INTELLIGENT INTRUSION DETECTION SYSTEM USING MACHINE LEARNING

MINOR PROJECT-1 REPORT

Submitted by

MANINDRA. J

ESWARA VENKATA SAI. J

YASWANTH SAI. N

Under the Guidance of

Dr. MUTHUKUMARAN. D

in partial fulfillment for the award of the degree

of

BACHELOR OF TECHNOLOGY

in

ELECTRONICS & COMMUNICATION ENGINEERING

Vel Tech
Rangarajan Dr. Sagunthala
R&D Institute of Science and Technology
(Deemed to be University Estd. u/s 3 of UGC Act, 1956)

APRIL 2024

Vel Tech
Rangarajan Dr. Sagunthala
R&D Institute of Science and Technology
(Deemed to be University Estd. u/s 3 of UGC Act, 1956)

BONAFIDE CERTIFICATE

Certified that this Minor project-1 report entitled **"INTELLIGENT INTRUSION DETECTION SYSTEM USING MACHINE LEARNING"** is the bonafide work of **"MANINDRA. J (21UEEC0109), ESWARA VENKATA SAI. J (21UEEC0112) and YASWANTH SAI. N (21UEEC0209)"** who carried out the project work under my supervision.

SUPERVISOR

Dr. D. MUTHUKUMARAN
Assistant Professor
Department of ECE

HEAD OF THE DEPARTMENT

Dr.A. SELWIN MICH PRIYADHARSON
Professor
Department of ECE

Submitted for Minor project-1 work viva-voce examination held on:_____

INTERNAL EXAMINER

EXTERNAL EXAMINER

ACKNOWLEDGEMENT

We express our deepest gratitude to our Respected Founder President and Chancellor **Col. Prof. Dr. R. Rangarajan**, Foundress President **Dr. R. Sagunthala Rangarajan**, Chairperson and Managing Trustee and Vice President.

We are very thankful to our beloved Vice Chancellor **Prof. Dr. S. Salivahanan** for providing us with an environment to complete the work successfully.

We are obligated to our beloved Registrar **Dr. E. Kannan** for providing immense support in all our endeavours. We are thankful to our esteemed Dean Academics **Dr. A. T. Ravichandran** for providing a wonderful environment to complete our work successfully.

We are extremely thankful and pay my gratitude to our Dean SoEC **Dr. R. S. Valarmathi** for her valuable guidance and support on completion of this project.

It is a great pleasure for us to acknowledge the assistance and contributions of our Head of the Department **Dr. A. Selwin Mich Priyadharson**, Professor for his useful suggestions, which helped us in completing the work in time and we thank him for being instrumental in the completion of third year with his encouragement and unwavering support during the entire course. We are extremely thankful and pay our gratitude to our Minor project -1 coordinator **Dr. Kanimozhi T**, for her valuable guidance and support on completing this project report in a successful manner.

We are grateful to our supervisor **Dr. D. MUTHUKUMARAN**, Associate Professor ECE for providing me the logistic support and his/her valuable suggestion to carry out our project work successfully.

We thank our department faculty, supporting staffs and our family and friends for encouraging and supporting us throughout the project.

MANINDRA. J

ESWARA VENKATA SAI. J

YASWANTH SAI. N

TABLE OF CONTENTS

ABSTRACT

With the increasing sophistication of cyber threats, the need for robust and adaptive intrusion detection systems (IDS) has become paramount in safeguarding sensitive digital assets. Traditional rule-based IDS are inadequate in detecting novel and evolving attacks, leading to a demand for intelligent systems capable of learning and adapting to dynamic threats. This paper presents an intelligent intrusion detection system leveraging machine learning (ML) algorithms for enhanced detection accuracy and efficiency.The proposed system integrates various ML techniques, including supervised, unsupervised, and semi-supervised learning, to analyze network traffic patterns and identify anomalous behavior indicative of intrusion attempts. By training on labeled datasets containing both normal and malicious activities, supervised learning algorithms such as Support Vector Machines (SVM), Random Forest, and Neural Networks can effectively classify incoming network traffic in real-time.Furthermore, unsupervised learning algorithms such as K-means clustering and Principal Component Analysis (PCA) are employed for anomaly detection, enabling the system to detect previously unseen threats without explicit training data. The incorporation of semi-supervised learning techniques enhances the system's adaptability by leveraging both labeled and unlabeled data, thereby improving detection capabilities for emerging threats. The system's architecture is designed to handle large volumes of network traffic efficiently, utilizing parallel processing and distributed computing frameworks for scalability. Additionally, feature selection and dimensionality reduction techniques are employed to optimize model performance and reduce computational overhead. Evaluation of the proposed IDS is conducted using standard benchmark datasets as well as real-world network traffic traces. Results demonstrate superior detection accuracy compared to traditional rule-based systems, with the ability to detect previously unknown threats effectively. Moreover, the system exhibits low false positive rates, ensuring minimal disruption to legitimate network activities.

