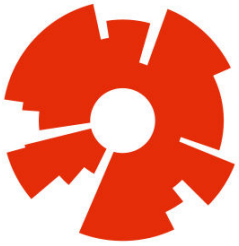

HÁSKÓLINN Í REYKJAVÍK
REYKJAVIK UNIVERSITY

Computer Science department

Machine Learning in Cybersecurity

T-710-MLCS

Report 3 - NIDS experiments

Saldana Giorgio
Email: giorgio24@ru.is

26th September 2024

Contents

1 Introduction

In the current digital landscape, the rapid expansion of network systems has led to an increasing vulnerability to cyber-attacks. Among these threats, network intrusions pose a significant risk to data security, necessitating the development of efficient detection mechanisms. One promising approach to identifying these threats is through the use of machine learning algorithms. This report focuses on employing decision trees and random forest classifiers to detect network intrusions using the CIC-IDS2017 dataset, a widely recognized dataset in intrusion detection research.

The primary objective of this study is to preprocess the CIC-IDS2017 dataset, apply classification techniques, and evaluate the performance of decision trees and random forests in identifying various types of network intrusions. These classifiers will be tested on two different dataset splits: a random 60%-40% split and a split based on specific days of the week. By comparing the performance of both classifiers, this report aims to highlight the strengths and weaknesses of decision trees and random forests in terms of accuracy and feature importance. Furthermore, it seeks to identify key features that contribute to the detection of different types of network attacks, offering insights into the decision-making process of these classifiers.

2 Decision Tree Results

In this section, we present the performance results of the decision tree classifier using two different dataset splits: a 60%-40% random split and a split based on the days of the week (Monday-Wednesday for training and Thursday-Friday for testing). We assess the classifier's performance using precision, recall, and F1-score metrics for the four aggregated traffic classes: Benign, DoS, PortScan, and Exploit. Additionally, we include confusion matrices for a more detailed understanding of the classification outcomes.

Decision Tree Classifier Performance on float-based split

Here, particularly will be displayed the results from the dataset splitting into 60% for training and 40% for testing. The confusion matrix helps visualize the classifier's performance by showing the exact counts of true and predicted labels for each class.

The confusion matrix provides a detailed breakdown of the decision tree classifier's predictions compared to the true labels. Each cell in the matrix represents the number of instances predicted for a particular class versus the actual class.

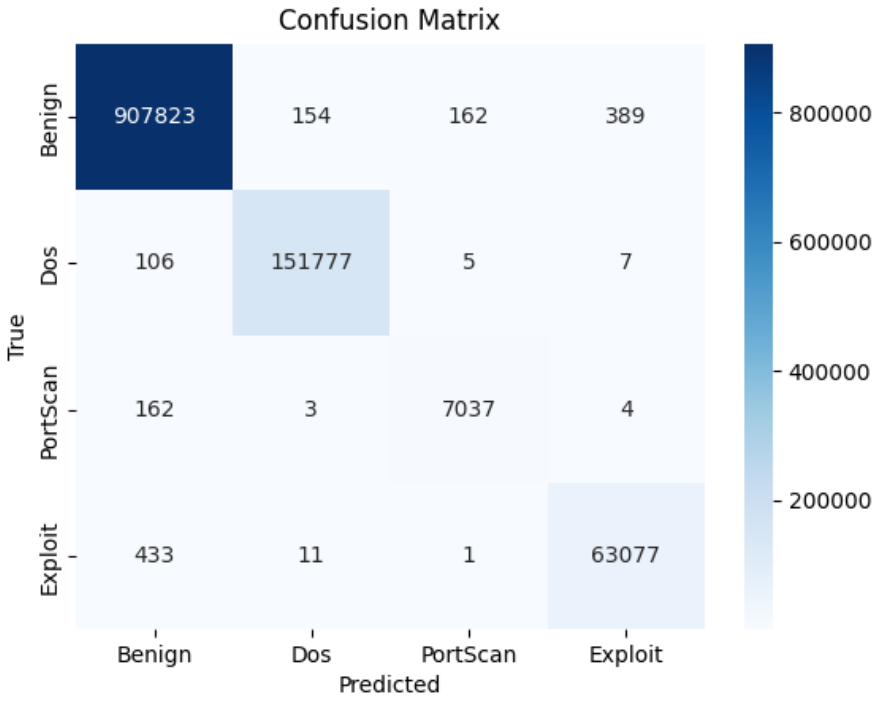

Figure 1: DT Confusion Matrix by 60/40 split

By analizing this matrix we can retrieve some useful information about our model:

- **Benign**: Of the 908,528 true benign instances, 907,823 were correctly classified. Only 154 benign instances were misclassified as DoS, 162 as PortScan, and 389 as

Exploit. This very low misclassification rate is consistent with the high precision and recall values for benign traffic seen in the classification report.

