Article

Detecting Malicious Insider Threats through Anomaly-Based User Behaviour Analytics in Enterprise Networks: Machine Learning Approach

Idris Olanrewaju Ibraheem
https://orcid.org/0009-0002-3677-2478
Fountain University, Nigeria
ibraheem.idris@fuo.edu.ng

Adepoju Temilola Morufat
https://orcid.org/ 0009-0002-7478-97
Federal Polytechnic Ayede, Nigeria
atemilola@gmail.com

Samuel Kwabla Segbefia
https://orcid.org/0000-0002-8246-5287
University of Cape Coast, Ghana
samsegbefia329@gmail.com

Ahmed Abiodun Abdulrasaq
https://orcid.org/0009-0001-4569-9375
Al-Hikmah University, Nigeria
abiodun005ng@gmail.com

Abstract

Detecting malicious insider threats within enterprise networks is essential for robust cybersecurity. Insiders with authorised access present significant risks that traditional security measures often fail to address. This paper explores the application of anomaly-based User Behavior Analytics (UBA) to identify these threats by examining a comprehensive dataset of user activities. The study assesses the performance of three machine learning models: Isolation Forest, One-Class SVM, and Autoencoder. Rigorous evaluation demonstrates the Autoencoder model's superior performance compared to other models, as evidenced by higher precision, recall, F1-score, and ROC-AUC metrics. These findings underscore the Autoencoder's effectiveness in accurately detecting insider threats, highlighting its potential as a valuable tool in enhancing enterprise network security. The results indicate that leveraging anomaly-based UBA with advanced machine learning techniques can significantly improve the detection and mitigation of insider threats, providing a more proactive and efficient approach to safeguarding sensitive information within organisations.

Keywords: malicious insider; machine learning; User Behaviour Analytics

UNISA university of south africa PRESS
Southern African Journal of Security
Volume 23 | 2025 | #18099 | 23 pages

https://doi.org/10.25159/3005-4222/18099
ISSN 3005-4222 (Online)

Introduction

In the realm of enterprise security, insider threats represent a critical and growing concern. These threats, stemming from individuals with legitimate access to sensitive systems and data, challenge traditional security mechanisms that primarily focus on external adversaries. Whether intentional or accidental, malicious insider activities can lead to significant financial and reputational damage. As organisational infrastructures become more complex, the need for robust detection mechanisms that can adapt to the subtle and often unpredictable nature of insider threats has become paramount.

Anomaly-based User Behavior Analytics (UBA) has emerged as a promising approach to addressing these challenges. By leveraging advanced machine learning algorithms, UBA systems analyse user activities to detect deviations from established behavioural baselines (Al Mansur and Zaman 2023). Unlike traditional rule-based systems, which are limited to known attack patterns, anomaly-based UBA excels at identifying novel or evolving threats. This makes it particularly well-suited for detecting malicious insiders who operate within their authorised privileges, often leaving minimal traces of their intent (Desetty 2024).

This paper explores the integration of machine learning techniques with anomaly-based UBA to enhance the detection of insider threats in enterprise environments. Specifically, the study evaluates the performance of three models, which are Isolation Forest, One-Class SVM, and Autoencoder, on a curated dataset of user activity. By examining their precision, recall, F1-scores, and ROC-AUC metrics, the paper aims to identify the most effective approach for real-time detection of anomalous behaviour indicative of insider threats.

Furthermore, the research contributes to the field by addressing key gaps in existing literature. While much attention has been given to anomaly detection in general cybersecurity contexts, few studies have focused on the unique challenges posed by insider threats. This paper not only highlights these challenges but also provides a comprehensive analysis of machine learning models tailored for anomaly-based UBA. By doing so, it offers practical insights for researchers and practitioners seeking to strengthen enterprise security against malicious insiders. Traditional security measures, such as firewalls and intrusion detection systems, are often ineffective against insiders because these individuals are already within the security perimeter. Anomaly-based detection, which focuses on identifying unusual patterns of behaviour that deviate from established baselines, offers a promising approach to addressing this challenge. By leveraging advanced machine learning techniques, organisations can analyse vast amounts of data to detect subtle indicators of insider threats that might otherwise go unnoticed (Zewdie, Girma, and Sitote 2024).

Research Objectives

i) anomalies affecting user behaviour for malicious insider threats;

ii) characteristics and behaviours of malicious insider threat actors;

iii) integration of machine learning algorithms for the identification of abnormal activities; and

iv) the development of a capable real-time malicious threat detection system.

Novelty of the Study

This study introduces a novel approach to insider threat detection by combining advanced machine learning techniques with user behaviour analytics in a way that is tailored specifically for enterprise environments. Unlike previous studies that often focus on external threats or use generic anomaly detection methods, this research emphasises the uniqueness of insider threats and the complexity of detecting them in real-time. Additionally, the comparative analysis of different machine learning models within this context adds value by identifying the most effective techniques for various scenarios, contributing to both the academic field and practical cybersecurity implementations.

Research Purpose

The purpose of this research was to address the growing concern of insider threats within enterprise networks, which have proven to be one of the most challenging aspects of cybersecurity. By leveraging anomaly-based user behaviour analytics, this study aimed to provide a more proactive approach to threat detection, as traditional signature-based methods often fail to identify novel or sophisticated attacks. The research sought to enhance the security posture of organisations by focusing on the subtle and often overlooked patterns of user behaviour that may indicate malicious intent.

Literature Review

Insider threats continue to be a significant concern for organisations, especially given the increasing sophistication of cyber-attacks and the volume of sensitive data being handled. According to Cappelli et al. (2012), insider threats are broadly categorised into three types: IT sabotage, theft of intellectual property, and fraud. Their research highlights the need for comprehensive monitoring and robust detection mechanisms to identify and mitigate insider threats effectively. Insider threats pose a significant challenge for organisations, as highlighted by Cappelli et al. (2012). These threats are primarily categorised as IT sabotage, theft of intellectual property (IP), and fraud. IT sabotage involves insiders damaging or disrupting an organisation’s information systems, often motivated by revenge or dissatisfaction. Theft of intellectual property typically involves the unauthorised acquisition of proprietary or confidential information, driven by personal gain or to assist competitors. Fraud encompasses various deceptive activities conducted for financial gain, often involving falsification of records or manipulation of processes.

Insider threats pose a significant challenge to enterprise cybersecurity, primarily because malicious actors exploit their legitimate access to systems. Behavioural anomalies, such as unusual login times, excessive file access, or irregular data transfer volumes, often serve as key indicators of insider threats. Anomaly detection, leveraging deviations from baseline behaviours, has emerged as a pivotal method to address these challenges (Kim et al. 2019; Yuan et al. 2023).

