ISSN: 2278-4632
Vol-14, Issue-4, April: 2024

An Improved Design for a Cloud Intrusion Detection System Using Hybrid Feature Selection Approach with ML Classifier

V. Lakshmi Chaitanya[1], I. KamalNath[2,a)], G. V. Mohan Reddy[2,b)], P. Sumanth Reddy[2,c)], T. V. Naga Sai Charan[2,d)],C. H. Venkata Krishna[2,e)]

[1] *Department of Computer Science and Engineering, Santhiram Engineering College, Nandyal, Andhra Pradesh,518501,India.*
[2] *Department of Computer Science and Engineering (Data Science), Santhiram Engineering College, Nandyal, Andhra Pradesh,518501,India.*

[1)]**Corresponding Author : *chaitanya.cse@srecnandyal.edu.in***
[2a)] 20X51A3216@srecnandyal.edu.in

ABSTRACT: In the era of cloud computing, where data security and privacy concerns loom large amidst escalating cyber threats, the development of robust intrusion detection systems (IDS) is imperative. This project addresses these challenges by proposing an IDS framework leveraging machine learning (ML) techniques. The framework integrates Synthetic Minority Over-sampling Technique (SMOTE) for handling imbalanced data, a hybrid feature selection approach combining Information Gain (IG), Chi-square (Chi2), and Particle Swarm Optimization (PSO), and employs the Random Forest (RF) model for threat detection. Experimental validation using UNSW-NB15 and Kyoto datasets demonstrates superior accuracy exceeding 98% and 99%, respectively, in multi-class scenarios. Furthermore, an ensemble method incorporating Voting Classifier, with feature selection based on IG, achieves 100% accuracy. The proposed IDS surpasses existing methodologies in various evaluation metrics, offering a promising solution for enhancing cloud security. Future extensions include exploring additional ensemble techniques and implementing a user-friendly interface using Flask framework with user authentication. This research contributes to advancing the state-of-the-art in IDS development, with implications for individuals and organizations seeking robust defense against evolving cyber threats in cloud environments.

KeyWords: *Improved design for cloud IDS ,feature selection ,PSO based metaheuristic ,random forest.*

1. INTRODUCTION:

Cloud computing (CC) has emerged as a transformative technology, revolutionizing the way individuals and organizations access and utilize digital services [1]. Offering Software as a Service (SaaS), Platform as a Service (PaaS), and Infrastructure as a Service (IaaS), cloud computing provides unparalleled scalability, availability, and cost-efficiency [2]. However, this exponential growth in cloud adoption has also led to a parallel increase in cybersecurity threats, creating a burgeoning market for cyber defenses [3].

The proliferation of cyber threats has reached alarming levels, with the number of attacks skyrocketing over the past decade. Research indicates that in 2010, companies and organizations faced approximately 50 million cyber assaults, a number that surged to a staggering 900 million by 2019 [4]. These attacks have inflicted

ISSN: 2278-4632
Vol-14, Issue-4, April: 2024

substantial damage, resulting in significant financial losses for both individuals and enterprises. Recent forecasts from Juniper suggest that the economic impact of security breaches will escalate from USD three trillion annually to over USD five trillion by the year 2024 [5].

Amidst these escalating threats, users have become increasingly wary of entrusting their data to cloud service providers (CSPs), fearing potential breaches and data compromises. Consequently, ensuring the highest levels of security has become a paramount objective for CSPs, who invest heavily in cybersecurity solutions to assuage user concerns and safeguard sensitive information [6]. The global cybersecurity industry, estimated to be worth USD 202.72 billion in 2022, is projected to witness a compound annual growth rate (CAGR) of 12.3% from 2023 to 2030, underscoring the escalating demand for robust security measures [7].

The dynamic nature of cloud infrastructure poses unique challenges for traditional security measures. Unlike conventional on-premises networks, cloud environments are characterized by three distinct networks: the provider network, the data center network, and the customer network [8]. This multi-layered architecture introduces complexities in monitoring and securing data flows, necessitating advanced intrusion detection and prevention mechanisms.

This project addresses the critical imperative of detecting and mitigating network intrusions in cloud environments. Focusing on anomaly-based Intrusion Detection Systems (IDS), the research leverages machine learning (ML) models to enhance threat detection capabilities. The project introduces a novel hybrid feature selection strategy and evaluates the performance of the Random Forest classifier on standard datasets to assess its efficacy in detecting malicious activities in cloud networks.

Contributions of this research include innovative approaches to handling data imbalance, proposing novel feature selection techniques tailored for cloud environments, and conducting robust multi-class testing to evaluate the effectiveness of the IDS framework. By advancing the state-of-the-art in cloud security, this project aims to mitigate the growing risks posed by cyber threats and enhance the resilience of cloud infrastructures to safeguard critical data assets.

