

Enhancing Intrusion Detection with Dimensionality Reduction Methods Using Machine Learning

M.Rathika¹, P.Sivakumar², V.Bhuvaneshwari³, S. Sandosh⁴

¹Department of Electronics and Communication Engineering, Kingston Engineering College, Vellore, India, rathikame@gmail.com

²Department of Electronics and Communication Engineering, Dr.NGP Institute of Technology, Coimbatore, India, sivakumar.poruran@gmail.com

³Department of Electronics and Communication Engineering, Dr.NGP Institute of Technology, Coimbatore, India, bhuvi1402@gmail.com

⁴School of Computer Science and Engineering, VIT University, Chennai Campus, India, sandosh.s@vit.ac.in

This project investigates the efficacy of machine learning algorithms, including Naive Bayes (NB), NLP, and K-Nearest Neighbors (KNN), in the context of Intrusion Detection Systems (IDS), with a focus on dimensionality reduction techniques. By leveraging datasets such as USNW NB 15, NSL-KDD, and UNSWNB15, the research aims to assess the accuracy and efficiency of these algorithms in identifying malicious network activities while integrating Dimensionality reduction methods. Through systematic experimentation and evaluation, the project seeks to determine the most effective algorithm for intrusion detection, considering both traditional metrics and the impact of Dimensionality reduction. The findings contribute to enhancing internet security measures by identifying robust solutions for detecting and preventing intrusions amidst complex network traffic. This research underscores the importance of incorporating dimensionality reduction techniques to improve the performance of IDS and mitigate potential threats effectively. The project provides insights for practitioners and researchers in selecting appropriate algorithms and integrating dimensionality reduction strategies for more efficient intrusion detection systems.

Keywords: NB (Naive Bayes), NLP (Natural Language Processing), KNN (K-Nearest Neighbours), Intrusion Detection.

1. Introduction

The proliferation of the internet over the past few decades has heralded a new era of communication, commerce, and connectivity. However, with this unprecedented level of interconnectedness comes an inherent vulnerability to various security threats. Malicious

actors constantly exploit weaknesses in network infrastructure, seeking to gain unauthorized access, steal sensitive data, or disrupt services. In response to these evolving threats, Intrusion Detection Systems (IDS) have emerged as essential components of cybersecurity strategies, aiming to safeguard networks from unauthorized access and malicious activities.

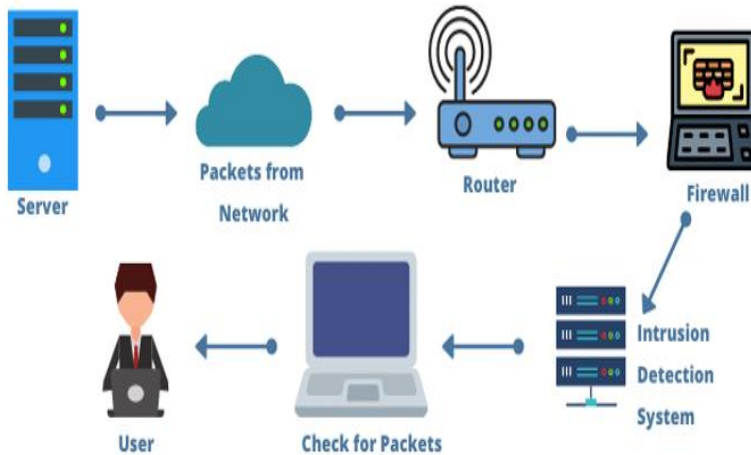


Fig. 1.1. Intrusion Detection System

IDS functions as vigilant sentinels, continuously monitoring network traffic for signs of suspicious or anomalous behaviour. Traditional IDS rely on predefined rules or signatures to identify known threats, but these methods struggle to keep pace with the rapidly evolving tactics of cyber attackers. As a result, there has been a growing interest in leveraging machine learning algorithms to enhance the effectiveness of IDS. By analysing vast amounts of network data and identifying patterns indicative of malicious behaviour, machinelearning-based IDS offer the potential for more adaptive and proactive threat detection. This evaluation will use benchmark datasets such as USNW NB 15, NSL-KDD, and UNSWNB15.

2. Review of Literature

The significance of robust intrusion detection systems (IDS) in cybersecurity is underscored by several studies focusing on various techniques and methodologies to enhance their effectiveness. Kiran et al. (2023) [1] highlight the increasing reliance on computers and the internet, which necessitates improved information security to prevent intrusions by hackers and unauthorized users. Their research introduces IDS as crucial for identifying and responding to security breaches, explaining its function of detecting deviations from established behavior patterns and generating management reports. They emphasize the potential of machine learning to enhance IDS capabilities, noting positive experimental results, though the study is limited in depth and lacks counterarguments.

Agora Moorthy et al. (2023) [2] delve into the performance and limitations of signature-based components in hybrid IDS, combining signature-based and anomaly-based methods to improve accuracy and reduce false positives. Their analysis of signature databases assesses

Nanotechnology Perceptions Vol. 20 No. S8 (2024)

their capacity to identify and guard against current threats, suggesting the integration of signature-based techniques with anomaly-based methods to enhance hybrid IDS efficacy. However, their discussion on anomaly-based detection is limited, raising concerns about the generalizability of their findings. Hocine and Zitouni (2023) [3] propose a collaborative IDS based on a multi-agent system that uses dynamic load balancing for traffic analysis, particularly effective against distributed denial of service attacks. Their system employs decision trees, support vector machines, and neural networks, showing improved performance and scalability with the NSL-KDD dataset compared to state-of-the-art techniques, although it lacks comparative analysis. Maheswaran et al. (2023) [4] address the security and privacy concerns in cloud computing environments, advocating for effective IDS to detect and mitigate suspicious activities.

They discuss the three main IDS methods—signature-based, anomaly-based, and hybrid detection—highlighting hybrid detection for its superior results. The study outlines various IDS types, including host-based, network-based, hypervisor-based, and distributed IDS, each with distinct characteristics and benefits. The importance of datasets like CICIDS2017 for developing and evaluating IDS is emphasized, though the coverage of hybrid detection techniques is limited. Lastly, Gururaj et al. (2022) [5] introduce the Collaborative Energy-Efficient Routing Protocol (CEERP) for 5G/6G wireless sensor networks, addressing energy constraints and enhancing system effectiveness.

CEERP employs reinforcement learning for cluster formation and residual energy-based cluster head selection, with a multi-objective improved seagull algorithm optimizing performance. While CEERP reduces energy consumption and improves network lifespan and efficiency, its implementation requires sophisticated optimization techniques, posing challenges in real-world scenarios, and its effectiveness may diminish with network scalability, especially in large-scale deployments. Collectively, these studies underscore the necessity of integrating advanced machine learning and hybrid approaches in IDS frameworks to effectively counteract the increasing complexity and volume of cyber threats, addressing specific challenges such as scalability, performance, and balancing detection accuracy with false positives. They highlight the evolving landscape of cybersecurity and the continuous need for innovative solutions to protect against sophisticated and emerging cyber threats, emphasizing the critical role of IDS in maintaining the security, privacy, and availability of information systems and cloud-based environments.

3. Existing System

Wireless sensor networks (WSNs) consist of numerous small devices called sensor nodes that collect and transmit data wirelessly. These sensor nodes usually have limited energy sources like batteries, which affects their operational lifespan and network performance. Researchers have developed various energy-saving techniques to enhance the efficiency and longevity of WSNs, especially in modern 5G and 6G networks. One effective approach is clustering, where nodes are grouped and managed by a leader or cluster head to optimize data transfer, thus saving energy and improving efficiency.