The foundation of anomaly detection lies in modelling normal behaviour patterns and flagging deviations as potential threats. Nazir et al. (2021) emphasise the role of machine learning algorithms, such as one-class classification and autoencoders, in identifying unusual activities. These methods work effectively with imbalanced datasets, where malicious activities are rare. Yuan et al. (2023) extend this view by highlighting the importance of integrating historical behavioural patterns for enhanced prediction accuracy. Deep learning approaches, particularly those incorporating temporal data, have proven to be highly effective. LSTM networks, for instance, have been employed to predict user actions based on historical data, flagging low-probability events as anomalies (Villarreal-Vasquez 2020). These models outperform traditional statistical approaches by adapting to evolving user behaviour patterns.

Several studies validate the efficacy of anomaly-based insider threat detection. For example, Kim et al. (2019) applied multiple anomaly detection algorithms to user log data, demonstrating high accuracy in detecting threats within an imbalanced dataset. Similarly, Nazir et al. (2021) used LSTM-based autoencoders to reconstruct user activity, identifying anomalies with a precision of 92 per cent. The detection of behavioural anomalies provides a robust framework for identifying insider threats. Advances in machine learning and deep learning, particularly models leveraging temporal and historical data, have significantly enhanced detection capabilities. However, addressing challenges such as data sparsity, false positives, and privacy concerns remains critical for the effective deployment of these systems (Kim et al. 2019; Nazir et al. 2021).

The Behavioral Rhythm Insider Threat Detection (BRITD) framework introduces time-aware anomaly detection, aligning detection models with users’ natural behaviour cycles. This approach has been shown to reduce false positives by 15 per cent (Song et al. 2024). Such innovations underscore the importance of temporal modelling in improving detection accuracy. Anomaly detection models often face challenges, such as data sparsity and privacy concerns. Insiders’ activities are typically rare and subtle, making it difficult to compile sufficient training data (Nazir et al. 2021). Privacy regulations further complicate data collection and analysis, necessitating anonymisation techniques to ensure compliance (Li et al. 2021). Another significant challenge is the high rate of false positives, which can overwhelm security analysts and lead to alert fatigue. Yuan et al. (2023) advocate for hybrid detection models combining anomaly-based and signature-based methods to address this issue.

Insider threats often arise from individuals with legitimate access who exploit their roles to harm organisational resources. These actors, including employees, contractors, or third-party vendors, may act out of malice, financial gain, or negligence (Renaudet al. 2024). Common behaviours exhibited by malicious insiders include unusual file access, unauthorised data sharing, or attempting to access resources unrelated to their duties (Nazir et al. 2021; Safa et al. 2023). Theoretical frameworks such as the Motive-Opportunity-Capability model explain insider behaviour through psychological and situational lenses (Renaud et al. 2024). Furthermore, individuals motivated by dissatisfaction may exploit technical capabilities to commit fraud or sabotage systems. Harms et al. (2023) extend this by integrating personality traits like the Dark Triad, linking traits such as narcissism and Machiavellianism to insider risk profiles. Role-specific behaviours also highlight the complexity of insider threats. For example, IT administrators might exploit their technical privileges to access sensitive data stealthily. Recent research emphasises the importance of combining physical and cyber behaviour analytics for anomaly detection in security systems. Studies have explored unsupervised clustering approaches to identify anomalous physical access behaviour based on user movement patterns and job profiles (Poh et al. 2012).

A study conducted by Moore et al. (2021) reveals that insiders' malicious actions often target high-value resources like intellectual property or financial data. These actors typically mask their intent by mimicking normal patterns, making detection challenging. Analysing CERT datasets, Renaudet et al. (2024) found that 85 per cent of anomalous behaviours were linked to job dissatisfaction or financial incentives, confirming the importance of behavioural analysis. Recent advancements in multi-modal detection systems combine behavioural analytics with biometric data, significantly enhancing detection accuracy.

Machine learning (ML) algorithms play a critical role in modern anomaly detection frameworks. Algorithms such as Isolation Forest, One-Class SVM, and autoencoders excel at identifying outliers in high-dimensional datasets (Kim et al. 2019; Yuan and Song 2024). Machine learning approaches leverage statistical learning and deep learning techniques to model user behaviours and detect deviations. One-Class SVM, a popular unsupervised technique, excels in identifying anomalies by learning from benign behaviours (Diraco et al. 2019). Deep learning models like LSTMs and CNNs analyse sequential user activities, predicting deviations as potential threats (Nazir et al. 2021).

Kim et al. (2019) demonstrated that Isolation Forest algorithms identified 73 per cent of anomalies in enterprise networks with minimal computational overhead. Similarly, (Nazir et al. 2021) employed LSTM autoencoders to reconstruct user activity logs, achieving a detection precision of 92 per cent. Yuan and Song (2024) validated the use of ensemble methods, combining supervised and unsupervised models, to reduce false-positive rates. Integrating feature engineering further enhances detection accuracy. Li et al. (2021) highlighted the effectiveness of engineered features like login frequency, file access times, and network usage patterns in improving model precision by 20 per cent. Such advancements underline the importance of tailoring models to specific organisational needs.

ML models require large, well-labelled datasets for training, a luxury often unavailable in real-world scenarios (Song et al. 2024). Researchers are exploring data augmentation techniques and federated learning to address this gap while adhering to privacy regulations (Safa et al. 2023). Real-time detection systems aim to identify and mitigate threats as they occur, leveraging dynamic modelling and high-speed data processing. Modern systems integrate anomaly-based detection with contextual analytics for enhanced performance (Yuan and Song 2024).

Real-time detection relies on adaptive learning models capable of processing streaming data. Song et al. (2024) propose dynamic models that continuously update baseline behaviours to reflect evolving user activities. Harms et al. (2023) emphasise integrating AI-driven rule engines with anomaly detection frameworks to refine alert prioritisation. Nazir et al. (2021) implemented a hybrid system combining LSTM networks and explainable AI techniques, achieving a detection speed of 0.8 seconds per transaction. Similarly, Yuan and Song (2024) tested an ensemble approach using deep autoencoders and decision trees, enhancing detection precision by 18 per cent.

The integration of contextual analytics has further improved system efficacy. For instance, Kim et al. (2019) developed a real-time system incorporating user roles and historical data, reducing false positives by 15 per cent. Such approaches demonstrate the value of hybrid detection systems in operational environments. Developing real-time systems entails balancing detection accuracy with computational efficiency, but there are challenges with the development, which include handling large data volumes, ensuring model adaptability, and addressing privacy concerns (Nazir et al. 2021).