- **Dos**: Out of 151,895 true DoS instances, 151,777 were correctly predicted, and only 106 were mistakenly classified as Benign. A very small number of DoS instances were misclassified as PortScan (5) or Exploit (7). This high level of accuracy confirms the decision tree's effectiveness in identifying DoS attacks, as seen in the high recall and precision scores.
- **PortScan**: For the PortScan class, there were 7,206 true instances, of which 7,037 were correctly classified. However, 162 were misclassified as Benign, and 3 were misclassified as DoS. The lower recall and precision scores for this class reflect this slight increase in misclassifications, indicating that the decision tree had more difficulty accurately detecting all PortScan attacks compared to the other classes.
- **Exploit**: The classifier correctly identified 63,077 out of 63,522 exploit instances. A small number of exploit instances were misclassified as Benign (433), DoS (11), and PortScan (1). These misclassifications are minimal, further reinforcing the strong performance for this class, as indicated by the high F1-score in the classification report.

Decision Tree Classifier Performance on Day-Based Split

The confusion matrix provides further insight into the decision tree classifier's performance when the dataset is split by days:

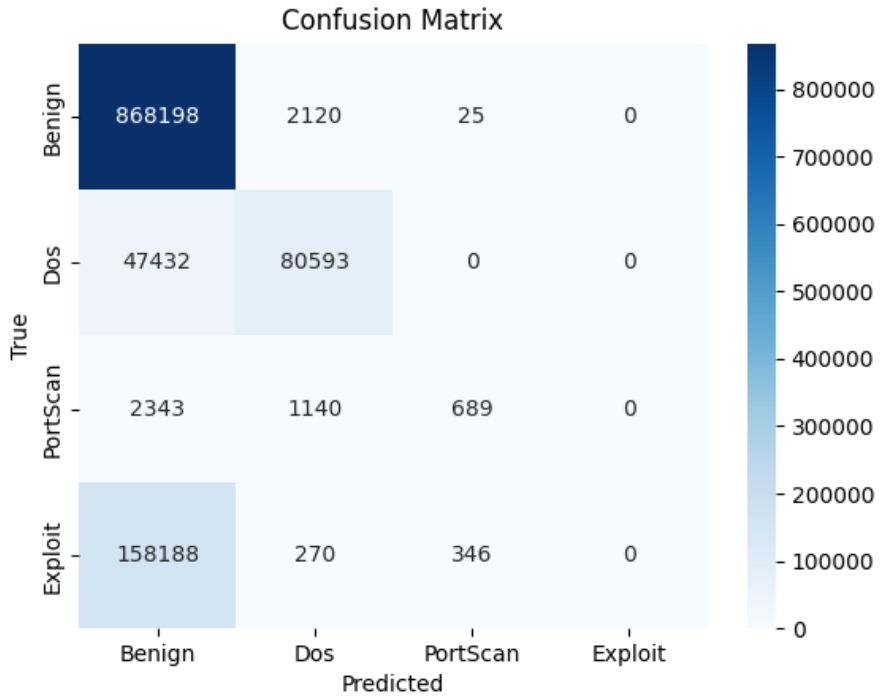

Figure 2: DT Confusion Matrix by Days split

- **Benign Traffic:** Of the 870,343 true benign instances, 868,198 were correctly classified, with 2,120 misclassified as DoS, 25 as PortScan, and none as Exploit. This explains the very high recall of 0.9975 but the lower precision due to some misclassifications.
- **DoS Attacks:** Out of 128,025 true DoS instances, only 80,593 were correctly predicted, while 47,432 were misclassified as Benign. This explains the low recall (0.6295) for DoS, as the model missed a significant number of attacks and instead labeled them as benign traffic.
- **PortScan:** For the PortScan class, only 689 out of 4,172 true instances were correctly identified, while 2,343 were misclassified as Benign and 1,140 as DoS. The significant number of misclassifications explains the low recall of 0.1651.
- **Exploit Attacks:** The classifier completely failed to detect any of the 158,804 Exploit instances, with 158,188 misclassified as Benign. This failure is reflected in the zero precision and recall for the Exploit class.