The collaborative energy-efficient routing protocol (CEEPR) is a novel technique designed to improve energy efficiency in 5G/6G WSNs, reportedly using 50% less energy and

Nanotechnology Perceptions Vol. 20 No. S8 (2024)

enhancing network lifespan and efficiency compared to current protocols. Simultaneously, the field of cybersecurity has seen the development of Intrusion Detection Systems (IDS), which have become crucial in protecting against cyber threats as the internet has become ubiquitous.

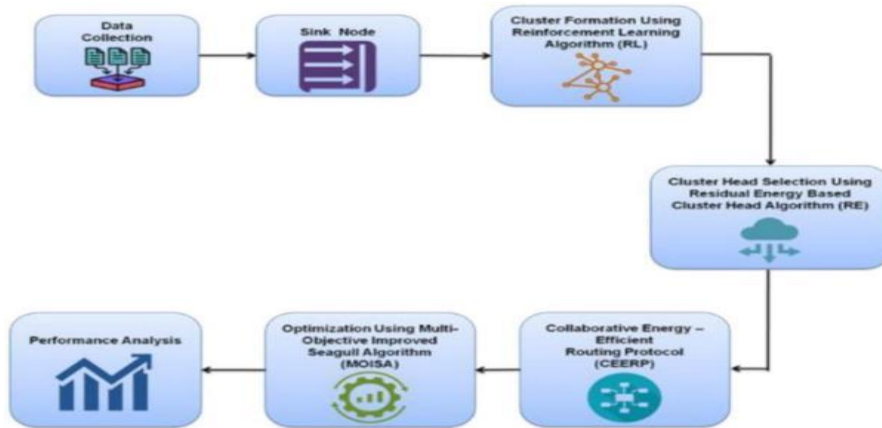


Fig. 3.1. Block Diagram of CEE

Traditional security measures like firewalls and antivirus software often fall short in detecting sophisticated and stealthy intrusion attempts. IDSs, which began emerging in the 1980s, primarily aimed to detect and alert administrators to known patterns of malicious behaviour through predefined rules or signatures. While these signature-based approaches could identify specific types of attacks such as port scanning or denial-of-service attacks, they struggled to keep up with the rapidly evolving tactics of cyber attackers. This gap in detection allowed attackers to employ novel techniques or modify their methods to bypass signature-based detection mechanisms, highlighting the need for more advanced and adaptive security measures in the ever-evolving landscape of cybersecurity.

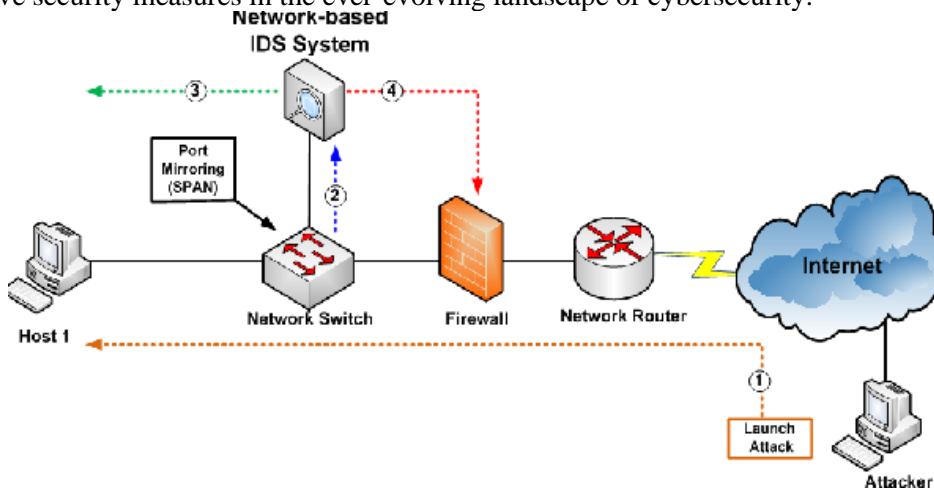


Fig. 3.2. Network-based IDS system

Problem statement

In today's interconnected world, the Internet is essential for communication, commerce, and collaboration, but this connectivity also exposes individuals and organizations to numerous security threats. Malicious actors continuously exploit network vulnerabilities to gain unauthorized access, steal sensitive data, or disrupt services. Traditional security measures, such as firewalls and antivirus software, often fail to detect sophisticated and stealthy intrusion attempts, necessitating the development of Intrusion Detection Systems (IDS) as a crucial component of modern cybersecurity strategies. IDS monitor network traffic for suspicious or anomalous behaviour and alert administrators to potential threats, yet face challenges in accurately detecting intrusions amidst legitimate activity. The dynamic nature of cyber threats, the volume and complexity of network traffic, and the rise of cloud computing [6] and IoT further complicate IDS effectiveness. Moreover, IDS must balance detection accuracy with minimizing false positives and false negatives to avoid unnecessary alerts and undetected intrusions. The problem statement highlights the need for IDS capable of efficiently detecting intrusions within the evolving threat landscape by leveraging advanced techniques like machine learning and dimensionality reduction to develop robust, scalable solutions for network protection.

4. Proposed System

4.1 System Analysis

In the realm of cyber security, system analysis and design play a crucial role in developing effective and robust solutions for detecting and mitigating intrusions. Intrusion Detection Systems (IDS) [7] serve as the front-line defence against malicious activities in network traffic, and their design requires careful consideration of various factors, including system requirements, data analysis techniques, and user interface design.

The process of system analysis begins with a thorough understanding of the problem domain and the specific requirements of the IDS. This involves identifying the types of intrusions that the system needs to detect, the characteristics of normal network traffic, and the operational environment [8] in which the IDS will be deployed. Additionally, stakeholders' needs and expectations must be taken.

4.2 System Design

Once the system requirements have been defined, the next step is system design, [9] which involves translating these requirements into a detailed blueprint for the IDS. This includes defining the system architecture, selecting appropriate algorithms and data analysis techniques, and designing [10] the user interface for interacting with the system.

4.3 System Architecture

In designing the system architecture, considerations must be given to scalability, reliability, and performance. [11] The IDS should be able to handle large volumes of network traffic efficiently, adapt to changes in the network environment, and provide timely alerts to administrators when suspicious activity is detected. This may involve deploying the IDS

across multiple network nodes or utilizing distributed processing techniques to distribute the computational workload.

The selection of algorithms and data analysis techniques is a critical aspect of IDS design. Machine learning algorithms, such as Naive Bayes, Support Vector Machine, and K-Nearest Neighbors, are commonly used for Intrusion Detection due to their ability to analyze large volumes of data and identify patterns indicative of malicious behaviour. However, the choice of algorithm depends on factors such as the nature of the data, [12] the computational resources available, and the desired level of detection accuracy.

In addition to selecting algorithms, dimensionality reduction techniques may be employed to streamline the analysis of high-dimensional data and improve the efficiency of the IDS. Dimensionality reduction methods, such as Principal Component Analysis (PCA) or t-distributed Stochastic Neighbor Embedding (t-SNE), can help reduce the computational complexity of the IDS by reducing the number of features used to describe network traffic while preserving the relevant information.

User interface design is another important aspect of IDS design, as it determines how administrators interact with the system and interpret the results of intrusion detection. The user interface should provide clear and intuitive visualizations of network activity, highlight suspicious events, and facilitate the configuration of system parameters and alerts. Usability testing and feedback from stakeholders are essential to ensure that the user interface meets the needs of its intended users and facilitates effective decision-making in response to detected intrusions.

In conclusion, system analysis and design are critical stages in the development of effective Intrusion Detection Systems. By carefully defining system requirements, selecting appropriate algorithms and data analysis techniques, and designing a user-friendly interface,[13] developers can create IDS that effectively detect and mitigate intrusions in network traffic, thereby enhancing internet security and protecting valuable assets from cyber threat.

