iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 **Issue:** VIII **Month of publication:** August 2025

DOI: https://doi.org/10.22214/ijraset.2025.73581

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue VIII Aug 2025- Available at www.ijraset.com

An Efficient NIDS Using Ensemble Techniques for Multinominal Classification

Kundarapu Viveka[1], Dr. M. Dhanalakshmi[2]
[1]*M. Tech, Data Science, Department of Information Technology, Jawaharlal Nehru Technological University Hyderabad, UCESTH, India*
[2]*Professor of IT Dept & Deputy Director of DILT, Department of Information Technology, Jawaharlal Nehru Technological University Hyderabad, UCESTH, India*

Abstract: ***Ensuring robust network security is a critical priority in modern network administration, as systems are constantly exposed to both known vulnerabilities and emerging threats. Intrusion detection plays a vital role in safeguarding these systems by identifying malicious activities that compromise confidentiality, integrity, or availability. To address these challenges, this work presents the development of a Network Intrusion Detection System (NIDS) capable of accurately detecting and classifying multiple categories of cyberattacks, including Denial of Service (DoS), Probe, User to Root (U2R), and Remote to Local (R2L), while maintaining a low false alarm rate, even under high network traffic. The proposed system leverages ensemble learning and advanced data mining classification techniques to enhance detection accuracy and efficiency. Using the benchmark NSL-KDD dataset, which includes attack-type labels and difficulty levels, the model is trained to recognize diverse attack patterns and determine their specific categories. Comprehensive parameter tuning further optimizes performance, enabling the NIDS to serve as a reliable and scalable security solution for real-world network environments.***

Keywords: Network Intrusion Detection System (NIDS), Host-based NIDS, Signature-based NIDS, Ensemble Learning, Multinomial Classification, DoS, Probe, U2R, R2L, Parameter Tuning.

I. INTRODUCTION

Network security has become increasingly critical, making intrusion detection a key focus area to safeguard systems against known vulnerabilities. An intrusion detection system (IDS) inspects all inbound and outbound network communications and identifies suspicious patterns that may indicate a network or system attack from someone trying to break into or compromise a system's security. An IDS can also be used to detect indecorous use and policy violations, like a user downloading large quantities of confidential data. The system then cautions administrators about a possible security breach so that they can take action to stop it. IDS solutions can generally be categorized into two main types: host-based (HIDS) and network-based (NIDS). HIDS is software installed on an end device that analyses system activities, logs, and events to identify suspicious gestures on that device. NIDS functions as a network-wide monitoring mechanism that tracks activity across all connected devices to detect malicious actions. It protects the entire network structure. The NIDS can be further classified as (i) Signature-based NIDS: In this, known attack patterns are identified. When incoming or outgoing traffic aligns with predefined attack signatures, it is flagged as a potential intrusion. (ii) Anomaly-based NIDS: In Anomaly-based NIDS, normal behaviour is established, and if there's a deviation from it, it's considered an intrusion. (iii) Hybrid NIDS: A combination of Signature-based NIDS and Anomaly-based NIDS is Hybrid NIDS. Strengths of both networks can be combined to provide a more effective approach. Various classifier algorithms and ensemble ways can be used for effective discovery. The intrusion detection system can be deployed over a network. It examines network packets and identifies attacks. The four major attacks that cause intrusions are: PROBE, DoS, R2L, and U2R.

II. RELATED WORK

Research says that intrusion detection causes a huge loss of money. Hence many efforts were made to detect intrusions. The intrusion detection system is built using Machine Learning techniques. Unsupervised approach was used in intrusion detection. K-means method is used in the unsupervised approach [4]. Improved Genetic K-Means Algorithm is also an unsupervised technique which is proved to be better than K-means [1]. Modified K-means algorithm builds a high-quality training dataset that contributes significantly to improving the performance of classifiers [15]. The dataset used for intrusion detection is NSL-KDD. It is one of the benchmark datasets. The dataset used previously was KDDCUP'99. Statistical analysis on the KDDCUP'99 data set was conducted and resulted in poor anomaly detection evaluation [2].

