International Journal of Creative and Open Research in Engineering and Management
ISSN: **3108-1754 (Online)**
Volume 02 Issue 04 April-2026 | Impact Factor: 3.5

Anomaly Detection using Machine Learning to Improve Security in Cloud System Logs

K Naresh[1], Pujala Pavan Kumar[2]

[1]Assistant Professor, Department of MCA, Annamacharya Institute of Technology and Sciences, Tirupati, Andhra Pradesh, India.

[2]Postgraduate, Department of MCA, Annamacharya Institute of Technology and Sciences, Tirupati, Andhra Pradesh, India.

How to Cite this Article:

Kumar, P. P. (2026). Anomaly Detection using Machine Learning to Improve Security in Cloud System Logs. International Journal of Creative and Open Research in Engineering and Management, 02(04). https://doi.org/10.55041/ijcope.v2i4.072

https://doi.org/10.55041/ijcope.v2i4.072

Abstract

During regular operations, cloud computing environments produce a significant amount of system logs and event recordings. It is challenging and time-consuming to manually monitor these logs in order to identify suspect activity. Conventional security measures frequently rely on rule-based methods that might not be able to identify new or developing cyberthreats. This study suggests a machine learning-based method for identifying questionable system activity from event log data in order to overcome this difficulty. The suggested system looks for unusual behaviour patterns in system activities by analysing attributes including process ID, user ID, event ID, and return values. Before the machine learning model is trained, the dataset is cleaned and prepared using data preparation techniques. System events are divided into two categories by the trained model: suspicious and safe. Python and web technologies are used to provide a web-based interface that enables users to view prediction results and effectively analyse system activities. The suggested method can successfully detect suspicious occurrences and enhance the general security monitoring procedure in cloud settings, according to experimental data. The approach improves the ability to identify possible security risks while reducing the need for manual examination.

Keywords

Machine Learning, Cloud Security, Anomaly Detection, System Log Analysis, Intrusion Detection, Cybersecurity, Suspicious Activity Detection, Data Security

I. Introduction

Cloud computing has emerged as a key technology for contemporary information systems, allowing enterprises to use distributed infrastructures to store, process, and manage massive amounts of data. The swift expansion of cloud services has greatly enhanced accessibility, scalability, and flexibility for both individuals and enterprises. However, because cloud systems are often targeted by cyberattackers looking to take advantage of system flaws and obtain unauthorised access to sensitive data, this expansion has also brought up significant security challenges.

Finding unusual or suspicious system activity is one of the main problems with cloud security. Large amounts of event logs, which document a variety of functions such user activity, system processes, and application events, are constantly produced by cloud systems. These logs provide important information that can be used to spot possible security risks. However, it is challenging, time-consuming, and frequently ineffective to manually analyse such vast amounts of data. Conventional security monitoring techniques mostly rely on rule-based systems, which have a limited capacity to identify new or changing attack patterns.

Machine learning approaches have become effective tools for cybersecurity applications in order to get beyond these restrictions. Large datasets may be automatically analysed by machine learning models, which can also recognise abnormalities that can point to malicious activity and learn trends from past data. Compared to conventional security measures, it is possible to identify suspicious activity more precisely and effectively by utilising machine learning algorithms.

By identifying questionable activity in system event logs, a machine learning-based approach is suggested in this study to improve cloud system security. In order to categorise system events as either safe or suspicious, the suggested method examines important characteristics such process ID, user ID, event ID, and return values. Users can enter event data and obtain real-time predictions using the system's web-based application. This method seeks to improve the whole security monitoring procedure in cloud settings and help managers promptly spot possible attacks.

II. Literature Review

Cybersecurity has grown in importance as a field of study due to the quick development of cloud computing and digital infrastructures. Numerous academics have put forth different strategies to identify malicious activity and shield systems from possible dangers. Firewalls and rule-based intrusion detection systems are examples of traditional security systems that are frequently used to keep an eye on network activity and spot questionable activity. Nevertheless, these systems are less successful in identifying novel or unidentified attack patterns because they mostly rely on pre-established rules and signatures.

