IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

INTRUSION DETECTION USING MACHINE LEARNING TECHNIQUE

[1]Dr Shilpa V, [2]Srinidhi G, [3]Hannah Thomas, [4]Vishnu Singh, [5]Srinivas D

[1] Assistant Professor, [2,3,4,5] UG Student, [1,2,3,4,5] Computer Science and Engineering,

[1,2,3,4,5]Cambridge Institute of Technology, Bengaluru, India

Abstract: The internet connects the world, but also exposes it to numerous network threats. With the vast amount of information exchanged globally, ensuring the integrity and confidentiality of data has become increasingly challenging. Network security is essential in preventing easy breaches and unintended interference. One approach involves employing Intrusion Detection Systems, strategically positioned to monitor traffic from source to destination apps. However, balancing thorough screening with system efficiency is a concern. Integrating machine learning algorithms enhances flexibility and reliability in detecting and distinguishing between ordinary and malicious activities. The algorithms, logistic regression, Naive Bayes, K-Nearest Neighbour, and Decision Trees, are utilized in our research to optimize intrusion detection in Network Traffic Data, employing various evaluation methodologies to achieve the highest accuracy.

***Index Terms* – Cyber-attack, distance relay, graph theory, multi-agent system, distributed system, deep neural network.**

I. INTRODUCTION

The Intrusion Detection System project addresses the escalating complexity of cyber threats in today's digital landscape. Traditional security measures struggle in order to keep up with dynamic threats, leaving organizations vulnerable to attacks that compromise sensitive information and disrupt operations. By utilising cutting-edge technology such as machine learning and anomaly detection, NIDS offers proactive monitoring and real-time response capabilities, bolstering overall security infrastructure. Its goal is to empower organizations to adapt continually to evolving cybersecurity challenges, ensuring the integrity and confidentiality of digital assets.

II. BACKGROUND STUDY

The Internet connects the world, but it also brings threats of network attacks and heightened risks to integrity and confidentiality cause of the vast amount of information available globally. Improving network security is important to prevent unauthorized access and interference. Network Intrusion Detection Systems track network traffic from source to destination, using strategic placement to monitor for anomalies. ML algorithms enhance NIDS by distinguishing between normal and malicious activities, optimizing intrusion detection. Research focuses on utilizing methods such logistic regression, Naive Bayes, K-Nearest Neighbour, and Decision Trees to enhance accuracy in detecting attacks. As information systems become vital for enterprises of all sizes, ensuring their security is paramount. Intrusion Detection System plays a significant role in safeguarding against various attacks, reflecting the increasing importance of cybersecurity research and development.

III. LITERATURE REVIEW

Recent studies have shown how crucial network intrusion detection is to solving new security issues. Novel approaches have been presented to properly categorise intrusion types and increase detection accuracy by leveraging deep learning techniques. Validated on datasets like KDD, these approaches show notable improvements over conventional machine learning techniques. Furthermore, research investigates the application of recurrent neural networks for intrusion detection, assessing their performance in various classification scenarios and taking into account variables such as learning rate and neuron count. These results highlight how important deep learning is becoming for improving network security by developing stronger intrusion detection systems.

\

RESEARCH METHODOLOGY

4.1 System Architecture

The application's overall hypermedia structure is determined by the System Architecture design. The objectives set for a WebApp, the information to be displayed, the users who will visit, and the specified navigation philosophy are all factors in architecture design. Content items' organisation for presentation and navigation is the main focus of content architecture. WebApp architecture deals with how the programme is set up to control user interaction, perform internal operations, facilitate navigation, and display content.

