IJIRIS: International Journal of Innovative Research in Information Security
Volume 11, Issue 03, May 2025
E-ISSN: 2349-7017
P-ISSN: 2349-7009
https://www.ijiris.com/archives

Cyber based Intrusion Detection Monitoring System Using Machine Learning

Mohammad Qamaruddin, Dodla Vamshi Krishna
Department of Information Technology,
Guru Nanak Institutions Technical Campus,
Hyderabad, Telangana, INDIA
mqamaruddin07@gmail.com, vamshikrishnadodla4@gmail.com
Faris Faiyaz Abdul Gaffar, Mohammed Sajid
Department of InformationTechnology,
Guru Nanak Institutions Technical Campus,
Hyderabad, Telangana, INDIA
farisfaiyaz1@gmail.com , sajidmohammed8897@gmail.com
Dr.M.I.Thariq Hussan
Professor and Head, Department of Information Technology,
Guru Nanak Institutions Technical Campus,
Hyderabad, Telangana, INDIA
hodit.gnitc@gniindia.org

Publication History:
Manuscript Reference No: IJIRIS/RS/Vol.11/Issue03/MYIS10080
Research Article | Open Access | Double-Blind Peer-Reviewed | Article ID: IJIRIS/RS/Vol.11/Issue03/MYIS10080
Received: 20,April 2025 Revised: 29, April 2025 Accepted: 05, May 2025 Published Online: 15, May 2025, Volume 2025 Article ID MYIS10080 https://www.ijiris.com/volumes/Vol11/iss-03/01.MYIS10080.pdf
Article Citation: Mohammad,Dodla,Faris,Sajid,Dr.Thariq(2025). Cyber based Intrusion Detection Monitoring System using Machine Learning: IJIRAE:: International Journal of Innovative Research in Information Security, Volume 11, Issue 02, Pages 157-161 **doi:>** https://doi.org/10.26562/ijiris.2025.v1103.01
BibTex key: Dr.Thariq@2025Cyber

Abstract: The rise of cyber-attacks presents significant challenges in securing digital networks. This paper proposes an Intrusion Detection System (IDS) leveraging Machine Learning (ML) techniques to classify network traffic as normal or malicious. The system utilizes the NSL-KDD dataset, a benchmark for evaluating IDS performance, to train and test various ML models, including K-Nearest Neighbors, Naïve Bayes, Decision Tree, Random Forest, and Logistic Regression. Among these, Random Forest is anticipated to achieve optimal accuracy due to its robust ensemble learning capabilities. A user-friendly web interface enables real-time interaction, allowing users to input network traffic data and obtain security insights. Future enhancements include optimization techniques, cloud-based deployment, and real-time monitoring to transition from prototype to practical application. This IDS aims to strengthen cyber security defenses while minimizing costs associated with cyber intrusions.

Keywords: Intrusion Detection System, Machine Learning, Cyber security, Network Security, NSL-KDD Dataset, Random Forest, Anomaly Detection, Signature-Based Detection, Real-Time Monitoring, Cyber Threats

I. INTRODUCTION

With the rapid growth of internet-connected devices and increasing cyber security threats, the need for efficient and accessible Intrusion Detection Systems (IDS) has become crucial. Traditional IDS solutions often require complex software installations and configurations, limiting their usability for individuals and small organizations. To address this challenge, we propose a novel web-based IDS frame work that enables users to detect network intrusions without the need for specialized software. Our system utilizes the Random Forest algorithm, trained on the well-known KDD dataset, to accurately identify malicious activities. By offering an intuitive web interface for manual data entry and CSV file uploads, combined with a light weight packet capture script, the proposed solution bridges the gap between accessibility and effective cyber security.

