IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

PRIVILEGE ESCALATION ATTACK DETECTION AND MITIGATION IN CLOUD USING MACHINE LEARNING

S. Nagendrudu [1a)], S. J. Moen [1b)], K. Rakesh Kumar Reddy [1c)], A. Veera Yugandhar Reddy [1d)], and M. Yaseen Basha [1e)].

[1] Asst.Professor,Department of Computer Science & Engineering, Santhiram Engineering College, Nandyal-518501, AndhraPradesh, India

[2,3,4,5] Student,Department of Computer Science & Engineering, Santhiram Engineering College, Nandyal-518501, AndhraPradesh, India

Abstract: Significant cybersecurity challenges have been caused by the development of smart goods due to the recent exponential rise in attack frequency and complexity. Although the huge developments that cloud computing has brought to the corporate sector, because of its centralization, using distributed services like security systems may be difficult. Due to the large amount of data that is sent between companies and cloud service providers, both maliciously and accidentally, valuable data breaches may occur. The malicious insider becomes a crucial threat to the organization since they have more access and opportunity to produce significant damage. Unlike outsiders, insiders possess privileged and proper access to information and resources. So, we proposes a machine learning-based system for insider threat detection and classification, which identifies various anomalous occurrences that may point to anomalies and security problems associated with privilege escalation. Multiple studies have been presented regarding detecting irregularities and vulnerabilities in network systems to find security flaws or threats involving privilege escalation. But these studies lack the proper identification of the attacks. We conclude that incorporating more than one machine learning algorithm can obtain a stronger classification in multiple internal attacks.

Keywords:

Security, Machine Learning algorithms, Cloud Computing, Data models.

INTRODUCTION

An innovative way of enabling and providing services via the Internet is cloud computing.

According to IEEE Transaction on Machine Learning Volume:11, Issue Date:08. May 2023, the financial crisis and rising computing needs have resulted in significant changes to the data processing, storage, and display of the current Cloud Model. By leveraging cloud infrastructure, cloud computing helps users save money on equipment purchases and upkeep. For their systems and the data, they manage, cloud storage providers make use of fundamental security features including encryption, access control, and authentication. Any kind of data can have access to nearly endless storage space on the cloud, contingent only on how quickly, readily, and frequently it is accessed. The capacity of information that is transmitted among companies and cloud facility workers, in cooperation purposely and inadvertently, increases potential of sensitive data breaches. The same characteristics that make modern internet facilities easy for staff members and IT systems to use also make it more challenging for businesses to forbid unauthorized usage. Businesses that use cloud services are subject to fresh security vulnerabilities such as open interfaces and authentication. Expert cybercriminals use their expertise to attack Cloud systems. Machine learning solves security issues and enhances data management through a variety of methods and algorithms. Many datasets lack important statistical features or are proprietary and cannot be made public for privacy reasons. With the cloud computing sector increasing at a rapid pace, there are legal dangers to security and privacy. It is conceivable that even if an employee moves inside the Cloud Company, their access credentials won't always change. As an outcome, outmoded rights are recklessly utilized to giveaway and tamper with vital information. Every version that utilizes a computer has some form of permission. Private files, server databases, and extra facilities are normally only accessible to authorized operators. Using or extending authorization, a hostile attacker can take control of an elevated user account and access a private system.

Attackers have two options: they can move horizontally to take over more computers, or they may move vertically to get admin and root access and ultimately total control over the system.

By exploiting horizontal privilege escalation, an attacker can get access to data that isn't always connected to him. A poorly built Web application may include vulnerabilities that an invader can practice to obtain admission to other users' personal information. Now that a horizontal elevation of privileges exploit has been successfully accomplished, the attacker can see, edit, and replicate confidential data. Figure 1 shows an example of a horizontal privilege escalation attack spanning multiple organizational divisions. For this type of attack to be successful, the attacker usually needs to be extremely skilled at employing malicious software and exploiting weaknesses in particular operating systems. Privilege elevation attacks are defined as giving a person, piece of software, or other object more privileges or privileged access than it already has. Transitioning from a low level of privileged access to a higher level of special access is the attacker's primary goal. The attacker may need to get beyond security protections in order to gain control over vertical access. Vertical privileges controls, which are more complex security model versions that help businesses achieve goals like least privilege and job separation, are shown in Figure 2. An invader might, for illustration, try to get root or administrative access to a network by taking over a legitimate user account. Behavioral analytics can identify unusual activity on company computers or user accounts, which may point to a security breach or a rise in privileges.

We require smart algorithms, like ML algorithms, to detect and foresee insider threats if we are to have improved security protection systems. Furthermore, being aware of how well ML algorithms work when it comes to insider attack classification allows you to select the best algorithm for every scenario and identify which ones could use some improvement. As a result, you can offer a more robust degree of safety. Applying efficient and effective machine learning algorithms to insider assault situations is the goal of this study to improve and speed up the findings. Ada Boost, Light-GBM, XG Boost, and Random Forest are among of the ML algorithms that have been tested and assessed for this purpose. The basic premise of the boosting technique is to educate a poor classifier to provide better results by increasing the classification algorithm's prediction. In terms of insider threat classification, Random Forest, Ada Boost, and XG Boost performed admirably and swiftly. This research intends to make the following contributions:

- The work here assumes a realistic setting for training ML models so that the results can be more accurately reflected in the real world. Afterward this, the effort underlines the distinctions from exercise lower than standard Machine Learning circumstances.

