WEB PHISHING DETECTION
A PROJECT REPORT

Submitted by

Team ID: PNT2022TMID27012
Team Leader: PITCHMA PRIYA K (310819104706)
Team member 2 : GIRISHA MV (310819104713)
Team member 3 : SURIYA LEKSHMI RM (310819104086)
Team member 4 : SHRUTHI R (310819104081)

DEPARTMENT OF COMPUTER SCIENCE
AND ENGINEERING
JEPPIAAR ENGINEERING COLLEGE
ANNA UNIVERSITY CHENNAI 600025

TABLE OF CONTENTS

ACKNOWLEDGEMENT

The Nalaiya Theren opportunity I had with IBM in the Data Analytics domain was a great chance for learning and professional development. Therefore, I consider myself as a very lucky individual for being a part of it. I am also grateful to all the professionals who led me though this internship period.
Bearing in mind previous I am using this opportunity to express my deepest gratitude and special thanks to My Principal who helped me in spite of being extraordinarily busy with his duties, allowing me to carry out my internship at the esteemed organization. I express my deepest thanks to my industry mentor Mr. Shanawaz Anwar for taking part in useful decision & giving necessary advices and guidance and arranged all facilities to make internship easier. I choose this moment to acknowledge his contribution gratefully. It is my radiant sentiment to place on record my best regards, deepest sense of gratitude to our College faculty mentor, VIDHYA Assistant Professor, Department of CSE for his careful and precious guidance which were extremely valuable for my study both theoretically and practically.
I perceive as this opportunity as a big milestone in my career development. I will strive to use gained skills and knowledge in the best possible way, and I will continue to work on their improvement, in order to attain desired career objectives.
Sincerely,
Team Leader : PITCHMA PRIYA K
Team member 2: GIRISHA MV
Team member : SURIYA LEKSHMI R.M
Team member 4: SHRUTHI R

1.INTRODUCTION

1.1PROJECT OVERVIEW

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.

1.2 PURPOSE

The main aim of this project is to prevent web phishing attacks and protecting the user's sensitive information from getting leaked online. Inorder to detect and predict e-banking phishing websites we

- Propose a flexible system using classification algorithms
- Implement techniques to extract the phishing datasets criteria to classify legitimacy
- Analyze the url
- Examine the domain identity
- Analyze the security and encryption criteria in the final phishing detection rate

- Use a data mining algorithm to detect whether the e-banking website is a phishing website or not

- Type of transmission which the customer prefers like Automatic or Manual

2.LITERATURE SURVEY

2.1 EXISTING PROBLEM

- Web phishing steals private information, such as usernames, passwords, and credit card details, by way of impersonating a legitimate entity.

- It will lead to information disclosure and property damage.

- Large organizations may get trapped in different kinds of scams.

2.2 References

- LongfeiWu etal..., "Effective Defense Schemes for Phishing Attacks on Mobile Computing Platforms, " IEEE 2016, pp.6678-6691.

- Surbhi Gupta etal., "A Literature Survey on Social Engineering Attacks: Phishing Attacks," in International Conference on Computing, Communication and Automation(ICCCA2016),2016, pp. 537-540

- Guardian Analytics, "A Practical Guide to Anomaly Detection Implications of meeting new FFIEC minimum expectations for layered security". [Accessed : 08 Jan 2015]

- SANS Institute, "Phishing : An Analysis of a Growing Problem", 2007. 1417[Accessed : 23 May 2017]

- J. Phys.: Conf. Ser. "A literature survey on Retraction : Phishing website detection using machine learning and deep learning techniques" 1916 (2021) 012407.

