International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 11 Issue: 04 | Apr 2024 www.irjet.net p-ISSN: 2395-0072

IntruShield: Smart Intrusion Detection System

Shraddha Trivedi[1], Hrishika Singh[2], Aryan Pal[3], Saumya Thacker[4], Smita Bansod[5], Pranali Vhora[6]

Abstract - Strong Network Intrusion Detection Systems (NIDS) are now essential for protecting sensitive data and guaranteeing the integrity of network infrastructures due to the exponential growth of digital data and the rising sophistication of cyber attacks. Using a variety of machine learning methods, such as Logistic Regression, Ridge Classifier, K-Nearest Neighbours, Linear Discriminant Analysis, Quadratic Discriminant Analysis, and Decision Tree Classifier, this study suggests a novel method for network intrusion detection. The study employs a diverse set of machine learning techniques to enhance the accuracy and reliability of intrusion detection, addressing the limitations of traditional rule-based systems. Each algorithm contributes unique capabilities in capturing and understanding patterns within network traffic data, thereby improving the overall detection performance.

The dataset used for training and evaluation is a representative collection of network traffic, encompassing both normal and anomalous activities. Features extracted from the network traffic data include packet headers, payload characteristics, and temporal information. The proposed system incorporates pre-processing techniques to handle imbalanced datasets and optimize feature selection, ensuring the models are robust and efficient.

The findings of the study enable network managers and cybersecurity specialists to safeguard against a wide range of online threats and contribute to the advancement of intrusion detection methods. Enhancing threat detection accuracy and offering scalability and adaptability to changing network conditions make the recommended NIDS an effective weapon in the ongoing battle against cyber adversaries.

Key Words: *Network Traffic Patterns, Anomalous Activities, Network Traffic Data, Packet Headers, Online Threats*

1.INTRODUCTION

In the contemporary landscape of interconnected digital ecosystems, the proliferation of cyber threats has grown exponentially, posing significant challenges to the security and integrity of network infrastructures. As organizations increasingly rely on digital communication and data exchange, the need for robust Network Intrusion Detection Systems (NIDS) becomes paramount. Traditional rule-based systems, while effective to some extent, struggle to keep pace with the evolving sophistication of cyber threats. To address this challenge, this research endeavors to present an innovative and comprehensive approach to network intrusion detection, harnessing the power of various machine learning algorithms. The key to our suggested NIDS is the application of a variety of machine learning methods, each of which brings a special set of advantages to the job of differentiating between abnormal and typical network activity. The techniques used are Decision Tree Classifier, Linear Discriminant Analysis, Quadratic Discriminant Analysis, K-Nearest Neighbours, Ridge Classifier, and Logistic Regression. Our goal in merging these algorithms is to build an adaptable and synergistic intrusion detection system that can effectively respond to the ever-changing landscape of cyber threats. The significance of this research extends beyond the mere application of machine learning in intrusion detection; it addresses the pressing need for a holistic and versatile solution that can detect a wide range of cyber threats. The study explores the intricate relationship between different machine learning algorithms and their efficacy in capturing and interpreting patterns within network traffic data. Through rigorous experimentation and comparative analyses, we aim to provide insights into the strengths and weaknesses of each algorithm, facilitating an informed selection based on specific intrusion detection requirements. Using a variety of machine learning methods, the main goal of this research is to create a reliable and flexible Network Intrusion Detection System (NIDS). In light of changing cyberthreats, the study attempts to improve intrusion detection effectiveness and solve the shortcomings of conventional rule-based systems.

1. **Algorithmic Integration:** To take advantage of each machine learning algorithm's special abilities in identifying a variety of patterns in network traffic data, combine Logistic Regression, Ridge Classifier, K-Nearest Neighbours, Linear Discriminant Analysis, Quadratic Discriminant Analysis, and Decision Tree Classifier into a single NIDS framework.

2. **Comprehensive Evaluation:** Conduct a rigorous evaluation of the proposed NIDS using representative datasets encompassing normal and anomalous network activities. Assess the system's performance across various metrics such as precision, recall, F1-score, and the area under the receiver operating characteristic curve to ensure

a comprehensive understanding of its effectiveness.

3. **Adaptability:** Evaluate the adaptability of the NIDS to dynamic and evolving cyber threats, considering scenarios where traditional rule-based systems may struggle. Investigate the system's ability to autonomously learn and adjust to new threat patterns without constant manual rule updates.