- Develop and test a user-centred insider threat detection process that includes data collection, preprocessing, and Machine Learning model-founded data analysis.

- Provide a thorough result reporting mechanism that shows example and user-built results and evaluates damaging situations in order to better realize insider attack scenarios.

To the greatest of our familiarity, this is the initial paper to compare the presentation of 4 machine learning algorithms (Ada Boost, XG Boost, Random Forest, and Light- GBM) in categorizing insider threats and consuming algorithm act to rapidly determine appropriate define measures that raise the bar for security protection. Current insider risk detection and classification experiments included a variety of models and ensemble approaches.

Those trainings individually applied the representations to dissimilar datasets and then published the categorization findings.

LITERATURE SURVEY

One of the biggest challenges to businesses and government organizations these days is malevolent insider assaults. This study presents a novel methodology for building an insider threat detection system on different data granularity levels that is built on user-centered machine learning. In addition to individual data instances, normal and malicious insiders are also subjected to system assessments and analyses, wherein outcomes related to insider scenarios and detection delays are published and debated. Our findings demonstrate that the machine learning-based detection system can identify new malicious insiders with a high degree of accuracy, even with minimal ground truth.

The on-demand availability of PC framework resources is known as cloud computing. in particular, the ability to store and handle information without direct, individual customer management. Customers can now access data storage and public and private computing on a single platform over the Internet. In addition, it confronts a number of security risks and problems that could impede the uptake of cloud computing models. This article discusses security dangers, challenges, tactics, and solutions related to cloud computing. In a previous study, several respondents voiced concerns about security. A number of the surveys go into detail about security issues and solutions, and another examines the architectural paradigm of cloud computing. All of the security-related issues, challenges, strategies, and fixes are compiled in one article.

One of the biggest risks to government and corporate networks is the occurrence of malicious insider attacks. Insider threat identification presents a distinct set of difficulties due to severely imbalanced data, a lack of ground truth, and behavioral shifts and drifts. In this work, a machine learning-based system for user-centered insider threat detection is proposed and evaluated. In order to identify hostile insiders as well as malicious activities, data is analyzed at various degrees of granularity under realistic situations using machine learning. In-depth examinations of well-known insider threat scenarios utilizing several performance metrics are showcased to enable practical system performance assessment. The evaluation's findings demonstrate that the machine learning-based detection system has a high accuracy rate for identifying new malevolent insiders in unseen data, even with little ground truth. In particular, only 0.78% of false positives are discovered for up to 85% of hostile insiders. Additionally, the system may identify malicious activities as soon as 14 minutes after the initial malicious action. When analyzing insider threat scenarios, analysts can obtain important insights from the system thanks to its comprehensive result reporting.

Internal network security has seen significant hurdles in recent years, as events involving insider threats and losses of businesses or organizations have increased. Insiders' malevolent actions are not detectable by conventional intrusion detection techniques. Insider threat detection technology has received a lot of attention and research due to its effectiveness. In this work, we first evaluate user behavior using the tree structure approach, then we combine it with the Copula Based Outlier Detection (COPOD) method to detect the difference between feature sequences and identify users that are abnormal. We conducted experiments using the CERT-IT insider threat dataset and compared the results with widely used techniques like Isolation Forest.

Wide-area measurement-based damping controllers are used in an interconnected multi-area power system to dampen inter-area oscillations that risk grid stability and limit power flows below their transmission capacity. The impact of wide-area damping control (WADC) is heavily dependent on both electrical and cyber systems. At the cyber system layer, an attacker can disrupt the WADC process by interfering with measurement signals, control signals, or both. Stealthy and coordinated cyber-attacks may overcome traditional cybersecurity safeguards, disrupting WADC's smooth operation. This work provides an anomaly detection (AD) approach that employs supervised machine learning and model-based reasoning for mitigation. The proposed AD method takes measurement signals (input of WADC) and control signals (output of WADC) as input to determine the type of activity: normal, perturbation (little or big signal defects), attack, and perturbation-and-attack. When an abnormality is detected, the mitigation module tunes the WADC signal and selects one of two control status modes: wide-area or local. The suggested anomaly detection and mitigation (ADM) module does away with the requirement for ADMs at widely dispersed actuators by integrating with the WADC at the control center to detect attacks on both measurement and control signals. We investigate coordinated and rudimentary data-integrity attack routes such as pulse, ramp, relay-trip, and replay assaults. The suggested ADM algorithms' performance was examined on a testbed environment for a 2-area 4-machine power system using various attack vector scenarios.

Computer network intrusions and attacks frequently exhibit unique traits and behaviors that call for expert assistance. Attack volume is increasing in tandem with computer network development. In actuality, the need for hiring an experienced individual arises from the fact that specialist knowledge is eroding with time and needs to be assessed and made available in the system. An innovation in IT, cloud computing offers consumers the newest, most sought-after virtual services with cheap infrastructure costs, tremendous flexibility, and little upkeep. One of the biggest security issues facing cloud computing is protecting against network intrusions, which has an impact on the confidentiality, accessibility, and integrity of cloud services. Since cloud computing operates in a shared environment, it is susceptible to several types of risks. Because creating robust access controls is crucial to preserving cloud security, but because cyberattacks are becoming more frequent, this is still a difficult objective. This study proposed an intrusion detection system (IDS) based on a network-based cloud computing architecture using a machine-learning based methodology. We selected features from the CICDDoS2019 database using a feature selection technique in the preceding processing stage. We have employed several well-known categories, including Random Forest, Naive Bayes, Decision Tree, Supporting, and Logistic Regression classifiers. The random forest technique outperforms these five distinct classifiers in terms of overall accuracy, precision, recall, and F1 score in the simulation results. We describe how different aspects of machine learning methods can be leveraged to create effective IDS. The purpose of this research is to improve intrusion detection accuracy in cloud computing by creating a novel technique based on intrusion detection systems and their various designs.

