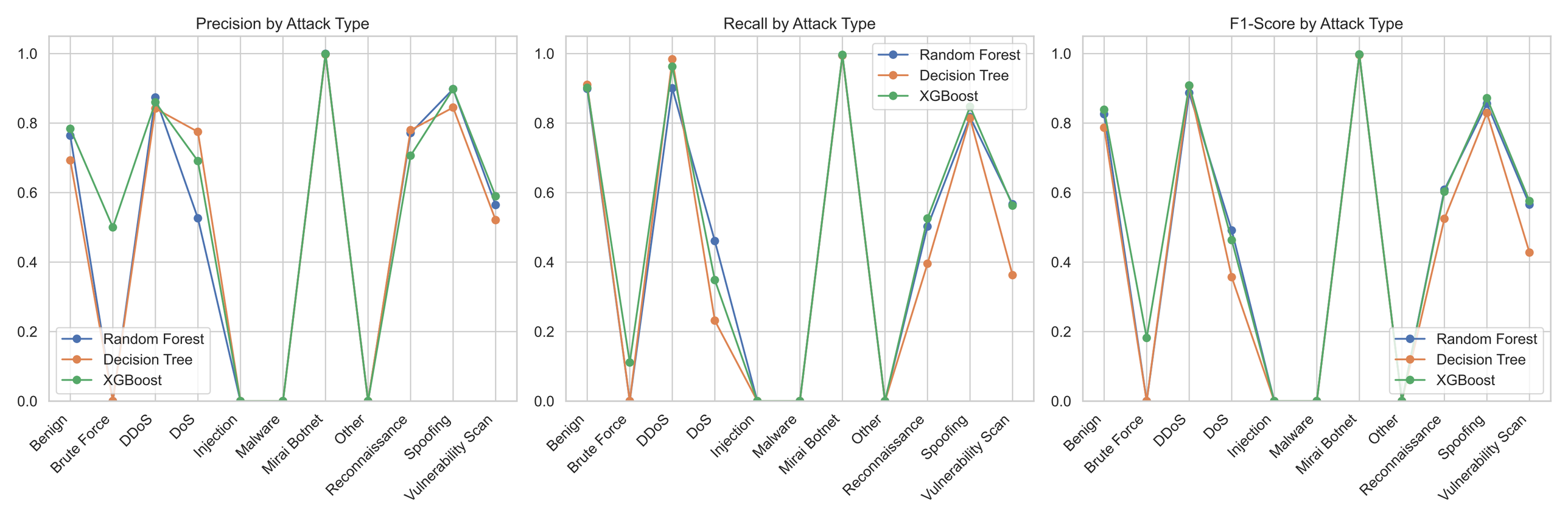
IoT Security Threat Detection for SMEs:

A Machine Learning Approach Using CIC-IoT Dataset

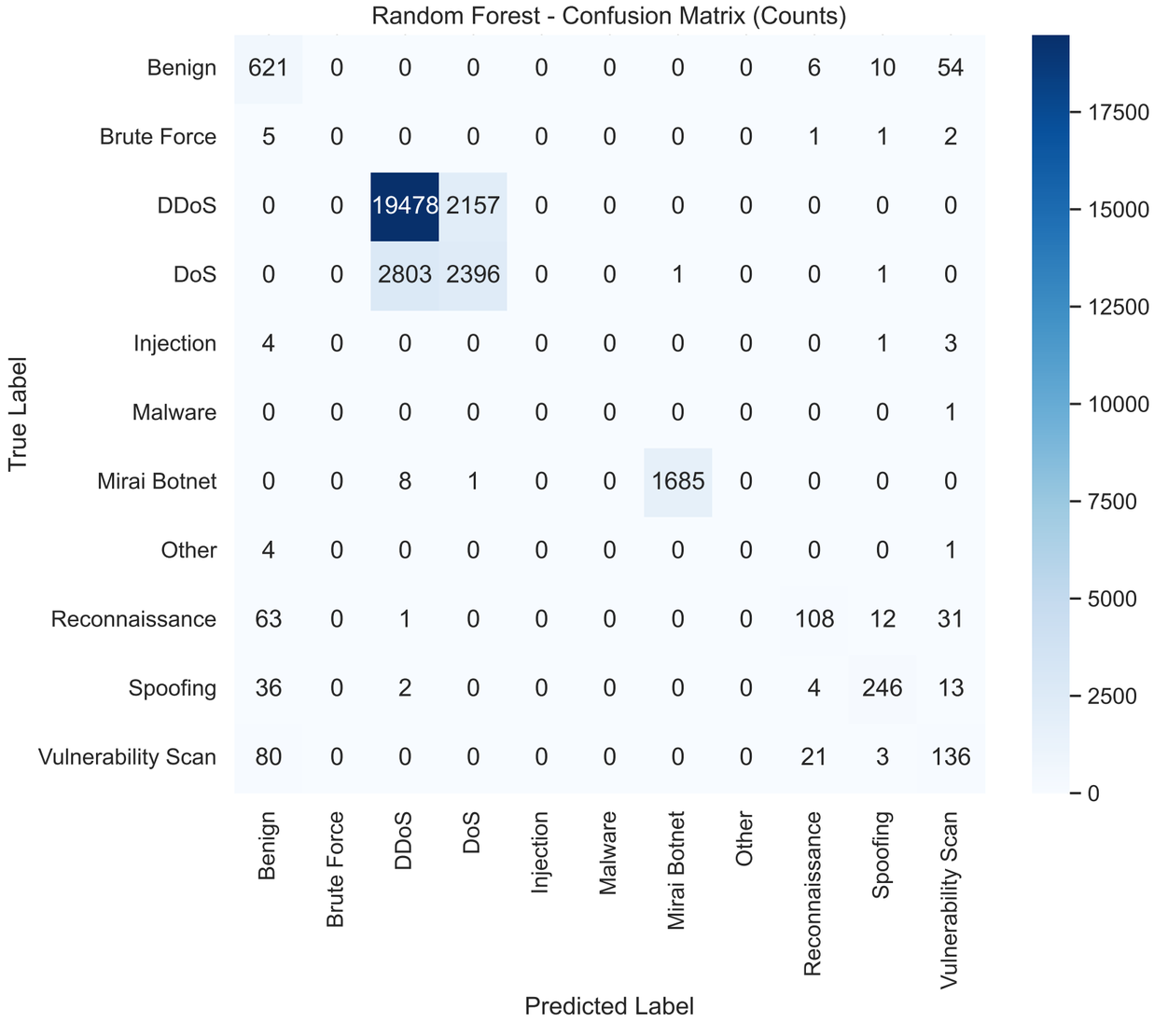
STAGE 5, STEP 1: PERFORMANCE METRICS

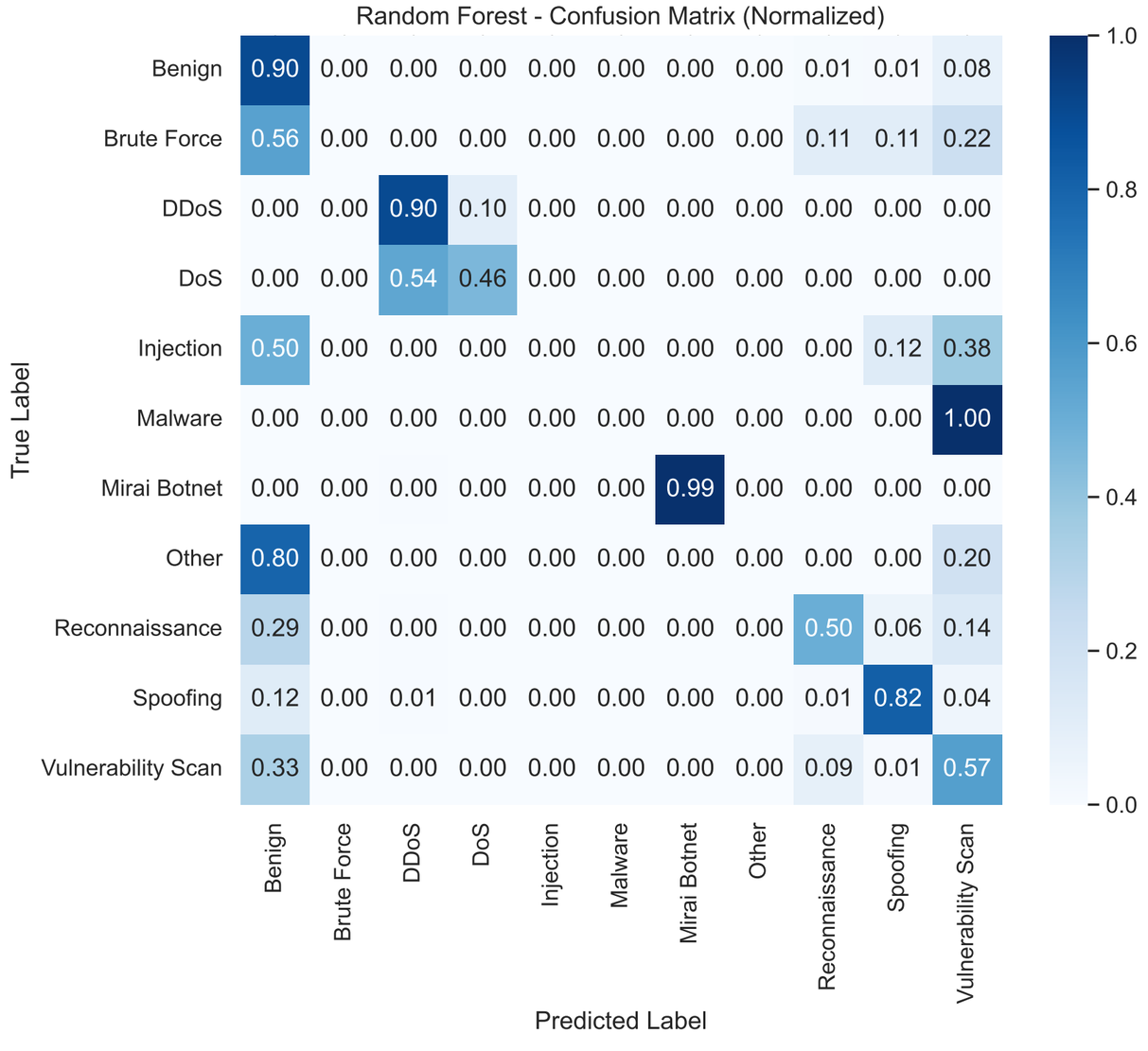
This report evaluates the performance of IoT security threat detection models, focusing on precision, recall, F1-score, confusion matrix analysis, detection latency, false positive impact, and resource requirements.


Precision by Attack Type
Recall by Attack Type
F1-Score by Attack Type
Random Forest
Decision Tree
XGBoost
1.0
0.8
0.6
0.4
0.2
0.0
Benign
Brute Force
DDoS
DoS
Injection
Malware
Mirai Botnet
Other
Reconnaissance
Spoofing
Vulnerability Scan

These charts compare precision, recall, and F1-scores across different attack categories for three machine learning models: Random Forest, Decision Tree, and XGBoost. The metrics reveal significant performance variations
across attack types. For common attacks like DDoS and DoS, all models perform well with high precision and recall,
making them reliable for detecting the most frequent threats. However, for rarer attack categories like Injection
and Brute Force, performance drops noticeably, particularly for Decision Trees, which struggle with class imbalance.

Random Forest consistently provides the best overall performance, especially for minority attack classes, highlighting its robustness for comprehensive threat detection in SME environments. The performance gaps for rare
but dangerous attacks emphasize the importance of targeted model tuning and specialized detection approaches for
critical threats. For resource-constrained SMEs, these metrics help prioritize which attack types require additional
detection mechanisms beyond ML-based approaches, ensuring comprehensive security coverage without overwhelming
limited IT resources.