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Due to the rapid increase in the number of applications and organizations using computer networks, security is becoming increasingly important. Most companies use network security tools like antivirus and anti-spam software to protect themselves from network attacks. These tools can't detect complex or new attacks. An IDS enables computer networks and computers to detect and eliminate unwanted intrusions. Identifier systems can collect and process information from various sources within a network or computer, identifying threats that can make people vulnerable, such as misuse and intrusion. IDSs (Intrusion Detection Systems) are systems that continuously monitor and analyze events occurring on a network to detect malicious activity. IDS are now regarded as an important element of the security infrastructure in most companies. By detecting intrusions, companies can deter attacks on their networks. Security professionals could use this method to reduce current network security risks and the complexity of current threats. The procedure of gaining extra approval by gaining access to a database is how attackers to compromise databases, approved users who abuse their assent are how approved users gain access to databases. An IDS identifies assaults that appear to be unusual or harmful in purpose . The existence of various types of intrusions has been identified using different techniques, but there are no heuristics to confirm their accuracy.

The majority of traditional IDS rely on human analysts to distinguish between invasive and noninvasive network data. Because of the considerable time frame necessary to notice an assault, fast attacks are not practical. For network owners and operators, access to the internet is a particularly delicate topic. As the online world offers different hazards, several solutions are designed to avoid internet assaults. The data mining method is used to derive models from massive data sets. The technology behind machine learning and deep learning has enabled a wide range of data mining techniques in recent years. Intrusion detection research uses a variety of techniques, including classifiers, link assessments, and sequence analysis. Data mining is essential for detecting intrusions by machine learning. It can provide insights into possible behaviors based on prior experiences. The most common data mining techniques are a hybrid association, clustering, and classification.

1.1 INTRUSION DETECTION SYSTEM

Intrusion detection is the operation of monitoring events occurring on a network/computer system and analyzing them for warnings of potential events, such as threats or violations of usage policies or standard security practices. IDS mainly focus on detecting potential events, recording information about these events, and reporting recorded information to security administrators. In addition, IDSs are used for other aims such as detecting issues on security policies, reporting present threats, and discouraging individuals from security attacks.

