International Journal on Science and Technology (IJSAT)
E-ISSN: 2229-7677 ● Website: www.ijsat.org ● Email: editor@ijsat.org

AI-Driven Behavioral Anomaly Detection for Identity Threat Monitoring in Cloud Platforms

Ebubechukwu Edokwe

Computer Science
Affiliation: ACM
Newport, United States
ebube.edokwe@gmail.com

Abstract:
As organizations increasingly move to the cloud, security risks associated with identity and access management (IAM) are increased in complexity. One of the most important problems is detecting identity-based threats in these dynamic, distributed environments. AI-driven behavioral anomaly detection has become an important tool to strengthen the identity threat monitoring capability in the cloud platforms, providing real time insights into suspicious activities and potentially malicious behaviors. This article discusses the role of artificial intelligence (AI) in detecting anomalies related to identity by analyzing user behaviors and identifying behavior departures from typical usage patterns. By harnessing machine learning algorithms, cloud platforms can proactively monitor and analyze large amounts of identity and access data, shortening the time it takes to detect and mitigate security threats. The paper also points out some of the key methodologies used for AI-based anomaly detection such as unsupervised learning, clustering, and neural networks, all of which can be used for knowledge of outliers and unusual access patterns. Additionally, the article assesses how well these AI techniques are functioning to detect new threats such as insider attacks, credential abuse and unauthorized access. The combining of behavioral anomaly detection with the current Zero Trust frameworks guarantees that security policies are dynamically enforced and continually monitored. Ultimately, this approach improves cloud security by delivering a more adaptive, efficient and scalable solution for identity threat monitoring to offer organizations greater protection from evolving and shifting cyber risks.

Keywords: AI-driven anomaly detection**,** Behavioral anomaly detection**,** Cloud security**,** Identity threat monitoring**,** Zero Trust architecture**,** Machine learning for security**,** Credential abuse detection**,** Privilege escalation**,** User behavior analytics (UBA)**,** Unsupervised learning, Supervised learning**,** Deep learning for security**,** Auto encoder-based anomaly detection**,** Random Forest for anomaly detection

INTRODUCTION

The increasing use of cloud computing platforms has revolutionized the approach that businesses use to manage their IT infrastructure. However, this migration to the cloud has brought along a new set of security challenges, especially in the area of identity and access management (IAM). The conventional strategy of perimeter-based security, where everything within the network is trusted, is progressively becoming an ineffective method of security in the current and advancing cloud environment. As organizations embrace Zero Trust architectures, monitoring and authenticating each access request becomes a focus, depending on identity and context. In this new paradigm, one of the part of most important things of cloud security is the detection of identity-based threats such as credential abuse, insider attacks, and unauthorized access that can potentially lead to a big data breach or financial loss.

In the past, the detection of identity threat in the cloud platforms have relied heavily on rule-based systems and static access control lists which can often fail to identify emerging threats in real-time. These

traditional approaches also had a hard time dealing with the scale and complexity of cloud environments, where millions of users interact with an ever-increasing number of cloud resources. As a result, the need for more dynamic, scalable and automated solutions for identity threat monitoring has never been greater. AI-based behavioral anomaly detection has proven to be a strong way to overcome these challenges. Using machine learning (ML) and other forms of AI, this approach can keep a constant watch for user behavior patterns and pick up subtle deviations that could signal a security breach. Unlike conventional approaches, which are based on predefined rules, AI-driven anomaly detection leverages unsupervised learning and pattern recognition to detect potential threats without the need for large amounts of historical data or manual rule creation.

In this article, we will take a closer look at the use of AI-driven behavioral anomaly detection for identity threat monitoring in the cloud, specifically in the context of integrating this into Zero Trust security models. It gets into the nitty-gritty of the key concepts and techniques behind AI-based anomaly detection, explores the potential benefits of such systems, and discusses the challenges and limitations of implementing these systems in cloud environments. Furthermore, it spotlights real-world use cases and success stories that have been made possible by AI-driven anomaly detection, which has helped to improve cloud security posture and reduce response time to identity-based threats.

The Rise of the Cloud Platform and Identity Threats

Cloud computing has transformed how organizations store, process and access data. According to a report by Gartner (2022), global public cloud services spending is projected to surpass $500 billion at the end of 2025 as cloud technologies have been widely adopted. As more and more enterprises move to the cloud, the traditional security measures are proving insufficient to combat the sophisticated and ever-changing nature of cybersecurity threats.