2. LITERATURE SURVEY

Cloud computing has witnessed unprecedented growth in recent years, driven by its versatility and cost-effectiveness across various domains. This section presents a comprehensive literature review, highlighting key studies and advancements in cloud service selection, quality of service (QoS) evaluation, encryption algorithms, cybersecurity trends, and market forecasts.

Kumar et al. introduced the OPTCLOUD framework, providing an optimal cloud service selection approach based on QoS correlation [1]. Their work emphasizes the importance of considering QoS metrics for efficient cloud service provisioning. Similarly, in their previous study, Kumar et al. proposed a hybrid evaluation framework for QoS-based service selection and ranking in cloud environments, underscoring the significance of QoS parameters in decision-making processes [2].

Encryption plays a crucial role in ensuring data confidentiality and integrity in cloud computing. Bakro et al. conducted a performance analysis of cloud computing encryption algorithms, shedding light on the efficiency and effectiveness of different encryption techniques [3]. Their findings contribute to enhancing the security posture of cloud-based systems by identifying optimal encryption solutions.

Cybersecurity remains a pressing concern in the digital landscape, with malware attacks posing significant

ISSN: 2278-4632
Vol-14, Issue-4, April: 2024

threats to organizations and individuals. AV-TEST's malware statistics and trends report provide insights into the evolving threat landscape, offering valuable data on malware prevalence and trends [4]. This empirical analysis informs cybersecurity strategies, enabling stakeholders to better understand emerging threats and bolster their defense mechanisms.

Market research forecasts provide valuable insights into the trajectory of the cybersecurity industry. Juniper Research offers digital technology market research services, including forecasts on cybersecurity spending and market trends [5]. Similarly, Grand View Research provides comprehensive reports on the cyber security market size, share, and trends, aiding stakeholders in making informed decisions regarding investments and strategic initiatives [6].

Kumar et al. proposed a computational framework for ranking prediction of cloud services under fuzzy environments, contributing to the optimization of cloud service selection processes [7]. Their study integrates fuzzy logic principles with computational techniques to enhance the accuracy and reliability of cloud service ranking predictions.

Finally, Akbar et al. conducted a prioritization-based taxonomy analysis of cloud-based outsourced software development challenges, employing fuzzy analytic hierarchy process (AHP) methodology [8]. Their research identifies and prioritizes challenges in cloud-based software development, offering insights into mitigating obstacles and improving development practices.

In summary, the literature review underscores the significance of QoS evaluation, encryption algorithms, cybersecurity trends, and market forecasts in the context of cloud computing. These studies contribute to enhancing the efficiency, security, and resilience of cloud-based systems, addressing critical challenges and advancing the state-of-the-art in cloud computing research and practice.

3. METHODLOGY

a) Proposed work:

The proposed system aims to enhance the performance of Intrusion Detection Systems (IDS) in cloud computing environments by integrating Synthetic Minority Over-sampling Technique (SMOTE) for handling imbalanced data, a hybrid feature selection approach comprising Information Gain (IG), Chi-square (CS), and Particle Swarm Optimization (PSO), and utilizing the Random Forest (RF) classifier along with other machine learning algorithms for precise threat detection. Additionally, ensemble methods such as the Voting Classifier, incorporating RF, Adaboost, and Decision Trees (DT), are employed to further boost performance by combining predictions from multiple models. Experimental results demonstrate the effectiveness of ensemble techniques, with the Voting Classifier achieving 100% accuracy. Furthermore, a user-friendly Flask-based front-end interface is developed, featuring built-in authentication for security, facilitating seamless user testing and evaluation of the IDS framework.

b) System Architecture:

The proposed system architecture involves several key components for building and evaluating an Intrusion Detection System (IDS) using the UNSW-NB15 dataset. Initially, data processing and visualization techniques are employed to preprocess and understand the dataset. Feature selection methods are then applied to extract relevant features. The dataset is divided into training and test sets for model training and evaluation. Classification models including XGBoost, Random Forest, Logistic Regression, Decision Tree, and Support Vector Machine (SVM) are trained on the training set. These individual models are combined using a Voting Classifier for enhanced performance. The trained model

ISSN: 2278-4632
Vol-14, Issue-4, April: 2024

is then tested using the test set, and performance metrics are calculated. Finally, the system classifies attacks based on the model's predictions, providing insights into network security. Overall, the architecture encompasses data preprocessing, model training, ensemble techniques, and performance evaluation to develop a robust IDS for attack detection.