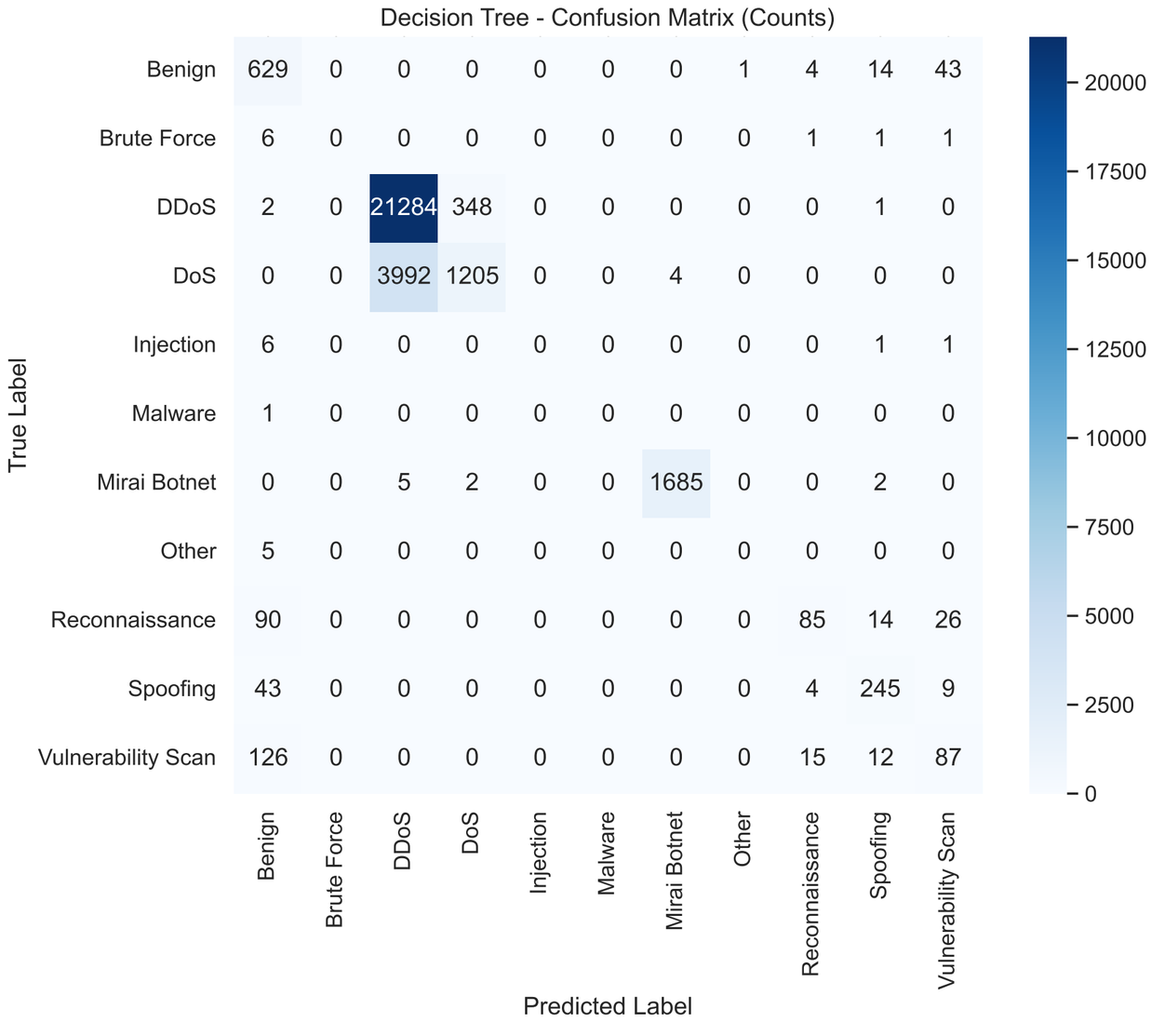

Random Forest - Confusion Matrix (Counts)
True Label
Predicted Label
Benign 621 0 0 0 0 0 0 0 6 10 54
Brute Force 5 0 0 0 0 0 0 0 1 1 2
DDoS 0 0 19478 2157 0 0 0 0 0 0 0
DoS 0 0 2803 2396 0 0 1 0 0 1 0
Injection 4 0 0 0 0 0 0 0 0 1 3
Malware 0 0 0 0 0 0 0 0 0 0 1
Mirai Botnet 0 0 8 1 0 0 1685 0 0 0 0
Other 4 0 0 0 0 0 0 0 0 0 1
Reconnaissance 63 0 1 0 0 0 0 0 108 12 31
Spoofing 36 0 2 0 0 0 0 0 4 246 13
Vulnerability Scan 80 0 0 0 0 0 0 0 21 3 136
Benign
Brute Force
DDoS
DoS
Injection
Malware
Mirai Botnet
Other
Reconnaissance
Spoofing
Vulnerability Scan
17500
15000
12500
10000
7500
5000
2500
0

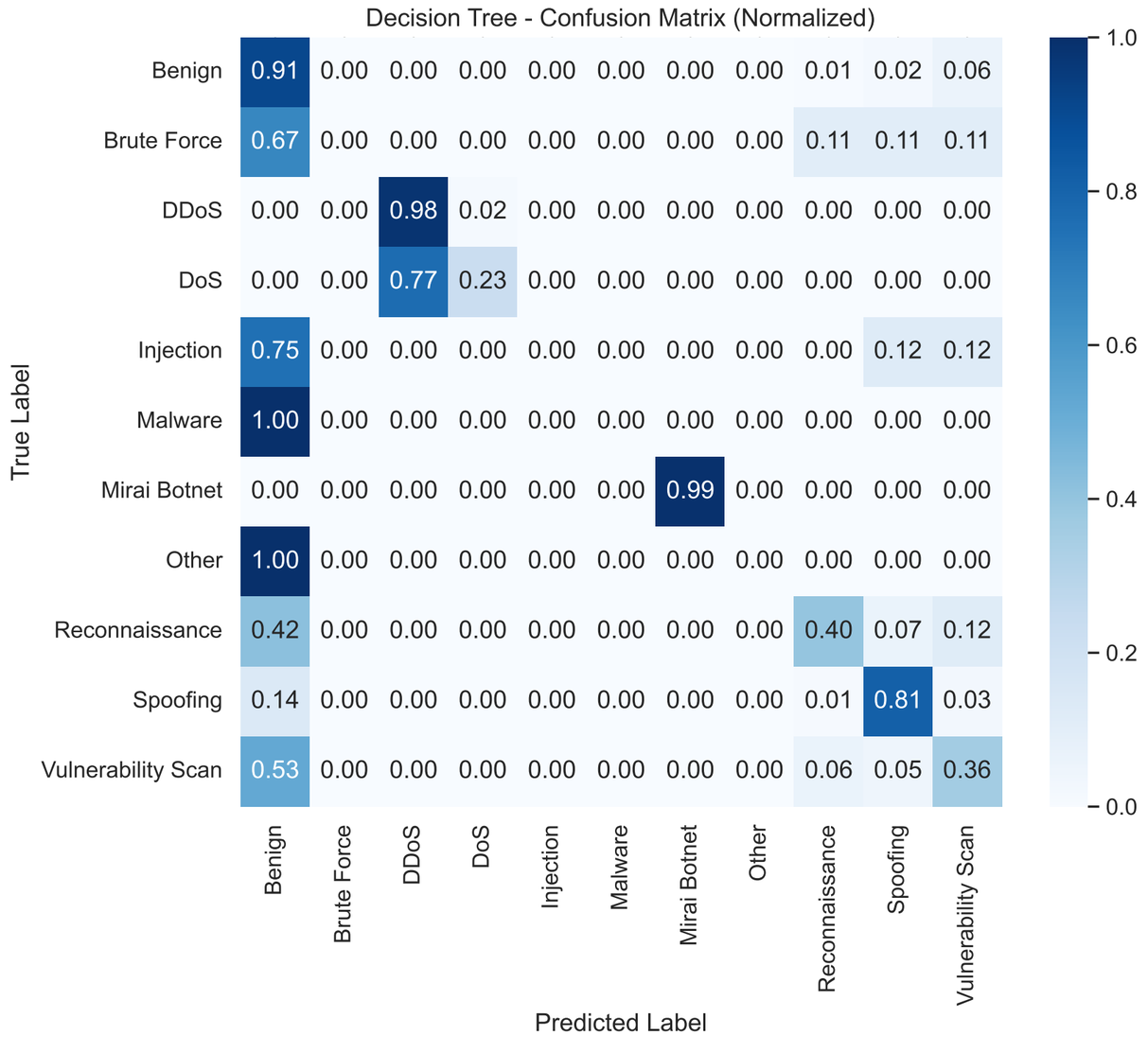

Random Forest - Confusion Matrix (Normalized)
True Label
Predicted Label
Benign 0.90 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.01 0.01 0.08
Brute Force 0.56 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.11 0.11 0.22
DDoS 0.00 0.00 0.90 0.10 0.00 0.00 0.00 0.00 0.00 0.00 0.00
DoS 0.00 0.00 0.54 0.46 0.00 0.00 0.00 0.00 0.00 0.00 0.00
Injection 0.50 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.12 0.38
Malware 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 1.00
Mirai Botnet 0.00 0.00 0.00 0.00 0.00 0.00 0.99 0.00 0.00 0.00 0.00
Other 0.80 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.20
Reconnaissance 0.29 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.50 0.06 0.14
Spoofing 0.12 0.00 0.01 0.00 0.00 0.00 0.00 0.00 0.01 0.82 0.04
Vulnerability Scan 0.33 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.09 0.01 0.57
Benign
Brute Force
DDoS
DoS
Injection
Malware
Mirai Botnet
Other
Reconnaissance
Spoofing
Vulnerability Scan
1.0
0.8
0.6
0.4
0.2
0.0

The confusion matrices for the Random Forest model reveal its classification performance across different attack types.
The left matrix shows raw counts, useful for understanding the actual number of samples in each category, while
the right matrix shows normalized values, highlighting the proportion of correctly classified instances for each
true label.

Strong diagonal elements indicate high classification accuracy, while off-diagonal elements reveal misclassifications. For this model, benign traffic and common attacks like DDoS show high accuracy, suggesting
reliable detection of normal traffic and the most frequent attack types. However, some attack categories exhibit
notable misclassifications, particularly among similar attack families (e.g., DoS being classified as DDoS) or
rare attack types with limited training samples.

For SMEs with limited security expertise, these matrices help identify which attack types might trigger false
alarms or go undetected, informing where additional verification steps or complementary detection methods should
be implemented to reduce false positives and ensure reliable threat detection in resource-constrained environments.