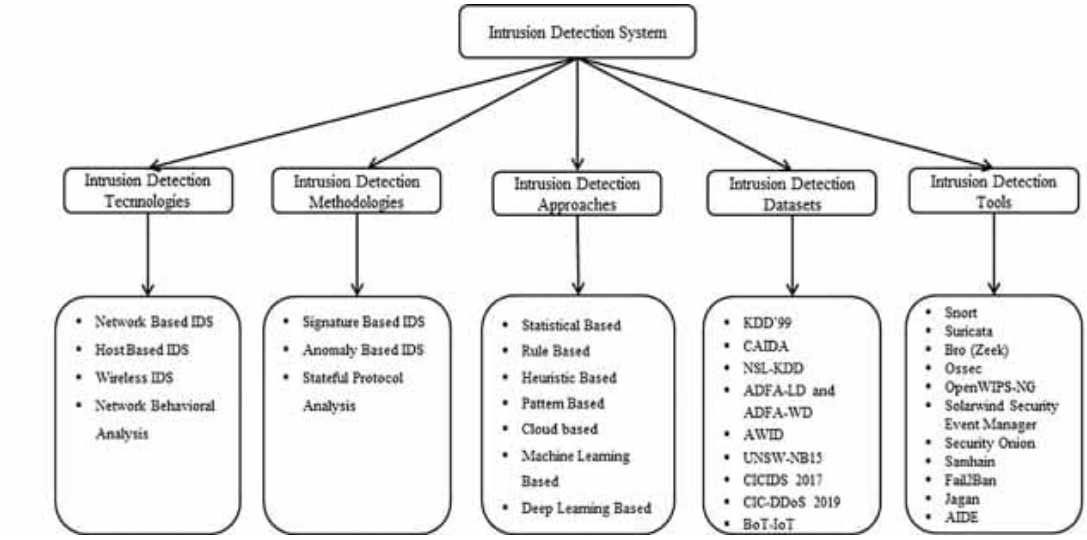

Figure 1.1: **INTRUSION DETECTION SYSTEM**

Generally, for an efficient and effective IDS where components must be properly secured. IDS consist of various components including users, sensors, database servers, management servers and networks. Securing IDSs components is crucial because they are targeted by attackers who want to prevent IDSs from accessing important information, known vulnerabilities or attack detection. The operating systems and applications of all components must be up-to-date, and all software-based IDS components must be protected against threats. It may also be an option to use multiple IDS technologies for comprehensive and high-accuracy detection of attacks. There are various IDS technologies being used such as network-based, wireless, and host-based. Each of them offers fundamentally different information gathering, recording, detection and prevention capabilities. Furthermore, each technology offers advantages such as detecting certain events more efficiently, or detecting with higher accuracy. For example, host-based and network-based IDSs can be integrated to provide an efficient solution. In other words, when choosing IDS technologies, different features and advantages of each technology should be considered. The most common technologies, approaches and methodologies of intrusion detection systems

1.2 MACHINE LEARNING TECHNIQUES

Machine learning techniques are algorithms and methods used to enable machines to learn from data and make predictions or decisions without being explicitly programmed for each task. The creation of analytical models is automated through the use of machine learning, a data analysis tool. A type of artificial intelligence that uses data analysis to detect patterns, recognize trends, and take action based on minimal human involvement. When machines learn from sufficient data and develop models capable of detecting attack variants and new attacks, intelligent intrusion detection systems are able to achieve satisfactory detection levels. Our study is primarily focused on identifying and summarizing IDSs that have utilized machine learning in the past. Here's a breakdown of some common machine learning techniques and their types:

1.2.1 DECISION TREE

Decision Trees are a class of Supervised Learning algorithms that can be used for predicting categorical or continuous variables. It works by breaking data from the root node into smaller and smaller subsets while incrementally building an associated decision tree. Decision nodes create a rule and leaf nodes deliver a result. In addition to CART, C4.5, and ID3, there are numerous other DT models available. Multidecision tree methods such as RF and XGBoost are used to build advanced learning algorithms.

1.2.2 K-NEAREST NEIGHBOR

In KNN, "feature similarity" is used to predict a data sample's class based on its features. By calculating the distance between it and its neighbors, it identifies samples based on their neighbors. A parameter called k affects KNN algorithm performance. The model can overfit with small values of k. Karatas et al. CSE-CIC-IDS2018 has been used as a benchmark dataset to compare the performance of ML algorithms. Selecting very large k values led to incorrect classification of the sample instances. An improvement in detection rate for minority class attacks resulted from using Synthetic Minority Oversampling Technique (SMOTE) to resolve dataset imbalance.

1.2.3 SUPPORT VECTOR MACHINE

A supervised machine learning algorithm that consists of an n-dimensional hyperplane whose elements are spaced closer than the distance between them. Both linear and nonlinear problems can be solved with SVM algorithms. A kernel function is typically applied to nonlinear problems. With the kernel function, an input vector is transformed into a high-dimensional feature space first. After that, the support vectors are used as a decision boundary to determine the maximal marginal hyperplane. NIDS can be improved by using the SVM algorithm to correctly identify normally occurring and malicious traffic.

1.3 PRINCIPLES OF IDS

Intrusion detection is the process of observing events occurring in a computer system or network, and analyzing these events to determine intrusions. There are various threats including malware, DoS-DDoS attacks, unauthorized access, escalation of privileges or probe attack. Although many events that appear to be harmful on the system are indeed attacks, there are some exceptions; for example, the user may mistype the computer's address or unknowingly connect to the wrong system. The system must correctly separate intrusions from the normal network traffic. In conclusion, an IDS is software that simplifies and automates the process of detecting attacks.

There are some important factors for an effective attack resolution when applying IDS technologies:

- System durability/reliability;
- Fast detection;
- Minimal false positives;
- Maximum detection rate;
- Usage minimum software/hardware;
- Ability to accurately detect the location of intrusion;
- Ability to work with other technologies.

1.4 CHALLANGES OF IDS

Intrusion detection systems can be defined as security systems that monitor computer systems and network traffic and use this information to identify external attacks, system abuses, or internal attacks. Today, IDSs are seen as one of the basic security products that should be used in corporate systems. IDSs can be used as a layered security architecture when used with other security products. For example, many use IDSs alongside firewalls and anti-virus software. In this way, IDSs can be used to detect attacks that other security products cannot detect.