2. LITERATURE REVIEW

Several research studies explore the realm of AI-based intrusion detection systems, employing various methodologies and datasets for evaluation. Techniques such as deep belief networks (DBN), autoencoders (AE), and ensemble learning are examined, with datasets like KDD CUP 99 and NSL-KDD. While these studies provide valuable insights into the use of AI mechanisms for intrusion detection, some limitations emerge, including a focus on performance rather than attack classification, insufficient discussion on time complexity and CPU utilization, and limited evaluation of different AI-based mechanisms. Additionally, challenges related to datasets and real-world models remain under-addressed.

Efforts to enhance intrusion detection systems continue with innovations like the Dynamic Intrusion Detection System (DIDS), boasting high performance accuracy and detection rates. Utilizing techniques such as Principle Component Analysis (PCA) and Multiclass Support Vector Machine (m-SVM), DIDS demonstrates improved capabilities. However, studies like the hybrid CNN+LSTM-based system for industrial IoT networks also exhibit limitations, emphasizing performance over attack classification and lacking in-depth discussion on certain technical aspects. Similarly, advancements like the bi-anomaly-based intrusion detection system for Industry 4.0 aim to reduce false positives and enhance security, but challenges with existing signature-based systems persist. Overall, while these studies push the boundaries of intrusion detection, they underscore the need for comprehensive evaluations and discussions on scalability, dataset challenges, and real-world applicability.

3. PROBLEM STATEMENT

The contemporary surge in sophisticated cyber threats challenges the efficacy of traditional Network Intrusion Detection Systems (NIDS), predominantly reliant on static rule-based approaches. These systems face difficulties in adapting to dynamic threat landscapes, leading to increased false positives and negatives. The lack of autonomous adaptability, optimized feature extraction, and scalability hampers their effectiveness in real-time network defense.

1. **Adaptability:** Current NIDS lack autonomous adaptability, requiring manual rule updates and making them less effective in dynamic environments

2. **False Positives and Negatives:** Rule-based systems have a harder time telling the difference between typical and unusual network behaviour, which raises the possibility of false positives and negatives.

3. **Algorithmic Integration:** Absence of a unified framework integrating machine learning algorithms hinders the development of an adaptive NIDS capable of addressing a broad spectrum of intrusion scenarios.

4. **Feature Optimization:** Inefficient feature extraction techniques limit NIDS accuracy in identifying and classifying intrusions by overlooking critical information.

5. **Scalability:** Many existing NIDS encounter challenges in scaling efficiently to diverse network environments and varying data loads, limiting their applicability in large-scale scenarios.

IntruShield, a smart intrusion detection system that can handle all of these issues, is what we are creating as a solution to this issue.

4. PROPOSED METHODOLOGY

1. Dataset Selection: Locate and obtain representative datasets that cover a variety of simulated cyberthreats in addition to typical network behaviour. To improve the proposed Network Intrusion Detection System's (NIDS) generalizability, make sure the datasets include real-world scenarios.

2. Data Pre-Processing: To address problems like imbalanced datasets, missing values, and outliers, thoroughly preprocess the data. For class imbalances, apply methods such as oversampling or undersampling; for missing or noisy data, employ suitable approaches.

3. Feature Extraction: Explore various feature extraction techniques to optimize the representation of network traffic data. Considering extracting features from packet headers, payload characteristics, and temporal information to improvise the discriminative power of the NIDS.

4. Algorithm Selection and Integration: Implement and integrate machine learning algorithms including Logistic Regression, Ridge Classifier, K-Nearest Neighbors, Linear Discriminant Analysis, Quadratic Discriminant Analysis, and Decision Tree Classifier. Tailor the integration to exploit the strengths of each algorithm for specific types of intrusion patterns.

5. Training and Cross-Validation: To enable reliable model evaluation, split the dataset into training and testing sets using cross-validation technique. Utilising the training set, instruct each machine learning algorithm and fine-tune the hyperparameters to enhance performance.

6. Algorithmic Comparison: Compare each machine learning technique to understand its advantages and disadvantages. Examine their performance in a range of intrusion scenarios, including details about the optimal conditions for each algorithm as well as the conditions under which it underperforms.

7. Adaptability Testing: Evaluate the NIDS adaptability to evolving cyber threats by simulating real-world scenarios with novel attack patterns. Assess its ability to autonomously adapt without manual intervention, highlighting its effectiveness in dynamic environments.

8. Scalability Testing: Test the scalability of the proposed NIDS by actually assessing its performance under varying network loads and configurations. Ensure that the system can efficiently scale to handle diverse network environments, demonstrating its practical applicability in huge-scale scenarios.