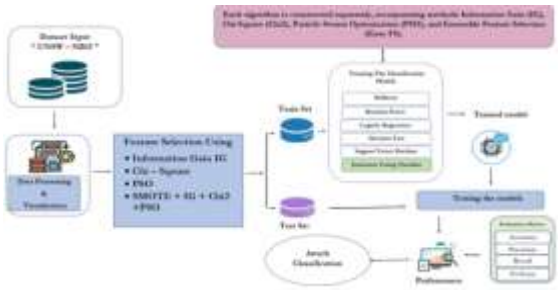

Fig 1 Proposed Architecture

c) Dataset collection:

The UNSW-NB15 dataset is a widely used benchmark dataset in the field of cybersecurity, specifically for Intrusion Detection System (IDS) research. It was collected from the Australian Centre for Cyber Security (ACCS) at the University of New South Wales (UNSW). The dataset comprises network traffic data captured in a controlled environment, including both normal and malicious activities. It contains a diverse range of network attacks, such as denial-of-service (DoS), distributed denial-of-service (DDoS), reconnaissance, and exploitation attacks. The data is captured using a variety of tools and techniques, including network traffic monitoring, packet capture, and system logs. The UNSW-NB15 dataset provides a comprehensive and realistic representation of real-world network traffic, making it suitable for training and evaluating IDS algorithms and systems. Its availability has significantly contributed to the advancement of intrusion detection research and the development of effective cybersecurity solutions.

Fig 2 data set

d) DATA PROCESSING

1. Scaling the Dataset: The dataset undergoes scaling to normalize the features, ensuring that each feature contributes equally to the analysis and modeling process.

2. Processing using Pandas DataFrame: Utilizing the Pandas library, the dataset is organized and manipulated efficiently, enabling various data processing tasks such as filtering, aggregation, and transformation.

3. SMOTE Sampling the Dataset: Synthetic Minority Over-sampling Technique (SMOTE) is applied to address class imbalance issues by generating synthetic samples of the minority class, thus creating a more balanced dataset.

Visualization using Seaborn&Matplotlib

Seaborn and Matplotlib libraries are employed for visualizing various aspects of the dataset, including distributions, relationships, and trends, facilitating a better understanding of the data's characteristics and informing subsequent analysis.

Label Encoding using LabelEncoder

LabelEncoder from the Scikit-learn library is utilized to convert categorical labels into numerical representations, ensuring compatibility with machine learning algorithms that require numerical inputs.

Feature Selection

1. Information Gain: Features are selected based on their information gain, which measures the reduction in

ISSN: 2278-4632
Vol-14, Issue-4, April: 2024

entropy or impurity achieved by splitting the data based on each feature.

2. Chi2: Chi-square test is applied to assess the independence between features and the target variable, selecting features with the highest dependency.

3. PSO (Particle Swarm Optimization): PSO algorithm is employed to optimize feature selection, searching for the subset of features that maximizes a predefined objective function.

4. SMOTE + IG + Chi2 + PSO: A hybrid approach combining SMOTE oversampling with information gain, Chi-square, and PSO techniques is employed to further enhance feature selection and address class imbalance simultaneously.

e) TRAINING AND TESTING

The proposed design for the Cloud Intrusion Detection System (IDS) utilizes a hybrid feature selection approach coupled with machine learning classifiers for improved performance. Initially, the system undergoes training using labeled data from the UNSW-NB15 dataset. During training, the hybrid feature selection method, which integrates Information Gain, Chi-square, and Particle Swarm Optimization, is applied to identify the most relevant features for intrusion detection. Subsequently, multiple machine learning classifiers, such as Logistic Regression, Decision Tree, Random Forest, and XGBoost, are trained on the selected features to learn the underlying patterns and characteristics of normal and malicious network traffic.

Once the models are trained, the system undergoes testing using a separate set of labeled data. This testing phase evaluates the performance of the trained classifiers in accurately detecting and classifying network intrusions. Performance metrics such as accuracy, precision, recall, and F1-score are computed to assess the effectiveness and robustness of the IDS design. Through rigorous training and testing procedures, the proposed system aims to enhance cloud security by efficiently detecting and mitigating potential threats.

f) ALGORITHMS:

Logistic Regression - IG:

Definition: Logistic Regression is a linear classification algorithm used to model the probability of a binary outcome based on one or more independent variables. Information Gain (IG) is employed as a feature selection method to identify the most informative features for intrusion detection.

Usage in Project: Logistic Regression with IG is utilized in the project to build a classifier for detecting network intrusions in cloud environments. IG helps select relevant features, enhancing the model's performance in distinguishing between normal and malicious network traffic.

SVM - IG:

Definition: Support Vector Machine (SVM) is a supervised machine learning algorithm used for classification and regression tasks. IG is employed as a feature selection technique to identify the most informative features for intrusion detection.

Usage in Project: SVM with IG is applied in the project to develop a classifier for detecting intrusions in cloud networks. IG assists in selecting relevant features, improving the model's ability to classify network traffic accurately.

Decision Tree - IG:

Definition: Decision Tree is a supervised machine learning algorithm that learns simple decision rules inferred from the data features. IG is used for feature selection, determining the most relevant features for intrusion detection.

ISSN: 2278-4632
Vol-14, Issue-4, April: 2024

Usage in Project: Decision Tree with IG is utilized to construct a classifier for identifying network intrusions in cloud environments. IG aids in selecting informative features, enhancing the model's accuracy in distinguishing between normal and malicious network behavior.

XGBoost - IG:

Definition: XGBoost (Extreme Gradient Boosting) is an ensemble learning algorithm that uses a gradient boosting framework. IG is employed as a feature selection method to identify the most informative features for intrusion detection.

