www.ijprems.com
editor@ijprems.com

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)
(Int Peer Reviewed Journal)
Vol. 04, Issue 11, November 2024, pp : 2444-2449

e-ISSN : 2583-1062
Impact Factor : 7.001

# CLOUD SECURITY: DETECTING PRIVILEGE ESCALATION WITH MACHINE LEARNING A REVIEW

**J. Pradeep**[1]

[1]GMRIT, India

DOI: https://www.doi.org/10.58257/IJPREMS37277

## ABSTRACT

The rapid expansion of cloud computing and smart device usage has significantly heightened cybersecurity challenges, particularly related to privilege escalation attacks. Centralized cloud services create critical vulnerabilities, making systems prone to both accidental and malicious data breaches. Among potential threats, malicious insiders pose the greatest risk due to their legitimate access, which provides numerous opportunities to inflict substantial damage. An advanced machine learning-based system is proposed to detect and mitigate privilege escalation attacks within cloud environments. This system leverages ensemble learning techniques, including Random Forest, AdaBoost, XG Boost, and Light GBM, to systematically identify and classify anomalous activities that may signal insider threats. A customized dataset derived from the CERT dataset is employed to evaluate the effectiveness of these models, with Light GBM achieving the highest accuracy. Despite the strong performance of Light GBM, it is essential to incorporate multiple machine learning algorithms to ensure robust detection across various insider attack scenarios. Insights from a systematic review of cloud security threats and machine learning methodologies further underscore the necessity of hybrid approaches to improvedetection and mitigation strategies. Integrating past decision data with current machine learning outcomes, applied through supervised learning models across different datasets, offers a promising path toward enhancing the resilience of cloud security frameworks.

**Keywords:** Cloud computing, Machine Learning, Cloud Security, Insider Threat Detection, Privilege Escalation.

## 1. INTRODUCTION

Cloud computing has become a key component of modern IT, but its growing use has also increased security concerns, such as privilege escalation attacks and data breaches. Research has focused on using machine learning (ML) and artificial intelligence (AI) techniques to address these challenges. Studies highlight the effectiveness of ML in detecting and mitigating privilege escalation attacks, while other works survey broader cloud security issues, such as intrusion detection and prevention. Advancements in supervised learning and anomaly detection continue to improve cloud security by providing intelligent, scalable solutions for protecting cloud environments.

## 2. RELATED WORK

Expanding how the methods of machine learning can be utilized in the discovery of unusual user activities. The article discusses the important benefits of ensemble methods such as Random Forest, AdaBoost, and XGBoost in detecting insider threats based on the behavioral analysis of the user [1].

Another paper is discussing using machine learning for the detection and response of privilege escalation attacks that can be committed using a cloud infrastructure. The paper offers an end-to-end machine learning framework based on the integration of several algorithms and techniques, such as anomaly detection and classification algorithms, to successfully face escalating privileges and unauthorized access attempts within cloud environments [2].

Finally, a taxonomy of cloud security attacks is proposed, describing several mechanisms of intrusion detection and prevention, focusing on the cloud-based security models that rely on machine learningin real-time threat detection. In this study, the layover approach to cloud security by integrating machine learning models with standard security protocols for better strategies in defending against cloud vulnerabilities [3].

Overview of machine learning application in cloud security develops the methods such as clustering, classification, and regression algorithms. The survey will be presented here about how machine learning can enhance the discovery process of various security threats especially when applied in changing and large scale environment of cloud computing [4].

Another research does a review on machine learning-based security solutions for cloud computing, analyzing diverse approaches, including supervised learning, that have been fruitful in the identification and mitigation of security threats in cloud infrastructure. In this analysis, other challengesranged from data privacy concerns to effective model training [5].

The application of machine learning algorithms for intrusion detection and fraud prevention in cloud security is discussed, focusing on how machine learning techniques are scalable and adaptable in cloud environments. Such techniques have been observed to give real-time solutions to the increasing cyberattacks that target the cloud-based infrastructures [6].