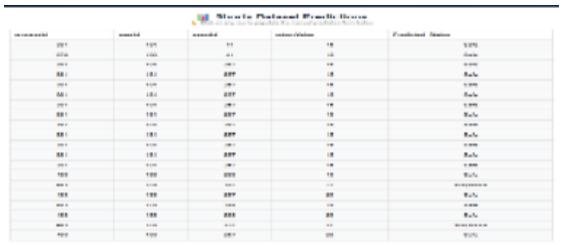
The application of machine learning techniques to improve anomaly and intrusion detection systems has been the subject of several studies. Large datasets can be analysed, hidden patterns can be found, and anomalous behaviours can be automatically detected by machine learning algorithms. For example, classification algorithms like Naïve Bayes, Decision Trees, and Support Vector Machines have been used to identify cyberthreats and harmful network activity. In security monitoring systems, these methods have demonstrated encouraging outcomes in terms of increasing detection accuracy and lowering false alarm rates.

Analysing event data and system logs to spot questionable activity has also been a subject of recent research. System logs can be useful for identifying unusual behaviour since they include comprehensive information about user actions, system operations, and processes. From past log data, machine learning models can identify patterns and categorise future occurrences as suspicious or normal. This method lets security administrators react swiftly to possible attacks and allows for automated monitoring. Although machine learning-based security solutions have advanced, there are still a number of obstacles to overcome. Large-scale log data processing and maintaining high accuracy in real-time detection scenarios are challenges for many current systems. As a result, intelligent systems that can efficiently examine system event logs and more accurately spot suspicious activity are required.

In order to improve security monitoring in cloud environments, the proposed research develops a machine learning-based framework that evaluates system event properties and categorises activities as safe or suspicious.

III. Dataset Description

System event log records that record different operational actions inside a computing environment make up the dataset used in this study. These logs include organised data regarding user interactions, system operations, and



event results. The machine learning model for spotting suspicious or unusual system behaviour is trained and assessed using the dataset. A system event produced during system operation is represented by each record in the dataset.

A number of significant attributes that characterise each event's features are included in the dataset. These characteristics offer helpful data that aids in the machine learning model's ability to identify patterns connected to both typical and anomalous system activity. ProcessId, userId, eventId, and returnValue are the primary attributes utilised in this investigation. While the userId identifies the user connected to that specific action, the processId is the identity of the system process that caused the event. The returnValue shows the response or status that the system returned following the execution of the event, while the eventId indicates the kind of system event that took place.

A number of data pretreatment procedures are carried out to enhance the dataset's quality prior to machine learning model training. These procedures involve feature preparation, data cleansing, and the elimination of inconsistent or missing values. The model is then trained to identify patterns in system behaviour using the cleaned dataset. The system determines whether an occurrence is safe or suspicious based on these patterns.

Fig: Dataset

The suggested system can spot odd behaviour patterns that can point to possible security risks by examining these event properties. The machine learning model's ability to identify significant connections between system operations and the security risks they provide depends heavily on the dataset.

IV. Results and Discussion

Several models were used to assess the efficacy of the suggested machine learning-based method in identifying questionable activity in cloud system event logs. Common evaluation criteria like Accuracy, Precision, Recall, and F1-Score were used to gauge the models' performance. These metrics aid in assessing the model's ability to minimise erroneous predictions while identifying suspicious system events. This study used a Q-learning prototype to assess three machine learning techniques: Random Forest, Deep Neural Network, and Reinforcement Learning. The prepared dataset, which included system event attributes like process ID, user ID, event ID, and return value, was used to train and test each model. The models' categorisation performance was then used to compare them.

According to the experimental results, both the Random Forest model and the Deep Neural Network model attained flawless precision, recall, and F1-score values in addition to an accuracy of 1.00. This shows that, on the provided dataset, these models were highly reliable in accurately classifying the system events into Safe and Suspicious categories. With an accuracy, precision, recall, and F1-score of 0.97—all of which still indicate a high degree of detection capability—the Reinforcement Learning model again showed impressive performance.

The findings imply that deep learning models and ensemble-based models like Random Forest are quite successful at identifying unusual system behaviours in cloud environments. These algorithms are capable of correctly classifying suspicious activity and learning intricate patterns from system event logs. By automatically evaluating event data and spotting possible vulnerabilities, the suggested approach offers an effective way to monitor cloud security. By enabling users to browse dataset details, do exploratory data analysis, and quickly acquire prediction results, the usage of a web-based interface further improves usability.

