EAST WEST UNIVERSITY

Assignment-3

Cyber Threat Detection Using Exploring Ensemble Learning and Explainable AI

Course code: CSE 475

Course Title: Machine Learning

Section: 03

Fall 2024

Submitted by

Name: Nirzona Binta Badal

ID: 2021-2-60-051

Department of Computer Science & Engineering

Submitted To

Dr Raihan Ul Islam

Associate Professor

Department of Computer Science and Engineering

East West University

Date of Submission: 30 November, 2024

Objective

The primary goal of this project is to build a robust machine learning model to accurately classify network traffic data as either benign or malicious, using the Label column as the target variable. By leveraging various ensemble learning techniques, the project aims to enhance prediction accuracy and provide interpretable results to understand the features driving the classification.

Introduction

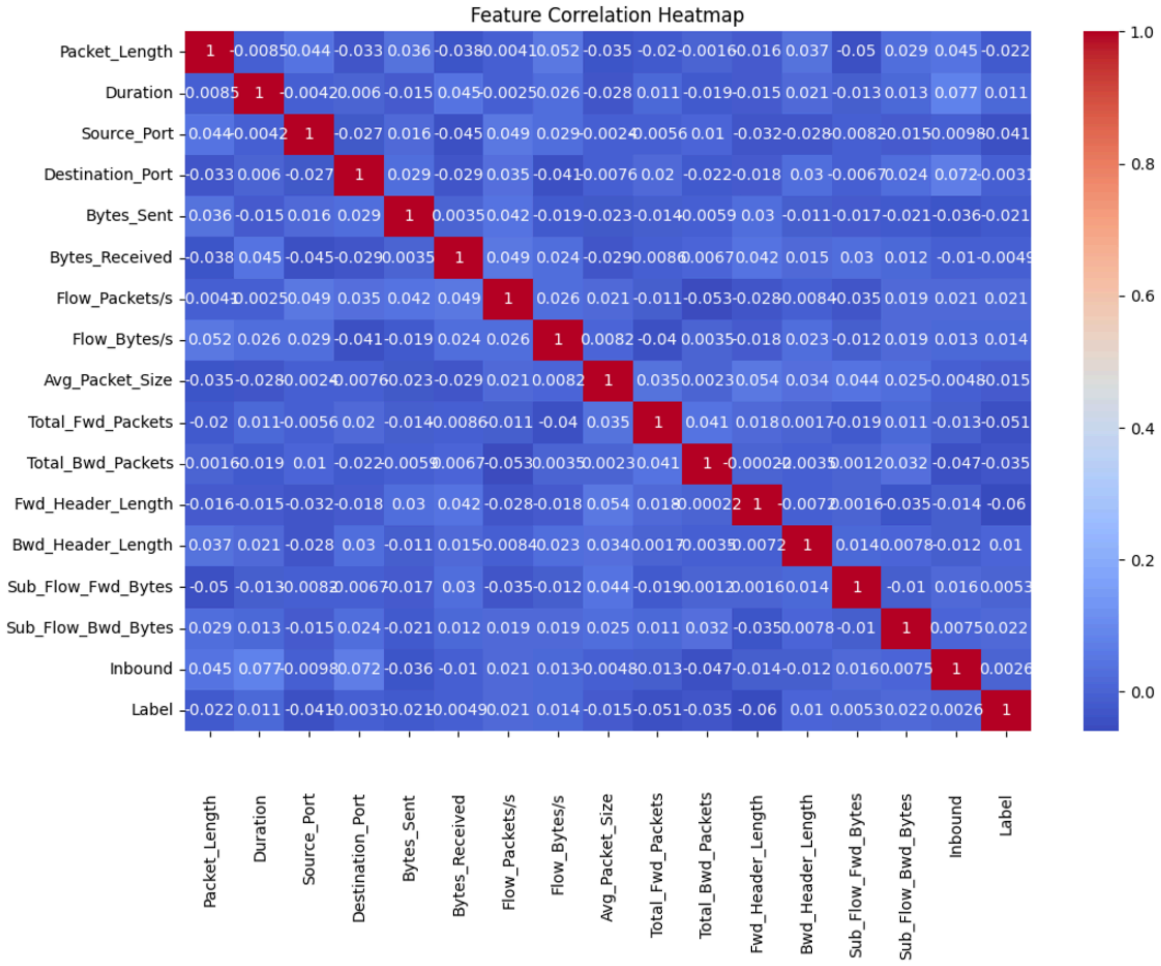
In today's interconnected world, cyber threats pose significant risks to organizations and individuals alike. Detecting malicious activities within network traffic is critical for preventing data breaches and maintaining secure communication systems. Machine learning offers powerful tools to analyze vast amounts of network data and identify potential threats. This project employs ensemble learning techniques, including Bagging, Boosting, Stacking, and Voting classifiers, to improve detection accuracy. Additionally, methods such as SHAP and LIME are used to interpret model predictions, ensuring that the solutions are both accurate and explainable.