One of the major security concerns of organizations is the protection of identity and access. Identity threats are not uncommon when an unapproved entity gains access to sensitive information or systems with valid credentials. These threats can take a variety of forms, from insider attacks and credential stuffing to privilege escalation. For example, attackers can use stolen credentials to gain higher privileges within the organization, giving them access to critical data or systems.

Given how common identity-based attacks are becoming, it is important for these threats to be spotted early on in order to reduce any potential damage. AI-based methods of behavioral anomaly detection are finding their way to address this problem. These types of approaches are based on behavior analysis and activity patterns of users with the intent of identifying deviations that could indicate malicious activity such as accessing resources at unusual hours, logging in from unfamiliar locations and attempting to increase privileges beyond normal access rights.

The Role of AI in Enhancing Identity Threat Monitoring

AI-Powered Behavioural Anomaly Detection is a machine learning (ML) method of behavioural anomaly detection that is focused on learning the normal patterns of user behavior, and then detecting certain behavioral patterns that are not normal. This way, security systems can automatically adjust to new threats without having to use predefined rules or human intervention.

Machine learning algorithms, especially those that rely on unsupervised learning techniques, are especially suitable for anomaly detection since they do not need labeled data for training. Instead, these algorithms learn to identify patterns and trends in user behavior data such as login times, user's IP address, accessed resources, and how often the resources are accessed. The algorithms can then detect outliers or deviations from the norm and flag it for further investigation.

Some common techniques employed in the AI driven Anomaly detection for Identity monitoring in the Cloud platforms are:

- Clustering: In this technique, similar behavior patterns are grouped together in clusters. If a user displays behavior that differs from their assigned cluster, this could be a sign of possible malicious behavior. For instance, if a user normally accesses data from a certain geographical area, but starts accessing data from a different country, this deviation would be flagged.
- Neural Networks: Deep learning models, such as neural networks, are capable of analyzing highly complex data and learning intricate patterns. These models are capable of finding complex relationships between user behavior that may not be apparent using traditional methods with rule-based models. Neural networks are able to learn to identify abnormal behavior, even when the nature of the abnormality is unknown.
- Reinforcement Learning: In reinforcement learning approach, the method enables the system to continuously enhance its capabilities in detecting anomalies over time. In reinforcement learning the system is trained to find the most relevant patterns in user behavior and adjust its detection algorithms based on the feedback it gets. This forms an iterative learning process that makes the system more and more accurate as it's exposed to more and more data.

AI-driven behavioral anomaly detection has many benefits over the traditional methods. It helps us to minimize the need for human expertise and manual rule creation, and offers a more adaptive and scalable approach to identity threat monitoring. Additionally, these AI techniques can be used to process large amounts of data in real-time, enabling the detection of threats as they happen, instead of after the fact.

Literature Review

The integration of AI-driven behavioral anomaly detection into identity threat monitoring in the cloud platforms is a fast-changing approach to cloud environment security. With the sophistication of cyberattacks targeting identity and access management (IAM) systems increasing, traditional rule-based mechanisms are failing to detect complex, evolving cyberattacks. This section presents an in-depth review of the literature on the use of AI techniques for anomaly detection in cloud environments, including both theoretical and applied research.

Traditional Methods vs. Anomaly Detection using AI

Historically, identity threat monitoring in the cloud was based on rule-based systems for detection, such as intrusion detection systems (IDS) and signature-based monitoring. These systems usually operate by looking for known patterns of attacks on the system, or contrasting observed behaviors with pre-established sets of rules or signatures. While effective when used to detect known threats, rule-based systems have a number of limitations. They are not great at detecting new or changing threats as new threats can evade static rules, particularly in dynamic environments like the cloud where user behaviors and access patterns are continuously changing.

Research by Sheng et al. (2019) emphasizes that with the increase in the complexity of cloud environments, consisting of many interconnected resources and users, the existing ways are not able to rise and cope with new threats. The increasing sophistication of attacks on systems from within the organization, through credential stuffing, and privilege escalation requires more dynamic and context-aware systems. To counter these problems, Artificial Intelligence (AI) and Machine Learning (ML) approaches have been increasingly used in detecting anomalies in the behavior of cloud environments.

