INTRUSION DETECTION SYSTEM

Prof. Sheetal S R[1], Shradha[2], Shravani S Raju[3], Vaishnavi B A[4],
Pamarthi Ravi Kiran[5]

[1] *Assistant Professor, CSE, AMCEC, Karnataka, India*

[2] *Student, CSE, AMCEC, Karnataka, India*

[3] *Student, CSE, AMCEC, Karnataka, India*

[4] *Student, CSE, AMCEC, Karnataka, India*

[5] *Student, CSE, AMCEC, Karnataka, India*

ABSTRACT

Our project aims to identify various attacks that are taking place on user's device using Machine Learning. After identifying the type of attacks, various suggestions are provided to user to secure his device against attacks. At the core our project is a Machine Learning project that uses various pre available datasets that are available on the internet and also our custom datasets. After that for suggestion various methods are fetched from internet via an API. Intrusion detection system refers to monitoring of network traffic of a device and alerts the concerned authority when such activity is discovered. It uses input from multiple sources and the uses specific filtering algorithms to differentiate normal activities from malicious ones. An intrusion detection system is used to detect and monitor network traffic and computer data that a typical firewall cannot detect. It is also used to monitor attacks and hacks against susceptible services, data-driven assaults on computers, protocols, application etc. and host-based and network-based attacks including stealing your information, illegal logins, trying to access sensitive file. It can also be placed in front of a server which can be used to monitor attacks like SQL injection etc. An IDS will typically consist of an agent that will collect all the necessary information from the packets and data that are incoming and outgoing, and then analysis engine will detect the signs of intrusion on the system, tree based ids using machine learning.

Keywords: - *intrusion detection, decision tree, web scraping, tokenization, classification, regression*

1. INTRODUCTION

Intrusion detection system refers to monitoring of network traffic of a device and alerts the concerned authority when such activity is discovered. It uses input from multiple sources and the uses specific filtering algorithms to differentiate normal activities from malicious ones. An intrusion detection system is used to detect and monitor network traffic and computer data that a typical firewall cannot detect. It is also used to monitor attacks and hacks against susceptible services, data-driven assaults on computers, protocols, application etc. and host-based and network-based attacks including stealing your information, illegal logins, trying to access sensitive file. It can also be placed in front of a server which can be used to monitor attacks like SQL injection etc.

2. PROBLEM STATEMENT

Design and develop a technological solution for identifying and rectifying obscene intrusion (image/ video) at the user end.

3. BACKGROUND WORK

Hoque, M. S., Mukit, M. A., & Bikas, M. A. N. aimed to discuss and implement the parameters and evolution of a genetic algorithm (GA) into an intrusion detection system (IDS) to filter traffic data and reduce complexity, using the KDD99 dataset. Their paper highlighted advantages such as improved detection rates and lower false positive rates due to the utilization of better heuristics, facilitating more statistical analysis and complex equations. However, the authors identified limitations including the increased usage of system resources by the IDS even in intrusion-free scenarios, susceptibility to tampering which could render the system unreliable, and the potential for data modification by attackers during transmission, leading to misinterpretation of results. The techniques and methods employed in their approach included mutation, crossover, and fitness function evaluation within the GA framework. This paper offers valuable insights into the integration of genetic algorithms into intrusion detection systems, addressing both advantages and limitations associated with this approach.

