International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 10 Issue: 04 | April - 2026 SJIF Rating: 8.659 ISSN: 2582-3930

INTELLIGENT NETWORK TRAFFIC ANOMALY DETECTION USING ML ALGORITHMS

Dr.A.Rajesh[1], Ankolu Chandra Sekhar[2], Badam Lakshmi Prasad[3], Chappidi Sai Kiran[4]
Associate Professor, Department of CSE[1]
Students, Department of CSE[2,3,4]
Guru Nanak Institute of Technology, Hyderabad, Telangana, India

Abstract: With the increased internet services and digital communication, network security has emerged as a significant issue. Conventional rule-based systems of intrusion detection find it difficult to detect emerging cyber threats and thus machine learning is a more viable approach towards detecting anomalies. The system is an innovative network traffic anomaly detection system based on the latest machine learning algorithms. This benchmark dataset is the KDD Cup 1999 dataset, which includes both malicious and normal network traffic records. Preprocessing of data, feature selection, and training of models is done to enhance the accuracy and efficiency of detection. Several algorithms such as Decision Tree, Random Forest, Gradient Boosting, and CatBoost are compared in terms of performance. CatBoost has the highest accuracy of more than 99% on test data, performing better than the techniques in the base paper. The system is also suitable in the real world by developing a Flask-based web application to perform user authentication, prediction of anomalies, and visualization of performance.

Keywords: Network Traffic Anomaly Detection, Machine Learning, Intrusion Detection System, CatBoost, KDD Cup 1999 Dataset, Cyber Security, Network Security.

I.INTRODUCTION

As the internet services and digital communication rapidly increases, network security has become a significant issue. Conventional rule based intrusion detection systems are not effective in detecting changing and unknown cyber-attacks. Machine learning methods are extensively employed to detect intelligent anomalies in network traffic in order to overcome this limitation. The application is dedicated to the creation of an intrusion detection system, based on machine learning, which can sort network traffic by classifying it as regular or malicious. It uses the KDD Cup 1999 dataset, which is a benchmark dataset, as it consists of normal and attack traffic records. Decision Tree, Random Forest, Gradient Boosting, and CatBoost are some of the algorithms that are analyzed in terms of performance. CatBoost is among them which attains better accuracy and effectively manages categorical and numerical variables.

II. LITERATURE SURVEY

S. Wang, Y. Zhang, and X. Chen (2021) The paper provides a review of machine learning algorithms in network anomaly detection including supervised, unsupervised, and deep learning models. It addresses questions like scalability, class imbalance and interpretability and emphasizes the contribution of boosting algorithms to enhance detection accuracy.

Aswathy and Rajkumar (2024) This article provides a comparative study of CatBoost, LightGBM, and XGBoost in detecting anomalies in real-time network traffic. Of these, CatBoost was more accurate, precise, recalled and inferred faster, and it can be scaled to more cybersecurity applications.

III. EXISTING SYSTEM

The current system applies the classic machine learning models like Decision Tree, Random Forest, and Logistic Regression in the detection of anomalies within the network traffic. These models categorize traffic as normal or malicious according to historical patterns, and predefined features. Although they are reasonably

effective in the case of known attacks, they need large amounts of preprocessing and they tend to become ineffective with large-scale and high-dimensional network data.

Existing System Disadvantages

- Prone to overfitting
- High preprocessing complexity
- Mishandling of categorical features
- Reduced support with large datasets
- Increased training and testing period
- Poor flexibility to new attacks.

Proposed System

The system is based on an intelligent network traffic anomaly detection using CatBoost, a state-of-the-art gradient boosting machine learning algorithm. It can easily numerically and categorically process both without a lot of preprocessing. The model has been found to be accurate in categorizing traffic as normal and attack with the KDD Cup 1999 dataset. It is optimized to boost architecture giving it high accuracy and fast processing and is thus applicable in real-time cybersecurity applications.

