International Journal of Advance Research in Science and Engineering
Volume. No. 15, Issue No. 05, May, 2026
ISSN: 2319-8354

Anomaly Detection in Network Traffic Using Deep Learning

Mr. K. Vinod Kumar
Assistant Professor Department of IT
Tirumala Engineering College
Andhra Pradesh, India
vinodkorsipati07@gmail.com

Shaik Sameera Jasmine
Department of IT
Tirumala Engineering College
Andhra Pradesh, India
shaiksameerajasmine9@gmail.com

Gummalla Nagalakshmi
Department of IT
Tirumala Engineering College
Andhra Pradesh, India
gummallanagalakshmi59@gmail.com

Jetty Pavan Sai
Department of IT
Tirumala Engineering College
Andhra Pradesh, India
pavansaijetty88@gmail.com

Annapareddy Puneeth Vijay
Department of IT
Tirumala Engineering College
Andhra Pradesh, India
punithvijay117@gamil.com

Abstract **- In recent years, with the enormous explosion of web based learning resources, personalization has become a critical factor for the success of services that wish to leverage the power of Web 2.0. However, the relevance, significance and impact of tailored content delivery in the learning domain is still questionable. Apart from considering only interaction based features like ratings and inferring learner preferences from them, if these services were to incorporate innate user profile attributes which affect learning activities, the quality of recommendations produced could be vastly improved. Recognizing the crucial role of effective guidance in informal educational settings, we provide a principled way of utilizing multiple sources of information from the user profile itself for the recommendation task. We explore factors that affect the choice of learning resources and explain in what way are they helpful to improve the pedagogical accuracy of learning objects recommended. Through a systematical application of machine learning techniques, we further provide a technological solution to convert these indirectly mapped learner specific attributes into a direct mapping with the learning resources. This mapping has a distinct advantage of tagging learning resources to make their metadata more informative. The results of our empirical study depict the similarity of nominal learning attributes with respect to each other. We further succeed in capturing the learner subset, whose preferences are most likely to be an indication of learning resource usage. Our novel system filters learner profile attributes to discover a tag that links them with learning resources.**

Keywords - Apriori principle, Clustering, NMF, Learning Pathway, Pedagogical accuracy, Technology enhanced learning.

I. INTRODUCTION

This project focuses on crime rate analysis and prediction using machine learning techniques applied to real-world crime datasets. The system utilizes data preprocessing and feature selection methods to extract meaningful information from raw crime records, thereby improving data quality and model efficiency. Various machine learning algorithms are trained and evaluated to accurately classify and predict different types of crimes. To enhance usability, a web-based interface is developed that allows users to input data and obtain crime predictions in real time. In addition, visualization techniques such as charts and graphs are employed to represent crime trends and patterns effectively, enabling the identification of crime-prone areas and critical time periods.

With the rapid advancement of modern technologies, criminals are increasingly adopting sophisticated methods to carry out illegal activities, resulting in a continuous rise in crime rates. According to reports from the Crime Records Bureau, crimes such as burglary, arson, murder, sexual assault, and gang-related offenses have shown a significant increase in recent years. Crime data collected from multiple sources, including news portals, online blogs, and official government websites, is stored in a centralized database for analysis. By applying data mining and machine learning techniques to this dataset, the proposed system assists law enforcement agencies in detecting crimes more efficiently, identifying affected regions, and locating crime hotspots with high concentrations of criminal activities. This approach supports proactive decision-making and enhances public safety through intelligent crime prediction and analysis.

A. Project Overview

This project focuses on network anomaly detection using machine learning and deep learning techniques. Data preprocessing and feature selection are applied to extract meaningful patterns from network traffic data. A Feed Forward Neural Network (FFNN) model is used to classify traffic as normal or anomalous using benchmark datasets such as KDDCup99 and UNSW-NB15. Visualization techniques are also used to analyze traffic patterns and model performance.