In both academia and business, cloud computing research is now being extensively utilized. Customers and cloud service providers (CSPs) both profit from cloud computing. Numerous studies have been conducted in the literature regarding the security issues related to cloud computing. The purpose of this systematic literature review (SLR) is to examine the extant literature on cloud computing security, dangers, and difficulties. This SLR looked at research papers that were published in popular digital libraries from 2010 and 2020. After a careful review of available studies, we chose 80 papers that address the suggested research questions. This SLR's findings identified seven significant security risks to cloud computing systems. The findings indicated that among the most talked-about subjects in the selected literature were data manipulation and leaks. Additional security threats were linked to data storage and data infiltration in cloud computing environments. The outcomes of this SLR also showed that cloud users and CSPs continue to have difficulties with consumer data outsourcing. The blockchain was mentioned in our survey paper as a collaborating technology to allay security worries. The results of the SLR provide some recommendations for more research to be conducted in order to improve data availability, confidentiality, and integrity.

METHODOLOGY

i) Proposed Work:

This work uses a modified dataset made up of several files from the CERT dataset. Using that dataset improves the performance of machine learning algorithms Random Forest, AdaBoost, XGBoost, and LightGBM. Giving a general overview of the process for identifying and categorizing insider threats is the primary objective. Four methods are used in the suggested model, and they are all applied to the CERT dataset. In previous stages, when constructed models included both technical knowledge and mathematical formulae, the mathematical and technical analysis were predefined. To improve a model's performance, ensemble learning is primarily used. As part of ensemble learning, strategies like bagging, boosting, and stacking are used. Our approach identifies and describes insider risks by employing bagging and boosting algorithms. We can retrieve data that provides pertinent information about the performance of the models through data aggregation during the data pre-processing stage.

During the information preparation phase, data standardization is an extremely useful technique for changing features to be on a comparable size. As a result, the model performs better while maintaining training stability. The technique of feature extraction is critical in reducing the amount of duplicate information gathered throughout the information gathering process.

Finally, data reduction accelerates the learning and generalization stages of machine learning, allowing for model development with less computational work. Boosting is an ensemble learning technique that converts a group of base learners into strong learners in order to reduce training errors. In this regard, the ensemble learning-based boosting technique is used. LightGBM, with the highest accuracy of these four algorithms, is the best.

ii) System Architecture:

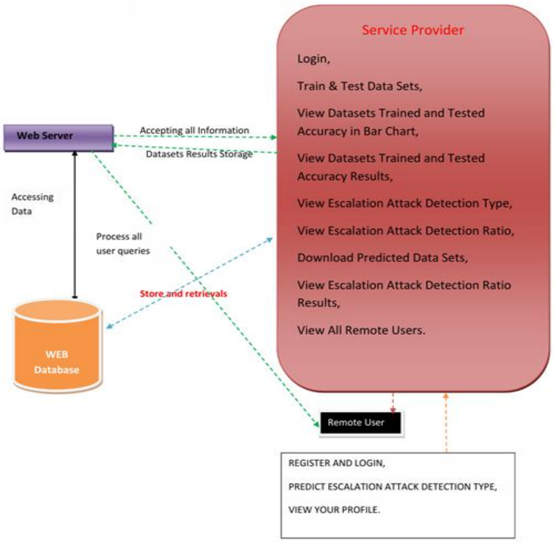

Service Provider
Login,
Train & Test Data Sets,
View Datasets Trained and Tested Accuracy in Bar Chart,
View Datasets Trained and Tested Accuracy Results,
View Escalation Attack Detection Type,
View Escalation Attack Detection Ratio,
Download Predicted Data Sets,
View Escalation Attack Detection Ratio Results,
View All Remote Users.
Web Server
Accepting all Information
Datasets Results Storage
Accessing Data
Process all user queries
Store and retrievals
WEB Database
Remote User
REGISTER AND LOGIN,
PREDICT ESCALATION ATTACK DETECTION TYPE,
VIEW YOUR PROFILE.

Fig 1 Proposed architecture

iii) Dataset collection:

A number of procedures and factors need to be taken into account while building a dataset for cloud-based privilege escalation attack detection and mitigation:

1. Data Sources: Compile information from a range of cloud-based sources, such as:

- Logs: From cloud services like AWS CloudTrail, Azure Monitor, or Google Cloud Audit Logs, access logs, authentication logs, system logs, and network logs are collected.

- Configuration Data: Details gleaned from configuration management systems or cloud provider APIs regarding user roles, permissions, and system configurations.

- User Activity: Data gathered from cloud management consoles or APIs that includes records of user activity, resource access patterns, and authentication events.

- Network traffic: Information gathered by network monitoring tools or cloud-native traffic logs, including data packets, network flow records, and communication patterns among cloud resources.

