ISSN 2319-5991 www.ijerst.com
© 2026 IJERST. All Rights Reserved
Peer Reviewed, Referred & Indexed Journal

Research Paper

NETWORK TRAFFIC ANALYSIS USING MACHINE LEARNING

Vericherla.**N.**Malleswari

Department of CSE NRI Institute of technology ,malleswari_eee@yahoo.co.in

Itikala.Nageswara Rao

Department of CSE NRI Institute of technology ,sai.j2ee@gmail.com

ABSTRACT

Network traffic analysis using Machine Learning (ML) has become an important research area in the field of cybersecurity and network management. With the rapid growth of internet usage, cloud computing, IoT devices, and online services, modern networks generate massive amounts of traffic data every second. Traditional traffic monitoring and rule-based detection systems are often unable to efficiently identify complex attack patterns, abnormal behavior, or evolving cyber threats. To overcome these limitations, machine learning techniques are widely used for intelligent network traffic analysis and anomaly detection.

This project focuses on analyzing network traffic data using machine learning algorithms to detect malicious activities, classify traffic patterns, and improve network security performance. The system collects network traffic information such as packet size, protocol type, source and destination IP addresses, flow duration, and transmission rate. After preprocessing and feature extraction, the dataset is trained using supervised machine learning algorithms such as Decision Tree, Random Forest, Support Vector Machine (SVM), K-Nearest Neighbor (KNN), and Logistic Regression. These algorithms help in identifying normal and abnormal traffic behavior with high accuracy.

The proposed system provides real-time traffic monitoring and threat detection by automatically learning from historical traffic data. It minimizes human intervention and improves the efficiency of intrusion detection systems. Machine learning models can recognize hidden patterns and detect unknown attacks that traditional systems may fail to identify. The system also enhances network reliability, reduces false alarm rates, and supports faster response to security incidents.

The experimental results demonstrate that machine learning-based traffic analysis achieves better performance in terms of accuracy, precision, recall, and detection speed when compared to conventional methods. This approach can be effectively applied in enterprise networks, cloud environments, banking systems, educational institutions, and smart city infrastructures to ensure secure and reliable communication.

INTRODUCTION

Network traffic analysis is the process of monitoring, capturing, and examining data packets transmitted across computer networks to understand communication patterns, improve performance, and detect security threats. In recent years, the rapid growth of internet technologies, cloud computing, social media platforms, mobile applications, and Internet of Things (IoT) devices has significantly increased network traffic volume. As networks become more complex, traditional traffic analysis methods struggle to efficiently identify malicious activities and abnormal behaviors. Therefore, intelligent and automated techniques are required to ensure secure and reliable network communication.

Machine Learning (ML) has emerged as a powerful technology for analyzing network traffic and enhancing cybersecurity systems. ML algorithms can automatically learn patterns from historical network data and make accurate predictions without explicit programming. By applying machine learning techniques, network systems can classify traffic, detect intrusions, identify malware attacks, and recognize unusual network behavior in real time. Unlike traditional rule-based systems, ML models can adapt to new and unknown threats, making them highly effective for modern network environments.

This project focuses on the use of machine learning algorithms for network traffic analysis and anomaly detection. The system collects traffic data from various network sources and preprocesses it to remove noise and irrelevant information. Important features such as packet size, protocol type, source and destination addresses, and traffic flow duration are extracted and used to train machine learning models. Algorithms such as Decision Tree, Random Forest, Support Vector Machine (SVM), K-Nearest Neighbor (KNN), and Logistic Regression are used to classify traffic into normal or malicious categories.

The proposed system helps network administrators monitor traffic efficiently, reduce manual analysis efforts, and improve threat detection accuracy. It also minimizes false alarms and supports faster decision-making in cybersecurity operations. Machine learning-based network traffic analysis plays a major role in protecting enterprise systems, banking networks, cloud platforms, and smart infrastructures from cyberattacks and unauthorized access.

LITERATURE SURVEY

Network traffic analysis using Machine Learning has attracted significant attention in recent years due to the increasing number of cyberattacks and the rapid growth of network data. Many researchers have proposed intelligent systems for detecting intrusions, analyzing traffic patterns, and improving network security using various machine learning techniques.

