SYMBIOSIS INSTITUTE OF TECHNOLOGY (SIT)

DEPARTMENT OF ROBOTICS AND AUTOMATION

TITLE: Web Phishing Detection using Machine Learning Algorithms

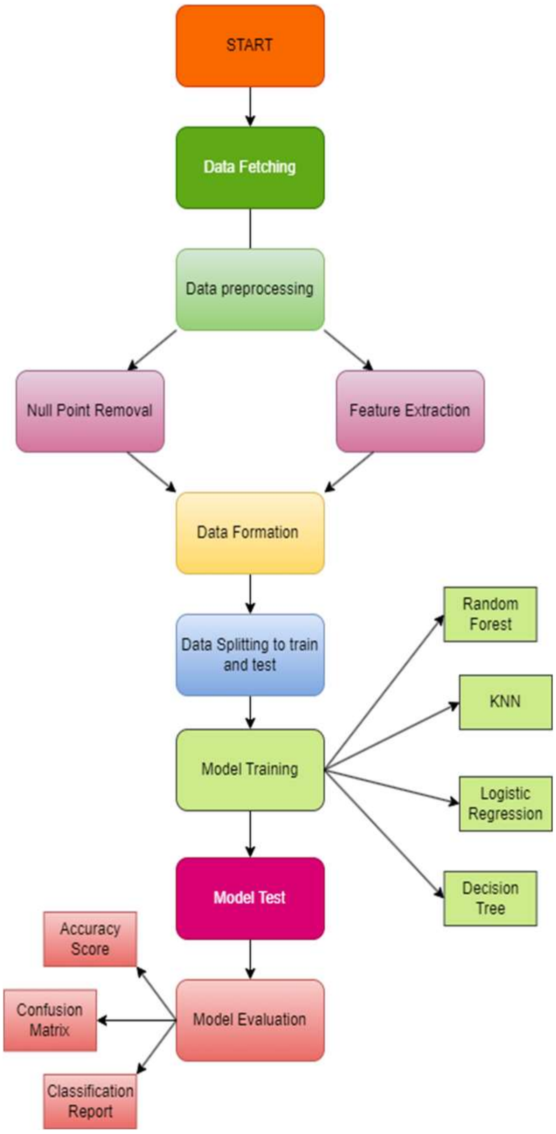
Name of Students: Ansh Sharma, Priyanshu Lathi

INTRODUCTION:

Web phishing poses a significant threat to online security. Leveraging machine learning algorithms enhances our ability to detect and mitigate phishing attacks effectively. This poster explores the application of machine learning techniques in the detection of web phishing incidents.

METHODOLOGY

Diagrams



START

Data Fetching

Data preprocessing

Null Point Removal

Feature Extraction

Data Formation

Data Splitting to train and test

Model Training

Random Forest

KNN

Logistic Regression

Decision Tree

Model Test

Model Evaluation

Accuracy Score

Confusion Matrix

Classification Report

OBJECTIVES & AIMS:

The objective of this study is to develop a machine learning-based system for detecting web phishing attempts. Our aims include identifying key features indicative of phishing websites, evaluating the performance of various machine learning algorithms, and ultimately providing a reliable solution to enhance online security for users by accurately identifying and mitigating phishing threats.

Dataset Description:

The dataset comprises 11,430 instances with 89 features, including URL attributes like length and presence of certain characters, and webpage characteristics such as hyperlinks and domain age. Features like 'ip' and 'phish_hints' help discern potential phishing attempts.

Data Preprocessing:

- Null Point Removal
- Feature Extraction

Model Training & Testing:

- Data Splitting
- Model Training
- Parameter Tuning

RESULTS:

Performance Metrics:

Random Forest: 97% accuracy

K-Nearest Neighbors (KNN): 95% accuracy

Logistic Regression: 92% accuracy

Decision Tree: 96% accuracy

CONCLUSIONS:

In conclusion, this research highlights machine learning's effectiveness in detecting web phishing using a diverse dataset. Models like K-Nearest Neighbors, Logistic Regression, and Random Forest achieved high accuracy, emphasizing the importance of feature selection and data quality. Future advancements and real-time monitoring can bolster cybersecurity against evolving threats

ACKNOWLEDGEMENT: Dr. Sachit T.S