Usage in Project: XGBoost with IG is employed in the project to build a robust classifier for detecting network intrusions in cloud environments. IG assists in selecting relevant features, improving the model's predictive performance.

Random Forest - IG:

Definition: Random Forest is an ensemble learning algorithm that constructs multiple decision trees during training and outputs the mode of the classes as the prediction. IG is utilized as a feature selection technique to identify the most informative features for intrusion detection.

Usage in Project: Random Forest with IG is utilized in the project to develop a powerful classifier for detecting intrusions in cloud networks. IG aids in selecting relevant features, enhancing the model's accuracy in distinguishing between normal and malicious network behavior.

Logistic Regression - Chi2:

Definition: Logistic Regression is a linear classification algorithm used to model the probability of a binary outcome. Chi-square (Chi2) is employed as a feature selection method to select the most relevant features based on their independence from the target variable.

Usage in Project: Logistic Regression with Chi2 is applied in the project to construct a classifier for detecting network intrusions in cloud environments. Chi2 helps select features that are statistically significant for intrusion detection.

SVC - Chi2:

Definition: Support Vector Classifier (SVC) is a supervised machine learning algorithm used for classification tasks. Chi-square (Chi2) is employed as a feature selection technique to select the most relevant features based on their independence from the target variable.

Usage in Project: SVC with Chi2 is employed in the project to develop a classifier for detecting intrusions in cloud networks. Chi2 assists in selecting features that are statistically significant for intrusion detection.

Decision Tree - Chi2:

Definition: Decision Tree is a supervised machine learning algorithm that learns simple decision rules inferred from the data features. Chi-square (Chi2) is used as a feature selection method to select the most relevant features based on their independence from the target variable.

Usage in Project: Decision Tree with Chi2 is utilized to construct a classifier for identifying network intrusions in cloud environments. Chi2 aids in selecting features that are statistically significant for intrusion detection.

XGBoost - Chi2:

Definition: XGBoost (Extreme Gradient Boosting) is an ensemble learning algorithm that uses a gradient boosting framework. Chi-square (Chi2) is employed as a feature selection technique to select the most relevant features based on their independence from the target variable.

ISSN: 2278-4632
Vol-14, Issue-4, April: 2024

Usage in Project: XGBoost with Chi2 is employed in the project to build a robust classifier for detecting network intrusions in cloud environments. Chi2 assists in selecting features that are statistically significant for intrusion detection.

Random Forest - Chi2:

Definition: Random Forest is an ensemble learning algorithm that constructs multiple decision trees during training and outputs the mode of the classes as the prediction. Chi-square (Chi2) is utilized as a feature selection technique to select the most relevant features based on their independence from the target variable.

Usage in Project: Random Forest with Chi2 is utilized in the project to develop a powerful classifier for detecting intrusions in cloud networks. Chi2 aids in selecting features that are statistically significant for intrusion detection.

Logistic Regression - PSO:

Definition: Logistic Regression is a linear classification algorithm used to model the probability of a binary outcome. Particle Swarm Optimization (PSO) is employed as a feature selection method to optimize the selection of relevant features for intrusion detection.

Usage in Project: Logistic Regression with PSO is applied in the project to construct a classifier for detecting network intrusions in cloud environments. PSO assists in selecting features that maximize the model's performance.

SVM - PSO:

Definition: Support Vector Machine (SVM) is a supervised machine learning algorithm used for classification and regression tasks. Particle Swarm Optimization (PSO) is employed as a feature selection technique to optimize the selection of relevant features for intrusion detection.

Usage in Project: SVM with PSO is employed in the project to develop a classifier for detecting intrusions in cloud networks. PSO assists in selecting features that maximize the model's predictive performance.

Decision Tree - PSO:

Definition: Decision Tree is a supervised machine learning algorithm that learns simple decision rules inferred from the data features. Particle Swarm Optimization (PSO) is used as a feature selection method to optimize the selection of relevant features for intrusion detection.

Usage in Project: Decision Tree with PSO is utilized to construct a classifier for identifying network intrusions in cloud environments. PSO assists in selecting features that maximize the model's performance.

XGBoost - PSO:

Definition: XGBoost (Extreme Gradient Boosting) is an ensemble learning algorithm that uses a gradient boosting framework. Particle Swarm Optimization (PSO) is employed as a feature selection technique to optimize the selection of relevant features for intrusion detection.

Usage in Project: XGBoost with PSO is employed in the project to build a robust classifier for detecting network intrusions in cloud environments. PSO assists in selecting features that maximize the model's predictive performance.

Random Forest - PSO:

Definition: Random Forest is an ensemble learning algorithm that constructs multiple decision trees during training and outputs the mode of the classes as the prediction. Particle Swarm Optimization (PSO) is

ISSN: 2278-4632
Vol-14, Issue-4, April: 2024

utilized as a feature selection technique to optimize the selection of relevant features for intrusion detection.