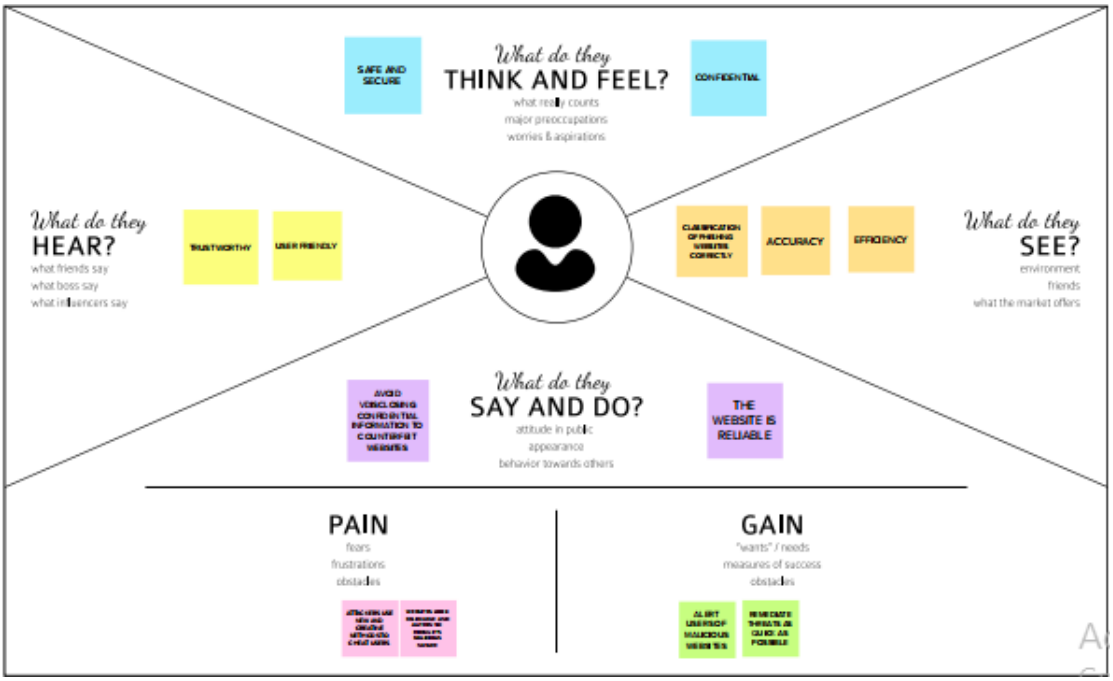
- Phishing Website Detection Based on Deep Convolutional Neural Network and Random Forest Ensemble Learning" ,This research was funded by the National Key R & D Program of China Grant Numbers 2017YFB0802800 and Beijing Natural Science Foundation (4202002)

2.3 PROBLEM STATEMENT DEFINITION

Phishing is a common attack on credulous people by making them to disclose their unique information using counterfeit websites. The objective of phishing website URLs is to purloin the personal information like user name, passwords and online banking transactions

3. IDEATION & PROPOSED SOLUTION

3.1 EMPATHY MAP CANVAS



3.2 IDEATION & BRAINSTORMING

1

Define your problem statement

What problem are you trying to solve? Frame your problem as a How Might We statement. This will be the focus of your brainstorm.

5 minutes

PROBLEM

WEB PHISHING DETECTION

type of social engineering attack often used to steal user datas..

Key rules of brainstorming

To run an smooth and productive session

- Stay in topic.
- Encourage wild ideas.
- Defer judgment.
- Listen to others.
- Go for volume.
- If possible, be visual.

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!

PITCHMA PRIYA K

- Fear of using websites
- Scam
- Usage of more Passwords
- Need to verify everything
- Install security softwares
- Obstades
- Avoid calls that ask confidential datas
- Rewards and Coupon offers
- Cost efficient

GIRISHA M V

- Open source
- Registration
- Login/OTP
- Two Step Verification
- verifying Captchas
- Online transactions
- UPI IDs
- Lucky Draw messages
- Pick strong Passwords

SHRUTHI R

- Avoid getting cheated by increasing amount
- Email phishing
- Unnecessary ADS in between
- Fake Website running automatically
- Don't click unknown linkd
- Messages by announcing winners

SURIYALEKSHMI RM

- Urgent Deadlines
- Usefriendly
- Never open Unexpected Attachments
- Use Anti spam filters
- Faster work completion
- Alert users of malicious websites

3

Group ideas

Take turns sharing your ideas while clustering similar or related notes as you go. Once all sticky notes have been grouped, 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 mind.

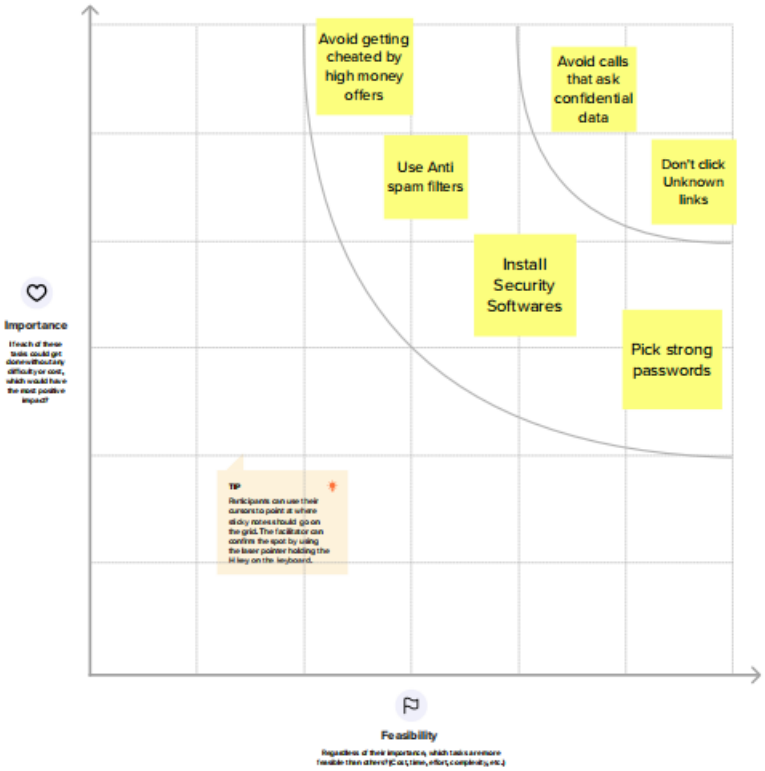
- Install Security Softwares
- Pick Strong Passwords
- Don't click Unknown links
- Use Anti-Spam Filters
- Avoid getting cheated by high money offers
- Avoid calls that ask Confidential datas

