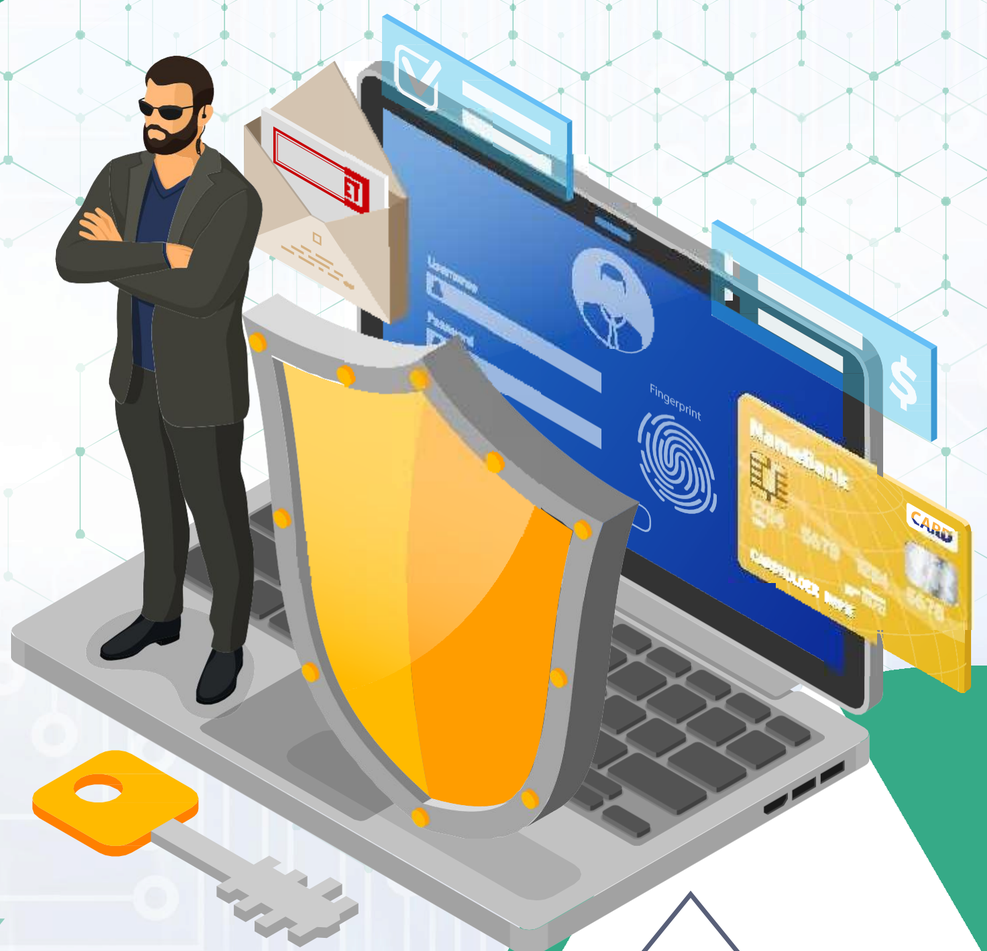
Introduction

- Phishing is a cyber-attack technique used to steal sensitive user data like passwords, credit card numbers, and login credentials.
- This project uses Machine Learning to classify websites as Phishing or Legitimate using URL features.
- A Flask-based web app provides real-time phishing detection for users.

Objective

- Automatically detect fake (phishing) websites.
- Lightweight, fast approach using only URL features.
- Real-time prediction via a Flask web application for demo.