4.4 Use Case Diagram

A use case diagram is a Graphical representation of the interactions between different users and a system under an Intrusion detection mechanism. It's a type of behavioural diagram to describe the functionality provided by a system and the interactions between them.

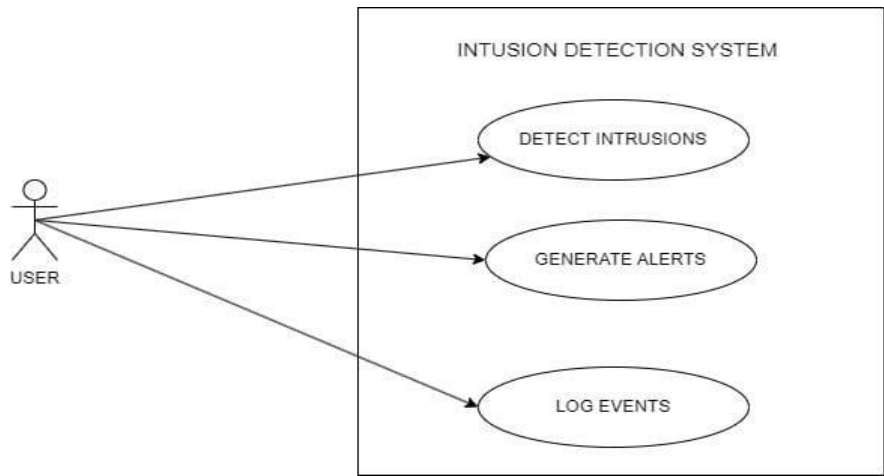


Fig 4.1.Use Case Diagram

4.5 Class Diagram

Machine Learning Algorithm Class:

This class represents the cumulative decision of NB,SVM,and KNN algorithms. It encapsulates the functionality related to training and using the decision tree model for predictions.

train(): This method is responsible for training the data using the provided dataset.

predict(): This method takes input data and returns predictions using the trained decision tree classifier.

Dataset Class:The Dataset class represents the USNW NB 15, NSL KDD datasets used for processing the classifier. It handles operations related to preprocessing the dataset.

preprocess(): This method preprocesses the dataset, which may involve tasks like handling missing values, encoding categorical variables, or scaling numerical features.

Dimensionality reduction class:It aims to reduce the number of features in the dataset while preserving its essential characteristics.

apply(): This method takes input data and applies dimensionality reduction to it, resulting in a dataset with reduced dimensions.

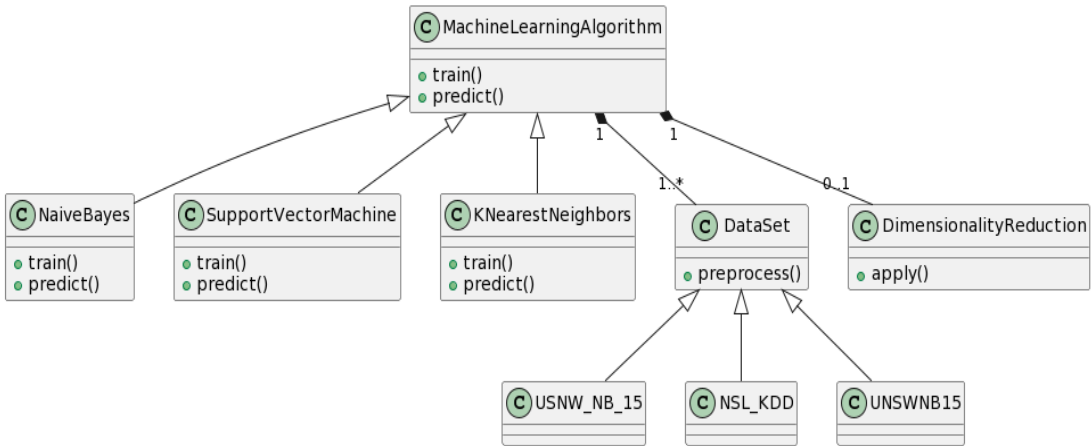


Fig. 4.2.Class Diagram

4.6 Sequence Diagram

It depicts the flow of messages and the temporal ordering of interactions between the user and the system. User can Request Intrusion Detection to the System then the system can Retrieve Data from the Data source and Performa Sequence of Actions like Data Preprocessing, Feature Extraction, Dimensionality Reduction, Algorithm Selection, Model training and Evaluation to detect the Intrusion and Provide the result back to the User.

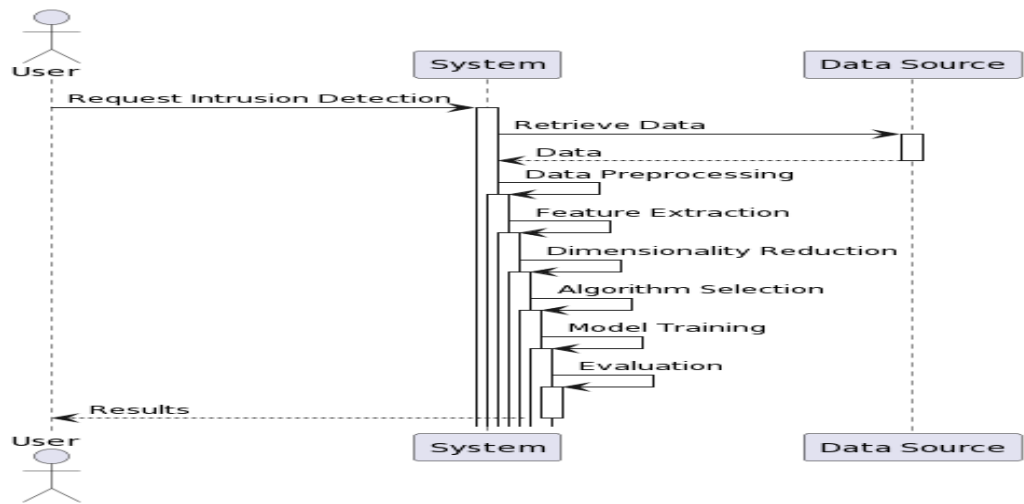


Fig. 4.3.Sequence Diagram

4.7 Activity Diagram

The Activity Diagram represents the dynamic aspects of a system, showing the sequence of actions or steps needed to accomplish Dimensionality reduction to enhance the efficiency of the Intrusion Reduction System.