2. Data Diversity: Make sure that a variety of scenarios and actions, such as both typical behavior and possible privilege escalation assaults, are included in the dataset. This diversity aids in the effective training of machine learning models that differentiate between benign and malevolent behavior.

3. Labeling: Label each action in the dataset to indicate if it is an example of legitimate behavior or a privilege escalation attack. Security professionals can manually assign labels or use automated methods based on known attack patterns and breach signs.

4. Data Privacy: Anonymize or pseudonymize personally identifiable information (PII) and other sensitive data elements in the dataset to preserve data privacy. Achieve adherence to data protection laws like HIPAA and GDPR when managing and archiving private information.

5. Data Quality: Use data cleansing, outlier identification, and normalizing techniques to confirm the dataset's quality and integrity. Resolve mistakes and inconsistencies in the data to guarantee its dependability when developing machine learning models.

6. Data Volume: Gather enough data to accurately depict the range of activities and possible attack scenarios existent in the cloud environment. The dataset ought to be sizable enough to offer a sufficient sample size for successfully training machine learning models.

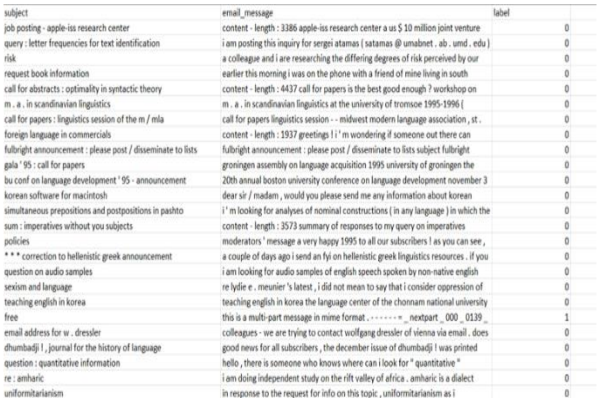
7. Data Representation: Provide the dataset in a format that can be used for machine learning, including feature vectors that have been taken from raw data sources or structured data (like CSV or JSON). In order to extract pertinent features and attributes from the dataset for model training, feature engineering techniques may be used.

8. Data Splitting: To accurately assess the performance of machine learning models, split the dataset into training, validation, and test sets. To make sure that the distribution of malicious and normal activity is the same in every set, use stratified sampling techniques.

9. Data Augmentation: To improve the diversity and resilience of the training data, add more data to the dataset by creating artificial samples or altering already-existing data points.

10. Data Retention: To properly manage the dataset's lifecycle, establish policies for data storage and retention. Respect data governance and compliance regulations while keeping past data for model retraining and analysis.

Organizations can use machine learning techniques to detect and mitigate privilege escalation attacks in the cloud by creating a thorough dataset by adhering to these guidelines and concerns. This dataset is used as the basis for developing and assessing machine learning models that can recognize and appropriately address security threats.



subject	email_message	label
job posting - apple-iss research center	content - length : 3386 apple-iss research center a us $ 10 million joint venture	0
query : letter frequencies for text identification	i am posting this inquiry for sergei atamas (satamas @ umabnet . ab . umd . edu)	0
risk	a colleague and i are researching the differing degrees of risk perceived by our	0
request book information	earlier this morning i was on the phone with a friend of mine living in south	0
call for abstracts : optimality in syntactic theory	content - length : 4437 call for papers is the best good enough ? workshop on	0
m . a . in scandinavian linguistics	m . a . in scandinavian linguistics at the university of tromsoe 1995-1996 (	0
call for papers : linguistics session of the m / mla	call for papers linguistics session - - midwest modern language association , st .	0
foreign language in commercials	content - length : 1937 greetings ! i ' m wondering if someone out there can	0
fulbright announcement : please post / disseminate to lists	fulbright announcement : please post / disseminate to lists subject fulbright	0
gala ' 95 : call for papers	groningen assembly on language acquisition 1995 university of groningen the	0
bu conf on language development ' 95 - announcement	20th annual boston university conference on language development november 3	0
korean software for macintosh	dear sir / madam , would you please send me any information about korean	0
simultaneous prepositions and postpositions in pashto	i ' m looking for analyses of nominal constructions (in any language) in which the	0
sum : imperatives without you subjects	content - length : 3573 summary of responses to my query on imperatives	0
policies	moderators ' message a very happy 1995 to all our subscribers ! as you can see ,	0
* * * correction to hellenistic greek announcement	a couple of days ago i send an fyi on hellenistic greek linguistics resources . if you	0
question on audio samples	i am looking for audio samples of english speech spoken by non-native english	0
sexism and language	re lydie e . meunier 's latest , i did not mean to say that i consider oppression of	0
teaching english in korea	teaching english in korea the language center of the chonnam national university	0
free	this is a multi-part message in mime format . - - - - - - = _ nextpart _ 000 _ 0139 _	1
email address for w . dressler	colleagues - we are trying to contact wolfgang dressler of vienna via email . does	0
dhumbadji ! , journal for the history of language	good news for all subscribers , the december issue of dhumbadji ! was printed	0
question : quantitative information	hello , there is someone who knows where can i look for " quantitative "	0
re : amharic	i am doing independent study on the rift valley of africa . amharic is a dialect	0
uniformitarianism	in response to the request for info on this topic , uniformitarianism as i	0

