**A**

**PROJECT REPORT**

**ON**

**AI IN SOCIAL ENGINEERING DEFENSE**

**SUBMITTED TO**

**SHIVAJI UNIVERSITY, KOLHAPUR**

**IN THE PARTIAL FULFILLMENT OF THE REQUIREMENT**

**FOR THE AWARD OF DEGREE**

**BACHELOR OF TECHNOLOGY IN COMPUTER SCIENCE AND ENGINEERING**

**SUBMITTED BY**

|  |  |  |
| --- | --- | --- |
| **MISS.** | **APURVA ANIL JOSHI** | **23UAD301** |
|  |  |  |
|  |  |  |

**UNDER THE GUIDANCE OF**

**Mr. S. P. Pise**



**DEPARTMENT OF ARTIFICIAL INTELLIGENCE AND DATA SCIENCE ENGINEERING**

**DKTE SOCIETY’S TEXTILE AND ENGINEERING INSTITUTE, ICHALKARANJI**

**(AN EMPOWERED AUTONOUMOUS INSTITUTE)**

**2024-2025**

**D.K.T.E. SOCIETY’S**

**TEXTILE AND ENGINEERING INSTITUTE, ICHALKARANJI**

**(AN EMPOWERED AUTONOUMOUS INSTITUTE)**

**DEPARTMENT OF ARTIFICIAL INTELLIGENCE AND DATA SCIENCE ENGINEERING**



**CERTIFICATE**

**This is to certify that, project work entitled**

**AI IN SOCIAL ENGINEERING DEFENSE**

**is a bonafide record of project work carried out in this college by**

|  |  |  |
| --- | --- | --- |
| **MISS.** | **APURVA ANIL JOSHI** | **23UAD301** |
|  |  |  |
|  |  |  |

**is in the partial fulfillment of award of degree Bachelor of Technology in Artificial Intelligence and Data Science Engineering prescribed by Shivaji University, Kolhapur for the academic year 2024-2025.**

**MR. S. P. PISE**

**(PROJECT GUIDE)**

**PROF. (DR.) T. I. BAGBAN PROF.(DR.) L.S.ADMUTHE**

**(HOD AI & DS DEPT.) (DIRECTOR)**

**EXAMINER: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**DECLARATION**

We hereby declare that, the project work report entitled  “AI in Social Engineering Defense” which is being submitted to D.K.T.E. Society’s Textile and Engineering Institute Ichalkaranji, affiliated to Shivaji University, Kolhapur is in partial fulfillment of degree B.Tech.(AI & DS). It is a bonafide report of the work carried out by us. The material contained in this report has not been submitted to any university or institution for the award of any degree. Further, we declare that we have not violated any of the provisions under Copyright and Piracy / Cyber / IPR Act amended from time to time.

|  |  |  |  |
| --- | --- | --- | --- |
| **Title** | **Name of the Student** | **PRN** | **Signature** |
| MISS. | APURVA ANIL JOSHI | 23UAD301 |  |
|  |  |  |  |
|  |  |  |  |

**ACKNOWLEDGEMENT**

With great pleasure we wish to express our deep sense of gratitude to Mr. S. P. Pise for his valuable guidance, support, and encouragement in the completion of this project report.

Also, we would like to take the opportunity to thank our head of department Dr. T. I. Bagban for his cooperation in preparing this project report.

We feel gratified to record our cordial thanks to other staff members of the Artificial Intelligence and Data Science Department for their support, help, and assistance which they extended as and when required.