II. LITERATURE REVIEW

Intrusion Detection Systems (IDS) have been an active area of research for decades due to the critical need for network security. Traditional IDS implementations have largely relied on signature-based techniques, as seen in Snort and BroIDS, which depend on predefined attack patterns to detect intrusions. While effective for known threats, such systems struggle with detecting novel attacks. In response, machine learning-based IDS frameworks have gained significant popularity. Early studies such as by Lee et al. demonstrated the use of decision trees for anomaly detection in network traffic. The KDD Cup1999 dataset emerged as a standard benchmark for training and evaluating IDS models. Research by Breiman introduced the Random Forest algorithm, a powerful ensemble learning method known for its robustness, scalability, and high accuracy in classification tasks. Recent developments have seen IDS models transitioning towards accessibility and user-friendliness. Web-based IDS solutions have been proposed, but most require software installations or server-side configurations, making them less approachable for individuals and small organizations.

Furthermore, concerns over data privacy, compliance with standards like GDPR and HIPAA, and real-time detection have been central in recent IDS research discussions. Our work builds upon these foundations by designing a light weight, web-accessible IDS using Random Forest trained on the KDD dataset. It eliminates installation barriers and offers an intuitive interface while maintaining strong detection capabilities and addressing privacy considerations.

III. METHODOLOGY

A. Dataset Selection:

The KDD Cup 1999 dataset was selected for training and evaluating the machine learning model. This data set contains a wide range of intrusion patterns, including DoS, U2R, R2L, and probing attacks, providing a comprehensive base for building a robust IDS.

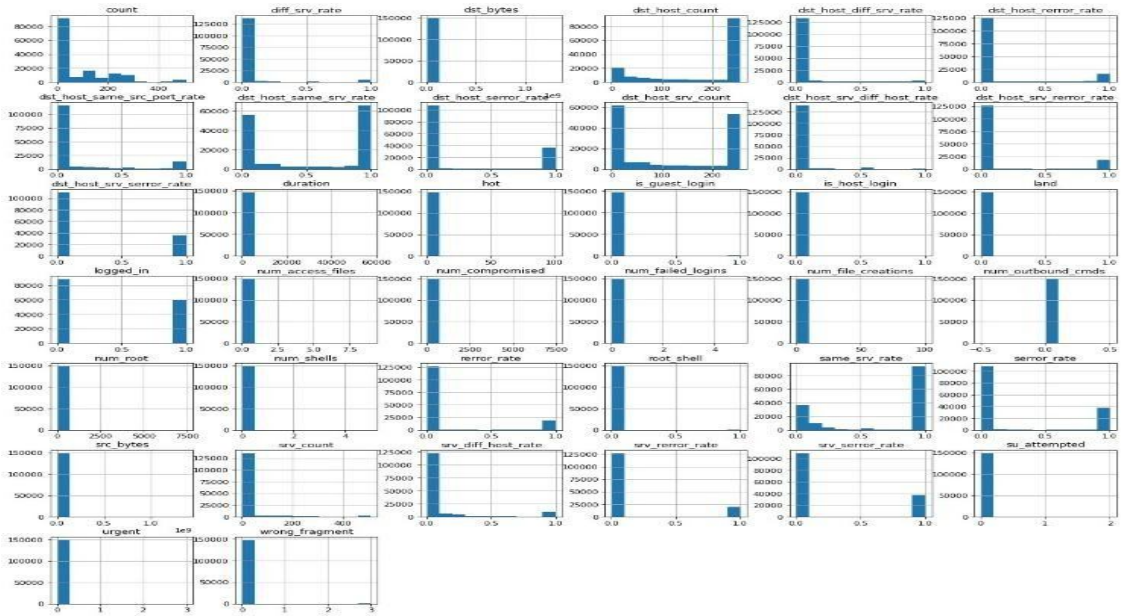

Fig.1-Feature Distribution

Title: Distribution of Key Features in the Dataset

Caption: This figure displays histograms of selected continuous features such as duration, src_bytes, dst_bytes, and count. These distributions reveal skewness, presence of outliers, and scale differences, indicating the need for normalization and log transformations to enhance model training performance.