4

Prioritize

Your team should all be on the same page about what's important moving forward. Place your ideas on this grid to determine which ideas are important and which are feasible.

20 minutes



After you collaborate

You can export the mural as an image or pdf to share with members of your company who might find it helpful.

Quick add-ons

Share the mural
Share a view link to the mural with stakeholders to keep them in the loop about the outcomes of the session.

Export the mural
Export a copy of the mural as a PNG or PDF to attach to emails, include in slides, or save in your drive.

Keep moving forward

Strategy blueprint
Define the components of a new idea or strategy.
Open the template →

Customer experience journey map
Understand customer needs, motivations, and obstacles for an experience.
Open the template →

Strengths, weaknesses, opportunities & threats
Identify strengths, weaknesses, opportunities, and threats (SWOT) to develop a plan.
Open the template →

Share template feedback

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 Accout Information etc. There are a number of users who purchase products online and make payments through e-banking. There are e-banking websites that ask users to provide sensitive data such as username, password & credit card details etc. This type of e-banking website is known as a phising website. Web phising is one of many security threats to web services on the internet.
2.	Idea / Solution description	To use anti-phishing protection and anti-spam software to protect yourself. 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. 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. Regularly change the passwords to online account which prevents many attacks.Finally never share your personal details
3.	Novelty / Uniqueness	Machine learning technology consists of many algorithms which requires past data to make a decision or prediction of future data. Using this technique, algorithm will analyze various backlisted and legitimate URL's and their features to accurately detect the phishing websites including zero-hour phishing websites.

4.	Social Impact / Customer Satisfaction	Phishing website has a list of effects on a business, including loss of money, loss of intellectual property, damage of reputation, and disruption of operational activities. **Example:** Facebook and Google between 2013 and 2015 facebook and google were tricked out of $100 million due to an extended phishing campain.

3.4 PROBLEM SOLUTION FIT

4.REQUIREMENT ANALYSIS

4.1 FUNCTIONAL REQUIREMENT

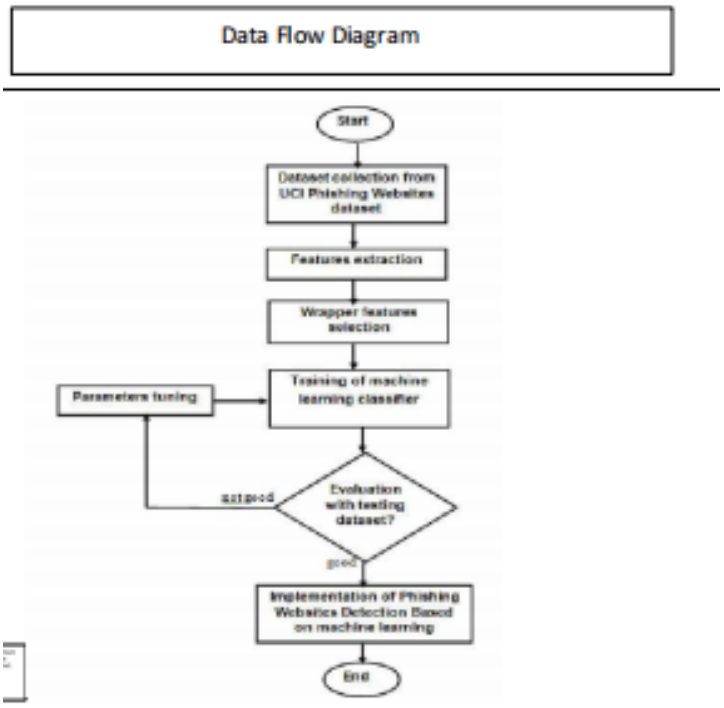
FR No.	Functional Requirement (Epic)	Sub Requirement (Story / Sub-Task)
FR-1	User Registration	Registration through Form.
FR-2	User Confirmation	Confirmation via Email.
FR-3	User Authentication	Authentication via Password.
FR-4	User Input	User input an URL to check it is legal or phishing site.
FR-5	Website Comparison	Model comparing the entered URL with the help of Blacklist and Whitelist.
FR-6	Feature extraction	After comparing, if none found on comparison the it extracts feature using heuristic and visual similarity approach.
FR-7	Prediction	Model Predicts the URL using Machine Learning algorithm such as Logistic Regression, KNN.
FR-8	Classifier	Model sends output to classifier and it produce final result.
FR-9	Announcement	Model the displays whether the website is a legal or phishing site.
FR-10	Events	Model needs the capability of retrieving and displaying accurate result for a website.