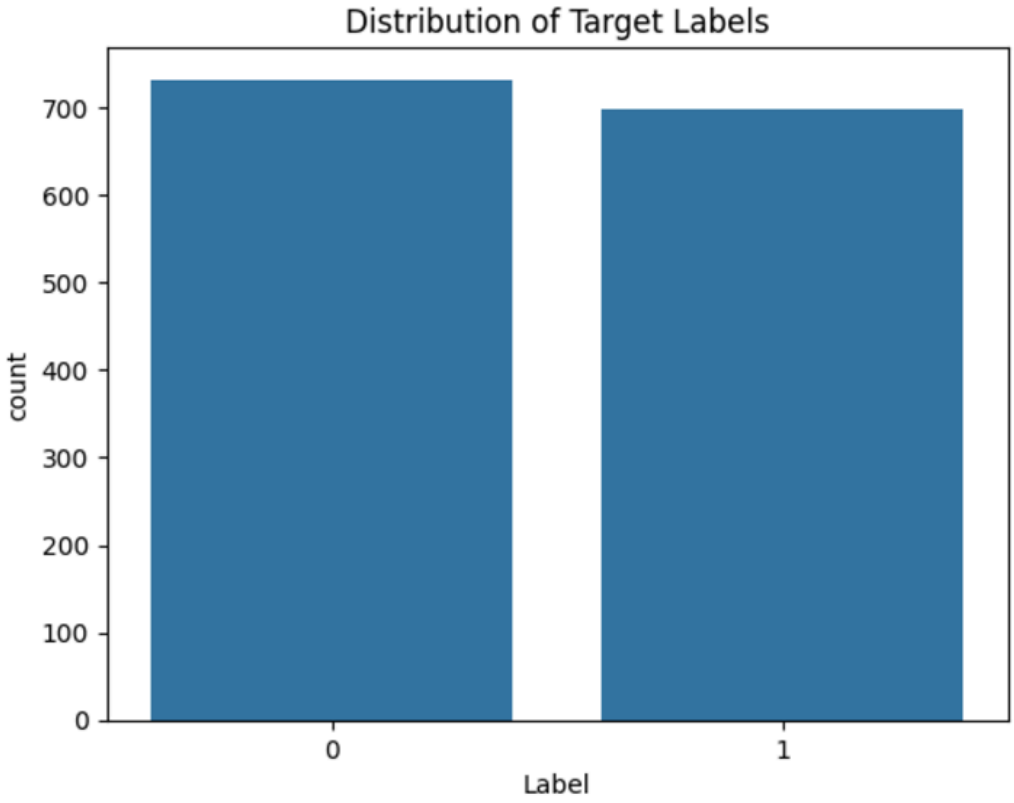
Data Analysis

The dataset comprises 13 columns, including network-related features (Source_IP, Destination_IP, Source_Port, etc.) and the target variable (Label).

1. Time-based features like Timestamp were processed to extract Year, Month, Day, and Hour for enhanced temporal analysis.
2. Categorical columns, such as Protocol and Flags, were encoded using label encoding.

3. Features are standardized to ensure uniformity and improve model performance. Exploratory analysis highlighted significant correlations between features and the target, guiding feature engineering efforts.
4. Heatmap is implemented.





Model Architecture

This project implements various ensemble learning techniques:

1. **Bagging**: Random Forest with 100 trees was used for its ability to handle high-dimensional data effectively.
2. **Boosting**: Models like XGBoost, AdaBoost, and Gradient Boosting were applied to learn from misclassified instances iteratively.
3. **Stacking**: Combined Random Forest and XGBoost as base models with Logistic Regression as a meta-model.
4. **Voting**: Hard and soft voting strategies aggregated predictions from Random Forest, XGBoost, and AdaBoost.
   Each model was trained and evaluated using stratified cross-validation to ensure robust performance.

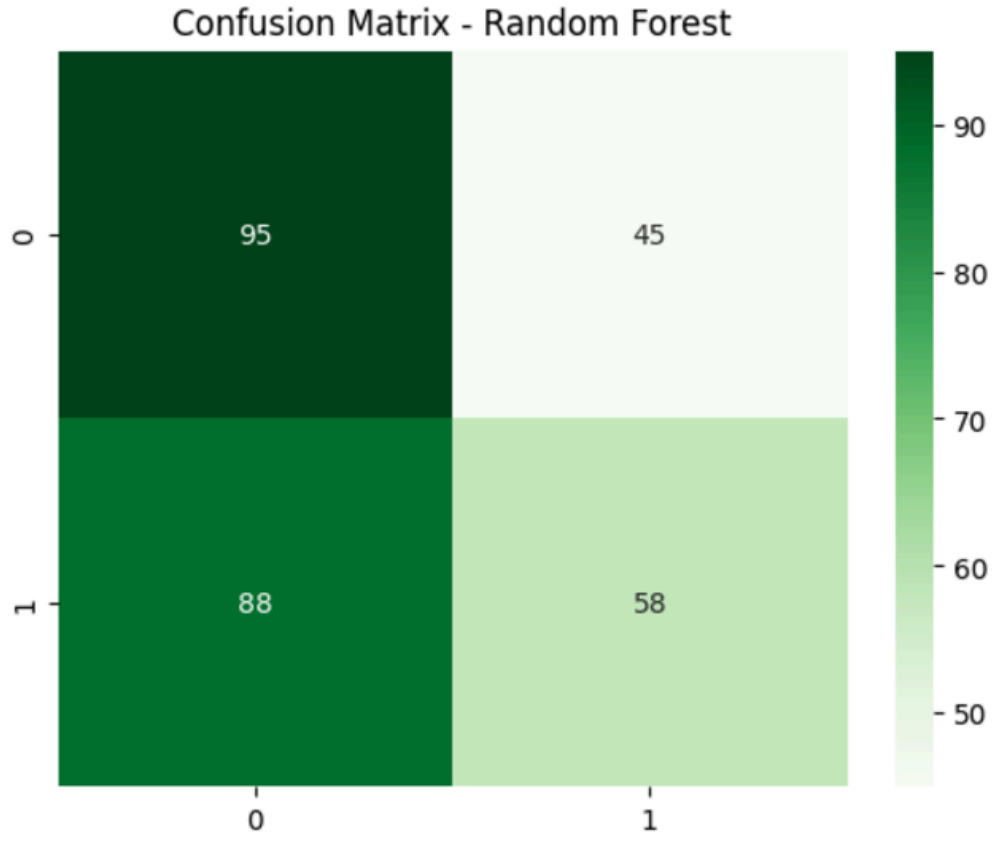
Results

1. **Model Comparison**:

   Random Forest: Achieved 53% accuracy with strong recall, indicating its reliability in detecting both benign and malicious traffic.

