**PROJECT CODE: 2024-P04**

**Title : Phishing-Website-Detection-using-Machine-Learning**

**Abstract**

**Novelty Added in this Project : Implementation of Logistic Regression, Decision Tree and XGBoost is added.**

Phishing websites pose a significant threat to online security, aiming to deceive users into providing sensitive information by masquerading as legitimate entities. This project focuses on detecting phishing websites using machine learning techniques, leveraging a dataset containing 10,000 rows and 18 features. We evaluate the efficacy of various supervised learning algorithms in accurately classifying websites as either phishing or legitimate.

The study employs five distinct machine learning models: **Decision Tree, Logistic Regression, Random Forest Classifiers, XGBoost, and K-Neighbors Classifier**. Each model is trained and tested on the dataset to identify patterns and features indicative of phishing websites. The Decision Tree model provides an intuitive approach by constructing a tree-like structure of decisions based on the dataset features. Logistic Regression, a statistical method, evaluates the probability of a website being phishing by fitting a logistic curve to the data. The Random Forest Classifier, an ensemble learning method, aggregates the decisions of multiple decision trees to enhance prediction accuracy. XGBoost, a scalable and efficient gradient boosting framework, improves prediction performance through iterative training and optimization. The K-Neighbors Classifier classifies websites based on the majority class among the nearest neighbors in the feature space.

Our experiments include preprocessing steps such as handling missing values, feature scaling, and splitting the dataset into training and testing sets. Model performance is assessed using metrics such as accuracy, precision, recall, F1-score, and ROC-AUC score. Comparative analysis reveals the strengths and limitations of each algorithm in the context of phishing website detection.

The findings demonstrate that machine learning models, particularly ensemble methods like Random Forest and XGBoost, are highly effective in identifying phishing websites, offering a robust tool for enhancing cybersecurity measures. This project highlights the potential of machine learning in mitigating phishing attacks and underscores the importance of selecting appropriate algorithms for security applications.