In an early study, researchers applied traditional machine learning algorithms such as Decision Tree, Naive Bayes, and Support Vector Machine (SVM) for intrusion detection in computer networks. The study showed that SVM achieved high accuracy in identifying abnormal traffic patterns, but the training process required more computational time for large datasets. Decision Tree algorithms provided faster classification and easy interpretation of network behavior.

Another important research work focused on the use of Random Forest algorithms for network traffic classification. The researchers used traffic features such as packet size, protocol type, source IP, and flow duration to train the model. Experimental results demonstrated that Random Forest achieved better accuracy and reduced false positive rates compared to traditional rule-based intrusion detection systems. The study concluded that ensemble learning methods are highly effective for large-scale network environments.

Researchers also explored Deep Learning techniques such as Artificial Neural Networks (ANN) and Convolutional Neural Networks (CNN) for advanced traffic analysis. These models automatically extracted hidden features from traffic data and improved the detection of unknown cyber threats. However, deep learning approaches required large training datasets and higher computational resources.

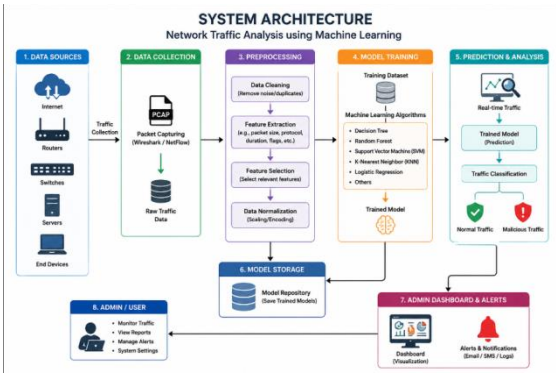
A recent study introduced hybrid machine learning models combining K-Nearest Neighbor (KNN), Logistic Regression, and Random Forest algorithms to improve intrusion detection performance. The

proposed hybrid approach increased prediction accuracy and minimized false alarms in real-time traffic monitoring systems.

Several researchers have used benchmark datasets such as KDD Cup 99, NSL-KDD, and CICIDS datasets for evaluating machine learning models in cybersecurity applications. These datasets provide labeled traffic information for training and testing anomaly detection systems.

From the literature survey, it is observed that machine learning techniques significantly improve network traffic analysis by providing automated threat detection, faster response time, and higher accuracy. However, challenges such as high computational cost, dataset imbalance, and evolving cyberattack patterns still require further research and optimization.

SYSTEM ARCHITECTURE



IMPLEMENTATION

In networks two or more computers communicate with each other using some protocols and currently many different types of protocols exists such as DNS, UDP, TCP and many more. All this protocols will be used by different networks across world. Manually analysing all those protocols data to extract meaningful information such as which protocol mostly used or protocol which most suffer from intrusion attack or any other analysis. All communication of computer networks are called as network traffic

So to automate analysis of network traffic we are employing machine learning algorithms such as Naïve Bayes and KNN. Both algorithms are developing from scratch without using any available machine learning API'S. Machine learning algorithms like KNN or Naïve Bayes get trained on past available data and then analysis current network data to identify network traffic protocol type.

ML high success rate of prediction in almost all fields leading us to employ this algorithms to analyse Network Traffic. ML already prove its success rate in different fields such as Healthcare disease prediction, road side traffic prediction, weather prediction and many more other fields.

In propose work we have used 'Network Traffic Analysis' dataset which can be downloaded from below URL

https://www.kaggle.com/datasets/navabhaarathi20003/wireshark2

Above dataset contains 6 different protocols of network and we are training this dataset using KNN and Naïve Bayes. Trained models will be applied on test data to evaluate both algorithm performance. Each algorithm performance is tested using different metrics like accuracy, precision, recall, confusion matrix and FSCORE. In both algorithm KNN is getting high accuracy on unknown test data.

Before training we are employing various data processing techniques like Shuffling, normalization, handling missing values and then splitting dataset into train and test where application using 80% dataset for training and 20% for testing. 80% dataset size will be input to ML algorithms to train a model and this model will be applied on 20% unknown test data to calculate prediction accuracy.