Thank you,

|  |  |  |
| --- | --- | --- |
| **Title** | **Name of the Student** | **PRN** |
| MISS. | APURVA ANIL JOSHI | 23UAD301 |
|  |  |  |
|  |  |  |

**ABSTRACT**

This project presents an AI-based defense system against social engineering attacks, specifically targeting phishing URLs and SMS spam detection. Social engineering is a growing threat where attackers manipulate individuals into revealing confidential information. Our system integrates multiple AI models to analyze and detect phishing URLs and spam SMS messages in real-time. It uses a Python-based machine learning backend (Random Forest for phishing detection, Naive Bayes for SMS spam) integrated with a Node.js API. MongoDB is used for storing detection results. A simple frontend displays the classification results and stores history. The system is scalable and can be integrated with broader cybersecurity frameworks.

**INDEX**

1. **Introduction PgNo 01** 
   1. Problem definition
   2. Aim and objective of the project
   3. Scope and limitation of the project
   4. o>>
2. **Background study and literature overview PgNo 02** 
   1. Literature overview
   2. Investigation of current project and related work
3. **Requirement analysis PgNo 3-4**

a. Requirement Gathering

b. Requirement Specification

c. Use case Diagram

1. **System design PgNo 5-6** 
   1. Architectural Design
   2. Flow Chart
   3. System Modeling
      1. Dataflow Diagram
2. **Implementation PgNo 07**

a. Agile Methodologies

b. Development Model

1. **Future Scope PgNo 08**
2. **References (Public repository GitHub source code links) PgNo 09**

**1. Introduction**

1. **Problem definition**

To build an AI-powered system capable of detecting phishing URLs and spam SMS messages by analyzing their content and structure, thereby preventing users from falling victim to social engineering attacks.

1. **Aim and objective of the project**

* To identify and classify phishing URLs using machine learning models.
* To detect spam SMS messages using natural language processing techniques.
* To integrate a real-time detection system with a user-friendly web interface.
* To store and manage detection results using a NoSQL database (MongoDB).
* To provide a scalable, secure, and efficient solution for social engineering defense.

1. **Scope and limitation of the project**

**Scope:**

* The system provides real-time detection of phishing URLs and spam SMS messages.
* Integration with a web-based frontend allows easy user interaction.
* Backend supports modular AI models for future extensibility.
* MongoDB integration allows storage and retrieval of detection logs.
* Can be deployed on cloud for broader accessibility.

**Limitations:**

* Models rely on training data; unseen types of attacks may reduce accuracy.
* Does not cover image-based or voice-based phishing in current version.
* Requires internet access for full functionality (cloud interaction).
* Limited to English-language inputs for accurate SMS spam detection.

**2. Background study and literature overview**

1. **Literature overview**

The use of AI in cybersecurity has been widely explored in recent years. Several research papers highlight the effectiveness of machine learning techniques like Decision Trees, Random Forest, Support Vector Machines, and Naive Bayes for phishing and spam detection. Natural Language Processing (NLP) methods have also proven effective in analyzing textual content of SMS and emails. The combination of these models helps in developing robust detection systems that can process large volumes of data in real-time.

1. **Investigation of current project and related work**

There are many existing anti-phishing and anti-spam systems, including browser-based warnings and email filters. However, most of these rely on blacklists or predefined rules, which are not effective against new, evolving attacks. Our system goes beyond by using AI to generalize detection to unseen data.

For instance, Google Safe Browsing API helps flag malicious links, but lacks real-time learning. SMS spam filters in mobile devices often require manual feedback. Our AI models trained on large datasets can detect patterns and make autonomous decisions.

Compared to traditional methods, our system's advantage lies in its learning capability, real-time analysis, cross-platform compatibility, and cloud-integrated architecture.

**3. Requirement analysis**

1. **Requirement Gathering**

The requirements were gathered through:

* Analysis of existing social engineering threats (phishing URLs, spam SMS).
* User needs for a responsive and easy-to-use interface.
* Technical feasibility for real-time classification.
* Selection of appropriate machine learning models based on past literature.