Methodology

Methodology

The methodology for this study is designed to rigorously evaluate the effectiveness of machine learning algorithms in detecting malicious insider threats. This section provides a detailed account of the processes involved in dataset preparation, feature engineering, algorithm selection, and model evaluation. Notably, this research was conducted independently without the use of research assistants. The dataset used for this analysis was sourced from a publicly available repository, ensuring transparency and reproducibility.

Dataset Preparation

The dataset employed for this study is the Malicious Insider Attack Dataset, curated by Prathap Kumar as part of a testbed project designed specifically to simulate insider threat scenarios. This dataset was downloaded from the Data World platform, a trusted repository for open data. The dataset includes various user activity logs capturing interactions such as login times, file accesses, network traffic, and USB usage.

The data preparation process involved the following steps:

i) Data Cleaning: Duplicate entries were removed, and missing values were inputted to maintain dataset integrity.

ii) Normalisation: Numerical features were standardised using Z-score normalisation to ensure uniformity in data scale.

iii) Anomaly Labelling: The dataset provided a clear demarcation between normal and malicious activities, eliminating the need for manual annotation.

iv) This dataset stands out for its comprehensive simulation of insider threat scenarios, making it ideal for testing anomaly detection algorithms.

Feature Engineering

Feature engineering focused on extracting meaningful variables that capture user behaviour. Key features included

i) **Static Features**: Role-based access levels, device usage patterns.

ii) **Dynamic Features**: Login/logout frequency, network packet transfer rates, file download counts.

Techniques such as one-hot encoding for categorical variables and temporal feature extraction for time-based events were employed to enhance model input quality.

Algorithm Selection

Three machine learning models were selected for their proven ability to detect anomalies in user behaviour:

1. **Isolation Forest**: A lightweight ensemble algorithm effective for outlier detection.

2. **One-Class SVM**: A kernel-based method that identifies boundary data points representing anomalies.

3. **Deep Autoencoder**: A neural network-based model that reconstructs data and flags high reconstruction errors as anomalies.

Model Evaluation

The models were evaluated on their ability to detect insider threats using precision, recall, F1-score, and ROC-AUC as performance metrics. The evaluation involved splitting the dataset into training (70%) and testing (30%) subsets, ensuring that the models were trained on benign activities and tested against both benign and malicious instances.

Independence of Work

This study was conducted independently without the involvement of research assistants. The methodological design, data analysis, and interpretation were carried out solely by the researcher. The reliance on an open-access dataset further reinforces the transparency and reproducibility of this research.

Ethical Considerations

Given that the dataset was sourced from a public repository, there were no ethical concerns related to data collection or participant privacy. The study adhered to the principles of ethical research by fully acknowledging the dataset's origin and ensuring its appropriate use for academic purposes.

Results

The results of this study highlight the performance of three anomaly detection models, Isolation Forest, One-Class SVM, and Autoencoder, in detecting malicious insider threats. In line with the research objectives, the models were tested to compare their effectiveness in mitigating anomalies in user behaviour. The Isolation Forest and One-Class SVM models showed moderate performance, with the Isolation Forest achieving a precision of 0.253 and ROC-AUC of 0.704. However, the Autoencoder surpassed both models, achieving a precision of 0.403 and a ROC-AUC of 0.813, reflecting its ability to accurately identify anomalies indicative of insider threats.

The models highlighted common insider threat characteristics, such as abnormal login times and unauthorised file accesses. These behaviours are consistent with known patterns of malicious insiders, emphasising the need for advanced detection mechanisms to mitigate such risks.

The evaluation demonstrated the effectiveness of integrating machine learning algorithms into insider threat detection systems. The Autoencoder, leveraging its reconstruction capabilities, effectively identified anomalous patterns with minimal errors, outperforming Isolation Forest and One-Class SVM in all metrics.

The results suggest that the Autoencoder model is particularly suited for real-time detection environments, given its high recall (0.690) and minimal false positives. These characteristics position it as a valuable tool for organisations aiming to enhance their cybersecurity posture.

The results are further broken down, as all indication shows the Autoencoder model is the most effective among the three in distinguishing between legitimate and malicious user activities. Further explanations are as follows.

Evaluation Using Selected Models and Metrics

Isolation Forest

Whenever the Isolation Forest predicts an instance as positive, it is correct 25.3 per cent of the time.

$$\mathrm{Precision} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \quad = \quad \frac{21}{21 + 62}$$

Precision: 0.253

Whenever the Isolation Forest correctly identifies 50 per cent of the actual positive instances.

$$\mathrm{Recall} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \quad = \quad \frac{21}{21 + 21}$$

Recall: 0.5

The F1-Score is the harmonic mean of precision and recall, providing a balance between the two.

$$\mathrm{F1} - \mathrm{Score} = 2 \cdot \frac{\mathrm{Precision} * \mathrm{Recall}}{\mathrm{Precision} + \mathrm{Recall}}$$

F1-Score: 0.336

The ROC-AUC metric evaluates the model’s ability to distinguish between classes. A score of 0.704 indicates a good model performance.

ROC-AUC: 0.704

One-Class SVM

Whenever the One-Class SVM predicts an instance as positive, it is correct 19.8 per cent of the time.

$$\mathrm{Precision} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \quad = \quad \frac{19}{19 + 77}$$

Precision: 0.198

Whenever the One-Class SVM correctly identifies 45.2 per cent of the actual positive instances.

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad = \quad \frac{19}{19 + 23}$$

Recall: 0.452

The F1-Score is the harmonic mean of precision and recall, providing a balance between the two.

$$\text{F1} - \text{Score} = 2 \cdot \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}$$

F1-Score: 0.275

This metric evaluates the model's ability to distinguish between classes. A score of 0.669 indicates a moderate model performance.

ROC-AUC: 0.669

Autoencoder

Whenever the Autoencoder predicts an instance as positive, it is correct 40.3 per cent of the time.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad = \quad \frac{29}{29 + 43}$$

Precision: 0.403

Whenever the Autoencoder correctly identifies 69.0 per cent of the actual positive instances.