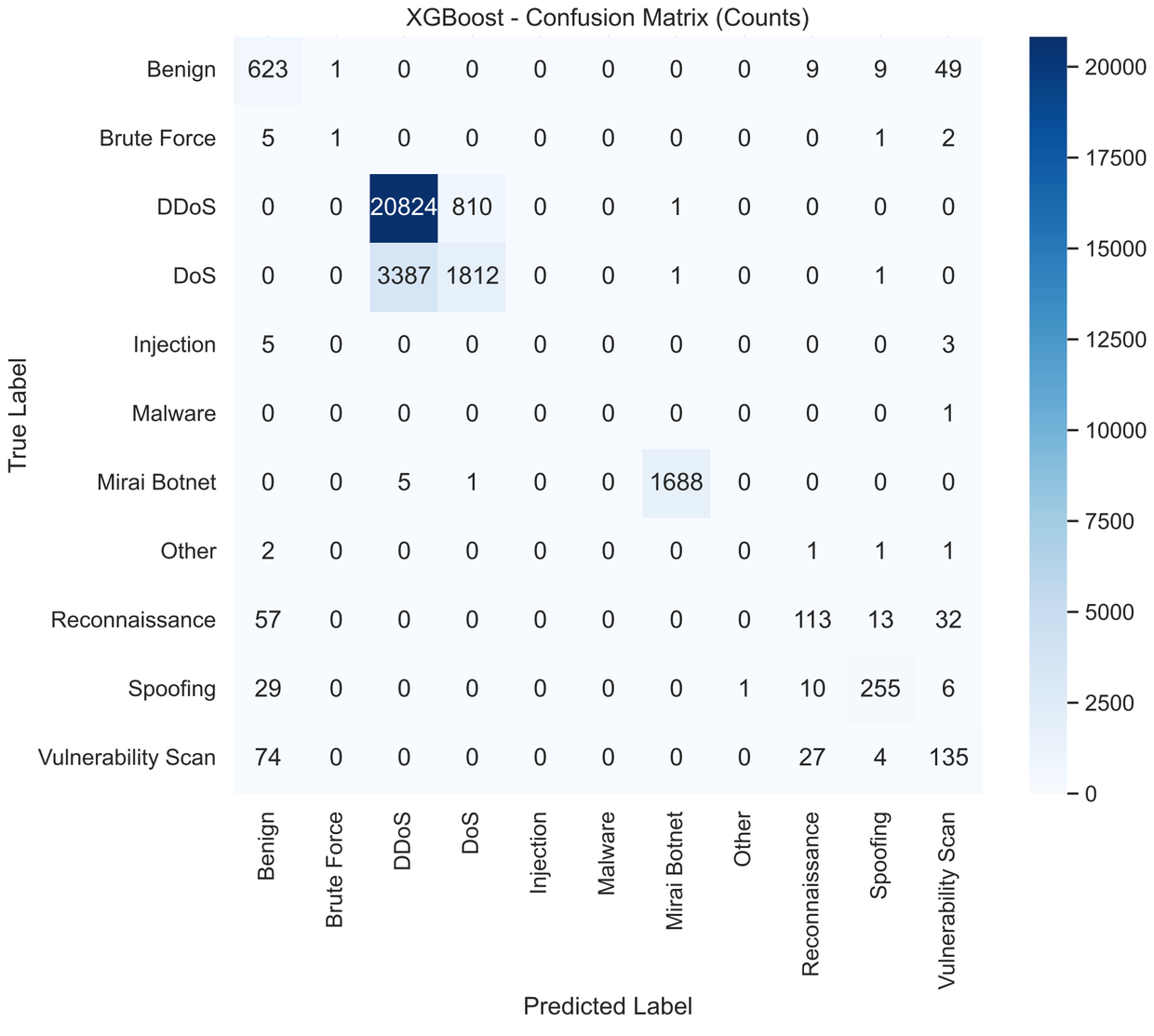

Decision Tree - Confusion Matrix (Counts)
True Label
Predicted Label
Benign 629 0 0 0 0 0 0 1 4 14 43
Brute Force 6 0 0 0 0 0 0 0 1 1 1
DDoS 2 0 21284 348 0 0 0 0 0 1 0
DoS 0 0 3992 1205 0 0 4 0 0 0 0
Injection 6 0 0 0 0 0 0 0 0 1 1
Malware 1 0 0 0 0 0 0 0 0 0 0
Mirai Botnet 0 0 5 2 0 0 1685 0 0 2 0
Other 5 0 0 0 0 0 0 0 0 0 0
Reconnaissance 90 0 0 0 0 0 0 0 85 14 26
Spoofing 43 0 0 0 0 0 0 0 4 245 9
Vulnerability Scan 126 0 0 0 0 0 0 0 15 12 87
Benign
Brute Force
DDoS
DoS
Injection
Malware
Mirai Botnet
Other
Reconnaissance
Spoofing
Vulnerability Scan
20000
17500
15000
12500
10000
7500
5000
2500
0