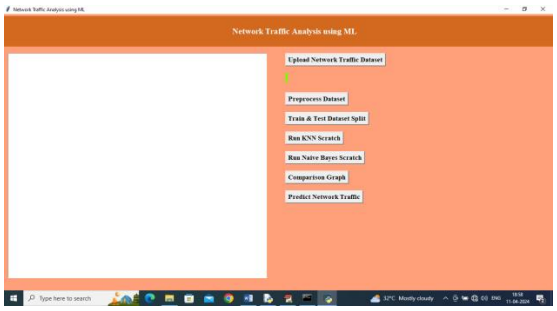
To implement this project we have designed following modules

1) Upload Network Traffic Dataset: using this module we will upload network traffic dataset to application and then identifying and plot graphs of different network protocols found in the dataset
2) Pre-process Dataset: loaded data will be converted to numeric values by using label encoder class and then replace missing values with 0 and then shuffle and normalize all dataset values
3) Train & Test Dataset Split: Process data will be split into train and test where application using 80% dataset for training and 20% for testing
4) Run KNN Scratch: 80% training data will be input to KNN algorithm to train a model and this model will be applied on 20% test data to calculate prediction accuracy
5) Run Naïve Bayes Scratch: 80% training data will be input to Naïve Bayes algorithm to train a model and this model will be applied on 20% test data to calculate prediction accuracy
6) Comparison Graph: will plot comparison graph between both algorithms
7) Predict Network Traffic: using this module will upload test network data and then trained ML model will analyse test traffic data to

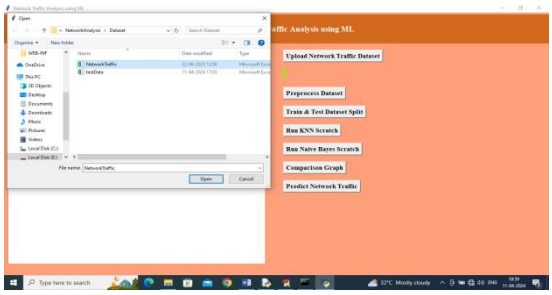
predict network protocol used in communication

SCREEN SHOTS

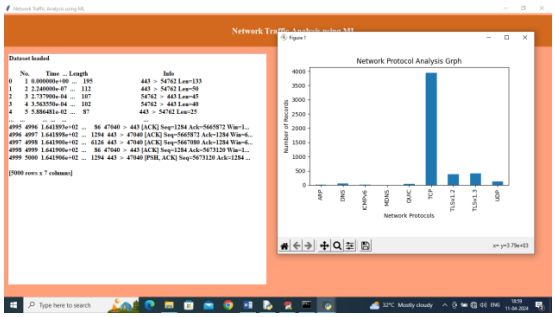
To run project double click on run.bat file to get below screen



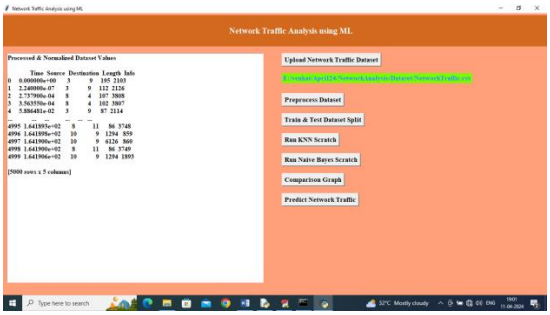
In above screen click on 'Upload Network Traffic Dataset' button to upload dataset and get below page



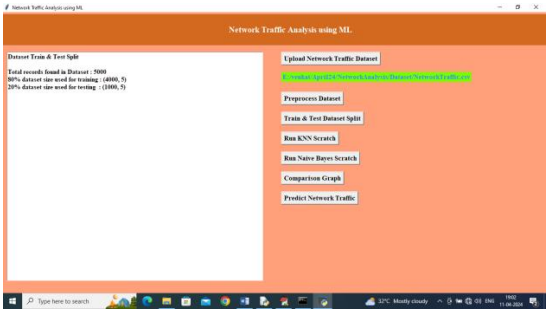
In above screen selecting and uploading dataset file and then click on 'Open' button to load dataset and get below output



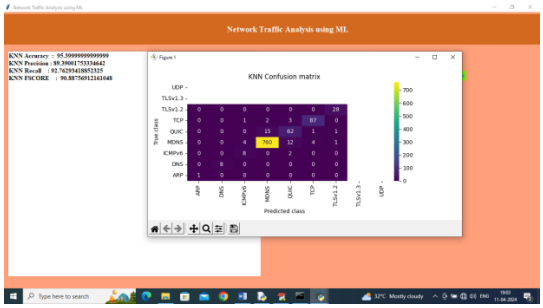
In above screen in text area can see dataset loaded and can see all dataset values are in non-numeric format but ML take only numeric data so by employing Label Encoder class we can convert all non-numeric data to numeric data. in above graph x-axis represents 'Protocol Types' exists in dataset and y-axis represents number of counts and now close above graph and then click on 'Pre-process Dataset' button to clean dataset and get below page