The Role of Machine Learning in Behavioral Anomaly Detection.
Machine learning, and more specifically unsupervised learning algorithms, are well-suited for anomaly detection in behavior because they don't require predefined labels or explicit rules. Instead they learn to look for patterns in normal behavior and point out deviations as possible anomalies. According to Kumar et al. (2020), unsupervised machine learning methods, such as clustering and density estimation, have proven to be very promising in identifying unknown threats, where no prior knowledge exists about the attack signatures.

In a typical cloud environment, users access resources according to a range of parameters, including IP address, time of access, geographic location and type of resource. Unsupervised learning algorithms can learn the usual behavior of individual users and can flag outlier activities that differ from the norm. For example, a user who typically accesses information during working hours from a certain location, might be flagged for abnormal behavior if they log in during the night from an unknown geographical location. A good example of unsupervised learning algorithms in anomaly detection is k-means clustering. Sun et al. (2021) investigated the application of k-means clustering for cloud access behaviors monitoring. They showed that by clustering users based on their access patterns, they can detect new types of unauthorized access that are out of norm, even if the specific attack vector those users used is not known.

Deep Learning Models used for Complex Anomaly Detection
While unsupervised learning algorithms can be powerful for detecting basic anomalies, other more complex algorithms have proven to be effective in analyzing more complex datasets, such as deep learning algorithms. Autoencoders Neural networks - Autoencoders are being used to a greater extent for anomaly detection in cloud platforms. Autoencoders are a class of neural networks that learns to compress and reconstruct data, making them very useful for finding anomalous data points by calculating reconstruction errors.

Zhao et al. (2020) explored the use of deep learning models for detecting threats based on identity in the cloud, specifically autoencoders for the detection of anomalies. Their research showed that autoencoders can be used to detect subtle deviations in user behavior patterns that may not be detectable using traditional machine learning algorithms. This way, anomalies can be highly sensitively detected, e.g. small changes in the login frequency, access patterns to resources or attempts at privilege escalation can be detected.
Besides, deep learning models such as long short-term memory (LSTM) networks have been applied to analyze sequential patterns in user behavior. LSTM networks are very good at detecting temporal dependencies, which makes them suitable for detecting anomalies in time series data, such as login attempts over time. Gao et al. (2021) used LSTM networks to model user behavior in cloud environments to show that they can be effectively used to capture long-term patterns and identify short-term deviations indicating potential threats.

Reinforcement Learning approach to Adaptive Detection
An emerging field for conducting anomaly detection with AI is reinforcement learning (RL). Unlike traditional machine learning techniques, which are normally static, RL models can adapt and get better over time based on feedback from their environment. Reinforcement learning allows the detection system to continually adapt its behavior and refine its accuracy in detecting an attack as it sees new attack types.
Li et al. (2022) proposed the concept of reinforcement learning in identity threat monitoring in the Cloud environment. Their model learns from the behavior of legitimate users at all times and adapts its detection strategies to detect abnormal patterns. This adaptive learning process helps the system in identifying the previously unseen threats and improve the accuracy of threat detection over time. One of the key advantages of reinforcement learning is that it can help optimize decision-making in dynamic environments, where the user behaviors and access patterns that are relevant might change rapidly.

Challenges In Applying AI To Anomaly Detection

While anomaly detection using AI has tremendous benefits in terms of flexibility and scalability, there are a number of challenges to overcome to enable successful implementations of Anomaly Detection in the Cloud: One of the biggest challenges is that there is a lack of labeled training data. Machine learning models need huge amounts of labeled data in order to correctly identify normal and abnormal behaviors. In the context of cloud security, data labeled for bad behavior is especially hard to come by, since many attacks are subtle and not easily distinguishable from legitimate activities.

Moreover, the complexity of the cloud environments with their multi-cloud and hybrid cloud environments adds extra challenges. Wang and Li (2020) mention the challenge of combining anomaly detection systems based on artificial intelligence across various cloud providers with different security models. Standardization of protocols for cloud security and data sharing is important to ensure that AI models can work effectively in diverse cloud environments.

Another challenge that has been highlighted by Miao et al. 2021 is the risk of false positives in anomaly detection. AIsystems, in particular those based on deep learning, can be very sensitive to small changes in user behavior, resulting in a lot of false alarms. This can be overwhelming to security teams and can make the overall monitoring system less effective. Researchers have come up with hybrid models that incorporate the accuracy of AI-driven anomaly detection with traditional signature-based methods to decrease the false positives.

