A Project Report ON PHISHING ATTACK DETECTION

Submitted to the Apex University, Jaipur

In Partial fulfillment of the requirement for the degree of

Bachelor of Computer Applications

Submitted by

SUNIL YADAV (23CA2591057)
ANIL YADAV (23CA2591005)
RUPESH SAIN (23CA2591047)
LOKESH KUMAR (23CA2591026)

Submitted To: YOGESH RAO
Department of Computer Science

Apex University

Sec-5, V.T. Road, Mansarovar, Jaipur

Admission Year: 2023-2026

(Deposition date: FEBURARY 2026)

DECLARATION

I, SUNIL YADAV hereby declare that the work presented in this project entitled **“PHISHING ATTACK DETECTION** in partial fulfillment of the requirements for the award of Bachelor of Computer Applications, submitted in the **Department of Computer Science** at Apex University, Jaipur, is an authentic record of my own research work under the supervision of YOESH SIR

I also declare that the work embodied in the present thesis is my original work/extension of the existing work and has not been copied from any Journal/thesis/book, and has not been submitted by me for any other Degree/Diploma.

NAME: SUNIL YADAV
Enrolment No. 23CA2591057
Date: 21 FEB 2026

CERTIFICATE

This is to certify that **SUNIL YADAV (23CA2591057)**
ANIL YADAV (23CA2591005) , RUPESH SAIN (23CA2591047),

LOKESH KUMAR (23CA2591026) have successfully completed the project work titled
“PHISHING ATTACK DETECTION”

submitted in partial fulfillment of the requirements for the award of the degree of
Bachelor of Computer Applications (BCA)
from **Apex University, Jaipur**.

The work embodied in this project report is original and has been carried out by the students under my supervision during the academic session 2023–26.

To the best of my knowledge, this work has not been submitted to any other university or institution for the award of any degree or diploma.

ACKNOWLEDGEMENT

I would like to express my sincere gratitude to my project guide **“YOGESH SIR”** for giving me the opportunity to work on this topic. It would never be possible for us to take this project to this level without his innovative ideas and his relentless support and encouragement.

Name of Students:-

1. SUNIL YADAV (23CA2591057)

2. ANIL YADAV (23CA2591005)

3. RUPESH SAIN (23CA2591047)

4. LOKESH KUMAR (23CA2591026)

ABSTRACT

This project focuses on detecting phishing attacks, which are one of the most common cyber threats. Phishing attacks involve fake websites or emails designed to steal sensitive information such as passwords and bank details.

The system analyzes URLs and identifies suspicious patterns such as unusual domain names, use of special characters, and insecure protocols. Based on these factors, it determines whether a website is safe or phishing.

This project helps in understanding cyber security threats and provides a basic solution to detect phishing attempts.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

□ CHAPTER 1

INTRODUCTION

The rapid growth of internet services has made online transactions and communication easier. However, it has also increased cyber crimes. Among various cyber attacks, phishing is one of the most dangerous and commonly used techniques.

Phishing attacks are designed to trick users into revealing confidential information by pretending to be a trusted source.

Phishing attack is one of the most common types of cybercrime in today's digital world. It is a technique used by attackers to trick users into revealing sensitive information such as passwords, bank details, and personal data by pretending to be a trusted source.

In a phishing attack, the attacker usually sends fake emails, messages, or links that appear to come from legitimate organizations like State Bank of India or Google. These messages create urgency or fear, forcing users to click on malicious links or provide confidential information.

Phishing attacks are dangerous because they target human psychology instead of technical systems. Even a well-secured system can be compromised if a user unknowingly shares their information.

With the rapid growth of internet usage, online banking, and digital services, phishing attacks are increasing day by day. Therefore, it is important to understand how phishing works and how to protect against it.

This project focuses on explaining phishing attacks, their types, working methods, impacts, and prevention techniques in detail.

1.1 CYBER SECURITY

Cyber security refers to the practice of protecting computers, networks, systems, and data from cyber attacks, unauthorized access, and damage. In today's digital world, where most activities like banking, communication, shopping, and education are done online, cyber security plays a very important role in ensuring safety and privacy.

Cyber security is designed to protect both individuals and organizations from threats such as hacking, phishing, malware, data theft, and identity fraud. It involves the use of technologies, processes, and policies to secure digital information.

For example, when you use services like Google or Amazon, cyber security measures ensure that your personal data, passwords, and transactions remain safe and confidential.

Goals of Cyber Security

The main goals of cyber security are:

- **Confidentiality:** Data should be accessed only by authorized users.
- **Integrity:** Data should not be altered or modified without permission.
- **Availability:** Data and systems should be available when needed.

These three principles are also known as the **CIA Triad**

Cyber Security and Phishing

Cyber security plays a key role in preventing phishing attacks. It helps in:

- Detecting fake emails and websites
- Protecting login credentials
- Alerting users about suspicious activities
- Implementing secure authentication methods

Types of Cyber Security

Cyber security can be divided into different types:

1. **Network Security** – Protects computer networks from intruders and attacks.
2. **Application Security** – Secures software and applications from threats.
3. **Information Security** – Protects data from unauthorized access.
4. **Cloud Security** – Secures data stored on cloud platforms.
5. **Endpoint Security** – Protects devices like computers and mobiles.

1.2 PHISHING ATTACK

A phishing attack is a type of cyber attack in which an attacker tries to steal sensitive information by pretending to be a trusted source. This information may include usernames, passwords, credit/debit card details, OTPs, or other personal data.

In phishing, attackers usually send fake emails, messages, or create fraudulent websites that look exactly like real ones. These messages often appear to come from trusted organizations such as State Bank of India, Paytm, or Google to gain the victim’s trust.

How Phishing Attack Works

The process of a phishing attack generally follows these steps:

1. **Impersonation** – The attacker pretends to be a legitimate organization or person.
2. **Creating Fake Message** – A fake email, SMS, or notification is created.
3. **Sending the Link** – Victim receives a message containing a malicious link.
4. **User Action** – User clicks the link and enters sensitive details.
5. **Data Theft** – The attacker collects and stores the information.
6. **Misuse of Data** – The stolen data is used for fraud or unauthorized access.

Common Features of Phishing Attacks

Phishing attacks usually have the following characteristics:

- Fake sender email or phone number
- Urgent or threatening message (e.g., “Account will be blocked”)
- Suspicious links or attachments
- Spelling and grammatical errors
- Request for confidential information

Example of Phishing Attack

A user receives an email saying:

“Your bank account has been suspended. Click the link below to verify your details.”

