Chapter Five: Testing and Evaluation

# 5.1 Introduction

The purpose of this chapter is to present the testing methodologies and evaluation metrics used to assess the performance, accuracy, and reliability of the phishing detection system. Testing was conducted at various stages of development to ensure the system's functionality, usability, and security. This chapter provides an overview of the different types of tests carried out, including unit testing, integration testing, and system testing. Furthermore, it covers the evaluation metrics used to measure the effectiveness of the machine learning model in detecting phishing emails.

# 5.2 Testing Methodologies

Various testing strategies were employed throughout the development process to ensure the system met the functional and non-functional requirements. The testing methodologies can be categorized into the following:

## 5.2.1 Unit Testing

Unit testing was conducted to ensure that individual components of the system functioned as expected. The main components tested included:

• User Authentication: The login, registration, and password reset functionalities were tested to ensure users could securely access the system.  
• Phishing Detection Model: The prediction function of the machine learning model was tested using a variety of phishing and non-phishing emails to confirm its accuracy in identifying phishing attempts.  
• Text Preprocessing: The text preprocessing functions, including tokenization, stopword removal, and TF-IDF vectorization, were tested to verify their correct application on email content.

## 5.2.2 Integration Testing

Integration testing was performed to validate that the different modules of the system worked together seamlessly. This involved testing the interaction between the frontend, backend, and the machine learning model. Key integration points tested included:  
• Frontend and Backend Communication: Ensuring that user inputs from the frontend were correctly sent to the backend API for processing and that responses were accurately reflected on the frontend.  
• Database Integration: Testing database operations such as storing user data, retrieving phishing detection history, and maintaining session integrity.  
• Model Deployment: Verifying that the machine learning model could be loaded into the Flask application and used to make real-time predictions without performance issues.

## 5.2.3 System Testing

System testing was conducted to evaluate the system’s performance as a whole, ensuring that all features functioned correctly from the user’s perspective. The following scenarios were tested:  
• User Registration and Authentication: Testing whether users could sign up, log in, and reset their passwords successfully. Different password strengths and email formats were used to test input validation.  
• Phishing Detection: Submitting both text input and file uploads (PDFs, DOCX, TXT) to the system to check the accuracy of phishing predictions.  
• Detection History: Testing the functionality of the detection history feature, ensuring that users could view past phishing checks, along with timestamps and results.

# 5.3 Evaluation Metrics for Phishing Detection

The performance of the phishing detection model was evaluated using standard classification metrics. The following metrics were used to assess the effectiveness of the system:

## 5.3.1 Accuracy

Accuracy is a commonly used metric for classification models and is defined as the proportion of correct predictions (both phishing and non-phishing) to the total number of predictions made by the model.

Accuracy = (TP + TN) / (TP + TN + FP + FN)