Fig 2 ESCALATION ATTACK DETECTION dataset

iv) Data Processing:

Data processing is the process of turning unprocessed data into information that is useful to organizations. Processed data is typically collected, arranged, cleaned, verified, analyzed, and put into understandable formats like documents or graphs by data scientists. There are three ways to process data: mechanically, electronically, and manually. Enhancing the value of information and making decision-making easier are the goals. As a result, companies are able to enhance their operations and make critical decisions on time. This is mostly due to automated data processing technologies, such computer software development. It can assist in transforming vast volumes of data, including big data, into insightful understandings for decision-making and quality control.

v) Feature selection:

The process of identifying the most reliable, pertinent, and non-redundant features to employ in the creation of a model is known as feature selection. As the quantity and diversity of datasets increase, it is crucial to gradually reduce their size. Reducing modeling's computational cost and enhancing predictive model performance are the primary objectives of feature selection.

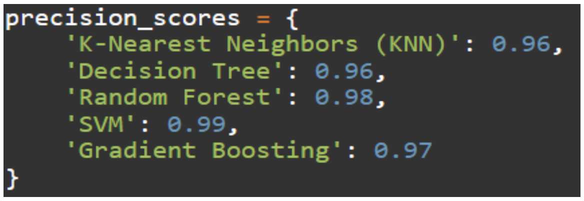
The act of choosing the most crucial features to include in machine learning algorithms is known as feature selection, and it is one of the key elements of feature engineering. By removing redundant or unnecessary features and condensing the set of features to those that are most pertinent to the machine learning model, feature selection approaches are used to decrease the number of input variables. The primary advantages of selecting features proactively as opposed to relying on the machine learning model to determine their relative importance.

EXPERIMENTAL RESULTS

Precision measures the percentage of correctly categorized samples or instances among the positive samples. Consequently, the following is the formula to determine the precision:

True positives/(True positives + False positives) = TP/(TP + FP) is the formula for precision.

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

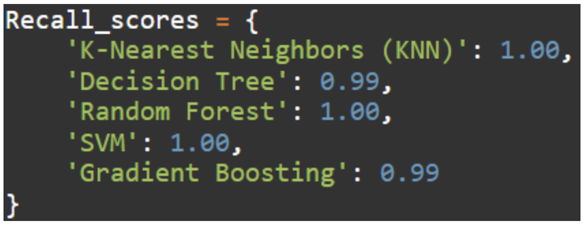


```
precision_scores = {
    'K-Nearest Neighbors (KNN)': 0.96,
    'Decision Tree': 0.96,
    'Random Forest': 0.98,
    'SVM': 0.99,
    'Gradient Boosting': 0.97
}
```

Fig 10 Precision score

Recall: In machine learning, recall is a metric that assesses a model's capacity to locate all pertinent instances of a given class. It is a measure of how well a model captures examples of a particular class: the ratio of correctly predicted positive observations to the total number of real positives.

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

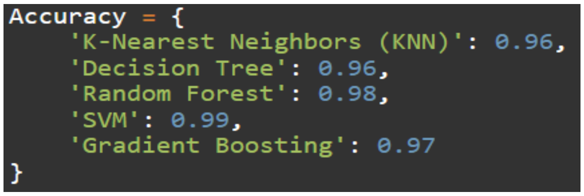


```
Recall_scores = {
    'K-Nearest Neighbors (KNN)': 1.00,
    'Decision Tree': 0.99,
    'Random Forest': 1.00,
    'SVM': 1.00,
    'Gradient Boosting': 0.99
}
```

Fig 11 Recall score

Accuracy: The percentage of accurate predictions made in a classification task is known as accuracy, and it indicates how accurate a model's predictions are overall.

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

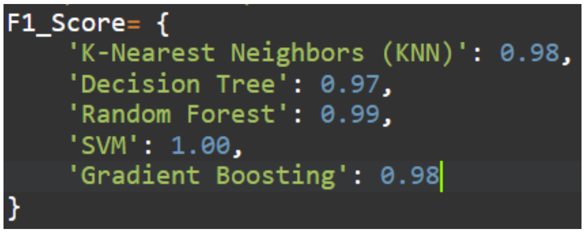


```
Accuracy = {
    'K-Nearest Neighbors (KNN)': 0.96,
    'Decision Tree': 0.96,
    'Random Forest': 0.98,
    'SVM': 0.99,
    'Gradient Boosting': 0.97
}
```

Fig 12 Accuracy score

F1 Score: The F1 Score is appropriate for imbalanced datasets because it provides a balanced metric that takes into account both false positives and false negatives. The harmonic mean of recall and precision is used to calculate it..

$$\text{F1 Score} = 2 * \frac{\text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}} * 100$$



```
F1_Score= {
    'K-Nearest Neighbors (KNN)': 0.98,
    'Decision Tree': 0.97,
    'Random Forest': 0.99,
    'SVM': 1.00,
    'Gradient Boosting': 0.98
}
```

