Progress Report: Random Forest Model Training

March 21, 2025

1 Introduction

This report presents the progress in training a Random Forest model for network traffic classification. The model was trained using a labeled dataset to distinguish between benign and attack traffic.

2 Training Process

The dataset was preprocessed by encoding categorical variables, normalizing numerical features, and performing a train-test split. The Random Forest classifier was trained with 100 decision trees, using a random seed for reproducibility. The model learned patterns from the training data and was evaluated on the test set.

3 Evaluation Results

The model achieved high accuracy in distinguishing between benign and attack traffic. The key metrics obtained from the evaluation are:

- **Accuracy**: 99.42%
- **Precision**:
  - Benign: 1.00
  - Attack: 0.91
- **Recall**:
  - Benign: 1.00
  - Attack: 0.94
- **F1-score**:
  - Benign: 1.00
  - Attack: 0.93

4 Confusion Matrix

Figure 1 shows the confusion matrix, illustrating the correct and incorrect classifications.

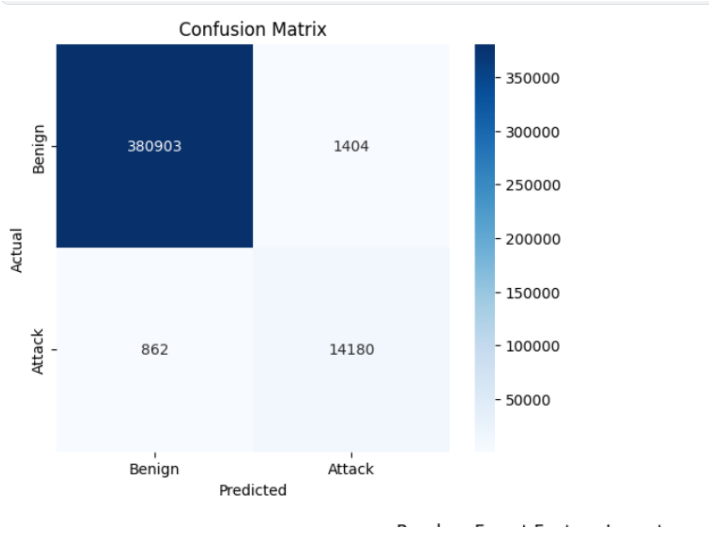

Figure 1: Confusion Matrix of the Random Forest Model

5 Feature Importance

The most influential features in the classification process were analyzed. Figure 2 presents the feature importance ranking based on the model's decisions.

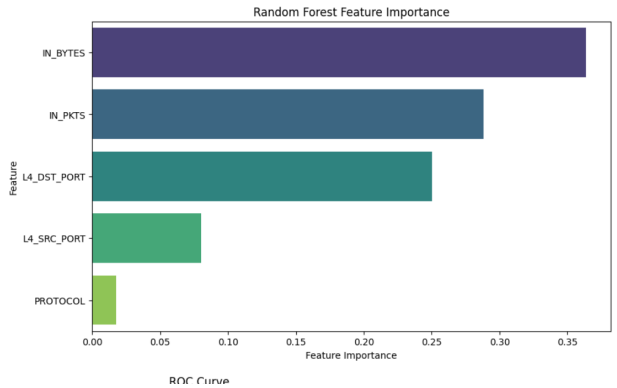

Figure 2: Feature Importance in the Random Forest Model

6 Conclusion

The Random Forest model demonstrated strong classification performance, achieving high accuracy and precision. Future work includes fine-tuning hyperparameters and exploring additional models for further improvement.