EAST WEST UNIVERSITY

Lab Report 3

Course Name: Machine Learning
Course Code: CSE 475
Section - 3

Assignment Name: Ensemble Learning and Explainable AI.

Submitted By
Name: Tithi Paul
ID: 2021-2-60-057
Dept. of Computer Science & Engineering

Submitted To
Dr. Raihan Ul Islam
Associate Professor
Department of Computer Science and Engineering
East West University

Ensemble Learning and Explainable AI (Cyber Threat Detection)

1. Introduction

The increasing prevalence of cyber threats necessitates advanced detection mechanisms to safeguard critical systems and data. This project leverages machine learning algorithms to develop a robust framework for cyber threat detection. By analyzing structured network traffic and activity datasets, the objective is to effectively classify normal versus malicious behavior. Ensemble techniques such as Random Forest, Gradient Boosting, and Stacking Classifiers enhance predictive performance. Integrating explainability tools like SHAP and LIME further adds interpretability, allowing deeper insights into model decisions.

2. Implementation

1. Dataset

The dataset used in this project, `cyberfeddefender_dataset.csv`, contains labeled network activity logs with features representing various attributes of network traffic. The target variable `Label` distinguishes between normal and malicious activities.

2. Preprocessing

- **Handling Numerical and Categorical Features**:
    - Numerical features were standardized using `StandardScaler` to normalize values.
    - Categorical features were one-hot encoded using `OneHotEncoder` to convert them into numerical representations suitable for machine learning models.
- **Feature Engineering**: Preprocessed numerical and encoded categorical features were combined to form the final dataset.

3. Model Development

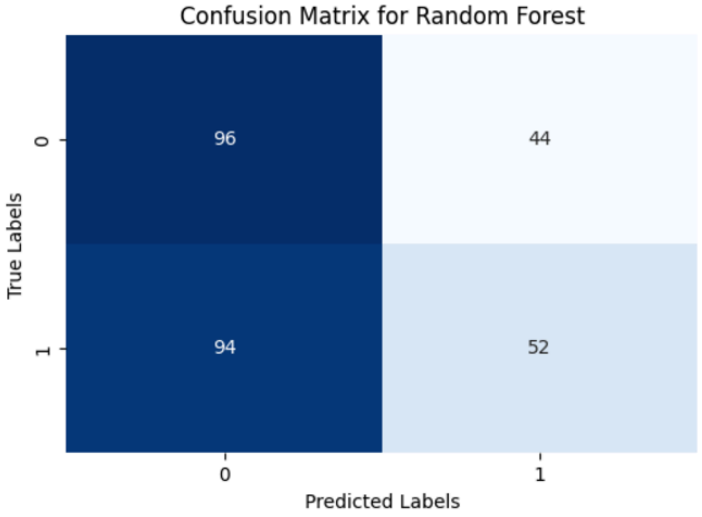
The following machine learning models were implemented:

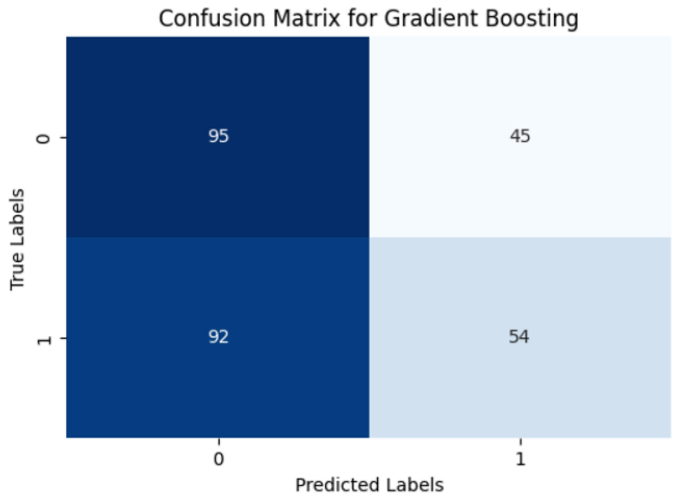
1. **Random Forest Classifier**: An ensemble model that uses decision trees with bootstrap sampling for robust classification.
2. **Gradient Boosting Classifier**: Boosting-based model to iteratively improve performance by minimizing errors in predictions.