Fig 13 F1_Score

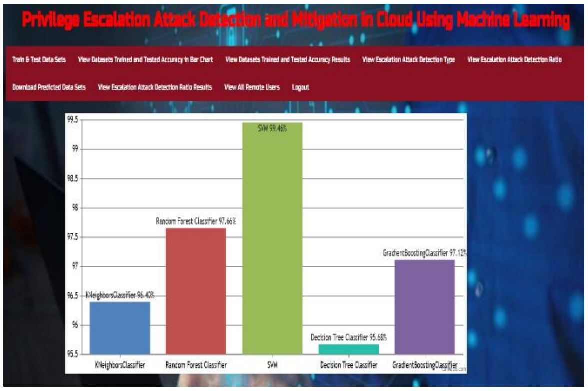
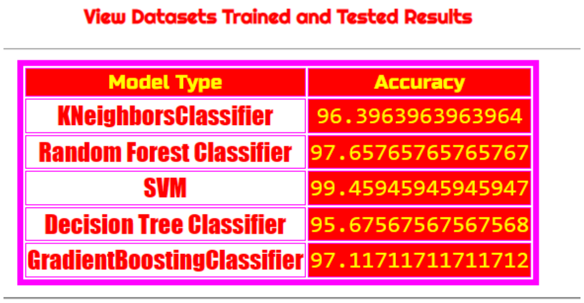


Fig 14 Performance Evaluation in Graph format



View Datasets Trained and Tested Results

Model Type	Accuracy
KNeighborsClassifier	96.3963963963964
Random Forest Classifier	97.65765765765767
SVM	99.45945945945947
Decision Tree Classifier	95.67567567567568
GradientBoostingClassifier	97.11711711711712

Fig 15 Datasets tested results

Privilege Escalation Attack Detection and Mitigation in Cloud Using Machine Learning
Login Using Your Account
User Name
Password
Login
Login Using Your Account
CLOUD
REGISTER

Fig 16 Home page

Privilege Escalation Attack Detection and Mitigation in Cloud Using Machine Learning
Login Using Your Account
User Name
Password
Login
Login Using Your Account
CLOUD
REGISTER

Fig 17 User login page

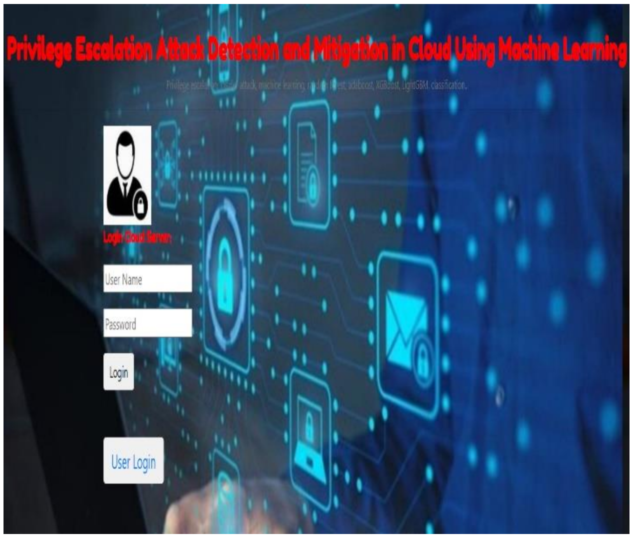

Privilege Escalation Attack Detection and Mitigation in Cloud Using Machine Learning
Login Cloud Server;
User Name
Password
Login
User Login

Fig 18 Cloud login page

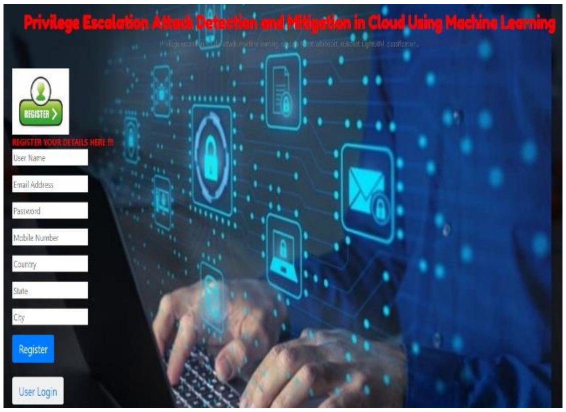

Privilege Escalation Attack Detection and Mitigation in Cloud Using Machine Learning
REGISTER
REGISTER YOUR DETAILS HERE !!!
User Name
Email Address
Password
Mobile Number
Country
State
City
Register
User Login

Fig 19 New user register page

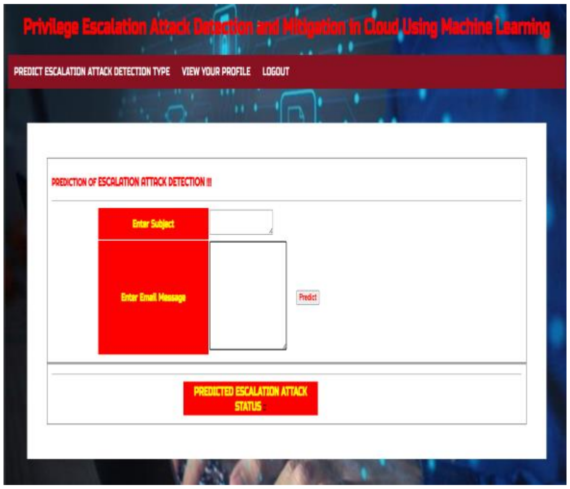
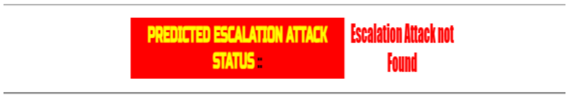