The link redirects to a fake website that looks like the original bank website. When the user enters login details, the attacker steals the information.

Why Phishing is Effective

Phishing is very effective because:

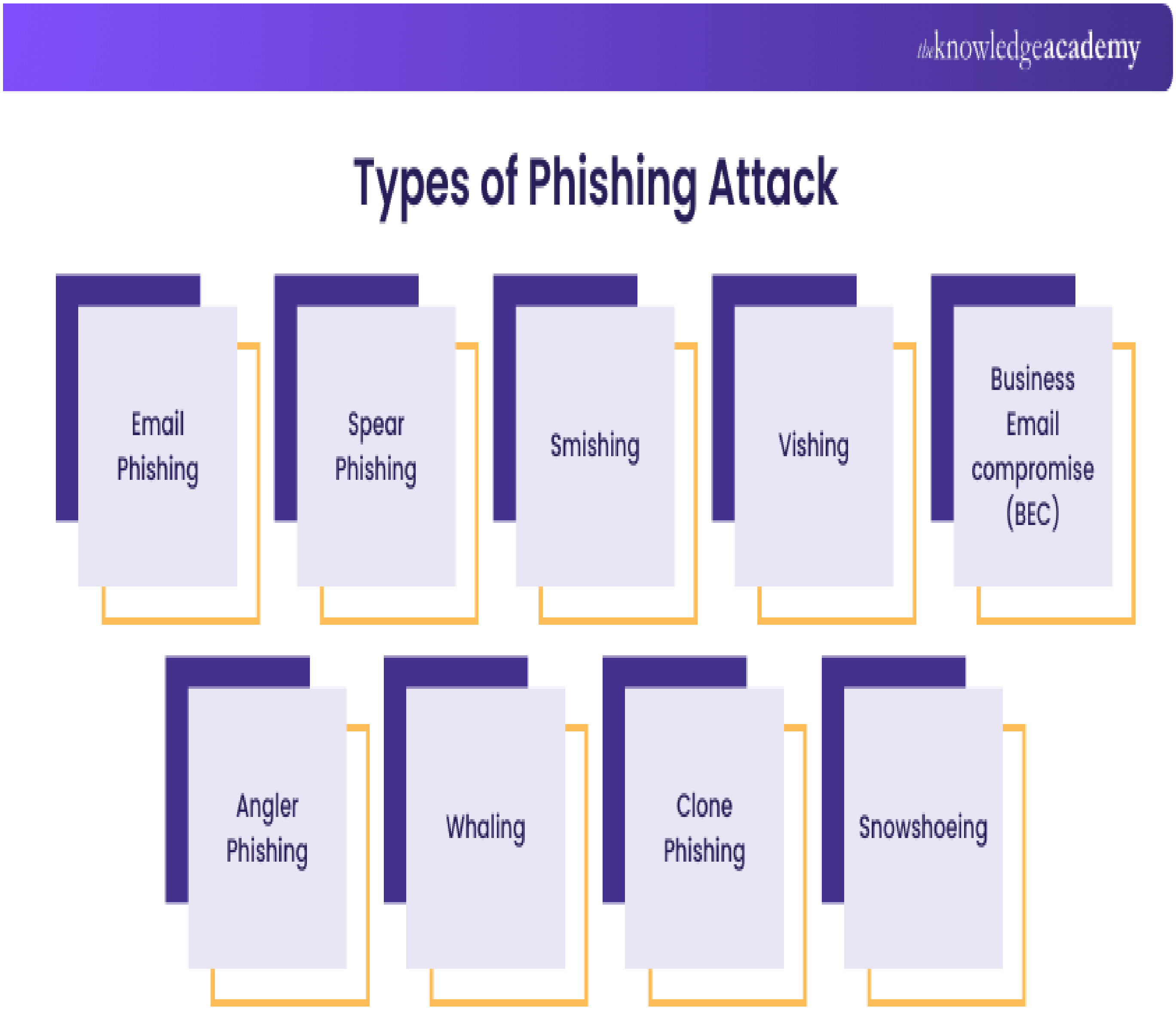
- It targets human emotions like fear and urgency
- Many users are not aware of such attacks
- Fake websites look very real
- It does not require advanced technical skills

Impact of Phishing

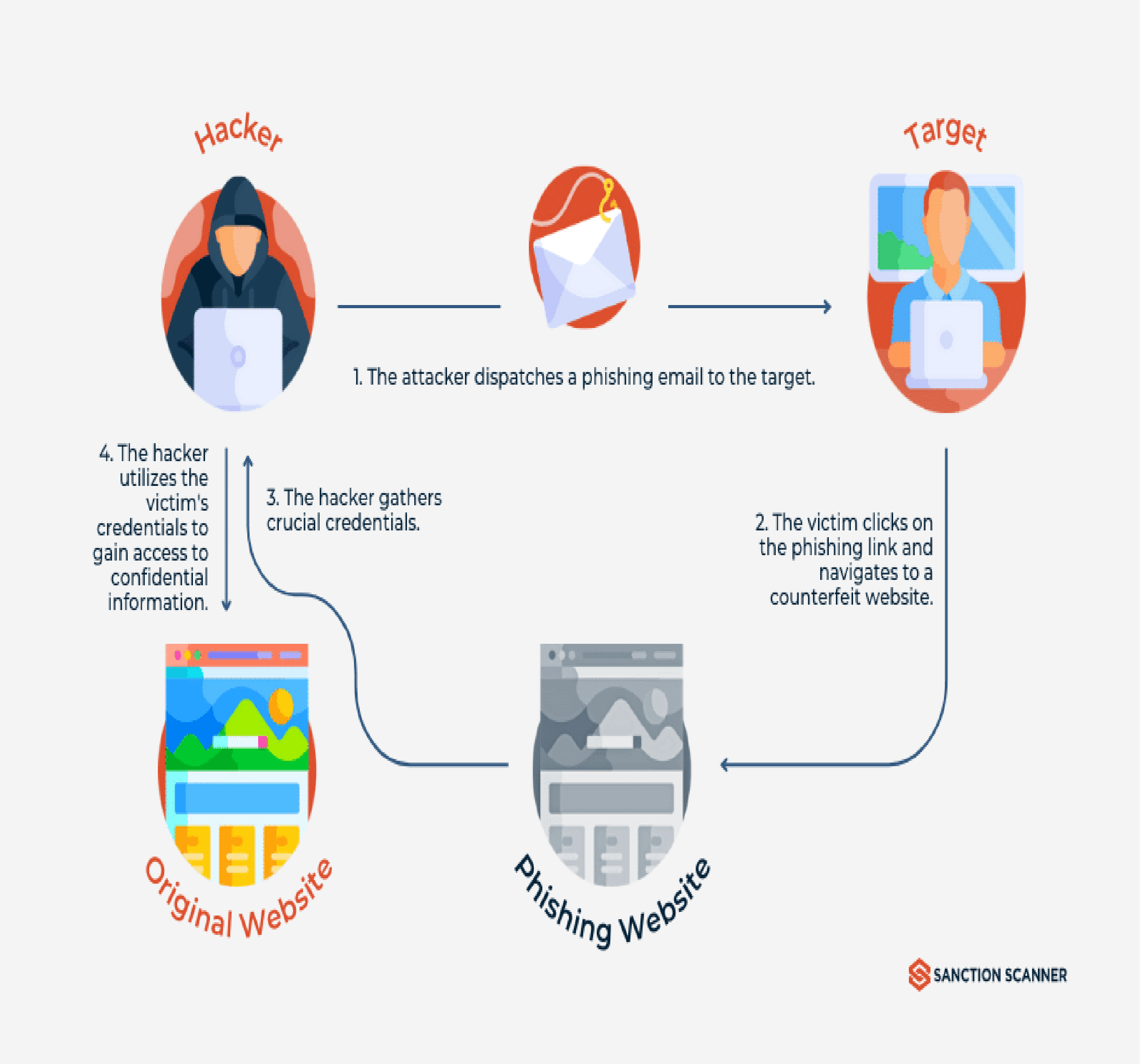
Phishing attacks can cause serious damage such as:

- Financial loss
- Identity theft
- Unauthorized account access
- Data breaches
- Loss of trust in digital platforms

1.3 Types of Phishing



Phishing attack process:


Hacker
Target
1. The attacker dispatches a phishing email to the target.
2. The victim clicks on the phishing link and navigates to a counterfeit website.
3. The hacker gathers crucial credentials.
4. The hacker utilizes the victim's credentials to gain access to confidential information.
Original Website
Phishing Website
SANCTION SCANNER

1.4 NEED OF DETECTION SYSTEM

With the rapid increase in phishing attacks, it has become very important to develop systems that can detect and prevent such attacks before they cause harm. A phishing detection system is a security mechanism designed to identify fake emails, malicious links, and fraudulent websites.

Phishing attacks are increasing day by day due to the growth of internet usage, online banking, and digital services. Attackers are becoming more advanced and their fake messages look very real.

For example, users may receive fake messages that appear to come from trusted organizations like State Bank of India or Google, making it difficult to identify whether the message is real or fake.

Main Reasons for Using Detection Systems:

1. Protection of Sensitive Data
2. Prevention of Financial Loss
3. Increasing Cyber Threats
4. Protection for Organizations
5. User Awareness and Safety

1.5 Objectives

The main objective of this project is to design and develop a system that can detect phishing websites and protect users from cyber fraud. The project focuses on analyzing URLs and identifying whether a website is safe or malicious.

Main Objectives:

1. To Detect Phishing Websites
2. To Improve User Security
3. To Analyze URL Features
4. To Classify Websites
5. To Develop a Simple and Efficient System

- **Table: Phishing Detection Criteria**

S. No.	Criteria	Description	Example
1	**URL Structure**	Checks if the website URL is suspicious, too long, or contains unusual characters	`http://secure-bank-login.xyz`
2	**Domain Age**	New or recently created domains are often used for phishing	Domain created few days ago
3	**HTTPS Security**	Verifies if the website uses secure HTTPS protocol	Fake site without HTTPS
4	**Website Design**	Checks if the website looks similar to trusted sites like State Bank of India	Duplicate login page
5	**Spelling & Grammar**	Identifies errors in email or website content	“Updte your acount now”
6	**Email Sender Address**	Verifies if sender email is genuine or fake	support@bank-secure.xyz
7	**Urgency in Message**	Detects threatening or urgent language	“Your account will be blocked immediately”
8	**Use of Shortened URLs**	Short links may hide actual malicious URLs	bit.ly/xyz123
9	**Pop-up Windows**	Fake pop-ups asking for login details	Login popup on unknown site
10	**Request for Sensitive Data**	Asking for OTP, password, or card details	“Enter OTP to verify”
11	**IP Address in URL**	Use of IP instead of domain name	http://192.168.1.1/login

□ CHAPTER 2

LITERATURE REVIEW

Literature review is an important part of any project. It includes the study and analysis of previous research, articles, and techniques related to phishing attacks and their detection methods. This helps in understanding what has already been done and what improvements are needed.

Overview of Existing Research

Many researchers have worked on phishing detection and prevention techniques. Their studies mainly focus on:

- Identifying phishing websites
- Detecting fake emails
- Using machine learning algorithms
- Improving user awareness

These studies provide a strong base for developing better phishing detection systems.

Research on Email-Based Detection

Some research focuses on detecting phishing emails by analyzing:

- Sender email address
- Content of the email
- Suspicious links and attachments

These methods help in filtering fake emails before they reach users.

2.1 EXISTING SYSTEMS

Existing systems refer to the tools, techniques, and technologies that are currently used to detect and prevent phishing attacks. These systems are designed to identify suspicious emails, fake websites, and malicious links before they can harm users.

1. Email Filtering Systems

These systems are used to detect phishing emails before they reach the user’s inbox.

- Analyze sender address, subject, and content
- Detect spam and malicious links
- Automatically block suspicious emails

Popular email services like Gmail use advanced filtering techniques to protect users.

2. Blacklist-Based Systems

These systems maintain a list of known phishing websites.

- If a user tries to open a blacklisted website, access is blocked
- Easy to implement
- Fast detection of known threats

 Limitation: Cannot detect new or unknown phishing websites.

3. Heuristic-Based Systems

These systems detect phishing based on rules and patterns.

- Check URL length, special characters, and domain name
- Identify suspicious behavior
- Use predefined rules

□ Limitation: May give false results sometimes.

---

4. Machine Learning-Based Systems

These systems use algorithms to detect phishing attacks.

- Learn from previous data
- Identify patterns in phishing attacks
- Improve detection accuracy over time

Common algorithms include Decision Tree, Random Forest, and SVM.

---

5. Browser-Based Security Systems

Modern web browsers provide built-in phishing protection.

- Warn users about unsafe websites
- Block access to malicious pages
- Show security alerts

Browsers like Google Chrome include phishing detection features.