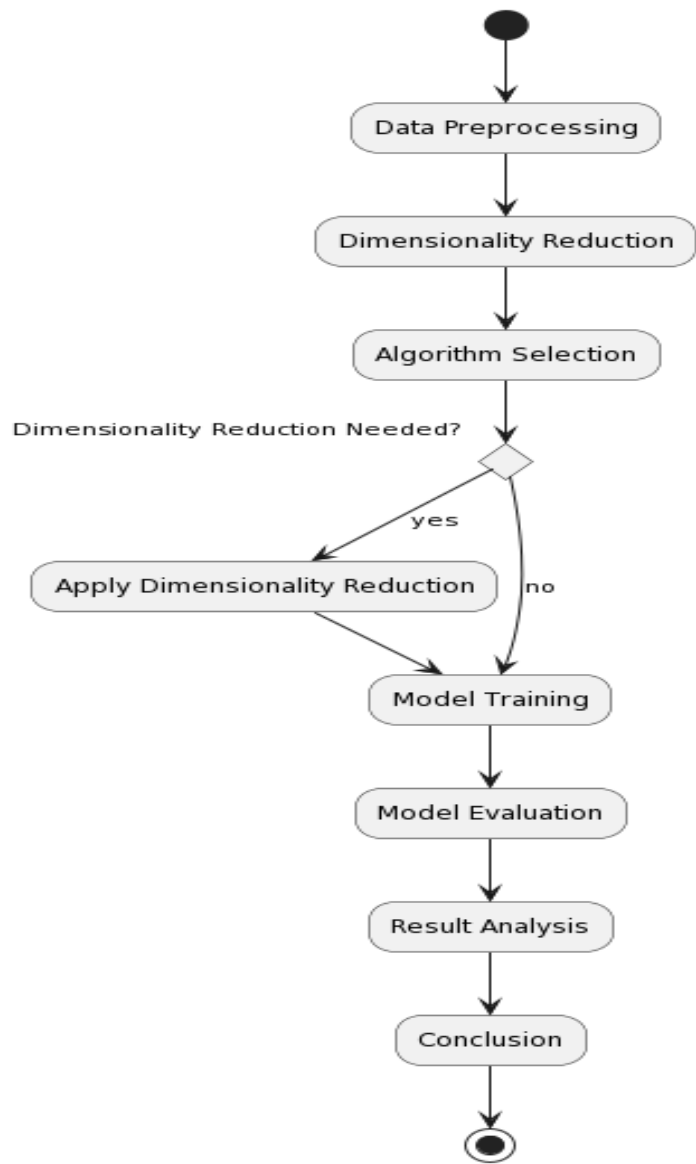


Fig. 4.4.Activity Diagram

4.8 Methodology

The methodology encompasses a systematic evaluation of machine learning algorithms—Naive Bayes (NB), NLP(NLP), and K-Nearest Neighbors (KNN)—for intrusion detection. It includes data preprocessing, Dimensionality reduction techniques, algorithm selection, model training, and evaluation using datasets such as USNW NB 15, NSL-KDD, and UNSWNB15. The study aims to assess the impact of Dimensionality reduction on algorithm performance and provide insights into selecting effective solutions [14] for enhancing internet security through robust Intrusion Detection Systems (IDS).

- **Data Collection:**

The project utilizes three distinct datasets for evaluating the performance of machine learning algorithms in intrusion detection: USNW NB 15, NSL-KDD, and UNSWNB15. These datasets encompass a diverse range of network traffic scenarios and intrusion types, providing comprehensive coverage for experimentation.

- **Data Preprocessing:**

Before model training, the datasets undergo preprocessing steps to ensure data quality and suitability for algorithmic analysis. Preprocessing tasks may include data cleaning, handling missing values, encoding categorical variables, and scaling numerical features[16].

- **Dimensionality Reduction:**

In this project, dimensionality reduction techniques are incorporated to address the challenges posed by high-dimensional feature spaces in intrusion detection datasets. Methods such as Principal Component Analysis (PCA) or Feature Selection are applied to reduce the number of features while preserving relevant information and minimizing computational overhead.

- **Algorithm Selection:**

Three machine learning algorithms—Naive Bayes (NB), NLP(NLP), and K-Nearest Neighbors (KNN)—are chosen for evaluation in intrusion detection. These algorithms represent diverse approaches to classification tasks and are widely used in security applications.

- **Model Training and Evaluation:**

Each selected algorithm is trained on the preprocessed datasets using both original feature sets and reduced dimensions obtained through dimensionality reduction techniques. The performance of each model is evaluated using standard metrics such as accuracy, precision, specificity, sensitivity, and F1 score[z. Cross-validation techniques may be employed to ensure robustness [15]and mitigate overfitting.

- **Comparative Analysis:**

The results obtained from experiments with both original and reduced feature sets are compared to assess the impact of dimensionality reduction on algorithm performance. The strengths and weaknesses of each algorithm in terms of intrusion detection effectiveness are analyzed, considering factors such as accuracy, computational efficiency, and interpretability.

- **Software and Tools:**

The implementation of the methodology is carried out using Python programming language within the Anaconda Navigator environment. Standard machine learning libraries such as scikit-learn are utilized for data preprocessing, dimensionality reduction, algorithm implementation, and performance evaluation.

By following this systematic approach, the study aims to provide valuable insights into the effectiveness of machine learning algorithms for intrusion detection, considering the integration of dimensionality reduction techniques [17] to enhance model performance and efficiency.

4.9 Block Diagram of Intrusion Detection System

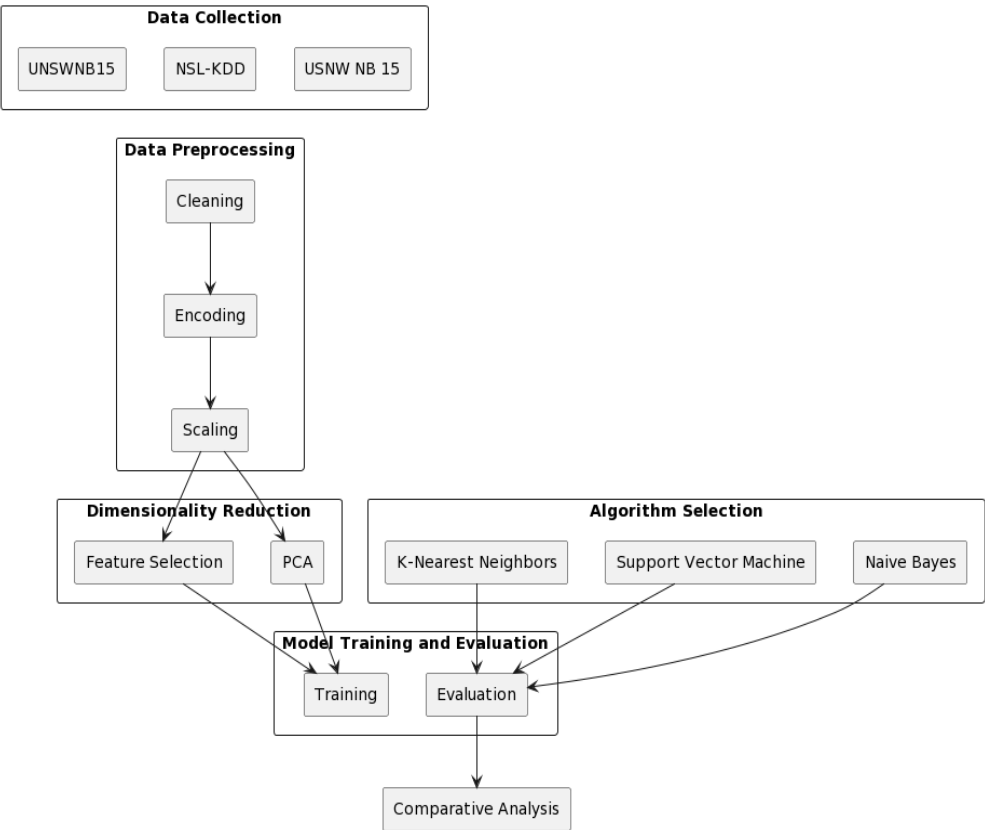


Fig. 4.5. Block Diagram of Intrusion Detection System

5. Results and Discussions

The experiments conducted on the,

USNW NB 15

NSL-KDD and

UNSWNB15

Naive Bayes (NB)

NLP(NLP) and

K-Nearest Neighbors (KNN)

5.1 Overall Output

The experiments conducted on the USNW NB 15, NSL-KDD, and UNSWNB15 datasets reveal NLP's effectiveness as the most reliable intrusion detection algorithm. NLP consistently demonstrated balanced performance across accuracy, precision, specificity, sensitivity, and F1 score metrics, ideal for critical real-world applications. While Naive Bayes (NB) and K-Nearest Neighbors (KNN) showed competitive accuracy, their precision rates suffered, especially with complex network traffic patterns. NB's assumption of feature independence and KNN's curse of dimensionality sensitivity limit their effectiveness, emphasizing algorithm selection importance[18]. Incorporating dimensionality reduction techniques, like Principal Component Analysis (PCA), improved machine learning algorithm performance by mitigating dimensionality curse effects.