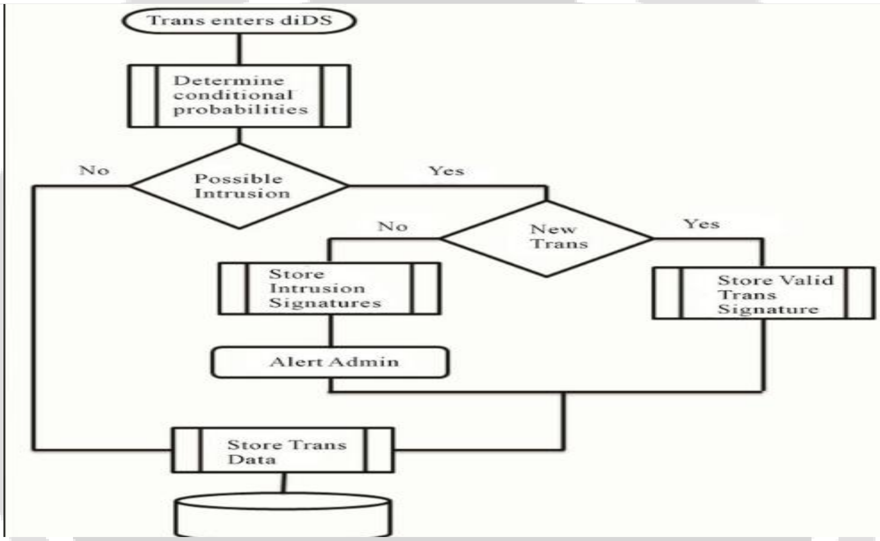

Trans enters diIDS
Determine conditional probabilities
Possible Intrusion
No
Yes
New Trans
No
Yes
Store Intrusion Signatures
Store Valid Trans Signature
Alert Admin
Store Trans Data

Fig 2.1 flow chart

4. OBJECTIVE

The objective of an Intrusion Detection System (IDS) utilizing decision trees is to proactively detect and respond to potential security threats within a computer network. By analyzing network traffic and system events in real-time, the IDS aims to identify anomalous patterns or behaviors that may indicate unauthorized access, malicious activity, or potential breaches. Through the use of decision tree algorithms, the IDS seeks to classify incoming data into normal or suspicious categories, generating alerts when suspicious activity is detected. These alerts prompt system administrators to take appropriate action, such as implementing security measures, isolating affected systems, or conducting further investigation, with the ultimate goal of mitigating the impact of security incidents and preserving the integrity of the network. Additionally, the IDS aims to continuously adapt and improve its detection capabilities through periodic retraining sessions, ensuring its effectiveness against evolving cyber threats over time. Overall, the objective of an IDS using decision trees is to enhance the overall cybersecurity posture of an organization by providing proactive threat detection and response capabilities.

5. LITERATURE SURVEY

Alqatani, H., Sarker, I. H., Kalim, A., Hossain, S. M. M., Ikhlaq, S., & Hossain, S. (2020) conducted a study with the objective of developing an accurate defense technique to identify inconsistencies and abnormal behavior on a network or system by determining the effectiveness of various models in these scenarios. The research aimed to evaluate how different models handle data and compare their classification accuracy in detecting intrusions. The study utilized Bayesian networks, decision trees, decision tables, random forests, random trees, and artificial neural networks as tools for analysis. While the research provided insights into the effectiveness of different models in detecting intrusions, it also highlighted limitations associated with cybersecurity datasets comprising various categories of cyber-attacks with relevant features. Consequently, some classifiers may exhibit lower accuracy and prediction rates due to the diversity of attack categories and features within the datasets.

Surana, J., Sharma, J., & Saraf, I. explored the landscape of Intrusion Detection System (IDS) classifications, aiming to delineate the components and techniques associated with Host-based, Network-based, and Application-based IDS. Their paper elucidated the advantages of each type: Host-based IDS verifying attack outcomes and monitoring system activities, Network-based IDS enabling real-time detection and rapid response, and Application-based IDS facilitating observation of user-application interactions. However, the authors highlighted several limitations, including the potential for false positives from IDS sensors, challenges in configuring IDS across diverse operating systems, and the incapacity to analyze encrypted packets. Notably, the paper discussed two primary detection techniques: Anomaly-based IDS and Signature-based IDS, providing insights into their respective methodologies and applications within the realm of intrusion detection.