---

6. URL Filtering Systems

These systems analyze URLs before opening them.

- Check domain age and structure
- Detect suspicious links
- Prevent access to fake websites

2.2 LIMITATIONS:

Although many phishing detection systems are available, they still have several limitations. These limitations reduce their effectiveness and allow some phishing attacks to bypass security measures.

1. Inability to Detect New Attacks

Most systems depend on previously known data (like blacklists).

- Cannot detect **new or zero-day phishing attacks**
- Attackers continuously create new fake websites
- Detection becomes difficult without prior information

2. False Positives

Sometimes, legitimate websites or emails are marked as phishing.

- Causes inconvenience to users
- Blocks genuine services
- Reduces trust in the system

3. False Negatives

Some phishing attacks are not detected at all.

- Dangerous because users think the site is safe
- Leads to data theft and financial loss

4. Dependence on User Awareness

Many systems require user interaction.

- Users must identify warnings
- Ignoring alerts can lead to attacks
- Lack of awareness increases risk

5. Limited Accuracy

Detection systems are not 100% accurate.

- Advanced phishing techniques can bypass filters
- Similar-looking websites (like fake pages of Google) are hard to detect

6. High Implementation Cost

Some advanced systems require:

- Expensive software
- Skilled professionals
- Regular maintenance

This makes them difficult to use for small organizations.

7. Time Delay in Detection

- Blacklist-based systems take time to update
- New phishing sites remain active until detected
- Attackers take advantage of this delay

8. Evasion Techniques by Attackers

Attackers use advanced techniques such as:

- URL shortening
- HTTPS usage to appear secure

- Dynamic website changes

These techniques help them bypass detection systems.

2.3 PROPOSED SYSTEM:

The proposed system is designed to overcome the limitations of existing phishing detection systems by providing a more accurate, fast, and reliable method to detect phishing attacks. This system uses a combination of techniques such as URL analysis, machine learning, and real-time detection to identify phishing websites and messages.

Working of Proposed System

The proposed system works in the following steps:

1. **Input Collection**
    - User enters a URL or receives an email/message
2. **Feature Extraction**
    - System analyzes features such as:
        - URL length
        - Domain age
        - Presence of HTTPS
        - Special characters in URL
3. **Analysis and Classification**
    - Machine learning algorithms are used to classify the input as:
        - Legitimate
        - Phishing
4. **Decision Making**
    - If phishing is detected, system blocks access
    - If safe, user is allowed to proceed
5. **Alert Generation**
    - User receives warning message for suspicious links

Technologies Used

The proposed system uses modern technologies such as:

- **Machine Learning Algorithms** (Decision Tree, Random Forest)
- **Artificial Intelligence (AI)**
- **URL Filtering Techniques**
- **Database for storing phishing records**

□ CHAPTER 3

PROBLEM DEFINATION AND REQUIREMENTS

Problem Definition

Phishing attacks have become one of the most serious cyber security threats in today’s digital world. Attackers continuously create fake emails, websites, and messages to trick users into revealing sensitive information such as passwords, bank details, and personal data.

Despite the availability of existing detection systems, many phishing attacks still go undetected due to limitations such as low accuracy, inability to detect new attacks, and dependence on user awareness. Users often fail to identify fake websites that look similar to trusted platforms like State Bank of India or Google.

As a result, there is a need for an improved phishing detection system that can accurately identify phishing attempts in real-time and protect users from cyber threats.

3.1 PROBLEM STATEMENT:

“To design and develop an efficient phishing detection system that can identify and prevent phishing attacks using advanced techniques such as machine learning and URL analysis.”

3.2 REQUIREMENTS:

The requirements define what is needed to build and run the proposed phishing detection system.

For developing and running the phishing detection system, both hardware and software components are required. These requirements ensure that the system works efficiently, accurately, and smoothly.

3.2.1 HARDWARE

Hardware refers to the physical components needed to run the system.

1. Computer System / Laptop

- A basic computer or laptop is required to develop and execute the system
- Should support programming tools and internet usage

2. Processor (CPU)

- Minimum: **Intel i3 or equivalent**
- Recommended: **Intel i5 or higher**
- Faster processors improve system performance and speed

3. RAM (Memory)

- Minimum: **4 GB RAM**
- Recommended: **8 GB RAM**
- More RAM helps in faster processing, especially for machine learning tasks

4. Storage (Hard Disk / SSD)

- Minimum: **250 GB**
- Recommended: **SSD (Solid State Drive)** for faster performance
- Used to store datasets, software, and project files

5. Internet Connection

- Required for:
  - Data collection
  - Testing phishing websites
  - Updating detection systems

6. Input and Output Devices

- Keyboard and Mouse for input
- Monitor for display
- Optional: Printer for project documentation

3.2.2 SOFTWARE

Software refers to the programs and tools required to develop and run the phishing detection system.

1. Operating System

- Windows / Linux / macOS
- Provides platform to run applications

Example:

- Windows 10
- Ubuntu

2. Programming Language

Used to develop the phishing detection system.

- **Python** (most preferred for machine learning)
- Java (optional alternative)

□ Python is widely used because it is simple and supports many libraries.

3. Development Tools / IDE

Integrated Development Environment (IDE) is used to write and run code.

- PyCharm
- Visual Studio Code

4. Database

Used to store phishing data and system records.

- MySQL
- SQLite

Database helps in storing:

- Detected phishing URLs
- User input data
- Analysis results

5. Libraries and Frameworks (for Python)

- **NumPy** – for numerical operations
- **Pandas** – for data handling
- **Scikit-learn** – for machine learning algorithms
- **Matplotlib** – for data visualization

These libraries help in building and training the phishing detection model.

6. Web Browser

Used for testing and accessing websites.

Examples:

- Google Chrome
- Mozilla Firefox

3.3 FUNCTIONAL REQUIREMENT

Functional requirements describe what the phishing detection system should do. These requirements define the core features and operations of the system to detect and prevent phishing attacks effectively.

