IBM Innovation Center for Education

PHEMESOFT Software That Matters
TRUSTED DELIVERY PARTNER for IBM ICE

YENEPOYA UNIVERSITY

RABBI ZIDNI 'ILMA
YENEPOYA
(DEEMED TO BE UNIVERSITY)

Phishing Detection using AI

PROJECT SYNOPSIS

Phishing Detection using ai

BACHELOR OF COMPUTER APPLICATIONS
Cyber Forensics, Data Analytics and
Cyber security

SUBMITTED BY
MUHAMMED ADIL RABEEH - 22BCACDC39

GUIDED BY

SUMITH K SHUKLA

Table of Contents

1. Introduction

Phishing Detection System Using AI is a web-based cybersecurity tool designed to proactively detect, analyze, and report phishing attempts targeting an organization's digital infrastructure. With the surge in online communication and data exchange, phishing attacks—where attackers impersonate trusted entities to steal sensitive information—have become increasingly frequent and sophisticated. Early detection is essential to protect users and critical assets.

This system integrates two core functionalities — an Email Content Analyzer and a URL Reputation Checker. The Email Content Analyzer uses AI and natural language processing (NLP) to scan emails for deceptive language, suspicious patterns, and spoofed identities. It detects anomalies and flags potentially harmful messages for review.

The URL Reputation Checker evaluates embedded links in emails by querying trusted sources like VirusTotal and PhishTank. These services maintain databases of known phishing domains, helping assess whether a link is malicious or safe to interact with.

Built with the Django framework for its scalability and security, the system uses SQLite as a lightweight database and Chart.js for visualizing phishing trends and alerts. This platform empowers security teams to automate phishing detection, reduce manual effort, and respond swiftly to potential threats.

2.Methodology / Planning of Work

The development of the AI-based Phishing Detection System adopts a structured, phased methodology to ensure that each component is thoroughly planned, developed, and validated for accuracy and efficiency. The workflow comprises the following key stages:

Requirement Analysis: Conduct in-depth research to collect both functional and non-functional requirements. This includes consulting cybersecurity frameworks, analyzing existing phishing detection systems, and engaging with prospective users to define system objectives, accuracy thresholds, and technical constraints.

UI and Database Design: Develop an intuitive and responsive user interface using HTML, CSS, and JavaScript, aimed at providing smooth navigation for both technical and non-technical users. Concurrently, design a secure and normalized SQLite database to store scanned emails, extracted URLs, threat classifications, and API results, ensuring data integrity and retrievability.

Development of Email Content Analyzer: Create an AI-powered module using machine learning and natural language processing (NLP) to evaluate email content. The module detects phishing characteristics such as urgent tone, deceptive language, and mismatched sender information, flagging suspicious emails for review.

API Integration: Integrate external threat intelligence services like VirusTotal and PhishTank to assess the reputation of embedded URLs and domains. The system securely queries these APIs and categorizes results based on their threat level to aid in automated decision-making.

Module Testing: Conduct thorough unit testing of each component to ensure accuracy, performance, and resilience under various phishing scenarios. Key focus areas include AI model precision, response latency, and handling of edge cases such as obfuscated links or spoofed addresses.

Final Integration and Deployment: Combine all modules into a cohesive application, perform integration testing to verify seamless interoperation, and deploy the system on a secure server environment. Full-system validation confirms the platform's reliability, scalability, and readiness for operational use.

3.Facilities Required

The successful design, development, testing, and deployment of the AI-based Phishing Detection System necessitate a combination of software tools, hardware resources, and networking facilities. Each component is selected for its compatibility, efficiency, and support for AI and cybersecurity-focused applications.

Software Requirements:

Django Framework (Python): A secure and scalable web framework ideal for rapid development of AI-integrated applications.

SQLite Database: A lightweight, embedded database used for storing analyzed email content, URLs, and classification results without complex configurations.

HTML, CSS, JavaScript: Essential for developing a responsive and user-friendly interface accessible via standard web browsers.

Chart.js Library: A JavaScript charting tool used to visualize phishing trends, detection rates, and threat classification outcomes.

VirusTotal and PhishTank APIs: External threat intelligence services that provide reputation scores for URLs and domains, aiding in phishing detection.

Python Libraries for AI: Libraries such as scikit-learn, TensorFlow/Keras, and NLTK or spaCy for building, training, and deploying machine learning and NLP models.

Hardware Requirements:

> **Computer System**: A development machine with at least 8GB RAM, a multi-core processor, and 500GB of storage is recommended to support machine learning tasks and local server deployment.
>
> **Stable Internet Connection**: High-speed connectivity is essential for accessing external APIs, downloading datasets or model updates, and real-time verification of suspicious links or domains.

These resources collectively ensure smooth development, effective AI processing, and reliable deployment, enabling proactive phishing detection and system scalability.

4.References

[1] Django Documentation: https://docs.djangoproject.com/
[2] SQLite Documentation: https://www.sqlite.org/docs.html
[3] Chart.js Documentation: https://www.chartjs.org/docs
[4] PhishTank API Documentation: https://www.phishtank.com/developer_info.php
[5] VirusTotal API Documentation: https://developers.virustotal.com/