Network Intrusion Detection Analysis Report

Performance Evaluation of Machine Learning Models on CICIDS2017 Dataset

Generated on: September 14, 2025

This report presents a comprehensive analysis of machine learning models for network intrusion detection using the CICIDS2017 dataset. The analysis compares the performance of four different algorithms in detecting malicious network activities.

Executive Summary

The analysis demonstrates that XGBoost outperforms other models with an accuracy of 96.18% and F1-score of 93.56%. The dataset exhibited significant class imbalance (5.88:1), which was addressed using SMOTE oversampling technique.

367,386	83,322
Training Samples	Testing Samples
30	**5.88:1**
Selected Features	Imbalance Ratio

Methodology

The analysis followed these steps:

1. Data loading and sampling (20% of original data)
2. Preprocessing and handling missing values
3. Binary classification (Benign vs Attack)

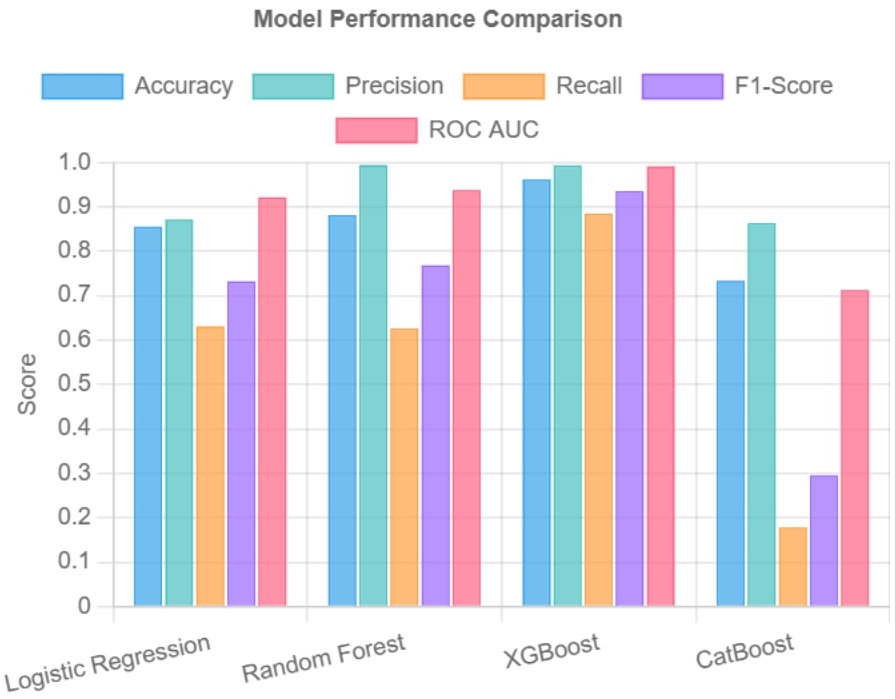
4. Feature selection and dimensionality reduction
5. Addressing class imbalance with SMOTE
6. Model training and evaluation
7. Performance comparison and analysis

Model Performance Comparison

Comprehensive evaluation of four machine learning models

Performance Metrics Across Models

Comparison of accuracy, precision, recall, F1-score, and ROC AUC for all evaluated models



The bar chart illustrates the performance metrics across all evaluated models. XGBoost demonstrates superior performance across all metrics, particularly excelling in accuracy and F1-score.

Detailed Performance Metrics

Model	Accuracy	Precision	Recall	F1-Score	ROC AUC	Training Time (s)
Logistic Regression	0.8554	0.8721	0.6310	0.7322	0.9215	484.88

Random Forest	0.8817	0.9945	0.6261	0.7684	0.9378	227.90
XGBoost	**0.9618**	**0.9932**	**0.8844**	**0.9356**	**0.9912**	**45.62**
CatBoost	0.7336	0.8634	0.1781	0.2952	0.7125	305.44

XGBoost achieved the best performance with the shortest training time, making it the most efficient and effective model for this intrusion detection task.