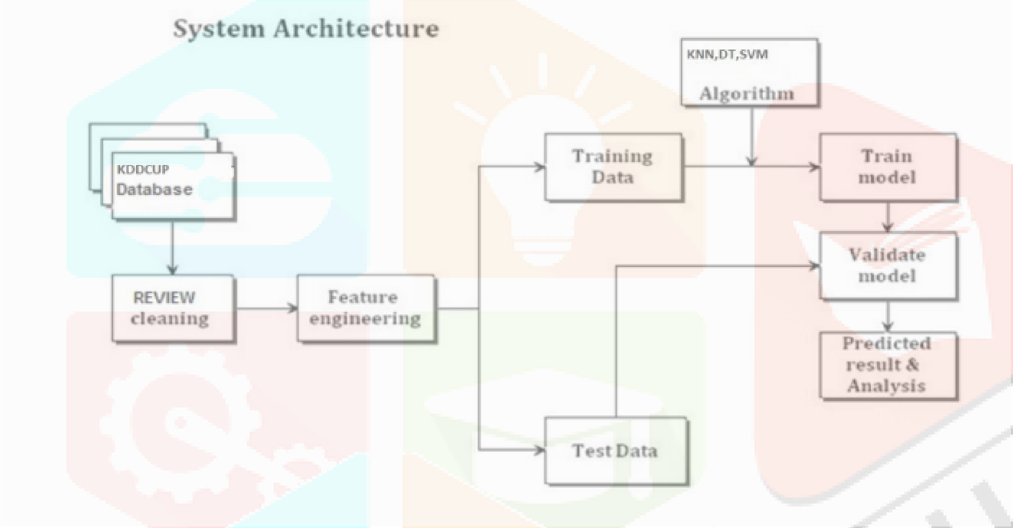


Figure 4.1: System Architecture Workflow

4.2 Flow of the Algorithm

The diagram below provides a comprehensive overview of the Rubik's Cube solving algorithm, incorporating various methodologies. It serves as a detailed guide, illustrating the sequential progression of the process from initiation to the eventual generation of solutions.

1. Dataset: This step involves acquiring the dataset that contains the relevant data needed to train the machine ML model. The dataset could include various attributes or features along with the target variable (the variable to be predicted).

2. Preprocessing: Once the dataset is obtained, it undergoes preprocessingThis involves activities such as data cleaning (addressing outliers and missing numbers),transforming the data into a suitable format (e.g., numerical encoding of categorical variables), and scaling the data if necessary.

3. Feature Analysis: In this step, the dataset's features are analyzed to gain insights into their characteristics. This analysis helps in understanding the distribution of features, identifying patterns, and assessing correlations between features and the target variable.

4. Feature Selection: Relevant features are chosen for the machine learning model's training based on the feature analysis. Reducing dimensionality and concentrating on the most useful qualities that make a substantial contribution to the target variable's prediction are the goals of feature selection.

5. Data Preparation: The data is ready for model training using the features that have been chosen. To do this, the dataset is usually divided into training and testing sets. The testing set is used to assess the model's performance after it has been trained using the training set..

6. Model Training: In this step, the model is trained using the prepared training dataset. The model learns to identify patterns and relationships within the data, optimizing its parameters to minimize errors and improve predictive accuracy.

7. Result: After the model is trained, it is used to generate predictions or outcomes for new or unseen data. The trained model's performance can also be evaluated using the test dataset to assess its accuracy and generalization ability.

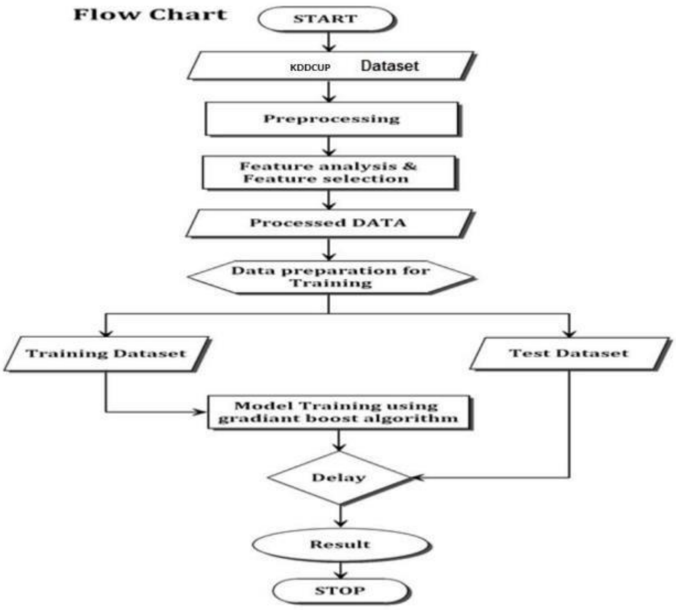

Flow Chart
START
KDDCUP Dataset
Preprocessing
Feature analysis & Feature selection
Processed DATA
Data preparation for Training
Training Dataset
Test Dataset
Model Training using gradiant boost algorithm
Delay
Result
STOP