IDSs detect attacks with different methods and techniques. Anomaly detection studies from system calls have been going on for many years. However, although a lot of work has been done in this area to produce universal datasets, there are still deficiencies in datasets that should theoretically model all normal behaviors. At the same time, anomaly-based approaches can detect unknown attacks as well as known attacks to a certain extent, while they can also identify normal behaviors as attacks. End users or system administrators should examine the behavior detected by IDS as an attack. Thus, it is possible to extract the correct signature for the application, which was detected through anomaly detection systems and determined to be an attack after analysis. Signature-based systems, on the other hand, can directly detect attacks with their signature, but they cannot detect unknown attacks.

Machine Learning techniques have recently received wide attention in the field of intrusion detection. There are many classification techniques that have proven to be effective in solving a wide variety of problems such as pattern recognition, image processing and cyber security, especially in the field of intrusion detection. However, ML techniques are more useful for estimating between two possible outcomes, such as normal or abnormal, for a given network traffic. The Software Defined Networking (SDN) architecture is based on a centralized control, separating the data plane from the control or management plane, thus providing the ability to program the network. All network devices can be monitored and managed from a central location. Centralized control of SDN can be leveraged to save and improve storage and processing as well. However, there are not any standardized security protocols for SDN. Even though there are some third party service providers, still there exists a security concern. In summary, existing IDSs cannot cope with the dynamic nature of the currently developing attack types.

In the studies to be carried out on these research areas, the development of new methods that will contribute to the literature, the generation of new datasets and the application of new technologies should be included. Another issue is that hybrid IDSs should be created to combine the strengths of IDS types to cover each other's weaknesses, and these systems should be used in real environments. In this study, detailed IDS examination and analysis were made for IDS types, strengths and deficiencies in order to contribute to new technologies that can be developed.

CHAPTER 2

LITERTAURE SURVEY

Wattanapongsakorn et al. proposed a network-based Intrusion Detection and Prevention System (IDPS). The purpose of this system is to effectively detect known attack types and to take immediate action against attacks. The proposed approach can be used with different machine learning techniques and tested on an online network environment. The results show that the proposed IDPS can recognize normal events from attacks within seconds with high accuracy and automatically block the victim's computer network against attacks. Additionally, they applied the C4.5 Decision Tree algorithm with a proposed approach to detect unknown attack types and this algorithm can work effectively when faced with unknown types of network attacks. However, this study can be further improved by developing the approach for the detection of unknown attacks as well as the detection of known attacks.

Amaral et al. proposed a network based intrusion detection system for IPv6-enabled wireless sensor networks. The proposed system detects attacks by using traffic signatures and abnormal behaviors. Proposed system consists of two components PPPSniffer and Finger2IPv6. In the proposed system, network nodes selected as observers are located by the intrusion detection system. In this way, packets exchanged in neighbors are observed and possible attack attempts are detected. The observed messages are compared with the rule set created by NIDS. If a match occurs, an alarm is generated and sent to the Event Management System. With this proposed system, possible misbehaviors can be detected instead of detecting predefined attacks.

Meftah et al. implemented an anomaly-based network intrusion detection approach using the UNSW-NB15 dataset. Their approach consists of two main stages. They use Recursive Feature Elimination and Random Forests among other techniques to select important features for machine learning purposes. Then they perform a binary classification to detect abnormal traffic using different data mining techniques such as Support Vector Machine, Gradient Boost Machine and Logistic Regression. They achieved the highest accuracy result of 82.11 with the Support Vector Machine. They then feed the output of the SVM into a set of polynomial classifiers to increase the accuracy of

detecting attack types. In particular, they evaluated the performance of Naive Bayes, Decision Trees and polynomial SVM. The application of the two-stage hybrid classification increased the accuracy of the results up to 86.04. This work can be further developed on different datasets by developing a new classification algorithm or using deep learning techniques.

Based on fuzzy entropy, Varma et al. proposed a features technique for real-time intrusion detection datasets using Ant Colony Optimization (ACO) techniques. Both discrete and continuous traffic characteristics could be extracted using this technique. In order to determine the most valuable characteristic among the detected characteristics, ACO uses the fuzzy entropy heuristic. Classifiers are therefore more accurate at detecting intrusions. A real-time intrusion threat detection technology provided the best solution for this task.

Using a support vector machine, Thaseenn and Kumar describe an intrusion detection method which is based on chi square properties. By calculating the largest variance for each feature, we improved the parameters of SVM. Reverse variance considerably reduces variance, which improves kernel parameters. Using variance balancing, the SVM parameters were improved in this intrusion detection model. The results improved classification accuracy.

Khammassi Krichen derived the best sample of characteristics from IDS using a featured selection method. To reduce the dataset size, the pre-processed dataset was first re-sampled, and then a wrapper method was used. Genetic algorithms and logistic regression were used in the method. The wrapping technique allows network intrusions to be identified using NBTree, Random Forest (RF), and C4.5 classifiers.

