CYBER SECURITYASSIGNMENT-2

REPORT

Name : BASSA SRILAKSHMI

RollNo:1601-23-737-005

Date :05/10/2025

Github Repository: https://github.com/BASSASRILAKSHMI/CS-ASSIGNMENT-2
Research Paper: Enhancing Cyber Security Through Predictive Analytics: Real-Time Threat Detection and Response
https://arxiv.org/abs/2407.10864

INTRODUCTION

Modern cybersecurity is hampered by traditional, reactive Intrusion Detection Systems (IDS) that fail against zero-day attacks and produce high false positives. This project implements a proactive, multi-class Network Intrusion Detection System (NIDS) using Machine Learning (ML). We selected and optimized the efficient Random Forest classifier on the CICIDS2017 dataset. The goal is to shift network defense from remedial measures to high-accuracy, real-time predictive analytics

RESEARCH GAP

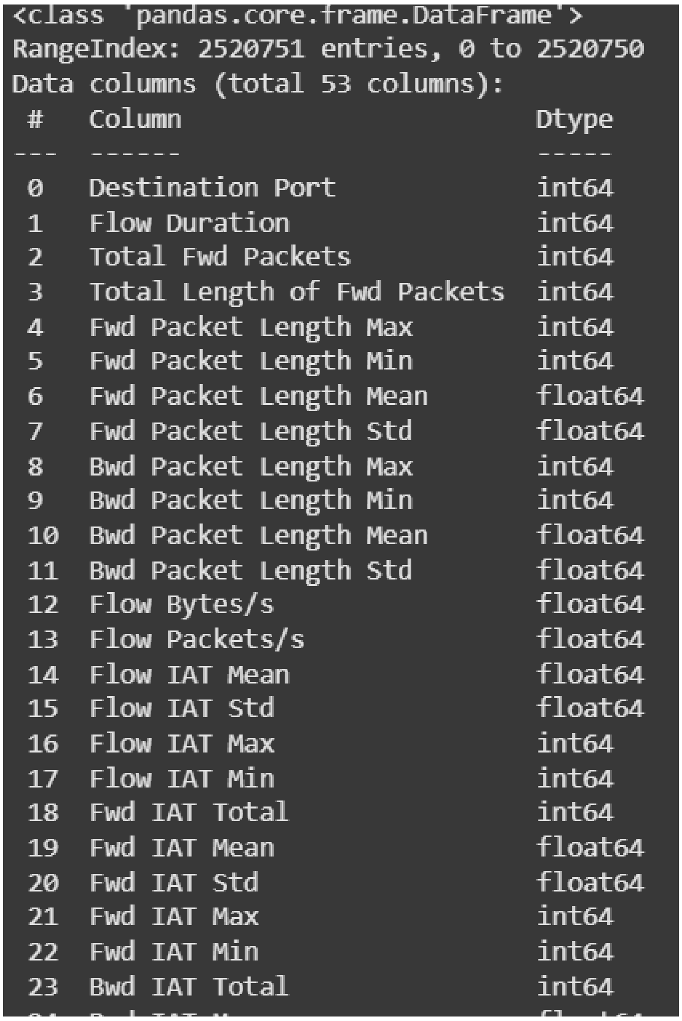
The primary gap is the lack of systems that combine highly accurate, multi-class threat classification with operational efficiency and an automated response capability. Traditional systems struggle with generalizing accurately to diverse, unseen attack patterns. This project fills the gap by optimizing the Random Forest model for efficiency and integrating its output with a simulated, tangible, real-time alert/response action (email alerts)

METHODOLOGY

We utilized a 200,000-row subset of the real-world CICIDS2017 network traffic dataset. Data preprocessing involved using Label Encoder for attack types and StandardScaler for feature normalization . The Random Forest model was efficiently optimized using RandomizedSearchCV to find the best hyperparameters . The solution culminates in a real-time simulation where the model predicts the threat and immediately triggers automated email alerts for non-benign traffic

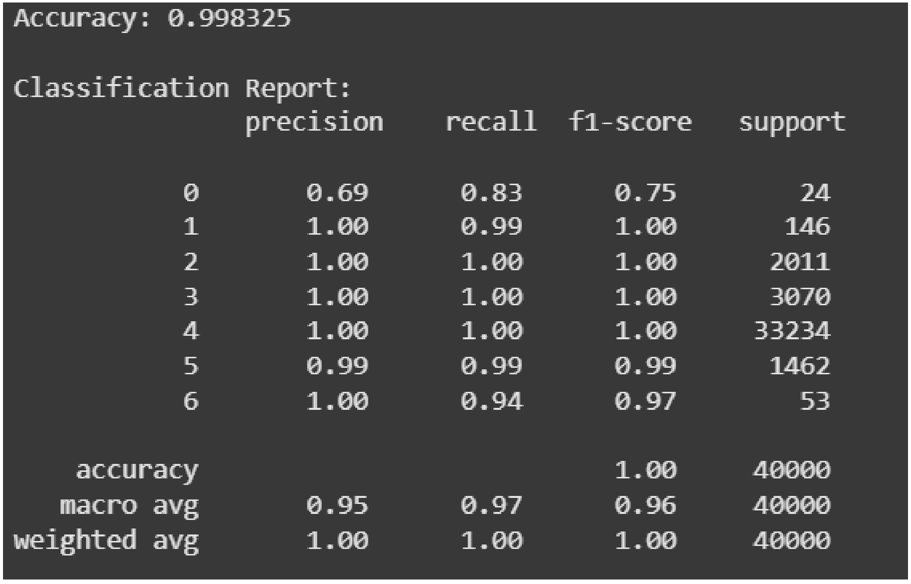
Screenshots:

Dataset screenshot:



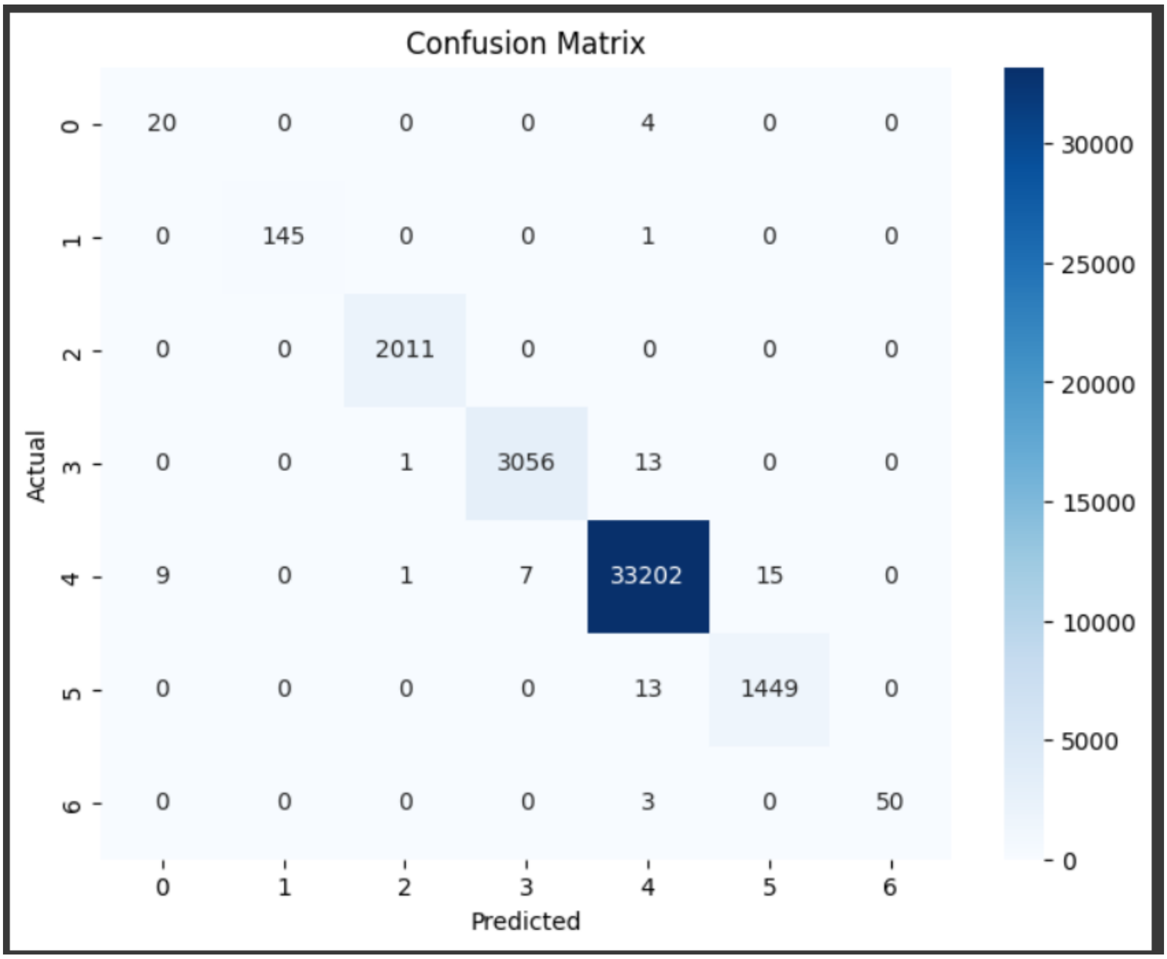
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2520751 entries, 0 to 2520750
Data columns (total 53 columns):
 #   Column                       Dtype
---  ------                       -----
 0   Destination Port             int64
 1   Flow Duration                int64
 2   Total Fwd Packets            int64
 3   Total Length of Fwd Packets  int64
 4   Fwd Packet Length Max        int64
 5   Fwd Packet Length Min        int64
 6   Fwd Packet Length Mean       float64
 7   Fwd Packet Length Std        float64
 8   Bwd Packet Length Max        int64
 9   Bwd Packet Length Min        int64
 10  Bwd Packet Length Mean       float64
 11  Bwd Packet Length Std        float64
 12  Flow Bytes/s                 float64
 13  Flow Packets/s               float64
 14  Flow IAT Mean                float64
 15  Flow IAT Std                 float64
 16  Flow IAT Max                 int64
 17  Flow IAT Min                 int64
 18  Fwd IAT Total                int64
 19  Fwd IAT Mean                 float64
 20  Fwd IAT Std                  float64
 21  Fwd IAT Max                  int64
 22  Fwd IAT Min                  int64
 23  Bwd IAT Total                int64
```

Accuracy screenshot:



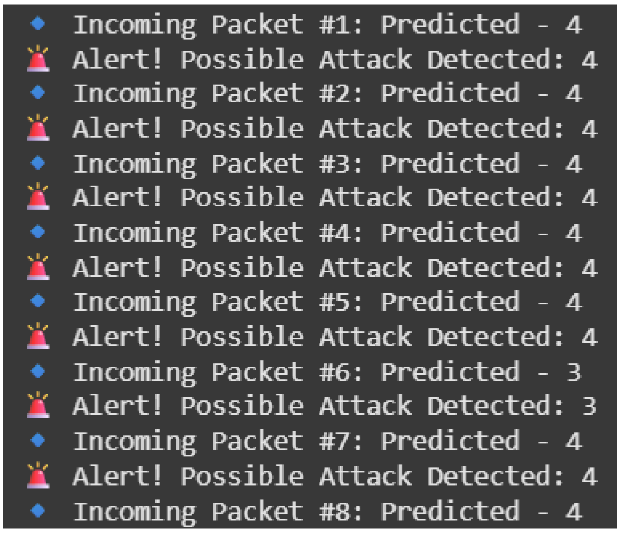
```
Accuracy: 0.998325