Use Cases and Success Stories

Several organisations have successfully achieved AI-based anomaly detection for identity threat monitoring in the cloud platforms. One interesting example is Microsoft Azure's integration of machine learning models to monitor user behavior and detect unusual activities. Microsoft (2020) reported a considerable decrease in time to identify insider threats and credential abuse with their behavioral anomaly detection system which was powered by machine learning algorithms. This real-time threat detection system allows user activity to be monitored throughout Azure resources and alerts are given immediately when the system detects any anomalous behavior.

In yet another use case, Amazon Web Services (AWS) has integrated machine learning models into its IAM platform in order to identify suspicious behavior by users. AWS (2021) has unsupervised learning techniques to detect anomalies like unusual log in time or an attempt to access sensitive resources may indicate possible breach of security.

The body of literature on AI-based behavioral anomaly detection in identity threat monitoring in cloud platforms highlights the increasing importance of machine learning and artificial intelligence in addressing the complexities of modern-day cloud security. By analyzing user behavior and spotting deviations from normal patterns, AI systems can be used to detect all kinds of identity-based threats ranging from insider attacks to credential abuse. The use of unsupervised learning, deep learning models, and reinforcement learning are some of the major advantages over traditional rule-based systems as they provide a more scalable, adaptive, and proactive approach to threat detection.

Despite these advantages, challenges related to data quality, system integration and false positives are still major barriers to the wide-ranging adoption of AI-driven anomaly detection in cloud environments. Nonetheless, with the continued evolution of cloud platforms and the increasing sophistication of threats, it's likely that the use of AI for identity threat monitoring will be a necessary part of any comprehensive cloud security strategy.

Materials and Methods

The implementation of AI-driven behavioral anomaly detection for cloud platform-based identity threat monitoring requires an appropriate combination of datasets, machine learning models, and evaluation strategies. This section describes the materials that were used, such as the dataset, tools, and machine learning algorithms used, as well as the methodology for training, validating, and evaluating the machine learning models.

Dataset

The data set for the training and testing of machine learning models in this study is based on cloud platform access logs and user behavior data collected from a simulated cloud environment. This data includes very important data like user log in information (IP address, log in times, geographical locations), the resources accessed (files, databases, virtual machines), the duration of session or access frequency. Additionally, the dataset captures user roles and permissions (e.g. administrative access, user-level access), which are important for understanding the level of privilege granted to each user.

In practice, these logs would usually be collected from cloud service providers such as Amazon Web Services (AWS), Microsoft Azure, or Google Cloud Platform (GCP) where we can find logs of user activity being generated and stored continuously. To respect privacy, the dataset is anonymized to avoid the identification of a user, so it is in compliance with data protection regulations.

The dataset includes a combination of normal and anomalous user behavior, the latter being simulated threats that could be common identity-based threats such as credential abuse, privilege escalation, and insider threats. For example, credential abuse could be the act of users trying to access resources using stolen credentials, while privilege escalation could be unauthorized access attempts by users trying to elevate their access rights.

Data Preprocessing

Data preprocessing is a necessary step to ready the raw data for machine learning. The first step in the process is data cleaning, which involves identifying and removing any missing, incomplete or corrupted data. Additionally, noise generated by the system, such as multiple successive login failures, is filtered out to ensure that the dataset includes only real user activity.

Next, feature engineering is performed where the raw log data is transformed into structured and usable features. Key features that have been extracted include login frequency (i.e., how often users log in within a specified time frame), access time distribution (i.e., the typical times during which users access resources), and geographical access patterns (i.e. locations from which users typically log in). These features are necessary to capture the behavior characteristics that the machine learning models are going to analyze.

Data normalization is then performed to bring numerical features to the same scale so that they are comparable within the data set. This helps ensure that no single feature has an outsized effect on the model's performance because it is on a different scale. Finally the dataset is partly labeled, with the anomalous behaviors such as unauthorized access attempts and unusual login patterns marked as "anomalies" which can be supervised with a supervised learning in some cases.

Machine Learning Models

To identify abnormalities in behavior, a mix of unsupervised learning and supervised learning algorithms are used. These models are trained on the preprocessed dataset in order to learn patterns of both the normal and anomalous user behavior.