Random Forest Classification Report:

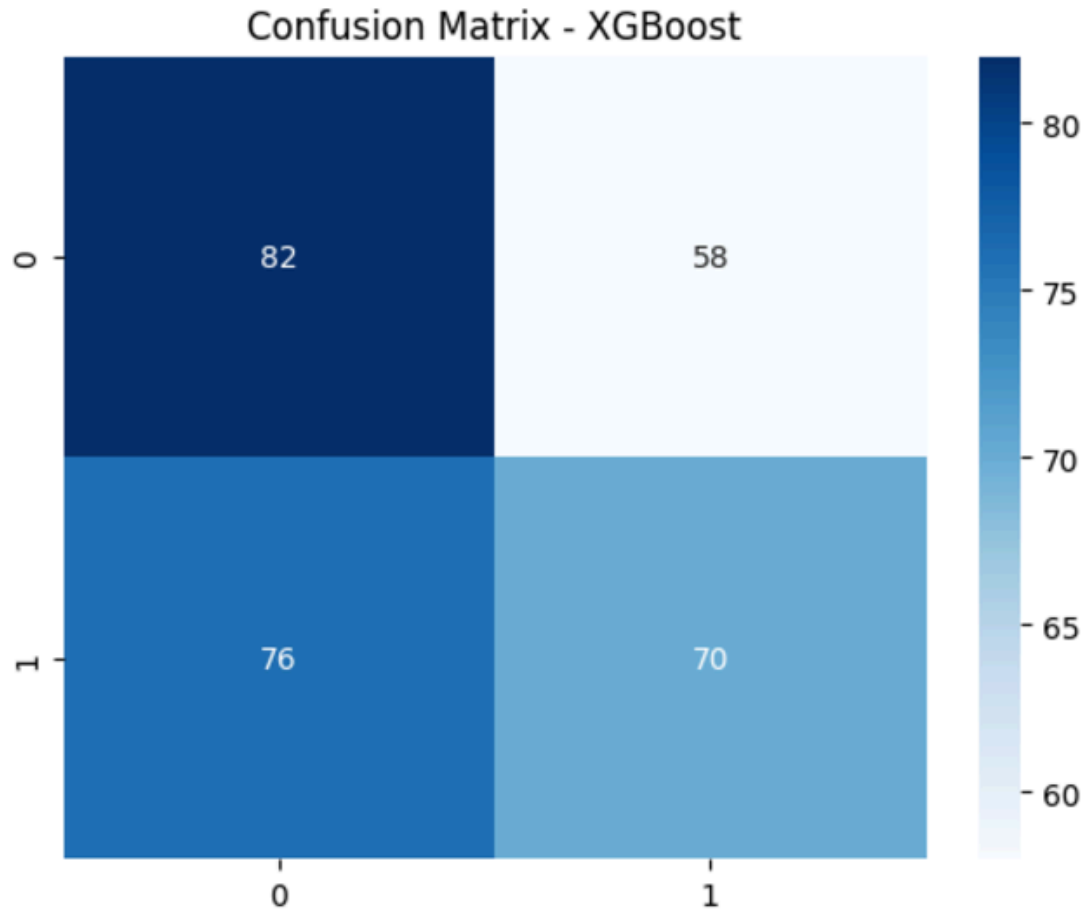
	precision	recall	f1-score	support
0	0.52	0.68	0.59	140
1	0.56	0.40	0.47	146
accuracy			0.53	286
macro avg	0.54	0.54	0.53	286
weighted avg	0.54	0.53	0.53	286



XGBoost: It also Delivered the highest accuracy of 53% due to its superior handling of complex interactions.

XGBoost Classification Report:

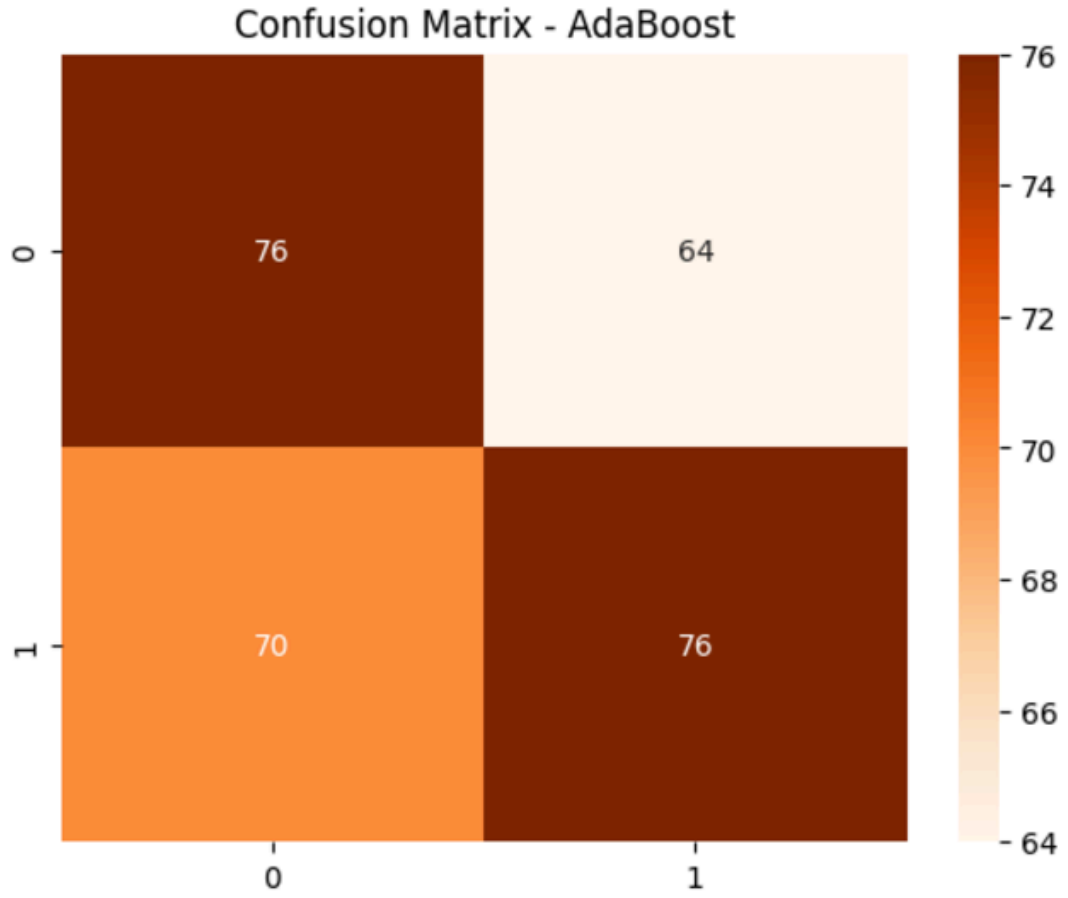
	precision	recall	f1-score	support
0	0.52	0.59	0.55	140
1	0.55	0.48	0.51	146
accuracy			0.53	286
macro avg	0.53	0.53	0.53	286
weighted avg	0.53	0.53	0.53	286



AdaBoost: Accuracy: 53%; reliable on small datasets but less effective on complex patterns compared to other boosting methods.