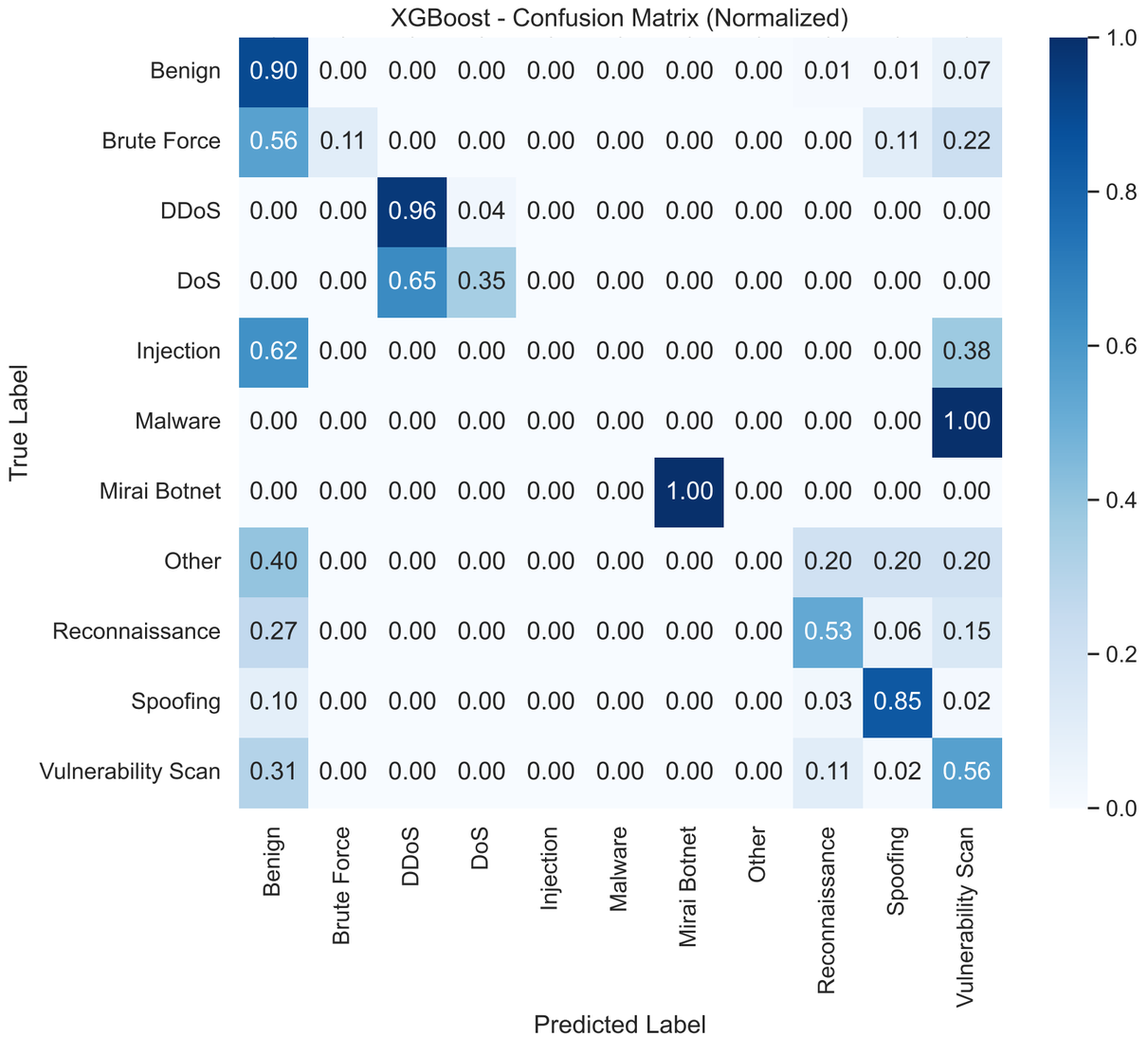

Decision Tree - Confusion Matrix (Normalized)
True Label
Predicted Label
Benign 0.91 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.01 0.02 0.06
Brute Force 0.67 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.11 0.11 0.11
DDoS 0.00 0.00 0.98 0.02 0.00 0.00 0.00 0.00 0.00 0.00 0.00
DoS 0.00 0.00 0.77 0.23 0.00 0.00 0.00 0.00 0.00 0.00 0.00
Injection 0.75 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.12 0.12
Malware 1.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
Mirai Botnet 0.00 0.00 0.00 0.00 0.00 0.00 0.99 0.00 0.00 0.00 0.00
Other 1.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
Reconnaissance 0.42 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.40 0.07 0.12
Spoofing 0.14 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.01 0.81 0.03
Vulnerability Scan 0.53 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.06 0.05 0.36
Benign
Brute Force
DDoS
DoS
Injection
Malware
Mirai Botnet
Other
Reconnaissance
Spoofing
Vulnerability Scan
1.0
0.8
0.6
0.4
0.2
0.0

The confusion matrices for the Decision Tree model reveal its classification performance across different attack types.
The left matrix shows raw counts, useful for understanding the actual number of samples in each category, while
the right matrix shows normalized values, highlighting the proportion of correctly classified instances for each
true label.

Strong diagonal elements indicate high classification accuracy, while off-diagonal elements reveal misclassifications. For this model, benign traffic and common attacks like DDoS show high accuracy, suggesting
reliable detection of normal traffic and the most frequent attack types. However, some attack categories exhibit
notable misclassifications, particularly among similar attack families (e.g., DoS being classified as DDoS) or
rare attack types with limited training samples.

For SMEs with limited security expertise, these matrices help identify which attack types might trigger false
alarms or go undetected, informing where additional verification steps or complementary detection methods should
be implemented to reduce false positives and ensure reliable threat detection in resource-constrained environments.

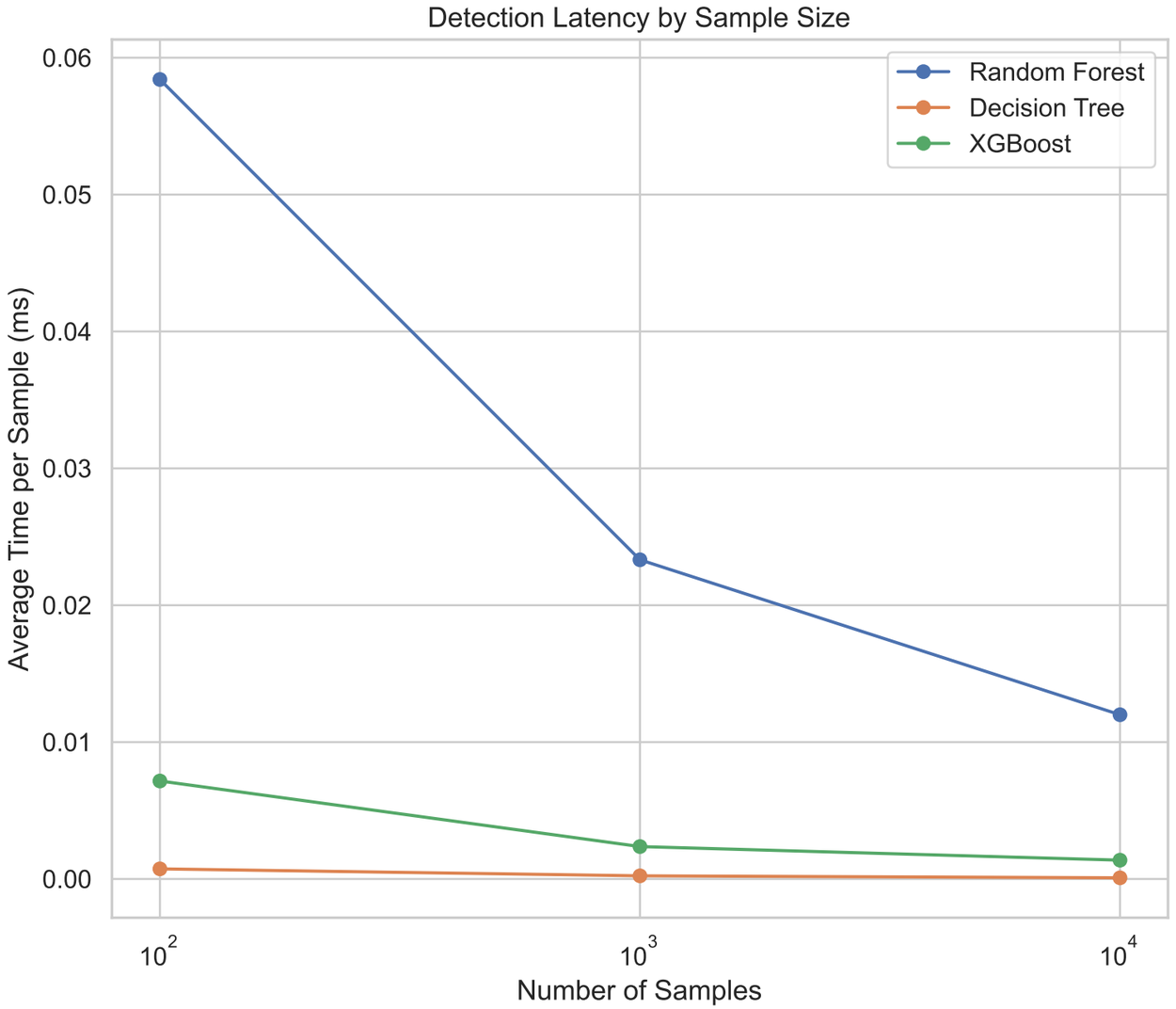

XGBoost - Confusion Matrix (Counts)
True Label
Predicted Label
Benign 623 1 0 0 0 0 0 0 9 9 49
Brute Force 5 1 0 0 0 0 0 0 0 1 2
DDoS 0 0 20824 810 0 0 1 0 0 0 0
DoS 0 0 3387 1812 0 0 1 0 0 1 0
Injection 5 0 0 0 0 0 0 0 0 0 3
Malware 0 0 0 0 0 0 0 0 0 0 1
Mirai Botnet 0 0 5 1 0 0 1688 0 0 0 0
Other 2 0 0 0 0 0 0 0 1 1 1
Reconnaissance 57 0 0 0 0 0 0 0 113 13 32
Spoofing 29 0 0 0 0 0 0 1 10 255 6
Vulnerability Scan 74 0 0 0 0 0 0 0 27 4 135
Benign
Brute Force
DDoS
DoS
Injection
Malware
Mirai Botnet
Other
Reconnaissance
Spoofing
Vulnerability Scan
20000
17500
15000
12500
10000
7500
5000
2500
0