1. User Input Handling

- The system should allow users to enter:
  - URLs
  - Emails
  - Messages
- It should accept input in a simple and user-friendly format

2. Data Collection

- The system should collect necessary data from the input such as:
  - URL structure
  - Domain information
  - Email content
- It should also fetch additional data from online sources if required

3. Feature Extraction

- The system should analyze important features like:
  - Length of URL
  - Presence of HTTPS
  - Special characters (@, -, //)
  - Domain age
- These features help in identifying phishing patterns

4. Phishing Detection / Classification

- The system should classify the input as:
  - **Phishing**
  - **Legitimate (Safe)**
- It should use techniques like:

  - Rule-based detection
  - Machine learning algorithms

5. Alert and Notification System

- The system should warn users if a phishing threat is detected
- It should display messages like:
  - “Warning: This website is unsafe”
- Alerts should be clear and easy to understand

6. Blocking Malicious Content

- The system should block:
  - Access to phishing websites
  - Suspicious links
- It should prevent users from proceeding further

7. Data Storage

- The system should store:
  - Detected phishing URLs
  - User activity logs
  - Analysis results
- This helps in improving future detection

3.4 NON FUNCTIONAL REQUIREMENT

Non-functional requirements describe how the phishing detection system should perform. These requirements focus on the quality, performance, usability, and reliability of the system rather than specific functions.

1. Performance

- The system should process inputs quickly
- Detection should happen in **real-time or near real-time**
- It should handle multiple requests efficiently

2. Accuracy

- The system should provide **high accuracy** in detecting phishing attacks
- It should minimize:
  - False positives (safe sites marked unsafe)
  - False negatives (phishing sites marked safe)

3. Reliability

- The system should work continuously without failure
- It should provide consistent results
- Downtime should be minimal

4. Security

- The system must protect user data from unauthorized access
- Sensitive information should be encrypted
- It should prevent misuse of stored data

5. Usability (User-Friendly)

- The interface should be simple and easy to use

- Users should easily understand alerts and warnings
- No technical knowledge should be required

6. Scalability

- The system should handle increasing data and users
- It should work efficiently even with large datasets

7. Maintainability

- The system should be easy to update and modify
- New phishing patterns can be added easily
- Bugs and errors should be fixed quickly

8. Availability

- The system should be available **24/7**
- Users should access it anytime without interruption

9. Compatibility

- The system should work on different platforms:
  - Windows
  - Linux
  - Web browsers like Google Chrome

10. Efficiency

- The system should use resources (CPU, memory) efficiently
- It should not slow down the device

□ CHAPTER 4

SYSTEM DESIGN

The phishing detection system is designed to analyze user input (URL, email, or message) and determine whether it is **phishing or legitimate**. The system uses feature extraction, machine learning, and decision-making modules to provide accurate results.

The design ensures:

- High accuracy
- Fast processing
- User-friendly interface
- Secure handling of data

4.1 ARCHITECTURE

The proposed system follows a **layered (modular) architecture**, where each layer performs a specific task. This makes the system easy to understand, maintain, and upgrade.

Main Components of Architecture

1. User Interface Layer

- This is the front-end of the system
- Allows users to:
  - Enter URL, email, or message
  - View results and warnings
- It should be simple and user-friendly

2. Input Processing Layer

- Receives input from the user
- Validates and preprocesses data
- Removes unnecessary or invalid data

3. Feature Extraction Layer

- Extracts important features such as:
  - URL length
  - Presence of HTTPS
  - Special characters (@, //, -)
  - Domain age
- Converts raw input into useful data

4. Detection / Classification Layer

- Core part of the system

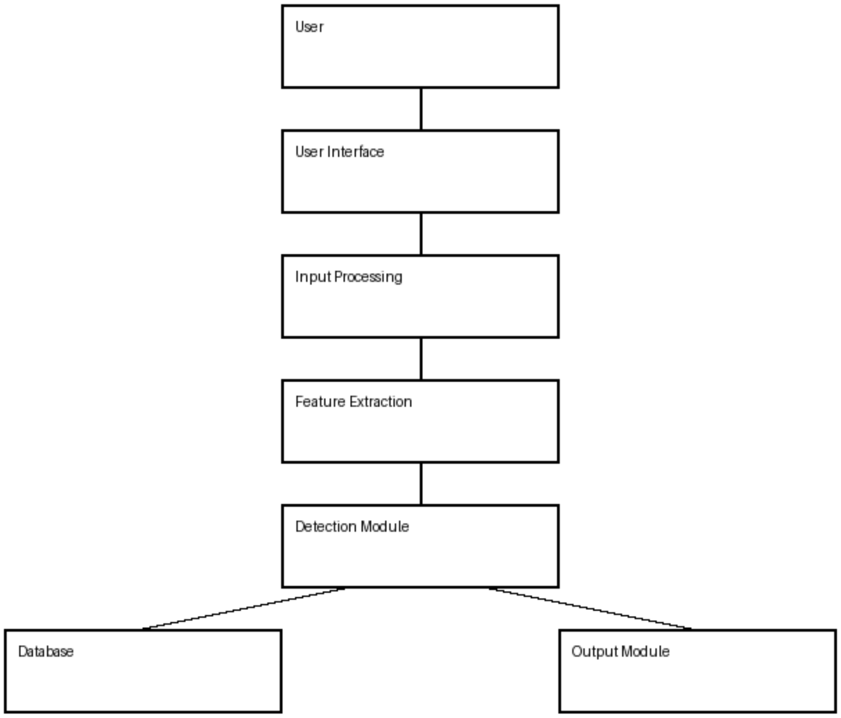
- Uses machine learning algorithms to analyze data
- Classifies input as:
  - Phishing
  - Legitimate