Privilege Escalation Attack Detection and Mitigation in Cloud Using Machine Learning
PREDICT ESCALATION ATTACK DETECTION TYPE
VIEW YOUR PROFILE
LOGOUT
PREDICTION OF ESCALATION ATTACK DETECTION !!!
Enter Subject
Enter Email Message
Predict
PREDICTED ESCALATION ATTACK STATUS

Fig 20 User input page



PREDICTED ESCALATION ATTACK STATUS : Escalation Attack not Found

Fig 21 Predict result for Escalation attack status

1. CONCLUSION

Due to their increased access and potential for serious harm, malevolent insiders pose a serious threat to the organization. Insiders have appropriate and privileged access to knowledge and resources, in contrast to outsiders. The machine learning techniques for identifying and categorizing insider attacks were presented in this work. In this work, a customized dataset comprising of various files from the dataset CERT is employed. When 4 machine learning approaches were deployed on that dataset, the outcomes improved. These algorithms are Light

GBM, XG Boost, Ada Boost, and Random Forest. This study used these supervised machine learning methods to show how effective the experimental findings were in terms of categorization report accuracy. The Light GBM algorithm offers the best accuracy of all the suggested algorithms, at 97%; the other accuracy figures are 86% for RF, 88% for Ada Boost, and 88.27% for XG Boost. Future improvements to the dataset's size and diversity of attributes, as well as emerging insider attack trends, could boost the efficacy of the suggested models. This could lead to new study directions in the identification and categorization of insider threats across a wide range of organizational domains. Businesses utilize models of machine learning to build believable business decisions, better prototypical findings translate into improved choices. Although errors might have a significant financial impact, this cost can be decreased by increasing model accuracy. With machine learning (ML) research, users can feed large volumes of data to algorithms on computers, which utilize the information to assess, suggest, and make decisions.

FUTURE SCOPE

In the evolving landscape of cloud security, leveraging machine learning presents a promising avenue for enhancing the detection and mitigation of privilege escalation attacks, with approaches such as behavioral analysis to discern abnormal patterns, anomaly detection for identifying suspicious activities, real-time monitoring for swift response, contextual analysis to factor in user roles and permissions, and automated response mechanisms to dynamically adjust access controls, collectively aiming to fortify defenses against increasingly sophisticated threats.

REFERENCES

1. Mahammad, F. S., & Viswanatham, V. M. (2020). Performance analysis of data compression algorithms for heterogeneous architecture through parallel approach. The Journal of Supercomputing, 76(4), 2275-2288.

2. Karukula, N. R., & Farooq, S. M. (2013). A route map for detecting Sybil attacks in urban vehicular networks. Journal of Information, Knowledge, and Research in Computer Engineering, 2(2), 540-544.

3. Farook, S. M., & NageswaraReddy, K. (2015). Implementation of Intrusion Detection Systems for High Performance Computing Environment Applications. Inter national journal of Scientific Engineering and Technology Research, 4(0), 41.

4. Sunar, M. F., & Viswanatham, V. M. (2018). A fast approach to encrypt and decrypt of video streams for secure channel transmission. World Review of Science, Technology and Sustainable Development, 14(1), 11-28.

5. Mahammad, F. S., & Viswanatham, V. M. (2017). A study on h. 26x family of video streaming compression techniques. International Journal of Pure and Applied Mathematics, 117(10), 63-66.

6. Devi,S M. S., Mahammad, F. S., Bhavana, D., Sukanya, D., Thanusha, T. S., Chandrakala, M., & Swathi, P. V. (2022).” Machine Learning Based Classification and Clustering Analysis of Efficiency of Exercise Against Covid-19 Infection.” Journal of Algebraic Statistics, 13(3), 112-117.

7. Devi, M. M. S., & Gangadhar, M. Y. (2012).” A comparative Study of Classification Algorithm for Printed Telugu Character Recognition.” International Journal of Electronics Communication and Computer Engineering, 3(3), 633-641.

8. Devi, M. S., Meghana, A. I., Susmitha, M., Mounika, G., Vineela, G., & Padmavathi, M. MISSING CHILD IDENTIFICATION SYSTEM USING DEEP LEARNING.

9. V. Lakshmi chaitanya. "Machine Learning Based Predictive Model for Data Fusion Based Intruder Alert System." journal of algebraic statistics 13, no. 2 (2022): 2477-2483.

10. Chaitanya, V. L., & Bhaskar, G. V. (2014). Apriori vs Genetic algorithms for Identifying Frequent Item Sets. International journal of Innovative Research &Development, 3(6), 249-254.

11. Chaitanya, V. L., Sutraye, N., Praveeena, A. S., Niharika, U. N., Ulfath, P., & Rani, D. P. (2023). Experimental Investigation of Machine Learning Techniques for Predicting Software Quality.

12. Lakshmi, B. S., Pranavi, S., Jayalakshmi, C., Gayatri, K., Sireesha, M., & Akhila, A. Detecting Android Malware with an Enhanced Genetic Algorithm for Feature Selection and Machine Learning.

13. Lakshmi, B. S., & Kumar, A. S. (2018). Identity-Based Proxy-Oriented Data Uploading and Remote Data Integrity checking in Public Cloud. International Journal of Research, 5(22), 744-757.

14. Lakshmi, B. S. (2021). Fire detection using Image processing. Asian Journal of Computer Science and Technology, 10(2), 14-19.

