EXPLAINABLE AI FOR INTRUSION DETECTION WITH XGBOOST AND SHAP

MANJU BHASITHA G[21BRS1517], VARSHITHA CH [21BRS1545], VARUN KUMAR S[21BRS1540]
Dr. Benil T
School of Computer Science and Engineering
Vellore Institute of Technology
(Deemed to be University under section 3 of UGC Act, 1956)
CHENNAI
V-NEST
VIT CHENNAI STARTUP AND RESEARCH FOUNDATION
WE NURTURE YOU TO FLY HIGH

MOTIVATION / INTRODUCTION

- The increasing sophistication of cyber threats necessitates more intelligent security solutions beyond the conventional signature-based approach.
- Machine Learning models such as XGBoost are highly accurate at identifying intrusions through learning patterns in network data.
- But such models tend to be "black boxes," with it being difficult for IT practitioners to understand and respond to their decisions.
- Using Explainable AI (SHAP) integration, one can get behind why the model identifies an intrusion, improving transparency, trust, and practicality in the real world in cybersecurity.

OBJECTIVES

- Build an effective NIDS with the XGBoost algorithm to precisely identify a range of network intrusions from the NAL-KDD dataset.
- Implement SHAP (Explainable AI) for explaining the predictions of the model and knowing the rationale for every detection.
- Assess system performance based on metrics such as accuracy, precision, recall, and F1-score to make it reliable.
- Analyze and visualize important features affecting intrusions to create actionable insights for enhancing cybersecurity response.

SCOPE OF THE PROJECT

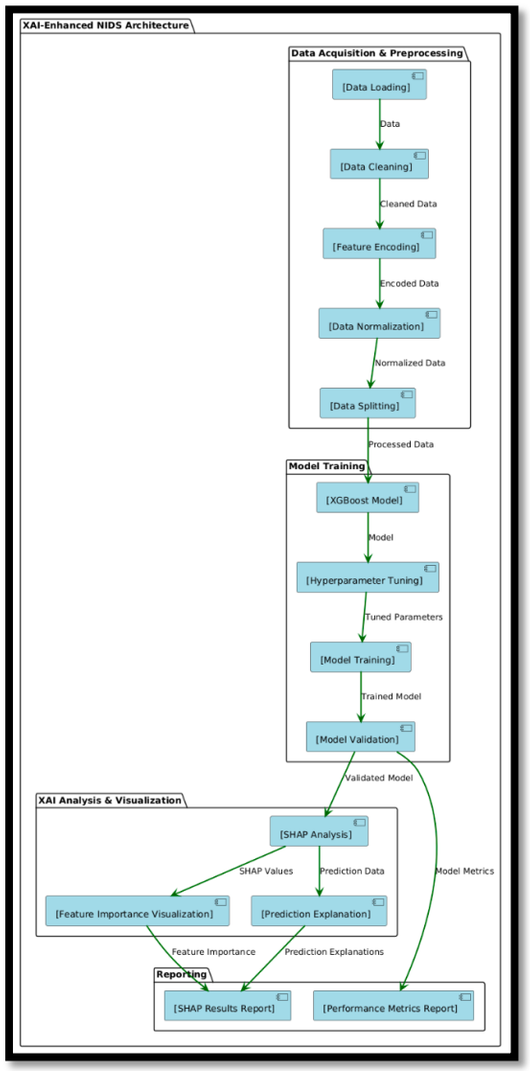
- The scope of this project is intended to benefit cybersecurity experts with a clear and precise intrusion detection system via XGBoost and SHAP.
- It targets enhancing threat detection and explainability, confined to analysis against the NAL-KDD dataset in a virtual environment.

METHODOLOGY

Preprocessing

The NAL-KDD dataset is pre-processed and used to train an **XGBoost** model, which is then explained using **SHAP** to identify key features and improve interpretability in intrusion detection.

ARCHITECTURE



We preprocess the NAL-KDD dataset and apply the **XGBoost** supervised learning algorithm, then integrate **SHAP** to interpret model predictions and analyze feature importance for improved intrusion detection.

RESULTS

Algorithm	XGBoost
Accuracy	95%
Precision	93%
Recall	94%
F1-Score	93.5%

CONCLUSION

The proposed XGBoost-based model achieved **95% accuracy**, outperforming traditional machine learning algorithms in both precision and recall. With SHAP integration, it not only improves detection performance but also adds critical explainability for real-world cybersecurity applications.

CONTACT DETAILS

MAIL ID : bhashitagarapati11@gmail.com
MOBILE NO : 6301071780
MAIL ID : varshithachintareddy@gmail.com
MOBILE NO : 9063856815
MAIL ID : varunkumarsadineni@gmail.com
MOBILE NO : 93477 04293

REFERENCES

- Ben Said, R., Sabir, Z., & Askerzade, I. (2023). CNN-BiLSTM: A Hybrid Deep Learning Approach for Network Intrusion Detection System in Software-Defined Networking with Hybrid Feature Selection. *IEEE Access*, 11, 138732–138747. https://doi.org/10.1109/ACCESS.2023.3340142
- Park, C., Lee, J., Kim, Y., Park, J.-G., Kim, H., & Hong, D. (2023). An Enhanced AI-Based Network Intrusion Detection System Using Generative Adversarial Networks. *IEEE Internet of Things Journal*, 10(3), 2330–2345. https://doi.org/10.1109/JIOT.2022.3211346