Figure 4.2: Algorithm Flowchart

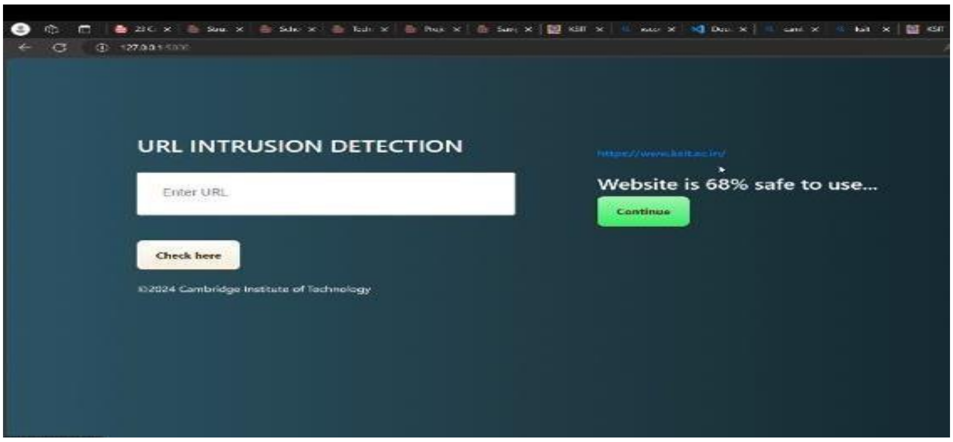
4.3 Proposed Algorithm

- Step-1: Select the number K of the neighbors
- Step-2: Calculate the Euclidean distance of K number of neighbors
- Step-3: Take the K nearest neighbors as per the calculated Euclidean distance.
- Step-4: Determine how many data items are in each category among these k neighbors.
- Step-5: Put the new data points in the category where the neighbor count is at its highest.
- Step-6: Our model is ready.

Results and Output

5.1 Software Output

In Figure 5.1.1, we present the initial view of the network intrusion detection system, depicting a website labeled as safe within the system's interface. Figure 5.1.2 showcases another view within the NIDS interface, illustrating a website flagged as unsafe. These screenshots demonstrate the ability to monitor and assess website security, empowering users to identify and respond to security risks effectively. The interactive interface and real-time threat detection features contribute to a comprehensive network security solution tailored for security enthusiasts and IT professionals.



URL INTRUSION DETECTION

Enter URL

Check here

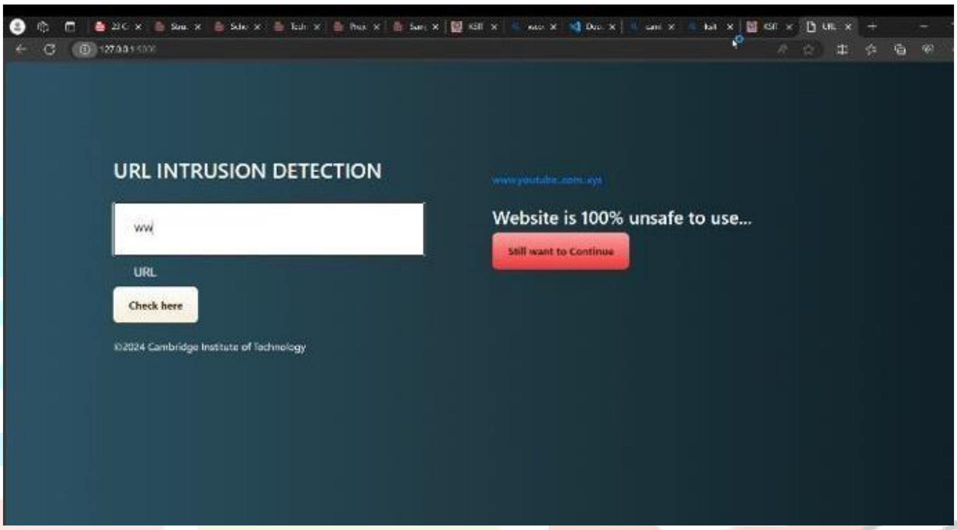
©2024 Cambridge Institute of Technology

https://www.ksit.ac.in/

Website is 68% safe to use...

Continue

Figure 5.1.1 Output detecting safe website



URL INTRUSION DETECTION

ww

URL

Check here

©2024 Cambridge Institute of Technology

www.youtube..com.xyz

Website is 100% unsafe to use...