5.2 Network Diagram

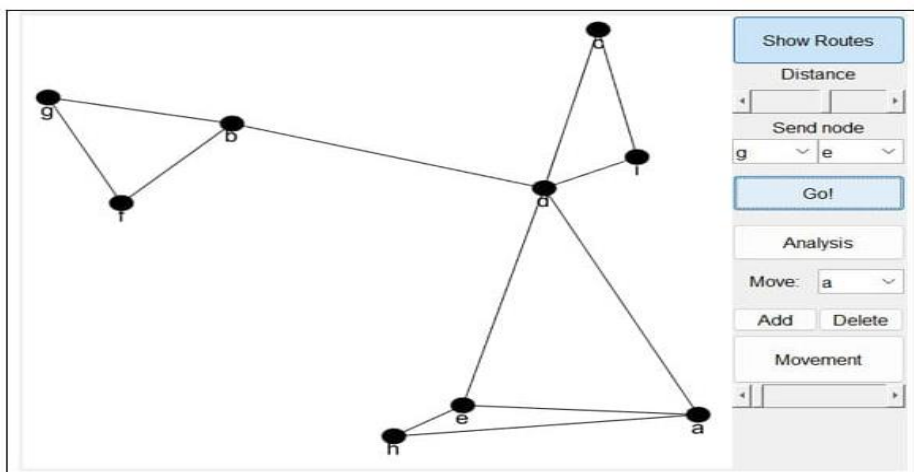


Fig.5.1. Network Diagram

A Network of nine interconnected nodes, denoted from "a" to "i", forms the backbone of an advanced intrusion detection system (IDS) of a Dynamic Network. Each node represents a distinct entity within the network, such as routers, servers, or endpoints, collectively orchestrating the flow of digital communication.

The network's dynamic nature reflects its continuous evolution, with nodes establishing and terminating connections, transmitting data packets, and adapting to varying traffic patterns in real time. This dynamism poses a formidable challenge to traditional intrusion detection methods, necessitating innovative approaches to safeguard the network's integrity.

5.2.1 Simulation Table

Simulation - Table View

<div>SeqNum: 20Node a</div> <table><tr><th>dest</th><th>nextHop</th><th>hopCnt</th><th>seqNum</th><th>lifetime</th></tr><tr><td>1</td><td>g</td><td>d</td><td>3</td><td>14</td><td>1</td></tr></table>	dest	nextHop	hopCnt	seqNum	lifetime	1	g	d	3	14	1	<div>SeqNum: 17Node b</div> <table><tr><th>dest</th><th>nextHop</th><th>hopCnt</th><th>seqNum</th><th>lifetime</th></tr><tr><td>1</td><td>g</td><td>g</td><td>1</td><td>14</td><td>1</td></tr><tr><td>2</td><td>w</td><td>d</td><td>2</td><td>19</td><td>2</td></tr></table>	dest	nextHop	hopCnt	seqNum	lifetime	1	g	g	1	14	1	2	w	d	2	19	2	<div>SeqNum: 15Node c</div> <table><tr><th>dest</th><th>nextHop</th><th>hopCnt</th><th>seqNum</th><th>lifetime</th></tr><tr><td>1</td><td>g</td><td>d</td><td>3</td><td>14</td><td>1</td></tr></table>	dest	nextHop	hopCnt	seqNum	lifetime	1	g	d	3	14	1
dest	nextHop	hopCnt	seqNum	lifetime																																					
1	g	d	3	14	1																																				
dest	nextHop	hopCnt	seqNum	lifetime																																					
1	g	g	1	14	1																																				
2	w	d	2	19	2																																				
dest	nextHop	hopCnt	seqNum	lifetime																																					
1	g	d	3	14	1																																				
<div>SeqNum: 23Node d</div> <table><tr><th>dest</th><th>nextHop</th><th>hopCnt</th><th>seqNum</th><th>lifetime</th></tr><tr><td>1</td><td>g</td><td>b</td><td>2</td><td>14</td><td>1</td></tr><tr><td>2</td><td>w</td><td>e</td><td>1</td><td>19</td><td>2</td></tr></table>	dest	nextHop	hopCnt	seqNum	lifetime	1	g	b	2	14	1	2	w	e	1	19	2	<div>SeqNum: 19Node e</div> <table><tr><th>dest</th><th>nextHop</th><th>hopCnt</th><th>seqNum</th><th>lifetime</th></tr><tr><td>1</td><td>g</td><td>d</td><td>3</td><td>14</td><td>1</td></tr></table>	dest	nextHop	hopCnt	seqNum	lifetime	1	g	d	3	14	1	<div>SeqNum: 18Node f</div> <table><tr><th>dest</th><th>nextHop</th><th>hopCnt</th><th>seqNum</th><th>lifetime</th></tr><tr><td>1</td><td>g</td><td>g</td><td>1</td><td>14</td><td>1</td></tr></table>	dest	nextHop	hopCnt	seqNum	lifetime	1	g	g	1	14	1
dest	nextHop	hopCnt	seqNum	lifetime																																					
1	g	b	2	14	1																																				
2	w	e	1	19	2																																				
dest	nextHop	hopCnt	seqNum	lifetime																																					
1	g	d	3	14	1																																				
dest	nextHop	hopCnt	seqNum	lifetime																																					
1	g	g	1	14	1																																				
<div>SeqNum: 14Node g</div> <table><tr><th>dest</th><th>nextHop</th><th>hopCnt</th><th>seqNum</th><th>lifetime</th></tr><tr><td>1</td><td>w</td><td>h</td><td>3</td><td>19</td><td>2</td></tr></table>	dest	nextHop	hopCnt	seqNum	lifetime	1	w	h	3	19	2	<div>SeqNum: 18Node h</div> <table><tr><th>dest</th><th>nextHop</th><th>hopCnt</th><th>seqNum</th><th>lifetime</th></tr><tr><td>1</td><td>g</td><td>g</td><td>4</td><td>14</td><td>1</td></tr></table>	dest	nextHop	hopCnt	seqNum	lifetime	1	g	g	4	14	1	<div>SeqNum: 17Node i</div> <table><tr><th>dest</th><th>nextHop</th><th>hopCnt</th><th>seqNum</th><th>lifetime</th></tr><tr><td>1</td><td>g</td><td>d</td><td>3</td><td>14</td><td>1</td></tr></table>	dest	nextHop	hopCnt	seqNum	lifetime	1	g	d	3	14	1						
dest	nextHop	hopCnt	seqNum	lifetime																																					
1	w	h	3	19	2																																				
dest	nextHop	hopCnt	seqNum	lifetime																																					
1	g	g	4	14	1																																				
dest	nextHop	hopCnt	seqNum	lifetime																																					
1	g	d	3	14	1																																				

Fig. 5.2. Simulation Diagram

Next Hop: The next hop is the next immediate router or gateway in the path taken by data packets as they travel through a network.

Sequence number (seq no): This indicates the order in which packets are transmitted or received. It helps in tracking the flow of data and ensuring that packets are delivered in the correct order.

Hop count: This represents the number of hops (intermediate devices or nodes) a packet must traverse to reach its destination. Lower hop counts generally imply more efficient routing.

Lifetime: Lifetime typically refers to the remaining battery life or energy level of a node in the network. It's an important metric in energy-constrained networks like wireless sensor networks or IoT devices. Nodes with longer lifetimes can continue to function and participate in the network for a longer duration.