A review possibility of machine learning in securing cloud environments illustrates the strength of some of the algorithms of machine learning such as decision trees, neural networks, and SVM in the detection and response of the threats to cloud security [7].

A hybrid machine learning model is presented, which integrates past security data with real-time monitoring to secure cloud environments. Arguably, historical data along with current observationswill enhance the accuracy of IDS systems in clouds [8].

This paper discusses the applicability of supervised machine learning techniques in cloud security, bringing forth their efficiency in the identification of anomalous behavior and integrity preservation in cloud systems. The challenges that come with processing large volumes of data in real-time and also the necessity for advanced algorithms to be in line with fast-changing threats is cited in reference [9].

A new intrusion detection approach in cloud environment takes the stacked contractive autoencoders further, combined with support vector machines. Such a hybrid model proves to be fairly effective in identifying anomalous user behaviors and privilege escalation attacks by analyzing the traffic patternsof the cloud [10].

The second is connected with the investigation of anomaly detection techniques that make use of machine learning as well as deep learning methods in cloud networks. More precisely, the focus here will be on discussing the state of the art in deep learning-based methods for anomaly detection in cloud infrastructures, especially with regard to those that are more effective for the discovery of complex, previously unknown threats [11].

It discusses several machine learning algorithms, including random forests and k-nearest neighbors (KNN), in terms of applying machine learning techniques to cloud-based intrusion detection systems. A good discussion has been made about their efficiency to detect security breaches in cloud environments [12].

A comprehensive review of network anomaly detection and the role of machine learning in cloud system security is presented and the possibility of the application of deep learning methods, including CNNs, in enhancing anomaly detection in order to prevent unauthorized access into cloud systems is discussed [13].

A systematic review on machine learning in cloud security identifies the obstacles faced while implementing machine learning in cloud environments. The problems faced in the reviewed article varied from data privacy, scalability of models, and other concerns for real-time detection, while providing potential research pathways in the future to address the afore-mentioned difficulties [14].

The review of the recent advances in securing a cloud environment with the use of machine learning focuses on theory-based models and experimental research. Several techniques of machine learning have been used for intrusion detection and prevention in cloud computing while emphasizing adaptive models to scale up according to the intensity of the system complexity [15].

## 3. METHODOLOGY

### 1.1 Problem Definition:

The central objective of this study is to evaluate a machine learning-based detection framework for privilege escalation attacks within cloud infrastructures by employing machine learning (ML) techniques, particularly ensemble models. which explored the application of ML for such security threats, by focusing on ensemble methods like Random Forest (RF), AdaBoost, XGBoost, and LightGBM to enhance detection precision and resilience against false positives.

### 1.2 Data Collection and Preprocessing:

The first phase of this methodology involves collecting a high-quality dataset focused on privilege escalation events within cloud environments. Real-world datasets can be challenging to obtain due to privacy restrictions; however, resources like the CERT Insider Threat Dataset, Cloud Security Alliance (CSA) logs, or anonymized logs from cloud providers can serve as valuable sources for training and testing. Once acquired, the data undergoes a rigorous preprocessing stage to optimize it for machine learning analysis.

Preprocessing includes addressing any missing data through techniques like mean imputation or forward-filling, depending on the data structure. Continuous features, such as login durations and access

frequencies, are normalized with Min-Max scaling to maintain consistency and model stability. Additionally, outliers—particularly those impacting user activity metrics—are detected and adjusted using Z-scores or the Isolation Forest algorithm, ensuring a balanced and accurate dataset.

### 1.3 Machine Learning Models:

This study employs an ensemble of machine learning models to improve the detection and mitigationof privilege escalation attacks in cloud environments. Ensemble methods are particularly advantageousin security applications, as they combine multiple models to enhance both accuracy and resilienceagainst false positives. By utilizing a combination of models, this approach aims to maximize detection precision while maintaining robustness.

Random Forest (RF): RF builds multiple decision trees and aggregates their results, helping to generalize predictions and reduce overfitting. This model's adaptability is useful for capturing diverse privilege escalation behaviors across various cloud settings.