With the rapid growth of cloud computing, IoT, and modern network systems, cyber threats such as DDoS attacks, malware, and unauthorized access have increased significantly. Traditional intrusion detection systems are no longer effective in handling complex and evolving threats. Therefore, the proposed system uses data mining and deep learning techniques to detect anomalies efficiently, identify suspicious traffic patterns, and enhance real-time network security.

B. Problem Definition

The rapid growth of network-based systems, including cloud computing, Internet of Things (IoT), and large-scale digital infrastructures, has led to a significant increase in the volume and complexity of network traffic. This has created major challenges in ensuring network security, as modern cyberattacks such as Distributed Denial of Service (DDoS), malware intrusions, and unauthorized access are becoming more frequent and sophisticated.

Traditional intrusion detection systems, which rely on signature-based or rule-based techniques, are limited in their ability to detect unknown or evolving threats. These systems often produce high false positive rates and fail to identify zero-day attacks, making them ineffective in dynamic network environments. Additionally, handling high-dimensional network traffic data and maintaining scalability in real-time scenarios remain critical issues.

Therefore, there is a need to develop an intelligent and efficient anomaly detection system that can accurately identify both known and unknown threats in network traffic. The system should be capable of processing large-scale data, reducing false alarms, and providing reliable detection performance. This project addresses these challenges by proposing a machine learning and deep learning-based approach using a Feed Forward Neural Network (FFNN) to improve anomaly detection accuracy and enhance network security.

C. Objectives

The main objectives of the proposed system are:

- To develop an effective anomaly detection system for network traffic.
- To preprocess and optimize network data for improved accuracy.
- To implement a Feed Forward Neural Network (FFNN) for classifying traffic as normal or anomalous.
- To detect cyber threats with high accuracy while reducing false alarms.

To evaluate and analyze model performance using standard metrics.

II. LITERATURE REVIEW

Several studies have explored the use of machine learning and deep learning techniques for network anomaly detection to improve cybersecurity systems. Chandola et al. [1] provided a comprehensive overview of anomaly detection methods, emphasizing statistical and machine learning approaches for identifying unusual patterns in large datasets. However, their work mainly focused on theoretical concepts and lacked practical implementation for real-time network environments.

Ahmed et al. [2] analyzed various network anomaly detection techniques and highlighted key challenges such as high false positive rates, scalability issues, and the difficulty of handling high-dimensional data. Their study showed that traditional approaches are often insufficient for modern network traffic analysis.

Liu et al. [3] introduced the Isolation Forest algorithm, which isolates anomalies using random partitioning of data. Although this method is effective for detecting outliers, its performance decreases when applied to high-dimensional datasets without proper feature engineering.

Usama et al. [4] explored unsupervised machine learning techniques for networking applications, demonstrating their ability to detect anomalies without requiring labeled data. However, these methods often lack classification accuracy and struggle to distinguish between different types of attacks.

With the advancement of deep learning, more sophisticated models have been developed for anomaly detection. Naseer et al. [5] proposed a deep neural network-based model that significantly improved detection accuracy compared to traditional methods. Similarly, Fotiadou et al. [6] utilized deep learning techniques to analyze network traffic patterns and detect complex cyber threats.

Hybrid approaches have also been introduced to enhance performance. Jihado and Girsang [7] proposed a model combining Convolutional Neural Networks (CNN) and Bi-directional Long Short-Term Memory (Bi-LSTM), which improved detection accuracy by capturing both spatial and temporal features. However, these models require high computational resources and increased training time.

III. EXISTING SYSTEM

Existing anomaly detection systems primarily rely on traditional machine learning and rule-based intrusion detection techniques to identify abnormal network activities. Methods such as Naïve Bayes, Support Vector Machine (SVM), Decision Trees, and Isolation Forest are commonly used to detect anomalies in network traffic. These approaches analyze historical data and classify traffic based on predefined patterns or statistical rules.