1. **Requirement Specification**

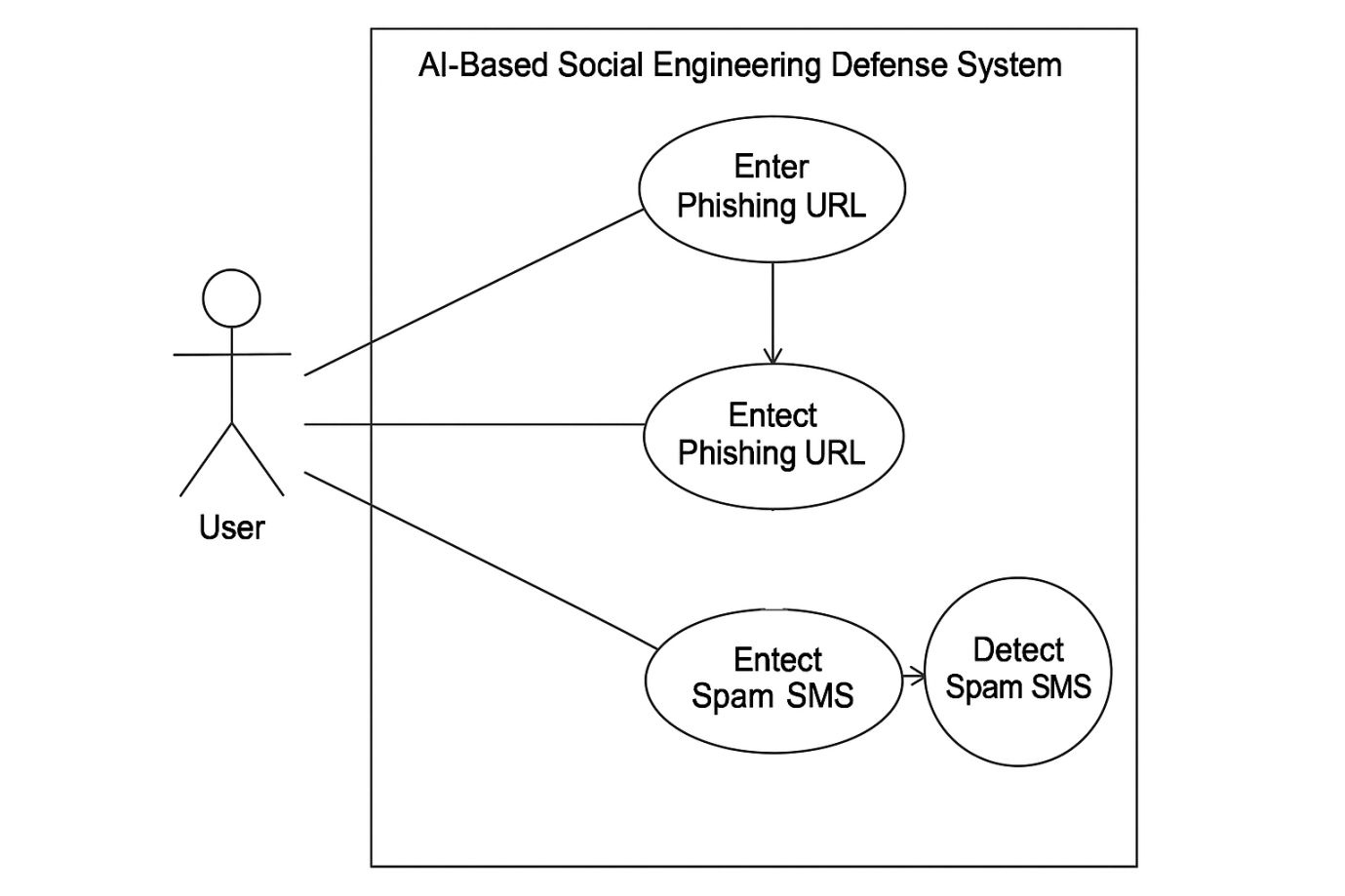
**Functional Requirements:**

* Accept user input (URL or SMS message).
* Process input using respective ML models.
* Display result as "Phishing/Safe" or "Spam/Ham".
* Store results in MongoDB.
* Provide frontend UI to interact with the system.