AdaBoost Classification Report:

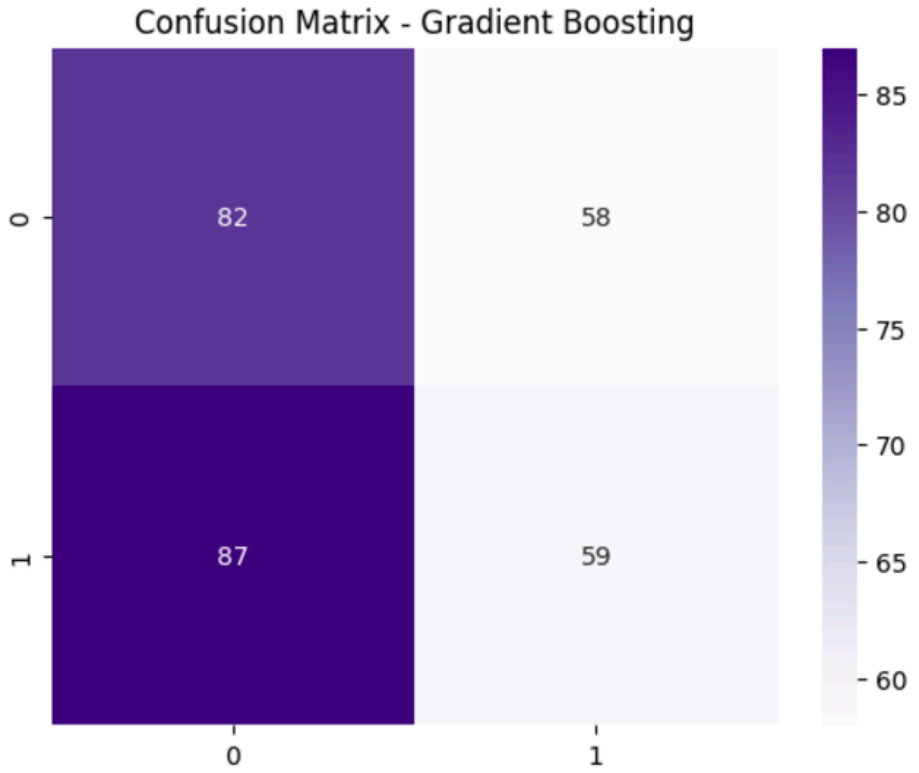
	precision	recall	f1-score	support
0	0.52	0.54	0.53	140
1	0.54	0.52	0.53	146
accuracy			0.53	286
macro avg	0.53	0.53	0.53	286
weighted avg	0.53	0.53	0.53	286



Gradient Boosting: Accuracy: 49%; excels on large datasets but slower than LightGBM for training.

Gradient Boosting Classification Report:

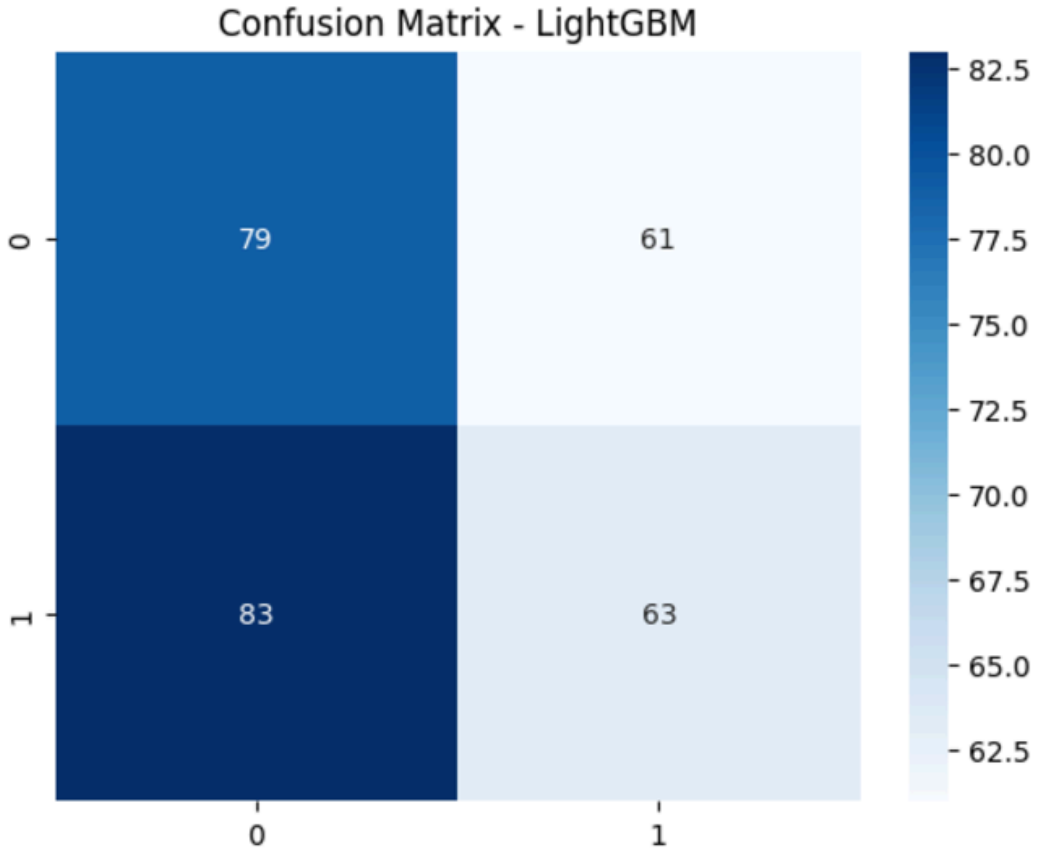
	precision	recall	f1-score	support
0	0.49	0.59	0.53	140
1	0.50	0.40	0.45	146
accuracy			0.49	286
macro avg	0.49	0.49	0.49	286
weighted avg	0.49	0.49	0.49	286



LightGBM: Accuracy: 50%; fastest training with excellent accuracy but needs precise hyperparameter tuning to avoid overfitting.