An anomaly-based IDS has been proposed by Aljawarneh et al.. In order to determine which characteristics were most important, a voting method led to an information gain. By doing so, basic learners' probabilities could be integrated. Several classifiers, including REPTree, AdaBoostM1, Meta Paging, Na*ve Bayes, and Random Tree, were implemented to identify network intrusions with the given attributes.

Kabir et al. developed LS-SVM (Least Square Support Vector Machine). There were two stages to this method. The dataset was divided based on arbitrary criteria into preset subgroups. The characteristics that distinguished these groups were then analyzed. They were listed together in the same order. For determining the most efficient allocation method, the variability of the data within subgroups was examined. To extract samples from a network, we used LS-SVM in phase two.

Kumar et al. [24] proposed and evaluated Network Based Intrusion Detection Systems based on machine learning to detect threats to the network. In this study, different supervised machine

learning classifiers are constructed using datasets including labeled examples of network traffic features created by various benign and malicious applications. The main goal of this study is Android-based malware due to the increase in mobile malware and its popularity among users. For testing the proposed approach, traffic was generated. Several malware examples such as Premium SMS sender, backdoor, spammer, bots, ransomware, information stealing and fake antivirus were used to generate this traffic. According to the obtained results, the proposed approach was able to detect unknown and known attacks up to 99.4 accuracy. This study can be improved by enlarging the created dataset and integrating it into the existing intrusion detection systems mentioned.

According to Khan et al. , intrusion detection should be performed in two stages. First, network traffic was categorised using likely score values. When determining if the intrusion was a routine or an assault, deep learning used this likelihood score value as a second measure. The probability score in step two was applied to avoid overfitting. By using this two-stage technique, it is possible to handle large volumes of unlabeled data effectively and automatically.

Based on Convolutional Neural Networks (CNNs) and feature reduction techniques, Xiao et al. developed an intrusion detection model. A step in the process of intrusion detection is the reduction of dimensionality by eliminating irrelevant or redundant characteristics. A CNN algorithm was used to extract features from the reduced data. A supervised learning approach was used to obtain the data that are more successful in detecting intrusions.

According to Qassim et al. [25], anomaly-based intrusion detection system (AIDS) can identify the network traffic that is detected as malicious. It raises an alarm each time when it detects an activity that is different from the normal behaviors. Therefore, managing IDS alarms and distinguishing false positives from true alarms becomes a major challenge. This study proposed an approach consisting of two steps. Firstly, they suggested a set of network traffic features that are supposed to be the most relevant features in detecting anomalies in the network. Secondly, an AIDS alarm classifier proposed to classify activities automatically by a packet header-based anomaly detection system. According to the authors, the proposed system based on machine learning algorithms is effective and efficient in terms of classifying malicious activities.

Among the approaches presented by Zhang et al. to detect network intrusion is a deep hierarchical network. Spatial and temporal aspects of flow were studied using LeNet-5 and LSTM. Various network cascade mechanisms were used to train the deep hierarchical network instead of two. It was also examined how the flow of information in the network varies.

Mazini et al. proposes a new hybrid network-based IDS approach to detect anomaly by using AdaBoost and artificial bee colony (ABC) algorithms. Feature selection was made using the ABC

algorithm. The AdaBoost algorithm was used to evaluate and classify the selected features. The proposed approach was applied to NSL-KDD and ISCXIDS2012 datasets to evaluate the accuracy of the method. 98.9 accuracy rate is achieved. According to the authors, the proposed method outperformed other IDSs on the same dataset. In the future studies, accuracy can be further improved and performance evaluation can be made on different datasets.

NIDSs trained on unstable data incline to offer inaccurate forecasts against small classes of attacks, resulting in undetected or misclassified intrusions. Previous studies have addressed this class imbalance problem using data-level approaches that increase minority-class instances or reduce majority class instances. Although these balancing approaches indirectly improve the performance of NIDSs, they fail to address the underlying problem. In the study of Bedi et al. [31], a two-layer Improved Siam-IDS (I-SiamIDS) approach was proposed to address the problem of class imbalance. I-SiamIDS defines both minority and majority classes as algorithms without using any data level balancing technique. The first layer of I-SiamIDS uses a binary ensemble of Siamese Neural Network, eXtreme Gradient Boosting and Deep Neural Network (DNN) for filtering of input samples. After that, these attacks are sent to the second layer to be classified into different attack classes using the multi-class eXtreme Gradient Boosting classifier (m-XGBoost). Compared to similar studies, I-SiamIDS showed important improvement in recall, accuracy, F1 score, precision and AUC values for both CIDDS-001 and NSL-KDD datasets. In order to present the results more clearly, the computational cost analysis of the proposed method is also given. At the same time, this study can be improved by examining the results on different datasets.

