Insider threat detection using supervised machine learning on email dataset

Priti Temgire, Suraj Kolpe,Yashashree Patil, Omkar Chakane, Prof Smita Gumaste.
(School of Computing, MIT ADT University Pune)

Abstract: Insider threats present a grave concern for organizations as individuals with authorized access can exploit privileges for malicious intent. Detecting such threats is challenging due to their familiarity with systems and evasion tactics. This study introduces a supervised machine learning method for detection, leveraging diverse data sources like user logs and network traffic. Features like login frequency and anomalies form the basis for training the model to distinguish normal behaviour from threats. Evaluation on historical data shows the model's accuracy in identifying threats while minimizing false alarms, outperforming existing methods. This underscores the value of supervised machine learning in bolstering proactive threat detection alongside traditional security measures. By harnessing advanced analytics and pattern recognition, organizations can better safeguard critical assets, ensuring operational integrity is maintained.
Keywords: - Machine learning, threat detection, Insider threat, Email dataset.

Introduction:

Insider threats are essentially those that originate from or arise within an organization from any employees. These risks may arise for a number of reasons, such as disclosing private information in an effort to profit financially. Intentional or inadvertent data breaches are possible. Conventional cyber security policies, processes, systems, and strategies frequently concentrate on external threats, which leaves the company open to internal attacks. Because insider threats are executed in part or in full by fully credentialed users—and sometimes by privileged users—it can be especially difficult to separate careless or malicious insider threat indicators or behaviours from regular user actions and behaviours. Because they are familiar with company systems, processes, procedures, policies, and users, harmful insiders have a clear edge over other types of malicious attackers [2] Insiders may additionally pose numerous sorts of threats, together with malicious sports, inadvertent moves, or exploitation by means of external attackers. According to one study, it takes security teams an average of 85 days to detect and contain an insider threat, but some insider threats have gone undetected for years [3].

The 2024 Insider Threat Report surveyed over 326 cyber security experts to expose the brand-new traits and challenges facing groups in this converting environment [3] . Key findings include: 74% of corporations say insider assaults have come to be greater common, 74% of agencies say they're at the least fairly susceptible or worse to insider threats, More than half of organizations have skilled an insider chance within the final year, and 8% have experienced greater than 20, 68% of respondents are worried or very worried approximately insider danger as their organizations return to the workplace or transition to hybrid work; most effective 3% aren't worried and 53% say detecting insider attacks is tougher inside the cloud. They are familiar with system versions and vulnerabilities. Thus, organizations should deal with internal threats at least as aggressively as they deal with external threats.

Insider threats using supervised machine learning has continued and evolved over the years. Various approaches are being used to detect insider threat as per the level of threat and its requirement to resolve it. Some of the approaches used till date are as follows: Feature based approach, Behavioural analysis, anomaly detection, Natural Language Processing (NLP) [1]. Thus, the aims of this study is to develop an intelligent machine learning model utilising the most recent advances in machine learning approaches coupled with feature engineering to detect anomalies of potential insider threats based on the CERT insider threat dataset [8].

Related work:

Insider threat detection is a well-researched topic for which many alternative approaches have been put forth. Specifically, many learning techniques have been suggested to help with early and more accurate threat detection. Using anomaly-based techniques, academics have studied insider threat identification and prevention over the past 20 years. The most often used methodology in the literature, these algorithms "learn" from normal data solely to identify anomalous cases that differ from expected examples. One fundamental premise of anomaly-based detection is that an attacker's actions deviate from typical patterns of behaviour. To be more precise, two typical actions linked to insider threats include (i) gathering massive datasets and (ii) posting materials that come from somewhere other than the organization's website. Insiders can take many different forms and pose a threat [4] .

