**PHISHING URL DETECTION**

**REPORT**

Name: Podichetty Samiksha

Date: 1/9/25

Github Repository: https://github.com/samikshaa0604/cybersecurityproject

**PHISHING URL DETECTION – REPORT**

## ****Abstract****

## Phishing websites pose a major threat to online security by tricking users into revealing sensitive information such as passwords, banking credentials, and personal data. This project aims to detect phishing URLs using **machine learning techniques**. A dataset of legitimate and phishing URLs was preprocessed and transformed using **TF–IDF vectorization**, which converts URL text into meaningful numerical features. A classification model was then trained to differentiate between phishing and legitimate sites. The solution was integrated with a **Gradio-based interface**, enabling users to input any URL and instantly receive a classification result. This project demonstrates how machine learning can provide a simple, accurate, and user-friendly approach to phishing detection.

## Approach

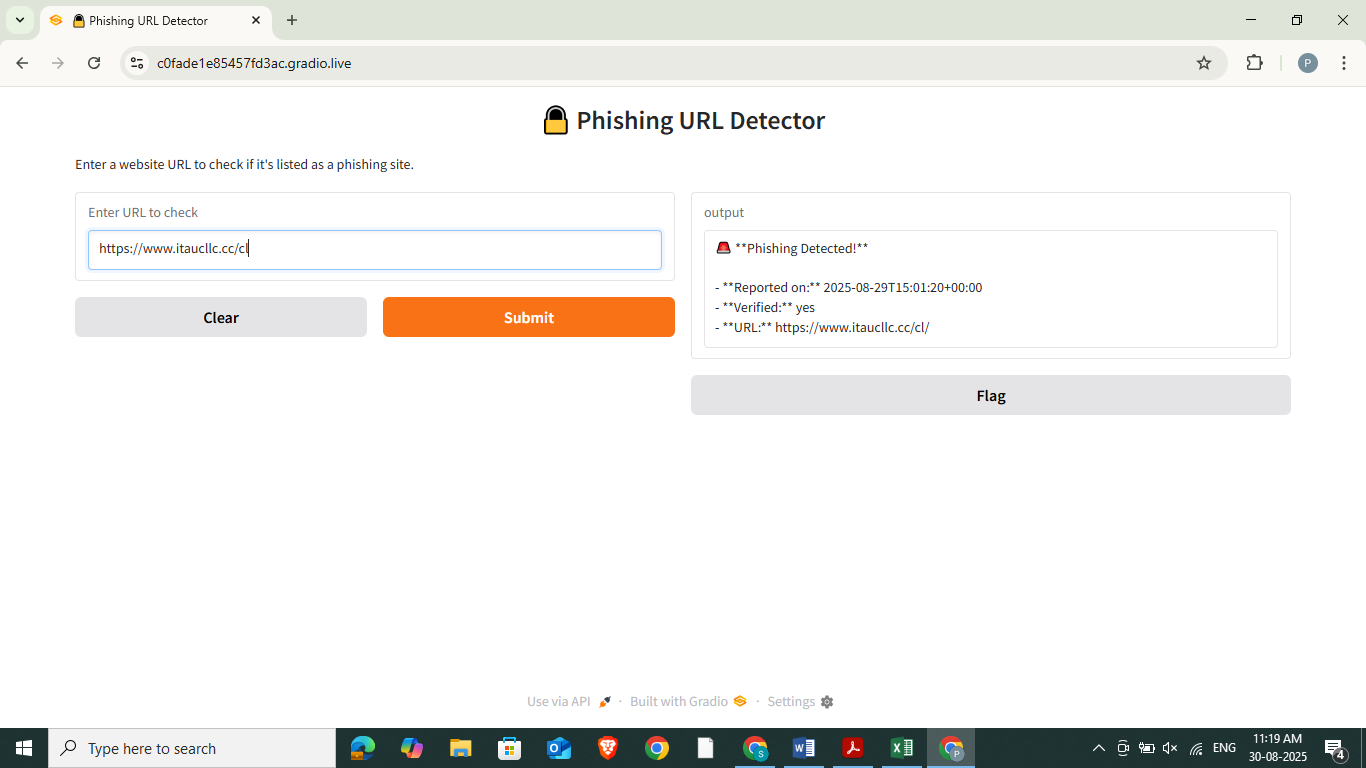
1. **Problem Definition:** The goal was to detect phishing URLs by analyzing their textual patterns and features.
2. **Dataset Preparation:** A dataset containing phishing and legitimate URLs was collected and labeled accordingly.
3. **Feature Extraction:** URLs were vectorized using **TF–IDF .**This technique converts URL text into numerical values, capturing the importance of different terms. For example, suspicious words like “login”, “secure”, or unusual symbols appear more frequently in phishing URLs and are given higher weight.
4. **Model Training:** A supervised machine learning model (e.g., Logistic Regression or Random Forest) was trained on the extracted features to classify URLs into **phishing** or **safe**
5. **User Interface (UI):** The model was connected to a **Gradio-based web interface**, allowing real-time interaction. Users can type or paste a URL into the text box, and the system outputs either “Phishing” or “safe.”