6. METHODOLOGY

The methodology of building an Intrusion Detection System (IDS) using decision trees involves several key steps. Firstly, comprehensive data collection is conducted, acquiring diverse and representative datasets containing network traffic logs, system events, or other relevant cybersecurity data. Subsequently, data preprocessing is performed to clean, normalize, and engineer features, preparing the data for model training. Following this, an appropriate decision tree algorithm, such as CART (Classification and Regression Trees) or Random Forests, is selected based on the dataset's characteristics. The dataset is then split into training and testing sets, and the decision tree model is trained using the training data. Hyperparameter tuning and cross-validation techniques are applied to optimize the model's performance and prevent overfitting. The trained model is evaluated using metrics such as accuracy, precision, recall, and F1-score to assess its effectiveness in detecting intrusions. Finally, the validated model is deployed in a production environment for real-time monitoring and analysis of network traffic and system events, integrating it with existing security infrastructure and configuring alerting mechanisms for timely threat detection and response.

7.ARCHITECTURE

The architecture diagram of an Intrusion Detection System (IDS) depicts the system's overall structure and components involved in detecting and responding to cyber threats. At its core, the IDS comprises three main layers: the Data Collection Layer, the Detection and Analysis Layer, and the Response and Reporting Layer. The Data Collection Layer encompasses various sources of data, including network traffic logs, system events, and security logs, which are collected from sensors, agents, or log files distributed across the network. These raw data streams are then fed into the Detection and Analysis Layer, where sophisticated detection algorithms, such as decision trees, machine learning models, or signature-based detection engines, analyze the data to identify anomalous behavior and potential security incidents. The detection results are further processed and correlated to assess the severity and impact of detected threats. Finally, the Response and Reporting Layer takes action based on the detected threats, which may include generating alerts, triggering automated responses, or initiating incident response workflows. Additionally, the layer generates comprehensive reports and visualizations to provide insights into the security posture of the system and facilitate decision-making by security analysts and stakeholders. Overall, the IDS architecture enables proactive threat detection, rapid incident response, and continuous improvement of security defenses to mitigate cyber risks and safeguard critical assets.

In the IDS architecture, each layer interacts with the others to form a cohesive and robust defense mechanism against cyber threats. The Data Collection Layer gathers data from various sources, including network devices, servers, and endpoints, ensuring comprehensive coverage of the network environment. This data is then processed and analyzed in the Detection and Analysis Layer, where advanced algorithms and techniques, such as decision trees, anomaly detection, and behavioral analysis, identify suspicious patterns and deviations from normal behavior. The Response and Reporting Layer plays a pivotal role in translating detection findings into actionable insights and responses, facilitating timely incident response and remediation efforts. Moreover, the IDS architecture incorporates feedback loops and continuous monitoring to adapt to evolving threats and improve detection capabilities over time. By integrating these layers into a unified architecture, the IDS enhances organizational resilience and empowers security teams to effectively detect, mitigate, and prevent cyber threats, thereby safeguarding critical assets and ensuring the integrity and availability of network resources**.**