CHAPTER 3

METHODOLOGY AND SIMULATION RESULT

3.1 OVERVIEW

Building an intelligent intrusion detection system (IDS) through machine learning involves a systematic approach encompassing various stages. Initially, requirements are defined, outlining the environment and types of intrusions to detect. Subsequently, labeled datasets comprising both normal and intrusive activities are collected. The collected data undergoes preprocessing, including cleaning, normalization, and partitioning into training and testing sets. Feature selection or extraction follows, aiming to identify pertinent characteristics for discrimination. With features in place, suitable machine learning algorithms are chosen, considering factors like classification accuracy and computational efficiency. Models are then trained using the labeled dataset, their performance evaluated against the testing set, and fine-tuned for optimal results. Post-training, deployment ensues, integrating the model into the IDS infrastructure, while ensuring scalability and real-time monitoring capabilities. Continuous monitoring and periodic updates are vital to adapt the IDS to evolving threats and changes in the network environment. Security and privacy concerns are addressed throughout the process, safeguarding both the IDS and the sensitive data it handles. Finally, comprehensive documentation and reporting capture the entire development journey, providing insights and facilitating future enhancements. This systematic methodology ensures the creation of an effective and robust intrusion detection system leveraging the power of machine learning. Data preprocessing is crucial for preparing the collected datasets for model training. This involves cleaning the data to remove noise and inconsistencies, as well as normalizing or standardizing it to ensure uniformity across features. Additionally, the dataset is partitioned into training and testing sets to facilitate model evaluation. Feature selection or extraction is then undertaken to identify the most relevant characteristics for distinguishing normal behavior from intrusions. This step is instrumental in enhancing the efficiency and effectiveness of the machine learning models.

3.2 Steps for implementing IDS system

1. Define Requirements: Understand the environment where the IDS will be deployed. Define what constitutes normal behavior and what constitutes an intrusion. Determine the types of attacks you want to detect.

2. Data Collection: Gather labeled datasets containing examples of both normal and intrusive activity.Data could include network traffic logs, system call traces, or any relevant data sources.

3. Data Preprocessing: Clean the data to remove noise and inconsistencies.Normalize or standardize the data to ensure that features are on a similar scale.Split the data into training and testing sets.

4. Feature Selection/Extraction: Identify relevant features that can distinguish between normal and intrusive behavior.Extract features from the raw data using techniques like PCA (Principal Component Analysis) or feature engineering.

5. Model Selection: Choose appropriate machine learning algorithms for classification.Common choices include Decision Trees, Random Forests, Support Vector Machines (SVM), Naive Bayes, and Neural Networks. Consider ensemble methods for improved performance.

6. Model Training: Train the selected models using the labeled training dataset.Tune hyperparameters using techniques like grid search or random search to optimize model performance.

7. Model Evaluation: Evaluate the trained models using the testing dataset.Metrics such as accuracy, precision, recall, and F1-score can be used to assess performance.Conduct cross-validation to ensure robustness of the model.

8. Tuning and Optimization: Fine-tune the model based on evaluation results. Experiment with different algorithms, feature sets, and hyperparameters to improve performance.

9. Deployment: Integrate the trained model into the intrusion detection system. Implement real-time monitoring for incoming data streams. Ensure scalability and efficiency of the deployed system.

10. Continuous Monitoring and Updating: Monitor the performance of the IDS in the production environment. Update the model periodically to adapt to new types of attacks or changes in the network environment. Implement feedback mechanisms to improve the model over time.

11. Security and Privacy Considerations: Ensure that the IDS itself is secure against attacks. Handle sensitive data with care and implement appropriate privacy-preserving techniques.

12. Documentation and Reporting: Document the entire process, including data collection, preprocessing, model selection, training, evaluation, and deployment. Report on the performance of the IDS and any insights gained during the process.

