Anomaly Detection in Network Traffic Using Autoencoders: A Study on the UNSW-NB15 Dataset

# Abstract

This paper investigates the use of autoencoder architectures for anomaly detection in network traffic, focusing on the UNSW-NB15 dataset. By training an unsupervised model on normal traffic, reconstruction errors were leveraged to identify malicious activity. The proposed approach achieved 95% accuracy and 98% recall for attack traffic, highlighting the utility of autoencoders in modern intrusion detection systems.

# Introduction

Detecting anomalies in network traffic is one of the cornerstones of modern cybersecurity. Attackers are constantly innovating, making it increasingly difficult to rely on static, rule-based detection systems. This is where deep learning methods, particularly autoencoders, prove valuable. Autoencoders can learn the underlying patterns of normal network traffic and highlight deviations that may signal malicious activity.  
  
The UNSW-NB15 dataset provides a realistic and challenging testbed, combining both benign and malicious network flows. By training an autoencoder on normal traffic, we can investigate how well reconstruction errors can separate harmless activity from potential intrusions. This approach contributes to proactive threat detection and can strengthen automated intrusion detection systems.

# Data Preprocessing

The UNSW-NB15 dataset includes a rich set of both categorical and numerical features describing network flows. To prepare the data for training:  
- Categorical features such as protocol, service, and state were one-hot encoded.  
- Numerical features were standardized using StandardScaler to stabilize training.  
- Missing values were handled by replacing them with feature means.  
- Only normal traffic records were used to train the autoencoder in an unsupervised manner.  
  
This preprocessing pipeline ensures that the model receives clean, consistent inputs and that each feature contributes equally during optimization.

# Model Building

We implemented a sequential autoencoder using TensorFlow/Keras. The model design included:  
- Gaussian noise at the input layer to improve robustness.  
- Multiple dense layers with LeakyReLU activations for non-linear representation learning.  
- L1 regularization applied to enforce sparsity in learned features.  
- A bottleneck latent representation serving as the compressed summary of normal network behavior.  
  
Hyperparameters such as hidden layer sizes, latent dimension, learning rate, and L1 strength were tuned using Keras Tuner (Random Search). Early stopping was applied during training to prevent overfitting and reduce unnecessary computation.

# Training

The autoencoder was trained on the normal samples of the UNSW-NB15 dataset, using mean squared error (MSE) as the reconstruction loss. Training proceeded with mini-batch gradient descent, and validation loss was monitored to determine when to stop.  
  
By the end of training, the model had learned to reconstruct normal traffic flows with high fidelity, while failing to accurately reconstruct anomalous traffic — a behavior crucial for anomaly detection.

# Model Evaluation

Evaluation was performed on the full UNSW-NB15 dataset, including both normal and malicious flows. Reconstruction error was calculated for each data point, and an anomaly threshold was set at the 95th percentile of training errors. Any flow exceeding this threshold was flagged as anomalous.  
  
The results were highly promising:  
- Accuracy: 95%  
- Precision (Normal): 100%  
- Precision (Attack): 74%  
- Recall (Normal): 95%  
- Recall (Attack): 98%  
- F1-score: 0.97 (Normal), 0.84 (Attack)  
  
The confusion matrix and classification report confirm the model’s effectiveness. Notably, the high recall for attack traffic (98%) demonstrates that the autoencoder captures the majority of intrusions, which is vital for cybersecurity applications.

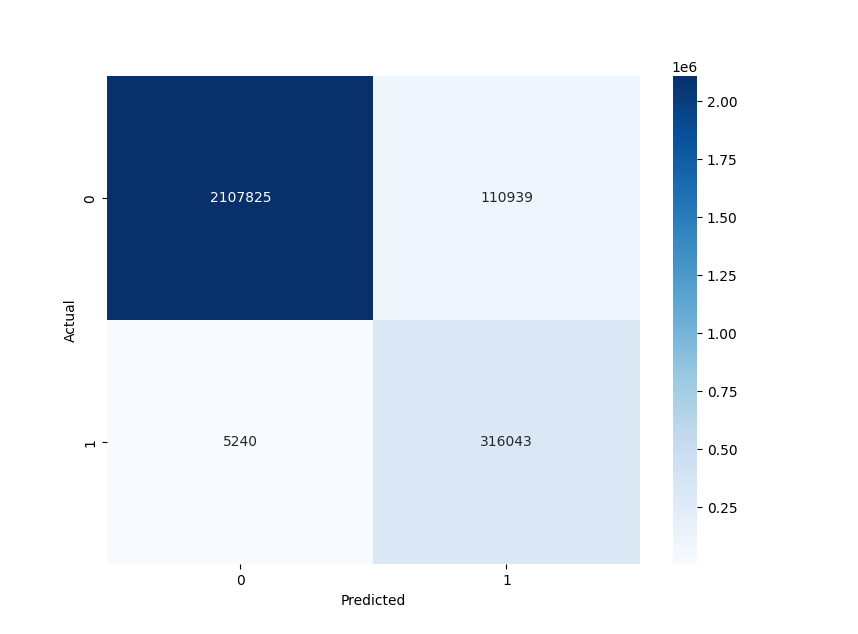


Figure 1: Confusion matrix of anomaly detection results.

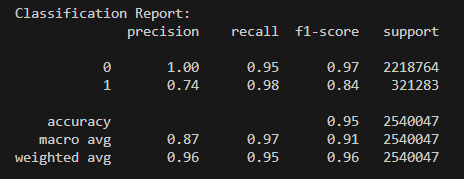


Figure 2: Classification metrics for normal and attack traffic.

# Conclusion

This study demonstrates the effectiveness of autoencoder-based anomaly detection in network intrusion detection using the UNSW-NB15 dataset. Through careful preprocessing, hyperparameter tuning, and robust training, the model achieved:  
- 95% overall accuracy  
- 98% recall for attack traffic, minimizing missed threats  
- Strong separation between normal and anomalous flows  
  
While the model performed well, some challenges remain. Precision on attack traffic was lower (74%), suggesting that some benign flows were misclassified as malicious. Future work could involve refining thresholding methods, exploring hybrid models (e.g., combining autoencoders with supervised classifiers), or extending the approach to other datasets to improve generalization.  
  
Ultimately, this research confirms that autoencoders are a powerful tool for proactive intrusion detection and hold significant promise in bolstering cybersecurity defenses against evolving threats.