STAR
Research Journal

Available online at www.starresearchjournal.com (Star International Journal)

DATA SCIENCE

ANOMALY DETECTION IN NETWORK SECURITY

[1]S.Nishidh, [2]K.Harish,

Department of Data Science, Sri Krishna Adithya College of Arts and Science, Coimbatore.

ABSTRACT:

It is one of the most vital paradigms in network security, especially in the area of fraud abuse in financial transaction processing. An anomaly detection system does proper analysis in transaction patterns and user behaviors and can identify deviations that may signal some fraud activities. This paper reviews the importance of anomaly detection in fraud prevention, pinning the discussion on two major areas: unauthorized access to financial accounts and suspicious patterns in transactions. Apart from this, the anomalies were supposed to contribute to organizations in strengthening their security controls and being able to take early action against emerging threats by detecting unusual login attempts and atypical transaction behaviour. The results strongly asserted the implementation of a robust system for the detection of anomalies that would save financial integrity and the users from fraudulent activities.

1.INTRODUCTION

The increasing dependence on online monetary transactions has brought the challenge of fraud detection to the forefront more than ever for organizations and consumers in today's digital world. With cybercriminals evolving more innovative ways of manipulating vulnerabilities, most security measures fail at identifying or neutralizing threats. Anomaly detection has become a strong arm in this fight against fraud, at the core of which lie advanced algorithms and data analysis techniques to spot irregular patterns in user behavior and transaction data. This would raise early warnings about possible fraudulent activities by monitoring deviations from established norms, thus helping an organization take immediate remedial action to safeguard its assets and reputation.

In this context of fraud prevention, underlined critical applications in the paper focus on unauthorized access to the financial account and suspicious transaction patterns. The present study thus seeks to explore how such anomalies are identified by systems whether it is several attempts at bad logins or unusual amounts for transactions-and stresses proactive fraud detection as a means to maintain financial system integrity and confidence in such systems by users.

2.WHAT IS ANOMALY DETECTION?

Anomaly detection is a form of data analysis used for identifying data points, entities, or events that are fairly different from other data. Anything that falls outside of the norm or that varies from normal or expected standards might be called an anomaly. It is instinctive behavior, both for people and animals, to reach for a fruit on the tree once they notice that it is ready for picking, or to pay attention to the movement in the grass if somehow it is singled out from the background-noise by signaling a possibility or a danger. For this reason, sometimes this concept is referred to as either anomaly detection or unusual-event detection. This task has traditionally been one of the most laborious tasks for statisticians, who for years have painstakingly gone through charts in order to find those atypical elements. The last couple of decades have seen a gradual change in that direction, as automation of the process has started, and various machine learning-based techniques are proposed by different researchers to find different types of outliers much faster. Advanced training methods nowadays are being used for more effective and time-efficient detection in various fields.

Anomaly detection is typically done to find instances that are unusual occurrences, unexpected opportunity or corrupted information in time-dependent datasets. These uncommon happenings may symbolize hacker attacks, fraud activities, criminal actions, health condition, and equipment failure. Unexpected opportunity may include finding a store, product, or salesperson that is far better than everybody else- it merits an investigation to learn ways to improve the business.

Other causes of anomalies include faulty machinery, malfunctioning sensors, or even network disruptions. In this instance, the data scientist might want to remove such anomalous records from further analyses so as not to compromise the creation of new algorithms.

3.WHAT IS ANOMALY?

An anomaly in anomaly detection is a data point that stands out from the rest of the data in a set and doesn't follow the

normal pattern. Anomalies can also be called outliers, standard deviations, noise, novelties, or exceptions.

4.TECHNIQUES

It is a very important process in detecting those patterns or data points that are significantly different from the rest. Starting with sophisticated machine learning and deep learning techniques, finishing with more conventional statistical methods, various approaches have been developed for solution. The key content of all major anomaly detection methods will be summarized in this section.

4.1. Statistical Methods

One of the earliest approaches to anomaly detection includes statistical methods. These depend on mathematical models and pre-set norms to identify outliers in data. Some common statistical methods that are regularly used include:

Z-Score: The Z-score of a data point reflects how many standard deviations it is from the mean. If the Z-score of a given data point is greater than some threshold, like ±3, it is considered an anomaly.

$$Q1-1.5\times IQR \text{ and } Q3+1.5\times IQR$$

Interquartile Range (IQR): In this approach, one picks the first quartile, Q1, and the third quartile, Q3, to estimate a normal range. Every value that falls outside the range defined by

Grubbs Test: The Grubbs test is designed to find outliers in a one-dimensional data set under the normal distribution assumption. It takes the farthest data point and tests it against all the rest to identify whether it's an outlier.

The statistical methods are simple and computational grease; however, they may run into complications when applied to complex datasets that do not totally meet the criteria set by basic statistical assumptions.

4.2. Machine Learning Approaches

That may be the reason why machine learning techniques have become the most popular in anomaly detection, as they can learn from data and adapt to various types of complex patterns. Again, these may be divided into another two categories of learned by supervised and those by unsupervised learning.