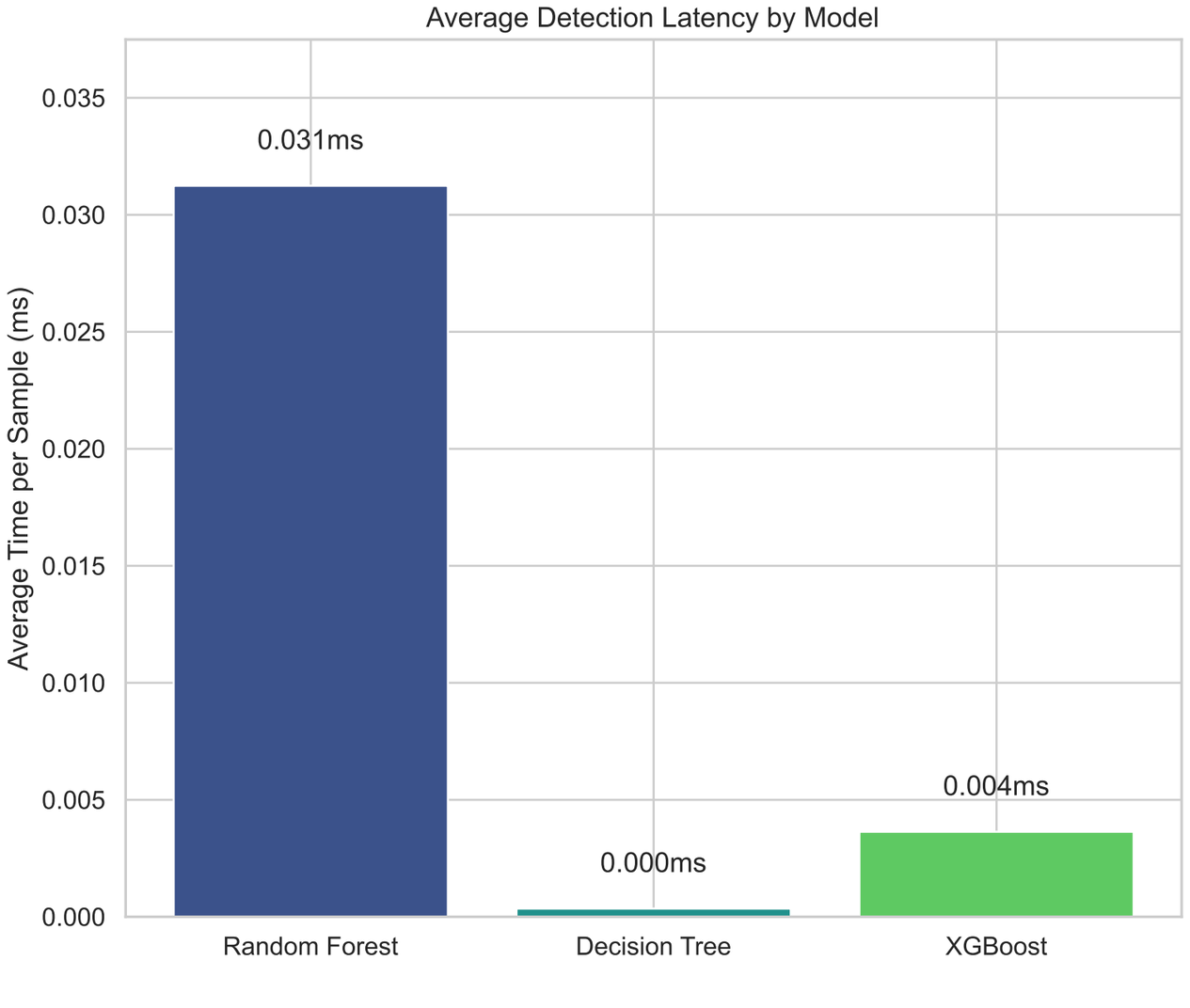

XGBoost - Confusion Matrix (Normalized)
True Label
Predicted Label
Benign 0.90 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.01 0.01 0.07
Brute Force 0.56 0.11 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.11 0.22
DDoS 0.00 0.00 0.96 0.04 0.00 0.00 0.00 0.00 0.00 0.00 0.00
DoS 0.00 0.00 0.65 0.35 0.00 0.00 0.00 0.00 0.00 0.00 0.00
Injection 0.62 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.38
Malware 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 1.00
Mirai Botnet 0.00 0.00 0.00 0.00 0.00 0.00 1.00 0.00 0.00 0.00 0.00
Other 0.40 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.20 0.20 0.20
Reconnaissance 0.27 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.53 0.06 0.15
Spoofing 0.10 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.03 0.85 0.02
Vulnerability Scan 0.31 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.11 0.02 0.56
Benign
Brute Force
DDoS
DoS
Injection
Malware
Mirai Botnet
Other
Reconnaissance
Spoofing
Vulnerability Scan
1.0
0.8
0.6
0.4
0.2
0.0

The confusion matrices for the XGBoost model reveal its classification performance across different attack types.
The left matrix shows raw counts, useful for understanding the actual number of samples in each category, while
the right matrix shows normalized values, highlighting the proportion of correctly classified instances for each
true label.

Strong diagonal elements indicate high classification accuracy, while off-diagonal elements reveal misclassifications. For this model, benign traffic and common attacks like DDoS show high accuracy, suggesting
reliable detection of normal traffic and the most frequent attack types. However, some attack categories exhibit
notable misclassifications, particularly among similar attack families (e.g., DoS being classified as DDoS) or
rare attack types with limited training samples.

For SMEs with limited security expertise, these matrices help identify which attack types might trigger false
alarms or go undetected, informing where additional verification steps or complementary detection methods should
be implemented to reduce false positives and ensure reliable threat detection in resource-constrained environments.