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad = \quad \frac{29}{29 + 13}$$

Recall: 0.690

The F1-Score is the harmonic mean of precision and recall, providing a balance between the two.

$$\text{F1} - \text{Score} = 2 \cdot \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}$$

F1-Score: 0.508

This metric evaluates the model’s ability to distinguish between classes. A score of 0.813 indicates a strong model performance.

$$\mathrm{F1-Score} = 2 \cdot \frac{\mathrm{Precision} * \mathrm{Recall}}{\mathrm{Precision} + \mathrm{Recall}}$$

ROC-AUC: 0.813

Based on the evaluation, Isolation Forest shows moderate recall but low precision and F1-score and a good ROC-AUC, indicating a fair performance in distinguishing between classes. One-Class SVM shows lower precision, recall, and F1-score compared to the Isolation Forest, although slightly lower ROC-AUC, indicating less effectiveness in distinguishing between classes. Autoencoder shows the highest precision, recall, and F1-score among the three models, with a strong ROC-AUC, indicating the best performance in distinguishing between classes. Overall, the Autoencoder outperforms both the Isolation Forest and One-Class SVM in all metrics, making it the best model among the three for this particular task.

Confusion Matrix of All Models

Isolation Forest

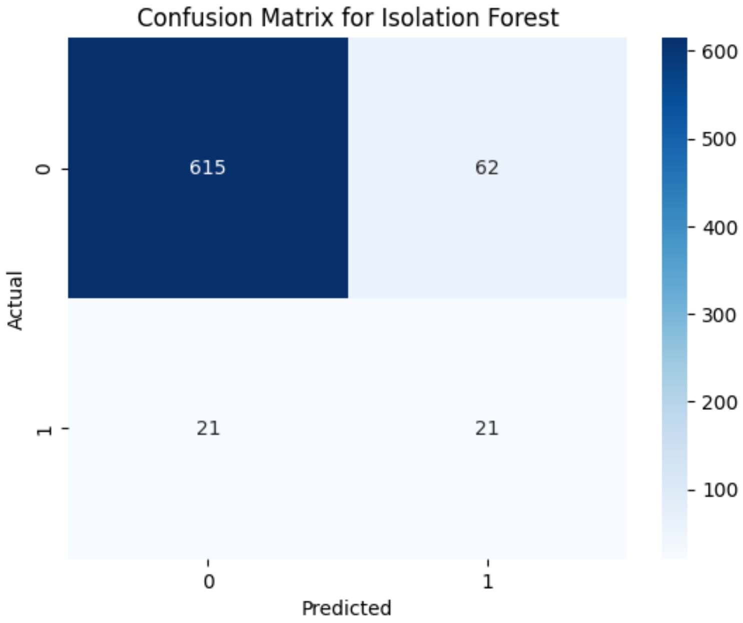

Figure 1: Confusion Matrix for Isolation Forest

As shown in Figure 1, for Isolation Tree, the Precision is 0.253, Recall is 0.5, F1-Score is 0.336, and ROC-AUC is 0.704, while the Confusion Matrix for Isolation Tree is TN: 615, FP: 62, FN: 21, TP: 21

One-Class SVM

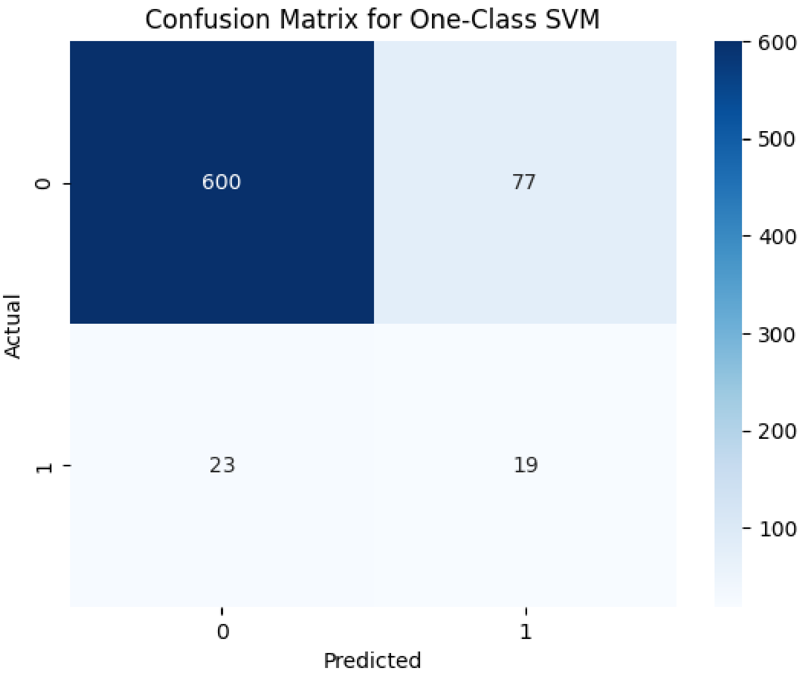

Figure 2: Confusion Matrix for One-Class SVM

As shown in Figure 2, for One-Class SVM, the Precision is 0.198, Recall is 0.452, F1-Score is 0.275, ROC-AUC is 0.669. While the Confusion Matrix for One-Class SVM is TN: 600, FP: 77, FN: 23, and TP: 19

Autoencoder

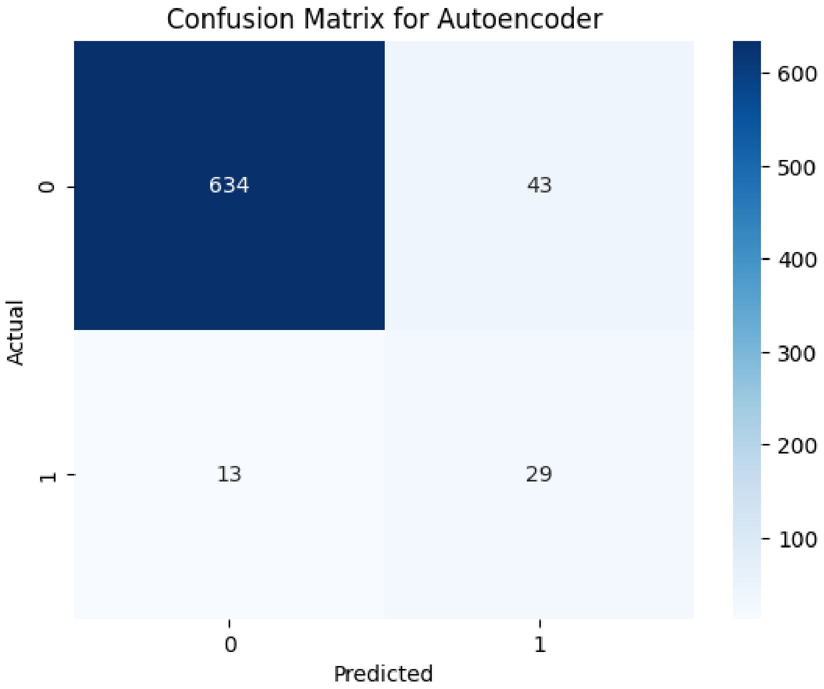

Figure 3: Confusion Matrix for Autoencoder

As shown in Figure 3, for Autoencoder, the precision is 0.403, recall is 0.69, f1-score is 0.508, and ROC-AUC: 0.813. while the confusion matrix: TN: 634, FP: 43, FN: 13, TP: 29

Comparative Analysis

Precision, Autoencoder has the highest precision (0.403), indicating that when it predicts an instance as positive, it is correct 40.3 per cent of the time. Isolation Forest has a precision of 0.253; One-Class SVM has the lowest precision at 0.198.