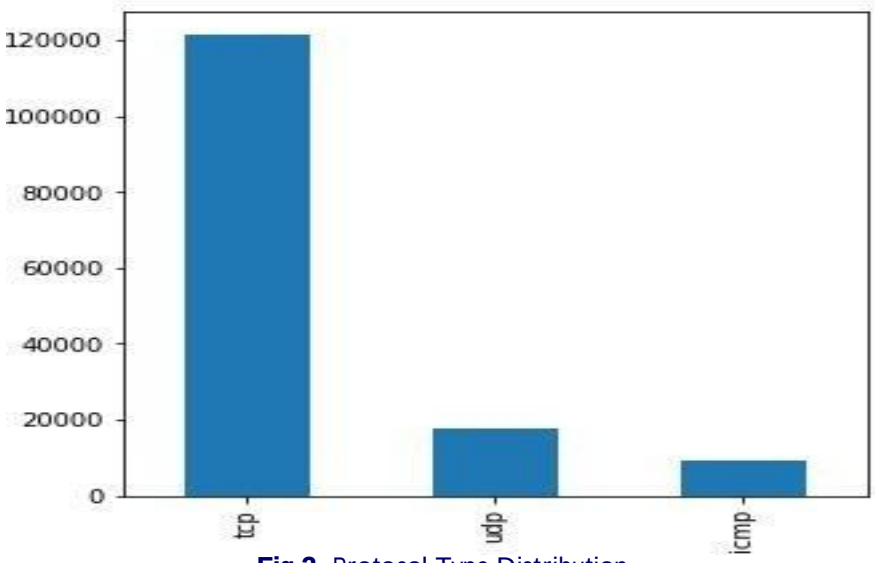

Fig.2- Protocol Type Distribution

Title: Frequency Distribution of Protocol Types

Caption: This bar chart shows the frequency of different network protocol types in the dataset, highlighting that TCP dominates, followed by UDP and ICMP. This helps understand protocol-level traffic behavior and its potential relation to intrusion patterns.

Title: Most Frequent Network Services

Caption: The figure visualizes the top 30 services accessed in the network traffic data, including HTTP, DNS, FTP, and SSH. This insight helps understand which services are more prone to attacks or serve as typical channels for malicious payloads.