Usage in Project: Random Forest with PSO is utilized in the project to develop a powerful classifier for detecting intrusions in cloud networks. PSO assists in selecting features that maximize the model's performance.

Extension Voting Classifier - PSO:

Definition: Extension Voting Classifier is an ensemble learning technique that combines predictions from multiple individual models using Particle Swarm Optimization (PSO) to optimize the selection of relevant features and enhance overall classification performance.

Usage in Project: Extension Voting Classifier with PSO is employed in the project to improve the classification performance of the IDS. By leveraging PSO for feature selection and model optimization, the Extension Voting Classifier enhances the accuracy and robustness of intrusion detection

4. EXPERIMENTAL RESULTS

Accuracy: The accuracy of a test is its ability to differentiate the patient and healthy cases correctly. To estimate the accuracy of a test, we should calculate the proportion of true positive and true negative in all evaluated cases. Mathematically, this can be stated as:

Accuracy = TP + TN TP + TN + FP + FN.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

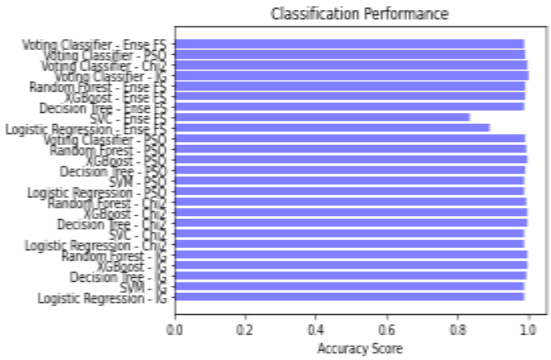

Fig 3 Accuracy Performance comparison Graph

Precision: Precision evaluates the fraction of correctly classified instances or samples among the ones classified as positives. Thus, the formula to calculate the precision is given by:

Precision = True positives/ (True positives + False positives) = TP/(TP + FP)

$$\text{Precision} = \frac{True\ Positive}{True\ Positive + False\ Positive}$$

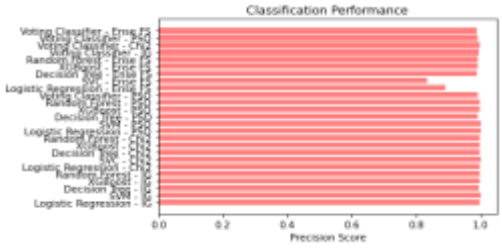

Fig 4 Precision Performance comparison Graph

Recall: Recall is a metric in machine learning that measures the ability of a model to identify all relevant instances of a particular class. It is the ratio of correctly predicted positive observations to the total actual positives, providing insights into a model's completeness in capturing instances of a given class.

ISSN: 2278-4632
Vol-14, Issue-4, April: 2024

$$Recall = \frac{TP}{TP + FN}$$

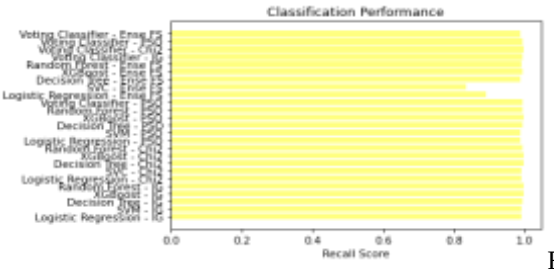

Fig 5 Recall Performance comparison Graph

F1-Score: F1 score is a machine learning evaluation metric that measures a model's accuracy. It combines the precision and recall scores of a model. The accuracy metric computes how many times a model made a correct prediction across the entire dataset.

$$\text{F1 Score} = \frac{2}{\left(\frac{1}{\text{Precision}} + \frac{1}{\text{Recall}}\right)}$$

