Journal of
Cyber Security, Privacy Issues and Challenges

www.matjournals.com

e-ISSN: 2583-7656
Volume-2, Issue-3 (September-December, 2023)

Machine Learning Approach Based on Synthetic Minority Over-Sampling Technique and Isolation Forest for Insider Threat Detection

Adeleke, Nafisa Damilola[1*], Umar, Suleiman Dauda[2], Ismaila Idris[3], Joseph Adebayo Ojeniyi[4]
[1] Post Graduate Student, [3]Associate Professor, [4]Senior Lecturer, Department of Cybersecurity Science, School of Information Communication Technology, Federal University of Technology, Minna, Nigeria
[2] Senior Lecturer, Department of Electrical Engineering, School of Electrical Engineering and Technology, Federal University of Technology, Minna, Nigeria

[*]Corresponding Author: adeleke.pg207787@st.futminna.edu.ng

Received Date:November 16, 2023 *Published Date:November 29, 2023*

ABSTRACT

Detecting insider threats is challenging due to insiders' deep familiarity with networks and security protocols, allowing them to bypass traditional security measures. While various methods combat insider threats, creating effective detection systems remains difficult. Research advocates using Machine Learning (ML) techniques, but handling imbalanced datasets reduces accuracy. To tackle this, this paper presents "SMOTE-IForest," merging SMOTE and IForest for insider threat detection. Testing on the CERT r6.2 dataset achieved 80.0% accuracy in detecting user behaviour. Additionally, it reached a 63.4% detection rate with a 67.0% false positive rate, boasting a high AUC of 96.0%, 93.30% precision, and 88.80% f-measure. This model addresses accuracy, detection, and false positive rate issues. SMOTE improves dataset balance by creating synthetic samples from the minority class, enhancing classification accuracy. IForest isolates anomalies, efficiently handling high-dimensional data without complex tuning, ideal for insider threat detection. The "SMOTE-IForest" model significantly strengthens insider threat detection systems by overcoming dataset imbalance and enhancing accuracy. Its precision and f-measure distinguish between normal and anomalous behaviour, aiding in addressing setbacks associated with existing studies' accuracy, detection, and false positive rates.

Keywords- Imbalance data, Insider threat detection, Isolation forest, Machine Learning (ML), Synthetic minority over-sampling technique

INTRODUCTION

A recent or previous worker, independent worker or freelancer, or business partner who presently or in the past used their access to harm a business network, computers, data, or facilities occasionally unintentionally [1] is regarded as an insider threat, in the opinion of Cybersecurity and Infrastructure Security Agency (CISA). These kinds of threats can come from within or outside the organisation. It is no more news that the global pandemic has increased the need for more employees to work from home. However, there has been an increase in the number of malicious threats and defending against this type of attack can be tedious as they are difficult to detect, prevent or predict mainly due to the main reason that the attack is from a trusted employee of the company [2].

Establishments and infrastructure are undoubtedly alerted to the risks posed by insider threats, yet it can be challenging to halt breaches caused by workers' intentional or unintentional activities. Why? To identify and stop the threat, it is difficult to keep complete visibility over data, both digital and physical, including who has access to it. Despite the efficiency of the current detection methods used to address insider threat issues, hackers can avoid using these approaches by altering their practices and behaviours. To guarantee that hostile insider danger is identified, it is essential to regulate insider threats, develop new tactics, and create new systems or enhance current ones. Various anomaly detection algorithms have been used to identify patterns and anomalies in the insider threat dataset [3-10]. The effects of insider threats may be operational, financial, or even fatal in extreme circumstances. As a result, it is