Network Traffic Anomaly Detection (UNSW-NB15 Dataset)

I. Introduction

The objective of this project was to develop a robust Network Intrusion Detection System (NIDS) using the UNSW-NB15 dataset. The goal was to accurately classify network traffic as either 'Normal' or 'Attack' using machine learning techniques. The project involved extensive data pre-processing, anomaly detection (unsupervised learning), and supervised classification. The outcome is a comprehensive comparison of multiple models to determine the most effective solution for real-world deployment.

II. Data Pre-processing

To ensure high-quality model input, the following steps were taken:

1) **Data Cleaning:** The dataset was inspected for missing values. Imputation strategies (median for numerical, mode for categorical) were established.
2) **Hybrid Encoding:** A custom 'Top-N' strategy was used for high-cardinality categorical features (proto, service, state). The top 5 most frequent categories were One-Hot encoded, while rare categories were grouped as 'Other'. This significantly reduced dimensionality from over 190 potential columns to a manageable 57.
3) **Column Alignment:** A robust alignment step was implemented to ensure the Test set columns exactly matched the Training set, filling missing columns with 0.

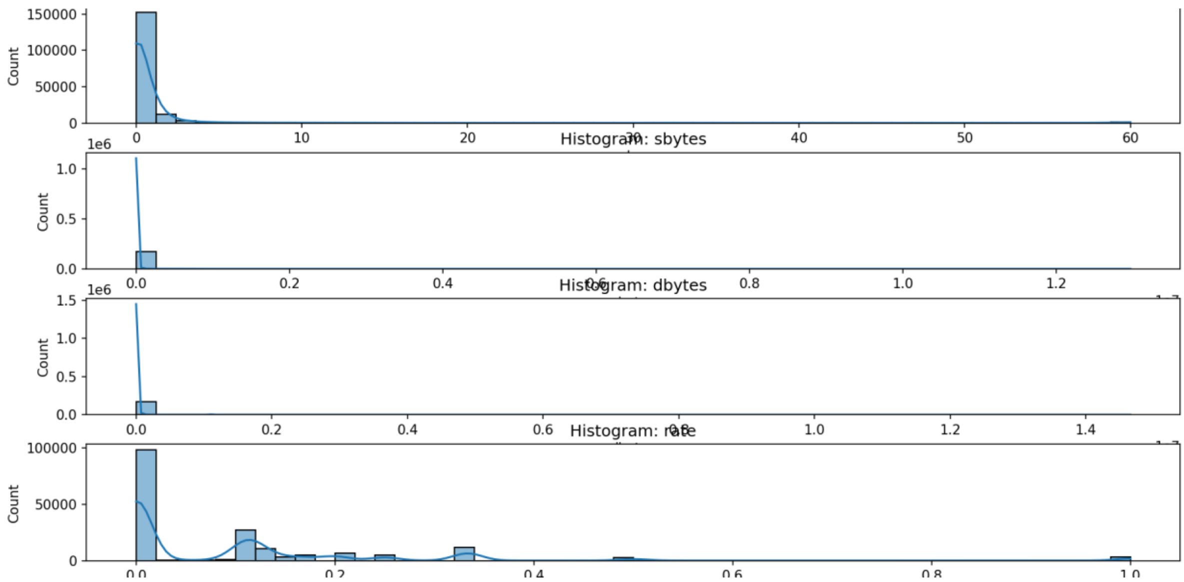

Figure 1: Histogram of feature distribution

We can observe something that is far from a normal distribution, most traffic is short tcp connections with little data being sent.