Confusion Matrix Analysis

Detailed examination of model predictions

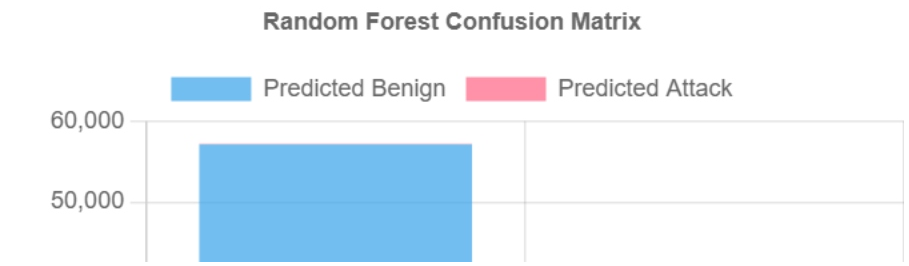
Logistic Regression Confusion Matrix

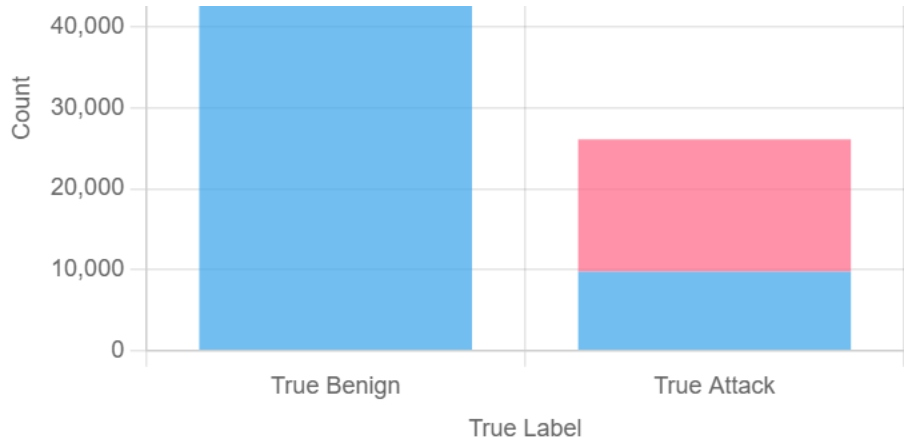
True vs. Predicted labels for Logistic Regression model

The Logistic Regression model correctly identified 54,792 benign samples but misclassified 2,416 as attacks. It detected 16,478 attacks correctly but missed 9,636 attacks (false negatives).

Random Forest Confusion Matrix

True vs. Predicted labels for Random Forest model





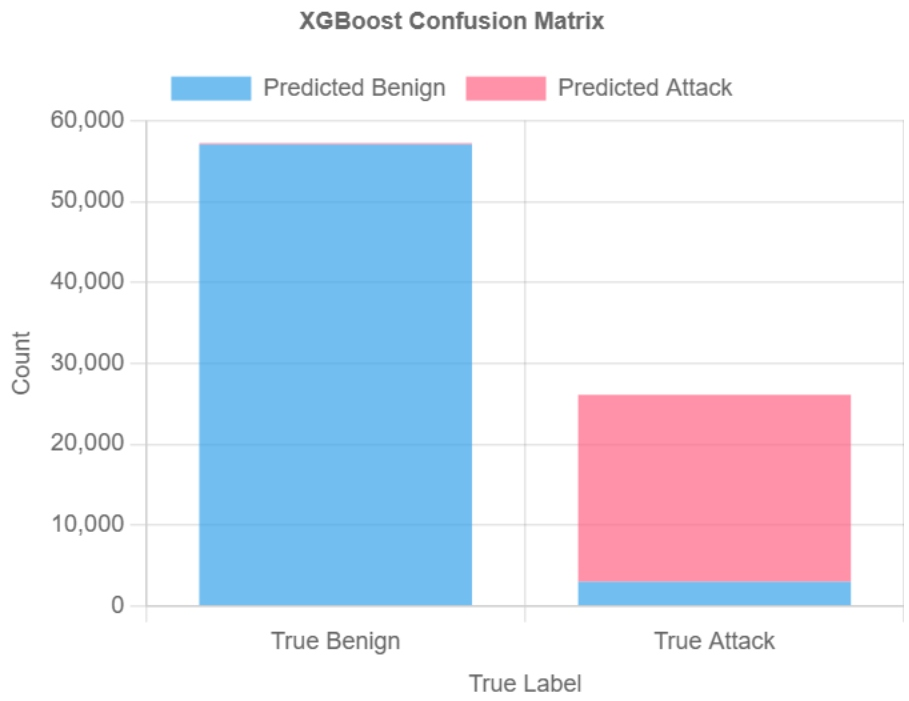
Random Forest showed excellent performance in identifying benign traffic (57,119 correct) but struggled with attack detection, missing 9,765 attacks while correctly identifying 16,349.