**Non-Functional Requirements:**

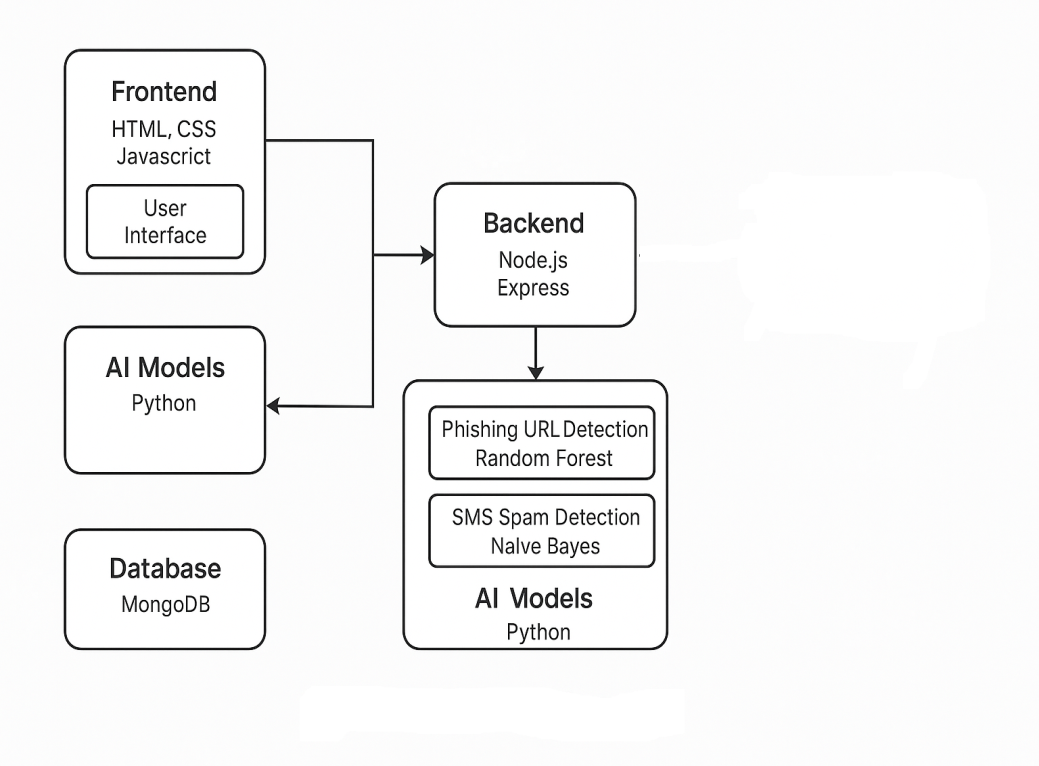
* High system accuracy and low latency.
* Scalable backend design.
* Secure API communication.
* Easy integration with cloud platforms.