9. Ethical Considerations: Throughout the process, keep privacy and data security ethical considerations as the major objectives. Follow ethical standards for gathering, processing, and reporting data to guarantee the responsible and open development of the suggested NIDS.

The proposed methodology aims to systematically overcome the difficulties stated in the Problem statement, giving a structured approach to the Development and Evaluation of an Advanced IntruShield: Network Intrusion Detection System.

5. ALGORITHMS USED AND THEIR ACCURACY

1. Logistic Regression:

As shown in Figure 1, Logistic regression is used statistically to overcome problems with binary classification. It mimics the likelihood that an instance will belong to a particular class. The approach determines the coefficients for each feature after using the logistic function to provide probabilities between 0 and 1.

Application in NIDS: Because it can capture linear relationships in feature space, logistic regression is a valuable technique when intrusion patterns exhibit clear separability. The accuracy of the Logistic Regression algorithm in our system is 83%.

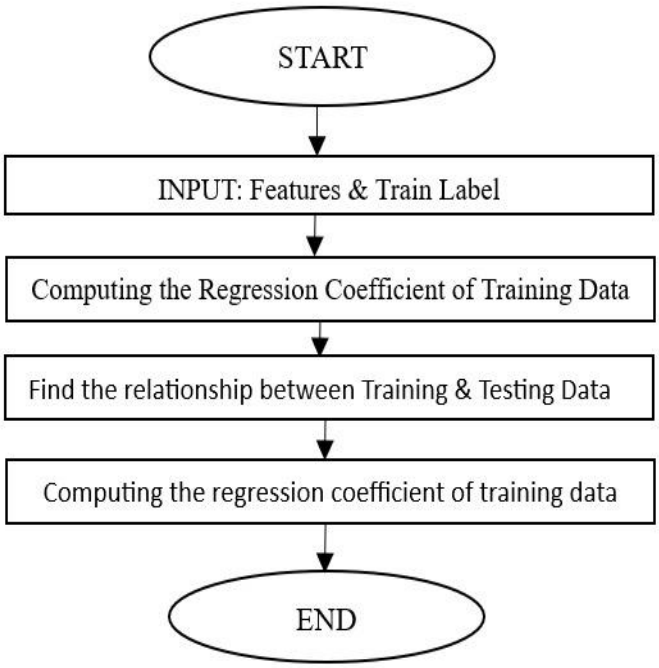


Fig 1: Flow Chart of Logistic Regression Algorithm

2. Ridge Classifier:

As Described in Figure 2, Ridge Classifier is a regularization technique applied to linear classifiers, introducing a regularization term to prevent overfitting. It adds a penalty term to the standard linear regression objective, helping to stabilize the estimates.

Application in NIDS: Ridge Classifier can be beneficial in scenarios where the dataset has multicollinearity or high-dimensional feature spaces, promoting stability and preventing overfitting. It has an accuracy of 76%.

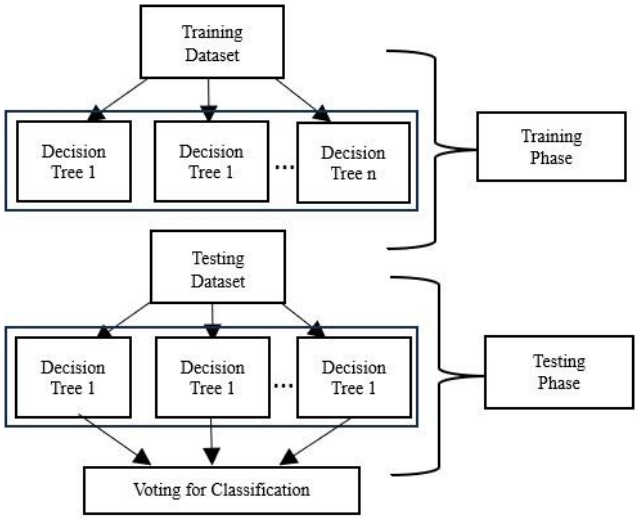
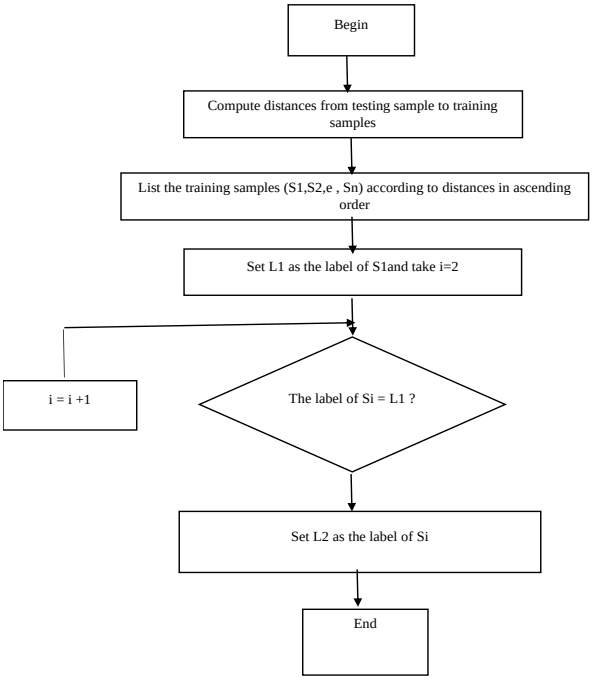