Proposed System Advantages:

- High detection accuracy
- Better generalisation
- Faster training performance
- Reduced preprocessing effort
- Real-time anomaly detection
- Scalable and practical

IV. SYSTEM ARCHITECTURE

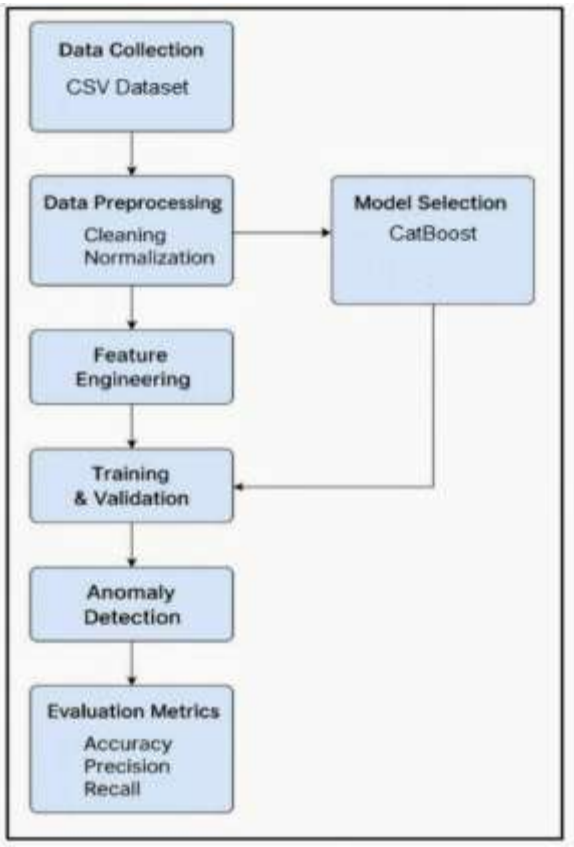

Figure 1 : SYSTEM ARCHITECTURE

The architecture of the system is based on the detection of network anomalies, which is efficient with the help of the CatBoost algorithm. It involves data gathering, data preprocessing, feature engineering, training a model and classifying anomalies to distinguish between normal and malicious traffic.

Methodology

1. Data Collection:

This module gathers network traffic data based on KDD Cup 1999 dataset which provides both normal and malicious traffic records with various categories of attacks. Training, validation and testing of the proposed anomaly detection model are done on the dataset.

2. Data Preprocessing:

This module will cleanse and normalize data, treat missing values and encode categorical features. This is because these steps enhance the quality of data, and the efficiency in the learning of the CatBoost model.

3. Feature Engineering:

The selected and transformed attributes of traffic include protocol type, service, source bytes, and destination bytes, which are all important attributes. The feature selection will decrease the complexity and enhance prediction accuracy.

4. Model Selection (CatBoost):

This module uses CatBoost, an advanced gradient boosting algorithm, for anomaly detection. The fact that it is efficient in working with both numerical and categorical data qualifies it to be largely viable in intrusion detection activities.

5. Model Training & Validation:

This module trains the CatBoost model on the basis of the supervised learning and tests its performance through testing data. Hyperparameter optimization is carried out with an aim to enhance accuracy and decrease overfitting.

6. Anomaly Detection:

The trained CatBoost model is used in this module to either identify normal or malicious network traffic in real-time to allow intrusion detection to occur.

7. Evaluation Metrics:

The Accuracy, Precision, Recall, F1-Score, and Confusion Matrix are used to assess the model in this module which guarantees the reliability of performance measurement.

8. Deployment:

This module deploy the trained CatBoost model as a Flask-based web application with prediction and visualization of real-time cybersecurity monitoring.

V. IMPLEMENTATION

The implementation stage transforms the suggested anomaly detection system into a functioning system. It consists of preprocessing the KDD Cup 1999 dataset, training the CatBoost model and deploying it to a Flask-based web application to classify traffic in real time.

Algorithm Used

Existing Algorithm

Random Forest:

Random Forest is used to classify the network traffic in the current system. Whereas it has a good level of accuracy, it has a lot of preprocessing and it exhibits lower scalability in large dataset.

Proposed Algorithm

CatBoost;

CatBoost is applied in the efficient detection of anomalies with high accuracy and less preprocessing. It is appropriate to do real-time intrusion detection because it can deal with both categorical and numerical features.

VI. EXPERIMENTAL RESULTS

HOME PAGE:

Figure 2 : HOME PAGE

The Home Page is the primary landing page of the application that has the following navigation options: Register, Login, Predict, and Charts, and a short description of the project purpose.