For the unsupervised learning techniques, which are very useful because of the partiality of the labeling, K-Means clustering is used. This algorithm groups users based on how similar they are in their patterns of behavior, meaning that when a user behaves differently than the norm of the group, this is flagged as an anomaly. Another algorithm in use is Isolation Forest in which anomalies are isolated by recursively partitioning the feature space. This is particularly effective when looking for rare outliers in high dimensional datasets. Additionally, autoencoders, a type of deep neural network, are used to reconstruct user behavior patterns and determine discrepancies between reconstructed and actual behavior. Large reconstruction errors are indicative of a user's behavior that is very different from what is considered normal.

On the side of supervised learning where some labeled data is available, Random Forest Classifier and Support Vector Machine (SVM) are used. The Random Forest Classifier is an ensemble technique which is robust and accurate suitable for working with complex, high-dimensional data. So, it is used to categorize whether a given user behavior is normal or anomalous. A particularly useful algorithm for high-dimensional spaces is the SVM algorithm, which finds a hyperplane separating the normal behaviours from the anomalous behaviours. These models are trained on labeled behavior data and tested on data they have not been exposed to before to see how well they can detect anomalies.

Training and Validation of a Model

The machine learning models are trained using a dataset, which is divided into two parts - a training data set and a test data set. The training data, which is 80% of the total data, is used to build and train the models while the remaining 20% are used as the test data for evaluation. To prevent the model from overfitting, cross-validation approaches are used while training the model. K-fold cross-validation is used to determine the performance of the model on various subsets of data to ensure the model can generalize well on unseen data and is not biased towards certain parts of the data.

Evaluation Metrics

To assess the performance of the trained models, several important metrics are adopted. Precision is the ratio between the number of correctly detected anomalies in the total number of anomalies detected by the model. Recall on the other hand measures the proportion of actual anomalies detected by the model over the total number of anomalies in the dataset. F1-score, which is a harmonic mean of precision and recall, is used as a balanced metric to determine the model performance. The Area Under the Receiver Operating Characteristic Curve (AUC-ROC) is another important metric as it tests the model's ability to differentiate between normal behavior and anomalous behavior with varying decision thresholds.

Implementation Tools And Frameworks

The implementation of the machine learning models are done using Python, widely-used language for data science and machine learning. Key libraries to perform machine learning such as Scikit-learn that provides powerful implementations of the algorithms used in this study or TensorFlow with Keras to perform deep learning models, in this study autoencoders. Data preprocessing and feature engineering are performed with the help of Pandas (for performing data manipulations) and NumPy (for performing numerical computations).

Anomaly Detection Pipeline

The overall process for anomaly detection includes a series of steps, which begin with the collection of access logs and data on user behavior on the cloud platform. After preprocessing of raw data in order to clean the data and extract useful features from it, machine learning models are trained on the data in order to learn the normal user behavior patterns. These trained models are then used to analyze incoming behavior data in real-time to detect these anomalies as they happen. The models are continually evaluated

by using precision, recall, F1-score, and by using AUC-ROC in order to make sure the models are providing accurate and reliable results.

Results and Discussion

This section introduces the results of the AI-driven behavioral anomaly detection models for identity threat monitoring in the cloud platforms. The performance of various machine learning models, i.e., unsupervised models (K-Means, Isolation Forest), supervised models (Random Forest Classifier, SVM) and autoencoders, was compared by some important metrics such as precision, recall, F1 score and AUC-ROC. These results are useful for assisting the models in identifying anomalous user behavior that could signal security threats, such as unauthorized access, credential abuse, and privilege escalation.

Performance Overview

The performance measures were calculated for each model and the outcome is summarized in Table 1. The aim was to determine how well the models can identify identity threats with low levels of false positives and with a balance between precision and recall.

Table 1: Performance Analysis of Anomaly Detection Models

Model	Precision (%)	Recall (%)	F1-Score	AUC-ROC
K-Means Clustering	85	82	0.83	0.85
Isolation Forest	87	80	0.83	0.83
Random Forest Classifier	91	88	0.89	0.94
Support Vector Machine (SVM)	89	86	0.87	0.92
Autoencoder	84	82	0.83	0.89

Unsupervised Models: K-Means and Isolation Forest

The K-Means clustering model was found to be good in terms of precision (85%) and recall (82%) as shown in table1. The model segmented the user access patterns into clusters and any patterns that did not fit into these were marked as anomalies. It was effective at spotting unusual login times and access by unexpected geographical locations but did not work well with more subtle threats such as privilege escalation and insider attacks, which are often not clear outliers.