Classification Report:
              precision    recall  f1-score   support

           0       0.69      0.83      0.75        24
           1       1.00      0.99      1.00       146
           2       1.00      1.00      1.00      2011
           3       1.00      1.00      1.00      3070
           4       1.00      1.00      1.00     33234
           5       0.99      0.99      0.99      1462
           6       1.00      0.94      0.97        53

    accuracy                           1.00     40000
   macro avg       0.95      0.97      0.96     40000
weighted avg       1.00      1.00      1.00     40000
```

Confusion matrix Screenshot:



	0	1	2	3	4	5	6
0	20	0	0	0	4	0	0
1	0	145	0	0	1	0	0
2	0	0	2011	0	0	0	0
3	0	0	1	3056	13	0	0
4	9	0	1	7	33202	15	0
5	0	0	0	0	13	1449	0
6	0	0	0	0	3	0	50

Result Screenshot



```
🔹 Incoming Packet #1: Predicted - 4
🚨 Alert! Possible Attack Detected: 4
🔹 Incoming Packet #2: Predicted - 4
🚨 Alert! Possible Attack Detected: 4
🔹 Incoming Packet #3: Predicted - 4
🚨 Alert! Possible Attack Detected: 4
🔹 Incoming Packet #4: Predicted - 4
🚨 Alert! Possible Attack Detected: 4
🔹 Incoming Packet #5: Predicted - 4
🚨 Alert! Possible Attack Detected: 4
🔹 Incoming Packet #6: Predicted - 3
🚨 Alert! Possible Attack Detected: 3
🔹 Incoming Packet #7: Predicted - 4
🚨 Alert! Possible Attack Detected: 4
🔹 Incoming Packet #8: Predicted - 4
```

DISCUSSION

The optimized Random Forest model achieved high performance in multi-class threat classification. However, the discussion highlights that purely supervised models risk a drastic drop in recall when facing entirely *novel* (unseen) attacks. Furthermore, simple network features like packet length alone are insufficient for precise classification, and the foundational CICIDS2017 dataset is known to contain duplicate data and mislabels, which risks skewing model training.

FUTUREIMPROVEMENTS

Future research should prioritize rigorous data cleaning to address known imperfections in the underlying dataset. We suggest exploring hybrid models that incorporate unsupervised techniques (e.g., Isolation Forest) to ensure high recall against novel or zero-day threats. Additionally, Deep Learning models (like CNNs or RNNs) should be investigated for their ability to automatically extract complex temporal dependencies from raw traffic flows, which current models may miss.

CONCLUSION

This project successfully implemented a highly accurate, optimized Random Forest-based NIDS for multi-class threat classification. The integration of a real-time detection and automated response mechanism successfully addresses a critical operational gap in cybersecurity . This work provides a functional blueprint for transitioning security operations from reactive logging to proactive, data-driven threat anticipation and mitigation.