Confusion Matrix Analysis (Continued)

Detailed examination of model predictions

XGBoost Confusion Matrix

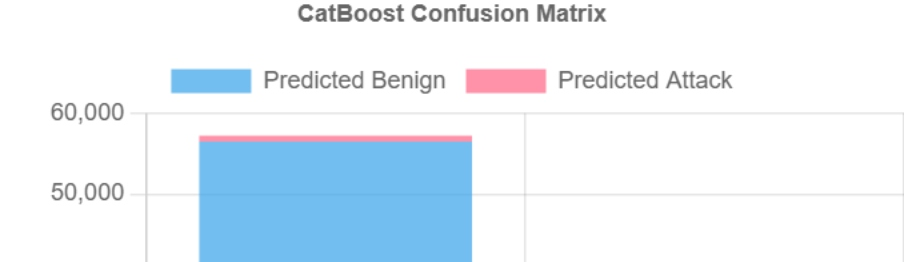
True vs. Predicted labels for XGBoost model

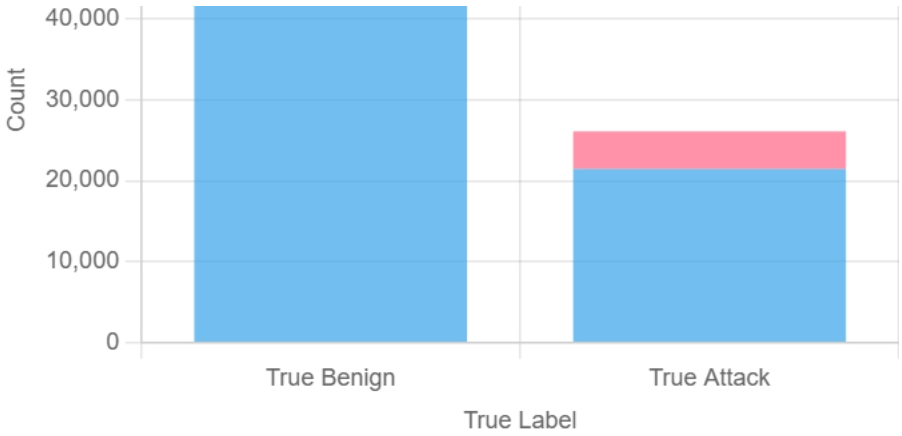


XGBoost demonstrated exceptional performance with 57,046 correctly identified benign samples and 23,097 correctly detected attacks. It had only 162 false positives and 3,017 false negatives.

CatBoost Confusion Matrix

True vs. Predicted labels for CatBoost model





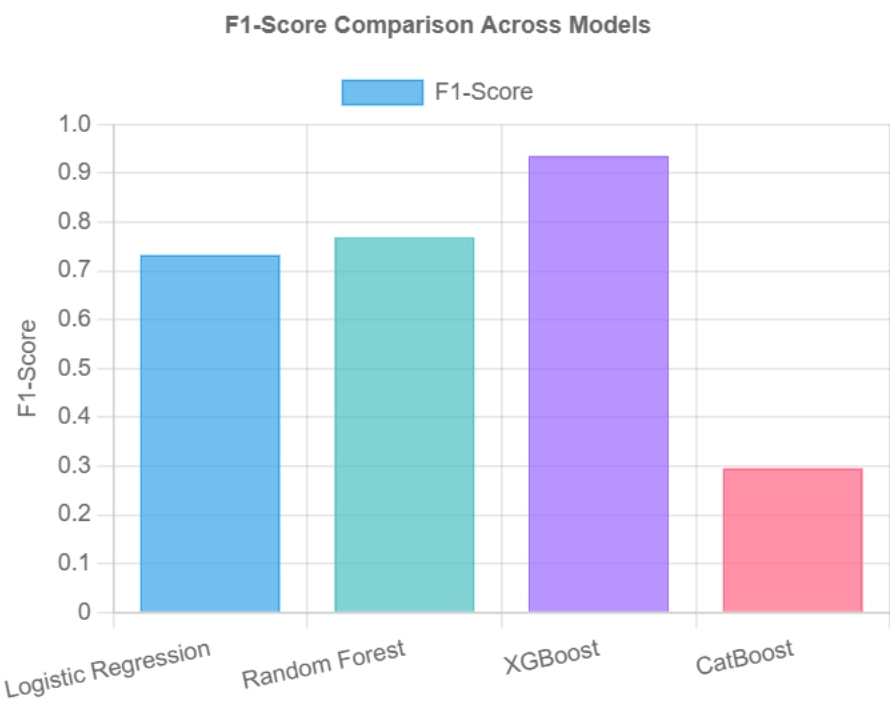
CatBoost struggled with this dataset, particularly with attack detection. It missed 21,465 attacks (false negatives) while correctly identifying only 4,649 attacks.

