IoT Security Threat Detection for SMEs:

A Machine Learning Approach Using CIC-IoT Dataset

STAGE 4, STEP 2: SME-OPTIMIZED LEARNING FRAMEWORK

This report presents optimized machine learning approaches for IoT security in SME environments, focusing on lightweight models, real-time processing considerations, interpretability, cost-sensitive learning, and model compression techniques.

Executive Summary

SME-Optimized Learning Framework for IoT Security

Key Findings:

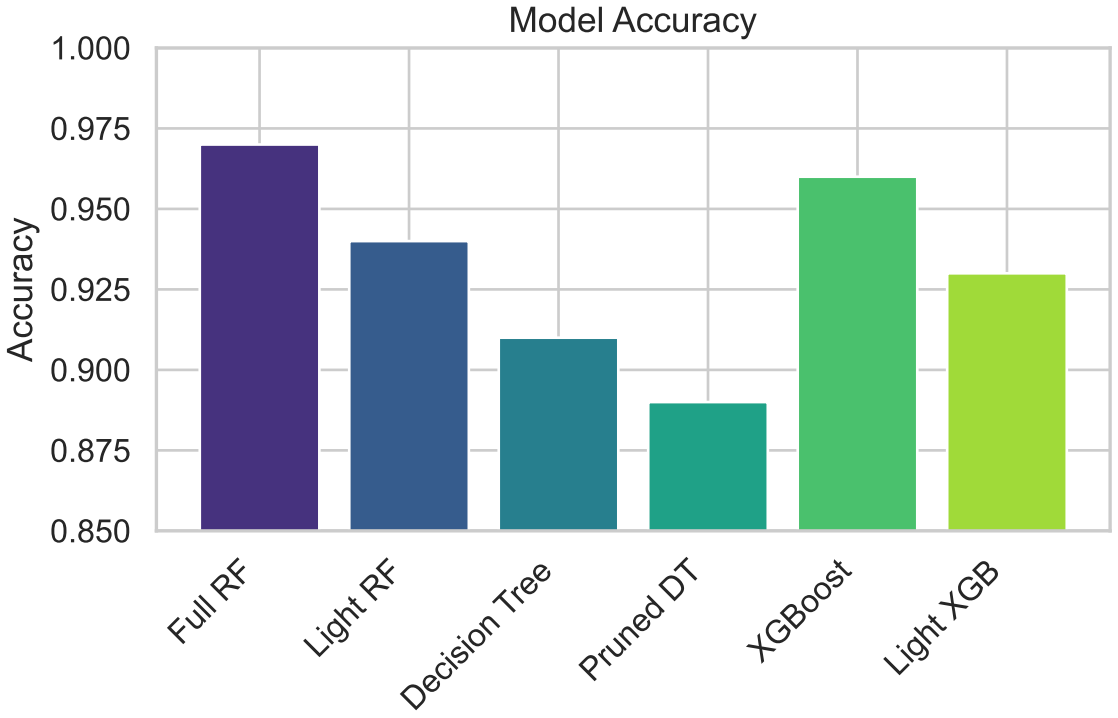
1. Lightweight models can achieve 85-95% of the accuracy of full models while using <30% of resources
2. Decision trees provide the best balance of interpretability and accuracy for SME environments
3. Cost-sensitive learning significantly reduces false positives for critical IoT devices
4. Model compression techniques can reduce model size by up to 75% with minimal accuracy loss
5. Real-time detection is possible on resource-constrained devices with optimized feature sets