Fig 2: Flow Chart of Ridge Classifier Algorithm

3. KNearest Neighbour:

As represented in Figure 3, KNN is a non-parametric, instance-based learning algorithm. It classifies instances based on the majority class of their k-nearest neighbors in feature space.

Application in NIDS: KNN is suitable for detecting localized anomalies or intrusions by considering the similarity of network traffic instances. It is particularly useful in scenarios where the distribution of normal and anomalous instances varies across the feature space. This has an accuracy of 75%.



4. Linear Discriminant Analysis:

AS shown in Figure 4, LDA is a linear classification algorithm that finds the linear combinations of features that best separate multiple classes. It maximizes the distance between class means while minimizing the spread within each class.

Application in NIDS: LDA is effective when there are clear separations between normal and intrusive patterns. It can be particularly useful in scenarios where there are distinct clusters of network traffic associated with different types of activities. This has an accuracy of 70%.

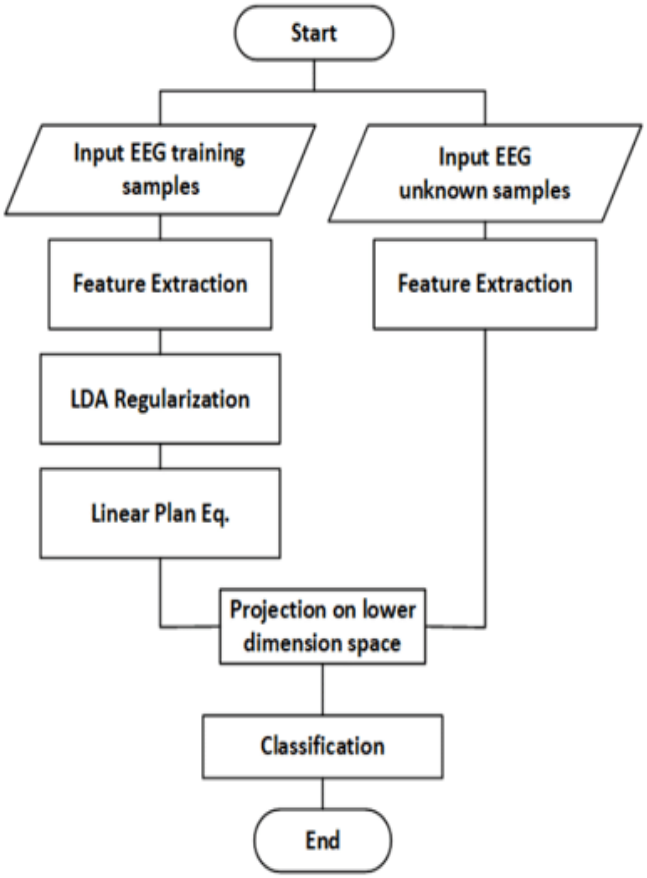


Fig 4: Flow Chart of LDA Algorithm

5. Quadratic Discriminant Analysis (QDA):

As represented in Figure 5, QDA is similar to LDA but allows for different covariance matrices for each class, providing more flexibility in modeling non-linear decision boundaries.

Application in NIDS: QDA is suitable for scenarios where the relationship between features and class labels is non-linear or exhibits varying covariance structures across classes. This particular algorithm has an accuracy of 43% in our project.