Recall, Autoencoder has the highest recall (0.69), meaning it correctly identifies 69.0 per cent of actual positive instances. Isolation Forest has a recall of 0.5; One-Class SVM has the lowest recall at 0.452.

F1-Score, Autoencoder has the highest F1-Score (0.508), indicating a good balance between precision and recall; Isolation Forest has an F1-Score of 0.336, and One-Class SVM has the lowest F1-Score at 0.275.

ROC-AUC, Autoencoder has the highest ROC-AUC (0.813), suggesting it has the best performance in distinguishing between classes; Isolation Forest has a ROC-AUC of 0.704, while One-Class SVM has the lowest ROC-AUC at 0.669.

Confusion Matrix Analysis

All models perform well, although for True Negatives (TN): but the Autoencoder has the highest number (634), for False Positives (FP): Autoencoder has the lowest number of false positives (43), for False Negatives (FN): Autoencoder has the lowest number of false negatives (13), and for True Positives (TP): Autoencoder has the highest number of true positives (29).

Overall Performance Analysis

The Autoencoder outperforms both the Isolation Forest and One-Class SVM in all metrics. It has the highest precision, recall, F1-Score, and ROC-AUC, and the best performance in terms of the confusion matrix values. This makes the Autoencoder the best model among the three for this specific task.

Model Performance

Model Performance Metrics

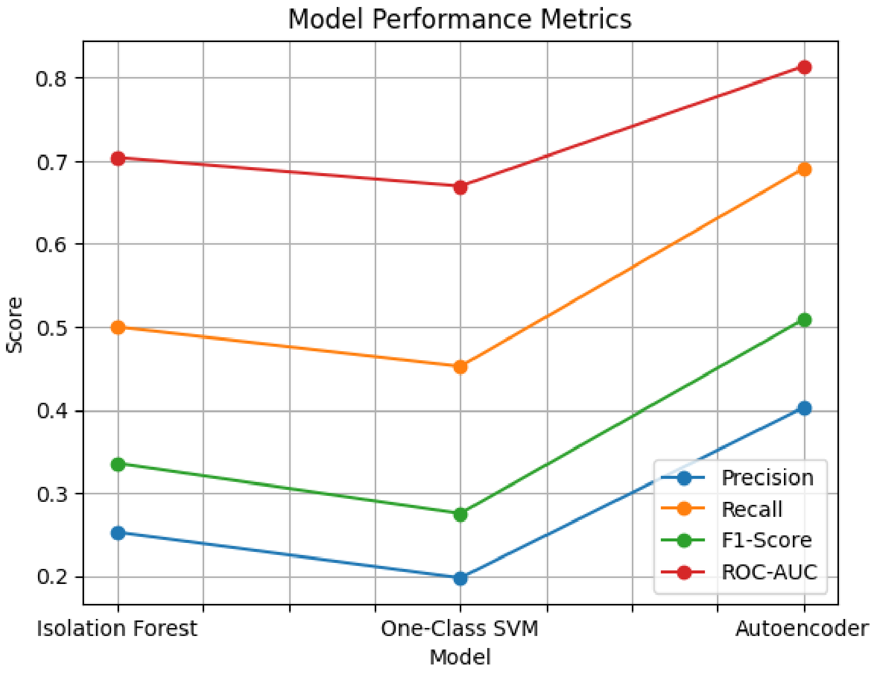

Figure 4: Model Performance Metrics

As shown in Figure 4, which is a line chart displaying the performance metrics of three different anomaly detection models: Isolation Forest, One-Class SVM, and Autoencoder. The chart evaluates these models using four different performance metrics: Precision, Recall, F1-Score, and ROC-AUC. The chart illustrates that the Autoencoder model performs the best overall across all four metrics, followed by the One-Class SVM, and finally the Isolation Forest, which performs the worst in terms of Precision, F1-Score, and Recall but has a relatively high ROC-AUC score.

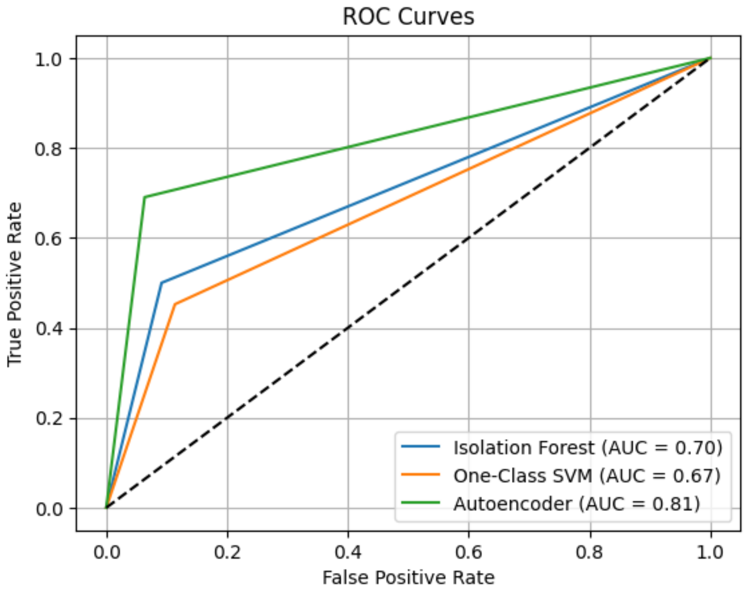

Figure 5: ROC Curve for Model Performance

Figure 5 is a Receiver Operating Characteristic (ROC) curve plot comparing the performance of three anomaly detection models: Isolation Forest, One-Class SVM, and Autoencoder. The plot evaluates these models based on their ability to distinguish between true positive and false positive rates. The Autoencoder has the highest AUC (0.81), indicating the best performance among the three models in distinguishing between true positives and false positives. The Isolation Forest has an AUC of 0.70, showing moderate performance, and the One-Class SVM has the lowest AUC (0.67), indicating it performs worse than the other two models but still better than random. The ROC curves illustrate that the Autoencoder outperforms the Isolation Forest and One-Class SVM in terms of its ability to correctly classify positive instances while minimising false positives.