AdaBoost: AdaBoost strengthens its detection capabilities by iteratively focusing on misclassified instances, increasing the model's sensitivity to slight deviations in user behavior. This quality is valuable for identifying subtle privilege escalations that may mimic normal activity.

XGBoost: Known for its high efficiency and scalability, XGBoost optimizes gradient boosting to deliver fast, accurate results, particularly in large datasets. Its rapid execution capabilities make it suitable for real-time detection in dynamic cloud environments.

LightGBM: A lightweight gradient-boosting model, LightGBM efficiently processes large feature sets with minimal memory usage. It is highly effective for real-time cloud security monitoring due to its ability to capture complex patterns within high-dimensional data while maintaining low computational costs.

Among the ensemble, LightGBM is highlighted as the primary model due to its superior balance of speed, scalability, and resource efficiency. LightGBM's design allows it to process large datasets quickly and identify intricate privilege escalation patterns without burdening system memory, makingit ideal for real-time monitoring in cloud environments. Its ability to handle extensive feature sets with low latency is particularly beneficial in detecting privilege escalation activities that involve complex, multidimensional behaviors.

This ensemble approach capitalizes on the unique advantages of each model, with LightGBM providing the core efficiency needed for effective real-time detection. By integrating these models, the study addresses the essential needs for high precision, recall, and operational feasibility in privilege escalation attack detection within cloud systems.