Project Plan: Step-by-step

- ❑ Data collection — demo CSV / Kaggle/UCI dataset.
- ❑ Feature engineering — URL metrics (length, https, keywords)
- ❑ Model training — baseline (Logistic), final (RandomForest)
- ❑ Save model (phish_model.joblib)
- ❑ Build Flask app: /predict endpoint + UI
- ❑ Local deployment — run demo; optional: gunicorn/nginx

Dataset & Features

- Dataset: Demo CSV (URLs + labels). (Recommendation: UCI / Kaggle larger dataset for final).
- Labels: phish / legit.
- URL-based features (fast, no rendering):
  - URL length, hostname length, subdomain count
  - Has HTTPS (yes/no)
  - Digit count, hyphen/special char count
  - Suspicious keywords (login, secure, verify, bank, PayPal)

Model and Checking

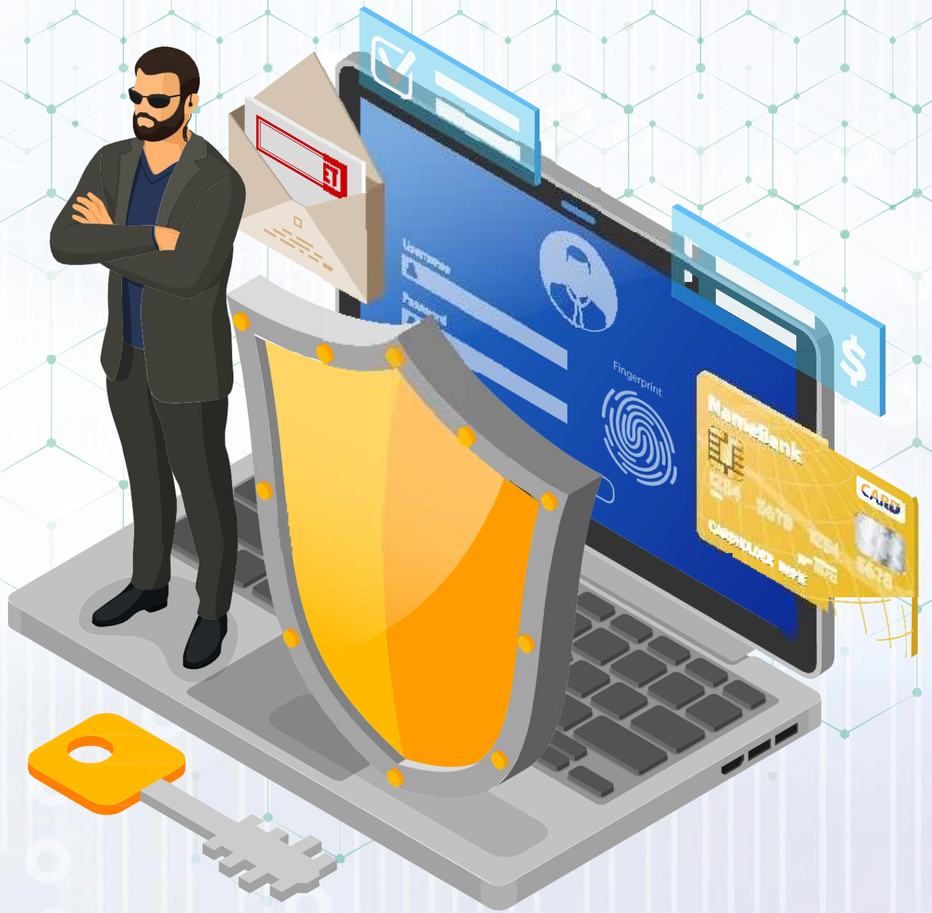
- Tried: Logistic Regression (basic), RandomForest (best)
- Used 80% data for training, 20% for testing
- Checked scores: accuracy, precision, recall
- Random Forest handles complex data without extra scaling.

Deployment & Demo

- Run: python3 app.py
- Open: http://127.0.0.1:5000
- Enter a link, app checks and shows if fake or real
- For real use: add security (rate limit, safety checks), use gunicorn/nginx

Results & Conclusion

- Model accuracy is around 90–95% on test data.
- Example URL paypal-login-update.com detected phishing with 94% confidence.
- URL-based features effectively detect phishing sites quickly.
- Complete ML pipeline built: data, features, model, and web deployment,
- Good project for beginners to learn machine learning and deployment skills.




Thank you