**RESULTS**

* The **Gradio UI** made the system accessible to non-technical users by providing a search-like interface.
* Sample test results:
* https://www.itaucllc.cc/cl → **Phishing**
* https://www.cbit.ac.in/ → **Safe**
* This project proves that combining text-based vectorization (TF–IDF) with machine learning can create an effective phishing detection system.
* The solution emphasizes **awareness and prevention**, enabling users to verify suspicious links before clicking them.

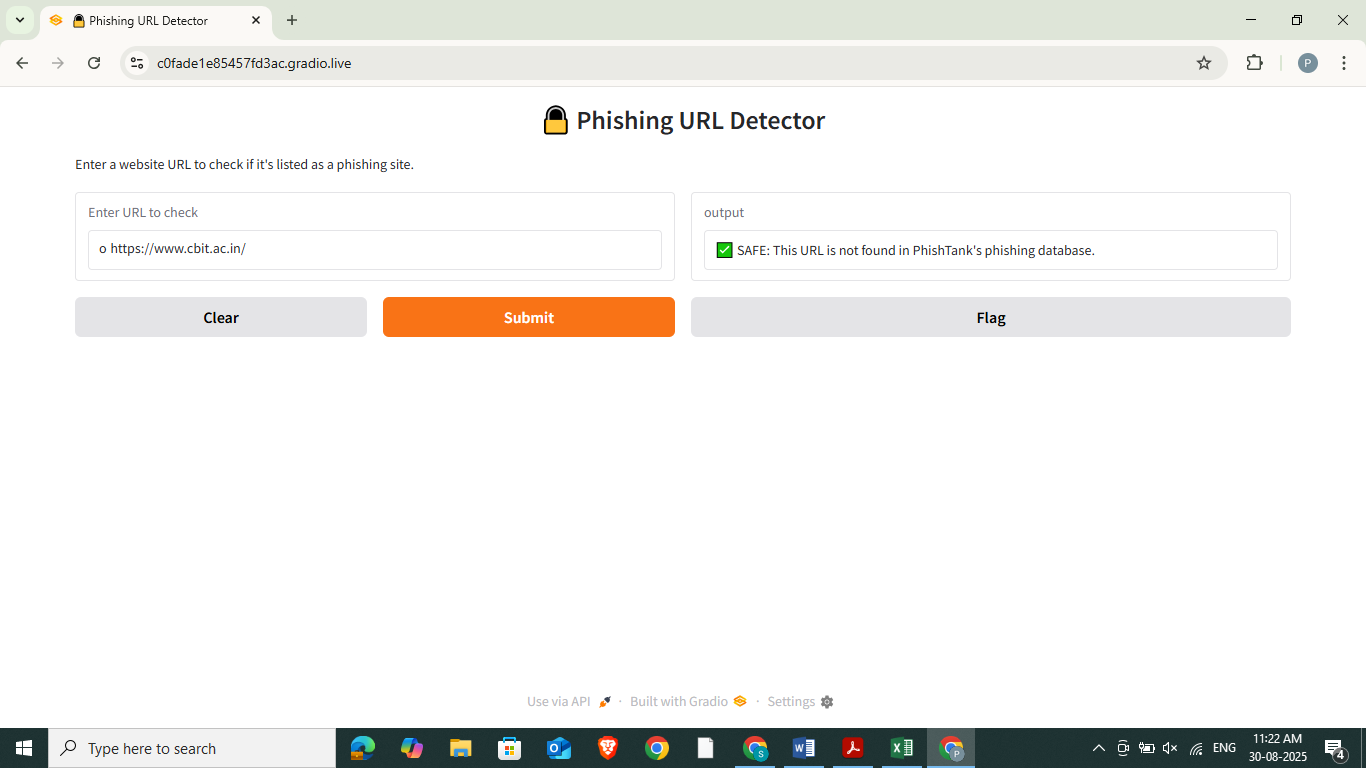
CASE 1:

Shows “Phishing Detected” because the URL was present in the dataset we have taken from PhishTank flagged URLS.

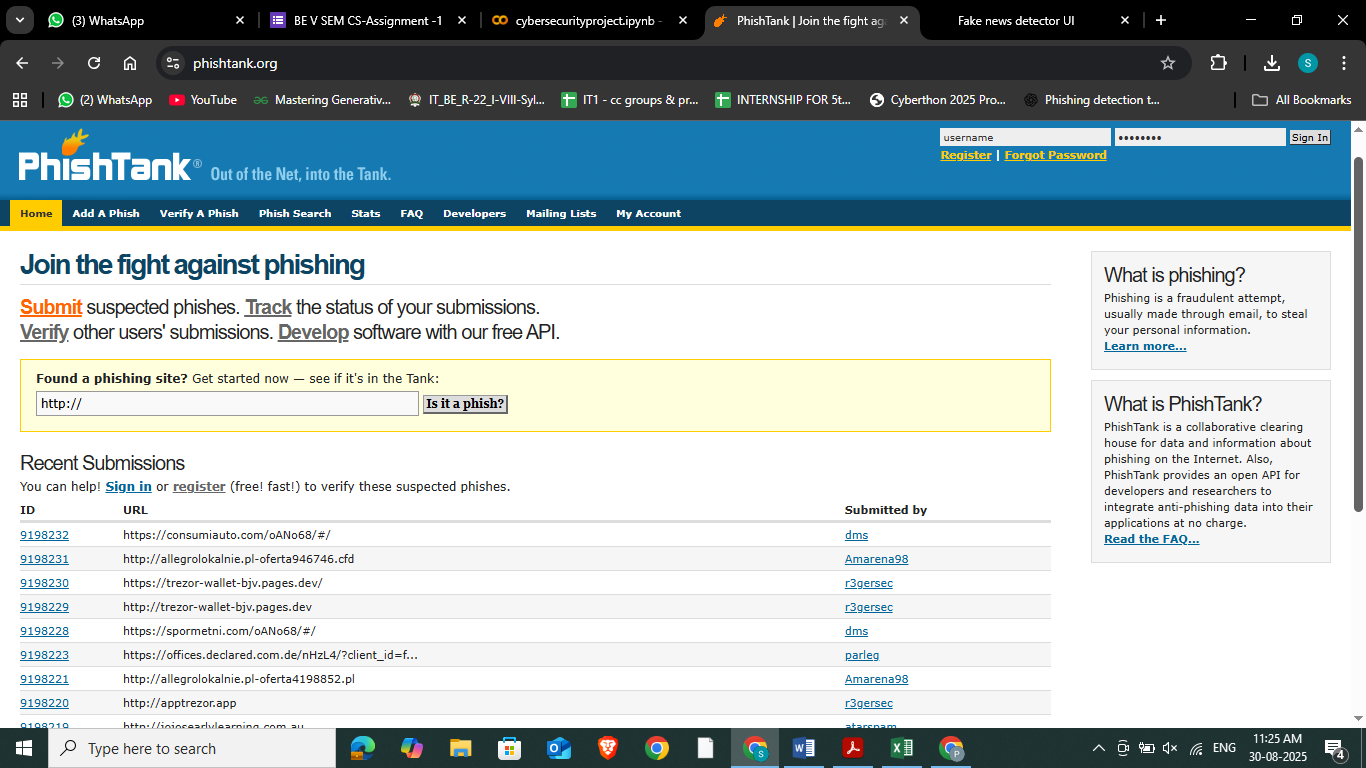


CASE 2:

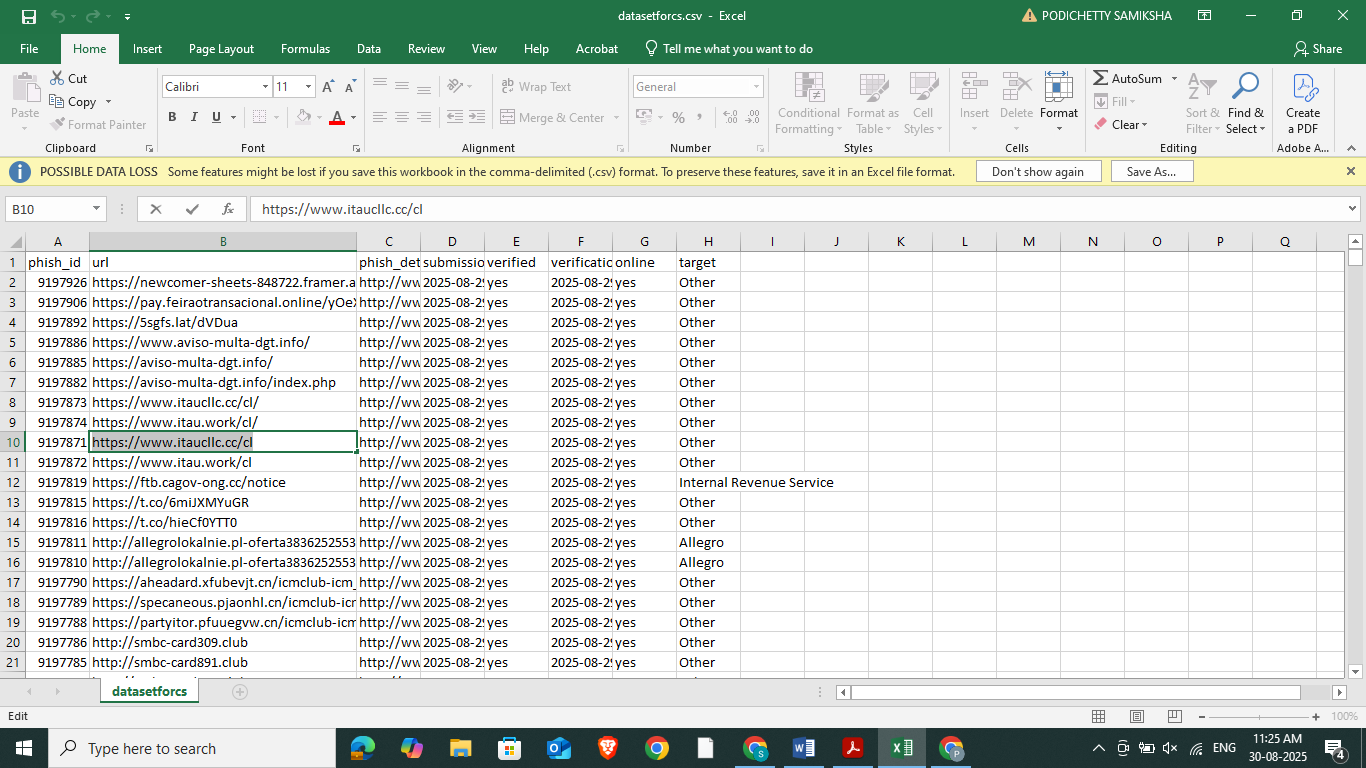
Shows “safe” because the URL hasn’t been flagged on PhishTank.



**The Cybersecurity tool explored:** PhishTank



The data used from the official PhishTank website in the form of csv file.



**CONCLUSION**

This project successfully demonstrated how machine learning, combined with cybersecurity tools, can strengthen phishing detection. By leveraging the **PhishTank dataset** for real-world malicious URLs, applying **ML classification models** for accurate prediction, and integrating the system with **Gradio** for an interactive interface, the solution achieved effective results in classifying phishing and legitimate URLs.