4.2 NON-FUNCTIONAL REQUIREMENTS

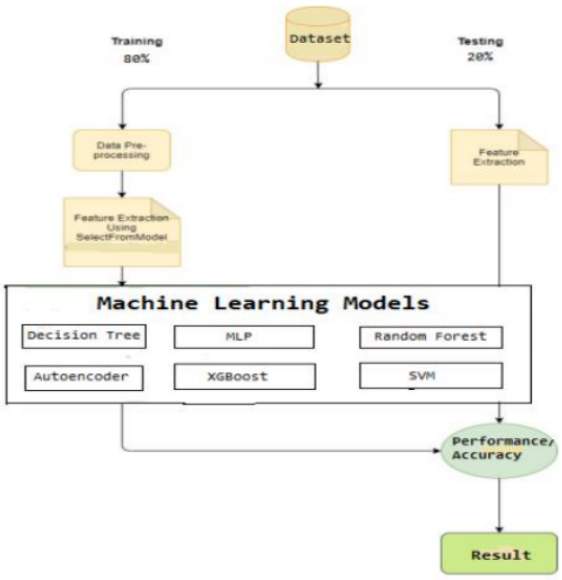
FR No.	Non-Functional Requirement	Description
NFR-1	**Usability**	A set of specifications that describe the system's operation capabilities and constraints and attempt to improve its functionality.
NFR-2	**Security**	Assuring all data inside the system or its part will be protected against malware attacks 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 efficiency according to rising demands.

5.PROJECT DESIGN

5.1 DATA FLOW DIAGRAMS



5.2 SOLUTION & TECHNICAL ARCHITECTURE



Solution Architecture for Website Phishing Detection

5.3 USER STORIES

User Type	Functional Requirement (Epic)	User Story Number	User Story / Task	Acceptance criteria	Priority	Release
Customer (Mobile user)	Registration	USN-1	As a user, I can register for the application by entering my email, password, and confirming my password.	I can access my account / dashboard	High	Sprint-1
		USN-2	As a user, I will receive confirmation email once I have registered for the application	I can receive confirmation email & click confirm	High	Sprint-1
		USN-3	As a user, I can register for the application through Facebook	I can register & access the dashboard with Facebook Login	Low	Sprint-2
		USN-4	As a user, I can register for the application through Gmail		Medium	Sprint-1
	Login	USN-5	As a user, I can log into the application by entering email & password		High	Sprint-1
	Dashboard					
Customer (Web user)						
Customer Care Executive						
Administrator						

6.PROJECT PLANNING & SCHEDULING

6.1 SPRINT PLANNING & ESTIMATION

Project Planning Phase

Project Planning Template (Product Backlog, Sprint Planning, Stories, Story points)

Date	18 October 2022
Team ID	PNT2022TMID27012
Project Name	Web Phishing Detection
Maximum Marks	8 Marks

Product Backlog, Sprint Schedule, and Estimation (4 Marks)

Use the below template to create product backlog and sprint schedule

Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority	Team Members
Sprint-1	Collecting Dataset	USN-1	Downloading the required dataset	1	Low	SURIYA LEKSHMI R M
Sprint-1	Pre-process data	USN-2	Import required libraries	1	Low	GIRISHA M V
Sprint-1		USN-3	Read and Splitting of data sets	2	Low	SHRUTHI R
Sprint-1		USN-4	Handling of Null Values, Split Data	2	Low	PITCHMA PRIYA K
Sprint-2	Model building	USN-1	Working with Logistic Regression Model with Split Data of dependent and independent variables	3	Medium	PITCHMA PRIYA K SURIYA LEKSHMI R M
Sprint-3	Application Building	USN-1	Build Flask-1, Flask-2	3	Medium	GIRISHA M V SHRUTHI R
		USN-2	Build HTML page	3	Medium	PITCHMA PRIYA K, GIRISHA M V
		USN-3	Execute and Testing	4	High	PITCHMA PRIYA K, SURIYA LEKSHMI R M,SHRUTHI R
Sprint-4	Training the model	USN-1	Train Machine Learning Model	5	High	PITCHMA PRIYA K,

Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority	Team Members
						GIRISHA M V, SURIYA LEKSHMI R M,SHRUTHI R
		USN-2	Integrate Flask with scored End Point	5	High	SURIYA LEKSHMI R M,SHRUTHI R,PITCHMA PRIYA K, GIRISHA M V

6.2 SPRINT DELIVERY SCHEDULE

Project Tracker, Velocity & Burndown Chart: (4 Marks)