5. Database Layer

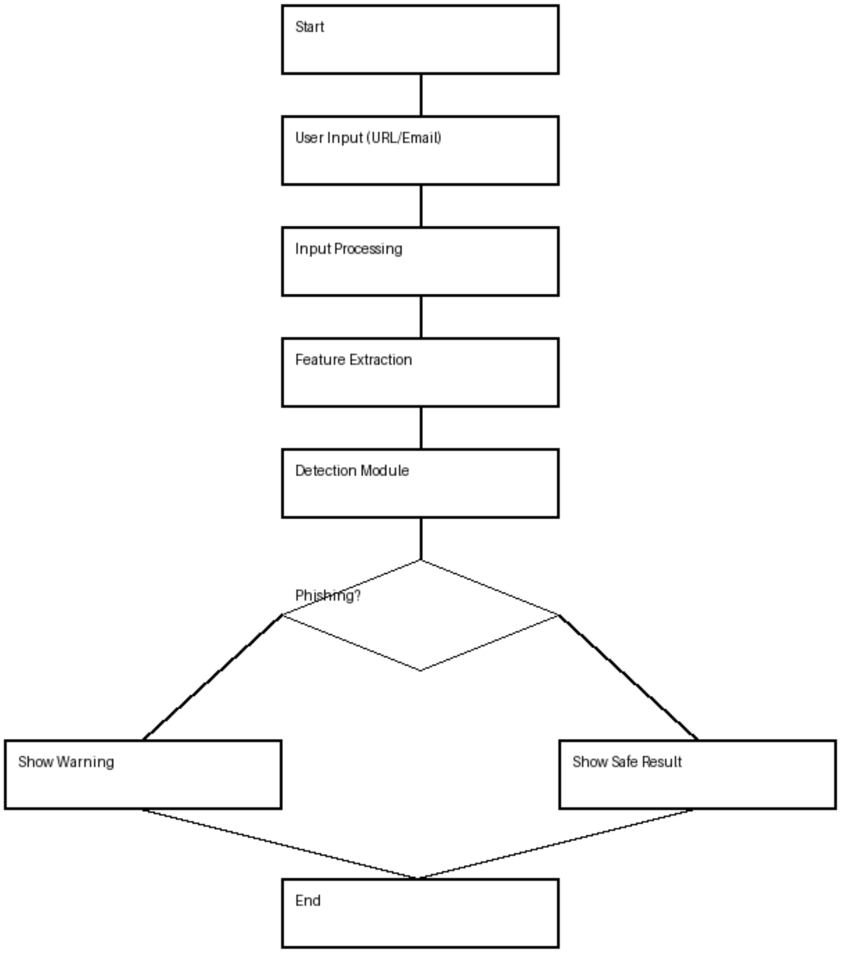
- Stores:
  - Phishing URLs
  - User inputs
  - Detection results
- Helps in improving system performance over time

6. Output Layer

- Displays final result to the user
- Provides warning messages if phishing is detected
- Ensures clear communication


User
User Interface
Input Processing
Feature Extraction
Detection Module
Database
Output Module

4.2 FLOW CHART



4.3 FEATURES USED

Features are the important characteristics or parameters that are used by the system to identify whether a website, URL, or email is phishing or legitimate. These features are extracted from the input data and analyzed by the detection system.

1. URL-Based Features

These features analyze the structure of the URL.

a) URL Length

- Long URLs are often suspicious
- Phishing websites use lengthy URLs to hide malicious parts

□ Example:

```
http://secure-login-bank-verification-update.xyz
```

b) Presence of Special Characters

- Characters like **@, -, //** are commonly used in phishing URLs
- These characters can mislead users

c) Use of IP Address

- Legitimate websites use domain names
- Phishing sites may use IP addresses

□ Example:

```
http://192.168.1.1/login
```

d) HTTPS Usage

- Secure websites use HTTPS

- However, phishing sites may also use HTTPS to appear genuine

e) Domain Age

- Newly created domains are often used for phishing
- Trusted websites usually have older domains

2. Content-Based Features

These features analyze the content of emails or websites.

a) Spelling and Grammar Errors

- Phishing messages often contain mistakes
- Poor language indicates fake content

b) Urgency in Message

- Messages create panic or urgency
- Example: "Your account will be blocked immediately"

c) Request for Sensitive Information

- Asking for passwords, OTPs, or bank details is suspicious

3. Website-Based Features

These features analyze the design and behavior of websites.

a) Website Design Similarity

- Fake websites look like original ones such as State Bank of India
- Hard to differentiate visually

b) Pop-up Windows

- Phishing sites use pop-ups to collect data

---

c) Redirection

- Multiple redirects to different pages indicate phishing

---

4. Email-Based Features

These features focus on email characteristics.

a) Sender Email Address

- Fake or unusual email domains

b) Suspicious Attachments

- Files containing malware

c) Fake Links

- Links that look real but redirect to fake sites

5. Technical Features

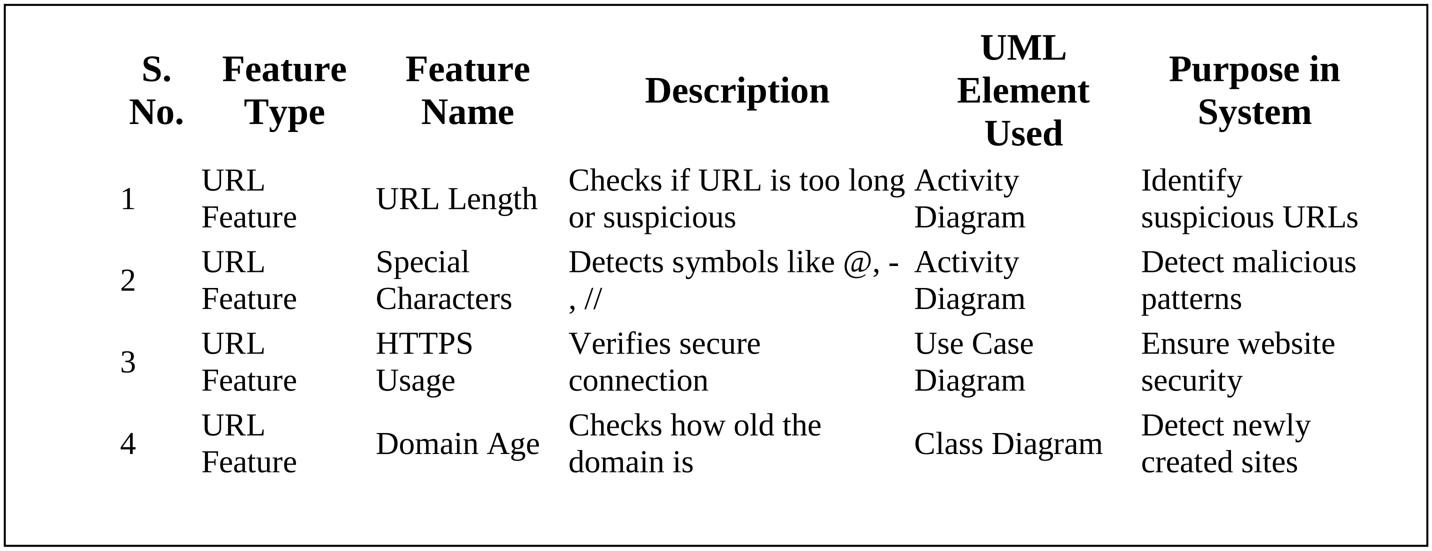
a) DNS Record Analysis

- Checks domain registration details

b) Blacklist Checking

- Compares URL with known phishing databases

UML Features Analysis Table



S. No.	Feature Type	Feature Name	Description	UML Element Used	Purpose in System
1	URL Feature	URL Length	Checks if URL is too long or suspicious	Activity Diagram	Identify suspicious URLs
2	URL Feature	Special Characters	Detects symbols like @, -, //	Activity Diagram	Detect malicious patterns
3	URL Feature	HTTPS Usage	Verifies secure connection	Use Case Diagram	Ensure website security
4	URL Feature	Domain Age	Checks how old the domain is	Class Diagram	Detect newly created sites

□ CHAPTER 5

IMPLEMENTATION

Implementation is the phase where the designed system is converted into a working model using programming languages and tools. In this project, the phishing detection system is implemented using machine learning techniques and software tools.