Performance Summary and Conclusions

Key findings and recommendations

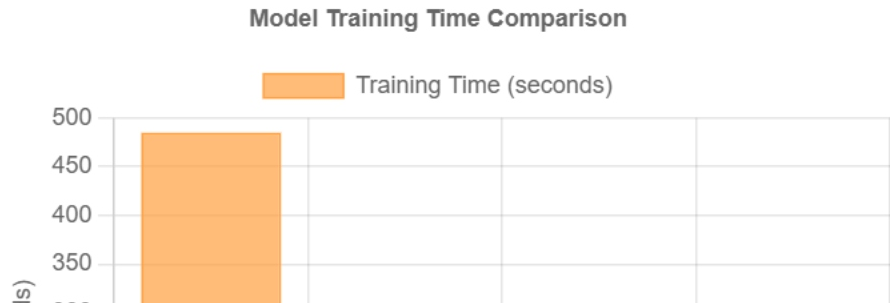
F1-Score Comparison Across Models

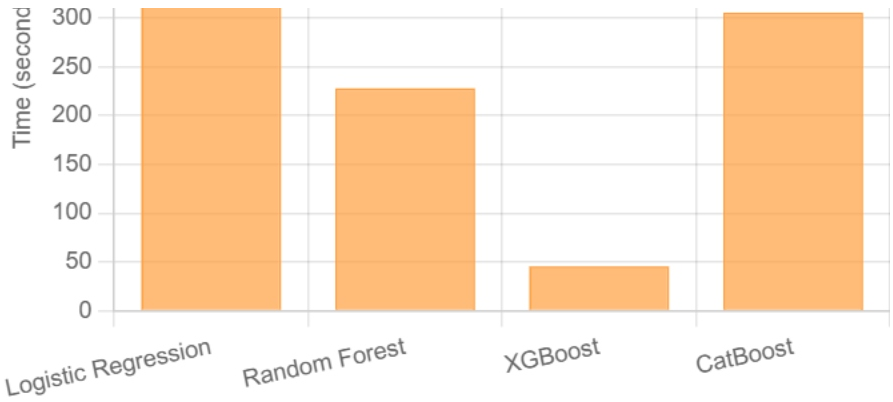
F1-Score values for each model, highlighting the balance between precision and recall



The F1-score comparison clearly shows XGBoost's superiority with a score of 0.9356, significantly outperforming other models. This metric is particularly important for imbalanced datasets like network intrusion detection.

Training Time Comparison





XGBoost not only achieved the best performance but also had the shortest training time (45.62 seconds), making it highly efficient for large-scale intrusion detection systems.

Key Findings

- XGBoost achieved the best overall performance with 96.18% accuracy and 93.56% F1-score
- Random Forest performed well but had limitations in detecting certain attack types
- Logistic Regression provided decent results but was outperformed by tree-based methods
- CatBoost underperformed, possibly due to suboptimal hyperparameter settings
- Class imbalance was successfully addressed using SMOTE oversampling
- Feature selection effectively reduced dimensionality from 79 to 30 features

Recommendations

Based on the analysis, we recommend:

1. **Implement XGBoost** as the primary intrusion detection model due to its superior performance and efficiency
2. **Continue feature engineering** to further improve model performance
3. **Implement real-time monitoring** with the trained model for network security
4. **Regularly update the model** with new data to maintain detection accuracy
5. **Explore ensemble methods** combining XGBoost and Random Forest for potentially even better performance