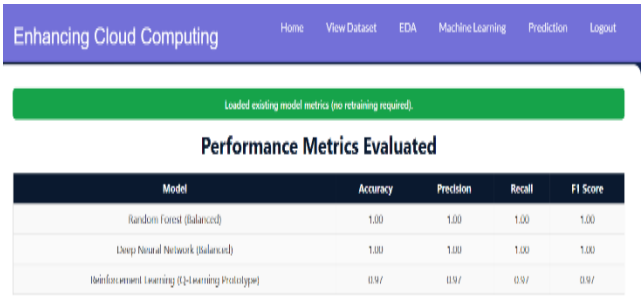


Fig: Evaluation of performance metrics

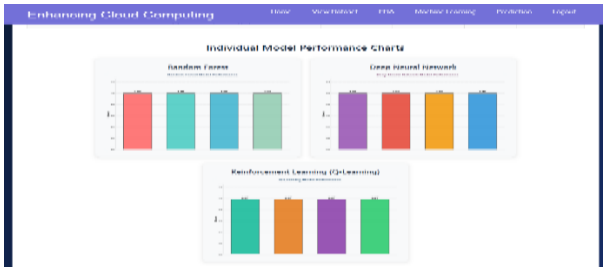


Fig: Individual Model Performance Charts

The evaluation outcomes of three machine learning models—Random Forest, Deep Neural Network (DNN), and Reinforcement Learning utilising Q-learning—used in the suggested system are depicted in the individual model performance charts. Four common assessment criteria were used to evaluate each model: F1 Score, Accuracy, Precision, and Recall.

The Random Forest and Deep Neural Network models both received perfect ratings of 1.00 across all evaluation metrics, as can be seen from the chart. This suggests that these models were highly effective in classifying system events into safe and suspicious categories. Conversely, with accuracy, precision, recall, and F1 score values of 0.97, the Reinforcement Learning (Q-learning) model performed marginally worse but was still very good. These findings show how well machine learning models can spot anomalous activity in cloud system event records.

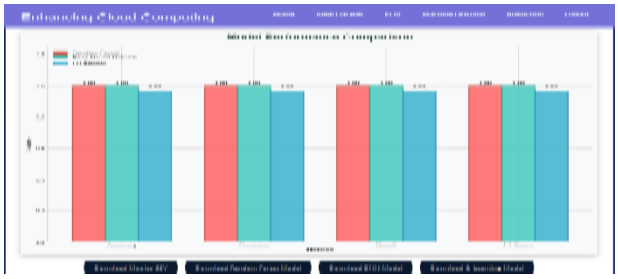


Fig: Model Performance Comparison

Based on the evaluation metrics, the three machine learning models are visually compared in the model performance comparison graphic. In terms of classification accuracy and consistency, the graphic unequivocally demonstrates that Random Forest and Deep Neural Networks perform better than the Reinforcement Learning model. Both models demonstrated a great ability to identify suspicious activity in system event data, as evidenced by their top performance scores across all measures. The Q-learning model's performance is nevertheless competitive and shows the potential of reinforcement learning techniques in security analysis tasks, even though its results were somewhat worse than those of the other two models. All things considered, the comparison shows that deep learning and ensemble-based models offer extremely accurate predictions for anomaly detection in cloud computing settings. These results demonstrate that by automatically identifying suspicious system events, the suggested approach can successfully improve cloud security.

V. Conclusion

By identifying suspicious system activity from event logs, a machine learning-based method was created in this study to improve security monitoring in cloud computing settings. To find unusual behaviour patterns, the suggested system examines important characteristics such process ID, user ID, event ID, and return value. The efficacy of several machine learning models, such as Random Forest, Deep Neural Network, and Reinforcement Learning (Q-learning), in identifying suspicious occurrences was assessed. The Random Forest and Deep Neural Network models performed the best with flawless assessment metrics, according to experimental results, while the reinforcement learning model also revealed outstanding detection capabilities. The addition of a web-based interface makes it easier for users to visualise performance metrics, analyse statistics, and get forecast results.

References:

[1] Sommer, R., & Paxson, V., "Outside the Closed World: On Using Machine Learning for Network Intrusion Detection," IEEE Symposium on Security and Privacy, 2010.

[2] Buczak, A. L., & Guven, E., "A Survey of Data Mining and Machine Learning Methods for Cyber Security Intrusion Detection," IEEE Communications Surveys & Tutorials, 2016.

[3] Chandola, V., Banerjee, A., & Kumar, V., "Anomaly Detection: A Survey," ACM Computing Surveys, 2009.

[4] Scikit-learn Developers, "Scikit-learn: Machine Learning in Python," Available: https://scikit-learn.org

[5] NIST, "Framework for Improving Critical Infrastructure Cybersecurity," National Institute of Standards and Technology, 2018.