$$\text{F1 Score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

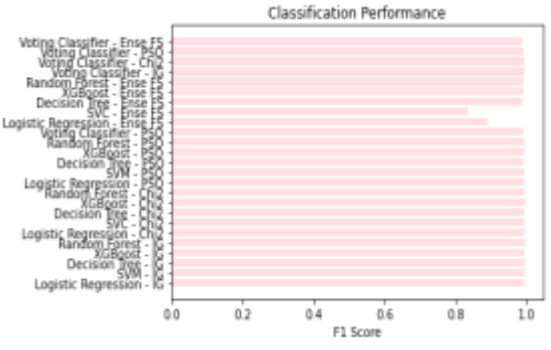

Fig 6 F1 Performance comparison Graph

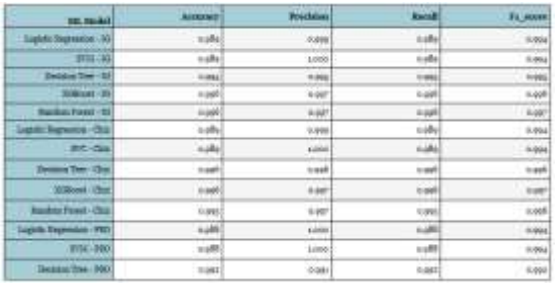

Fig 7 Performance Evaluation Table

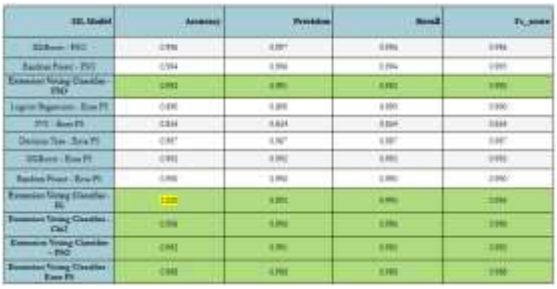

Fig 8 Performance Evaluation Table

Fig 9 Home Page

Fig 10 Sign Up

Fig 11 Sign In

ISSN: 2278-4632
Vol-14, Issue-4, April: 2024

Fig 12 Upload Input Data

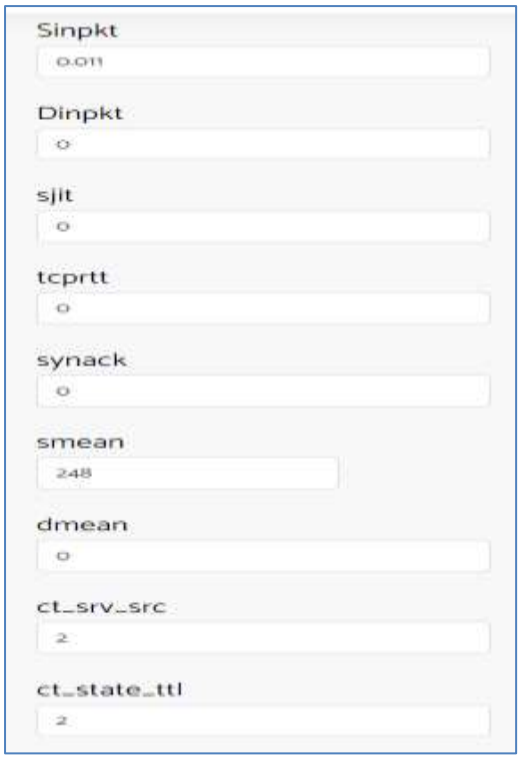

Fig 13 Upload Input Data

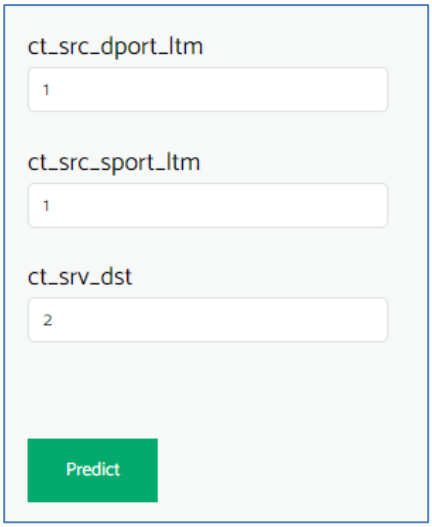

Fig 14 Upload Input Data

Fig 15 Predicted Result

Similarly we can try another inputs data to predict results for given input data

5. CONCLUSION

In conclusion, the project presents a comprehensive and advanced Cloud Intrusion Detection System (IDS) that addresses the critical challenges of imbalanced data and feature selection. By integrating Synthetic Minority Over-sampling Technique (SMOTE) and employing a hybrid approach combining Information Gain (IG), Chi-square (CS), and Particle Swarm Optimization (PSO), the system optimizes feature relevance and ensures precise intrusion detection. Leveraging the Random Forest (RF) model, exceptional accuracies are achieved, particularly in multi-class scenarios evaluated on the UNSW-NB15 dataset. The extension algorithm, exemplified by the Voting Classifier (RF + DT + AdaBoost), demonstrates remarkable accuracy in threat detection, affirming the effectiveness of ensemble methods. Furthermore, the incorporation of Flask and SQLite provides a user-friendly interface, enhancing practical applicability and accessibility for

ISSN: 2278-4632
Vol-14, Issue-4, April: 2024

cybersecurity professionals, cloud providers, and users. Overall, the project significantly enhances cloud intrusion detection, contributing to bolstering overall security in cloud environments.

6. FUTURE SCOPE

The feature scope of the improved design for a Cloud Intrusion Detection System (IDS) using a hybrid feature selection approach with machine learning classifiers encompasses several key components. Firstly, the system aims to address the challenge of imbalanced data through the integration of Synthetic Minority Over-sampling Technique (SMOTE), ensuring a more balanced and representative dataset for training. Secondly, the hybrid feature selection approach, combining Information Gain (IG), Chi-square (CS), and Particle Swarm Optimization (PSO), aims to optimize feature relevance and enhance the precision of intrusion detection. Thirdly, the system leverages machine learning classifiers such as Random Forest (RF), Decision Trees (DT), and AdaBoost to build robust models capable of accurately identifying and classifying network intrusions. Finally, the extension algorithm, including the Voting Classifier, further enhances classification accuracy by combining predictions from multiple individual models. Together, these features contribute to the development of a comprehensive and effective IDS for cloud environments.