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	11 Nov 2022
Sprint-2	20	6 Days	31 Oct 2022	05 Nov 2022	20	11 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	19 Nov 2022

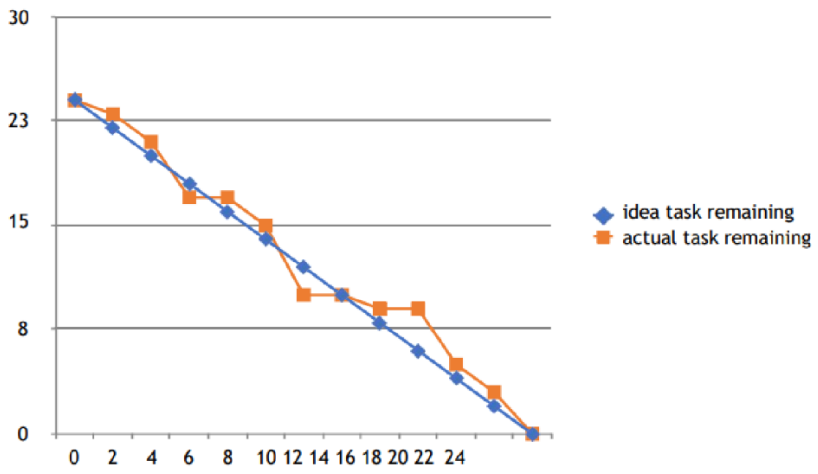
Velocity:

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

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

Burndown Chart:

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



7.CODING & SOLUTION

7.1 FEATURE 1:

.html

```
<!DOCTYPE html>
<html>

<head>
        <title>Web Phishing Detection</title>
        <link rel="stylesheet" type="text/css" href="./style1.css">
</head>

<body>
<div class="URL-form">
```

```
<center>
<h2>Detection of Phishing URL</h2>
<form action="next.html">

<input type="text" id="url" name="url" placeholder="Enter URL">
<button>Submit</button>
<p>If you want to see the result of mulltiple URLs you can upload a file containing URLs.</p>
<button>Upload File</button>

</form>
</center>
</div>
</body>

</html>
```

7.2 FEATURE 2

Main.py

```
import os
from os.path import join, dirname
from dotenv import load_dotenv
from functools import wraps
from http.client import HTTPException
import numpy as np
from flask import Flask, request, render_template,session, url_for,redirect,flash
import pickle
import inputScript
import pymongo
from passlib.hash import  pbkdf2_sha256
import json
import inputScript

app = Flask(_name_,template_folder='../Flask')
model = pickle.load(open('../Flask/Phishing_Website.pkl','rb'))


dotenv_path = join(dirname(_file_), '.env')
load_dotenv(dotenv_path)
MONGODB_URL = os.environ.get("MONGODB_URL")
SECRET_KEY = os.environ.get("SECRET_KEY")


mongoDB=pymongo.MongoClient(MONGODB_URL)
db=mongoDB['Web_Phishing_Detection']
account=db.account
app.secret_key= SECRET_KEY
carouselDataFile = open('./static/json/carouselData.json')
carouselData = json.load(carouselDataFile)
aboutDataFile = open('./static/json/aboutData.json')
aboutData = json.load(aboutDataFile)

def login_required(f):
    @wraps(f)
    def wrap(*args, **kwargs):
        if('logged_in' in session):
            return f(*args, **kwargs)
        else:
```

```
        return redirect('/')
    return wrap


def start_session(userInfo):
    if userInfo:
        userInfo['_id']=str(userInfo['_id'])
    else:
        raise HTTPException(status_code=404, detail=f"Unable to retrieve record")
    del userInfo['password']
    session['logged_in']=True
    session['user']=userInfo
    return redirect(url_for('index'))


@app.route('/login/',methods=['POST'])
def login():
    if request.method=="POST":
        email=request.form.get("email")
        password=request.form.get("password")
        if(account.find_one({"email":email})):
            user=account.find_one({"email":email})
            if(user and pbkdf2_sha256.verify(password,user['password'])):
                return start_session(user)
            else:
                flash("Password is incorrect","loginError")
                return redirect(url_for('index',loginError=True))
        flash("Sorry, user with this email id does not exist","loginError")
        return redirect(url_for('index',loginError=True))


@app.route('/signup/',methods=['POST'])
def signup():
    if request.method=="POST":
        userInfo={
        "fullName":request.form.get('fullName'),
        "email":request.form.get('email'),
        "phoneNumber":request.form.get('phoneNumber'),
        "password":request.form.get('password'),