5.2.2 Dynamic Network – [Different Movement of Nodes]

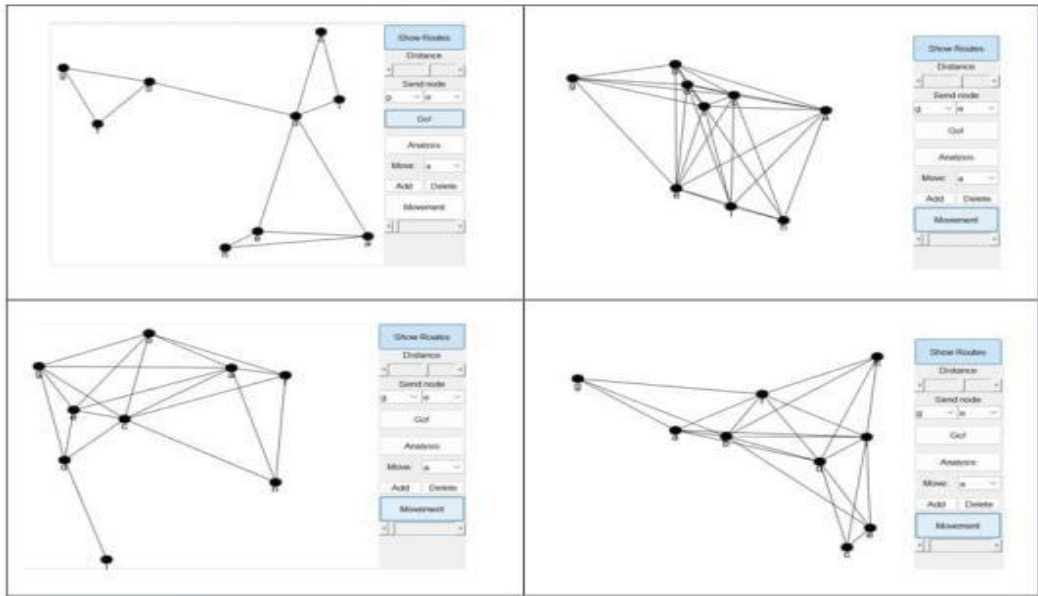


Fig. 5.3. Dynamic Network Diagram

The above diagram represents the Different movements of a Dynamic Network, here we can select the Source Node and Destination Node for the Packet transmission then we need to select the Go for performing the Analysis of Intrusion Detection and to deliver the right path to the Source node for the packet transmission to the Destination Node. Once, the Analysis is completed we can get the output with the comparison of Nodes with Clustering and Nodes without clustering. The output will be the Total Energy consumption, Packet Delivery Ratio, end-to-end delay, Throughput and Total Packets Received. Also, the Output of Accuracy, Precision and F1 score will be generated.

5.3 Output

5.3.1 Total Energy Consumption

It is the amount of energy consumed during the packet transmission by each node and calculates the overall energy of the whole network. Here the Comparison of Energy Consumption is given in terms of (J/m) between the Nodes with Clustering and Without Clustering. The Energy Consumption of Nodes with Clustering is low compared to the Nodes without Clustering.

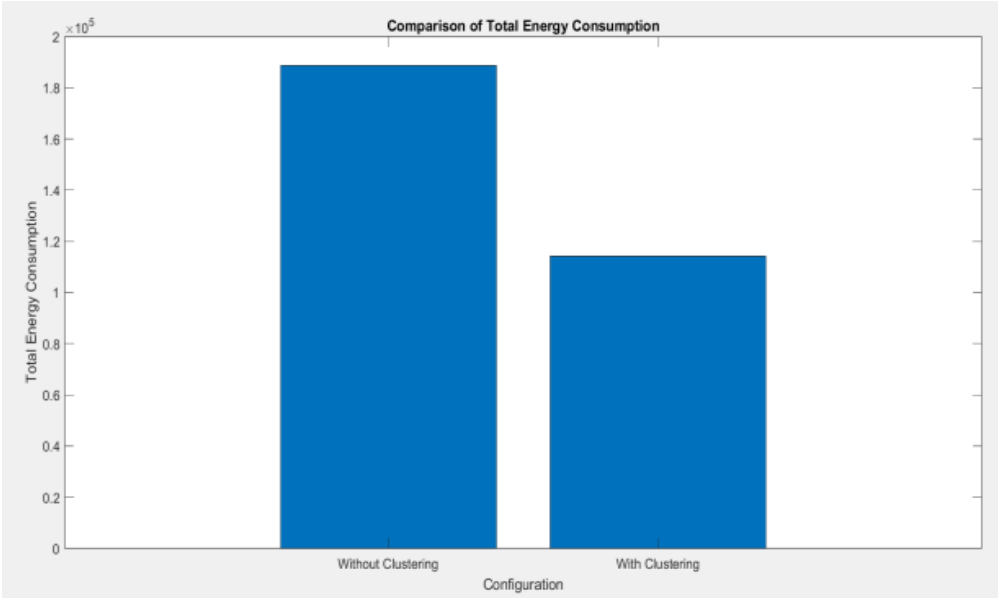


Fig. 5.4. Total Energy Consumption

5.3.2 Packet Delivery Ratio (Pdr)

Packet delivery ratio (PDR) can be measured as the ratio of several packets delivered in total to the total number of packets sent from the source node to the destination node in the network. It is desired that the maximum number of data packets has to be reached to the destination. In our project, the Packed delivery ratio of Nodes with Clustering is High Compared to the Nodes without Clustering.

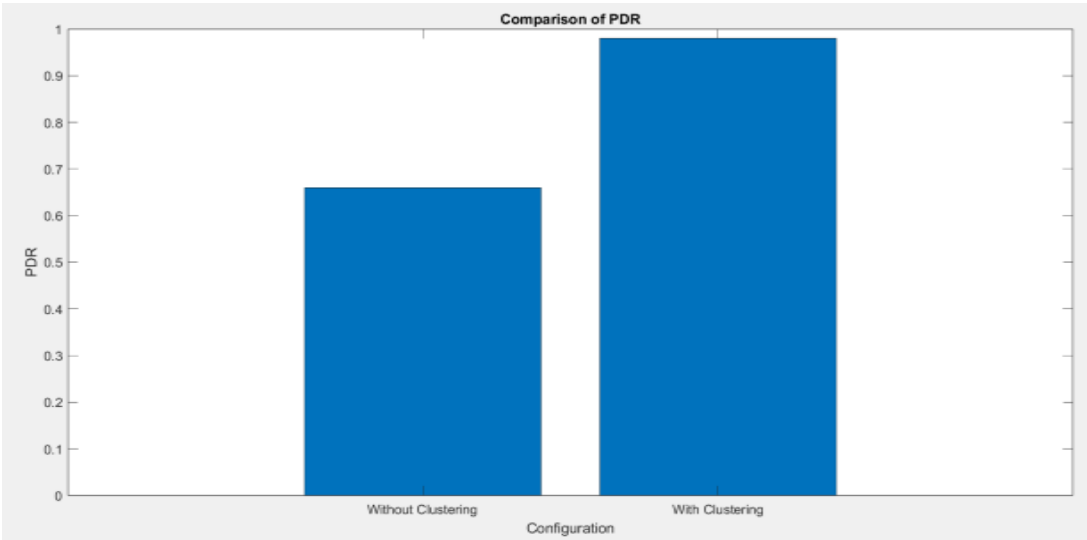


Fig. 5.5. Packet Delivery Ratio

5.3.3 End-To-End Delay (EED)

End-to-end delay, also known as one-way delay (OWD), is the time it takes for a packet to be transmitted from a source to a destination across a network. In our Project, the end-to-end delay in terms of (ms) is Low when the Nodes with Clustering and the end-to-end delay is High When the Nodes are Without Clustering.