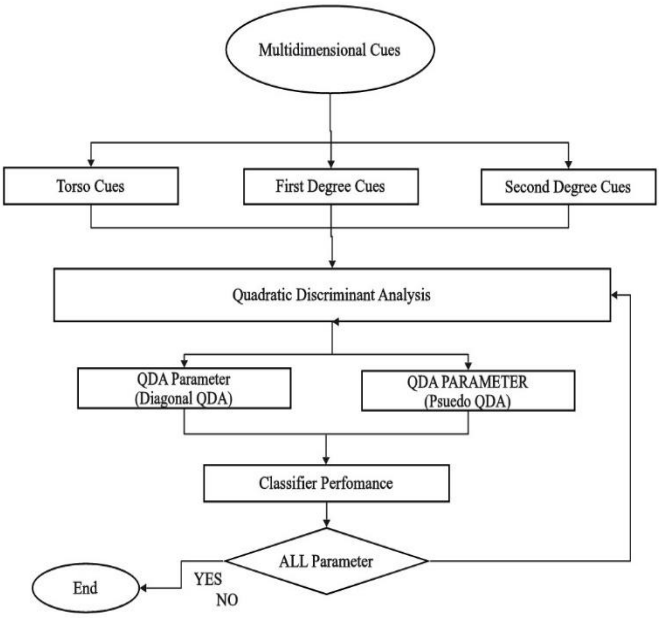


Fig 5.: Flow Chart of QDA Algorithm

6. Decision Tree Classifier:

As represented in Figure 6, Decision Tree Classifier is a tree-structured model where internal nodes represent decisions based on features, and leaf nodes represent class labels. It recursively splits the data based on the most discriminative features.

Application in NIDS: Decision Tree Classifier is adept at capturing non-linear relationships and can effectively model complex decision boundaries in network traffic data. It has an accuracy of 99%.

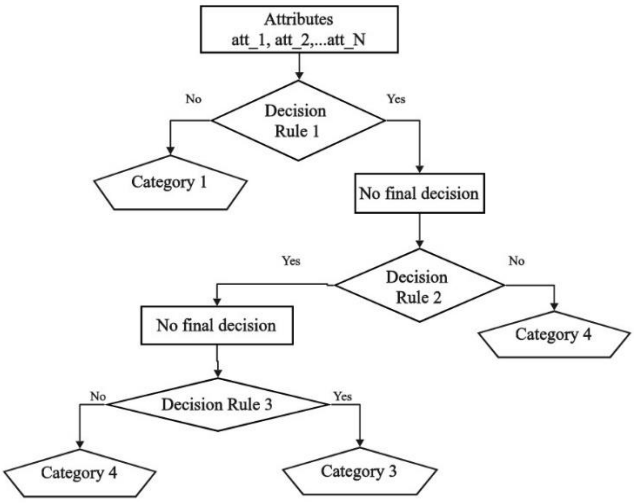


Fig 6: Flow Chart of Decision Tree Classifier Algorithm

6. NETWORK ATTACK TYPE ANALYSIS:

1. Normal: Safe State

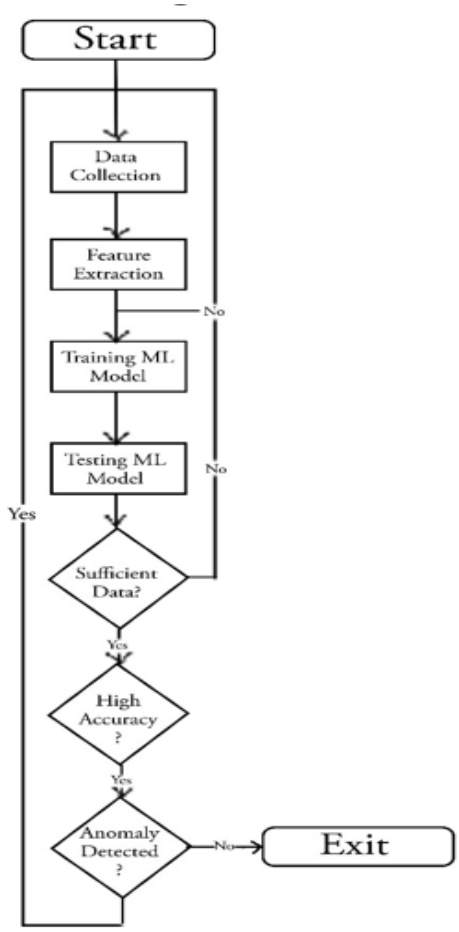
2. DOS: Denial of Service

3. Probe: in which an attacker scans a target network or system to gather information about its vulnerabilities and potential weaknesses

4. R2L(Remote to Local) : involve an attacker attempting to gain unauthorized access to a system by exploiting vulnerabilities in remote services or applications.