        }
        userInfo['password']=pbkdf2_sha256.encrypt(userInfo['password'])
        if(account.find_one({"email":userInfo['email']})):
            flash("Sorry,user with this email already exist","signupError")
            return redirect(url_for('index',signupError=True))
        if(account.insert_one(userInfo)):
            return start_session(userInfo)
    flash("Signup failed","signupError")
    return redirect(url_for('index',signupError=True))


@app.route('/logout/',methods=["GET"])
def logout():
    if request.method=="GET":
        session.clear()
    return redirect(url_for('index'))
@app.route('/')
def index():
```

```
    if(session and '_flashes' in dict(session)):
        loginError=request.args.get('loginError')
        signupError=request.args.get('signupError')
        if(loginError):
            return
render_template('./index.html',loginError=loginError,carousel_content=carouselData['carousel
_content'])
        if(signupError):
            return
render_template('./index.html',signupError=signupError,carousel_content=carouselData['caro
usel_content'])
    if(session and '_flashes' not in dict(session)):
        print(dict(session))
        if(session['logged_in']==True):
            return
render_template('./index.html',userInfo=session['user'],carousel_content=carouselData['carou
sel_content'])
        else:
            return
render_template('./index.html',carousel_content=carouselData['carousel_content'])
    else:
        return render_template('./index.html',carousel_content=carouselData['carousel_content'])



@app.route('/predict/', methods=['GET','POST'])
@login_required
def predict():
    if request.method == 'POST':
        url = request.form['url']
        checkprediction = inputScript.main(url)
        print(url)
        print(checkprediction)
        prediction = model.predict(checkprediction)
        print(prediction)
        output=prediction[0]
        if(output==1):
            pred="Safe,legitimate link"

        else:
            pred="Malicious URL alert!"
        if(session and session['logged_in']):
            if(session['logged_in']==True):
                return
render_template('./templates/prediction-result.html',userInfo=session['user'],pred=pred)
            # else:
            #     return render_template('./templates/prediction-result.html',pred=pred)
        # else:
        #     return render_template('./templates/prediction-result.html',pred=pred)
    elif request.method == 'GET':
        return render_template('./templates/predict-form.html',userInfo=session['user'])

@app.route('/about/')
def about():
    if(session and session['logged_in']):
        if(session['logged_in']==True):
```

```
        return
render_template('./templates/about.html',userInfo=session['user'],aboutContents=aboutData['
aboutContents'])
    else:
        return
render_template('./templates/about.html',aboutContents=aboutData['aboutContents'])
  else:
    return
render_template('./templates/about.html',aboutContents=aboutData['aboutContents'])



@app.route('/contact/')
def contact():
    if(session and session['logged_in']):
        if(session['logged_in']==True):
            return render_template('./templates/contact.html',userInfo=session['user'])
        else:
            return render_template('./templates/contact.html')
    else:
        return render_template('./templates/contact.html')
if _name_ == '_main_':
  app.run(host='127.0.0.1', debug=True)
```