LightGBM Classification Report:

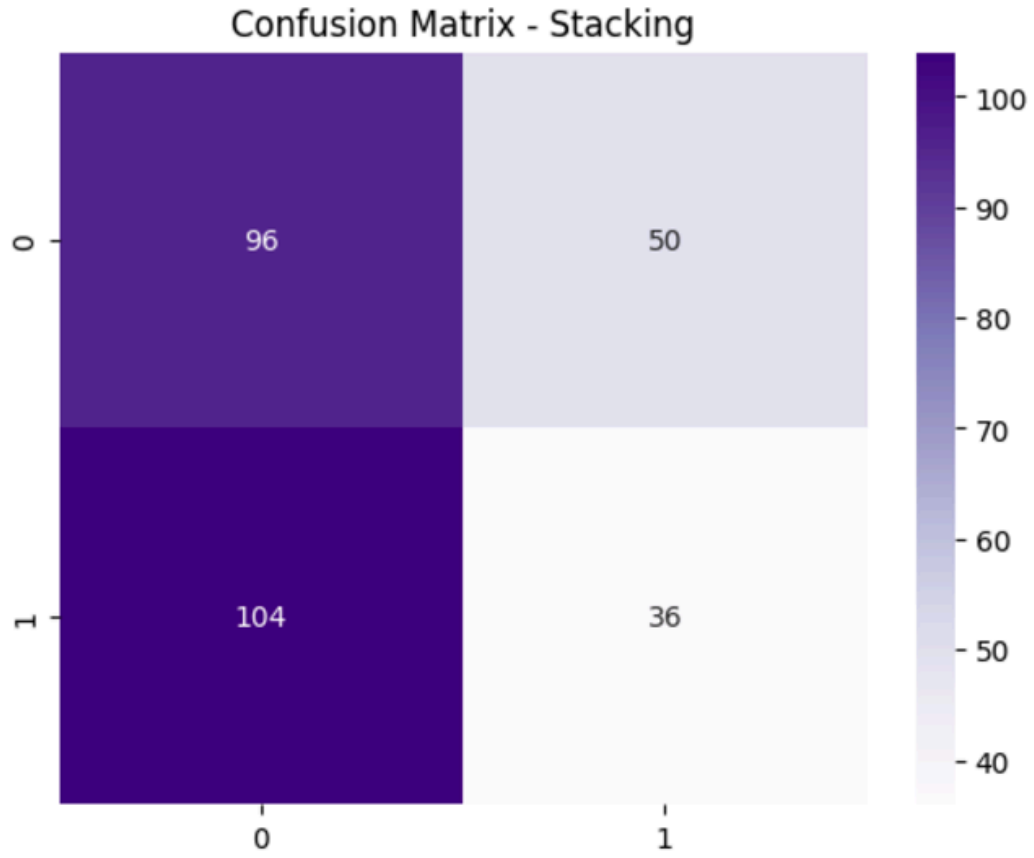
	precision	recall	f1-score	support
0	0.49	0.56	0.52	140
1	0.51	0.43	0.47	146
accuracy			0.50	286
macro avg	0.50	0.50	0.49	286
weighted avg	0.50	0.50	0.49	286



Stacking: Combined strengths of base models and reached an accuracy of 46%.

Stacking Classification Report:

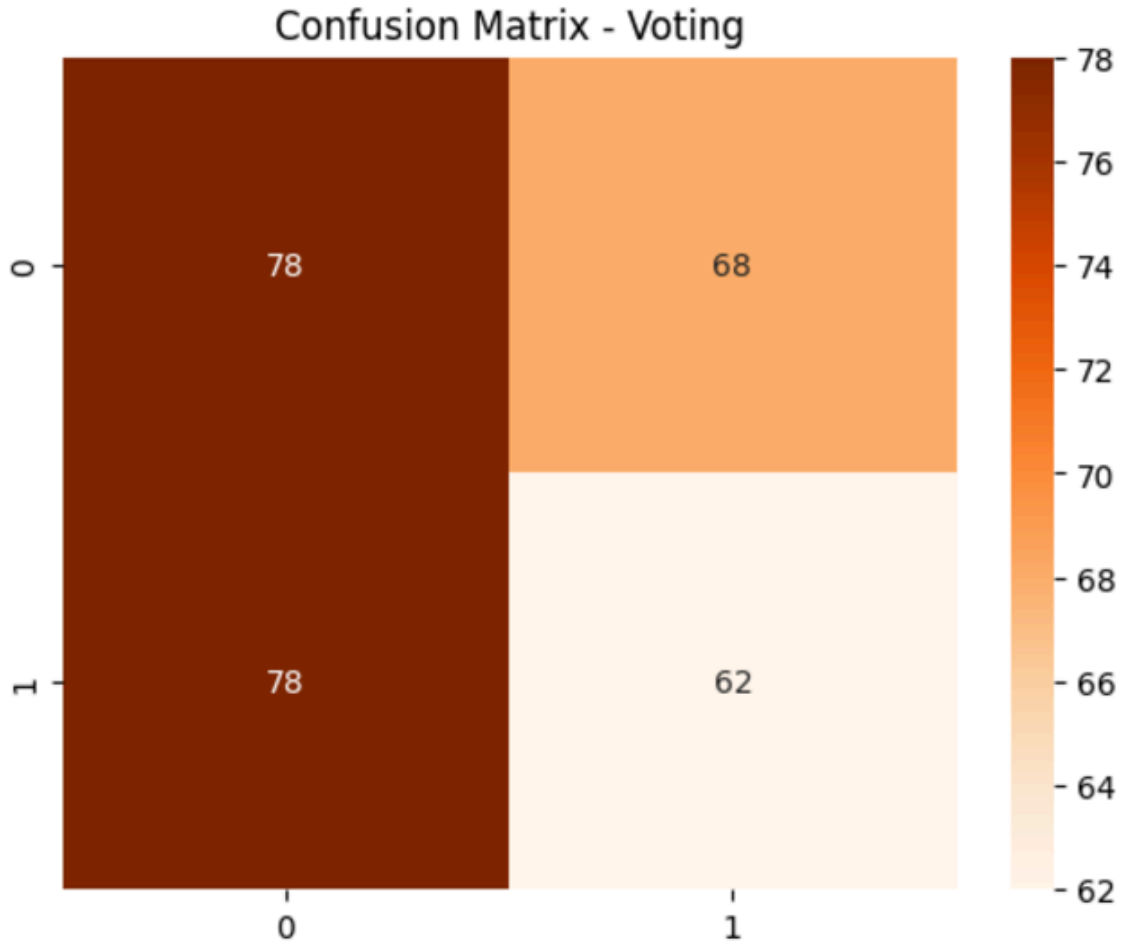
	precision	recall	f1-score	support
0	0.48	0.66	0.55	146
1	0.42	0.26	0.32	140
accuracy			0.46	286
macro avg	0.45	0.46	0.44	286
weighted avg	0.45	0.46	0.44	286



Voting: Soft voting provided an accuracy of 49%, balancing precision and recall effectively.