PREDICTION PAGE:

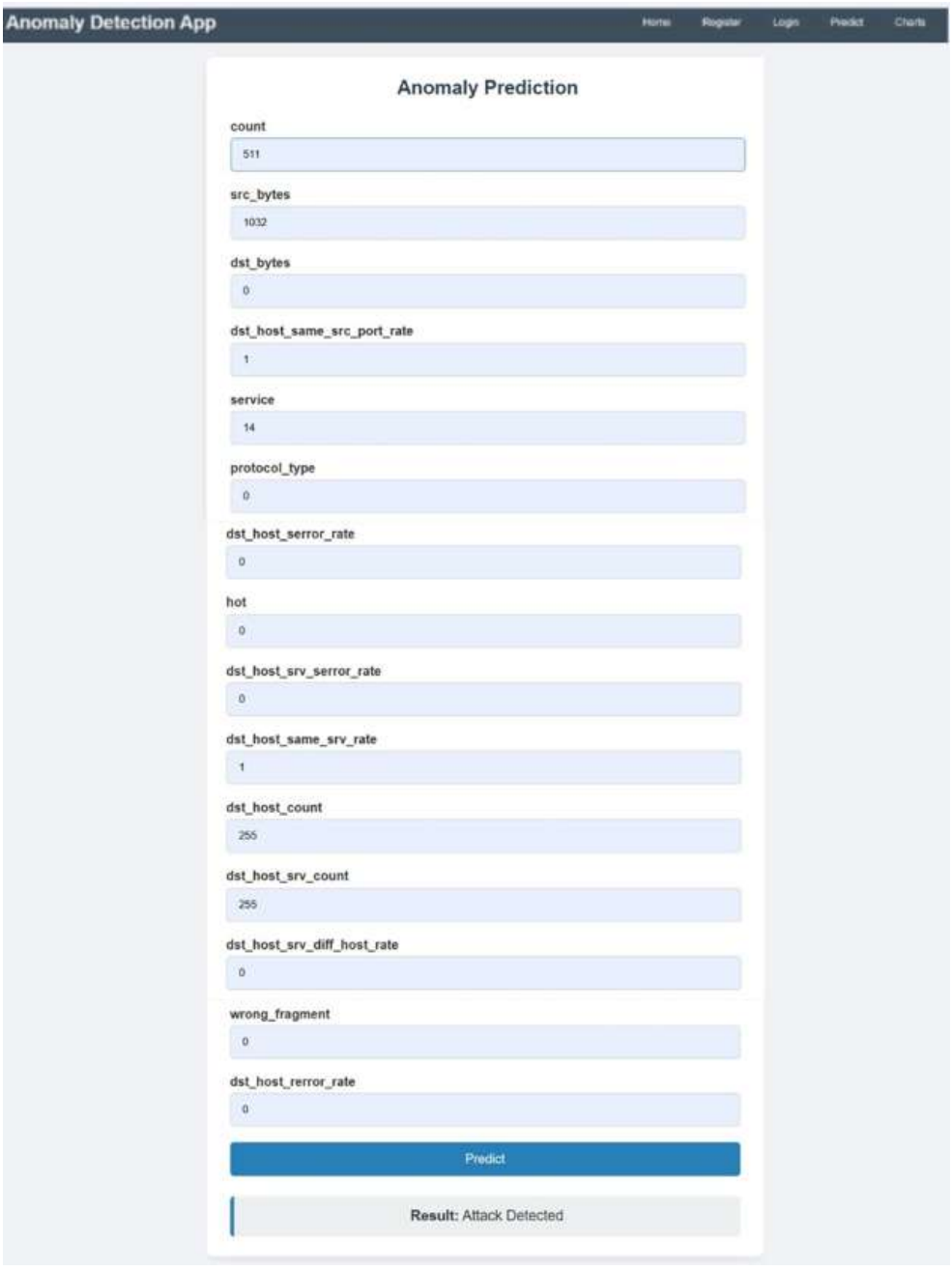

Figure 3: PREDICTION PAGE

The Prediction Result window confirms that the developed system is able to detect malicious network attacks in real time since the system classifies the provided information as “Attack Detected.