©IJRASET: All Rights are Reserved

Intrusion detection can also be done by tree-based classifiers. We can build an effective intrusion detection system using tree-based classification techniques like BF Tree, FT, NB Tree, Random Tree, Random Forest [3,8]. In this paper Towards Near-Real-Time Intrusion Detection for IoT Devices using Supervised Learning and Apache Spark [12], the performances of several machine learning algorithms in identifying cyber-attacks (namely SYN- DOS attacks) to IoT systems are compared. Supervised machine learning algorithms are used that are included in the ML library of Apache Spark, a fast and general engine for big data processing. NIDS was also deployed over the SDN (Software Defined Networks) controller. As NIDS listens to the network and actively compares all traffic against predefined attack signatures, it detects the attacker's scanning attempts [6]. Reinforcement Learning Approach for Anomaly Network Intrusion Detection System has the ability of self-updating to reflect new types of network traffic behavior [13]. Evaluation of machine learning techniques for network intrusion detection [14]. As described in this paper, early research work in this area and commercially available Intrusion Detection Systems (IDS) are mostly signature-based. In this, seven different machine learning techniques were applied with information entropy calculation to Kyoto 2006+ data set and evaluation of performances of these techniques was done. The recent trend is developing the IDS using Machine Learning and Deep Learning [10]. A NIDS developed using ML and DL methods usually involves following three major steps they are: Data pre-processing phase, Training phase, and Testing phase. ML algorithms used for IDS are Decision Tree, K-Nearest Neighbor (KNN), Artificial Neural Network (ANN), Support Vector Machine (SVM), K-Mean Clustering, Fast Learning Network, and Ensemble Methods [7,8]. The tuning of the ML model's parameters is a critical topic since it can improve detection quality. The procedure is called Hyper Parameter Optimization [5]. Deep Learning algorithms include recurrent neural networks, auto encoder, deep belief network and convolutional neural networks [7]. DL methods use deep confidence neural network to extract features of network monitoring data, and uses BP neural network as top-level classifier to classify intrusion types [13]. Deep Neural Network Based Real-Time Instruction Detection System, identifies intrusions by analyzing the inbound and outbound network data in real –time. It consists of a deep neural network (DNN) [9].

III. METHODOLOGY

1) Data Preprocessing and exploratory data analysis: The preprocessing of the data is done and as part of it, outliers are handled. In exploratory data analysis, the data is analyzed considering various features. After analysing and observing the data, KBest selects some features for final interpretation.

2) Building model: Various machine learning algorithms are used, and models are built. Cross validation is done on the models to find algorithms with greater accuracy. Parameter tuning is done improve the performance of the task. Parameter tuning refers to the process of selecting the optimal values for hyperparameters of a model Description of some of the algorithms is as follows.

3) Logistic regression: The logistic regression method is a widely used technique for predicting binary outcomes. Using this method, input variables are mapped to probability values between 0 and 1. Based on network traffic values which are predictor variables, the model predicts the intrusion.

4) K Nearest Neighbors: The k-nearest neighbors (k-NN) algorithm is a classification algorithm that assigns a label to a new data point based on the majority label of its k nearest neighbors in the training dataset. The value of k is a hyperparameter that can be tuned to optimize performance. KNN can be effective in detecting anomalies in network traffic by identifying instances that are significantly different from their neighbors.

5) Discriminant Analysis: Discriminant analysis is a classification method in machine learning that seeks to find a linear combination of features that maximize separation between classes. Discriminate analysis can be used in intrusion detection by finding a linear combination of features that maximally separates normal and intrusive instances in a training set.

6) Decision Tree: Decision trees are a type of algorithm used for classification and regression problems in machine learning. They work by recursively partitioning the data into subsets based on the values of input features. This model can hence detect the intrusion by the recursive partitioning.

7) Neural Network Model: A neural network model is a machine learning model inspired by the human brain structure and function. Neural networks can be used in intrusion detection by learning a non-linear mapping between input features (e.g., network traffic features) and output labels (e.g., normal or intrusive).

8) AdaBoost: Adaptive Boost is a boosting algorithm and a type of ensemble technique. Ensemble refers to combining two diverse algorithms, in order to improve prediction accuracy. AdaBoost combines multiple weak classifiers to create a strong classifier. Therefore, the model will have improved performance.