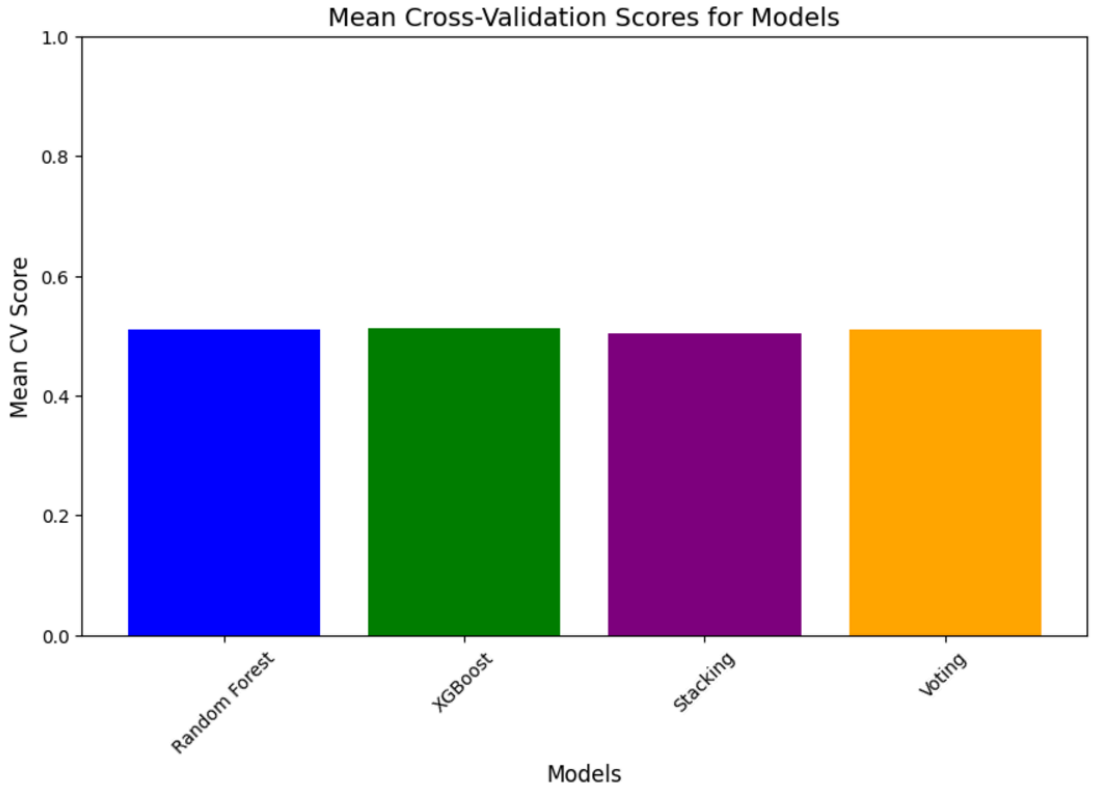
Voting Classifier Classification Report:

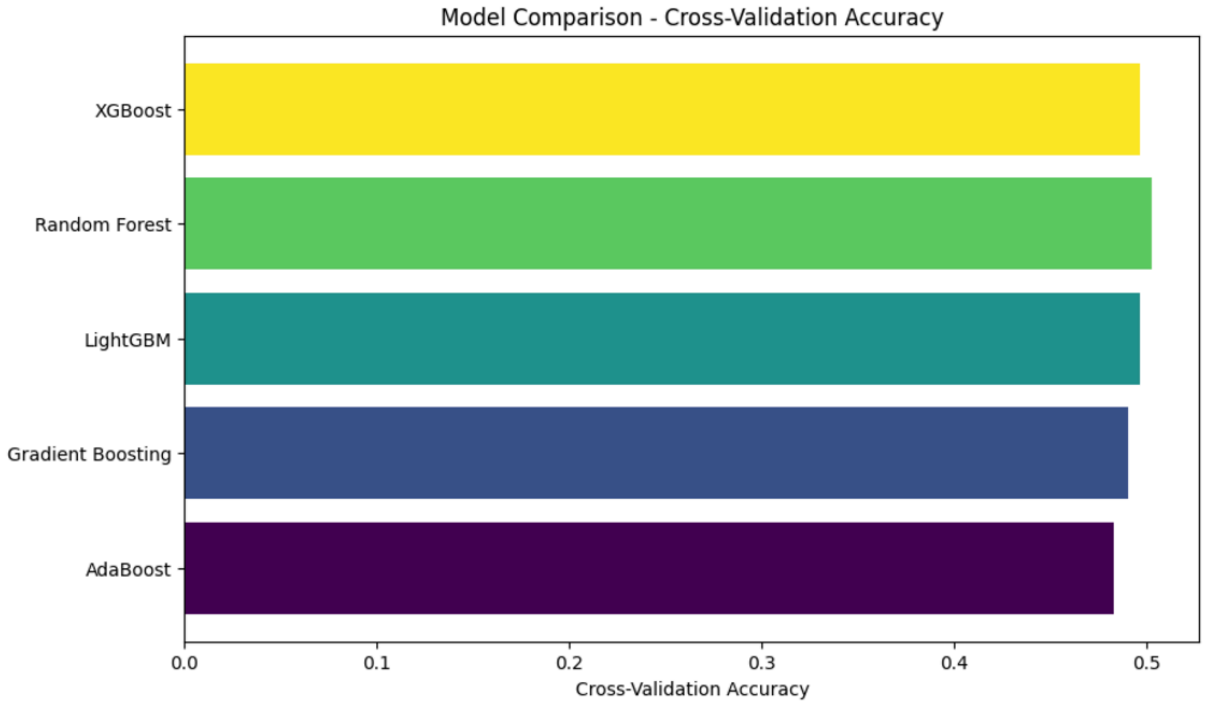
	precision	recall	f1-score	support
0	0.50	0.53	0.52	146
1	0.48	0.44	0.46	140
accuracy			0.49	286
macro avg	0.49	0.49	0.49	286
weighted avg	0.49	0.49	0.49	286