REFERENCES

[1] R. R. Kumar, A. Tomar, M. Shameem, and M. N. Alam, ‘‘OPTCLOUD: An optimal cloud service selection framework using QoS correlation lens,’’ Comput. Intell.Neurosci., vol. 2022, pp. 1–16, May 2022, doi: 10.1155/2022/2019485.

[2] R. R. Kumar, M. Shameem, R. Khanam, and C. Kumar, ‘‘A hybrid evaluation framework for QoS based service selection and ranking in cloud environment,’’ in Proc. 15th IEEE India Council Int. Conf., Oct. 2018, pp. 1–6, doi: 10.1109/INDICON45594.2018.8987192.

[3] M. Bakro, S. K. Bisoy, A. K. Patel, and M. A. Naal, ‘‘Performance analysis of cloud computing encryption algorithms,’’ in Advances in Intel ligent Computing and Communication, in Lecture Notes in Networks and Systems, vol. 202. Singapore: Springer, 2021, pp. 357–367, doi: 10.1007/978-981-16-0695-3_35.

[4] (2020). Malware Statistics & Trends Report | AV-TEST.Accessed: Jan. 21, 2023.[Online]. Available: https://www.av-test.org/en/statistics/ malware/

[5] Digital Technology Market Research Services | Juniper Research. Accessed: Jan. 21, 2023.[Online].Available:https://www.juniperresearch. com/home [6] Cyber Security Market Size, Share & Trends Report, 2030. Accessed: Jan. 21, 2023.[Online].Available:https://www.grandviewresearch. com/industry-analysis/cyber-security-market

[7] R. R. Kumar, M. Shameem, and C. Kumar, ‘‘A computational frame work for ranking prediction of cloud services under fuzzy environment,’’ Enterprise Inf. Syst., vol. 16, no. 1, pp. 167–187, Jan. 2022, doi: 10.1080/17517575.2021.1889037.

[8] M. A. Akbar, M. Shameem, S. Mahmood, A. Alsanad, and A. Gumaei, ‘‘Prioritization based taxonomy of cloud-based outsource software development challenges: Fuzzy AHP analysis,’’ Appl. Soft Comput., vol. 95, Oct. 2020, Art. no. 106557, doi: 10.1016/j.asoc.2020.106557.

[9]M.Bakro,R.R.Kumar,A.A.Alabrah,Z.Ashraf,S.K.Bisoy, N. Parveen, S. Khawatmi, and A. Abdelsalam, ‘‘Efficient intrusion detection system in the cloud using fusion feature selection approaches and an ensemble classifier,’’ Electronics, vol. 12, no. 11, p. 2427, May 2023, doi: 10.3390/electronics12112427.

[10] M. Bakro, S. K. Bisoy, A. K. Patel, and M. A. Naal, ‘‘Hybrid blockchain enabled security in cloud storage infrastructure using ECC and AES algorithms,’’ in Blockchain based Internet of Things. Singapore: Springer, 2022, pp. 139–170, doi: 10.1007/978-981-16-9260-4_6.

[11] Z. Ahmad, A. S. Khan, C. W. Shiang, J. Abdullah, and F. Ahmad, ‘‘Net work intrusion detection system: A systematic study of machine learning and deep learning approaches,’’ Trans. Emerg. Telecommun. Technol., vol. 32, no. 1, p. e4150, Jan. 2021, doi: 10.1002/ett.4150.

[12] I. F. Kilincer, F. Ertam, and A. Sengur, ‘‘Machine learning meth ods for cyber security intrusion detection: Datasets and comparative study,’’ Comput. Netw., vol. 188, Apr. 2021, Art. no. 107840, doi: 10.1016/j.comnet.2021.107840.

[13] I. Benmessahel, K. Xie, and M. Chellal, ‘‘A new evolutionary neural networks based on intrusion detection systems using multiverse optimization,’’ Int. J. Speech Technol., vol. 48, no. 8, pp. 2315–2327, Aug. 2018, doi: 10.1007/S10489-017-1085-Y.

[14] Y. Yang, K. Zheng, C. Wu, and Y. Yang, ‘‘Improving the classification effectiveness of intrusion detection by using improved conditional variationalAutoEncoder and deep neural network,’’ Sensors, vol. 19, no. 11, p. 2528, Jun. 2019, doi: 10.3390/s19112528.

[15] B. A. Tama, M. Comuzzi, and K. Rhee, ‘‘TSE-IDS: A two stage classifier ensemble for intelligent anomaly-based intrusion detection system,’’ IEEE Access, vol. 7, pp. 94497–94507, 2019, doi: 10.1109/ACCESS.2019.2928048.