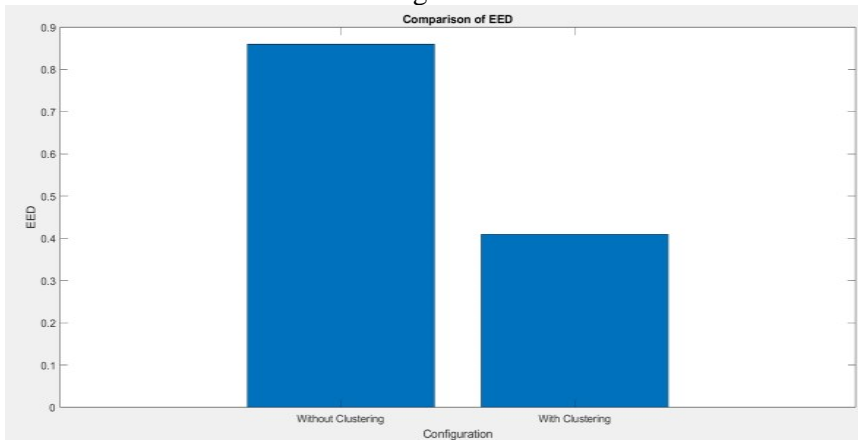


Fig. 5.6. End-to-End Delay

5.3.4 Throughput

The rate at which data is delivered over a network connection in a given period. It's a key performance metric that measures a network's efficiency and capacity. Throughput is usually measured in bits per second (bps). In our project Throughput is High when the Nodes without Clustering and Throughput is Low when the Nodes with Clustering.

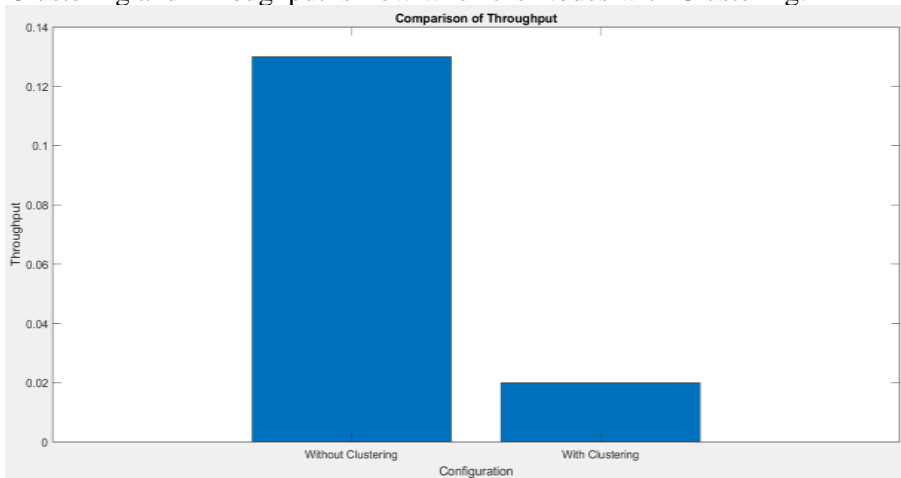


Fig. 5.7. Throughput

5.3.5 Total Packets Received

A packet is a small segment of a larger message that is divided into smaller packets for efficient and reliable transmission of data over computer networks. The receiving computer

or device reassembles the packets to recreate the original content. In our output the Total Packets Received are High when the Nodes are with Clustering and the Total Packets Received is Low when the Nodes are Without Clustering.

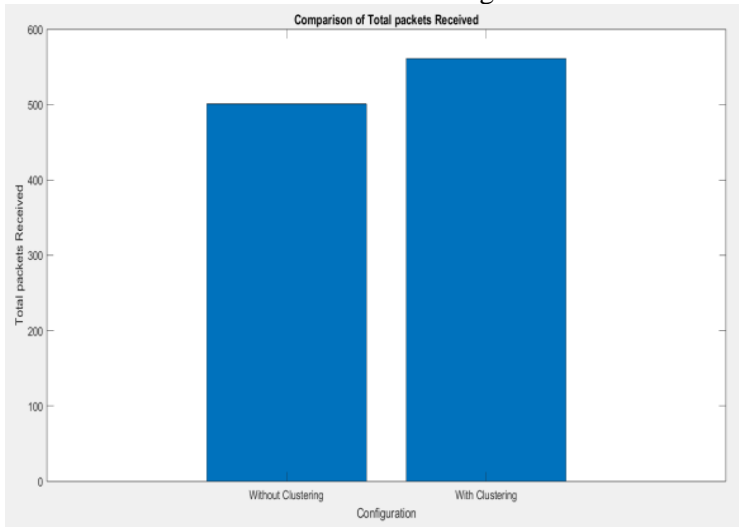


Fig. 5.8. Total Packets Received

5.4 Performance Analysis

Table 5.1. Comparison of Output

PARAMETERS	WITH CLUSTERING	WITHOUT CLUSTERING
TOTAL ENERGY CONSUMPTION (J/m)	38%	62%
PACKET DELIVERY RATIO	60%	40%
END-TO-END DELAY (ms)	32%	68%
THROUGHPUT (bps)	13%	87%
TOTAL PACKETS RECEIVED	53%	47%

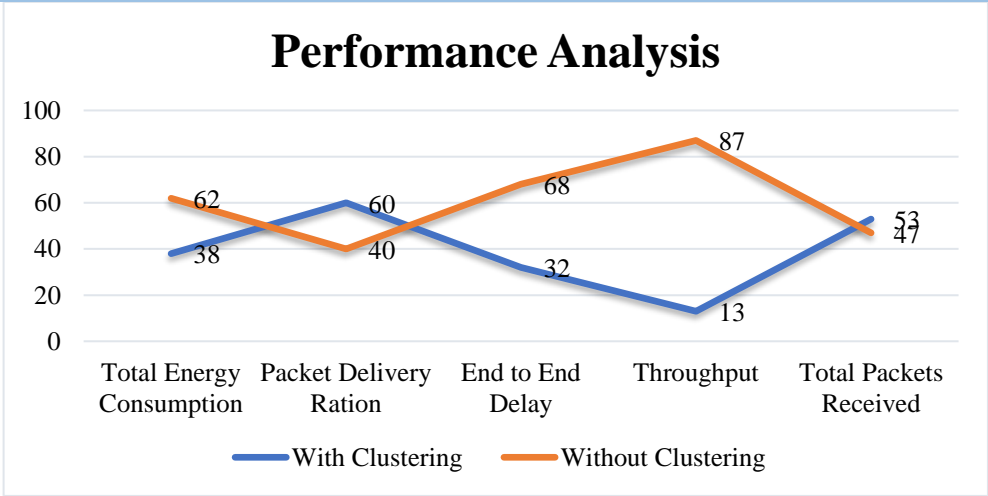


Fig. 5.9. Performance Analysis

5.5 Generated Output

Table 5.2.Output Parameters

PARAMETERS	PERCENTAGE
ACCURACY	99.4
PRECISION	99.61
SPECIFICITY	94.92
SENSITIVITY	99.74
F1 SCORE	99.68

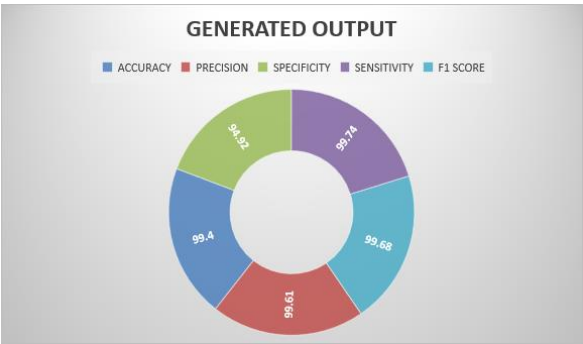


Fig. 5.10. Generation of Output

6. Conclusion

In conclusion, this project highlights the importance of robust algorithm selection and the integration of dimensionality reduction techniques in enhancing the effectiveness of Intrusion Detection Systems (IDS). Through comprehensive experimentation on diverse datasets including USNW NB 15, NSL-KDD, and UNSWNB15, NLP(NLP) emerged as the most reliable algorithm for intrusion detection, demonstrating balanced performance across accuracy, precision, specificity, sensitivity, and F1 score metrics. While Naive Bayes (NB) and K-Nearest Neighbors (KNN) showed competitive accuracy, their performance was hindered by lower precision, particularly in the presence of complex network traffic patterns. Overall, the results of the experiments conducted on the USNW NB 15, NSL-KDD, and UNSWNB15 datasets highlight the effectiveness of Natural Language Processing (NLP) as the most reliable algorithm for intrusion detection.