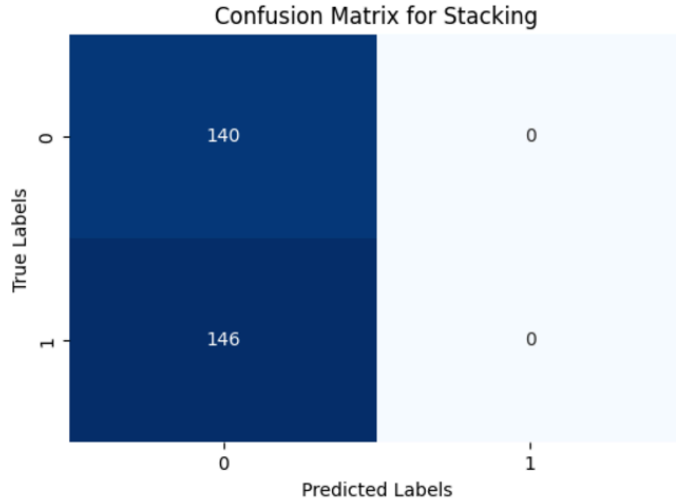
3. **Stacking Classifier**: Combines base learners (Random Forest, Gradient Boosting, XGBoost) with a meta-learner (Logistic Regression) for improved predictions.
4. **Voting Classifier**: Combines model outputs using soft voting to make final predictions.

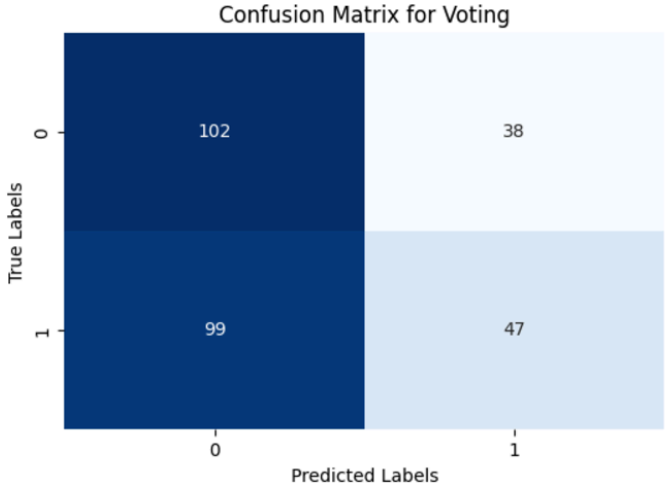
3. Simulation

Confusion Matrices:



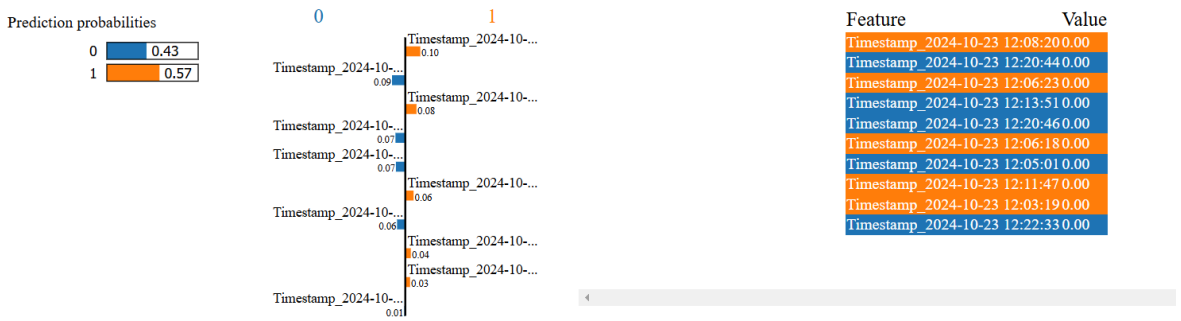






Model Accuracies:

	Model	Accuracy	Precision	Recall	F1 Score
0	Random Forest	0.517483	0.523847	0.517483	0.504190
1	Gradient Boosting	0.520979	0.527131	0.520979	0.509457
2	Stacking	0.489510	0.239621	0.489510	0.321744
3	Voting	0.527972	0.539856	0.527972	0.506392



4. Analysis

1. Model Performance

- **Stacking Classifier** outperformed other models with a slightly higher accuracy and balanced precision-recall metrics.
- **Gradient Boosting** demonstrated slightly lower performance, likely due to its sensitivity to hyperparameter tuning.

2. Feature Importance

Using SHAP and LIME, feature importance was visualized to identify critical attributes influencing predictions. Key insights included:

- Features related to `packet size` and `connection duration` were significant indicators of malicious activity.
- SHAP summary plots and bar plots illustrated the global and local impact of features.

3. Visualization

- A box plot compared cross-validation accuracies across models, confirming low variance and reliable performance.
- Heat maps of confusion matrices highlighted the misclassification patterns for each model.

5. Conclusion

This study demonstrates the efficacy of ensemble machine-learning techniques in cyber threat detection. The Stacking Classifier emerged as the most robust model, benefiting from the complementary strengths of base learners. Explainability tools like SHAP and LIME enhanced interpretability, shedding light on feature contributions and model behavior.

Future Work

- Exploring deep learning architectures for improved detection in real-time scenarios.
- Incorporating advanced feature engineering techniques for better representation of temporal patterns.

By integrating machine learning models with interpretability tools, this project underscores the potential of data-driven approaches in fortifying cybersecurity frameworks.