Phishing URL Detection using Machine Learning

Abstract:

This project detects phishing URLs using machine learning. It analyzes URL characteristics and uses a Random Forest Classifier to classify URLs as phishing or legitimate.

Introduction:

Phishing is a common cyberattack where attackers use fake websites to steal user information. This project helps in detecting phishing URLs to improve cybersecurity.

Objectives:

- Collect dataset of phishing and legitimate URLs
- Extract features such as URL length, presence of '@', HTTPS usage
- Train a machine learning model for classification
- Create a Flask web application for real-time URL detection

System Architecture:

1. Data Collection
2. Feature Extraction
3. Model Training
4. Flask Web App for Prediction

Dataset:

The dataset contains URLs labeled as phishing or legitimate, with features like URL length, special characters, and domain properties.

Phishing URL Detection using Machine Learning

Technologies Used:

- Python

- Scikit-learn

- Flask

- Pandas, NumPy

Code Overview:

The model is trained using Random Forest Classifier and saved as 'phishing_model.pkl'.

Flask is used to deploy a web interface where users can enter URLs for detection.

Sample Code:

```
model = RandomForestClassifier()

model.fit(X_train, y_train)

pickle.dump(model, open('phishing_model.pkl', 'wb'))
```

Flask Application:

A web page takes a URL input, extracts features, and predicts whether it's phishing or legitimate.

Conclusion:

The project successfully detects phishing URLs using machine learning and provides a simple web interface for real-time detection.