1. **Use case Diagram**

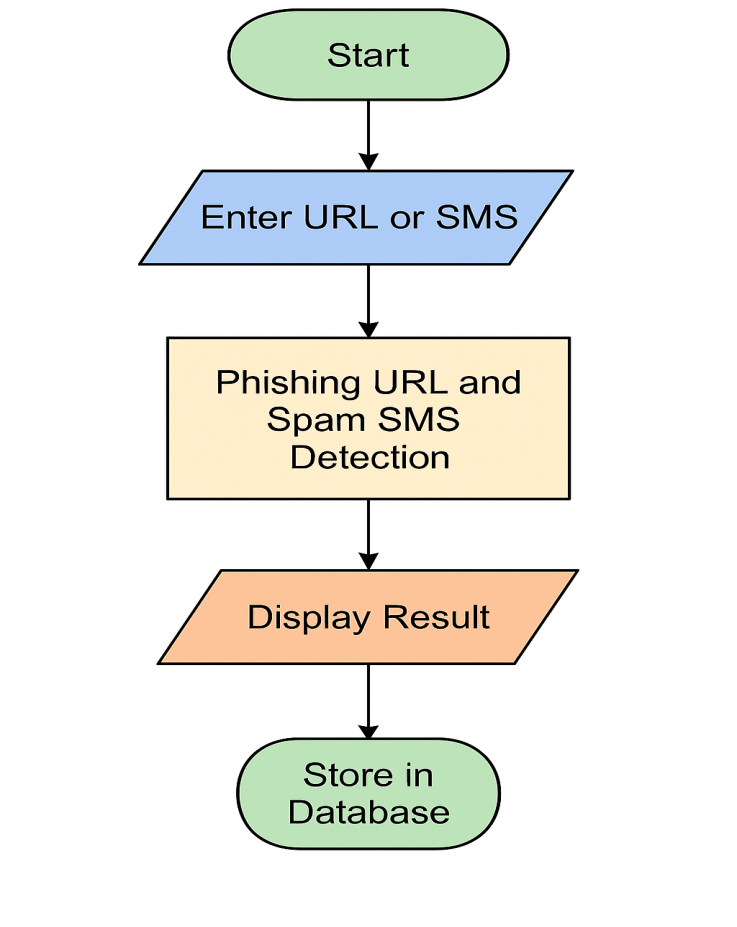


**4. System design**

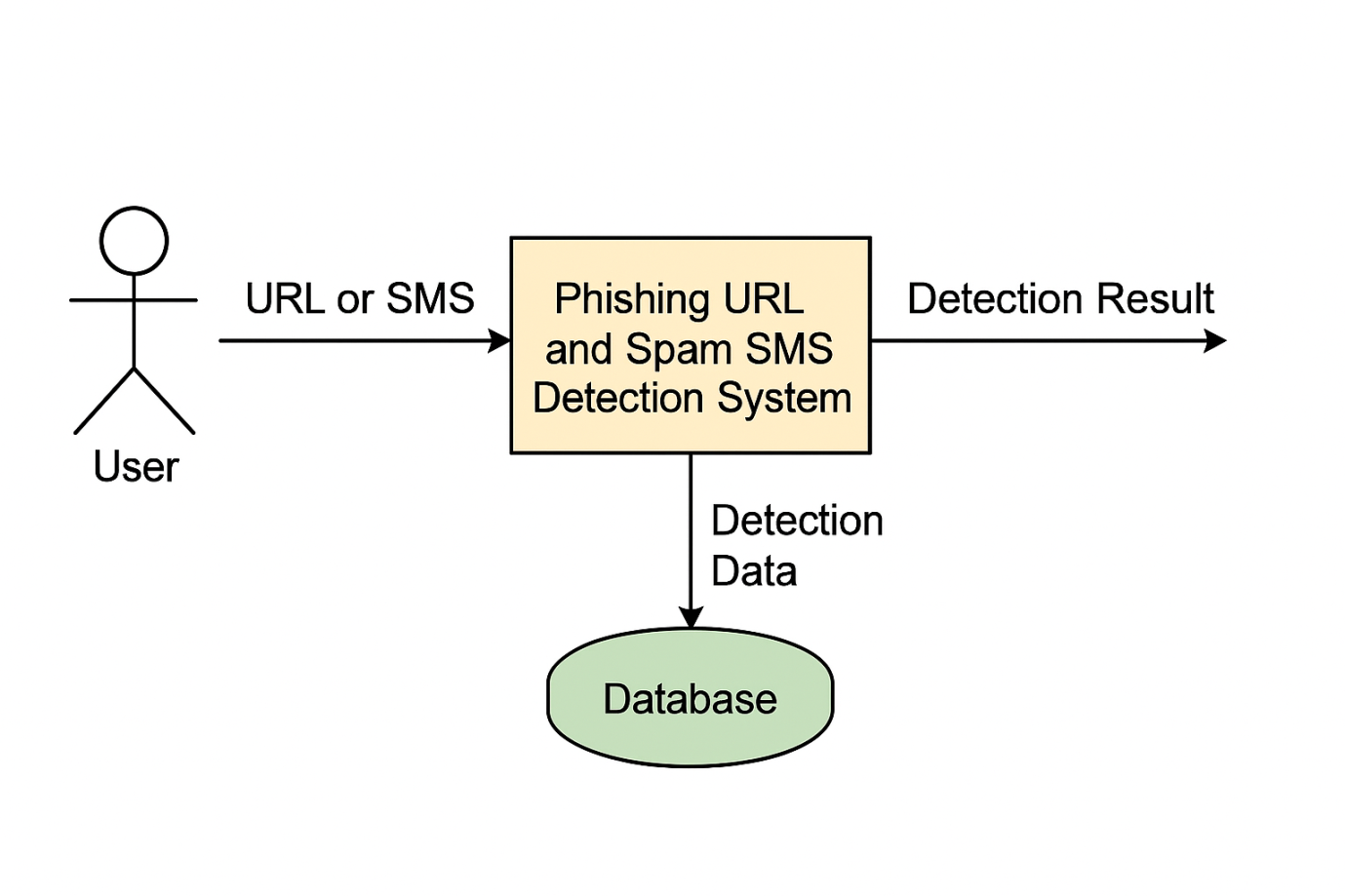
1. **Architectural Design**



1. **Flow Chart**



1. **System Modeling**
2. **Dataflow Diagram**



**5. Implementation**

1. **Agile Methodologies**

Agile methodology is a modern, iterative approach to software development that emphasizes flexibility, collaboration, and customer feedback. Instead of building the entire system at once, Agile encourages breaking down the project into smaller, manageable units called sprints.

**Type of Project: Research-Based**

 As I am using **machine learning models** to detect phishing and spam, which requires **experimenting, training, and testing** with different algorithms and datasets.

 The focus is on finding an **effective solution to a real-world problem** by researching phishing behavior, SMS patterns, and detection techniques.

 This includes collecting datasets, cleaning them, training models, tuning parameters, and evaluating performance.

In this project, Agile helped in managing the dual focus on phishing URL detection and SMS spam filtering, allowing parallel development and testing of both modules efficiently.

1. **Development Model**

#### Development Model Used: Spiral Model

**Spiral Model Phases in Your Project**

Each cycle of the spiral includes four major activities:

1. **Planning Phase:**
   * Requirement gathering for phishing and SMS modules
   * Choosing appropriate technologies (Python, Node.js, MongoDB, etc.)
2. **Risk Analysis:**
   * Evaluating performance of chosen ML models (e.g., Random Forest, Naive Bayes)
   * Testing data reliability and model overfitting
3. **Engineering:**
   * Model training and testing
   * Developing frontend and backend APIs
   * Integrating Python models with Node.js using PythonShell
4. **Evaluation:**
   * User and developer feedback
   * Adjusting the models, improving interface, and optimizing storage logic

**6. Future Scope**

* Integration of voice phishing (vishing) detection using speech-to-text and DeepSpeech.
* Behavioral anomaly detection based on usage patterns.
* Deployment on cloud platforms with real-time alerting systems.
* Browser extension or mobile app integration.

**7. References (public repository GitHub source code links)**

**Dataset:**

1. <http://data.phishtank.com/data/online-valid.csv>
2. https://www.kaggle.com/datasets/uciml/sms-spam-collection-dataset

**ML Models**

1. <https://scikit-learn.org/RandomForestClassifier>
2. https://scikit-learn.org/ naive\_bayes

**Github link:**

* https://github.com/Apurva-02/AI-in-Social-Engg-Defense