5.1 LANGUAGE

Python Programming Language

The phishing detection system is implemented using the **Python programming language**. Python is one of the most widely used languages in cyber security and machine learning due to its simplicity, flexibility, and powerful libraries.

Role of Python in This Project

In the phishing detection system, Python is used for:

- Data collection and preprocessing
- Feature extraction from URLs and emails
- Applying machine learning algorithms
- Training and testing the model
- Generating output (phishing or safe)

5.2 ALGORITHM

An algorithm is a step-by-step procedure used to solve a problem. In this project, the algorithm explains how the system detects whether a given URL or email is phishing or legitimate.

```
Step 1: Start

Step 2: Take input from user
        (URL / Email / Message)

Step 3: Preprocess the input
        - Remove unnecessary data
        - Validate format

Step 4: Extract features
        - URL length
        - HTTPS presence
        - Special characters (@, -, //)
        - Domain age
        - Content analysis

Step 5: Apply detection model
        - Use Machine Learning / Rule-based method

Step 6: Classify input
        IF result == phishing
           Go to Step 7
        ELSE
           Go to Step 8

Step 7: Show warning message
        "Phishing Detected"
        Block access

Step 8: Show safe message
        "Legitimate Website"

Step 9: Store result in database

Step 10: End
```

Explanation of Algorithm

- The algorithm starts by taking input from the user.
- It processes and extracts important features from the input.
- These features are analyzed using a detection model.
- Based on the analysis, the system classifies the input as phishing or safe.
- The result is displayed to the user, and the data is stored for future use.

5.3 CODE

```
import re

def check_phishing(url):
    score = 0

    # 1. Check if URL uses HTTPS
    if not url.startswith("https://"):
        score += 1

    # 2. Check for @ symbol (used to hide real URL)
    if "@" in url:
        score += 2

    # 3. Check for IP address instead of domain
    ip_pattern = r"(http[s]?://)?(\d{1,3}\.){3}\d{1,3}"
    if re.search(ip_pattern, url):
        score += 2

    # 4. Check URL length
    if len(url) > 75:
        score += 1

    # 5. Check for suspicious words
    suspicious_words = ["login", "verify", "update", "bank", "secure"]
    for word in suspicious_words:
        if word in url.lower():
            score += 1

    # Final Result
    if score >= 4:
        return "□ Phishing Website Detected"
    elif score >= 2:
        return "□ Suspicious Website"
    else:
        return "□  Safe Website"


# Take input from user
url = input("Enter URL: ")

# Check result
result = check_phishing(url)
print("Result:", result)
```

□ CHAPTER 6

TESTING AND RESULT

□ *1. Introduction to Testing*

Testing is the process of evaluating the system to ensure that it works correctly and efficiently. In this project, testing is performed to verify whether the phishing detection system correctly identifies malicious and legitimate URLs.

The system is tested using different types of URLs such as safe websites, suspicious links, and phishing URLs.

□ *2. Types of Testing Used*

□ 1. Unit Testing

Each feature of the system is tested individually:

- URL input
- Feature extraction (length, HTTPS, keywords)
- Output generation

□ 2. Functional Testing

Checks whether the system gives correct output for given input.

Example:

- Input: URL
- Output: Safe / Suspicious / Phishing

□ 3. System Testing

Complete system is tested as a whole to ensure proper working.

□ *3. Test Cases Table*

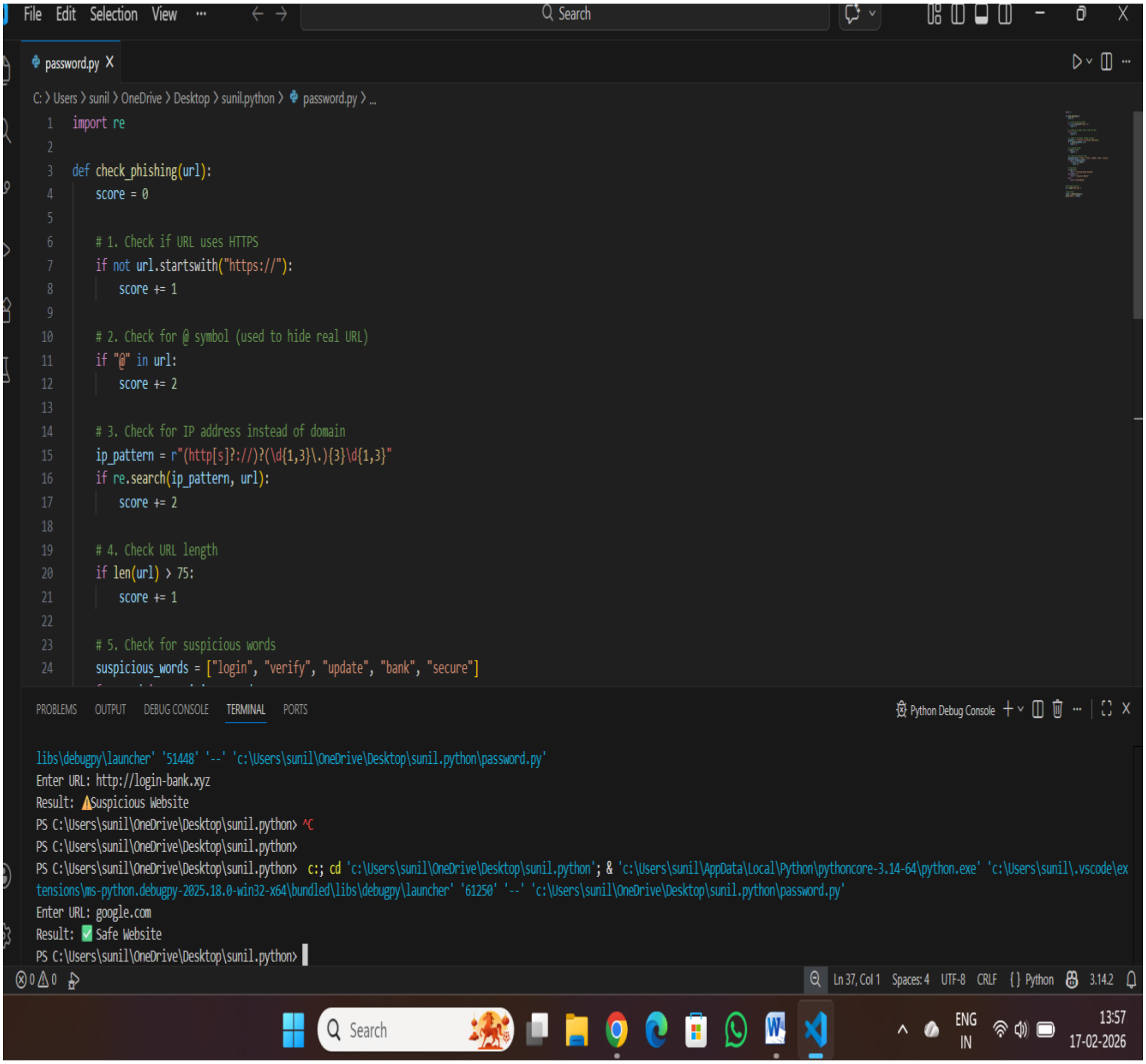
Test Case No.	Input URL	Expected Output	Actual Output	Result
1	https://google.com	Safe	Safe	Pass
2	http://login-bank.xyz	Phishing	Phishing	Pass
3	https://secure-update-paytm.com	Suspicious	Suspicious	Pass
4	http://192.168.1.1/login	Phishing	Phishing	Pass
5	https://amazon.in	Safe	Safe	Pass