The day-based split posed greater challenges for the decision tree classifier, particularly in identifying attack traffic. While the classifier maintained high accuracy for benign traffic, it struggled significantly with DoS, PortScan, and especially Exploit attacks. The sharp drop in recall for attack classes suggests that splitting the dataset by days introduced variations in the traffic patterns that the decision tree could not generalize from earlier days (Monday-Wednesday) to later days (Thursday-Friday).

In conclusion, while the decision tree performs reasonably well on benign traffic, its

ability to detect attacks, particularly exploits, is significantly limited in this day-based split. These results highlight the potential limitations of using a simple decision tree model for intrusion detection in temporally split datasets, where more sophisticated models may perform better.

3 Random Forest Results

Random Forest Classifier Performance on float-based split

This section discusses the results of the random forest classifier when the dataset was split randomly into 60% for training and 40% for testing. The confusion matrix provides a detailed visualization of the classifier's predictions compared to the true labels across the four aggregated traffic classes.

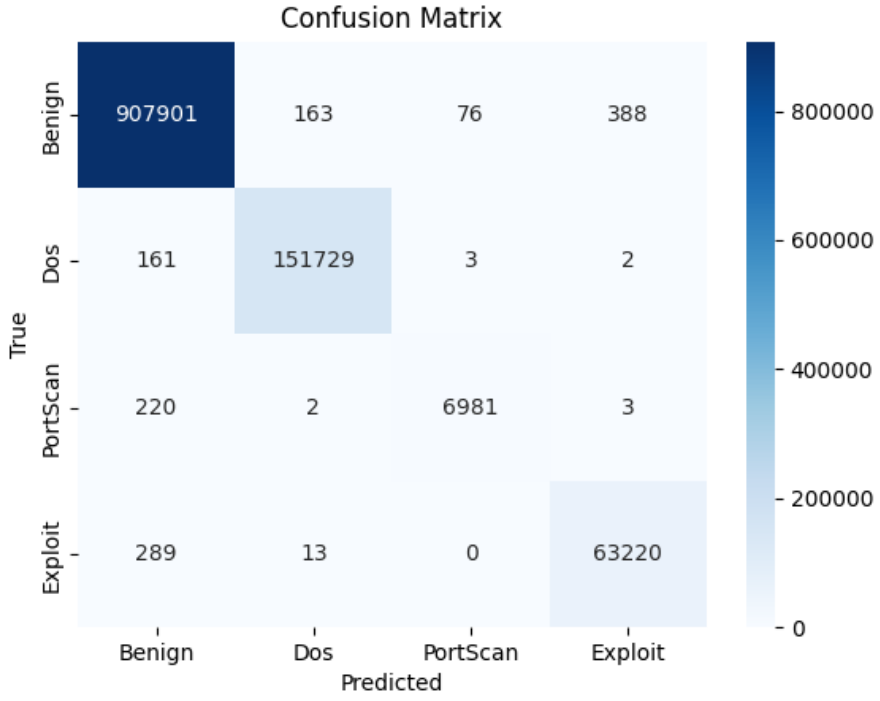

Figure 3: RF Confusion Matrix by 60/40-based split

The confusion matrix shows how the random forest classifier performed in classifying the network traffic for the 60%-40% split:

- **Benign Traffic:** Out of 908,528 true benign instances, 907,901 were correctly classified. Only 163 were misclassified as DoS, 76 as PortScan, and 388 as Exploit. The small number of misclassifications demonstrates the random forest's strong ability to detect benign traffic, significantly minimizing false positives for this class.

- **DoS Attacks:** Of the 151,895 true DoS instances, 151,729 were correctly classified, with only 161 instances misclassified as benign, and very few as PortScan (3) or Exploit (2). This shows a strong capacity to correctly classify DoS attacks with minimal errors.
- **PortScan:** For the PortScan class, 6,981 out of 7,206 instances were correctly identified, while 220 were misclassified as benign and 2 as DoS. Despite some misclassifications, the random forest classifier demonstrated high accuracy in detecting PortScan traffic.
- **Exploit Attacks:** Of the 63,522 true exploit instances, 63,220 were correctly identified, with 289 misclassified as benign and 13 as DoS. None were misclassified as PortScan. These results highlight the model's strong performance in detecting exploit traffic with minimal confusion between classes.