15. Devi, M. S., Poojitha, M., Sucharitha, R., Keerthi, K., Manideepika, P., & Vasudha, C. Extracting and Analyzing Features in Natural Language Processing for Deep Learning with English Language.

16. Kumar JDS, Subramanyam MV, Kumar APS. Hybrid Chameleon Search and Remora Optimization Algorithm-based Dynamic Heterogeneous load balancing clustering protocol for extending the lifetime of wireless sensor networks. Int J Commun Syst. 2023; 36(17):e5609. doi:10.1002/dac.5609

17. David Sukeerthi Kumar, J., Subramanyam, M.V., Siva Kumar, A.P. (2023). A Hybrid Spotted Hyena and Whale Optimization Algorithm-Based Load-Balanced Clustering Technique in WSNs. In: Mahapatra, R.P., Peddoju, S.K., Roy, S., Parwekar, P. (eds) Proceedings of International Conference on Recent Trends in Computing. Lecture Notes in Networks and Systems, vol 600. Springer, Singapore. https://doi.org/10.1007/978-981-19-8825-7_68

18. Murali Kanthi, J. David Sukeerthi Kumar, K. Venkateshwara Rao, Mohmad Ahmed Ali, Sudha Pavani K, Nuthanakanti Bhaskar, T. Hitendra Sarma, “A FUSED 3D-2D CONVOLUTION

NEURAL NETWORK FOR SPATIAL-SPECTRAL FEATURE LEARNING AND HYPERSPECTRAL IMAGE CLASSIFICATION," J Theor Appl Inf Technol, vol. 15, no. 5, 2024, Accessed: Apr. 03, 2024. [Online]. Available: www.jatit.org

19. Prediction Of Covid-19 Infection Based on Lifestyle Habits Employing Random Forest Algorithm FS Mahammad, P Bhaskar, A Prudvi, NY Reddy, PJ Reddy journal of algebraic statistics 13 (3), 40-45

20. Machine Learning Based Predictive Model for Closed Loop Air Filtering System P Bhaskar, FS Mahammad, AH Kumar, DR Kumar, SMA Khadar, ...Journal of Algebraic Statistics 13 (3), 609-616

21. Kumar, M. A., Mahammad, F. S., Dhanush, M. N., Rahul, D. P., Sreedhara, K. L., Rabi, B. A., & Reddy, A. K. (2022). Traffic Length Data Based Signal Timing Calculation for Road Traffic Signals Employing Proportionality Machine Learning. Journal of Algebraic Statistics, 13(3), 25-32.

22. Kumar, M. A., Pullama, K. B., & Reddy, B. S. V. M. (2013). Energy Efficient Routing In Wireless Sensor Networks. International Journal of Emerging Technology and Advanced Engineering, 9(9), 172-176.

23. Kumar, M. M. A., Sivaraman, G., Charan Sai, P., Dinesh, T., Vivekananda, S. S., Rakesh, G., & Peer, S. D. BUILDING SEARCH ENGINE USING MACHINE LEARNING TECHNIQUES.

24. "Providing Security in IOT using Watermarking and Partial Encryption. ISSN No: 2250-1797 Issue 1, Volume 2 (December 2011)

25. The Dissemination Architecture of Streaming Media Information on Integrated CDN and P2P, ISSN 2249-6149 Issue 2, Vol.2 (March-2012)

26. Provably Secure and Blind sort of Biometric Authentication Protocol using Kerberos, ISSN: 2249-9954, Issue 2, Vol 2 (APRIL 2012)

27. D.LAKSHMAIAH, DR.M.SUBRAMANYAM, DR.K.SATYA PRASAD," DESIGN OF LOW POWER 4-BIT CMOS BRAUN MULTIPLIER BASED ON THRESHOLD VOLTAGE TECHNIQUES", GLOBAL JOURNAL OF RESEARCH IN ENGINEERING, VOL.14(9),PP.1125-1131,2014.

28.RSUMALATHA, DR.M.SUBRAMANYAM, "IMAGE DENOISING USING SPATIAL ADAPTIVE MASK FILTER", IEEE INTERNATIONAL CONFERENCE ON ELECTRICAL, ELECTRONICS, SIGNALS, COMMUNICATION &AMP; OPTIMIZATION (EESCO-2015), ORGANIZED BYVIGNANS INSTITUTE OF INFORMATION TECHNOLOGY, VISHAKAPATNAM, 24 TH TO 26TH JANUARY 2015. (SCOPUS INDEXED)

29.P.BALAMURALI KRISHNA, DR.M.V.SUBRAMANYAM, DR.K.SATYA PRASAD, "HYBRID GENETIC OPTIMIZATION TO MITIGATE STARVATION IN WIRELESS MESH NETWORKS", INDIAN JOURNAL OF SCIENCE AND TECHNOLOGY,VOL.8,NO.23,2015. (SCOPUS INDEXED)

30.Y.MURALI MOHAN BABU, DR.M.V.SUBRAMANYAM,M.N. GIRI PRASAD," FUSION AND TEXURE BASED CLASSIFICATION OF INDIAN MICROWAVE DATA – A COMPARATIVE STUDY", INTERNATIONAL JOURNAL OF APPLIED ENGINEERING RESEARCH, VOL.10 NO.1, PP. 1003-1009, 2015. (SCOPUS INDEXED)