Many existing systems also use clustering and unsupervised learning techniques to detect anomalies without labeled data. While these methods can identify unusual behavior, they often lack accuracy in distinguishing between normal and malicious

traffic. Additionally, traditional intrusion detection systems are mainly signature-based, which limits their ability to detect unknown or evolving cyber threats.

Although these systems provide basic anomaly detection capabilities, they face several limitations. They struggle to handle high-dimensional and large-scale network data, resulting in reduced efficiency and scalability. High false positive rates are another major issue, leading to incorrect classification of normal traffic as malicious. Furthermore, most existing systems are not suitable for real-time detection and require significant manual intervention for feature selection and tuning.

A. Challenges in Existing System

Despite the availability of various anomaly detection techniques, existing systems face several critical challenges that limit their effectiveness in real-world network environments.

- **High-dimensional data complexity:** Network traffic data contains a large number of features, making it difficult for traditional models to process efficiently.
- **High false positive rate:** Many systems incorrectly classify normal traffic as anomalous, leading to unnecessary alerts and reduced reliability.
- **Inability to detect unknown attacks:** Signature-based and rule-based systems fail to identify new or evolving cyber threats such as zero-day attacks.
- **Scalability issues:** Existing methods struggle to handle large-scale and rapidly growing network traffic data.
- **Lack of real-time detection:** Most systems rely on ofline analysis and are not capable of detecting anomalies in real time.
- **Data quality issues:** Network datasets often contain missing values, noise, and inconsistencies, which affect model accuracy.
- **Computational overhead:** Some machine learning models require high computational resources and extensive tuning.

B. System Architecture Overview

The system architecture of the proposed anomaly detection framework is designed in a sequential and modular manner to ensure efficient processing of network traffic data. The architecture begins with the data collection module, where benchmark datasets such as KDDCup99 and UNSW-NB15 are used as input. These datasets contain both normal and malicious traffic records with features such as protocoltype, source bytes, destination bytes, and connection duration.

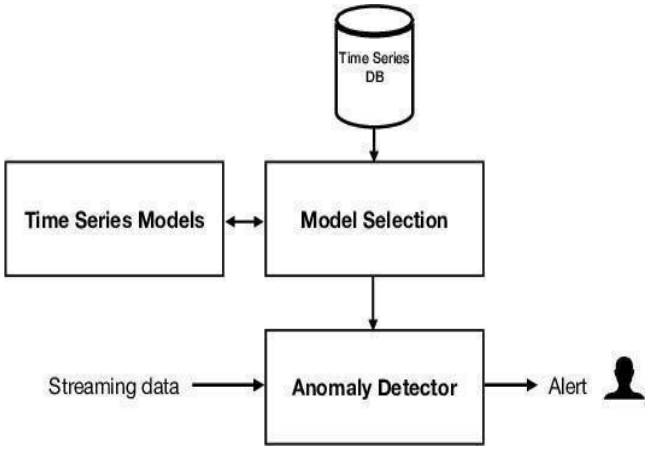

Fig1: Anomaly Detection

The next stage is the data preprocessing module, where raw network traffic data is cleaned and transformed into a suitable format for model training. In this stage, missing values and duplicate records are removed, categorical features are encoded into numerical values, and numerical data is normalized. This improves data quality and ensures consistency across all input features.

After preprocessing, the data is passed to the feature selection module, where only the most relevant attributes are retained. This step reduces dimensionality, eliminates redundant information, and improves computational efficiency. Techniques such as correlation analysis and Principal Component Analysis (PCA) can be used for this purpose.

The refined feature set is then provided to the FFNN model module, which is the core component of the system. The Feed Forward Neural Network consists of an input layer, one or more hidden layers, and an output layer. The hidden layers learn complex non-linear relationships in the network traffic data, while the output layer classifies the traffic as either normal or anomalous.