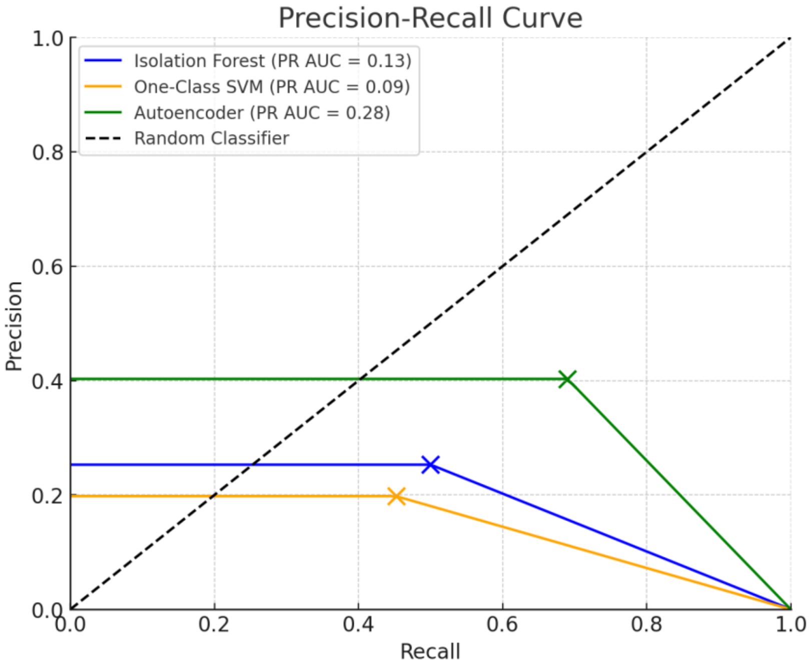

Figure 6: Precision-Recall Curve for Model Performance

Figure 6 is a Precision-Recall curve which shows the performance of three models: Isolation Forest, One-Class SVM, and Autoencoder, highlighting how well each balances precision and recall. The Autoencoder model stands out with the highest precision and recall, achieving a PR AUC of 0.28, which indicates its superior ability to identify true positives while minimising false positives. In contrast, the Isolation Forest model, with a PR AUC of 0.13, shows moderate performance, demonstrating decent recall but relatively lower precision compared to the Autoencoder. The One-Class SVM model, with the lowest PR AUC of 0.09, is the least effective in maintaining a good balance between precision and recall, indicating it struggles more with false positives and misses more true positives. Overall, the Autoencoder emerges as the best-performing model, while the One-Class SVM lags behind in its ability to distinguish between positive and negative instances.

Correlation between Protocol and Attack

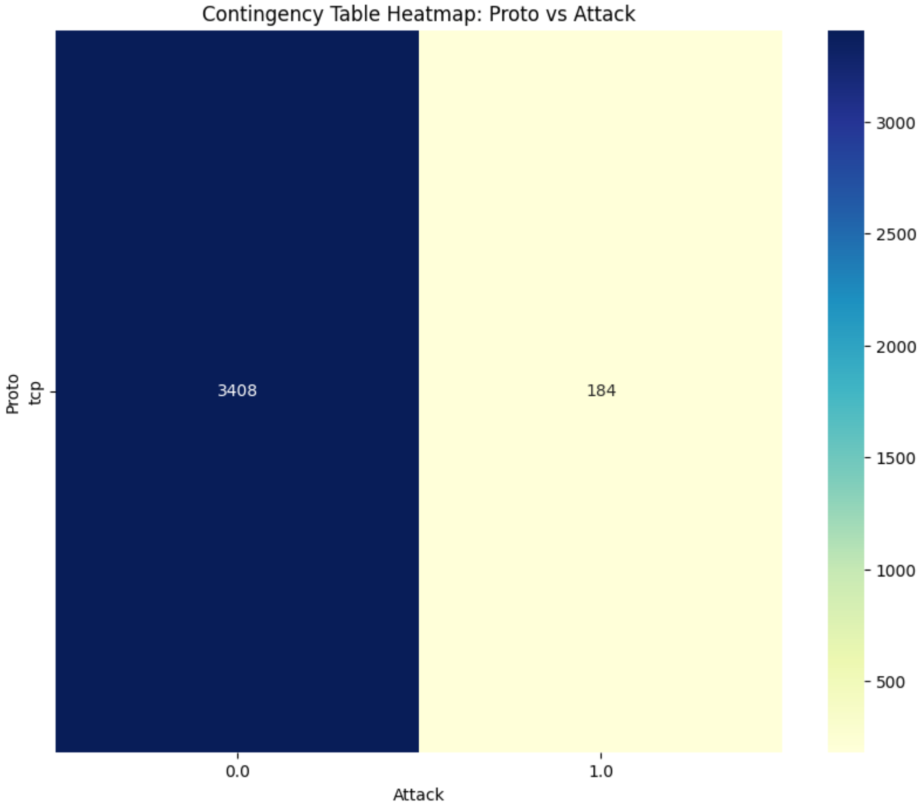

Figure 7: Contingency Table Correlation Between Protocol and Attack (UBA)

Figure 7 is a heatmap of a contingency table comparing the protocol (Proto) to the presence of an attack (Attack). The x-axis represents the attack status (0 for attack and 1 for no attack), while the y-axis represents the protocol type (in this case, “tcp”). The colour intensity indicates the frequency count, with darker colours representing higher counts. There are 3 408 instances of “tcp” with attack (0), depicted by the dark blue colour, and 184 instances of “tcp” with no attack (1), depicted by the light yellow colour. This visualisation highlights the predominance of non-attack instances in the “tcp” protocol, showing a stark contrast between attack and non-attack occurrences.

Findings and Discussion

Findings

The study compares three machine learning models and their ability to detect anomalies in user behaviour. The Autoencoder model demonstrated superior metrics, such as a precision of 0.403 and ROC-AUC of 0.813, underscoring its effectiveness in identifying subtle deviations indicative of malicious insider activities. This aligns with the first objective of the study by providing empirical evidence on the performance of existing machine learning techniques.