Random Forest Performance on Day-based split

Instead here is provided the results of the random forest classifier when the dataset was split based on days, where data from Monday to Wednesday was used for training and data from Thursday to Friday was used for testing. The confusion matrix provides a detailed visualization of the classifier's predictions compared to the true labels across the four aggregated traffic classes.

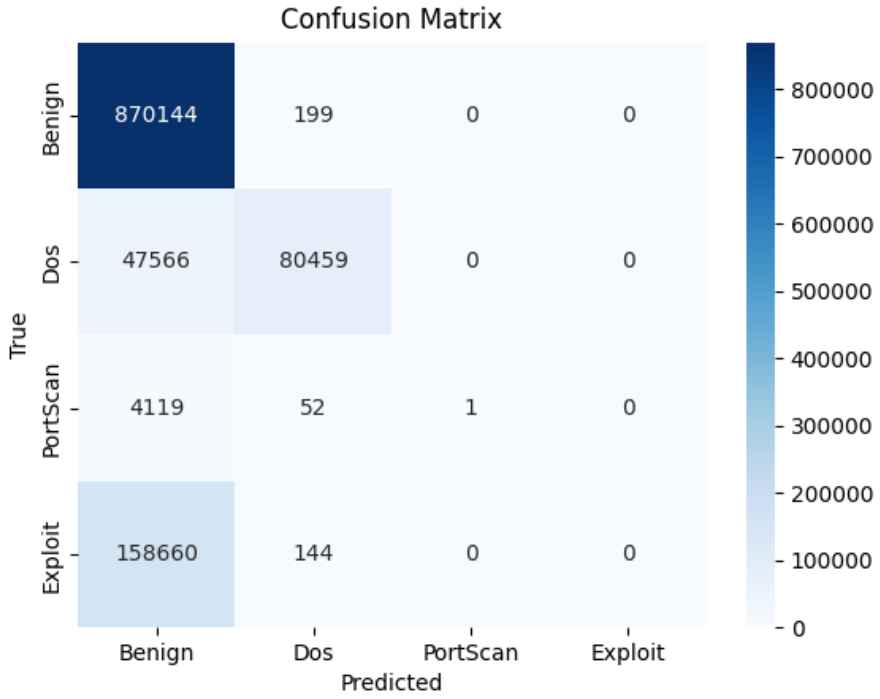

Figure 4: RF Confusion Matrix by Day-based split

The confusion matrix shows how the random forest classifier performed in classifying

the network traffic for the day-based split:

- **Benign Traffic:** Out of 870,343 true benign instances, 870,144 were correctly classified. Only 199 were misclassified as DoS, and none were misclassified as PortScan or Exploit. The minimal misclassifications demonstrate the random forest's high accuracy in detecting benign traffic, with almost no confusion between attack classes.
- **DoS Attacks:** Of the 128,025 true DoS instances, 80,459 were correctly classified, while 47,566 were misclassified as benign. No DoS instances were misclassified as PortScan or Exploit. The large number of DoS instances misclassified as benign traffic indicates that the model struggled to separate these two classes in the day-based split.
- **PortScan:** For the PortScan class, out of 4,172 true instances, only 1 was correctly classified. A large number (4,119) were misclassified as benign traffic, and 52 were misclassified as DoS. This shows that the model had significant difficulty detecting PortScan traffic under this split, misclassifying the majority of the PortScan instances.
- **Exploit Attacks:** Of the 158,804 true exploit instances, none were correctly classified. A total of 158,660 were misclassified as benign, and 144 as DoS. This reflects a complete failure of the model to detect exploit traffic, as all instances were confused with benign or DoS traffic.

4 Feature Importance Analysis

In this section, we discuss the main features identified by both the decision tree and random forest classifiers as discriminating factors. These features were extracted from both models under two different split configurations: a random 60%-40% split and a day-based split (Monday-Wednesday for training and Thursday-Friday for testing). The figures below show the importance values of the top features for each model and split.