6.1Future Enhancement

The Incorporation of Dimensionality Reduction methods have proven its benefits in improving the efficiency of classification algorithms and mitigating the curse of dimensionality. By reducing the dimensionality of input data while preserving relevant information, dimensionality reduction techniques helped to enhance the parameters like accuracy, precision, Specificity, sensitivity and F1 score of intrusion detection systems. This project contributes valuable insights into the Machine learning algorithms for Intrusion detection and features selecting appropriate algorithms and incorporating Dimensionality reduction for robust and scalable IDS solutions. More Research can be done in this work by Improving the Throughput of the Network

References

1. A. Kiran, S. W. Prakash, B. A. Kumar, Likhitha, T. Sameeratmaja and U. S. S. R. Charan, "Intrusion Detection System Using Machine Learning (2023)" "International Conference on Computer Communication and (ICCCI) Informatics, Coimbatore, India, pp. 1-4, doi: 10.1109/ICCCI56745.2023.10128363.
2. Agoramoorthy, A. Ali, D. Sujatha, M. R. T. F and G. Ramesh, "An Analysis of Signature-Based Components in Hybrid Intrusion Detection Systems (2023) " 2023 Intelligent Computing and Control for Engineering and Business Systems (ICCEBS), Chennai, India doi: 10.1109/ICCEBS58601.2023.10449209.
3. N. Hocine and C. Zitouni, "A Multi-Agent System Based on Dynamic Load Balancing for Collaborative Intrusion Detection (2023) " 2023 International Conference on Networking and Advanced Systems (ICNAS), Algiers, Algeria, pp. 1-6, doi:10.1109/ICNAS59892.2023.10330527.
4. Mojail, N. Disages K., et al. "Understanding Capacitance and Inductance in Antennas." National Journal of Antennas and Propagation 4.2 (2022): 41-48.
5. N. Maheswaran, S. Bose, S. Sonny, M. Araventh, G. Tharun and R. J, "Effective Intrusion Detection System using Hybrid Ensemble Method for Cloud Computing (2023) " 2023 Second International Conference on Advances in Computational Intelligence and Communication (ICACIC), Puducherry, India, pp.1-5, doi: .1109/ICACIC59454.2023.10435091
6. H.L.Gururaj, RajeshNatarajan, NoufAbdullahAlmujally, FrancescoFlammini, SujathaKrishna andShashiKantGupta (2023) "Collaborative Energy-Efficient Routing Protocol for Sustainable Communication in 5G/6G Wireless Sensor Network".
7. G. Kadiravan, P. Sathish, D. Preethi and S. R, "Dynamic Network Intrusion Detection System for Virtual Machine Environment (2023) " 2023 International Conference on System, Computation, Automation and Networking (ICSCAN), PUDUCHERRY, India, pp. 1-4, doi: 10.1109/ICSCAN58655.2023.10395829.
8. J. Chen, Y. Guo, K. Shi and M. Yang, "Network Intrusion Detection Method of Power Monitoring System Based on Data Mining (2022) " 2022 2nd International Conference on Algorithms, High-Performance Computing and Artificial Intelligence (AHPCAI), Guangzhou, China, pp. 255-259, doi: 10.1109/AHPCAI57455.2022.10087405.
9. M. Zhang, "Design of Network Intrusion Detection System Based on Data Mining (2022) " 2022 International Conference on Electronics and Devices, Computational Science (ICEDCS), Marseille, France, pp. 460-463, doi: 10.1109/ICEDCS57360.2022.00105
10. X. Li, "Research and Design of Network Intrusion Detection System (2022) " 2022 IEEE 2nd International Conference on Power, Electronics and Computer Applications (ICPECA), Shenyang, China, pp. 1069-1072, doi: 10.1109/ICPECA53709.2022.9718920.
11. J. Li, "Network Intrusion Detection Algorithm and Simulation of Complex System in Internet Environment (2022) " 2022 4th International Conference on Inventive Research in Computing Applications (ICIRCA), Coimbatore, India, pp. 520-523, doi: 10.1109/ICIRCA54612.2022.9985720.
12. C. Lu, "Research on the technical application of artificial intelligence in network intrusion detection system (2022) " 2022 International Conference on Electronics and Devices, Computational Science (ICEDCS), Marseille, France, pp. 109-112, doi: 10.1109/ICEDCS57360.2022.00031.
13. S. L. Rocha, G. Daniel AmvameNze and F. L. Lopes de Mendonça, "Intrusion Detection in Container Orchestration Clusters: A framework proposal based on real-time system call analysis with machine learning for anomaly detection (2022) " 2022 17th Iberian Conference on Information Systems and Technologies (CISTI), Madrid, Spain, pp. 1-4, doi: 10.23919/CISTI54924.2022.9820103.
14. GIBSON, KATHARINE, and Y. SALAMONSON. "Image processing application: Overlapping of Images for faster video processing devices." International Journal of

- communication and computer Technologies 11.1 (2023): 10-18.
15. G. Zachos, G. Mantas, I. Essop, K. Porfyrakis, J. C. Ribeiro and J. Rodriguez, "Prototyping an Anomaly-Based Intrusion Detection System for Internet of Medical Things Networks (2022) " 2022 IEEE 27th International Workshop on Computer Aided Modeling and Design of Communication Links and Networks (CAMAD), Paris, France, pp. 179-183, doi: 10.1109/CAMAD55695.2022.9966912.
 16. P. Widulinski and K. Wawryn, "Parameter Efficiency Testing for an Intrusion Detection System Inspired by the Human Immune System (2022) " 2022 29th International Conference on Mixed Design of Integrated Circuits and System (MIXDES),Wrocław,Poland,pp.208-212,doi: 10.23919/MIXDES55591.2022.9838210.
 17. J. Zhao, "Intrusion Detection and Prevention Algorithm of Teacher Management Information System in Private Higher Vocational colleges based on AvroAvro framework (2022) " 2022 6th International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India, pp. 790-794, doi: 10.1109/ICICCS53718.2022.9788118.
 18. B. Xiang, C. Zhang, J. Wang and B. Wang (2022) "Network Intrusion Detection Method for Secondary System of Intelligent Substation based on Semantic Enhancement," 2022 4th International Conference on Electrical Engineering and Control Technologies (CEECT), Shanghai, China, pp. 796-800, doi: 10.1109/CEECT55960.2022.10030264.
 19. W. Yongkang and Z. Meixia, "Research on the Application of Distributed Intrusion Detection System in Campus Network (2024) " 2024 International Conference on Optimization Computing and Wireless Communication (ICOCWC), Debre Tabor, Ethiopia, pp. 1-4, doi: 10.1109/ICOCWC60930.2024.10470621.
 20. Nandhini, C., and G. Vijaiprabhu. "Diverse Techniques on Digital Image Compression Techniques: A Meta Analysis on Medical Images." International Journal of Pharmacy Research & Technology (IJPRT) 14.2 (2024): 13-22.