Finally, the output and evaluation module generates the classification results and performance metrics such as accuracy, precision, recall, and F1-score. Visualization tools such as confusion matrix, ROC curve, and accuracy graphs are used to analyze the effectiveness of the model. This overall architecture enables reliable, scalable, and real-time anomaly detection in modern network environments.

C. Motivation for Proposed System

The motivation for the proposed system arises from the limitations of existing anomaly detection methods, such as high false positives, inability to detect unknown attacks, and lack of real-time capability. With the rapid growth of network traffic and increasing cyber threats, there is a need for an intelligent and scalable solution. The proposed FFNN-based system aims to improve detection accuracy,

reduce false alarms, and enhance network security through efficient anomaly detection.

D. Disadvantages of Existing System

The major drawbacks of the existing systems include:

- Requires high computational resources for training the model.
- Training time increases with large-scale network datasets.
- Model interpretability is low compared to traditional methods.
- Performance depends on quality of preprocessing and feature selection.
- May require frequent updates to adapt to new cyber threats.

IV. Proposed System

The proposed system uses a deep learning-based approach for network anomaly detection using a Feed Forward Neural Network (FFNN). The system is designed to classify network traffic as normal or anomalous by learning hidden patterns from network data. It overcomes the limitations of traditional intrusion detection systems, such as poor scalability, high false positive rates, and inability to detect unknown attacks.

In the proposed system, network traffic data is collected from benchmark datasets such as KDDCup99 and UNSW-NB15. The data is then preprocessed by removing missing values and duplicate records, encoding categorical features, and normalizing numerical attributes. Feature selection techniques are applied to reduce dimensionality and improve model efficiency.

The processed data is passed to the FFNN model, which consists of input, hidden, and output layers. The hidden layers learn complex non-linear relationships in the traffic data, while the output layer performs binary classification to identify whether the traffic is normal or anomalous. The model is trained using backpropagation and evaluated using performance metrics such as accuracy, precision, recall, and F1-score.

The proposed system provides improved detection accuracy, better scalability, and reduced false alarm rates, making it suitable for real-time network security applications.

V. DEEP Learning Algorithms Used

This section describes the machine learning and deep learning algorithm employed in the proposed network anomaly detection system. A Feed Forward Neural Network (FFNN) model is used to learn patterns from network traffic data and accurately classify it as normal or anomalous.

A. Feed Forward Neural Network (FFNN)

A Feed Forward Neural Network (FFNN) is a deep learning model used for classification and prediction tasks. It consists of multiple layers, including an input layer, one or more hidden layers, and an output layer. In this network, data flows in one direction from the input layer to the output layer without forming cycles, allowing the model to learn complex non-linear relationships between input features.

In the proposed system, the FFNN model is trained using network traffic features such as protocol type, source bytes, destination bytes, and connection duration. The model processes the input data through hidden layers, where each neuron applies weighted computations followed by activation functions to learn patterns in the data. The output layer performs binary classification to determine whether the traffic is normal or anomalous.

The FFNN model is trained using the backpropagation algorithm, which adjusts weights by minimizing the error between predicted and actual outputs. This enables the model to improve its accuracy over time. The use of FFNN helps in capturing hidden patterns in high-dimensional network data, making it effective for detecting both known and unknown cyber threats.

B. Working of FFNN Model

The FFNN model improves anomaly detection performance by learning complex relationships within network traffic data. It reduces the need for manual feature engineering and provides better generalization compared to traditional machine learning models. Additionally, the model is capable of handling large-scale datasets and adapting to dynamic network environments.

The experimental results show that the FFNN model achieves high accuracy, precision, recall, and F1-score, demonstrating its effectiveness in anomaly detection. Its ability to reduce false positives and false negatives makes it suitable for real-time network security applications.

C. Advantages of Proposed System

The proposed system offers several advantages:

- Provides high accuracy in detecting network anomalies.
- Identifies both known and unknown cyber threats.
- Reduces false positive and false negative rates.
- Handles large-scale network data efficiently.
- Suitable for real-time anomaly detection..

