WEB PHISHING DETECTION

RAVITEJA SANGEETHAM

GOKULAVASAN S

SAI TEJA N

SAI KUMAR M

SRIRAMAN M

DOMAIN: DATA SCIENCE

PROBLEM STATEMENT

robust web phishing detection system should aim at identifying malicious websites that attempt to deceive users and steal sensitive information. The challenges involve adapting to evolving phishing techniques, addressing data imbalance, achieving real-time detection capabilities, extracting relevant features, ensuring generalization across various contexts, and safeguarding user privacy. The project's scope covers data collection, feature engineering, model development, real-time detection implementation, performance evaluation, user education, and scalability. The ultimate goal is to create a dependable system that enhances web security by effectively identifying and thwarting phishing threats as users browse the internet.

ABSTRACT

With the rapid expansion of online services and the increasing sophistication of cyberattacks, web phishing has become a significant threat to individuals and organizations alike. This project proposes a comprehensive approach to detect and prevent web phishing attacks through a combination of machine learning techniques and domain analysis. The system employs a multi-layered model that leverages features extracted from URLs, web content, and user behavior to classify websites as legitimate or phishing attempts. By training on a diverse dataset, the model learns to identify subtle patterns and anomalies indicative of phishing activities. Furthermore, the project incorporates real-time domain analysis, checking for characteristics common to phishing domains such as deceptive URLs, SSL certificate inconsistencies, and domain age. This proactive approach enhances the accuracy of phishing detection and reduces false positives. The proposed solution aims to provide a robust defense against evolving phishing tactics, safeguarding users' sensitive information and bolstering cybersecurity in the digital landscape

TECHNOLOGY STACK:

Data Collection: Phishing Website Data, Genuine Website data

Data Preprocessing: Python, Pandas, NumPy, Scikit-learn, Feature engineering

Machine Learning Model: Random Forest Classifier

Model Evaluation: Enhancing the metrics like accuracy, precision, recall, F1-score will assess the model's performance

Web Application : Streamlit

IDE: VS Code, Jupyter Notebook, Pycharm