**Explainability of AIML Decisions**

\*\*Arghyajit Debnath\*\*

Member of AI/ML Intern Team 74

**Overview**

This report highlights the implementation of explainability methods to interpret AI model predictions in the context of financial loss prediction from cybersecurity incidents. Techniques used include SHAP and LIME, which help improve transparency and trust in AI decisions.

**Anomaly Detection**

An Isolation Forest model was used to detect anomalies with 5% contamination.

**Anomaly Detection Results**

Normal : 2850

Anomaly : 150

**Regression Model Comparison**

Four models were compared based on MSE and R². Linear Regression was chosen as the best model.

Linear Regression: MSE=0.98, R2=-0.004

Random Forest : MSE=1.06, R2=-0.085

Gradient Boosting: MSE=1.00, R2=-0.021

XGBoost : MSE=1.27, R2=-0.301

Best Model : Linear Regression

**3. SHAP Explainability**

SHAP was used to evaluate global and local feature impact on model output.

**4. LIME Explainability**

LIME was applied to explain individual predictions using local feature contributions.

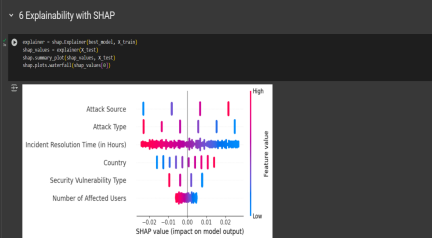
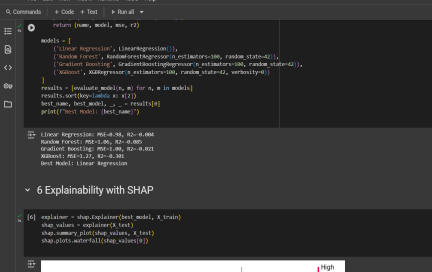
**Saved Artifacts**

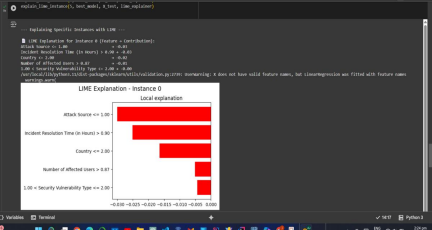
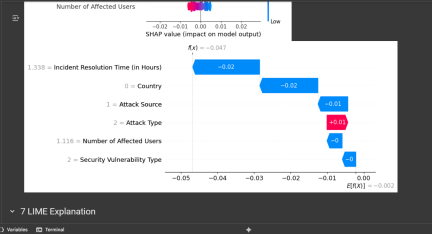
∙ best\_regressor\_model.joblib — trained model ∙ isolation\_forest\_model.joblib — anomaly detection model

∙ label\_encoders.pkl — saved encoders for categorical features ∙ scaler.pkl — saved StandardScaler for reproducibility

**6) SHAP & LIME Visualizations**

****

****

****

****