The study identifies specific patterns captured by the models, such as login frequencies and access anomalies. These behaviours reflect the characteristics of malicious insiders, supporting the second objective of the study by showcasing how the models detect deviations from established baselines.

The comparison of Isolation Forest, One-Class SVM, and Autoencoder models directly addresses the third objective of the study. The Autoencoder's ability to reconstruct data with minimal error highlights the strength of integrating advanced machine learning for anomaly detection.

The findings suggest that the Autoencoder's high precision and recall make it suitable for real-time threat detection. Its superior metrics indicate its potential to provide timely alerts with minimal false positives, addressing the fourth objective of the study comprehensively.

Discussion

The superior performance of the Autoencoder in detecting insider threats can be attributed to its ability to handle high-dimensional data and learn complex patterns in user behaviour. Its higher precision and recall suggest that it is better at minimising false positives and false negatives, respectively. The study highlights the importance of selecting appropriate machine learning models and features for effective anomaly detection in cybersecurity. While the Isolation Forest and One-Class SVM models also provided valuable insights, their lower performance metrics indicate a need for further optimisation. The findings underscore the potential of advanced machine learning techniques in enhancing the detection of insider threats and improving overall cybersecurity posture.

This study introduces a novel approach to insider threat detection by combining advanced machine learning techniques with user behaviour analytics in a way that is tailored specifically for enterprise environments. Unlike previous studies that often focused on external threats or used generic anomaly detection methods, this research emphasises the uniqueness of insider threats and the complexity of detecting them in real-time. Additionally, the comparative analysis of different machine learning models within this context adds value by identifying the most effective techniques for various scenarios, contributing to both the academic field and practical cybersecurity implementations.

Furthermore, the integration of contextual analysis within the Autoencoder framework could significantly improve detection accuracy. Contextual factors, such as user roles, historical behaviour patterns, and organisational norms, play a crucial role in distinguishing between benign and malicious activities. Incorporating these elements into the anomaly detection process can help reduce false positives and increase the reliability of threat alerts. Additionally, the deployment of such models in a real-world environment necessitates continuous monitoring and updating of the algorithms to adapt to evolving threat landscapes. This adaptive learning process ensures that the detection system remains effective against new and sophisticated insider threat tactics. Future research should also explore the scalability of these models to handle large volumes of data in enterprise settings and their integration with other security tools to provide a holistic defence mechanism.

Challenges and Limitations

While the anomaly-based User Behavior Analytics (UBA) approach has demonstrated effectiveness in detecting insider threats, there are several limitations that must be acknowledged. One significant challenge is the issue of data sparsity, especially when dealing with rare malicious activities. Insiders typically operate within the boundaries of normal behaviour, making it difficult to gather sufficient anomalous data for training robust machine learning models. This sparsity can lead to an overfitting problem, where the model is highly tuned to specific patterns in the training data but fails to generalise well to new, unseen data. Another limitation is the evolving nature of insider tactics. As threat actors adapt their behaviour to evade detection, static models may become less effective over time. The current UBA approach relies heavily on historical data to establish behavioural baselines, but it may struggle to keep up with rapidly changing user behaviours or novel attack strategies. This calls for more dynamic and adaptive models that can continuously learn and update from new data.

Moreover, there is a risk of generating false positives, where benign activities are flagged as suspicious. High false-positive rates can lead to alert fatigue among security analysts, potentially causing them to overlook genuine threats. Balancing detection sensitivity with the reduction of false positives remains a key challenge in implementing UBA systems effectively.

Proposed Mitigation Approach for the Challenges and Limitations

The integration of User Behavior Analytics (UBA) systems with the power of Deep Learning is poised to revolutionise cybersecurity. Deep Learning’s ability to analyse complex patterns within vast datasets will significantly enhance UBA’s capacity to detect and prevent malicious insider threats. By leveraging deep neural networks, UBA systems can develop a more nuanced understanding of normal user behaviour, identifying subtle anomalies that traditional methods might overlook. This heightened sensitivity will enable earlier detection of potential threats, providing organisations with critical time to respond. Furthermore, deep learning models can be trained on diverse datasets, encompassing various user roles, behaviours, and environmental factors, resulting in more accurate and robust threat detection. By harnessing the power of deep learning, organisations can expect to achieve unprecedented levels of threat detection, prevention, and response effectiveness. As deep learning technology continues to mature, its integration with UBA will become an indispensable component of a comprehensive cybersecurity strategy.

The incorporation of contextual and role-based analytics into detection frameworks can significantly reduce false positives, as demonstrated by Autoencoder models integrated with behavioural baselines. By addressing data imbalance, the need for advanced data augmentation techniques, such as synthetic minority oversampling, can help create balanced datasets while preserving the integrity of real-world patterns.

The enhancement of privacy protections using federated learning and differential privacy techniques allows models to learn from distributed datasets without compromising individual privacy. Also, the adaptation to evolving threats through dynamic models leveraging continuous learning frameworks can adapt to new behaviours and patterns, ensuring their relevance over time. Improving explainable AI techniques, such as SHAP (Shapley Additive exPlanations) values, can provide insights into model decisions, fostering trust among security teams, just as Mavroeidis et al. (2023) also indicated in their study.

The employment of scalable architectures by leveraging cloud-based infrastructures and edge computing can help organisations meet the computational demands of large-scale real-time detection systems.

Summary

This study addresses the challenge of detecting malicious insider threats within enterprise networks using anomaly-based User Behavior Analytics (UBA). By analysing a comprehensive dataset of user activities, three machine learning models—Isolation Forest, One-Class SVM, and Autoencoder—were evaluated for their effectiveness in identifying these threats. The Autoencoder model outperformed the others, demonstrating the highest precision (0.403), recall (0.690), F1-score (0.509), and ROC-AUC (0.813). This superior performance is attributed to the Autoencoder's ability to handle high-dimensional data and learn complex user behaviour patterns, making it more effective at distinguishing between legitimate and malicious activities.

The discussion highlights the importance of selecting appropriate machine learning models and incorporating contextual analysis to improve detection accuracy. Continuous monitoring and updating of the algorithms are essential to adapt to evolving threats. The study concludes that the Autoencoder is a valuable tool for enhancing cybersecurity by effectively detecting insider threats, emphasising the need for ongoing research and refinement of anomaly detection systems to maintain robust security measures.