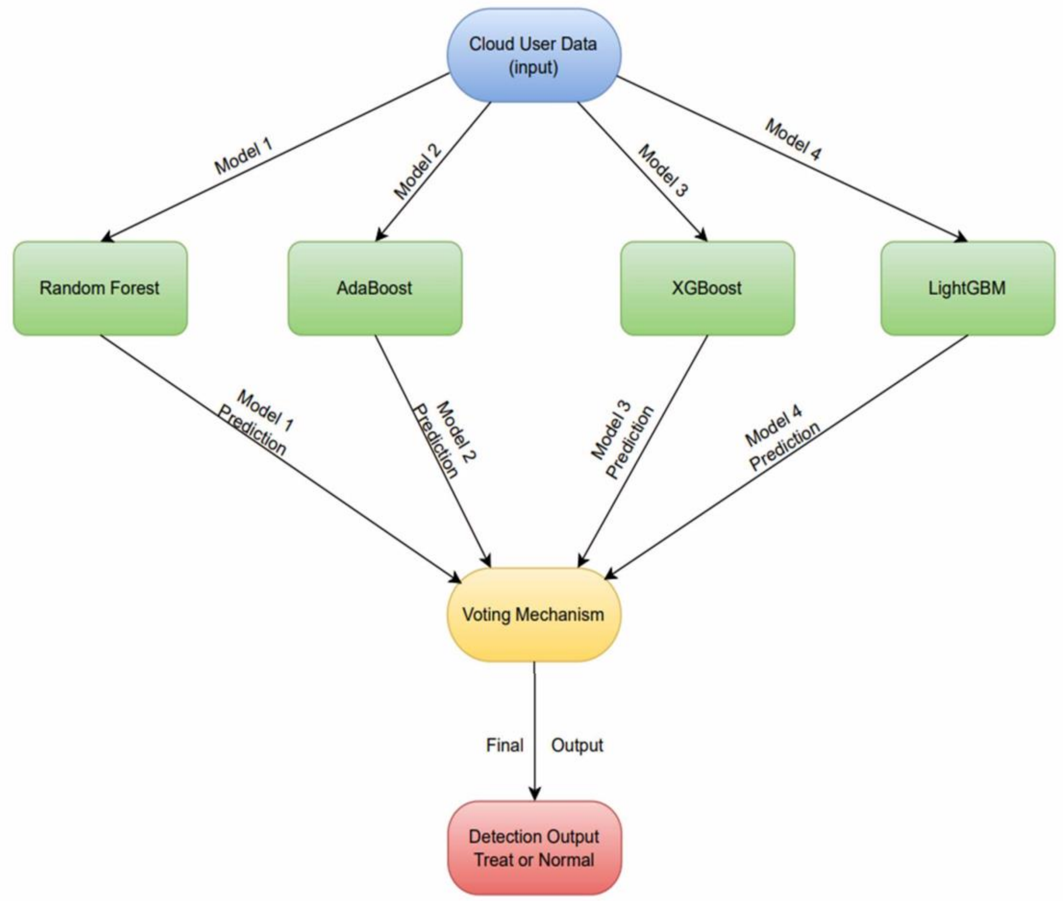


Fig - 1

1.4 Model Training and Evaluation:

This study employs a structured training and testing protocol to ensure model efficacy. The dataset is split into training and testing subsets to evaluate model performance comprehensively. Hyperparameter tuning for each model is conducted using cross-validation techniques, aiming to optimize parameters like the number of trees for RF, learning rate for AdaBoost, and maximum depth for XGBoost and LightGBM.

Accuracy provides an overall measurement of model correctness in detecting escalation events. Performance metrics are analyzed through confusion matrices, enabling a nuanced examination of each model's strengths and limitations, as recommended in the referenced studies. This approach allows fora balanced evaluation, ensuring that the chosen ensemble models deliver both accuracy and reliability.

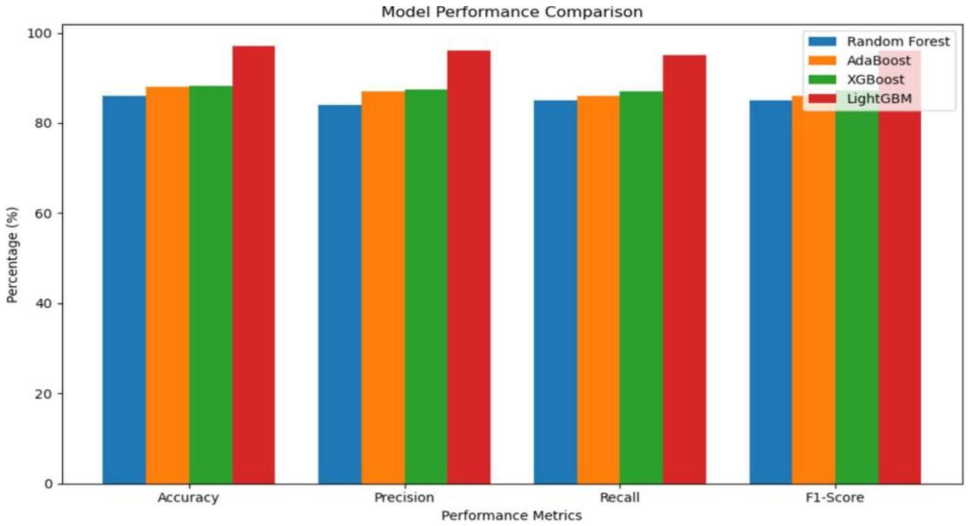


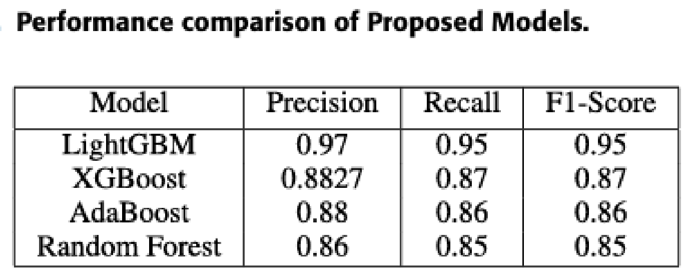
**Fig. 2**

1.5 Challenges and Limitations:

The implementation of machine learning for detecting. privilege escalation in cloud environmentsencounters several key challenges and limitations. First, data availability poses a constraint, as high- quality, real-world datasets on privilege escalation attacks are scarce. Although simulated data offers an alternative, it often lacks the complexity and variability necessary to fully generalize to real-world scenarios, potentially affecting model reliability. Second, model interpretability is a concern; while ensemble methods like LightGBM and XGBoost enhance detection accuracy, their inherent complexity can reduce transparency, which is critical in security-focused applications where insights into model decisions are essential. Additionally, real-time deployment introduces computational demands, as these models require considerable resources for continuous monitoring, which may challenge scalability and efficiency in cloud settings. Finally, adaptability to evolving cloud environments is essential, as insider threat patterns shift over time. Regular retraining is necessary to maintain model efficacy.

