Low Level Document (LLD)

Credit Card Fraud Detection System

Version number: 1.0

Last date of revision: 9 August 2024

Pranav Ghorpade

**DECLARATION**

We declare that this written submission represents us ideas is our own words and where others' ideas or words have been included, we have adequately cited and referenced the original sources.

We also declare that we have adhered to all principles of academic honesty

and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my submission.

We understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when

needed.

Table of Contents

* Introduction …………………………………………………………………………………..4

1.1 Scope of the document…………………………………………………………………4

1.2 Intended Audience………………………………………………………………………4

1.3 System Overview………………………………………………………………………..4

* Project Briefing……………………………………………………………………………….4
* Problem Statement…………………………………………………………………………..5
* Problem Solution……………………………………………………………………………..5
* Objective of Project………………………………………………………………………….5
* Scope of Project……………………………………………………………………………...5
* Requirements Gathering…………………………………………………………………….6
* Final Screenshot of Project Output………………………………………………………...6
* **Introduction:**

**1.1 Scope of the Document**

* This section will cover details regarding scope of the document
* Low level design document will be at component level i.e., for website portal there will be one LLD

**1.2 Intended Audience**

* This section will cover categories of audiences who will be referring/reviewing this document

**1.3 System Overview**

* This section will capture overview of system application i.e., for what system is being developed
* Who are the stake holders of system?
* What are other external Systems through which this will be interacting
* **Project Briefing:**

In proposed system, my project focuses on detecting phishing URLs using machine learning techniques to enhance cybersecurity measures. Initially, we experimented with logistic regression, which provided a baseline but lacked the complexity to capture intricate patterns in phishing URLs. To address this, we implemented a Random Forest Classifier, which significantly improved accuracy by leveraging an ensemble of decision trees. The primary goal was to develop a robust and scalable system capable of distinguishing between phishing and legitimate URLs with high reliability, ultimately protecting users from online threats.

Data Preprocessing and Feature Extraction

Data preprocessing involved cleaning the dataset by removing duplicates and handling missing values. We then extracted a variety of features from the URLs, including lexical features like URL length and number of special characters, host-based features such as domain age, and content-based features including the presence of forms asking for sensitive information. These features were crucial in capturing the characteristics that differentiate phishing URLs from legitimate ones. We also encoded the categorical labels and split the dataset into training and testing sets to evaluate the model's performance effectively.

Model Training and Evaluation

We trained the Random Forest Classifier on the extracted features, leveraging its ability to handle large datasets and capture complex patterns through multiple decision trees. The model was evaluated on a test set, achieving high accuracy and demonstrating its efficacy in phishing URL detection. Feature importance analysis provided insights into which features were most influential in the classification process. Future improvements could involve incorporating deep learning techniques for even greater accuracy and expanding the feature set with real-time threat intelligence data. This project highlights the potential of machine learning in enhancing cybersecurity and offers a practical solution for mitigating phishing attacks.

* **Problem Statement:**

To create the machine learning based solution to detect phishing Urls.

* **Problem Solution:**

To address this, we developed a phishing URL detection system utilizing a random forest model. This machine learning approach captures complex patterns in URL features, significantly improving detection accuracy. The model is trained on a comprehensive dataset of lexical, host-based, and content-based features, enabling it to effectively distinguish between legitimate and malicious URLs. This solution enhances cybersecurity measures, providing reliable, real-time protection against phishing attacks.

* **Objective of the Project:**

The objective is to create an accurate and efficient phishing URL detection system using a Random Forest Classifier. This involves extracting key features from URLs, training the model, and ensuring it performs well on new data. The project seeks to enhance cybersecurity by providing a tool that can be easily integrated for real-time threat protection.

* **Scope of Project:**

The scope of this project encompasses developing a machine learning-based system for detecting phishing URLs by extracting and analyzing various URL features. It includes implementing a Random Forest Classifier, evaluating its performance, and integrating the solution into existing cybersecurity infrastructures. The project aims to provide a scalable and reliable tool to enhance online security against phishing attacks.

* **Requirements Gathering:**
* Window 10 Operating system
* Visual studio software
* Project integration idea from IEEE website
* Few Github Non copyrighted source codes
* **Final Screenshot of Project Output**



**Fig1. SBI’s original URL**



**Fig2. SBI’s fake URL**