VI. Results and Discussion

The proposed anomaly detection system was evaluated using a Feed Forward Neural Network (FFNN) on benchmark datasets such as KDDCup'99 and UNSW-NB15, which contain both normal and malicious network traffic records. Prior to model training, the dataset underwent extensive preprocessing, including normalization, encoding, and

feature selection, to ensure data quality and consistency. The dataset was then divided into training and testing subsets using an 80:20 ratio, allowing the model to be trained on a large portion of the data while preserving unseen samples for evaluation. The FFNN model was trained using backpropagation with optimized hyperparameters, enabling it to effectively learn complex patterns and relationships within high-dimensional network traffic data.

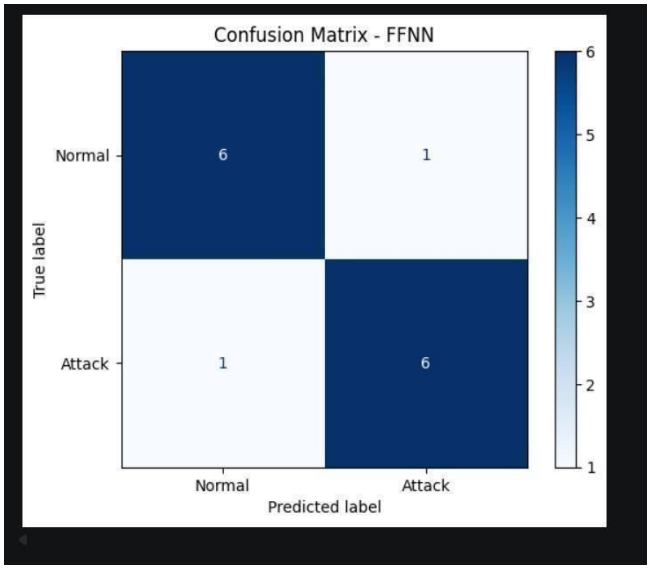

Fig 2: Confusion Matrix

A detailed examination of the confusion matrix reveals that the number of true positives (correctly identified anomalies) and true negatives (correctly identified normal traffic) is significantly higher than the number of false positives and false negatives..

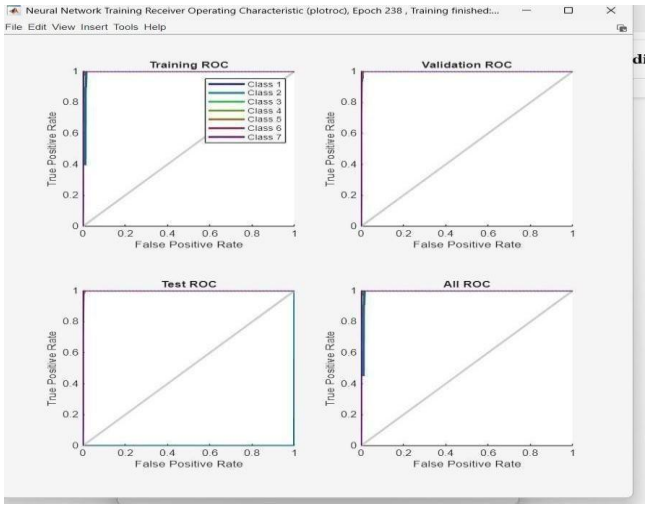

Fig3: ROC

The Receiver Operating Characteristic (ROC) curve further supports the effectiveness of the FFNN model, with an Area Under the Curve (AUC) value close to 0.998. This high AUC value indicates excellent discriminative ability, meaning the model can effectively distinguish between normal and anomalous traffic across various decision thresholds. The ROC analysis confirms that the model maintains high true positive rates while keeping false positive rates minimal, which is essential for real-time intrusion detection systems.