3.2.1 BLOCK DIAGRAM

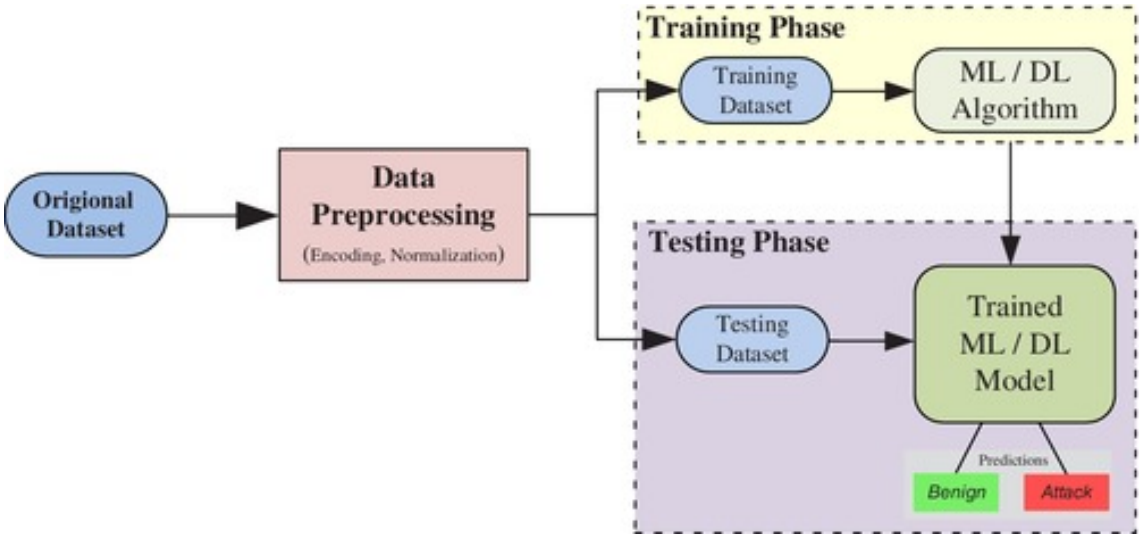

Figure 3.1: **BLOCK DIAGRAM OF IDS SYSTEM**

3.3 METHODOLOGY

- Collect the dataset from the kaggle, dataset must be a KDD-99 or IDS datasets.
- In this we are using the Goggle collab software or anconda navigator in that Jupiter note book software.
- we need to import python libraries like numpy, pandas, seaborn and matplotlib.
- After importing libraries we need to import our machine learning techiques like K-Nearest neighbour, support vector machine, decision tree, then run it.
- Then import the collected datasets in the goggle collab and check for no errors occured, then only we will get the output.
- Train the dataset, test the dataset.
- If the dataset contains string values means, convert it to the int values by one hot encoding method.
- After training the code we need to know the test score and train score.
- In this if we are getting highest accuracy rate for the best of the three techniques then we need to select that machine learning technique.
- Then select the model based on output.
- Record the precision scores for each machine learning techniques.