9) Parameter Tuning: For the improvement of efficiency, the model is tuned. The parameters are changed for a better performance of the model.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

10) Detection of Intrusion: The intrusion is detected based on various features. When the network traffic is given as input to the system, it detects abnormality in the system.

A. Architecture

Initially, data was collected from data set. Then, data is preprocessed by removing missing values, handling outliers. Followed by feature extraction, specific features are selected based on variable reduction where K-Best technique is used. Then data set has two categories: Train dataset and Test dataset. By using Train dataset, a Machine Learning Model is using various techniques like – Logistic Regression, K Neighbours Classifier, Decision Tree, Naïve Bayes and Ensemble predictions like boosting algorithms. Then the model is evaluated for test dataset. Parameter tuning is done in order to improve the efficiency of the model. Finally, intrusions are detected, based on the provided input.

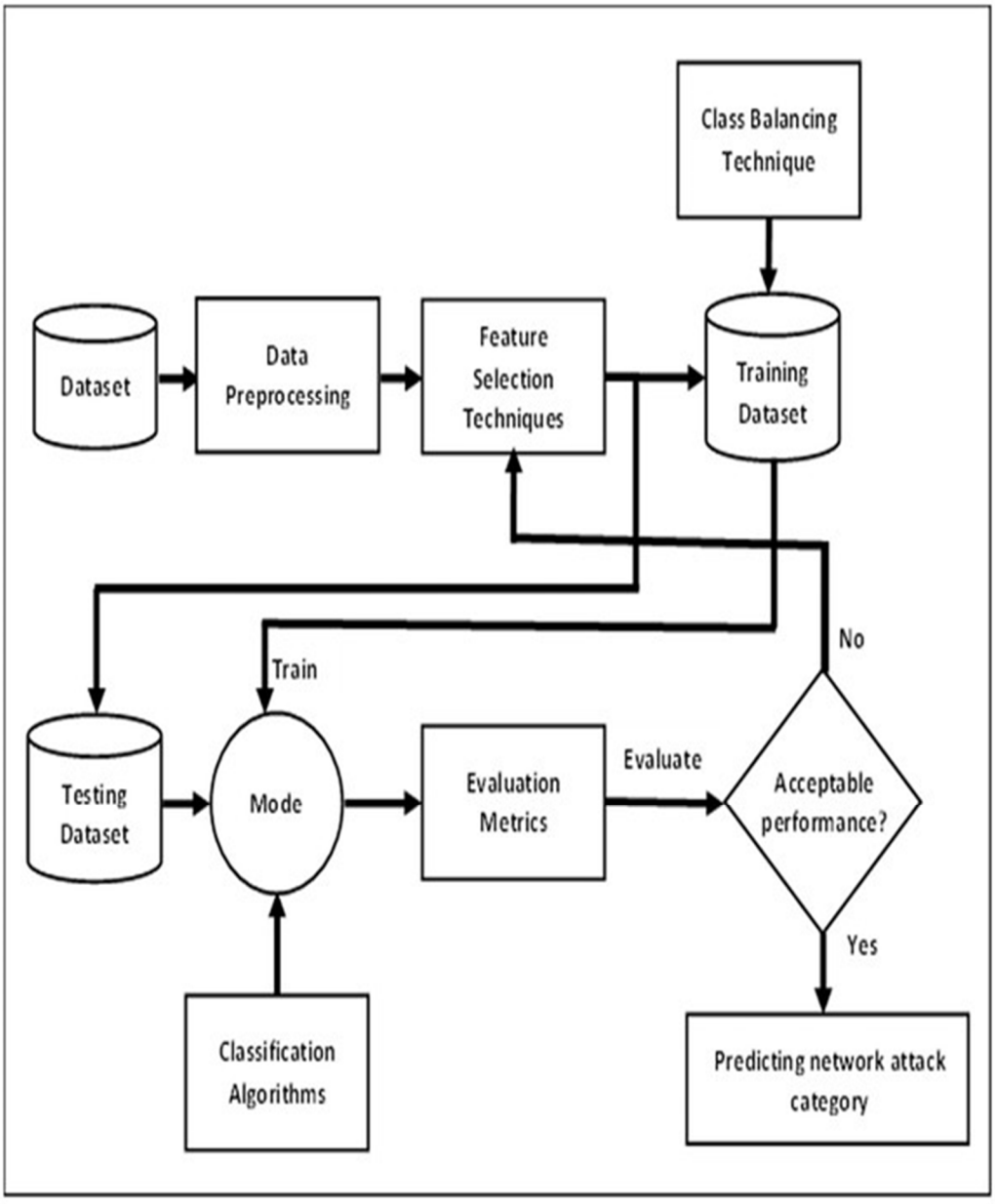


Fig 1 Network Intrusion Detection System Architecture

IV. RESULTS

When we launch the app and open the webpage

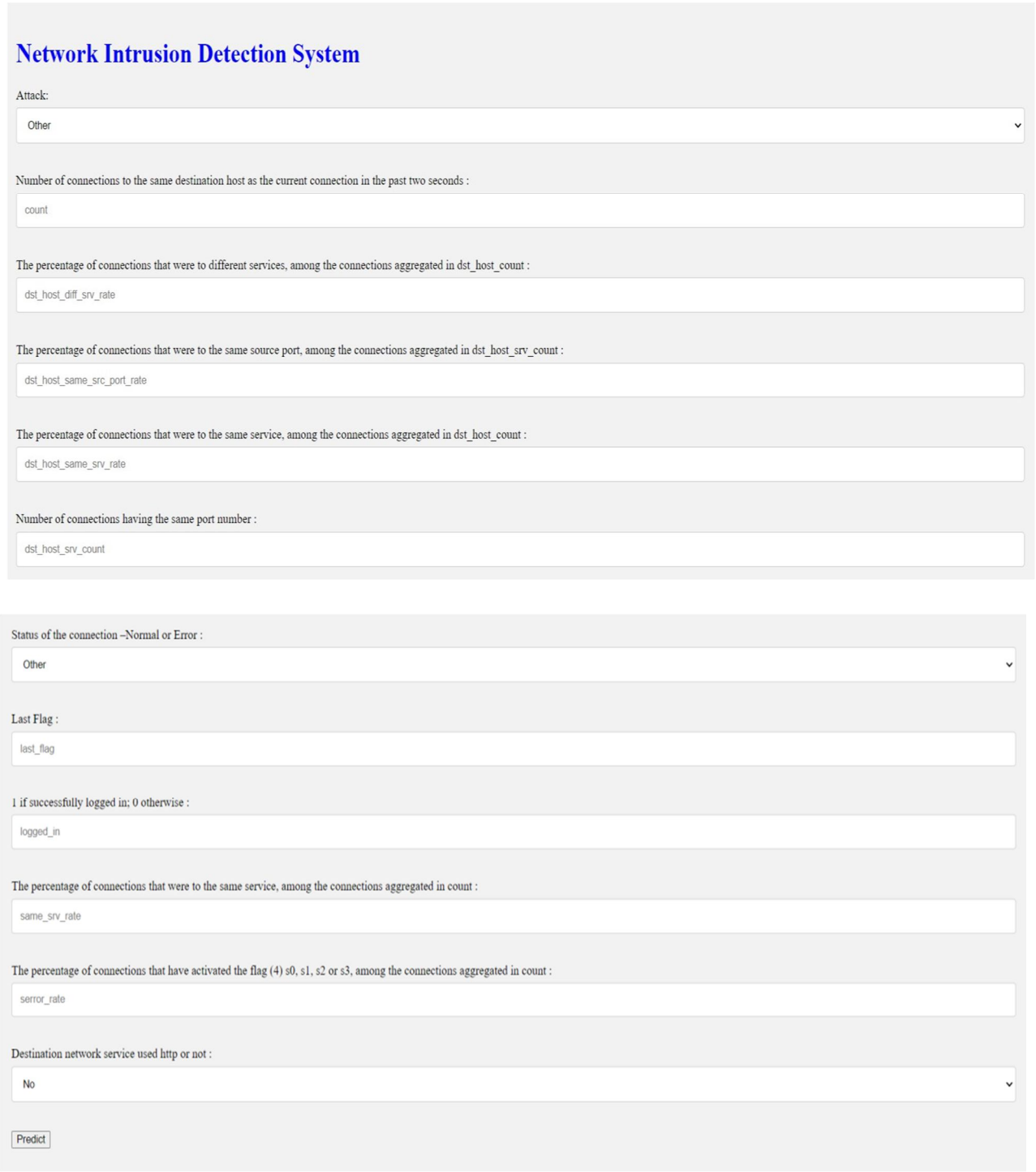


Fig 2 Web Interface

These are the final attributes that are selected using K-Best technique. From the features of dataset, 12 features are selected. These features are used to predict the type of attack. When input values are given about the network traffic, the output describing the attack will be given. Some sample values are as follows:

Case-1:

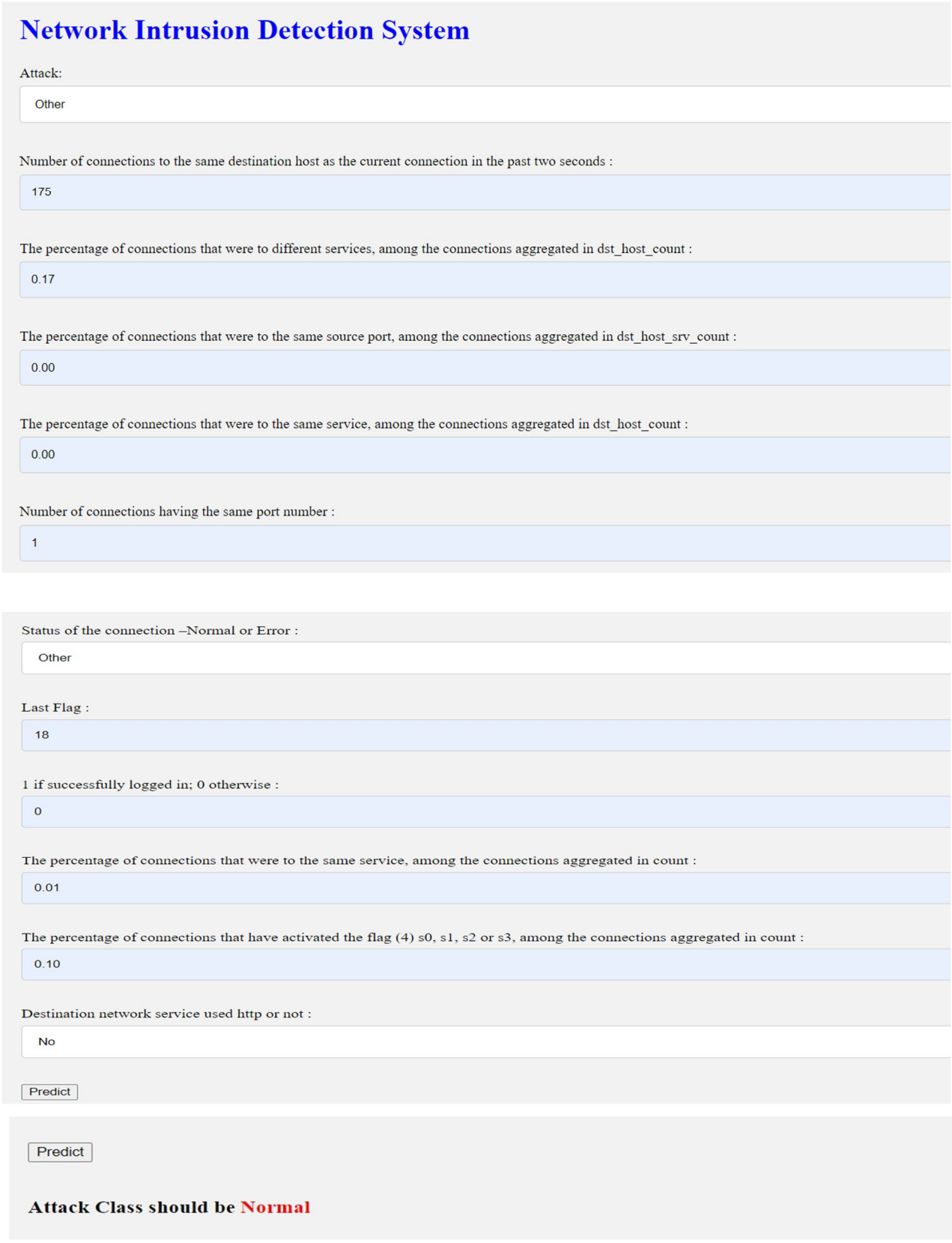


Fig 3 Network Attack 1

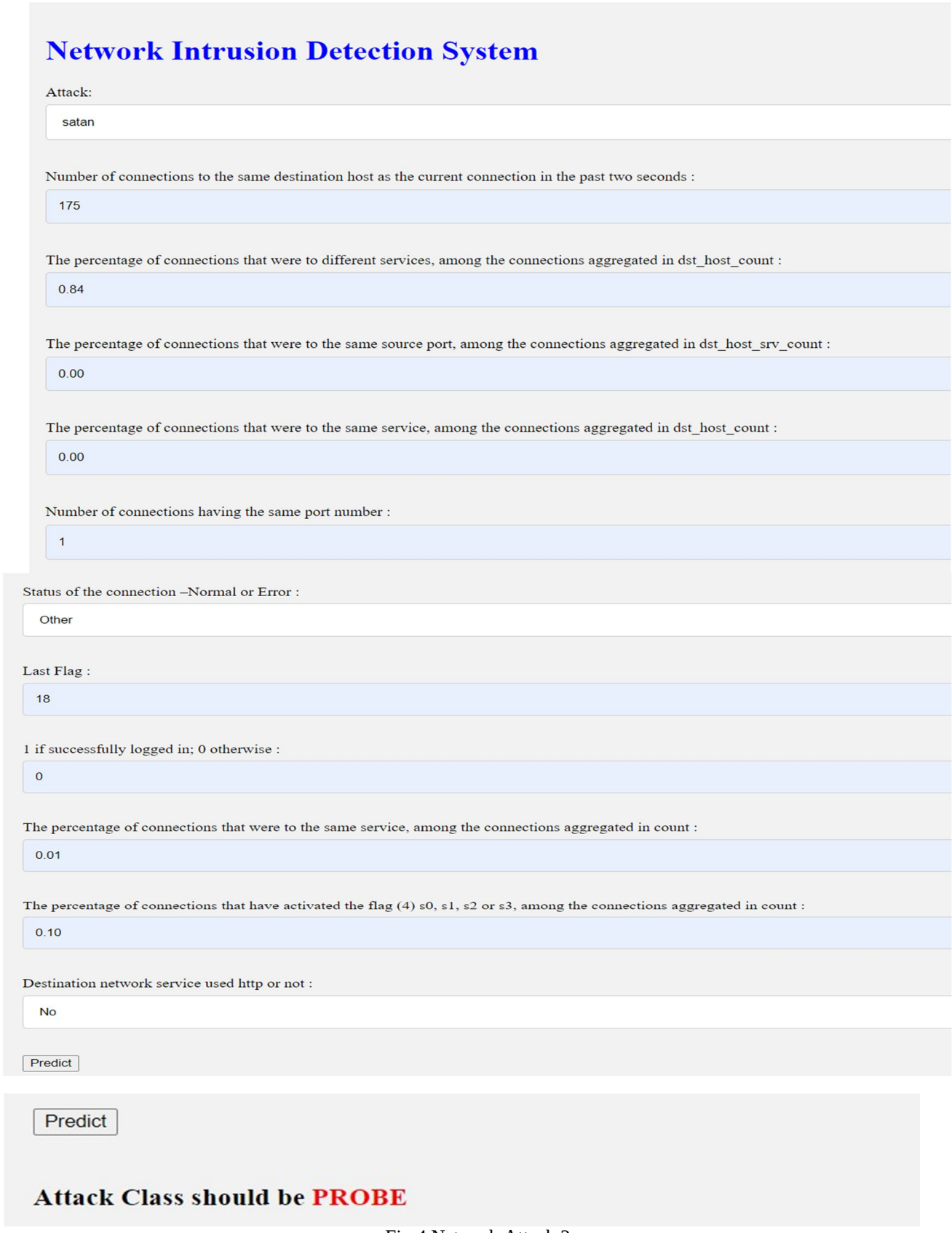


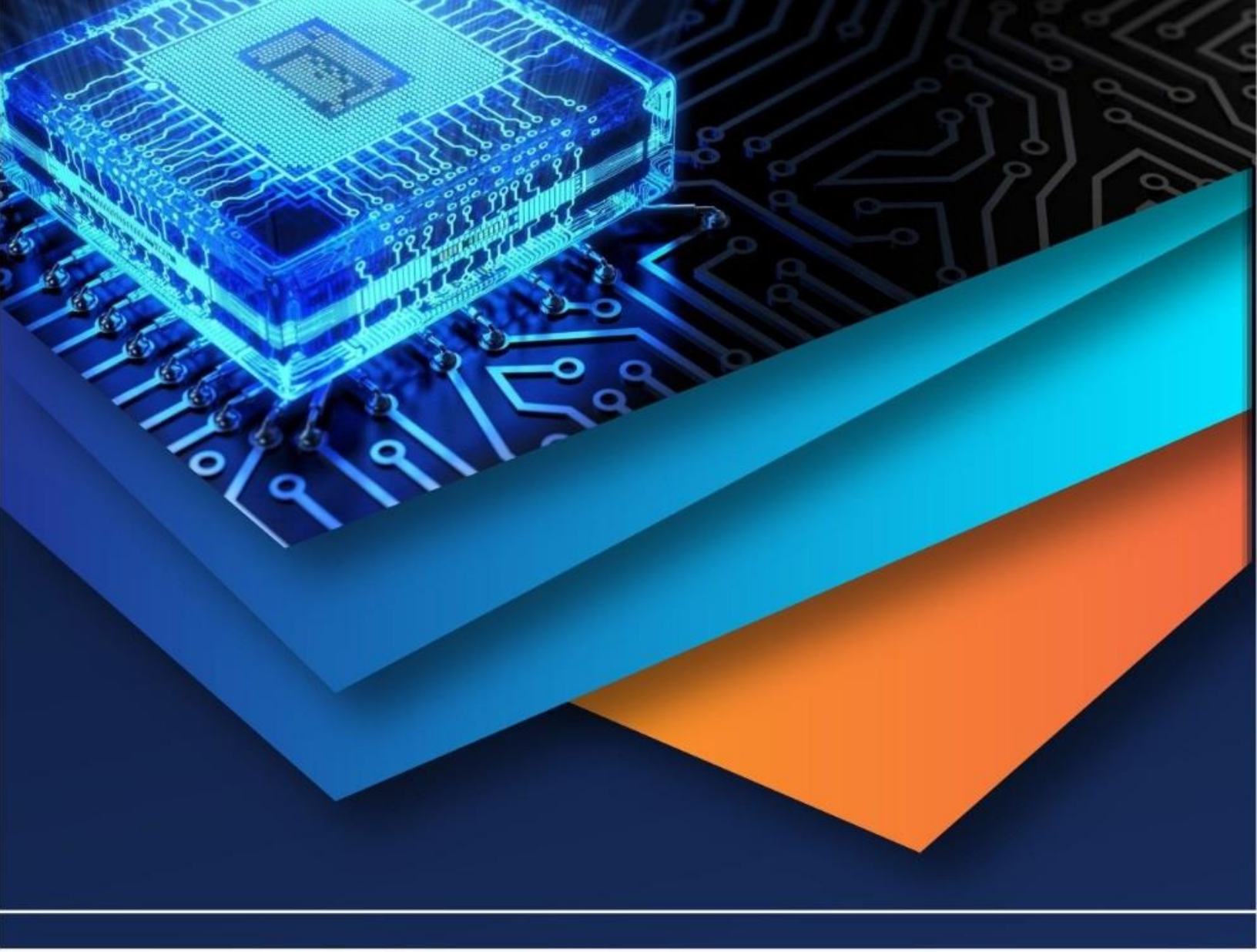
Fig 4 Network Attack 2

V. CONCLUSION

As part of the proposed NIDS, a dataset is taken, preprocessed, and analysed. ML models are built using different algorithms based on the training data. Classifier algorithms such as Decision Trees, Logistic Regression are used and ensemble techniques such as AdaBoost Voting Classifier are employed. The model is designed to classify whether there is an attack in the network. Additionally, it specifies the type of attack among Probe, DoS, R2L, and U2R attacks. In order to achieve optimal accuracy, learning models were trained and parameter-tuned according to network traffic details and configuration parameters. Some models have achieved a higher level of accuracy than others. The model is limited to intrusion detection. Further, the model can be developed and employed for other websites where networking is crucial. It can