[16] F. A. Khan, A. Gumaei, A. Derhab, and A. Hussain, ‘‘TSDL: A two-stage deep learning model for efficient network intrusion detection,’’ IEEE Access, vol. 7, pp. 30373–30385, 2019, doi: 10.1109/ACCESS.2019.2899721.

[17] R. Vinayakumar, M. Alazab, K. P. Soman, P. Poornachandran, A. Al-Nemrat, and S. Venkatraman, ‘‘Deep learning approach for intelligent intrusion detection system,’’ IEEE Access, vol. 7, pp. 41525–41550, 2019, doi: 10.1109/ACCESS.2019.2895334.

[18] R. Patil, H. Dudeja, and C. Modi, ‘‘Designing an efficient security framework for detecting intrusions in virtual network of cloud computing,’’ Comput. Secur., vol. 85, pp. 402–422, Aug. 2019, doi: 10.1016/j.cose.2019.05.016.

[19] A. I. Saleh, F. M. Talaat, and L. M. Labib, ‘‘A hybrid intrusion detection system (HIDS) based on prioritized k-nearest neighbors and optimized SVMclassifiers,’’ Artif. Intell. Rev., vol. 51, no. 3, pp. 403–443, Mar. 2019, doi: 10.1007/s10462-017-9567-1.

[20] J. Zhang, Y. Ling, X. Fu, X. Yang, G. Xiong, and R. Zhang, ‘‘Model of the intrusion detection system based on the integration of spatial temporal features,’’ Comput. Secur., vol. 89, Feb. 2020, Art. no. 101681, doi: 10.1016/j.cose.2019.101681.

[21] S. M. Kasongo and Y. Sun, ‘‘Performance analysis of intrusion detection systems using a feature selection method on the UNSW-NB15 dataset,’’ J. Big Data, vol. 7, no. 1, pp. 1–12, Dec. 2020, doi: 10.1186/s40537-020 00379-6.

[22] V. Kumar, D. Sinha, A. K. Das, S. C. Pandey, and R. T. Goswami, ‘‘An integrated rule based intrusion detection system: Analysis on UNSW NB15 data set and the real time online dataset,’’ Cluster Comput., vol. 23, no. 2, pp. 1397–1418, Jun. 2020, doi: 10.1007/s10586-019-03008-x.

[23] O. Almomani, ‘‘A feature selection model for network intrusion detection system basedonPSO,GWO,FFAandGAalgorithms,’’Symmetry,

vol.12, no. 6, pp. 1–20, 2020, doi: 10.3390/sym12061046.

[24] K. Jiang, W. Wang, A. Wang, and H. Wu, ‘‘Network intrusion detection combined hybrid sampling with deep hierarchical network,’’ IEEE Access, vol. 8, pp. 32464–32476, 2020, doi: 10.1109/ACCESS.2020.2973730.

[25] P. Rajesh Kanna and P. Santhi, ‘‘Unified deep learning approach for efficient intrusion detection system using integrated spatial–temporal features,’’ Knowl.-Based Syst., vol. 226, Aug. 2021, Art. no. 107132, doi: 10.1016/j.knosys.2021.107132.

[26] G. Sreelatha, A. V. Babu, and D. Midhunchakkaravarthy, ‘‘Improved security in cloud using sandpiper and extended equilibrium deep trans fer learning based intrusion detection,’’ Cluster Comput., vol. 25, no. 5, pp. 3129–3144, Oct. 2022, doi: 10.1007/s10586-021-03516-9.

[27] P. R. Kanna and P. Santhi, ‘‘Hybrid intrusion detection using MapReduce based black widow optimized convolutional long short-term memory neural networks,’’ Expert Syst. Appl., vol. 194, May 2022, Art. no. 116545, doi: 10.1016/j.eswa.2022.116545.

[28] S. Krishnaveni, S. Sivamohan, S. S. Sridhar, and S. Prabakaran, ‘‘Efficient feature selection and classification through ensemble method for network intrusion detection on cloud computing,’’ Cluster Comput., vol. 24, no. 3, pp. 1761–1779, Sep. 2021, doi: 10.1007/s10586-020-03222-y.

[29] K. Potdar, ‘‘A comparative study of categorical variable encoding tech niquesforneuralnetworkclassifiers,’’ Int. J. Comput.Appl.,vol.175,no.4, pp. 7–9, Oct. 2017, doi: 10.5120/ijca2017915495.

[30] M. Rashid, J. Kamruzzaman, T. Imam, S. Wibowo, and S. Gordon, ‘‘A tree-based stacking ensemble technique with feature selection for network intrusion detection,’’ Int. J. Speech Technol., vol. 52, no. 9, pp. 9768–9781, Jul. 2022, doi: 10.1007/s10489-021-02968-1.

[31] V.Lakshmi Chaitanya,Machine Learning Based Predictive Model for Data Fusion Based Intruder Alert System, journal of algebraic statistics, Vol. 13,no. 2,pages. 2477-2483, June 2022.