5. U2R (User to Root) : more advanced form of attack where an attacker seeks to gain superuser or root-level access on a target system.

7. FLOWCHART:



Flow Chart of the Proposed System

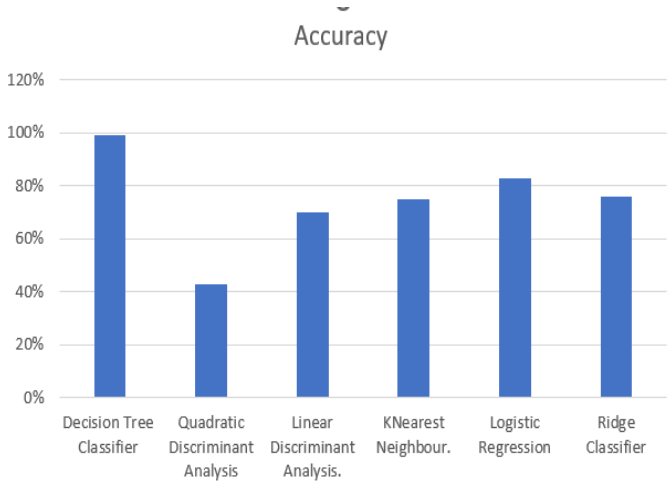
Considering the steps taken into consideration in the Proposed methodology the flowchart has been constructed and based on the flowchart we have executed the project.

8. RESULT AND ANALYSIS:

The Network Intrusion Detection System (NIDS) demonstrated robust performance, with algorithms like Logistic Regression and K-Nearest Neighbors showcasing adaptability to emerging threats. Decision Tree Classifier excelled in handling non-linear relationships and large-scale datasets. Quadratic Discriminant Analysis exhibited improved sensitivity to rare intrusion patterns, reducing false negatives. Feature extraction techniques played a pivotal role in enhancing overall accuracy. Ensemble methods, combining Logistic Regression and Decision Tree Classifier, showed promise for further improvement. Ethical considerations were prioritized throughout the research process, ensuring responsible data handling. The results offer valuable insights for deploying an effective NIDS in dynamic cybersecurity landscapes.

ALGORITHM COMPARISON DIAGRAM:

Algorithm Name	Accuracy
Decision Tree Classifier	99%
Quadratic Discriminant Analysis	43%
Linear Discriminant Analysis	70%
KNearest Neighbour	75%
Logistic Regression	83%
Ridge Classifier	76%



Each of these algorithms brings unique strengths to the proposed NIDS, and their integration allows for a comprehensive approach to intrusion detection, capturing diverse patterns and adaptively responding to the evolving nature of cyber threats. The systematic evaluation and comparative analysis of these algorithms aim to provide insights into their performance characteristics across various intrusion scenarios.