Still want to Continue

Figure 5.1.2 Output detecting unsafe website

5.2 Data Statistics

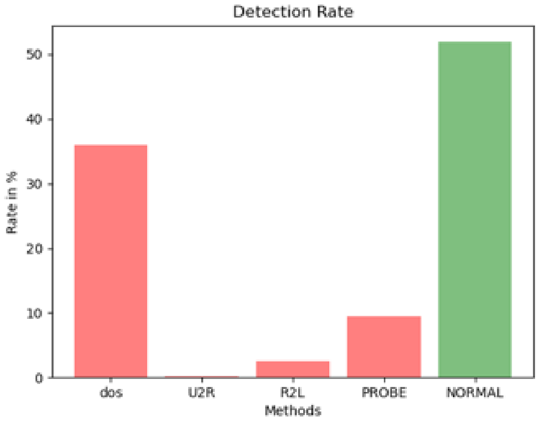


Figure 5.1: Detection Rate

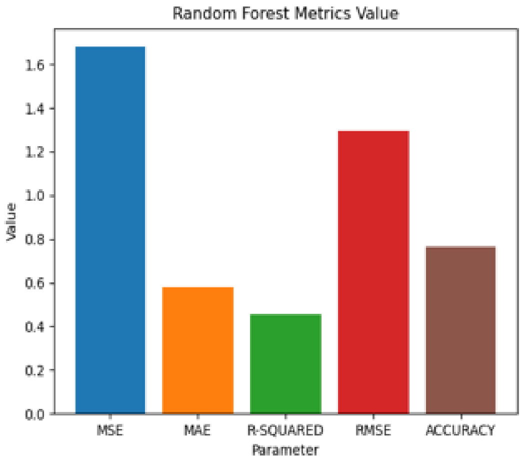


Figure 5.2: Random Forest Metrics Value

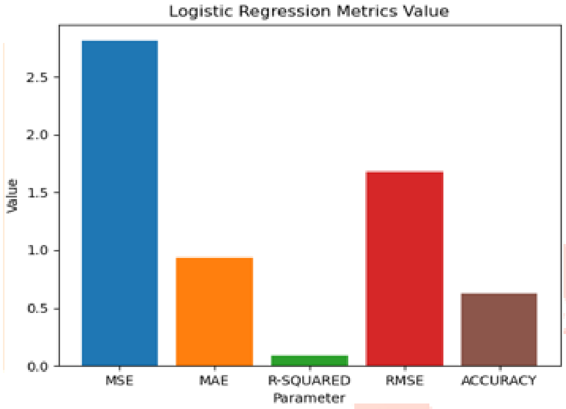


Figure 5.3: Logistic Regression Metrics Value

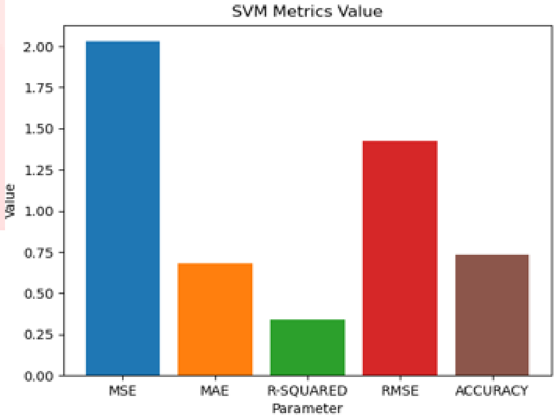


Figure 5.4: Support Vector Machine Metrics Value

CONCLUSION

In order to identify the optimal model, we have offered a variety of machine learning models that make use of various machine learning algorithms and feature selection techniques. With a 94.02% detection rate, the model constructed with ANN and wrapper feature selection beat all other models in accurately identifying network traffic, according to the examination of the data. We think that these results will further study in the area of developing a detection system that can identify both known and unknown assaults.

REFERENCES

- •P. Alaei and F. Noorbehbahani , ‘’Incremental anomaly –based intrusion detection system using limited labeled data,’’ in Web Research (ICWR),2017 3th International Conference on, 2017,pp. 178-184
- •M. Saber, S. Chadli, M. Emharraf, and I.EI Farissi,’’Modeling and implementation approach to evaluate the intrusion detection system,’’ in International Conference on Networked systems, 2015,pp.513-517.
- •M. Tavallaee, N. Stakhanova, and A. A. Ghorbani ,’’Toward credible evaluation of anomaly-based intrusion-detection methods,’’IEEE Transactions on Systems, Man, and Cybernetics,Part C (Applications and Reviews),vol. 40, no. 5,pp. 516- 524,2010.
- •A. S. Ashoor and S. Gore,’’Importance of intrusion detection