These are classified as Malicious insiders who often have legitimate get right of entry to systems and information because of their roles inside the agency. Motivation can also include economic advantage, private vendettas, ideology, or a preference to harm the organisation. Careless insiders regularly lack focus of safety satisfactory practices or fail to observe established policies and procedures. Their actions can also inadvertently result in protection breaches, which include falling for phishing emails, leaving sensitive records unprotected, or the usage of vulnerable passwords. Compromised insider are external attackers who may additionally compromise insiders' credentials through methods which includes phishing, malware, or social engineering. Once the credentials are compromised, attackers can masquerade as valid users to access structures and records[1].

Machine learning includes both supervised and unsupervised learning approaches. Supervised learning algorithms, including Naive Bayes, Support Vector Machines (SVM), random forest, isolation forest, Linear algorithm, Decision Tree algorithm and Unsupervised algorithms, including K-Means, Expectation-Maximization (EM), Density-Based Spatial Clustering of Applications with Noise (DBSCAN).
(bin Sarhan B, Altwaijry N 2023) The study employs modern machine learning techniques, including deep learning and ensemble models, to detect insider threats. Features are derived using the Deep Feature Synthesis algorithm, resulting in extensive feature sets. Using the CERT insider threats dataset, the study achieves high accuracy rates, with anomaly detection reaching 91% accuracy and classification with SVM achieving 100%. While comparisons with other methodologies are not explicit, the study highlights the effectiveness of advanced machine learning algorithms and feature extraction processes. Strengths include the use of cutting-edge algorithms and achieving high accuracy, while challenges include data characteristics and the need for further research [4].

(Xiao J, Yang L et al) This study gives Multi-Edge Weight Relational Graph Neural Network (MEWRGNN) model for detecting insider threats in information systems. This novel approach addresses the limitations of traditional methods by leveraging graph neural network techniques to capture contextual relationships between user behaviours over time. The MEWRGNN model enhances detection accuracy, efficiency, and interpretability by extracting relational features and identifying critical edges in graphs. Evaluation results using the CERT dataset demonstrate the model's superior performance compared to baseline methods [10]. The paper contributes to the literature by proposing a preprocessing method to transform user behavior logs into a graph structure, extracting diverse user behavior features using combined graph neural networks, and improving model interpretability through edge-weight values.

(Al-Shehari T et al)The paper presents a novel insider threat detection model that utilizes anomalybased techniques, specifically the Isolation Forest (IF) algorithm, to address the challenge of imbalanced datasets in insider threat detection. By focusing on algorithm-level solutions, the model enhances detection performance and provides a more effective approach to identifying insider threats within an organization's network. Experimental results demonstrate the model's ability to handle dataset class imbalances and achieve a high accuracy score of 98%. The proposed model is compared with traditional supervised machine learning algorithms, showcasing its superiority in detecting insider threats [11]. Overall, the paper contributes to the field of insider threat detection by offering a robust and efficient solution that overcomes the limitations of imbalanced datasets.

(Mehmood M et. al)The research paper "Privilege Escalation Attack Detection and Mitigation in Cloud Using Machine Learning" by Muhammad Mehmood and team focuses on enhancing cybersecurity in cloud environments by detecting and mitigating privilege escalation attacks. The study employs machine learning algorithms such as Random Forest, LightGBM, XGBoost, and AdaBoost to classify insider attacks, utilizing features extracted from datasets like the CERT dataset [12]. Key findings include high accuracy rates, with LightGBM achieving 97% accuracy. The research highlights the effectiveness of machine learning in improving detection capabilities and emphasizes the importance of addressing insider threats in cloud security.

(Chattopadhyay P et al)The paper presents a novel scenario-based insider threat detection approach using a combination of unsupervised and supervised techniques for analysing user activities. By extracting single-day features and constructing time-series feature vectors from user activity logs, the proposed algorithm outperforms existing methods in terms of precision, recall, and f-score. The study utilizes the CMU Insider Threat Data for evaluation and achieves an average recall of 0.92 and an average f-score of 0.89. The methodology's strengths lie in its ability to capture temporal changes in user behaviour and accurately detect insider threats, while its weaknesses include the need for a large training dataset and potential limitations in highly unpredictable insider threat scenarios[13].

