import Flask,render_template,url_for,request
import inputScript
#import pymongo
from passlib.hash import  pbkdf2_sha256
import json
import requests

# NOTE: you must manually set API_KEY below using information retrieved from your IBM Cloud account.
API_KEY = "46QJ_oRgPFuE2UcGhsleCwXBIk4R_m0unM7a87psbKT9"
token_response = requests.post('https://iam.cloud.ibm.com/identity/token', data={"apikey":
 API_KEY, "grant_type": 'urn:ibm:params:oauth:grant-type:apikey'})
mltoken = token_response.json()["access_token"]

header = {'Content-Type': 'application/json', 'Authorization': 'Bearer ' + mltoken}

app = Flask(__name__,template_folder='templates')


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

@app.route("/predicturl")
def predicturl():
    return render_template("/predict1.html")


@app.route("/predict" ,methods=["POST","GET"] )

def predict():
    url = request.form['url']
    checkprediction = inputScript.main(url)

    print(url)
    print(checkprediction)
    # NOTE: manually define and pass the array(s) of values to be scored in the next line
    payload_scoring = {"input_data": [{"fields":
[['f0','f1','f2','f3','f4','f5','f6','f7','f8','f9','f9','f10','f11','f12','f13','f14','f15','f15','f16','f17','f18','f19','f
20','f21','f22','f23','f24','f25','f26','f27']], "values":checkprediction }]}

    response_scoring = requests.post('https://us-south.ml.cloud.ibm.com/ml/v4/deployments/
62efb8db-e32e-4c70-bd7c-7f819762d9b7/predictions?version=2022-11-12',
json=payload_scoring,headers={'Authorization': 'Bearer ' + mltoken})
    print("Scoring response")
    print(response_scoring.json())
    pred = response_scoring.json()
    output = pred['predictions'][0]['values'][0][0]


    if output==1 :
        return render_template("/output1.html")

    elif output==-1 :
        return render_template("/output.html")

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

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

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


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

Prediction.html

```
<!doctype html>
<html lang="en">
<head>
  <link rel="stylesheet" type="text/css" href="{{url_for('static',filename='css/style.css')}}">
  <meta charset="utf-8">
  <meta name="viewport" content="width=device-width, initial-scale=1">
  <title>URL Prediction</title>

<script>
      function clearInput() {
            document.getElementById("Form").reset();
}
</script>
```

```
<link href="https://cdn.jsdelivr.net/npm/bootstrap@5.2.2/dist/css/bootstrap.min.css"
rel="stylesheet" integrity="sha384-
Zenh87qX5JnK2Jl0vWa8Ck2rdkQ2Bzep5IDxbcnCeuOxjzrPF/et3URy9Bv1WTRi"
crossorigin="anonymous">

</head>

<body class="bg-co">
  <div class="bg-nav text-light d-flex flex-column flex-md-row align-items-center pb-3 mb-4
border-bottom">
     <h5 class="my-0 mr-md-auto font-weight-bold mt-3" style="font-size:20px;opacity: 0.5;
font-family: Georgia, serif; font-weight: bold;  padding-left: 50px;">URL Prediction</h5>
   <nav class="d-inline-flex mt-2 mt-md-0 ms-md-auto ">
       <a class="me-3 py-2 text-light text-decoration-none mt-3" style="font-family: Georgia,
serif;font-weight: bold;margin-right:20px; "  href="/">Home</a>
       <a class="me-3 py-2 text-light text-decoration-none mt-3" style="font-family: Georgia,
serif;font-weight: bold;margin-right:20px; " href="/addurl">Add url</a>
       <a class="me-3 py-2 text-light text-decoration-none mt-3" style="font-family: Georgia,
serif;font-weight: bold; margin-right:20px;" href="/project_details">Project Details</a>
      <a class="py-2 text-light text-decoration-none mt-3" style="font-family: Georgia, serif;font-
weight: bold; margin-right:20px;" href="/about">About</a>
   </nav>
 </div>

<div class="bg-co" style="margin-top: 120px;  margin-left: 400px;margin-right: 400px;">
  <div class="card-body">
    <form id="Form" action="/predict" method='post'class="form">
        <img src="{{url_for('static', filename='code.png')}}" style="height: 20%; width: 20%;"
class="img img-responsive img-circle  mx-auto d-block" /><br>
       <label><b>Enter URL to predict</b></label><br>
         <input type="text" name="url" id= "myText" placeholder="Ex : https://google.com/"
class="form-control" required><br>
         <div class="w3-bar "><center>
         <input type="submit"  style="background-color: black; font-weight: bold; color: white;"
class="btn "  value="Predict URL" >
         <input type= "button" style="background-color: black; font-weight: bold; color: white;"
class=" btn " value= "Clear" onclick= "clearInput()">
         </div>
    </form>
   </div>
   </div>

<script src="https://cdn.jsdelivr.net/npm/@popperjs/core@2.11.6/dist/umd/popper.min.js"
integrity="sha384-oBqDVmMz9ATKxIep9tiCxS/Z9fNfEXiDAYTujMAeBAsjFuCZSmKbSSU
nQlmh/jp3" crossorigin="anonymous"></script>
<script src="https://cdn.jsdelivr.net/npm/bootstrap@5.2.2/dist/js/bootstrap.min.js"
integrity="sha384-
IDwe1+LCz02ROU9k972gdyvl+AESN10+x7tBKgc9I5HFtuNz0wWnPclzo6p9vxnk"
crossorigin="anonymous"></script>
<script src="https://cdn.jsdelivr.net/npm/bootstrap@5.2.2/dist/js/bootstrap.bundle.min.js"
integrity="sha384-OERcA2EqjJCMA+/3y+gxIOqMEjwtxJY7qPCqsdltbNJuaOe923+mo//
f6V8Qbsw3" crossorigin="anonymous"></script>
```

```
  </body>
</html>
```