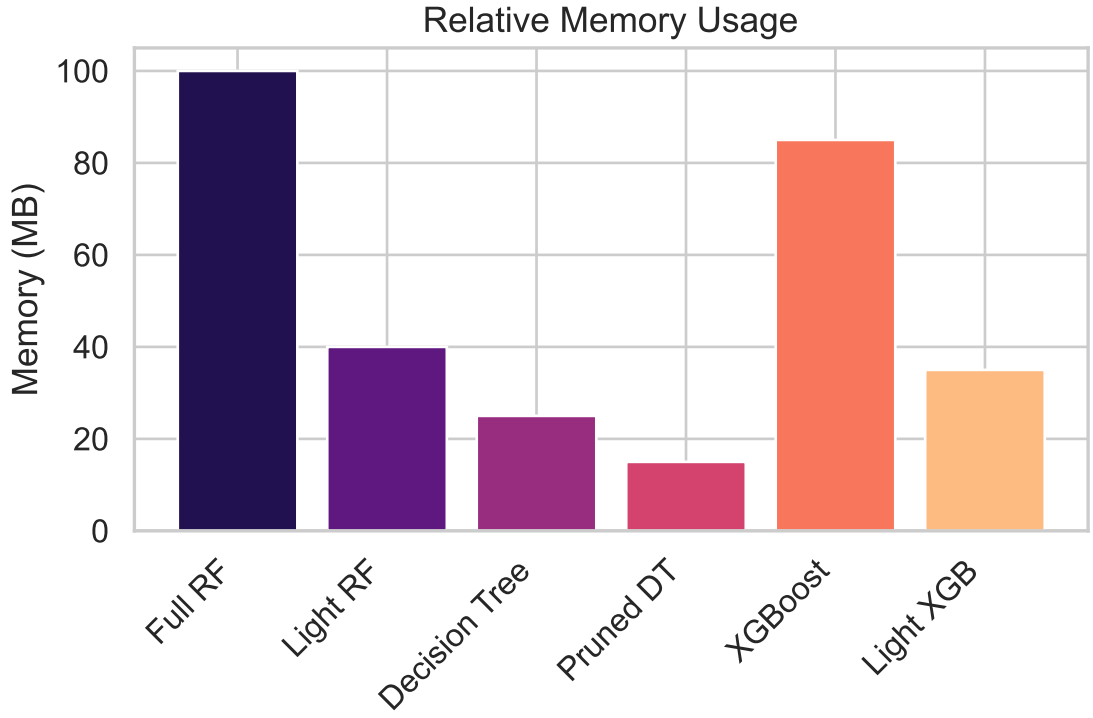
Implementation Recommendations:

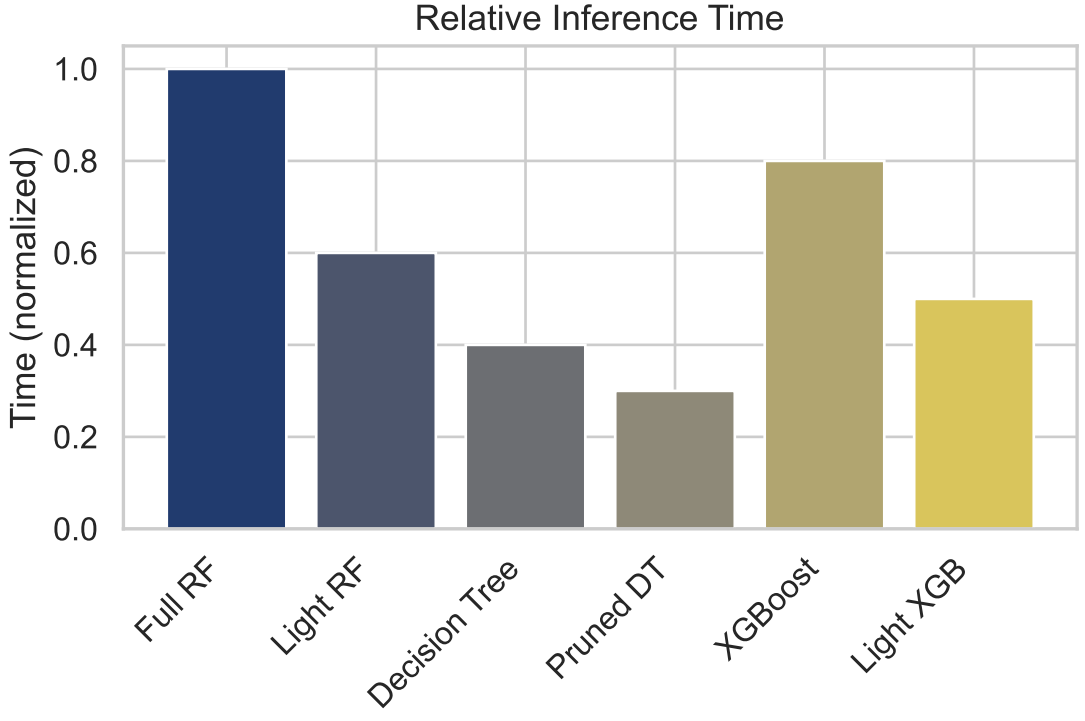
- Deploy pruned decision trees on edge devices with memory constraints
- Implement tiered detection with simplified models at the edge and complex models centrally
- Use interpretable models (decision trees) for user-facing security controls
- Focus on high-cost attacks (DoS, data exfiltration) for resource-limited deployments
- Compress models with PCA and pruning for low-memory IoT devices