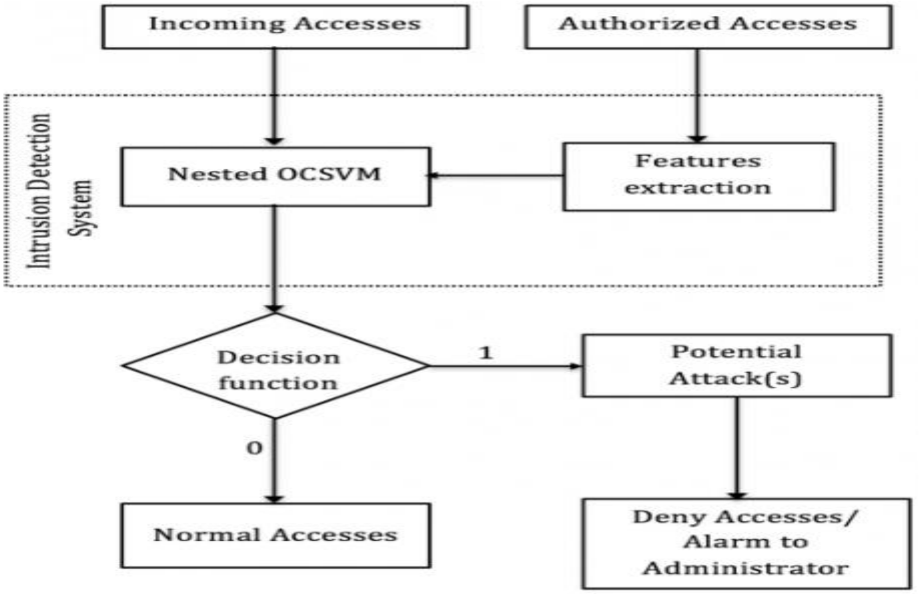
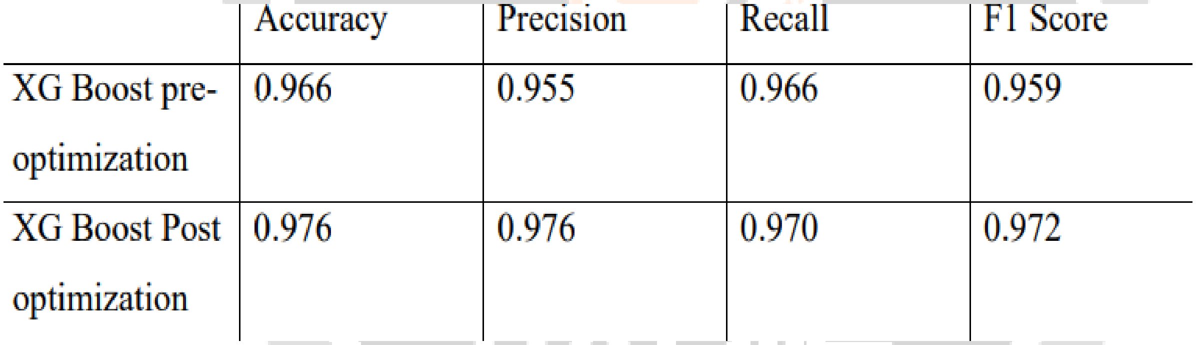
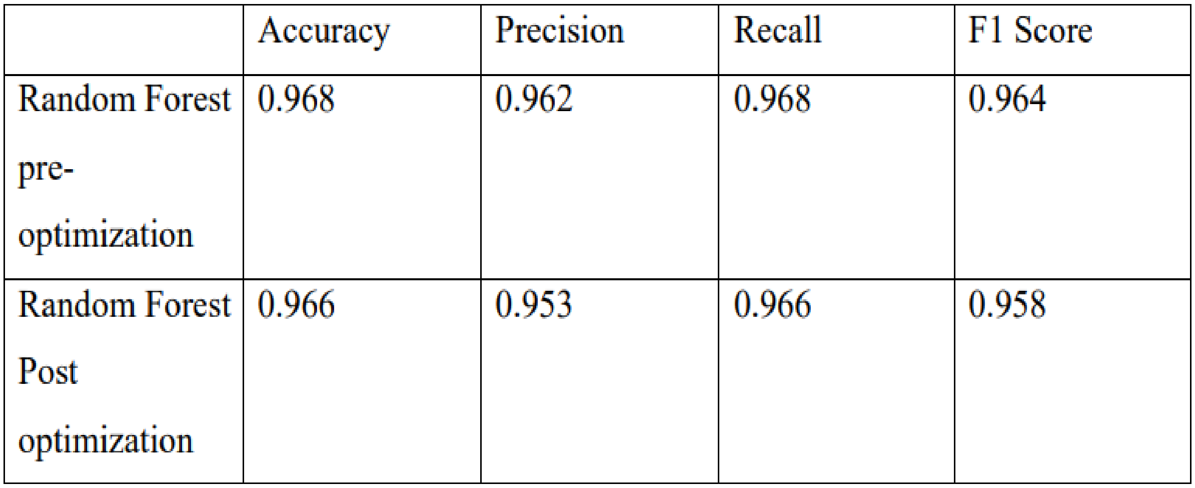


