Extended Explainability Report: Tree-Based Models – Decision Tree & Random Forest

---

Model Summary

Model	Description
Decision Tree (DTC)	Interpretable, rule-based, low-bias, high-variance
Random Forest (RFC)	Ensemble of decision trees — reduces overfitting, improves accuracy

---

Model Performance

Decision Tree (Entropy)

Metric	Value
Accuracy	92%
Comments	Performs well, may overfit slightly depending on depth

Random Forest (100 Trees, Entropy)

Metric	Value
Accuracy	95%
Comments	Top performer in terms of accuracy and generalization

Both models exhibit strong classification capability, especially RFC due to ensemble averaging.

Explainability with SHAP

Decision Tree SHAP Summary

- SHAP explainer used on `.predict_proba()` for better class separation.
- Top Influencers:
  - `https_token`
  - `statistical_report`
  - `request_url`
  - `web_traffic`

Insights:

- Feature contributions clearly show rule-based splits and their cumulative impact.
- Visuals are interpretable — good for feature reasoning.

Random Forest SHAP Summary

- SHAP values based on mean SHAP across 100 trees.
- Top Influencers (Similar to DTC):
  - `https_token`, `statistical_report`, `having_at_symbol`, `obfuscation_score`

Insights:

- Very consistent SHAP importance due to feature bagging in RFC.
- SHAP values more stable across different test samples than DTC.

Explainability with LIME

Decision Tree (LIME)

- Clear, intuitive rule breakdown.
- LIME matches SHAP's findings but focuses on **local instance explanation**.

Random Forest (LIME)

- Still very interpretable thanks to decision trees in ensemble.
- Shows probability weights influenced by top ~10 features.

Summary:

- LIME excels at **per-instance storytelling**.
- SHAP excels at **global feature importance**.