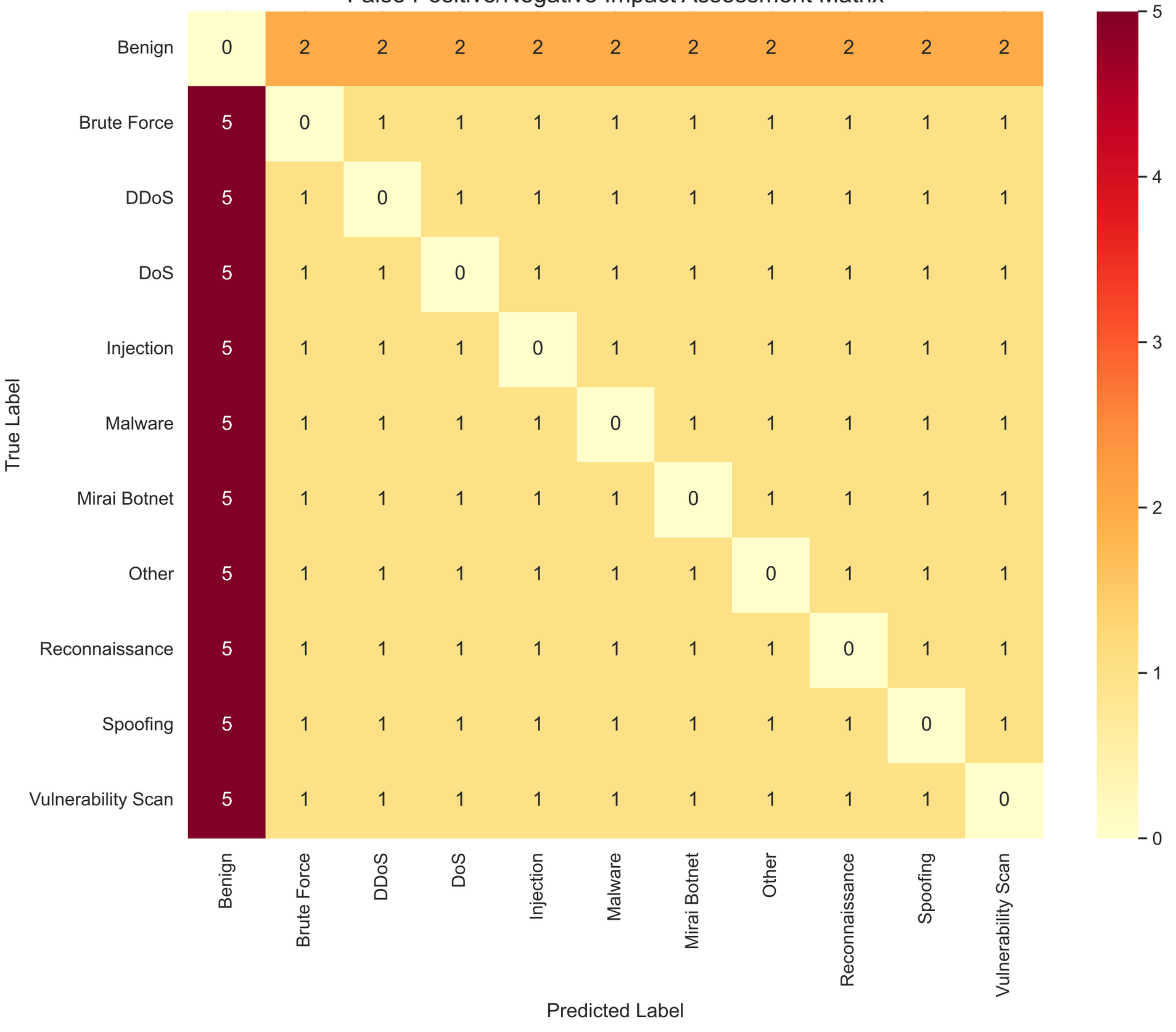

Detection Latency by Sample Size
Random Forest
Decision Tree
XGBoost
Average Time per Sample (ms)
0.06
0.05
0.04
0.03
0.02
0.01
0.00
10^2
10^3
10^4
Number of Samples


Average Detection Latency by Model
Average Time per Sample (ms)
0.035
0.030
0.025
0.020
0.015
0.010
0.005
0.000
0.031ms
0.000ms
0.004ms
Random Forest
Decision Tree
XGBoost

These visualizations examine detection latency, a critical performance metric for real-time IoT security monitoring. The left chart shows how latency scales with increasing sample sizes, while the right chart compares average latency across different models.

Decision Trees consistently demonstrate the lowest latency, making them ideal for resource-constrained edge
devices requiring immediate threat detection. Random Forests offer a balance between accuracy and speed, with
moderate latency suitable for gateway-level deployment. XGBoost exhibits the highest latency, suggesting it's
better suited for central processing where computational resources are less constrained.

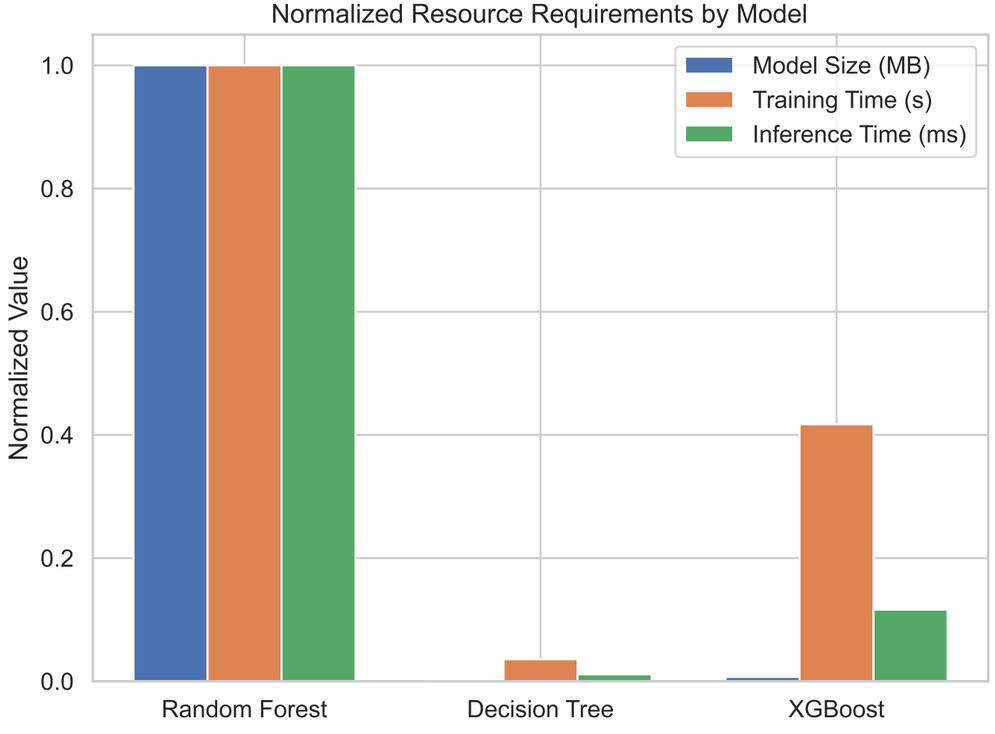
For SMEs, these latency differences translate directly to hardware requirements and deployment options. Edge
devices like sensors or cameras would benefit from the lightweight Decision Tree models, while more powerful
network appliances could leverage Random Forests for better accuracy with acceptable speed. These measurements
help organizations optimize model selection based on their specific infrastructure constraints and security response time requirements, ensuring effective threat detection without overloading limited resources.