MODEL CHARTS PAGE :

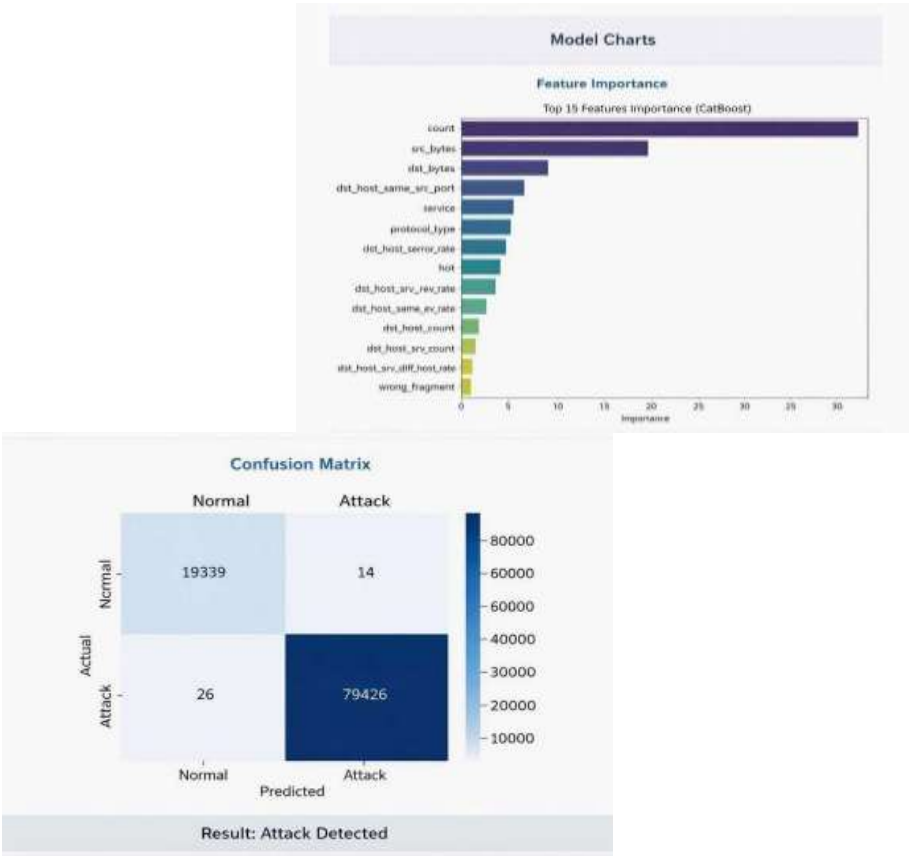

Figure 4: MODEL CHARTS PAGE

Model Charts includes feature importance chart and confusion matrix to prove the high accuracy and low number of false classification between normal and attack traffic types.

VII. CONCLUSION

An Intelligent Network Traffic Anomaly Detection system which can effectively classify traffic data into normal or attack categories using the CatBoost classification model with accuracy exceeding 99% has been developed. Real-time prediction is achieved through the development of web application based on Flask technology.

VIII. FUTURE ENHANCEMENT

Future enhancement of the anomaly detection system may involve integration of deep learning algorithms like CNNs, RNNs, and Autoencoders to ensure improved detection capabilities and efficiency. Further developments in the framework may enable real-time traffic analysis, online learning capabilities, and periodic training of the predictive model to cope with new attack types. Other improvements may include cloud/IoT deployment, detection of zero-day attacks, advanced dashboards, and immediate email/mobile alerts.

REFERENCES

[1] S. Wang, Y. Zhang, and X. Chen, “Machine learning in network anomaly detection: A survey,” IEEE Access, vol. 9, pp. 116725–116745, 2021.

[2] X. Liu, Y. Li, and L. Sun, “Research on anomaly network detection based on self-attention mechanism,” Sensors, vol. 23, no. 10, pp. 5123–5138, 2023.

[3] A. Gupta, R. Singh, and P. Sharma, “Robust anomaly detection in network traffic,” arXiv preprint, 2025.

[4] Y. Zhou, H. Yang, and J. Chen, “Network traffic anomaly detection via deep learning,” Information, vol. 12, no. 5, pp. 1–15, 2021.

[5] N. Shone, T. N. Ngoc, V. D. Phai, and Q. Shi, “A deep learning approach to network intrusion detection,” IEEE Transactions on Emerging Topics in Computational Intelligence, vol. 2, no. 1, pp. 41–50, 2018.