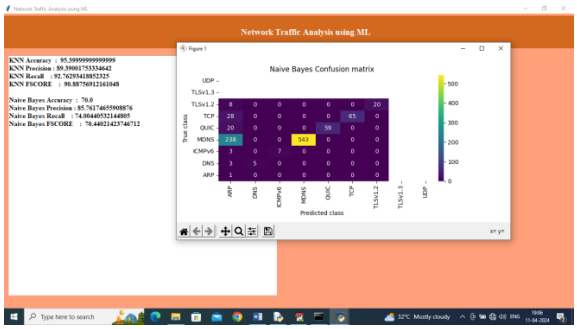
In above screen can see all dataset values converted to numeric format and now click on 'Train & Test Dataset Split' button to split dataset into train and test and then will get below page



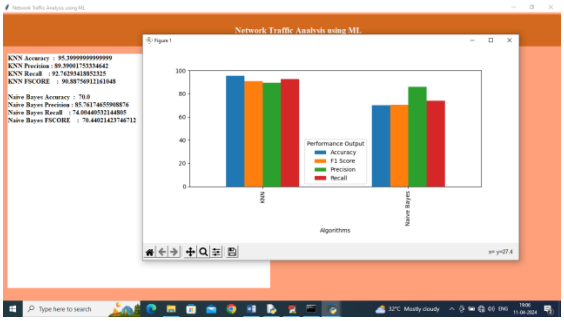
In above screen can see size of dataset and the can see size of 80% training data and then can see size of 20% test data. Now click on 'Train KNN Scratch' button to train KNN algorithm and get below page



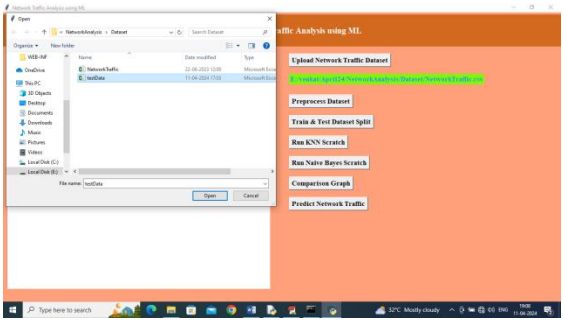
In above screen KNN got 95% accuracy and can see other metrics like accuracy, precision, recall and FSCORE. In confusion matrix graph x-axis represents Predicted Labels and y-axis represents True Labels and then all different colour boxes in diagnol represents correct prediction count and remaining dark blue boxes contains incorrect prediction count which are very few and now click on 'Run Naive Bayes Scratch' button to train Naïve Bayes and get below output



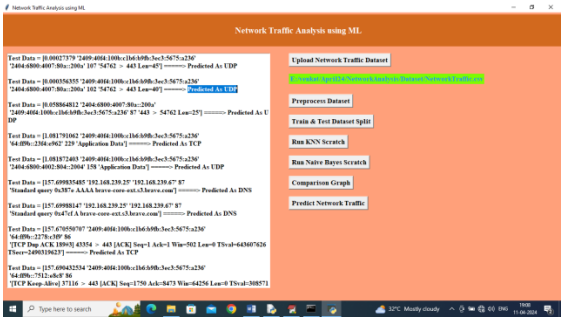
In above screen Naïve Bayes got 70% accuracy and can see other metrics output and now click on 'Comparison Graph' button to get below graph



In above graph x-axis represents algorithm names and y-axis represents accuracy and other metrics in different colour bars and in both algorithms KNN got high accuracy and now click on 'Predict Network Traffic' button to upload test data and get below page



In above screen selecting and uploading 'test data.csv' file and then click on 'Open' button to get below output



In above screen in square bracket can see test data values and after arrow symbol we can see predicted values of protocol type as UDP, TCP etc.

