Northeastern University

CS 6140 PROJECT PRESENTATION

PHISHING WEBSITE DETECTION USING MACHINE LEARNING

Radhika Khurana
Nishaant Sitendra Soni
Sahil Subodh Bane

PROBLEM

- Fraudulent websites mimic trusted brands to steal credentials and payments
- High incident rates across banking, e-commerce, and social platforms; short attack lifecycles with outsized financial and privacy risk

WHY IT MATTERS?

- Rule-based filters lag behind attackers’ evolving tactics
- Impact of detection: safer browsing, lower breach risk, better user trust

CHALLENGES

Rapidly evolving attacker behavior which can lead to model drift and evasion

Real world data imbalance introduces bias in the model

High False positive cost damages UX and business operations

Feature similarity complicates accurate classification

OUR APPROACH

Task Framing

Supervised Binary Classification: Phishing (−1) vs Legitimate (+1)

Methods to Incorporate:

- Logistic Regression: Strong baseline, interpretable coefficients
- SVM (RBF): captures nonlinear boundaries
- Random Forest: robust, handles interactions/outliers
- DNN: flexible nonlinear classifier
- Ensemble: combines model strengths for better generalization

- Imbalance handling (if present): class weights and/or SMOTE; stratified splits
- Hyperparameter tuning: LR – regularization; SVM (RBF) – C & gamma; RF – #/depth of trees; DNN – layers/neurons, dropout, learning rate, epochs, etc.
- Evaluation metrics: Accuracy, Precision, Recall, F1, ROC-AUC

DATASET

- Source: UCI Machine Learning Repository
- Samples: 11,055 website records
- Target classes: 2 → phishing = −1 and legitimate = +1
- Features: 30 engineered attributes, all numeric
  - URL-based: "@" presence, use of https, URL length, "//" position
  - Domain-based: domain age, DNS record availability, web-traffic rank
  - Content-based: iFrame usage, redirections, JavaScript behaviors
- Encoding scheme: values in {−1, 0, 1} where −1 = phishing, 0 = suspicious, 1 = legitimate (per feature semantics)
- Pre-processing needs are minimal as there are no missing values and features are already normalized
- Planned post-processing: deployable threshold + feature importance (SHAP)

THANK YOU