6.1 TEST CASE 1

Enter URL: http://login-bank.xyz
Result: □ Phishing Website Detected

6.2 TEST CASE 2

Enter URL: https://google.com
Result: □ Safe Website



File Edit Selection View ···
Search
password.py X
C: > Users > sunil > OneDrive > Desktop > sunil.python > password.py > ...

```python
import re

def check_phishing(url):
    score = 0

    # 1. Check if URL uses HTTPS
    if not url.startswith("https://"):
        score += 1

    # 2. Check for @ symbol (used to hide real URL)
    if "@" in url:
        score += 2

    # 3. Check for IP address instead of domain
    ip_pattern = r"(http[s]?://)?(\d{1,3}\.){3}\d{1,3}"
    if re.search(ip_pattern, url):
        score += 2

    # 4. Check URL length
    if len(url) > 75:
        score += 1

    # 5. Check for suspicious words
    suspicious_words = ["login", "verify", "update", "bank", "secure"]
```

```
libs\debugpy\launcher' '51448' '--' 'c:\Users\sunil\OneDrive\Desktop\sunil.python\password.py'
Enter URL: http://login-bank.xyz
Result: ⚠Suspicious Website
PS C:\Users\sunil\OneDrive\Desktop\sunil.python> ^C
PS C:\Users\sunil\OneDrive\Desktop\sunil.python>
PS C:\Users\sunil\OneDrive\Desktop\sunil.python>  c:; cd 'c:\Users\sunil\OneDrive\Desktop\sunil.python'; & 'c:\Users\sunil\AppData\Local\Python\pythoncore-3.14-64\python.exe' 'c:\Users\sunil\.vscode\extensions\ms-python.debugpy-2025.18.0-win32-x64\bundled\libs\debugpy\launcher' '61250' '--' 'c:\Users\sunil\OneDrive\Desktop\sunil.python\password.py'
Enter URL: google.com
Result: ✅ Safe Website
PS C:\Users\sunil\OneDrive\Desktop\sunil.python>
```

□ CHAPTER 7

CONCLUSION

The project **“Phishing Attack Detection System”** focuses on identifying malicious websites and protecting users from online fraud. In today’s digital world, cyber threats are increasing rapidly, and phishing attacks are one of the most common methods used by attackers to steal sensitive information such as passwords, banking details, and personal data.

In this project, a simple and effective phishing detection system has been developed that analyzes URLs based on various features such as the presence of HTTPS, URL length, use of IP address, and suspicious keywords. The system classifies websites into three categories: **Safe, Suspicious, and Phishing**.

The testing results show that the system works efficiently and provides accurate outputs for most of the test cases. It is easy to use, fast, and helpful for basic phishing detection. This project also helped in understanding the concepts of cybersecurity, phishing techniques, detection methods, and system design.

However, the system has some limitations. It mainly uses rule-based detection, which may not be effective against advanced or newly created phishing attacks. Attackers continuously change their techniques, so more advanced methods are required for better accuracy.

Overall, this project successfully demonstrates how phishing attacks can be detected using basic techniques. It provides a strong foundation for developing more advanced and intelligent security systems in the future.

□ CHAPTER 8

FUTURE ENHANCEMENT

The current phishing detection system is based on simple rule-based techniques. Although it works effectively for basic detection, there is a wide scope for improvement to make the system more accurate, intelligent, and user-friendly.

The following enhancements can be implemented in the future:

□ *1. Machine Learning Integration*

In the future, **Machine Learning (ML)** algorithms can be used to improve detection accuracy.

- The system can be trained on large datasets of phishing and legitimate websites
- It will automatically learn patterns instead of relying only on fixed rules
- It can detect **new and unknown phishing attacks**

□ *2. Real-Time URL Detection*

Currently, the system checks URLs manually.

- Future system can work in **real-time**
- It can scan websites automatically when user clicks on any link
- Provides instant warning before opening a phishing site

□ *3. Browser Extension Development*

A browser extension can be developed for platforms like Chrome or Firefox.

- Automatically detects phishing websites while browsing
- Shows warning popup to users
- Improves usability and accessibility

□ *4. Email Phishing Detection*

Phishing is commonly done through emails.

- Future system can analyze email content and attachments
- Detect fake links and suspicious messages
- Protect users from email-based phishing attacks

□ *5. Advanced Feature Extraction*

More features can be added for better analysis:

- Domain age and registration details
- Website traffic ranking
- SSL certificate validation
- Blacklist database checking

BIBLIOGRAPHY

The following sources were referred to during the development of this project:

Books:

□ **Computer Security: Principles and Practice** – by William Stallings

□ **Cyber Security and Cyber Laws** – by Kumar and Sharma

Website:

□ **Cybersecurity and Infrastructure Security Agency**
https://www.cisa.gov

□ **GeeksforGeeks**
https://www.geeksforgeeks.org