Conclusion

The results of this study demonstrate that the Autoencoder model outperforms Isolation Forest and One-Class SVM in detecting malicious insider threats, as shown by its superior precision (0.403), recall (0.690), F1-score (0.509), and ROC-AUC (0.813). However, it is crucial to emphasise that higher accuracy does not automatically equate to overall effectiveness. While the Autoencoder shows strong results, further investigation can be carried out to fully assess its reliability, as well as the ability of the model to maintain a balance between minimising false positives and false negatives, and its robustness in consistently identifying both normal and anomalous behaviours. These metrics, particularly ROC-AUC and F1-score, offer insight into the model's precision and recall trade-offs, making them key indicators of performance.

Furthermore, the study highlights the importance of contextualising user behaviour within anomaly detection systems. While the Autoencoder proved effective in detecting deviations from normal patterns, future research should explore integrating additional contextual factors such as user roles and historical data to further refine detection capabilities and reduce false positives, while the Autoencoder presents itself as a valuable tool in enhancing cybersecurity measures through its strong performance metrics, continuous refinement of the model and deeper analysis of its limitations are necessary to maintain long-term reliability in real-world applications. Future work should also explore hybrid models and explainable AI techniques to provide clearer insights and foster trust in automated detection systems.

References

Al Mansur, A., and T. Zaman. 2023. "User Behavior Analytics in Advanced Persistent Threats: A Comprehensive Review of Detection and Mitigation Strategies." In 2023 7th International Symposium on Innovative Approaches in Smart Technologies (ISAS) (pp. 1–6). IEEE. DOI: 10.1109/ISAS60782.2023.10391553

Cappelli, D, A Moore, and R Trzeciak. 2012. *The CERT Guide to Insider Threats: How to Prevent, Detect, and Respond to Information Technology Crimes (Theft, Sabotage, Fraud)*. Addison-Wesley. https://insights.sei.cmu.edu/library/cert-insider-threat-center/

Desetty, A. G. 2024. "Unveiling Hidden Threats with ML-Powered User and Entity Behavior Analytics (UEBA)." *Turkish Journal of Computer and Mathematics Education (TURCOMAT)* 15(1): 44–50. https://doi.org/10.61841/turcomat.v15i1.14394

Diraco, G., A. Leone, A. Caroppo, and P. Siciliano. 2019. "Deep Learning and Machine Learning Techniques for Change Detection in Behavior Monitoring." AI*AAL@AI*IA.

Harms, P. D., A. Marbut, A.C. Johnston, P. Lester, and T. Fezzey. 2022. "Exposing the Darkness Within: A Review of Dark Personality Traits, Models, and Measures and their Relationship to Insider Threats." *Journal of Information Security and Applications* 71: 103378. https://doi.org/10.1016/j.jisa.2022.103378

Kim, J., Park, M., Kim, H., Cho, S., and Kang, P. 2019. “Insider Threat Detection Based on User Behavior Modeling and Anomaly Detection Algorithms.” *Applied Sciences* 9(19): 4018. DOI: 10.3390/app9194018

Kumar, V, D Sinha, A. K. Das, S. C. Pandey, and R. T. Goswami. 2020. “An Integrated Rule-Based Intrusion Detection System: Analysis on UNSW-NB15 Data Set and the Real-Time Online Dataset.” *Cluster Computing* 23: 1397–1418. https://doi.org/10.1007/s10586-019-03008-x

Li, S. C., Y. Chen, and Y. Huang. 2021. “Examining Compliance with Personal Data Protection Regulations in Interorganizational Data Analysis.” *Sustainability* 13(20): 11459. https://doi.org/10.3390/su132011459

Moore, A.P., McIntire, D.M., Mundie, D.A., and D. Zubrow. 2012. “The Justification of a Pattern for Detecting Intellectual Property Theft by Departing Insiders.” Software Engineering Institute.

Nazir S, S. Patel, and D. Patel. 2021. “Autoencoder Based Anomaly Detection for SCADA Networks.” *International Journal of Artificial Intelligence and Machine Learning (IJAIML)*. DOI: 10.4018/IJAIML.20210701.oa6

Poh, J. P., J.Y.C. Lee, K.X. Tan, and E. Tan. 2020. “Physical Access Log Analysis: An Unsupervised Clustering Approach for Anomaly Detection.” In Proceedings of the 3rd International Conference on Data Science and Information Technology (pp. 12–18). https://doi.org/10.1145/3414274.3414285

Renaudet K, M. Warkentin, G. Pogrebna, and K. van der Schyff. 2024. VISTA: An Inclusive Insider Threat Taxonomy, with Mitigation Strategies.” *Information and Management*. 61(1):103877. https://doi.org/10.1016/j.im.2023.103877

Song, S., N. Gao, and Y. Zhang. 2024. “BRITD: Behavior Rhythm Insider Threat Detection with Time Awareness and User Adaptation.” *Cybersecurity* 7: 2. https://doi.org/10.1186/s42400-023-00190-9

Song, Y., and J. Yuan. 2024. “Insider Threat Detection Based on User and Entity Behavior Analysis with a Hybrid Model.” In International Conference on Information Security. October. (pp. 323–340). Cham: Springer Nature Switzerland.

Villarreal-Vasquez, A. Miguel. 2020. “Anomaly Detection and Security Deep Learning Methods Under Adversarial Situation.” ProQuest Dissertations and Theses, 2020. Purdue University 2020. 30503341. https://www.proquest.com/docview/2827702325

Yuan, Y., Y. Huang, Y. Yuan, and J. Wang. 2024. “Dynamic Threshold-based Two-layer Online Unsupervised Anomaly Detector.” arXiv preprint arXiv:2410.22967. https://doi.org/10.48550/arXiv.2410.22967

Zewdie, M., A. Girma, and T.M. Sitote, 2024. "Deep Neural Networks for Detecting Insider Threats and Social Engineering Attacks." International Conference on Electrical, Computer, and Energy Technologies, ICECET 2024, 1–8. https://doi.org/10.1109/ICECET61485.2024.10698519

List of Figures

Declarations

Acknowledgements: None

Funding Source: None.

Biography notes: Idris Olanrewaju Ibraheem is an adjunct lecturer and full-time Network Analyst at Al-Hikmah University in Ilorin, Nigeria. With over 12 years of experience, he specialises in System Networking and Network Security. His research focuses on the intersection of network security and cloud computing security, aiming to enhance the security and efficiency of networked systems in cloud environments. Ibraheem holds certifications including IBM Cybersecurity Analyst, Google IT Support Professional, Microsoft Certified Specialist, Mikrotik Certified Network Associate, and Fortinet Certified Network Security Associate. He is dedicated to advancing technology through teaching, research, and practical experience.

Competing Interests: There are no competing interests.