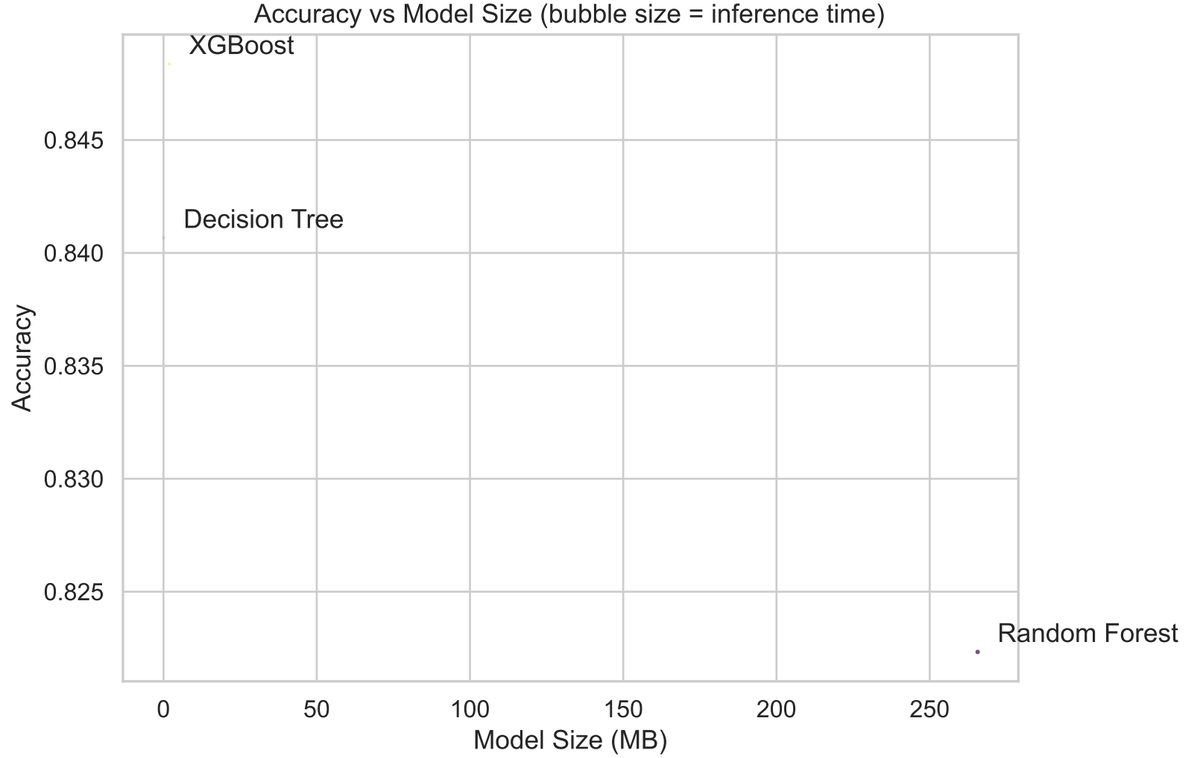

False Positive/Negative Impact Assessment Matrix
True Label
Predicted Label
Benign
Brute Force
DDoS
DoS
Injection
Malware
Mirai Botnet
Other
Reconnaissance
Spoofing
Vulnerability Scan
0 2 2 2 2 2 2 2 2 2 2
5 0 1 1 1 1 1 1 1 1 1
5 1 0 1 1 1 1 1 1 1 1
5 1 1 0 1 1 1 1 1 1 1
5 1 1 1 0 1 1 1 1 1 1
5 1 1 1 1 0 1 1 1 1 1
5 1 1 1 1 1 0 1 1 1 1
5 1 1 1 1 1 1 0 1 1 1
5 1 1 1 1 1 1 1 0 1 1
5 1 1 1 1 1 1 1 1 0 1
5 1 1 1 1 1 1 1 1 1 0
5
4
3
2
1
0

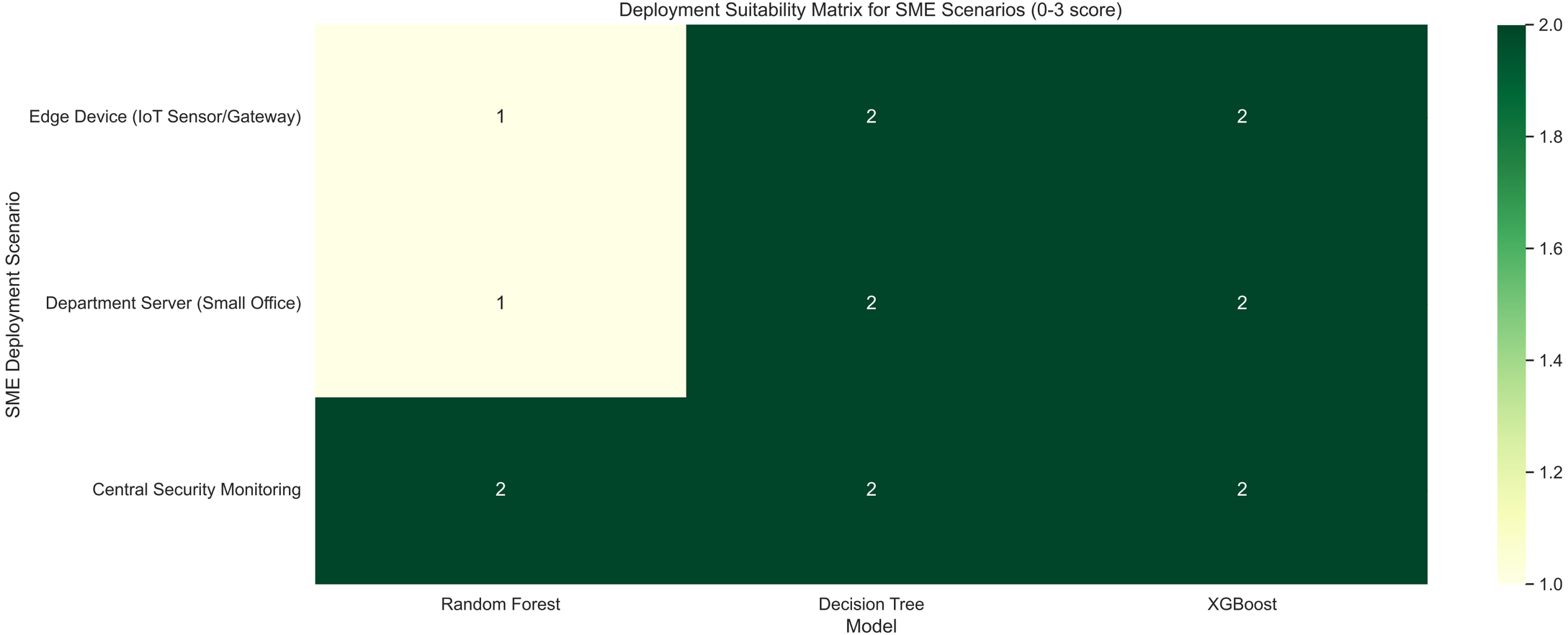
This heatmap visualizes the business impact of classification errors for SME environments through a cost matrix.
The rows represent true labels, while columns show predicted labels, with costs assigned based on the severity of
misclassification. Zero values along the diagonal represent correct classifications with no cost impact.

False negatives (attacks classified as benign traffic) carry the highest cost (5), as they represent security breaches that go undetected, potentially leading to data theft, service disruption, or reputation damage. False
positives (benign traffic classified as attacks) have a moderate cost (2), representing wasted investigation time
and potential business disruption from unnecessary security responses. Misclassification between different attack
types has the lowest cost (1), as the security team would still be alerted to investigate suspicious activity.

For resource-constrained SMEs, this matrix helps prioritize model optimization efforts to minimize costly errors.
It suggests that models should be tuned to minimize false negatives, even at the expense of slightly increased
false positives—a different approach than conventional ML optimization that treats all errors equally. This SME-
specific perspective ensures security resources are allocated to address the most impactful threats.


Normalized Resource Requirements by Model
Model Size (MB)
Training Time (s)
Inference Time (ms)
Normalized Value
1.0
0.8
0.6
0.4
0.2
0.0
Random Forest
Decision Tree
XGBoost


Accuracy vs Model Size (bubble size = inference time)
XGBoost
Decision Tree
Random Forest
Accuracy
0.845
0.840
0.835
0.830
0.825
0
50
100
150
200
250
Model Size (MB)


Deployment Suitability Matrix for SME Scenarios (0-3 score)
SME Deployment Scenario
Edge Device (IoT Sensor/Gateway)
Department Server (Small Office)
Central Security Monitoring
1
2
2
1
2
2
2
2
2
Random Forest
Decision Tree
XGBoost
Model
2.0
1.8
1.6
1.4
1.2
1.0

This comprehensive resource analysis helps SMEs match model capabilities to their infrastructure constraints.
The top-left chart compares normalized resource requirements across models, with Random Forest demanding the
most storage and training time, while Decision Trees offer the lightest footprint. The top-right chart visualizes the accuracy vs. resource tradeoff, with bubble size representing inference time—smaller bubbles
indicate faster prediction speed, ideal for real-time monitoring.

The deployment suitability matrix at the bottom evaluates each model against three common SME scenarios, scoring
them on a 0-3 scale (meeting memory, latency, and accuracy requirements). Decision Trees score highest for
resource-constrained edge devices like IoT sensors and gateways, where quick response and small footprint are
essential despite slightly lower accuracy. Random Forests excel in central monitoring deployments where accuracy
is paramount and resources are less constrained. XGBoost offers a balanced option for departmental servers with
moderate resources.

This analysis provides concrete deployment guidance for SMEs based on their specific infrastructure limitations.
Organizations can implement a tiered detection approach, using lightweight Decision Trees at the network edge for
immediate detection and more resource-intensive Random Forests at the center for comprehensive analysis and
verification, maximizing security effectiveness within existing resource constraints.