## 4. RESULT & DISCUSSIONS

**Result**



**Performance comparison of Proposed Models.**

| Model | Precision | Recall | F1-Score |
|---|---|---|---|
| LightGBM | 0.97 | 0.95 | 0.95 |
| XGBoost | 0.8827 | 0.87 | 0.87 |
| AdaBoost | 0.88 | 0.86 | 0.86 |
| Random Forest | 0.86 | 0.85 | 0.85 |

The table compares the performance of four machine learning models (LightGBM, XGBoost, AdaBoost, and Random Forest) based on three metrics: Precision, Recall, and F1-Score.

- LightGBM performs the best with a Precision of 0.97, Recall of 0.95, and F1-Score of 0.95.
- XGBoost follows, with a Precision of 0.8827, Recall of 0.87, and F1-Score of 0.87.
- AdaBoost has similar Recall (0.86) and F1-Score (0.86) to XGBoost but slightly lower Precision at 0.88.
- Random Forest scores the lowest across all metrics, with a Precision of 0.86, Recall of 0.85, and F1-Score of 0.85.

In summary, LightGBM demonstrates the highest performance among the models.

### Discussion

The results emphasize the suitability of LightGBM for real-time cloud security applications due to its superior precision and computational efficiency.

Its gradient-boosting framework enables rapid processing and accurate identification of privilege escalation patterns, even in large, dynamic datasets. This makes it an ideal choice for deployment in cloud environments where real-time detection is critical.

However, the ensemble approach integrating LightGBM with other models like XGBoost and AdaBoost is crucial for ensuring robustness against diverse attack scenarios. By leveraging the strengths of multiple algorithms, the system can adapt to different insider threat patterns and reduce false positives, enhancing overall system resilience.

### Advantages

High Detection Accuracy: The ensemble framework significantly reduces false positives, ensuring reliability in identifying insider threats.

Real-Time Monitoring: LightGBM's low computational overhead enables real-time anomaly detection, critical for dynamic cloud systems.

Scalability: The framework effectively handles large datasets with complex feature sets, making it scalable for diverse cloud environments.

### Challenges

- Data Availability: The scarcity of real-world datasets on privilege escalation attacks limits the model's ability to generalize across all potential scenarios. Simulated datasets may lack complexity and variability, affecting reliability.
- Model Interpretability: Ensemble methods like LightGBM and XGBoost, while accurate, pose challenges in interpretability, complicating the understanding of decision processes in security applications.
- Adaptability: The evolving nature of insider threats necessitates regular retraining and updating of models to maintain efficacy in dynamic cloud environments.
- Computational Overhead: Although LightGBM is efficient, integrating multiple models in real-time monitoring can introduce computational demands that may challenge resource-constrained cloud systems.

### Future Scope

Future research should focus on improving model interpretability through explainable AI (XAI) techniques, enhancing dataset diversity by collaborating with cloud service providers for real-world data, and optimizing resource usage for real-time applications. Adaptive learning models capable of evolving with changing threat landscapes could further enhance the framework’s resilience.

The proposed ensemble-based framework demonstrates the potential to significantly enhance privilege escalation detection in cloud environments, setting a strong foundation for secure, scalable, and adaptive cloud security systems.