Decision Tree Feature Importance

The decision tree model relies on key features to classify network traffic accurately. Below, we examine the most important features identified in both the 60%-40% split and the day-based split:

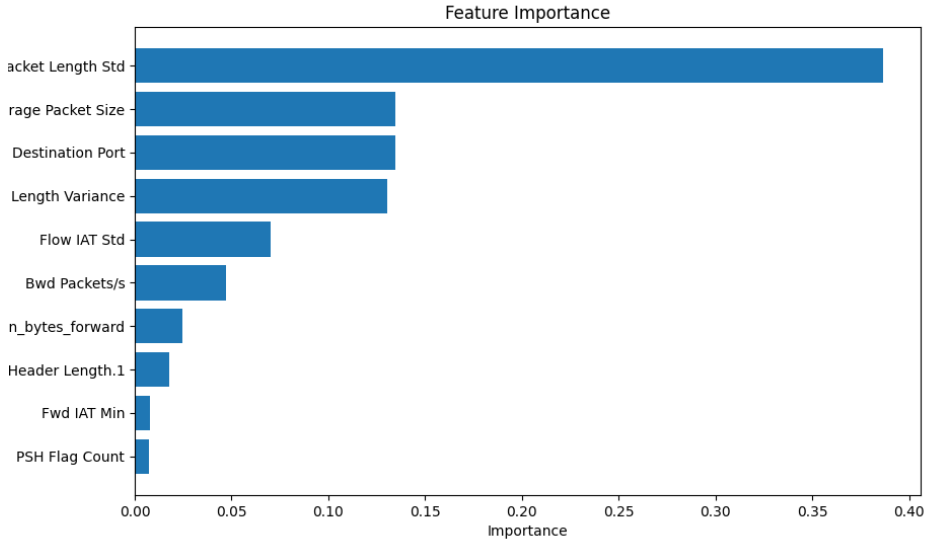

Figure 5: Decision Tree Feature Importance by 60/40 split

- **Packet Length Std:** The standard deviation of packet lengths is the most important feature in both splits, reflecting variability in traffic patterns which is crucial for distinguishing between benign and attack traffic.
- **Destination Port:** This feature is important because specific ports are commonly associated with certain types of network traffic and attacks, such as DoS or PortScan attacks.
- **Flow IAT Mean/Std:** Inter-arrival times (IAT) measure the time between packets in a flow. Higher variability or mean values can indicate anomalous behaviors such as network scans or DDoS attacks.

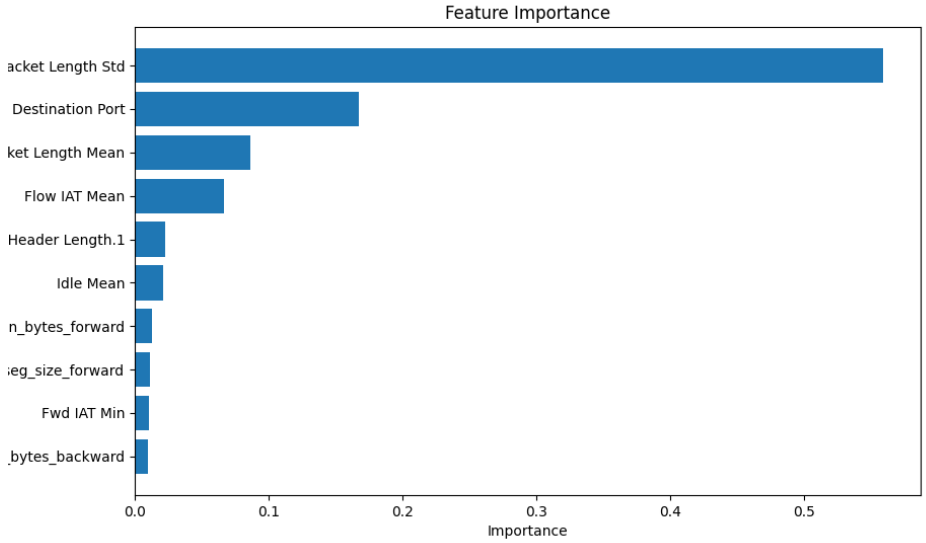

Figure 6: Decision Tree Feature Importance by Day-based split

Random Forest Feature Importance

Random forest classifiers leverage multiple features more evenly compared to decision trees. The following figures depict the importance of features for random forest models under both split configurations.