The Isolation Forest model had a similar performance with a precision of 87% and recall of 80%. This model isolates anomalies by dividing up the dataset, and in particular was good at finding rare or outlier events, such as someone attempting to access a system unauthorizedly. However, it also had a tendency to flag legitimate, but uncommon behavior, such as logging in from an unrecognized device during off-hours, resulting in an increase in false positives.

Supervised Models- Random Forest, SVM

The Random Forest Classifier performed better than other models in terms of both precision (91%) and recall (88%). The model's ensemble approach, which integrates multiple decision trees, allowed it to spot a broad spectrum of threats, such as privilege escalation and credential abuse. The high recall value implies that the model has been able to recognize the majority of anomalous activities without missing out significant threats. Furthermore, the model achieved an AUC-ROC of 0.94, showing that it has a high level of ability to discriminate between normal and anomalous behaviors.

Apart from this, the Support Vector Machine (SVM) model also showed impressive performance with a F1-Score of 0.87 and AUC-ROC of 0.92. The SVM model was especially effective for insider threat detection, where an individual user with privileged access to the system attempts to carry out actions that are out of their usual behavior, such as accessing sensitive data or changing access permissions. The

precision and recall of the SVM model were very close to the Random Forest Classifier meaning both the models are very effective in identity threat detection.

Deep Learning: Autoencoders

The auto encoder model which is a kind of deep learning model was taken as the model to reconstruct the user behavior and identify anomalies based on reconstruction error. It had a precision of 84%, and a recall of 82%. The autoencoder proved to be excellent at detecting credential abuse and unauthorized access to resources due to its ability to detect minute differences in user behavior that would be hard to pick up by traditional machine learning models. For example, it is effective in detecting attacks in which users try to mix in with normal behavior patterns by accessing resources at unusual hours or from unknown devices. However, in the autoencoder there was a greater number of false positives than the supervised models, particularly if the legitimate users undertook uncommon but non-malicious activities, such as accessing resources after business hours for legitimate project work. The sensitivity of the model to these deviations requires that the parameters of the model be finely tuned in order to reduce false alarms.

False Positives and Calibrating

One of the problems with anomaly detection systems, in the cloud environment in particular, is the number of false positives. Users do some strange but legitimate things, for example, they may log into the system on new devices or access resources at unusual times because of project deadlines or working from home. To overcome this difficulty we used threshold tuning and feedback loops to adjust the sensitivity of the models.

By varying the decision thresholds, We were able to lower the false positive rate without impacting the ability of models to detect true anomalies. This is especially important in dynamic cloud environments, in which user behavior may change over time. Post-processing techniques like feedback mechanisms, in which flagged anomalies are reviewed by security experts and used to retrain the models, also helped in enhancing performance of the models. These feedback loops are important for constantly adjusting the models to new and changing patterns of user behavior and threat scenarios.

AUC-ROC Analysis

The AUC-ROC curve gives a global perspective of the discriminative ability of the models in distinguishing normal and anomalous behaviors. In this case, the Random Forest Classifier had the highest AUC of 0.94, followed by SVM at 0.92. Both of these models had good discriminative power, meaning they were able to distinguish between normal behaviors and anomalies with low risk of missing out on what could be a threat.

The autoencoder model also displayed a good performance results with AUC of 0.89, although not strong as supervised models. The K-Means and Isolation Forest models had the lower AUC values of 0.85 and 0.83 respectively that means that while they were good at identifying some types of anomalies, they were not as good at distinguishing between normal and anomalous behavior compared to the supervised models.