## 5. CONCLUSION

In conclusion, this study underscores the effectiveness of ensemble machine learning models, particularly LightGBM, in enhancing privilege escalation detection in cloud environments. By leveraging the complementary strengths of models like Random Forest, AdaBoost, XGBoost, and LightGBM, the ensemble approach achieves both high accuracy and low false-positive rates, addressing the critical challenge of detecting insider threats with precision. LightGBM's efficiency in processing complex datasets in real-time makes it ideal for scalable, high-speed applications in dynamic cloud settings. This multi-model framework not only strengthens detection robustness but also highlights the potential for hybrid machine learning solutions in advancing adaptive, resilient cloud security systems. Future work can expand on adaptive and interpretable model techniques to further optimize detection in increasingly complex threat landscapes.

## 6. REFERENCES

[1] Mehmood, M., Amin, R., Muslam, M. M. A., Xie, J., & Aldabbas, H. (2023). Privilege escalation attack detection and mitigation in cloud using machine learning. IEEE Access, 11, 46561-46576.

[2] Pratyusha, A., & Kumar, M. N. N. LEVERAGING MACHINE LEARNING FOR DETECTING AND COUNTERING PRIVILEGE ESCALATION ATTACKS IN CLOUD.

[3] Iqbal, S., Kiah, M. L. M., Dhaghighi, B., Hussain, M., Khan, S., Khan, M. K., & Choo, K. K. R. (2016). On cloud security attacks: A taxonomy and intrusion detection and prevention as a service. Journal of Network and Computer Applications, 74, 98-120.

[4] Samunnisa, K., Vijaya Kumar, G. S., & Madhavi, K. (2021). Cloud Security Solutions Through Machine Learning-Approaches: A Survey. Int. J. of Aquatic Science, 12(2), 1958-1972.

[5] Saran, M., Yadav, R. K., & Tripathi, U. N. (2022). Machine learning based security for cloud computing: A survey. Int J Appl Eng Res, 17(4), 332-337.

[6] Butt, U. A., Mehmood, M., Shah, S. B. H., Amin, R., Shaukat, M. W., Raza, S. M., ... & Piran, M. J. (2020). A review of machine learning algorithms for cloud computing security. Electronics, 9(9), 1379.

[7] Subramanian, E. K., & Tamilselvan, L. (2019). A focus on future cloud: machine learning-based cloud security. Service Oriented Computing and Applications, 13(3), 237-249.

[8] Chkirbene, Z., Erbad, A., & Hamila, R. (2019, April). A combined decision for secure cloudcomputing based on machine learning and past information. In 2019 IEEE Wireless Communications and Networking Conference (WCNC) (pp. 1-6). IEEE.

[9] Bhamare, D., Salman, T., Samaka, M., Erbad, A., & Jain, R. (2016, December). Feasibility of supervised machine learning for cloud security. In 2016 International Conference on Information Science and Security (ICISS) (pp. 1-5). IEEE.

[10] [10].Wang, W., Du, X., Shan, D., Qin, R., & Wang, N. (2020). Cloud intrusion detection method based on stacked contractive auto-encoder and support vector machine. IEEE transactions on cloud computing, 10(3), 1634-1646.

[11] Abdallah, A., Alkaabi, A., Alameri, G., Rafique, S. H., Musa, N. S., & Murugan, T. (2024). Cloud Network Anomaly Detection Using Machine and Deep Learning Techniques-Recent Research Advancements. IEEE Access.

[12] Attou, H., Guezzaz, A., Benkirane, S., Azrour, M., & Farhaoui, Y. (2023). Cloud-based intrusion detection approach using machine learning techniques. Big Data Mining and Analytics, 6(3), 311-320. [13].Wang, S., Balarezo, J. F., Kandeepan, S., Al-Hourani, A., Chavez, K. G., & Rubinstein, B. (2021). Machine learning in network anomaly detection: A survey. IEEE Access, 9, 152379-152396.

[13] Nassif, A. B., Talib, M. A., Nasir, Q., Albadani, H., & Dakalbab, F. M. (2021). Machine learning for cloud security: a systematic review. IEEE Access, 9, 20717-20735.

[14] Om Prakash Suman, Lakshya Kumar Saini, Dipender Singh(2024). Securing Clouds with Machine Learning: Advancements in Theoretical and Experimental Research.