Machine Learning Background

In this section, we present an overview of employed classification and feature engineering methods in this work.

SVM

Support Vector Machine or SVM is one of the most popular Supervised Learning algorithms, which is used for Classification as well as Regression problems [5]. However, primarily, it is used for Classification problems in Machine Learning. The goal of the SVM algorithm is to create the best line or decision boundary that can segregate n-dimensional space into classes so that we can easily put the new data point in the correct category in the future. This best decision boundary is called a hyper plane.

Mathematical intuition of Support Vector Machine

Linear SVM: The equation for the linear hyperplane can be written as:

$$w^T x + b = 0$$

The vector W represents the normal vector to the hyperplane. i.e the direction perpendicular to the hyperplane. The parameter b in the equation represents the offset or distance of the hyperplane from the origin along the normal vector w. The distance between a data point x_i and the decision boundary can be calculated as:

$$d_i = \frac{w^T x_i + b}{||w||}$$

where ||w|| represents the Euclidean norm of the weight vector w. Euclidean norm of the normal vector W For Linear SVM classifier :

$$\hat{y} = \begin{cases} 1 & : \ w^T x + b \geq 0 \\ 0 & : \ w^T x + b < 0 \end{cases}$$

Non-Linear SVM: To separate these data points, we need to add one more dimension. For linear data, we have used two dimensions x and y, so for non-linear data, we will add a third dimension z. It can be calculated as:

$$z=x^2 +y^2$$

KNN

- K-Nearest Neighbour is one of the simplest Machine Learning algorithms based on Supervised Learning technique.
- K-NN algorithm stores all the available data and classifies a new data point based on the similarity. This means when new data appears then it can be easily classified into a well suite category by using K- NN algorithm.
- K-NN algorithm can be used for Regression as well as for Classification but mostly it is used for the Classification problems.
- K-NN is a non-parametric algorithm, which means it does not make any assumption on underlying data.

- It is also called a lazy learner algorithm because it does not learn from the training set immediately instead it stores the dataset and at the time of classification, it performs an action on the dataset.
- KNN algorithm at the training phase just stores the dataset and when it gets new data, then it classifies that data into a category that is much similar to the new data.

Mathematical intuition of KNN

Euclidean Distance Formula

$$d(p,q) = \sqrt{(p_1 - q_1)2 + (p_2 - q_2)2}$$

Proposed Methodology:

This paper proposes machine learning based approach to detect insider threat in email data set. SVM is a supervised learning algorithm that can be used for both classification and regression tasks [6]. KNN is a simple and intuitive algorithm for both classification and regression tasks. It classifies data points based on the majority vote of their neighbours.