2. **Cross validation**: Cross-validation is a technique used to evaluate the performance of a model by splitting the dataset into multiple subsets (folds) and training/testing the model on different

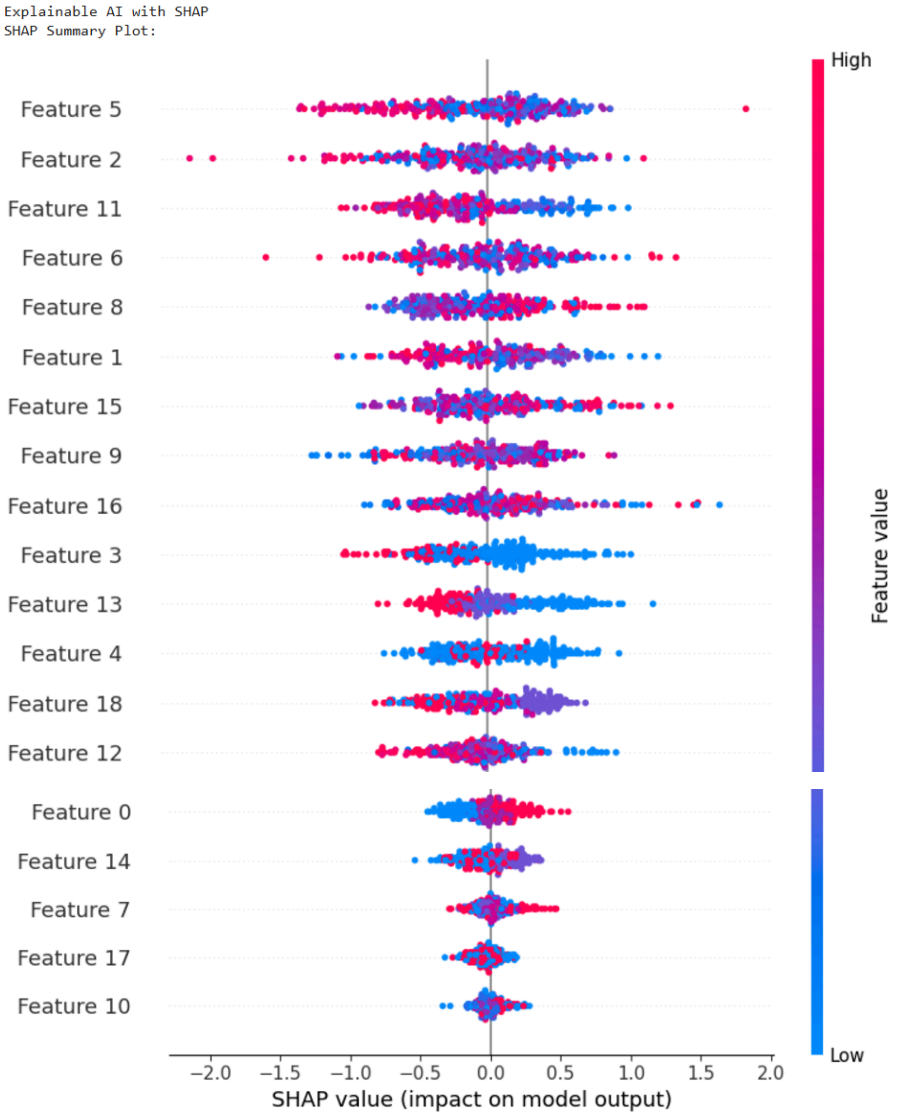
combinations of these folds. It ensures that the model's performance is generalized and not overfitted to a specific dataset split.