Figure 1. System Architecture Diagram

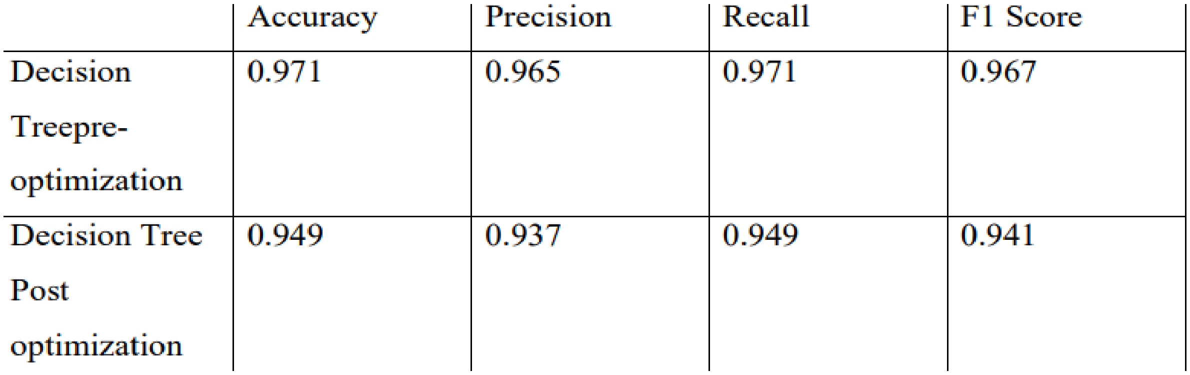
7. RESULTS



	Accuracy	Precision	Recall	F1 Score
XG Boost pre-optimization	0.966	0.955	0.966	0.959
XG Boost Post optimization	0.976	0.976	0.970	0.972



	Accuracy	Precision	Recall	F1 Score
Random Forest pre-optimization	0.968	0.962	0.968	0.964
Random Forest Post optimization	0.966	0.953	0.966	0.958



	Accuracy	Precision	Recall	F1 Score
Decision Treepre-optimization	0.971	0.965	0.971	0.967
Decision Tree Post optimization	0.949	0.937	0.949	0.941

8. CONCLUSION

We started doing literature survey. Each of us went through a couple of papers and perform extensive literature survey on 1- 2 papers. And we add a couple more of literature survey papers. We used different techniques, approaches, procedures and algorithms which were previously used by others to solve the problem. By doing so we were able to build our project. Taking into account the problems faced by them during their modelling, we could easily overcome our problems which we faced by studying various models and analysing them, we decided on the models that would be appropriate for our project. Taking into the account the various advantages and limitations of the model it really helped us understand and work on making the accuracy of model even better. After selecting 4 datasets, we worked on them using different ML models and optimizations i.e., XG Boost, Random Forest Classifier, Decision Tree, Extra Tree Classifier and Ensemble stacking model, applying Bayesian optimization at each stage. After finding our results, we were able to find great accuracy in certain models and others not so much.

REFERENCES

[1] Hamed Alqahtani, Iqbal H. Sarker, Asra Kalim and Syed Mohammod Minhaz Hossain, “CyberIntrusion Detection Using Machine Learning Classification Techniques”.

[2] Mohit Tiwary and Raj Kumar, “INTRUSION DETECTION SYSTEM”.

[3] Iqbal H. Sarker, Yoosef Abushark and Fawaz Alsomi, “IntruDTree: A Machine LearningBasedCyber Security Intrusion Detection Model”.

[4] Jayesh Surana, Jagrati Sharma and Ishika Saraf, “A Survey on Intrusion Detection System”.

[5] Lirim Ashiku and Cihan Dagli, “Network Intrusion Detection System using Deep Learning”.

[6] Mohamed Amine Ferrag, Leandros Maglaras, Sotiris Moschoyiannis, and Helge Janicke, “DeepLearning for Cyber Security Intrusion Detection: Approaches, Datasets, and Comparative Study”

[7] Mohammad Sazzadul Hoque, Md. Abdul Mukit, and Md. Abu Naser Bikas, “An Implementationof Detection System using Genetic Alogorithm”.

[8] Stefan Axelsson, “Intrusion Detection Systems: A Survey and Taxonomy”.

[9] Marcin Niemiec, Rafał Ko´sciej, and Bartłomiej Gdowski, “Multivariable Heuristic Approach toIntrusion Detection in Network Environments”