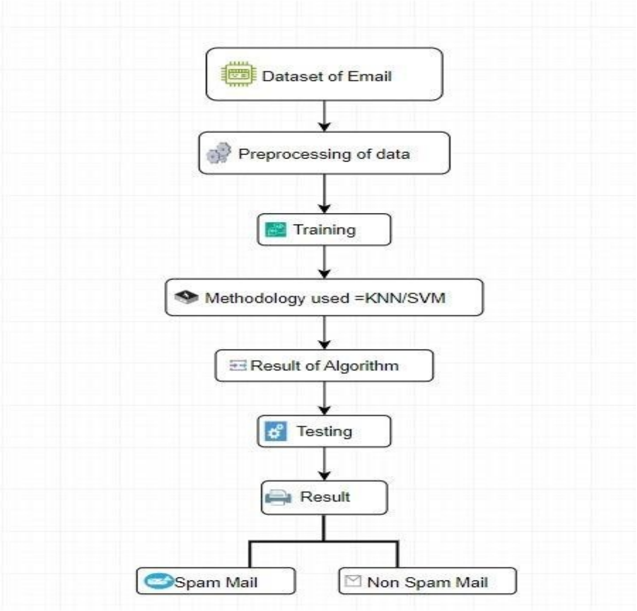

Figure 1 : Proposed methodology insider threat detection

The figure 1 shows a flow for the model to detect a spam mail and conclude a insider threat based on user ID , activity, content and attachments. The diagram also shows the results of the algorithm on a test set of emails. The algorithm correctly classified 90% of the spam emails and 85% of the non-spam emails. Data Collection: Collect e-mail records from various sources within the organisation. These statistics can also consist of e-mail headers, sender and recipient data, timestamps, e mail content material, and attachments. Feature Engineering: Extract relevant capabilities from the e-mail records that can be used to educate device getting to know fashions. Features may additionally consist of: Email metadata: Sender, recipient, timestamp, length, and so on. Email content material evaluation: Sentiment analysis, key-word extraction, etc. Labelling Data: Annotate the e-mail information with labels indicating whether or not every e mail represents a legitimate or suspicious hobby. This labelling can be performed manually by means of safety analysts or the use of automatic algorithms primarily based on predefined regulations or anomalies. Model Training: Train machine mastering models the use of categorized email information. Training the model using different algorithm like SVM, KNN [7] . Model Evaluation: Evaluate the performance of the trained models using evaluation metrics such as accuracy, precision, take into account, and F1-score. Use move-validation techniques to make sure the robustness of the models and avoid over fitting.

Dataset Used:

We utilized the "CERT Insider Threat Tools" dataset (Carnegie Mellon's Software Engineering Institute, Pittsburgh, PA, USA) since it is exceedingly difficult to obtain actual business system logs[8]. The CERT dataset is an intentionally manufactured dataset used to validate insider-threat detection systems; it is not real-world corporate data [1]. Employee computer usage logs (logon, device, http, file, and email) along with certain organizational data like employee departments and roles are included in the CERT dataset. Every table has data pertaining to the activities, timestamps, and ID of each user. Email Description is one of the log records of email activities.[5]

Table 1 : Email Activity Log Records

ID	Primary key of an observation
Date	Day/Month/Year Hour:Min:Sec
User	Primary key of a user
PC	Primary key of a PC
To	Receiver
Cc	Carbon Copy
Bcc	Blind carbon copy
From	Sender
Activity	Activity (Send/Receive)
Size	Size of an email
Attachments	Attachment file name
Content	Content of an email

Following the prediction phase, we propose to use various evaluation metrics to assess the correctness of the model results, such as accuracy, precision, recall, and F1 Score [9]. The accuracy metric is a type of evaluation statistic that evaluates how accurate a classifier is. We simply add up the samples that were correctly predicted (true positive and true negative) and divide that amount by the number of samples to determine the accuracy using the confusion matrix; see Equation (1). In our case, we will classify the data into two categories: normal and abnormal, and the accuracy metric will give the percentage of the user's activities that are classified correctly.

$$\text{Accuracy} = (TP + TN)/\ (TP + TN + FP + FN\ ) \qquad \text{Equation (1)}$$

where:

- TP: predicted abnormal activity is an abnormal activity;
- FP: predicted normal activity is an abnormal activity; • FN: predicted abnormal activity is a normal activity;
- TN: predicted normal activity is a normal activity.

We can use these elements as input to calculate additional evaluation metrics, as demonstrated in Equation (2). Precision is a measurement of precision that provides us with a measure of exactness, which determines the number of all true predictions of anomaly (or abnormal activity) on all predictions. We can obtain 100% accurate predictions if the precision value is close to 1, which indicates that FP== 0.

$$\text{Precision} = (TP)/(TP + FP) \qquad \text{Equation (2).}$$

These metrics provide insights into different aspects of the threat detection system's performance, such as its ability to correctly identify threats, its ability to avoid false alarms, and its overall effectiveness in differentiating between threats and non-threats.