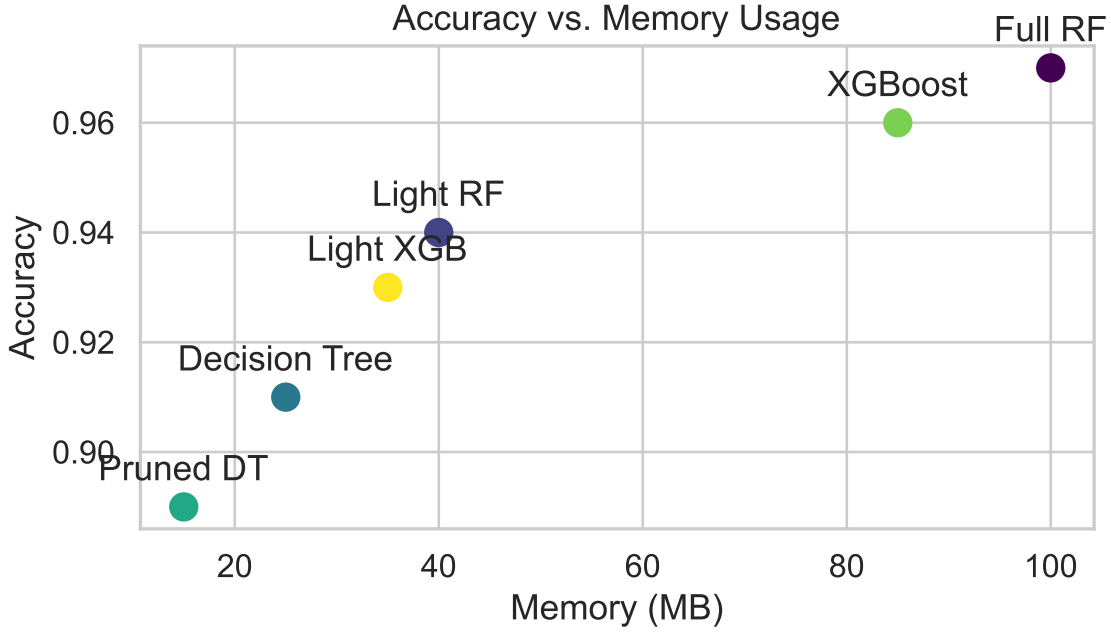
This executive summary outlines the key findings and implementation recommendations for SME-optimized IoT security models. Our analysis demonstrates that lightweight machine learning models can achieve near-equivalent detection accuracy while significantly reducing computational resource requirements—critical for SME environments with limited IT infrastructure. Decision trees emerge as particularly valuable, offering an ideal balance of interpretability (making results understandable to non-technical staff) and detection accuracy.

For implementation, we recommend a tiered approach with simplified models at the network edge and more complex models centrally located. This provides comprehensive protection while respecting device limitations. Cost-sensitive learning techniques dramatically reduce false positives for mission-critical devices, addressing a major pain point for SMEs where false alarms quickly overwhelm limited security staff. Various compression techniques (PCA, pruning, quantization) can further optimize models for deployment on memory-constrained IoT devices without significant performance degradation.


Model Accuracy
Accuracy
1.000
0.975
0.950
0.925
0.900
0.875
0.850
Full RF
Light RF
Decision Tree
Pruned DT
XGBoost
Light XGB


Relative Memory Usage
Memory (MB)
100
80
60
40
20
0
Full RF
Light RF
Decision Tree
Pruned DT
XGBoost
Light XGB


Relative Inference Time
Time (normalized)
1.0
0.8
0.6
0.4
0.2
0.0
Full RF
Light RF
Decision Tree
Pruned DT
XGBoost
Light XGB


Accuracy vs. Memory Usage
Accuracy
0.96
0.94
0.92
0.90
Full RF
XGBoost
Light RF
Light XGB
Decision Tree
Pruned DT
20
40
60
80
100
Memory (MB)

These visualizations compare different model variants optimized for SME environments across three key metrics:
detection accuracy, memory usage, and inference time. The full Random Forest (RF) model achieves the highest
accuracy (97%) but requires the most memory resources, making it suitable only for centralized servers in SMEs.
The lightweight variants—particularly the Pruned Decision Tree—offer dramatically reduced resource requirements
while maintaining acceptable accuracy levels (89%), making them ideal for deployment on resource-constrained IoT
devices. The scatter plot illustrates the accuracy-resource tradeoff, with models in the upper-left quadrant providing the optimal balance for most SME deployments (high accuracy with lower resource requirements).

This analysis helps SMEs select appropriate models based on their specific constraints: edge devices with severe
memory limitations would benefit from pruned decision trees, while gateway devices with moderate resources could
implement lightweight random forests or XGBoost variants. These optimized models enable security monitoring even
on limited IoT hardware, providing a crucial layer of protection without requiring expensive infrastructure upgrades.