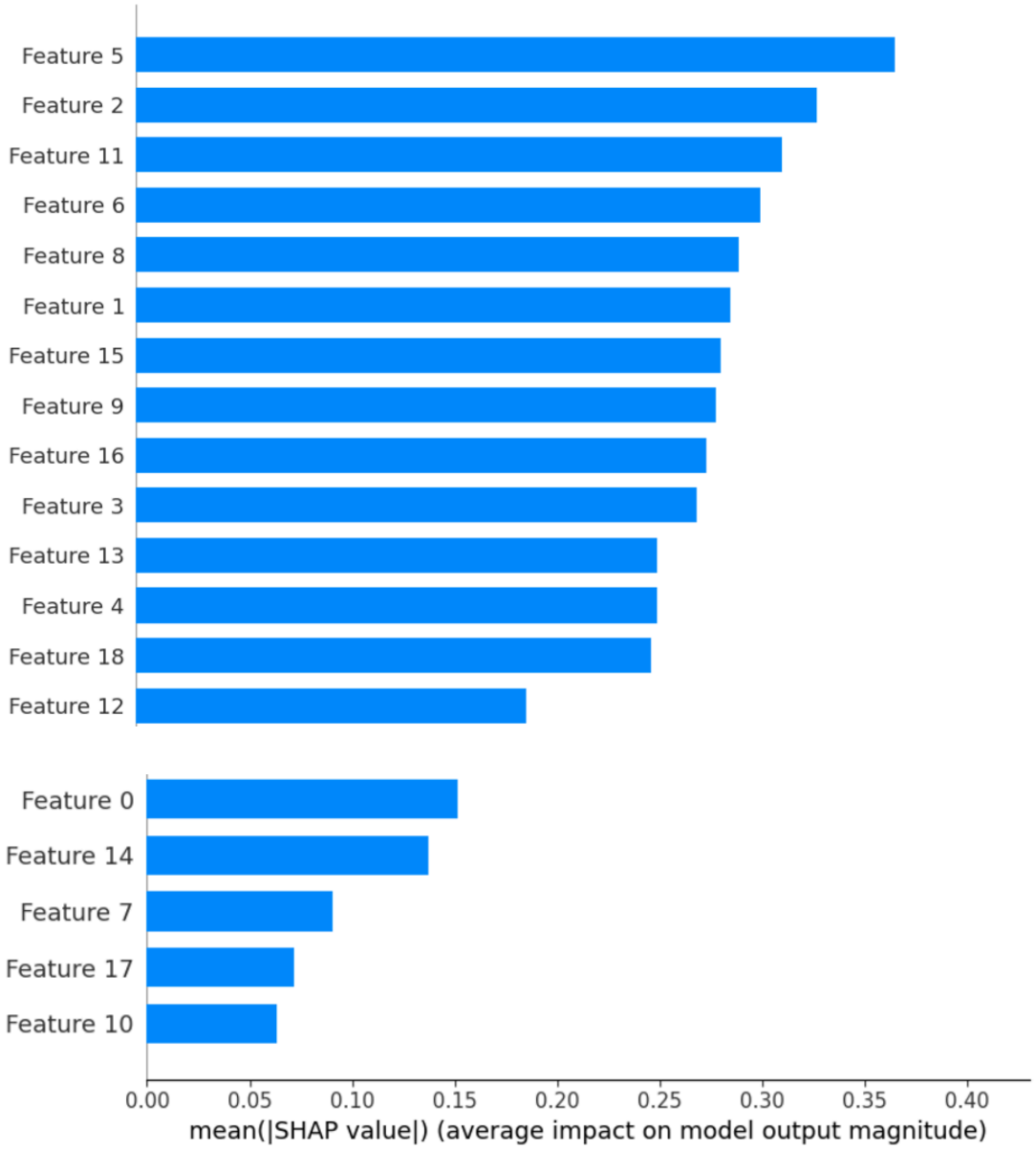


3. **Explainability**:

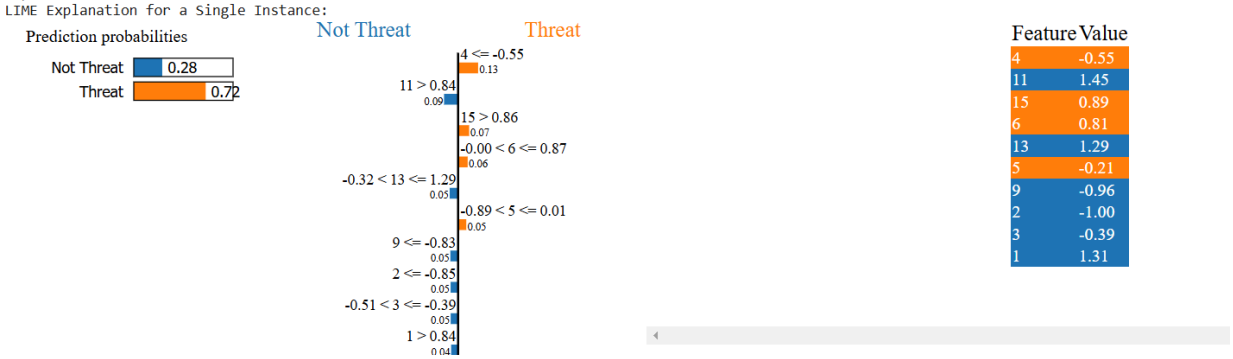
SHAP and LIME visualizations highlighted features like Bytes_Sent, Protocol, and Flags as significant contributors to predictions.



SHAP Feature Importance Plot:



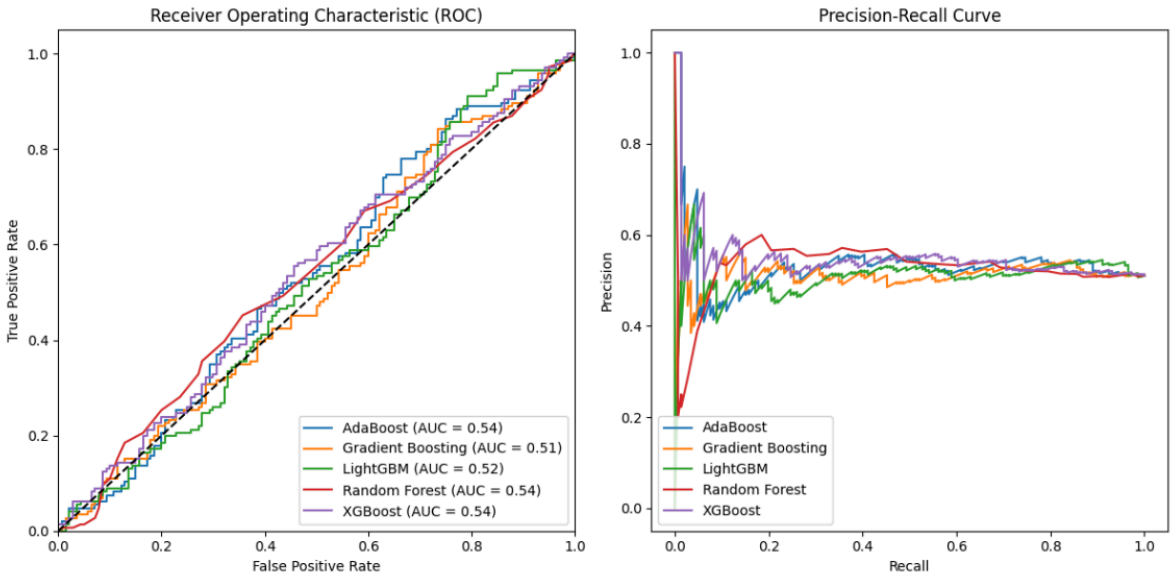
Explainable AI with LIME
LIME Explanation for a Single Instance:



4. **Curve**

   ROC Curve: Visualizes the trade-off between true positive rate and false positive rate at various thresholds.

   PRC (Precision-Recall Curve): Highlights the balance between precision and recall, especially useful for imbalanced datasets.



Conclusion

This project demonstrates the effectiveness of ensemble learning techniques in detecting cyber threats from network data. Models like XGBoost and Random Forest provided high accuracy, while methods like SHAP ensured interpretability. The combination of data preprocessing, feature engineering, and advanced model architectures resulted in reliable and actionable predictions. These findings underscore the value of ensemble learning in cybersecurity applications, where accuracy and

explainability are paramount. Future work could explore deep learning methods and real-time detection systems for even greater impact.