Result.html

```
<!doctype html>
<html lang="en">
<head>
  <link rel="stylesheet" type="text/css" href="{{url_for('static',filename='css/style.css')}}">
  <meta charset="utf-8">
  <meta name="viewport" content="width=device-width, initial-scale=1">
  <title>Result</title>
<link href="https://cdn.jsdelivr.net/npm/bootstrap@5.2.2/dist/css/bootstrap.min.css"
rel="stylesheet" integrity="sha384-
Zenh87qX5JnK2Jl0vWa8Ck2rdkQ2Bzep5IDxbcnCeuOxjzrPF/et3URy9Bv1WTRi"
crossorigin="anonymous">
</head>

       <body class="bg-co">
        <center>
        <div style="color: green;margin-top: 180px;">
           <img src="/static/trust.png" style="height: 100px; width: 100px; opacity: 1; " class="img
img-responsive img-circle  mx-auto d-block" />

            <h4> <b>TRUSTED SITE</b><br><b>Entered Site or URL is Not a phishing. So don't
worry about this site.<b></h4>
          <a href="/predicturl"><input type="button"  style="background-color: black; font-weight:
bold; color: white;" class="btn "  value="Predict Another URL" >

          </a>
          </div>
      </center>

         <script  src="https://cdn.jsdelivr.net/npm/@popperjs/core@2.11.6/dist/umd/popper.min.js"
integrity="sha384-oBqDVmMz9ATKxIep9tiCxS/Z9fNfEXiDAYTujMAeBAsjFuCZSmKbSSU
nQlmh/jp3" crossorigin="anonymous"></script>
                 <script    src="https://cdn.jsdelivr.net/npm/bootstrap@5.2.2/dist/js/bootstrap.min.js"
integrity="sha384-
IDwe1+LCz02ROU9k972gdyvl+AESN10+x7tBKgc9I5HFtuNz0wWnPclzo6p9vxnk"
crossorigin="anonymous"></script>
          <script  src="https://cdn.jsdelivr.net/npm/bootstrap@5.2.2/dist/js/bootstrap.bundle.min.js"
integrity="sha384-OERcA2EqjJCMA+/3y+gxIOqMEjwtxJY7qPCqsdltbNJuaOe923+mo//
f6V8Qbsw3" crossorigin="anonymous"></script>
  </body>
</html>
```

GitHub Link : https://github.com/IBM-EPBL/IBM-Project-42143-1660652065.git
Project Demo Link : https://youtu.be/qZwuTboPep4