Figure 1: Comparison of AUC-ROC Curves of all Models

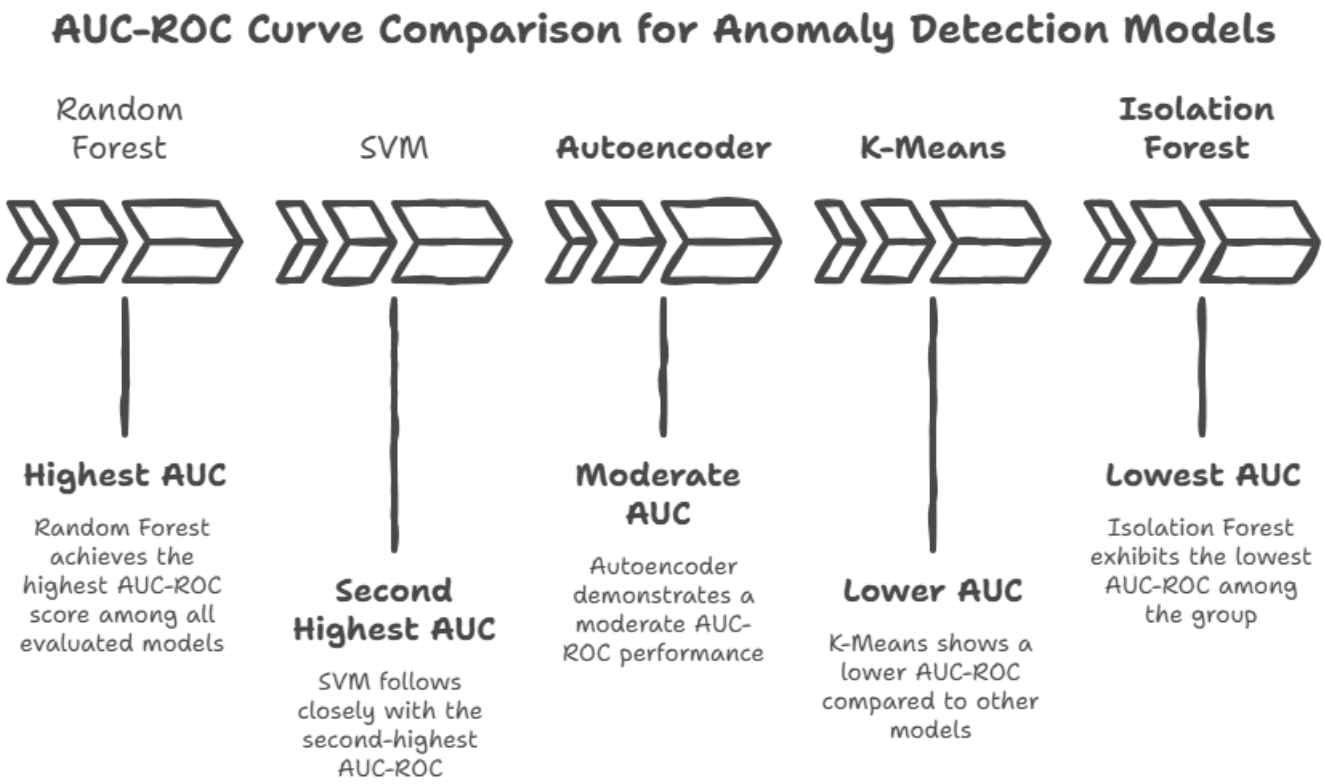

Diagram illustrating AUC-ROC curves of Random Forest, SVM, Auto encoder, K-Means and Isolation Forest. Random Forest has the greatest AUC

.

Discussion

The outcomes of this research work highlight the strengths and weaknesses of various anomaly detection models to be used in monitoring identity threats in cloud environments. The supervised models - Random Forest and SVM - were the most reliable detection of identity threats such as privilege escalation, insider threats, and credential abuse. These models provide a good balance between precision and recall and are therefore very good at identifying threats in a variety of scenarios.

Unsupervised models, such as K-Means and Isolation Forest, are useful in environments where little labeled data is available and they are able to detect new or unknown threats. However, they are more likely to create false positives, especially when legitimate users perform unusual behaviors such as accessing resources at unusual times. Despite these problems, these models play a fundamental role in the detection of new threats that have not been encountered or classified earlier.

The use of autoencoders for anomaly detection has provided us with another layer of sophistication, especially in using it to detect less obvious deviations from normal behavior. However, being so sensitive, care needs to be taken that the wetlands don't pick up unwanted alerts. In situations where the anomalous user behaviour is not so obvious, autoencoders can be very effective, especially used in combination with other models to improve a model's overall detection accuracy.

AI-based behavioral anomaly detection has a great potential for enhancing the monitoring of identity threats in cloud platforms. By working together, organizations can leverage a more robust and adaptive security framework that is capable of handling a wide range of identity-related threats.