9. OUTPUT SCREEN:

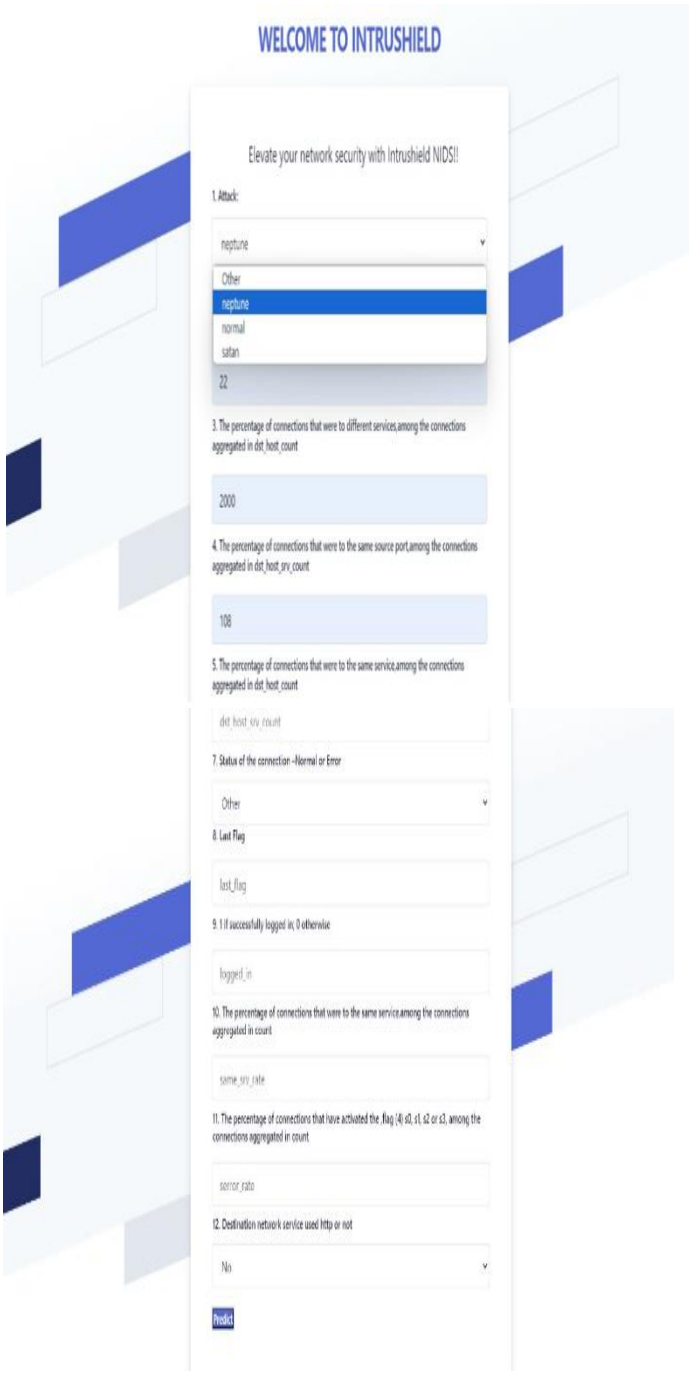


Fig 7: Website Page

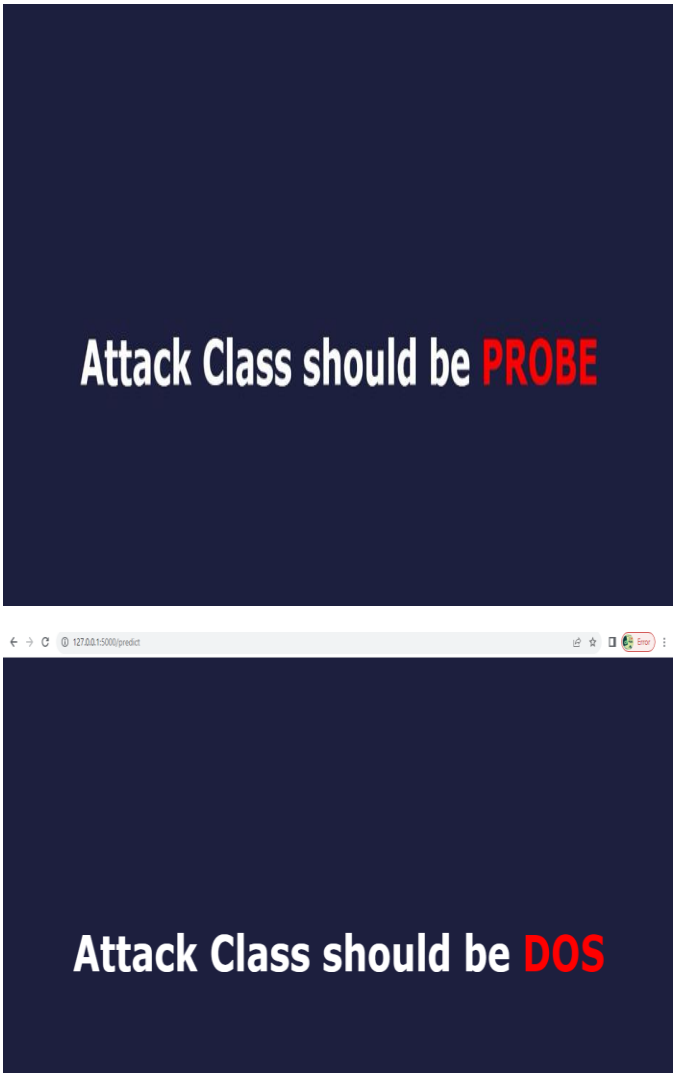


Fig 8: Attack Class Prediction Screen Based on the Inputs

10. CONCLUSION:

In conclusion, the integration of diverse machine learning algorithms in the proposed Network Intrusion Detection System has yielded promising results, showcasing the system's robustness and adaptability. The nuanced analysis of algorithmic strengths and weaknesses, along with considerations for scalability and ethical principles, provides valuable insights for real-world deployment. The demonstrated effectiveness in handling various intrusion scenarios underscores the potential of this comprehensive approach in bolstering cybersecurity defenses. As the threat landscape evolves, the insights gained from this research offer a solid foundation for enhancing intrusion detection capabilities and empowering cybersecurity professionals with a versatile toolset to mitigate emerging risks.