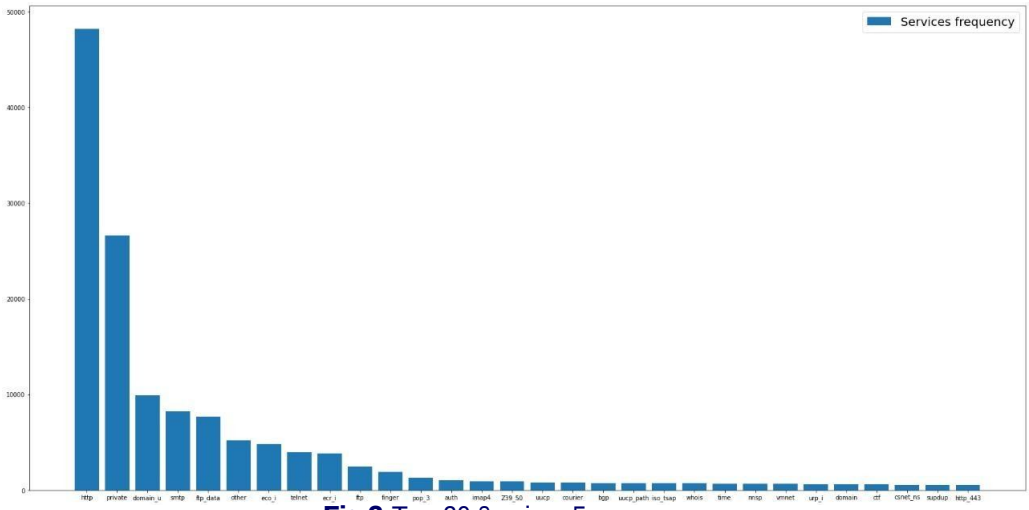

Fig.3-Top 30 Services Frequency

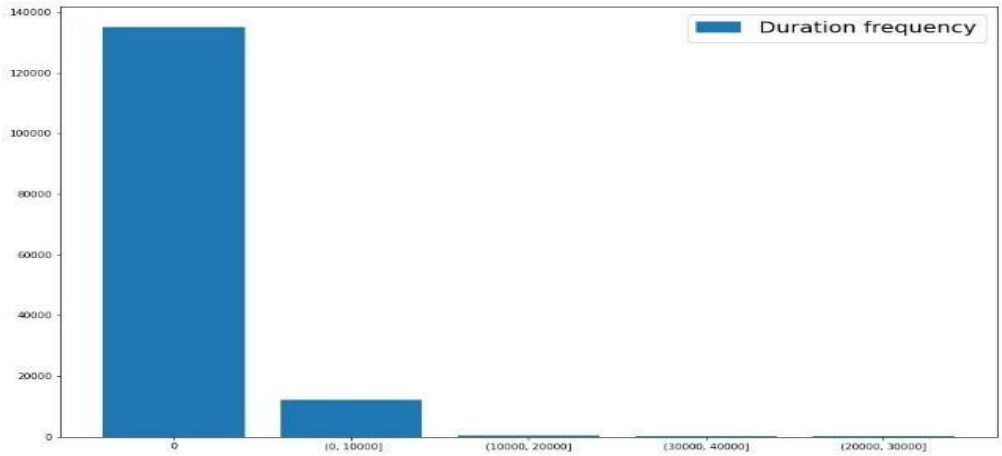

Fig.4- Connection Duration Distribution

Title: Distribution of Connection Duration

Caption: This histogram illustrates the wide range of connection durations. Most connections are short-lived, while a few outliers represent prolonged sessions, which could indicate data exhilaration or scanning activity. Log-scaling might be applied to better visualize long-tail distributions.

B. Machine Learning Model:

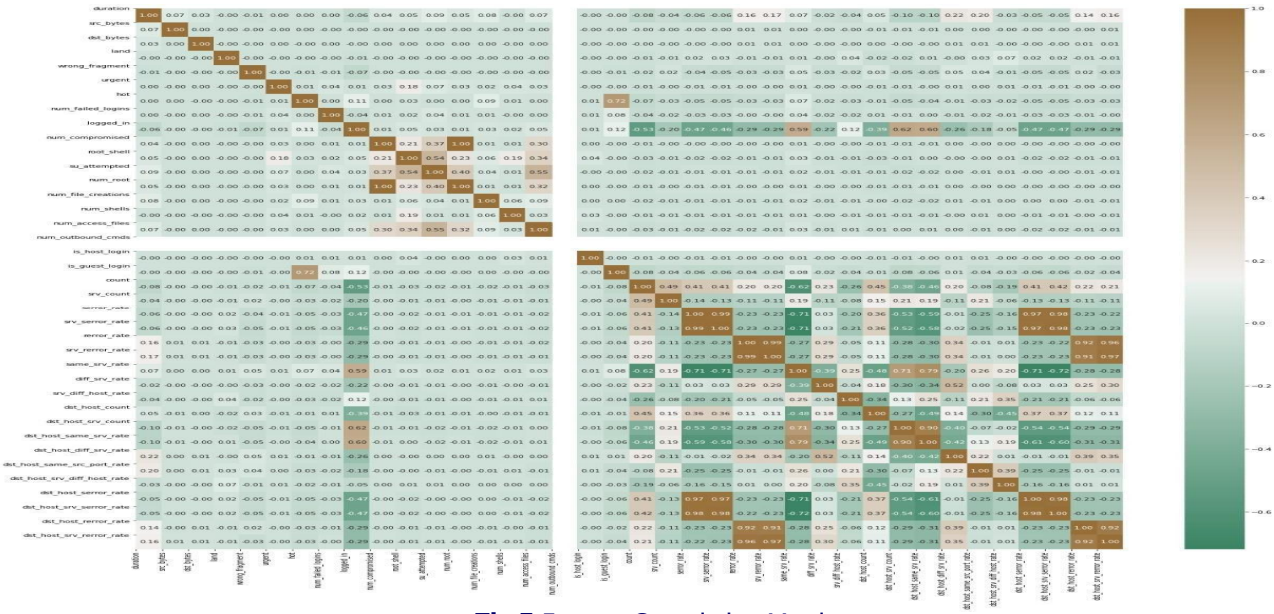

Fig.5-Feature Correlation Matrix

The Random Forest algorithm was chosen due to its high accuracy, ability to handle large datasets, and robustness against over fitting. The model was trained to classify network connections as either normal or attack based on multiple features extracted from the dataset.