8.TESTING

8.1 TEST CASES

Acceptance Testing
UAT Execution & Report Submission

Date	03 November 2022
Team ID	PNT2022TMID27012
Project Name	Web Phishing Detection
Maximum Marks	4 Marks

1. Purpose of Document

The purpose of 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).

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	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

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	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

8	TC_001	Functional	Login Page	Verify user when Update button is clicked		1.Enter username 2.Enter password 3.Click Login button	Username: priem Password: priem	Login successful	Login successful	Pass		N		PITCHMA PRIYA K
7	TC_002	Functional	Employee Attrition Details	Enter the details of the employee		1.Enter valid Education 2.Job Satisfaction 3.Monthly Salary 4.Age 5.Attrition 6.Business Travel 7.Department 8.Distance From Home 9.Education 10.Education Field 11.Employee Number 12.Gender 13.Hourly Rate 14.Job Involvement 15.Job Level 16.Marital Status 17.Monthly Income 18.Number Companies Worked 19.Over 18 20.Over Time 21.Percent Salary Hike 22.Performance Rating 23.Relationship Satisfaction 24.Standard Hours 25.Stock Option Level 26.Total Working Years 27.Training Times Last Year 28.Years At Company 29.Years Since Last Promotion 30.Years With Current Manager	1.Employee ID: 1 2.Education Field: Life Sciences 3.Job Satisfaction: 4 4.Monthly Salary: 2 5.Age: 41 6.Attrition: No 7.Business Travel: Travel_Rarely 8.Department: Sales 9.Distance From Home: 1 10.Education: 2 11.Education Field: Life Sciences 12.Employee Number: 1 13.Gender: F 14.Hourly Rate: 94 15.Job Involvement: 3 16.Job Level: 2 17.Marital Status: Married 18.Monthly Income: 5993 19.Number Companies Worked: 8 20.Over 18: Y 21.Over Time: Yes 22.Percent Salary Hike: 11 23.Performance Rating: 3 24.Relationship Satisfaction: 1 25.Standard Hours: 80 26.Stock Option Level: 0 27.Total Working Years: 8 28.Training Times Last Year: 0 29.Years At Company: 6 30.Years Since Last Promotion: 1 31.Years With Current Manager: 5	The details of the employees are submitted successfully	The details of the employees are submitted successfully	Pass		N		GIRISHA MV
8	TC_003	Functional	Database	Allow the user to store the data		Create the table	HR Analytics Case Study Dataset from Kaggle	The database has been created successfully	The database has been created	Pass		N		SURIYA LAKSHMI R
9	TC_004	Functional	Database	Allow the user to store the data		Update the table	HR Analytics Case Study Dataset from Kaggle	The database has been updated successfully	The database has been updated	Pass		N		SHRUTHI R
10	TC_005	Functional	Dashboard	Updation of dashboard		Update the charts and graphs in the dashboard	Database	The dashboard has been updated successfully	The dashboard has been updated successfully	Pass		N		PITCHMA PRIYA K
11	TC_006	Functional	Visualize the data	Attrition status (yes/no)		Create a pie chart based on attrition by attrition	Attrition	The data is visualized	The data is visualized	Pass		N		GIRISHA MV
12	TC_007	Functional	Visualize the data	Attrition in Business travel		Create a bar chart based on attrition by business travel	Attrition, Business travel	The data is visualized	The data is visualized	Pass		N		SURUTHI R
13	TC_008	Functional	Visualize the data	Attrition based on Department, Job role, Education, Marital status		1.Create a bar chart based on attrition by department 2.Create a column chart based on attrition by job role 3.Create a stacked chart based on attrition by education 4.Create a pie chart based on marital status	Attrition, Department, Job role, Education, Marital status	The data is visualized	The data is visualized	Pass		N		SURIYA LEKSHMI R M
14	TC_009	Functional	Visualize the data	Attrition based on Salary hike percent		Create a column chart based on attrition by salary hike percent	Attrition, Percent salary hike	The data is visualized	The data is visualized	Pass		N		PITCHMA PRIYA K
15	TC_010	Functional	Visualize the data	Attrition based on No.of companies worked		Create a pie chart based on attrition by no.of companies worked	Attrition, No.of companies worked	The data is visualized	The data is visualized	Pass		N		GIRISHA MV
16	TC_011	Functional	Visualize the data	Attrition based on Monthly income groups		Create a bar chart based on attrition by monthly income groups	Attrition, Monthly income	The data is visualized	The data is visualized	Pass		N		SHRUTHI RM
17	TC_012	Functional	Visualize the data	Attrition based on Employee working groups		Create a bar chart based on attrition by employee working	Attrition, Total working years	The data is visualized	The data is visualized	Pass		N		SURIYA LEKSHMI R M
18	TC_013	Functional	Visualize the data	Attrition based on Percent salary hike, No.of companies worked, Monthly income, Total working		Create a dashboard based on previous charts	Attrition, Percent salary hike, No.of companies worked, Monthly income, Total working years	The data is visualized	The data is visualized	Pass		N		PITCHMA PRIYA K
19	TC_014	Functional	Logistic Regression	Uploading the dataset and performing univariate analysis		1.Upload the dataset 2.Merge the dataset 3.Univariate Analysis	HR Analytics Case Study Dataset from Kaggle	1.The datasets are uploaded successfully 2.All the datasets are merged 3.Univariate Analysis is performed	1.The datasets are uploaded successfully 2.All the datasets are merged 3.Univariate Analysis is performed	Pass		N		SURIYA LEKSHMI R M
20	TC_015	Functional	Logistic Regression	Understanding the dataset and performing statistics		1.Statistical description 2.Replacing missing values	HR Analytics Case Study Dataset from Kaggle	1.Statistical description is performed 2.The missing values are replaced	1.Statistical description is performed 2.The missing values are replaced	Pass		N		GIRISHA MV
21	TC_016	Functional	Logistic Regression	Identifying the outliers and splitting the data		1.Identifying the outliers and replacing them 2.Splitting the independent and dependent values	HR Analytics Case Study Dataset from Kaggle	1.The outliers are identified and replaced 2.The data has been splitted into independent and	1.The outliers are identified and replaced 2.The data has been splitted into independent and	Pass		N		SURUTHI R

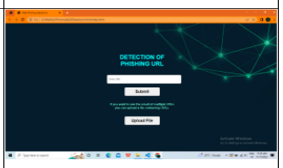
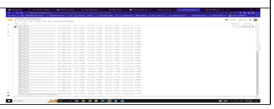
8.2 USER ACCEPTANCE TESTING

Project Development Phase
Model Performance Test

Date	10 November 2022
Team ID	PNT2022TMID27012
Project Name	Web Phishing Detection
Maximum Marks	10 Marks

Model Performance Testing:

Project team shall fill the following information in model performance testing template.