3.4 RESULT

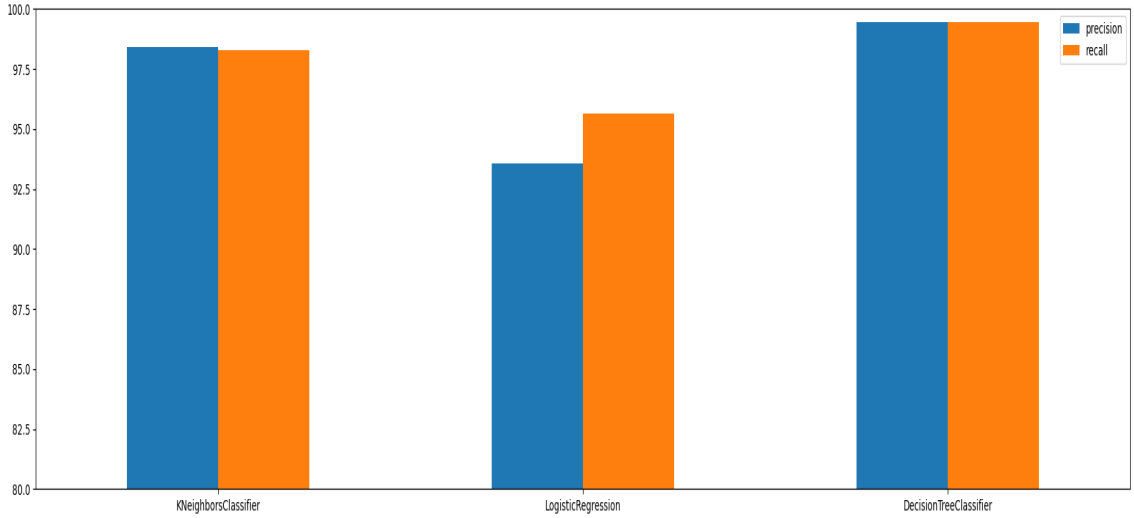

Figure 3.2: **RESULT 1**

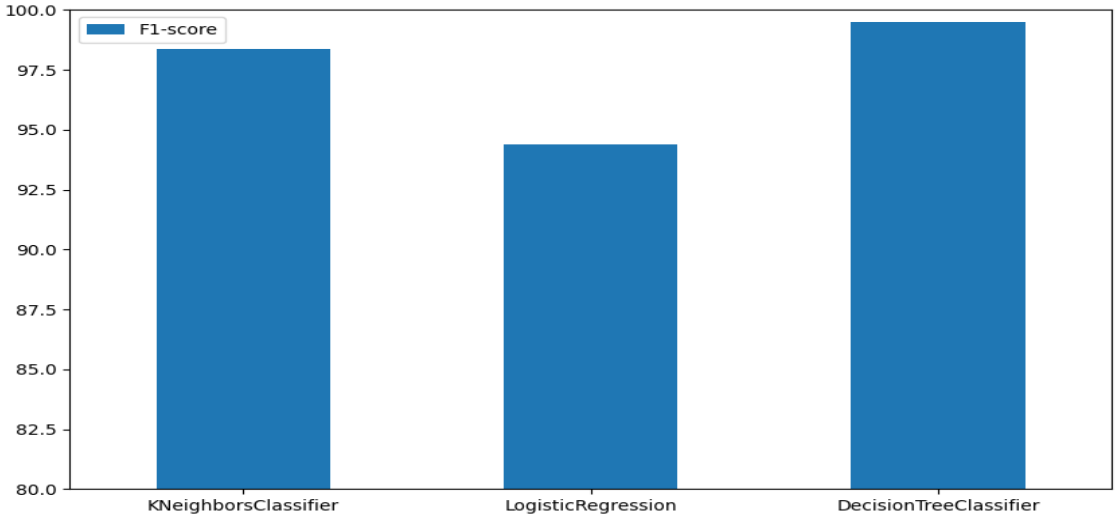

Figure 3.3: **RESULT 2**

CHAPTER 4

CONCLUSION

The integration of machine learning into Intrusion Detection Systems (IDS) marks a significant evolution in cybersecurity. By harnessing the capabilities of machine learning algorithms, IDS have gained the ability to analyze large volumes of data and detect nuanced patterns indicative of potential security threats. This advancement allows IDS to not only identify known attack signatures but also adapt to emerging threats in real-time, enhancing the overall security posture of organizations. A detailed overview of data mining strategies based on IDS mostly in network is offered in this study. The advantages and disadvantages of these strategies are also examined in order to offer future options for improving intrusion detection performance and thereby improving IDS.

The findings of the comparison investigation revealed that insider threat detection utilising deep hierarchical networks had greater accuracy, clarity, and recall. However, the intrusion detection system algorithm's training period is lengthy. Because the efficiency of the system of wireless intrusion detection systems are poorly quantified in the preceding comparisons, a machine learning-based network detection model is presented. Machine learning capabilities that automatically extract and select features reduce the difficulty of calculating domain-specific, manually generated features and allow you to skip the traditional attribute selection phase. Deep learning (DL) is also widely used in a variety of fields and has proven to be effective. Therefore, for the next few years, we will use machines and deep learning algorithms to prevent overfitting with zero elements, address model training issues with a limited percentage of attack classifications, and avoid DNNs. Increases the effectiveness of intrusion detection and prevention. Misunderstandings due to controversial input formation and ultimately solving the problem of instability in cyber attacks

REFERENCES

[1] Graham, J., Olson, R., Howard, R. (Eds.). (2011). Cyber security essentials. CRC Press.

[2] Liao, H. J., Lin, C. H. R., Lin, Y. C., Tung, K. Y. (2013). Intrusion detection system: A comprehensive review. Journal of Network and Computer Applications, 36(1), 16-24.

[3] Liu, H., Lang, B. (2019). Machine learning and deep learning methods for intrusion detection systems: A survey. applied sciences, 9(20), 4396

[4] Hamid, Y., Sugumaran, M., Balasaraswathi, V. R. (2016). Ids using machine learning-current state of the art and future directions. Current Journal of Applied Science and Technology, 1-22.

[5] Masdari, M., Khezri, H. (2020). A survey and taxonomy of the fuzzy signature-based intrusion detection systems. Applied Soft Computing, 106301.

[6] Milenkoski, A., Vieira, M., Kounev, S., Avritzer, A., Payne, B. D. (2015). Evaluating computer intrusion detection systems: A survey of common practices. ACM Computing Surveys (CSUR), 48(1), 1-41.

[7] Witten, I. H., Frank, E. (2002). Data mining: practical machine learning tools and techniques with Java implementations. Acm Sigmod Record, 31(1), 76-77