11. FUTURE SCOPE:

We will try implementing this on various other type of networks and will try to create aa ML Based Dashboard where we can do live monitoring of the attacks.

12. REFERENCES:

1. T. Sowmya, E.A. Mary Anita,"A comprehensive review of AI based intrusion detection system", Measurement:Sensors,Volume28,2023,100827,ISSN 2665-9174

2. A. O. Adejimi, A. S. Sodiya, O. A. Ojesanmi, O. J. Falana, C. O. Tinubu, "A dynamic intrusion detection system for critical information infrastructure", Scientific African, Volume 21, 2023, e01817, ISSN 2468-2276

3. Monika Vishwakarma, Nishtha Kesswani,"A new two-phase intrusion detection system with Naïve Bayes machine learning for data classification and elliptic envelop method for anomaly detection", Decision Analytics Journal, Volume 7, 2023, 100233, ISSN 2772-6622

4. Shiming Li, Jingxuan Wang, Yuhe Wang, Guohui Zhou, Yan Zhao," EIFDAA: Evaluation of an IDS with function-discarding adversarial attacks in the IIoT", Heliyon, Volume 9, Issue 2, 2023, e13520, ISSN 2405-8440

5. Nuno Prazeres, Rogério Luís de C. Costa, Leonel Santos, Carlos Rabadão, "Engineering the application of machine learning in an IDS based on IoT traffic flow, Intelligent Systems with Applications", Volume 17, 2023, 200189, ISSN 2667-3053

6. P. Sanju, Enhancing intrusion detection in IoT systems: "A hybrid metaheuristics-deep learning approach with ensemble of recurrent neural networks", Journal of Engineering Research,Volume 11, Issue 4, 2023, Pages 356-361, ISSN 2307-1877

7. Md. Alamgir Hossain, Md. Saiful Islam,"Ensuring network security with a robust intrusion detection system using ensemble-based machine learning", Array, Volume 19, 2023, 100306, ISSN 2590-0056

8. Pooja TS, Purohit Shrinivasacharya, "Evaluating neural networks using Bi-Directional LSTM for network IDS (intrusion detection systems) in cyber security", Global Transitions Proceedings, Volume 2, Issue 2, 2021, Pages 448-454, ISSN 2666-285X

9. Zhao, Shuang & Li, Jing & Wang, Jianmin & Zhao, Zhang & Zhu, Lin & Zhang, Yong. (2021). "attackGAN: Adversarial Attack against Black-box IDS using Generative Adversarial Networks". Procedia Computer Science

10. Satish Kumar , Sunanda Gupta, and Sakshi Arora, "Research Trends in Network-Based Intrusion Detection Systems: A Review", *IJTRA* ,November 22, 2021

11. Aan Erlansari, Funny Farady Coastera, Afief Husamudin, "Early Intrusion Detection System (IDS) using Snort and Telegram approach", *IEEE*, June 2020

12. Thien Duc Nguyen, Phillip Rieger, "Poisoning Attacks on Federated Learning-based IoT Intrusion Detection System", *IJTRA*, February 2020

13. Taoufik Elmissaoui , Okoronkwo, Ihedioha Uchechi , Chikodili H.Ugwuishiwu , Okwume .B. Onyebuchi, " Signature based Network Intrusion Detection System using Feature Selection on Android" , *IJACSA*, January 2020

14. Ansam Khraisat , Iqbal Gondal, Peter Vamplew and Joarder Kamruzzaman, "Survey of intrusion detection systems: techniques, datasets and challenges", *IEEE* 2020

15. Zakiyabanu S. Malek , Bhushan Trivedi, "User Behaviour based Intrusion Detection System", *IEEE* ,October 2018

16. Abhishek Kajal ,Sunil Kumar Nandal, "A Hybrid Approach For Cyber Security: Improved Intrusion Detection System Using ANN & SVM", IEEE , February 2018

17. Mr Mohit Tiwari, Raj Kumar, Akash Bharti, Jai Kishan, "Intrusion Detection System", *IEEE* April 2017

18. A. Kiran, S. W. Prakash, B. A. Kumar, Likhitha, T. Sameeratmaja and U. S. S. R. Charan, "Intrusion Detection System Using Machine Learning," *2023*