be made to notify users directly while communication is going on. In that case, not only detection, but prevention can be made so that the data does not lose its integrity and confidentiality, availability.

REFERENCES

[1] Network Intrusion Detection Using Improved Genetic k-means Algorithm. S. McElwee, "Active learning intrusion detection using k-means clustering selection", Conf. Proc. - IEEE SOUTHEASTCON, 2017
[2] Intrusion Detection Using Tree-Based Classifiers. Ahmim, M. Derdour, and M. A. Ferrag. An intrusion detection system based on combining probability predictions of a tree of classifiers, International Journal of Communication System, vol. 31, pp.1–14, 2018.
[3] A Survey of Intrusion Detection Models based on NSL-KDD Data Set. M. R. Parsaei, S. M. Rostami, and R. Javidan, “A Hybrid Data Mining Approach for Intrusion Detection on Imbalanced NSL-KDD Dataset,” Int. J. Adv. Comput. Sci. Appl., vol. 7, no. 6, pp. 20–25, 2016.
[4] Intrusion Detection Using Unsupervised Approach. Mirsky Y, Doitshman T, Elovici Y, Shabtai A (2018) Kitsune: an ensemble of autoencoders for online network intrusion detection. arXiv Preprint.
[5] Arindam Sarkar, Hanjabam Saratchandra Sharma & Moirangthem Marjit Singh. A supervised machine learning-based solution for efficient network intrusion detection using ensemble learning based on hyperparameter optimization International Journal of Information Technology volume 15, pages423–434 (2023)
[6] Abdulsalam O. Alzahrani and Mohammed J. F. Alenazi. Designing a Network Intrusion Detection System Based on Machine Learning for Software Defined Networks. Future Internet 2021, 13,111.
[7] Zeeshan Ahmad, Adnan Shahid Khan, CheahWai Shiang, Johari Abdullah, Farhan Ahmad. Network intrusion detection system: A systematic study of machine learning and deep learning approaches.
[8] J. Olamantanmi Mebawondu, OlufunsoD. Alowolodu , JacobO. Mebawondu , Adebayo O. Adetunmbi. Practical real-time intrusion detection using machine learning approaches.
[9] Tavallaee M, Bagheri E, Lu W, Ghorbani AA. Deep Neural network and Real-Time Intrusion detection system.
[10] Abdullah B, Abd-Alghafar I, Salama GI. The Machine Learning and Deep learning methods for intrusion detection system.
[11] Rong Wang, Yuansheng Dong, Juan He, P.R China. The Real-Time network intrusion detection using deferred decision and hybrid classifier.
[12] Valerio Morfino and Salvotore Ranpone, department of law, Economics, University of Sannio, I-82100 Benevento, Italy.
[13] Wang Peng, Xiangwei Kong, Guojin Peng, Xiaoya Li, Zhongjie Wang. Network Intrusion Detection Based on Deep Learning. 2019 International Conference on Communications, Information System and Computer Engineering (CISCE).
[14] Marzia Zaman,Chung-Horng. Evaluation of machine learning techniques for network Intrusion detection.
[15] Wathig Laftah AL-Yasena, Zulaiha Ali Othmana Mohd Zakree Ahmad. Multi-level Hybrid support vector machine and extreme learning machine based on modified k-means for intrusion detection system.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)