S.No.	Parameter	Values	Screenshot
1.	Model Summary	This model focuses on applying machine learning algorithms to detect phishing websites. It is based on classification algorithms. A data mining algorithm is used to detect whether the website is a phishing website or not.	
2.	Accuracy	Training Accuracy-0.9861 Validation Accuracy-0.9134	

9.RESULTS

9.1PERFORMANCE METRICS

We have collected unstructured data of URLs from Phishtank website, Kaggle website and Alexa website, etc. In pre-processing, feature generation is done where in features are generated from unstructured data. These features are length of an URL, URL has HTTP, URL has suspicious character, prefix/suffix, number of dots, number of slashes, URL has phishing term, length of subdomain, URL contains IP address. After, an organized dataset is made in which each detail incorporates the paired (0,1) which is then passed to the various classifiers. Next, we train the three unique classifiers and analyse their presentation based on exactness two classifiers utilized are Decision Tree and Random Forest algorithm. At that point, the classifier identifies the given URL dependent on the preparation information that is if the site is phishing it prompts the user that the website is phished and if genuine, it prompts the user that the website is legitimate. We look at the exactness of various classifiers and discovered Random Forest as the best classifiers which gives the most extreme precision.

10.ADVANTAGES & DISADVANTAGES

ADVANTAGES

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

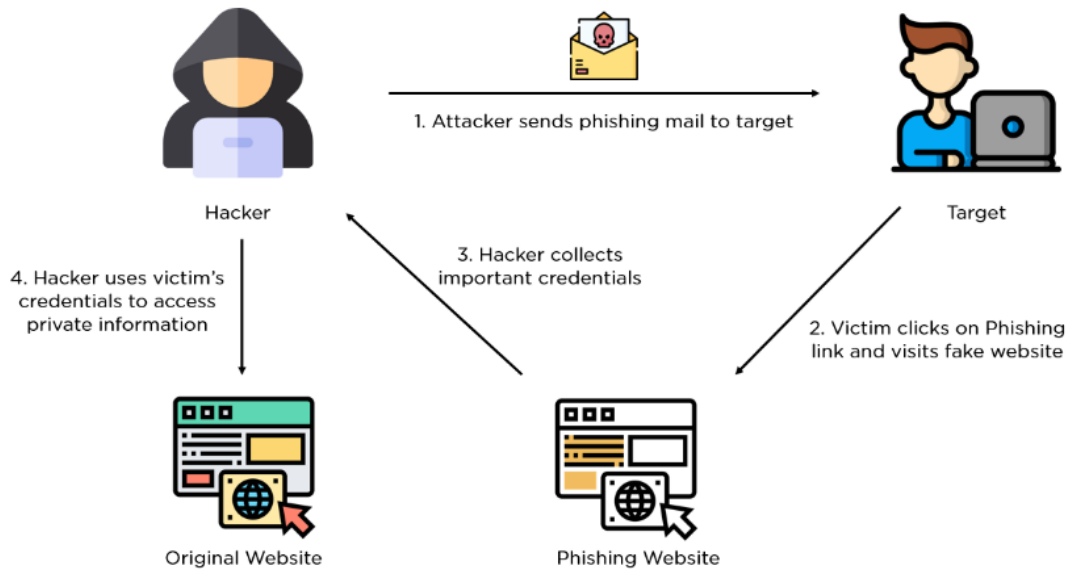
DISADVANTAGES

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. These effects work together to cause loss of company value, sometimes with irreparable repercussions.

11.CONCLUSION

The importance to safeguard online users from becoming victims of online fraud, divulging confidential information to an attacker among other effective uses of phishing as an attacker’s tool, phishing detection tools play a vital role in ensuring a secure online experience for users. Unfortunately, many of the existing phishing-detection tools, especially those that depend on an existing blacklist, suffer limitations such as low detection accuracy and high false alarm that is often caused by either a delay in blacklist update as a result of human verification process involved in classification or perhaps, it can be attributed to human error in classification which may lead to improper classification of the classes. These critical issues have drawn many

researchers to work on various approaches to improve detection accuracy of phishing attacks and to minimize false alarm rate. The inconsistent nature of attacks behaviors and continuously changing URL phish patterns require timely updating of the reference model. Therefore, it requires an effective technique to regulate retraining as to enable machine learning algorithm to actively adapt to the changes in phish patterns.



12.FUTURE SCOPE

Phishing website is an illegitimate website that is designed by dishonest people to mimic a real website. Those who are entering such a website may expose their sensitive information to the attacker who might use this information for financial and criminal activities. In this technological world, phishing websites are created using new techniques allows them to escape from most anti-phishing tool. So, the white list and blacklist based techniques are less effective when compared with the recent phishing trends. Advanced to that, there exist some tools using machine learning and deep learning approaches by examining webpage content in order to detect phishing websites. Along with the rapid growth of phishing technologies, it is needed to improve the effectiveness and efficiency of phishing website detection.

13.APPENDIX

GITHUB LINK

https://github.com/IBM-EPBL/IBM-Project-958-1658332500

YOUTUBE LINK

https://youtu.be/64x_YH4PZ7o