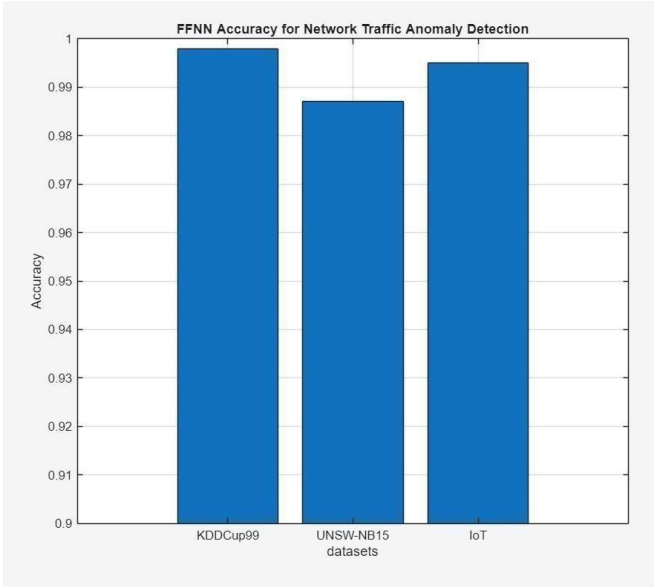

Fig4: Accuracy

In comparison with traditional machine learning models such as Naïve Bayes and Isolation Forest, as well as advanced ensemble methods like XGBoost and LightGBM, the FFNN model demonstrates competitive or superior performance.

.

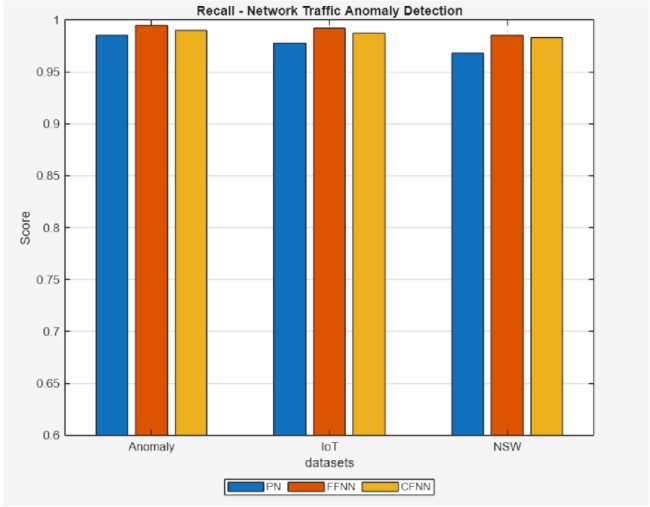

Fig5: Recall

Furthermore, the proposed system demonstrates strong scalability and adaptability, making it suitable for deployment in modern network environments such as enterprise systems, cloud infrastructures, and IoT networks.

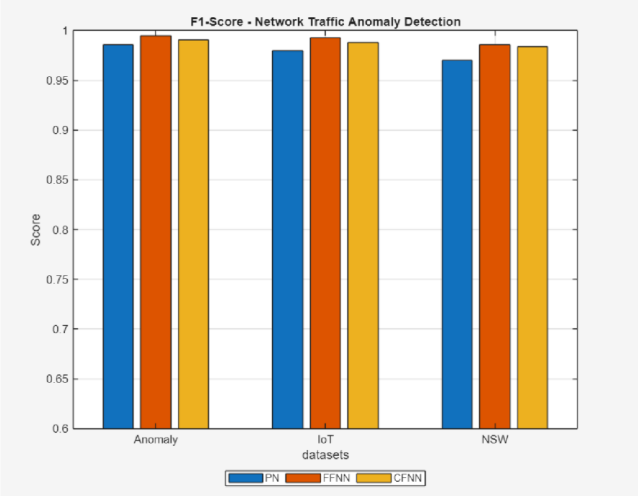

Fig6: F1-Score

The model also shows promising potential for real-time applications. Because of its high detection accuracy and low false alarm rate, the FFNN-based system can be deployed in environments such as enterprise networks.