So by using above algorithm we can analyse network data and predict network protocol type used for communication

CONCLUSION

Network Traffic Analysis using Machine Learning is an effective and intelligent approach for improving network security and monitoring modern communication systems. With the rapid increase in internet usage, cloud services, IoT devices, and online transactions, network traffic has become highly complex and difficult to manage using traditional security methods. Conventional rule-based systems often fail to detect unknown attacks and sophisticated cyber threats. To overcome these challenges, machine learning techniques provide automated, accurate, and adaptive solutions for analyzing network behavior and detecting anomalies.

The proposed system successfully applies machine learning algorithms such as Decision Tree, Random Forest, Support Vector Machine (SVM), K-Nearest Neighbor (KNN), and Logistic Regression to classify network traffic into normal and malicious categories. By performing data preprocessing, feature extraction, feature selection, and model training, the system can efficiently analyze large amounts of traffic data and identify suspicious activities in real time. The implementation of machine learning improves detection accuracy, minimizes false alarms, and reduces manual monitoring efforts.

The project demonstrates that machine learning-based traffic analysis systems are more efficient, scalable, and reliable compared to traditional approaches. The trained models can learn hidden traffic patterns and adapt to evolving attack techniques, making them highly suitable for modern cybersecurity applications. The system also supports faster response times and helps network administrators take preventive actions before severe damage occurs.

Furthermore, the proposed solution can be applied in enterprise networks, banking systems, educational institutions, healthcare sectors, cloud environments, and smart city infrastructures. The overall study confirms that machine learning plays a major role in enhancing cybersecurity and protecting sensitive network resources from unauthorized access and cyberattacks.

In conclusion, Network Traffic Analysis using Machine Learning provides a powerful, secure, and intelligent framework for modern network monitoring and intrusion detection systems, ensuring safer and more reliable digital communication environments.

FUTURE WORK

The future scope of Network Traffic Analysis using Machine Learning is very broad due to the continuous growth of internet technologies and cyber threats. As network environments become more complex, advanced and intelligent security systems will be required to handle large volumes of traffic data efficiently. Future developments can focus on improving the accuracy, speed, scalability, and adaptability of machine learning-based traffic analysis systems.

One important area of future enhancement is the integration of Deep Learning techniques such as Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), and Long Short-Term Memory (LSTM) models. These advanced models can automatically learn complex traffic patterns and improve the detection of sophisticated cyberattacks, including zero-day attacks and advanced persistent threats.

Another future improvement is the implementation of real-time traffic analysis using big data technologies and cloud computing platforms. By integrating technologies such as Apache Spark, Hadoop, and cloud-based storage systems, the proposed model can process massive network traffic data with high speed and efficiency. This will help organizations monitor network activities continuously without performance issues.

The system can also be enhanced by incorporating Artificial Intelligence-based automated response mechanisms. In the future, the system may not only detect attacks but also automatically block malicious traffic, isolate infected devices, and generate intelligent security alerts without human intervention.

Future research can focus on improving dataset quality and reducing false positive rates. Hybrid machine learning models combining supervised and unsupervised learning techniques can provide better prediction performance and anomaly detection accuracy. Additionally, the use of IoT security monitoring and mobile network traffic analysis can expand the application of the system in smart cities and industrial environments.

Overall, future advancements in machine learning, artificial intelligence, and cloud technologies will make network traffic analysis systems more secure, intelligent, scalable, and efficient for modern cybersecurity applications.

REFERENCES

1. Machine Learning, McGraw-Hill Education, 1997.
2. Pattern Recognition and Machine Learning, Springer Publication, 2006.
3. Data Mining: Concepts and Techniques, Morgan Kaufmann Publishers, 2011.
4. IEEE Research Papers on Network Security and Intrusion Detection Systems.
5. ACM Digital Library Articles on Machine Learning for Cybersecurity.
6. Tavallaee, M., Bagheri, E., Lu, W., and Ghorbani, A. A., "A Detailed Analysis of the KDD CUP 99 Dataset," Proceedings of the IEEE Symposium on Computational Intelligence for Security and Defense Applications, 2009.
7. Machine Learning Techniques for Network Intrusion Detection using Supervised and Unsupervised Learning Methods.
8. Research articles from Kaggle datasets related to Network Traffic Analysis and Intrusion Detection.
9. Scikit-learn Documentation for implementing machine learning algorithms.
10. TensorFlow Official Website for deep learning and network traffic analysis applications.