Title: Correlation Matrix of Dataset Features

Caption: This heat map visualizes Pearson correlation coefficients between all pairs of numeric features. Strong correlations (positive or negative) are observed among some variables, guiding the feature selection process to reduce redundancy and multi collinearity.

C. System Architecture:

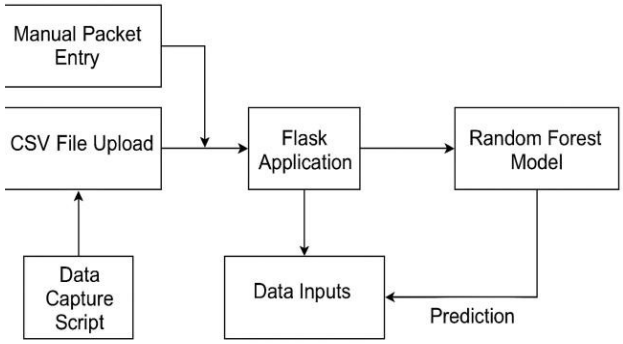

FIG. 1 – System Architecture

The system consists of a web-based front-end developed using Flask, a machine learning back-end for prediction, and a light weight Python-based packet capture script. The architecture enables users to interact with the IDS through a browser without needing to install any software.

IV. IMPLEMENTATION

A. Web Application:

The web application provides two main functionalities: manual packet entry and CSV file upload. Users can either enter individual packet details into a form or upload a file containing multiple packets to detect intrusions.

B. Data Capture Script:

A Python script using libraries such as Scapy captures live network packets and saves them into a CSV format. This script enables users to collect real-world data for testing without complex setup procedures.

C. Model Integration:

The trained Random Forest model is integrated into the Flask application. When data is submitted via the form or file upload, the model processes the input and predicts whether the traffic is normal or indicative of an intrusion.

V. RESULTS

Experimental evaluations demonstrate that the proposed IDS accurately identify intrusions with an approximate accuracy of 96% on the KDD dataset. The web-based system provides seamless user interaction and reliable intrusion detection without compromising user privacy.