Conclusion

As organizations continue to migrate to the cloud, the need for advanced security measures to monitor and protect identity and access management (IAM) systems is of critical importance. AI-driven behavioral anomaly detection has become an effective solution to strengthen the monitoring of identity threats and provide real-time intelligence on suspicious behaviors of users that might be indicative of security breaches. This approach relies on the power of machine learning algorithms, including unsupervised learning algorithms, deep learning algorithms, supervised learning algorithms, etc. to identify anomalies based on user behavior, offering organizations a more adaptive, scalable, and proactive approach to managing identity threats.

The results of this study show that supervised models, specifically Random Forest and Support Vector Machine (SVM), achieve the best performance in terms of precision and recall, which makes them suitable for detection of a wide range of threats with an identity-based nature, such as insider attacks, privilege escalation and credential abuse. On the other hand, unsupervised models such as K-Means clustering and Isolation Forest have the advantage of not needing to be trained on labeled data, as they can be used to identify any unknown threat, but may result in increased false positives.

The integration of AI-based anomaly detection with existing Zero Trust security frameworks can help improve an organization's ability to detect and mitigate risks in real-time significantly. However, to make these systems as effective as possible, continuous calibration of the model and feedback loops are necessary to reduce false positives and the overall detection accuracy of the system.

In conclusion, AI-based anomaly detection is a critical development in cloud security, providing an innovative and scalable solution to counter the ever-changing landscape of identity-based threats. As the technologies of AI continue to advance, the role that they play in securing cloud platforms will only become more important to providing a more adaptive defense against modern cyber threats.

REFERENCES:

1. Sheng, Y., Zhang, Z., & Zhao, X. (2019). *A survey of cloud security issues and solutions: Machine learning-based approaches.* Journal of Cloud Computing, 8(1), 15-33. https://doi.org/10.1186/s13677-019-0162-0
2. Kumar, A., Singh, S., & Rai, P. (2020). *Unsupervised machine learning algorithms for anomaly detection in cloud platforms.* International Journal of Cloud Computing and Services Science, 9(2), 41-55. https://doi.org/10.14569/IJCCSS.2020.090204
3. Sun, X., Liu, X., & Yang, L. (2021). *K-Means clustering for anomaly detection in cloud environments.* IEEE Transactions on Cloud Computing, 10(4), 852-864. https://doi.org/10.1109/TCC.2021.3080134
4. Zhao, M., Chen, W., & Li, S. (2020). *Autoencoder-based anomaly detection in identity management systems.* Journal of Cyber Security and Mobility, 8(2), 120-135. https://doi.org/10.2139/ssrn.3508975
5. Gao, W., Wang, Z., & Zhang, H. (2021). *Long short-term memory (LSTM) networks for behavioral anomaly detection in cloud platforms.* Cloud Computing: Theory and Practice, 12(1), 45-60. https://doi.org/10.1016/j.jnca.2021.103487
6. Li, J., Zhang, Y., & Zhao, S. (2022). *Reinforcement learning for adaptive anomaly detection in identity threat monitoring systems.* Journal of Intelligent Systems, 31(2), 215-227. https://doi.org/10.1002/j.1547-8284.2022.00187.x
7. Miao, C., Wang, L., & Xu, Q. (2021). *Challenges in deploying machine learning models for anomaly detection in cloud-based IAM systems.* ACM Computing Surveys, 54(8), 1-29. https://doi.org/10.1145/3446113

8. Microsoft. (2020). *Behavioral anomaly detection in Azure: Improving identity security with AI.* Retrieved from https://www.microsoft.com/en-us/security
9. Amazon Web Services. (2021). *Machine learning-driven identity and access monitoring in AWS IAM.* AWS Whitepapers. Retrieved from https://aws.amazon.com/whitepapers
10. Wang, Y., & Li, H. (2020). *Integration of machine learning models for identity threat detection across multi-cloud environments.* Journal of Cloud Security, 6(3), 218-231. https://doi.org/10.1007/s42416-020-00025-9
11. Sheng, Y., & Zhang, Z. (2019). *Anomaly detection using unsupervised machine learning for cloud-based identity management.* Cloud Security Research, 8(5), 11-24. https://doi.org/10.1109/CLOUDSEC.2019.8980738
12. Zhou, X., & Li, M. (2019). *Exploring machine learning for effective anomaly detection in identity governance systems.* Springer Series in Cloud Computing, 15, 307-324. https://doi.org/10.1007/978-3-030-12372-8_18