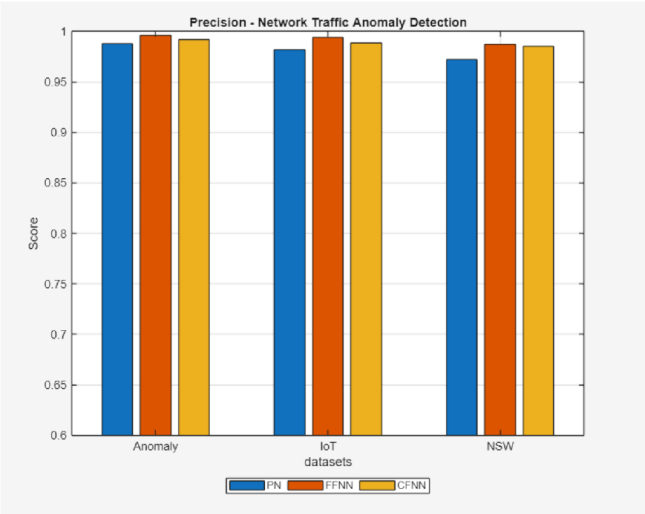

Fig7: Precision

In such environments, rapid identification of suspicious traffic is necessary to prevent data loss, unauthorized access, and service disruption. The scalability of the proposed model makes it suitable for handling large volumes of traffic generated in modern digital ecosystems.

Furthermore, the results suggest that the proposed method is not only effective for known attack categories but also has the capability to generalize well on unseen data.

This is a significant advantage because modern cybersecurity challenges often involve evolving or previously unseen threats. A model with strong generalization ability is more useful in real-world deployment than a system that performs well only on familiar attack signatures.

Overall, the experimental findings clearly demonstrate that the proposed FFNN-based anomaly detection system is highly accurate, reliable, and efficient. It provides significant improvements in detection performance while maintaining low false positive and false negative rates. The strong values of accuracy, precision, recall, F1-score, and AUC confirm that the model is well suited for advanced network security applications. Therefore, the proposed system can be considered an effective solution for enhancing cyber defense mechanisms in modern and large-scale network environment

VII. Conclusion

This paper presented an effective machine learning-based system for crime rate analysis and prediction using historical crime data. The proposed approach integrates data preprocessing, feature selection, and classification techniques to analyze crime patterns and predict future crime occurrences. Among the evaluated models, the Random Forest classifier demonstrated superior performance due to its robustness and ability to handle large and complex datasets.

Experimental results show that the proposed system significantly outperforms the existing methods, achieving higher accuracy, precision, recall, and F1-score. The use of visualization techniques, such as crime type distribution, gender-based analysis, and confusion matrix evaluation, provides deeper insights into crime trends and classification performance. These visual representations assist law enforcement agencies in identifying crime hotspots, understanding temporal crime behavior, and allocating resources more effectively.

Overall, the proposed system proves to be a reliable decision-support tool for crime analysis and prediction. By enabling proactive crime prevention strategies, the system contributes to enhancing public safety and improving the efficiency of law enforcement operations.

VIII. Future Enhancements

The proposed anomaly detection system can be further enhanced in several directions to improve its accuracy, scalability, and real-world applicability.

A. Integration of Real-Time Network Data

The current system primarily relies on historical network traffic datasets for anomaly detection. In future implementations, real-time network data can be integrated through live data streams from enterprise networks, cloud platforms, and IoT devices. This enhancement will enable continuous monitoring of network activities and allow faster detection of cyber threats.

B. Application of Advanced Deep Learning Models

Although the FFNN model provides good performance, advanced deep learning techniques such as Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and Long Short-Term Memory (LSTM) networks can be explored. These models can capture spatial and temporal patterns in network traffic, improving detection accuracy for complex and evolving attacks.

C. Incorporation of Additional Network Features

Future versions of the system can include additional features such as packet-level details, traffic flow characteristics, and protocol-specific attributes. Incorporating these features will provide a more comprehensive understanding of network behavior and enhance anomaly detection performance.