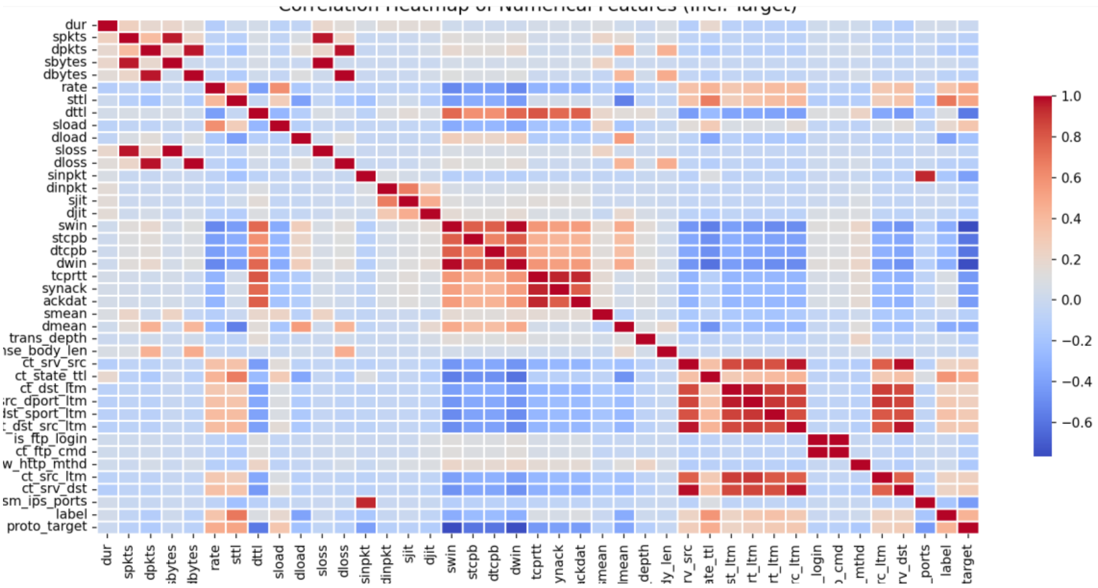

Figure 2: Correlation Matrix

The Correlation Matrix looks pretty good, we do not need to drop any columns, there is no redundance, expect for the one we created via encoding.

III. Modelling

Four distinct modelling paradigms were implemented to explore different approaches to detection:

1. **Isolation Forest (Unsupervised):** Trained **only** on 'Normal' traffic to learn a baseline of legitimate activity. It detects attacks as outliers (anomalies) that deviate mathematically from this baseline.

2. **Neural Autoencoder (Unsupervised):** A deep learning model designed to compress and reconstruct input data. Trained on normal traffic, it flags attacks based on high **Reconstruction Error** (the inability to accurately recreate the unknown attack patterns).

3. **Decision Tree (Supervised):** A rule-based classifier (CART) trained on labelled data. This model was chosen for its high interpretability, allowing analysts to understand the specific rules leading to an alert.

4. **Gaussian Naive Bayes (Supervised):** A probabilistic classifier used as a baseline. It assumes feature independence and required feature scaling (MinMax/StandardScaler) to function correctly.

IV. Evaluation

Models were evaluated based on **Recall (Attack Detection Rate)** and **False Positive Rate (False Alarms)**, prioritizing security.

1) Isolation Forest: Initially low Recall (~30%). After optimizing the anomaly threshold using Precision-Recall curves, Recall improved to **82%**, with ~13,500 False Alarms.

2) Autoencoder: Achieved a balanced performance with **81% Recall** and significantly fewer False Alarms (**~10,300**) compared to the Isolation Forest, making it the superior unsupervised model.

3) Naive Bayes: Achieved high Recall (**96%**) but suffered from an extreme number of False Alarms (~21,900), flagging nearly 60% of normal traffic as malicious.

4) Decision Tree: The standard Decision Tree achieved the highest security with **98% Recall** but initially generated ~11,800 False Positives.

V. Model Tuning

To address the False Positive issue in the Decision Tree, Hyperparameter Tuning was performed using GridSearchCV.

Process: We systematically tested combinations of max_depth (to control complexity) and min_samples_leaf (to force generalization).

Outcome: The optimal parameters were found to be {'max_depth': 20, 'min_samples_leaf': 100}.

Impact: This tuning successfully reduced the number of False Positives by over **2,000** (dropping to ~9,800) without sacrificing the Attack Recall (98%).

VI) Model Selection

Rank	Model	Type	Accuracy	Attack Recall (Security)	False Alarms (Annoyance)	Best Use Case
1	**Decision Tree (Tuned)**	Supervised	**87%**	**98%**	**Low (~9,800)**	**Primary Defense.** Best overall performance.
2	**Autoencoder**	Unsupervised	77%	81%	Moderate (~10,300)	**Zero-Day Detector.** Use alongside Decision Tree to catch new, weird attacks.
3	**Isolation Forest**	Unsupervised	73%	82%	High (~13,500)	Backup to Autoencoder.
4	**Naive Bayes**	Supervised	71%	96%	Extreme (~21,900)	Not recommended for this dataset.

Based on the evaluation metrics, the **Tuned Decision Tree** is selected as the **Champion Model**.

Security: It offers the highest detection rate (98%).

Usability: It has the lowest False Alarm rate of all models tested (~9,800).

Interpretability: The decision logic is transparent and verifiable.

The **Autoencoder** is recommended as a secondary "safety net" to run in parallel, specifically to detect novel Zero-Day attacks that might not match the Decision Tree's learned rules.

7. Conclusion

The project successfully demonstrated that a Supervised Decision Tree, when properly tuned, offers the most effective defense for the UNSW-NB15 dataset. Among unsupervised methods, the Neural Autoencoder proved superior to the Isolation Forest, offering a better balance of accuracy and outlier detection.