BATCH 13
PURUSHOTHAMAN V
ROSHAN A

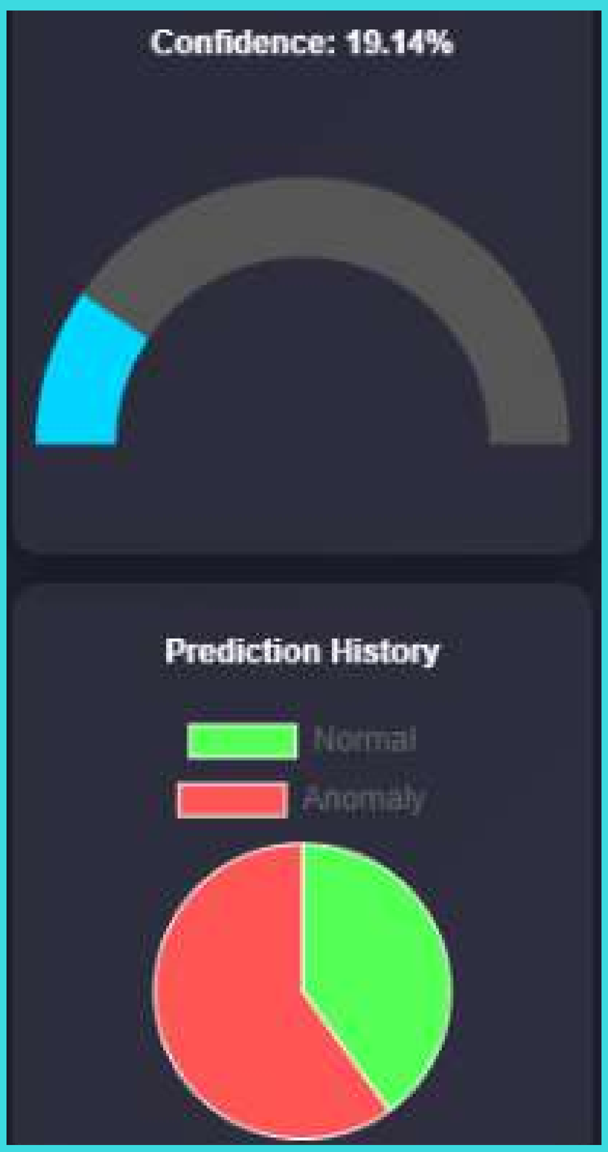
APRIL 2025

AUTOMATED THREAT DETECTION USING MACHINE LEARNING IN PYTHON

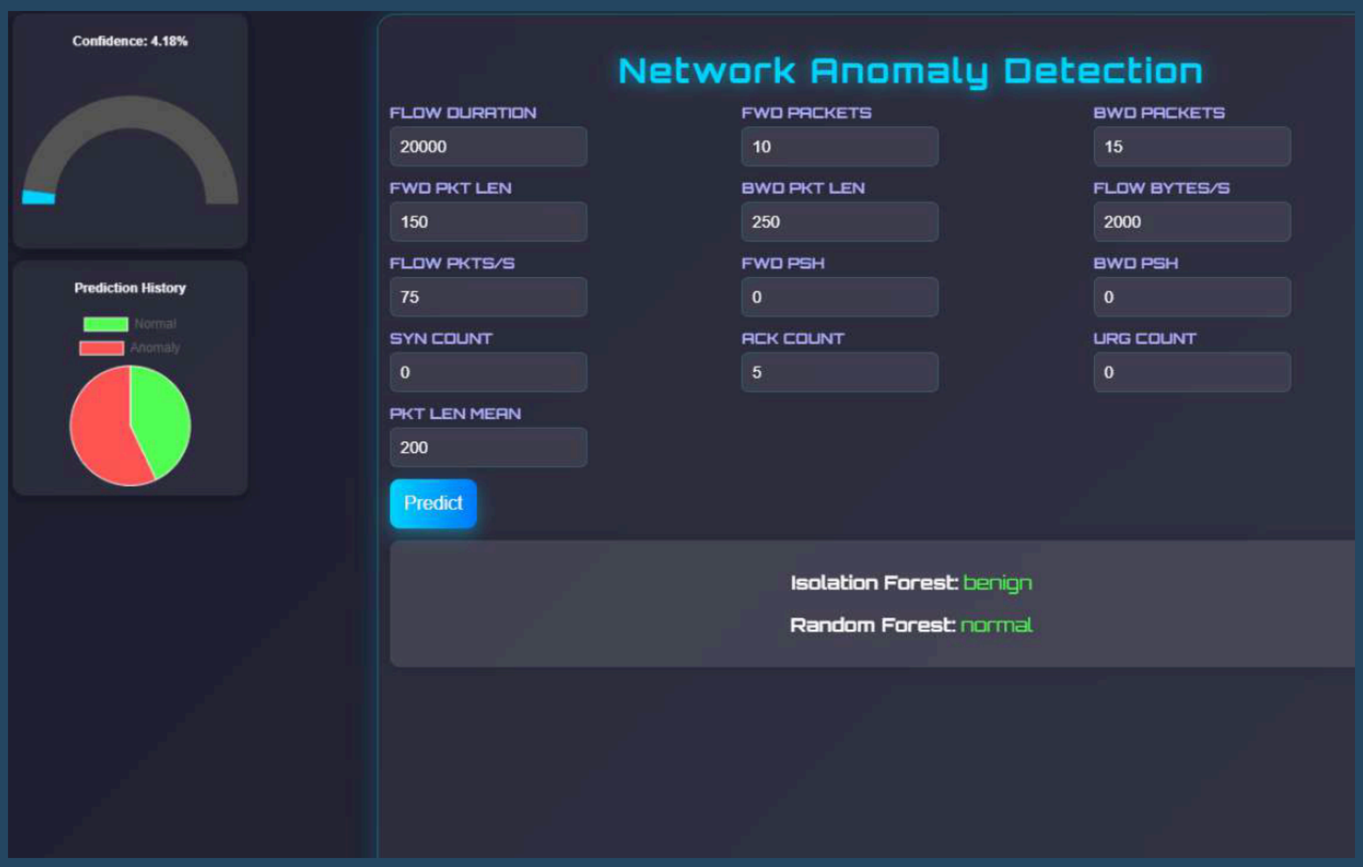
ABSTRACT

A hybrid machine learning system employing Random Forest and Isolation Forest models for real-time, scalable cyberthreat detection using PCAP analysis.

KEY EVIDENCE



- Hybrid ML detects threats: 98% accuracy, 20% FPR, CICIDS2017.
- Real-time PCAP analysis: ~1000 flows/second, scalable Flask API.
- Feedback loop adapts models, reduces errors, enhances cybersecurity.



Network Anomaly Detection

FLOW DURATION: 20000
FWD PACKETS: 10
BWD PACKETS: 15
FWD PKT LEN: 150
BWD PKT LEN: 250
FLOW BYTES/S: 2000
FLOW PKTS/S: 75
FWD PSH: 0
BWD PSH: 0
SYN COUNT: 0
ACK COUNT: 5
URG COUNT: 0
PKT LEN MEAN: 200

Isolation Forest: benign
Random Forest: normal

OBJECTIVES

Develop a real-time machine learning system for automated network threat detection.

Improve accuracy and minimize false positives using hybrid Random Forest and Isolation Forest models.

Implement a feedback loop for continuous model updates against evolving cyber threats.

METHODOLOGY

- Data Collection and Preprocessing: Capture PCAP logs, clean data, and extract important network features.
- Model Training: Train Random Forest and Isolation Forest models on labeled and unlabeled traffic data.
- Real-Time Detection: Deploy models via Flask API for real-time network traffic prediction and alerting.

REFERENCES

- Maseer et al. (2021): Traditional IDS struggle with cloud and IoT traffic scalability.
- Mbona et al. (2022): Apache Spark enables real-time detection but suffers from latency .
- Akgun, D. Deep learning for DDoS intrusion detection in cybersecurity.

CONCLUSION

The ATD system revolutionizes cybersecurity using hybrid ML, achieving 98% accuracy and scalability. Despite limitations, it offers robust protection, with future enhancements promising greater resilience in dynamic digital environments.