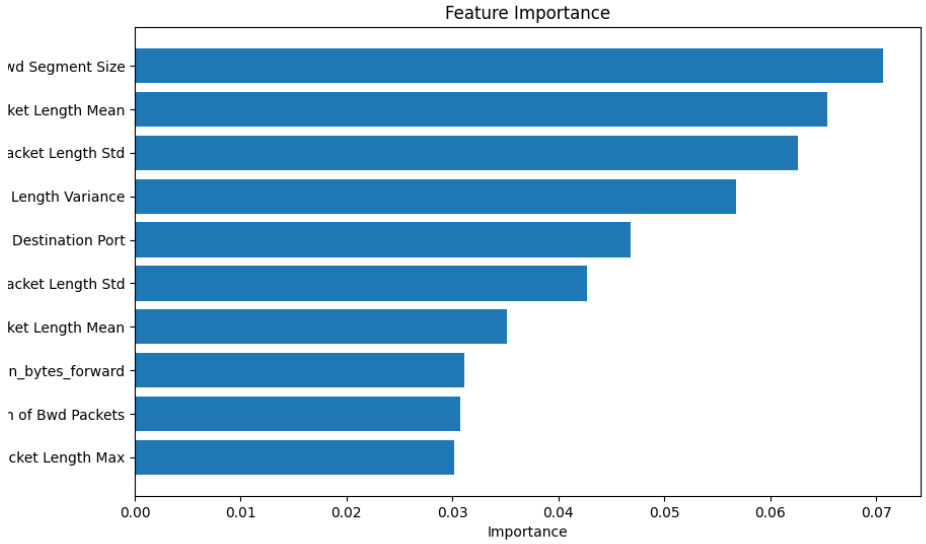

Figure 7: Random Forest Feature Importance by 60/40 split

- **Packet Length Std:** Similar to the decision tree, the standard deviation of packet

length is the most critical feature, underscoring its relevance across different model types and splits.

- **Average Packet Size:** This feature helps the model distinguish between various traffic patterns. For example, large packet sizes may be indicative of data exfiltration or file-sharing activities, while small packet sizes may be more common in benign traffic or reconnaissance attacks.
- **Flow IAT Std:** Variations in the time between packets often correlate with network attacks, where the attacker sends packets at irregular intervals to probe network defenses.

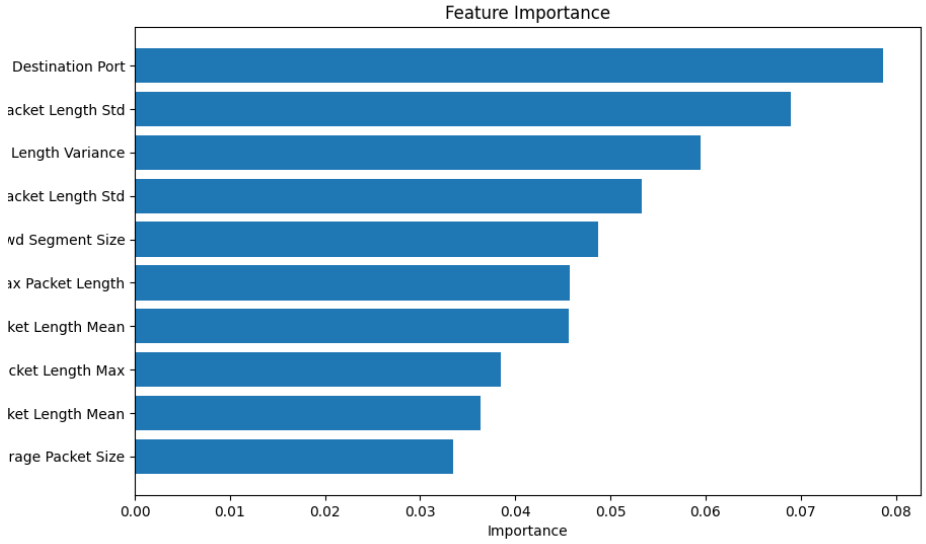

Figure 8: Random Forest Feature Importance by Day-based split

Interpretation of Feature Importance for Attack Detection

The identified features, such as the *Packet Length Std*, *Destination Port*, and *Flow IAT*, are highly indicative of different network traffic behaviors. For instance:

- **DoS and DDoS Attacks:** These attacks often flood the network with a large number of packets, which may show up as irregular packet lengths and high variability in inter-arrival times (IAT), making *Packet Length Std* and *Flow IAT Mean/Std* essential for their detection.
- **PortScan Attacks:** Port scanning involves probing different ports on a target machine, which makes the *Destination Port* feature highly relevant, as it allows the classifier to detect unusual traffic to specific ports commonly targeted in attacks.
- **Exploit Attacks:** These attacks may manifest in sudden changes in packet size or the number of bytes sent in the backward direction (*bwd_packets_s*). Detecting such variations helps in identifying these types of anomalies.

In summary, both decision trees and random forests use a similar set of features,

such as packet length variability and destination ports, to distinguish between benign and malicious traffic. Random forests, due to their ensemble nature, may balance the use of these features more effectively, resulting in better overall performance in detecting complex attack patterns across different dataset splits.