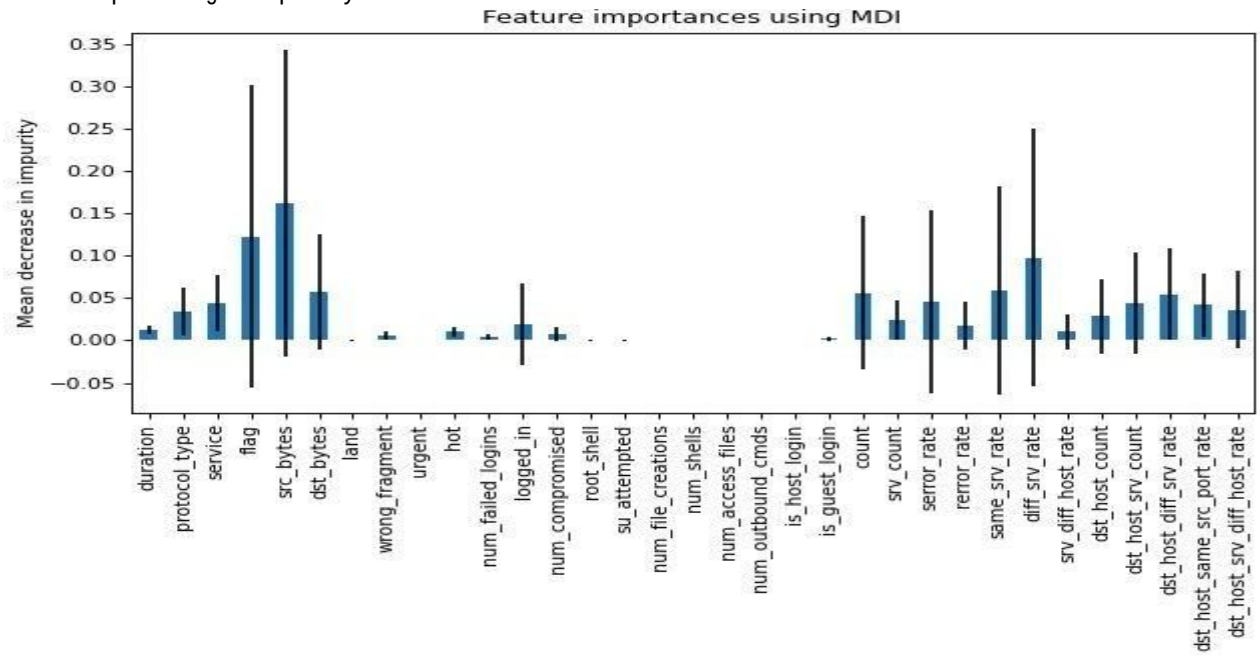

Fig.6- Feature Importance Ranking by Random Forest

Title: Feature Importance Ranking by Random Forest Model

Caption: The bar graph displays the top features influencing model predictions, as determined by the Random Forest classifier. Features like service, flag,src_bytes, and dst_bytes emerged as highly significant in detecting intrusions, guiding feature optimization for model improvement.

VI.CONCLUSION

In conclusion, the field of Managing Data and Quality Evaluation for Machine Learning-Based Cyber Intrusion Detection plays a vital role in enhancing the efficacy and reliability of cyber security measures. Cyber threats continue to evolve and pose a significant risk to organizations, making the development of accurate and efficient intrusion detection system scrucial. In a world where cyber threats are ever-present, the work on data curation and quality evaluation for machine learning- based cyber intrusion detection is a proactive and necessary step to protect sensitive data, critical infrastructure, and digital assets. Item powers organizations to build robust and adaptive intrusion detection systems that can effectively safeguard against a wide range of cyber threats and vulnerabilities. As the field of cyber security continues to evolve, ongoing research and development in data curation and quality evaluation will remain essential for staying one stepahead of cyber adversaries.

REFERENCES

1. Amini,M.,Ghaffari,A.,&Akbari,M.(2002).A review of intrusion detection systems and the KDD dataset. Journal of Computer Science and Technology, 17(3), 231-240.
2. Bagade,A. S.,Patil,R.D.,&Kulkarni,P.(2015). Intrusion detection using Random Forest and KDD dataset. International Journal of Computer Applications, 120(18), 42-47.
3. Chavan,S.,&Raut,D.(2019).A comparative study of Decision Trees and Machine Learning algorithms for intrusion detection. International Journal of Computer Science and Information Technologies, 10(2), 89-94.
4. Pham,V.H.,&Bui,X.(2018). Deep learning-based intrusion detection systems using KDD-99 dataset. Proceedings of the International Conference on Machine Learning and Data Engineering, 48-53.
5. Zhang,Z.,&Li,H.(2017). Feature selection for support vector machine in intrusion detection using KDD dataset. Neural Computing and Applications,28(4),745-752.
6. Xu,Z.,&Li,H.(2014). A survey of machine learning-based intrusion detection techniques. International Journal of Computer Science & Network Security, 14(8), 76-82.
7. Vapnik,V.(1998).Statistical learning theory. Wiley-Inter science.
8. Cortes,C.,&Vapnik,V.(1995).Support vector networks. Machine Learning,20(3),273-297.
9. Bishop,C.M.(2006).Pattern recognition and machine learning. Springer.
10. He,H.,&Wu,D.(2019).Real-time a nomaly detection using deep learning. IEEE Transactions on Neural Networks and Learning Systems,30(4),1-11