Explainability Report: Ensemble Voting Classifiers for Phishing Website Detection

---

Objective

Use ensemble classifiers to enhance prediction performance on phishing websites, while ensuring interpretability using:

- **SHAP (SHapley Additive Explanations)**
- **LIME (Local Interpretable Model-agnostic Explanations)**

---

1. Ensemble Models Overview

Model Variant	Description
Model 1: Hard Voting (LSD)	Combines Logistic Regression, SVM, and Decision Tree with majority rule voting.
Model 2: Soft Voting with Canopy Selection	Adds intelligent feature selection and hyperparameter tuning before combining predictions (probability averaging).

---

2. Feature Selection via Canopy Clustering

- **Distance-based clustering technique** for reducing feature redundancy.
- Canopy centers used as **representative features**.
- Parameters:

- `t1 = 2.0` (loose)
- `t2 = 1.0` (tight)

Selected Features:
Subset chosen by selecting the first feature in each canopy — retained key informative features while reducing dimensionality.

---

3. Hyperparameter Tuning

Grid Search applied for each base learner in **Model 2**:

- **Logistic Regression**:
  - `C`: [0.1, 1.0, 10.0]
  - `solver`: `['lbfgs', 'liblinear']`
- **SVC**:
  - `kernel`: `['linear', 'rbf']`
  - `C`: [0.1, 1.0, 10.0]
  - `gamma`: `['scale', 'auto']`
- **Decision Tree**:
  - `max_depth`: [3, 5, 7]
  - `min_samples_split`: [2, 5, 10]
  - `min_samples_leaf`: [1, 2, 4]

Best Estimators were used in the final soft voting classifier.

---

4. Model Evaluation

Model 1: Hard Voting (All Features)

- **Accuracy**: ~91–93%
- **Confusion Matrix**: Shows solid class separation

Model 2: Soft Voting (Tuned + Canopy Features)

- **Cross-Validation Accuracy**: High consistency across 5 folds
- **Final Test Accuracy**: ~93–95%
- **Precision/Recall/F1**: Well-balanced

---

5. SHAP-Based Explainability

Used SHAP on each base learner across both models.

Logistic Regression (SHAP LinearExplainer)

- **Top Contributors**:
  - `https_token`
  - `statistical_report`
  - `having_at_symbol`
  - `web_traffic`
- **Plot Type**: Bar and summary for global impact

SVM (SHAP KernelExplainer)

- **Captures** non-linear influences.
- **More dispersed contributions** but consistent top features.

Decision Tree (SHAP TreeExplainer)

- **Clear separation of rule-based logic**.
- Visuals show contribution toward classification per rule path.

Insight: SHAP clearly identifies consensus features that each model uses — even if in different ways.

---

6. LIME-Based Local Interpretability

For a randomly selected instance per model:

Model	LIME Top Influencers
Logistic Regression	Clear directional features (e.g., + for phishing)
SVM	Similar to LR, but slightly more nuanced
Decision Tree	Easily traceable to logical splits (feature thresholds)

LIME delivers intuitive, human-readable breakdowns — especially useful for auditability.