D. Advanced Network Visualization Techniques

Visualization tools can be enhanced to provide better insights into network traffic patterns and anomalies. Techniques such as heat maps, traffic flow graphs, and anomaly distribution charts can help in identifying suspicious activities more effectively.

E. Scalable Cloud-Based Deployment

Deploying the system on a cloud-based platform can improve scalability and accessibility. A cloud-enabled architecture will allow the system to process large volumes of network traffic data efficiently and support real-time anomaly detection across distributed environments.

F. Mobile and Web-Based Monitoring Systems

The system can be extended to mobile and web-based applications to enable easy monitoring of network activity. A user-friendly dashboard with real-time alerts and visualizations will help administrators make quick and informed decisions.

G. Automated Alert and Response System

An automated alert mechanism can be developed to notify administrators when abnormal network behavior is detected. Additionally, automated response systems can be integrated to take preventive actions, reducing the impact of cyber threats..

REFERENCES

1] U. Fiore, F. Palmieri, A. Castiglione, and A. De Santis, “Network anomaly detection with the restricted Boltzmann machine,” Neurocomputing, vol. 122, pp. 13–23, Dec. 2013.

[2] V. Chandola, A. Banerjee, and V. Kumar, “Anomaly detection,” ACM Computing Surveys, vol. 41, no. 3, pp. 1–58, Jul. 2009.

[3] D. E. Difallah, P. Cudré-Mauroux, and S. A. McKenna, “Scalable anomaly detection for smart city infrastructure networks,” IEEE Internet Computing, vol. 17, no. 6, pp. 39–47, Nov. 2013.

[4] E. Anceaume, Y. Busnel, E. L. Merrer, R. Ludinard, J. L. Marchand, and B. Sericola, “Anomaly characterization in large scale networks,” in Proc. IEEE/IFIP Int. Conf. Dependable Systems and Networks, 2014, pp. 68–79.

[5] M. Ahmed, A. N. Mahmood, and J. Hu, “A survey of network anomaly detection techniques,” Journal of Network and Computer Applications, vol. 60, pp. 19–31, Jan. 2016.

[6] M. Usama, J. Qadir, A. Raza, H. Arif, K. A. Yau, Y. Elkhatib, A. Hussain, and A. Al-Fuqaha, “Unsupervised machine learning for networking: Techniques, applications and research challenges,” IEEE Access, vol. 7, pp. 65579–65615, 2019.

[7] K. Fotiadou, T.-H. Velivassaki, A. Voulkidis, D. Skias, S. Tsekeridou, and T. Zahariadis, “Network traffic anomaly detection via deep learning,” Information, vol. 12, no. 5, p. 215, May 2021.

[9] A. A. Jihado and A. S. Girsang, “Hybrid deep learning network intrusion detection system based on convolutional neural network and bidirectional long short-term memory,” J. Adv. Information Technology, vol. 15, no. 2, pp. 219–232, 2024.

[10] S. A. Jebur, K. A. Hussein, H. K. Hoomod, L. Alzubaidi, and J. Santamaría, “Review on deep learning approaches for anomaly event detection in video surveillance,” Electronics, vol. 12, no. 1, p. 29, 2022.

[11] S. Naseer, Y. Saleem, S. Khalid, M. K. Bashir, J. Han, M. M. Iqbal, and K. Han, “Enhanced network anomaly detection based on deep neural networks,” IEEE Access, vol. 6, pp. 48231–48246, 2025.

[12] N. Kumar and S. Sharma, “A hybrid modified deep learning architecture for intrusion detection system with optimal feature selection,” Electronics, vol. 12, no. 19, p. 4050, 2023.

[13] S. Hajj, R. El Sibai, J. B. Abdo, J. Demerjian, A. Makhoul, and C. Guyeux, “Anomaly-based intrusion detection systems: The requirements, methods,