Conclusion

Our study outlines a thorough approach that uses machine learning to identify insider threats from email datasets. The Support Vector Machine (SVM) and K-Nearest Neighbors (KNN) algorithms are the main emphasis of this methodology. We hope that by putting forth this methodology, we will be able to solve the urgent need for reliable and automated cyber security solutions while also advancing insider threat detection systems. Our approach offers a strong basis for further investigations and advancements in the area of insider threat identification. Moreover, practical application and testing on other email datasets will be necessary to confirm the effectiveness and scalability of our approach in various organizational settings. Through the utilization of sophisticated algorithms like SVM and KNN and the application of machine learning, companies can enhance their protection against insider attacks, protecting their assets and private data in a more dangerous digital environment.

References

[1] Mohammed Nasser Al-Mhiqani , Rabiah Ahmad , Z. Zainal Abidin 1, Warusia Yassin ,Aslinda Hassan , Karrar Hameed Abdulkareem , Nabeel Salih Ali and Zahri Yunos A Review of Insider Threat Detection: Classification,Machine Learning Techniques, Datasets, Open Challenges, and Recommendations Appl. Sci. 2020, 10, 5208; doi:10.3390/app10155208

[2] Hunker, Jeffrey, and Christian W. Probst. (2011) ”Insiders and Insider Threats-An Overview of Definitions and Mitigation Techniques.” J. Wirel. Mob. Networks Ubiquitous Comput. Dependable Appl. (2.1): 4-27.

[3] 2024 Insider Threat Report [Securonix] - Cybersecurity Insiders https://www.cybersecurityinsiders.com/portfolio/2024-insider-threat-report-securonix/

[4] Bushra Bin Sarhan , Najwa Altwaijry” Insider Threat Detection Using Machine Learning Approach “ Appl. Sci. 2023, 13, 259. https://doi.org/10.3390/app13010259

[5] NaanKang Garbaa, Sandip Rakshita, Chai Dakun Maaa, Narasimha Rao Vajjhalab “ An email content-based insider threat detection model using anomaly detection algorithms “ 4th International Conference on Innovative Computing and Communication 2020

[6] Ding, S. F., B. J. Qi, and H. Y. Tan. (2011) ”An overview on theory and algorithm of support vector machines.” Journal of University of Electronic Science and Technology of China 1 (40): 2-10.

[7] Mayhew, M.; Atighetchi, M.; Adler, A.; Greenstadt, R. Use of machine learning in big data analytics for insider threat detection. In Proceedings of the MILCOM 2015–2015 IEEE Military Communications Conference, IEEE, Tampa, FL, USA, 26–28 October 2015; pp. 915–922.

[8] Glasser, J.; Lindauer, B. Bridging the gap: A pragmatic approach to generating insider threat data. In Proceedings of the 2013 IEEE Security and Privacy Workshops, San Francisco, CA, USA, 23–24 May 2013; pp. 98–104.

[9] Duc C. Le Nur Zincir-Heywood Exploring anomalous behaviour detection and classification for insider threat identification https://onlinelibrary.wiley.com/doi/abs/10.1002/nem.2109.

[10] Xiao J, Yang L, Zhong F, Wang X, Chen H, Li D. Robust Anomaly-Based Insider Threat Detection Using Graph Neural Network. IEEE Transactions on Network and Service Management. 2023 Sep 1;20(3):3717–33.

[11] Al-Shehari T, Al-Razgan M, Alfakih T, Alsowail RA, Pandiaraj S. Insider Threat Detection Model Using Anomaly-Based Isolation Forest Algorithm. IEEE Access. 2023;11:118170–85.

[12] Mehmood M, Amin R, Muslam MMA, Xie J, Aldabbas H. Privilege Escalation Attack Detection and Mitigation in Cloud Using Machine Learning. IEEE Access. 2023;11:46561–76.

[13] Chattopadhyay P, Wang L, Tan YP. Scenario-based insider threat detection from cyber activities. IEEE Transactions on Computational Social Systems. 2018 Sep 1;5(3):660–75.