Week 2 Model Comparison Report

CyberIDS Research Team

May 24, 2025

1 Objective

The goal of this study is to evaluate multiple binary classification models for network intrusion detection.

2 Models Evaluated

The following models were trained and compared:

- Logistic Regression
- Decision Tree (Pruned)
- Random Forest
- K-Nearest Neighbors (KNN)
- Support Vector Machine (SVM)

3 Evaluation Metrics

The following metrics were used to assess model performance:

- Accuracy
- Precision
- Recall
- F1-Score
- Cross-Validation Score

Model	**Accuracy**	**Precision**	**Recall**	**F1-Score**	**CV Score**
Logistic Regression	0.9997	0.9994	0.99999	0.9997	0.9996
Decision Tree (Pruned)	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**
Random Forest	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**
K-Nearest Neighbors	**1.0**	**1.0**	**1.0**	**1.0**	**1.0**
SVM	0.9998	0.9997	0.99999	0.9998	0.9998

Table 1: Model Evaluation Metrics

4 Key Insights

- **High accuracy across all models** indicates robust classification performance.
- **Possible overfitting** in Decision Tree and Random Forest (achieving perfect scores).
- **Cross-validation confirms generalization** but requires further unseen data validation.

5 Future Improvements

- Investigate **overfitting risks** in Decision Tree and Random Forest.
- Perform **hyperparameter tuning** for KNN and SVM to optimize performance.
- Validate models on a **completely fresh dataset** for real-world effectiveness.

6 Conclusion

This comparison highlights the performance of different classification models for intrusion detection. Tree-based models achieve high accuracy, but their real-world reliability must be further tested. Logistic Regression and SVM provide strong recall, making them effective for detecting malicious activity.

Next Steps:

- Fine-tune hyperparameters for KNN and SVM.

- Validate models with unseen datasets.
- Investigate feature importance for model optimization.

CyberIDS Research Team
May 24, 2025