4.2.1Supervised Learning:

Supervised learning algorithms require labelled data; instances should represent normal as well as anomalous data. Some of the popular anomaly detection algorithms, which come under supervised learning, are mentioned below:

Support Vector Machines:

A support vector machine in general is a classifier that, in its most basic form, finds the hyperplane that separates normal and abnormal data points in a high-dimensional space; data points falling outside of the boundary are considered to be anomalies.

Isolation Forest:

This will be done via an algorithm that randomly selects a feature and then a split value in order to isolate anomalies. The intuition is that anomalies will be more easily isolated, resulting in shorter paths in the tree structure.

4.2.2. Unsupervised Learning:

Unsupervised Anomaly Detection Techniques of unsupervised anomaly detection do not require labeled data and are mainly used when the anomalies are very rare. Default, some common

unsupervised anomaly detection algorithms are:

K-Nearest Neighbors:

KNN detects anomaly with respect to the distance between a point in a data set and its k-nearest neighbors. Those points that have significantly larger distances from their k-nearest neighbors are labeled as anomalies.

Local Outlier Factor:

The LOF is the estimation of any given information point approximately its neighborhood thickness versus that of its neighbours. The points that have comparatively lower density than their neighbors are anomalies.

4.3. Deep Learning

Deep learning techniques, especially neural networks have great potential in finding intricate patterns in large data. These techniques will also learn the hierarchical feature representation model and are appropriate to find the subtle anomalies.

Convolutional Neural Networks:

Although one-dimensional CNNs take image data as the main input, they also prove useful when applied to time-series data. They learn hierarchical feature representations and make them suitable for anomaly detection of subtle nature.

Recurrent Neural Networks:

RNNs are suitable for sequential data. In this type of network, to detect anomalies in time-series data, temporal dependencies are modeled.

4.4. Hybrid Approaches

Hybrid approaches seek to combine statistical techniques with machine learning methods for the best possible solution. Thus, for instance: Obviously, statistical pre-processing may include filtering anomalies by statistical methods to enhance machine learning techniques for better overall performance of the models.

Ensemble Methods:

The combination of different anomaly detection algorithms such as clustering and classification can provide better detection performance, reducing false positives and false negatives.

Generally speaking, anomaly detection techniques vary right from the conventional statistical methods to modern machine learning and deep learning methods. Application, nature of data, and availability of labeled instances normally influence their choice. Basically, mastering these techniques will be very helpful for the realization of efficient detection systems in these and other domains such as network security, finance, and healthcare.

5.APPLICATIONS

Anomaly detection is important in network security, in that it helps in the identification of unusual patterns or behaviors that would have otherwise signified any imminent threat to security. Given the increasing organization dependence on interconnected networks, the development of effective systems of anomaly detection becomes of utmost priority. The subsequent segment discusses with details various applications of anomaly detection in network security, including the importance of the processes, techniques involved, and the challenges developed accordingly.

5.1. IDS-Intrusion Detection Systems

Anomaly detection represents a core building block for IDS. IDS monitor network traffic for activities that cannot be

considered normal compared to some set baselines of normal behavior. Generally, speaking, using IDS, there are two major types:

Network-based IDS (NIDS):

This system is responsible for monitoring network flow for abnormal activities based on predefined rules or heuristics from data. It often uses statistical analysis and machine learning algorithms, including Support Vector Machines and Isolation Forests, which identify unusual patterns that might be indicative of unauthorized access attempts or malicious activities.

Host-based IDS (HIDS):

It will be more about monitoring individual devices or hosts for suspect activities. Anomaly detection techniques may inspect system logs, user behavior, and application activities for deviations that may indicate possible breaches.

5.2. Fraud Detection:

Anomaly detection also plays an important role in the network security domain for its huge value in fraud prevention, including fraud in financial transaction processing. The analysis of transaction patterns and user behavior enabled by anomaly detection systems may reveal irregularities that can indicate fraud activities, such as:

Unauthorized access to financial accounts:

This could be an anomaly in the login pattern, such as multiple failed attempts or a login from some strange location, that would serve to trigger an alert in cases of a potential account compromise.

Suspicious transaction patterns:

Unusual amounts, frequencies, or geographical locations of transactions may have the presence of fraud in them. For instance, an unusual increase in transactions coming from one account may point toward account takeovers or card fraud.

5.3. Real-time Monitoring:

Anomaly detection systems are of paramount importance when it comes to real-time network activity monitoring. The system detects anomalies at the time they occur through continuous analysis of network traffic and user behavior, thus making it possible for organizations to respond swiftly against any potential danger. The proactive approach allows:

Immediate incident response:

Anomalies, when detected, can thus automatically trigger notifications that allow security teams to investigate threats and perform mitigative actions in near real time.

Reduced dwell time:

The quicker it takes an organization to identify and contain an anomaly, the less time it gives attackers to exploit any vulnerability, hence helping to decrease the potential impact of security incidents.

5.4. Behavioral Analysis:

It builds up the baseline of normal user behavior within an organization. Continued monitoring of user activities-like times of login, resources accessed, and pattern of data transfer-may indicate deviations from established norms. These applications are useful for:

Insider Threat Detection:

It may also reveal other forms of insider threats, like an employee handling

data that he or she wouldn't normally use or even trying to exfiltrate data.

Compromised accounts:

These are usually detected by unusual patterns of access or data usage that flag an account for immediate investigation and remediation.

5.5. Network Traffic Analysis:

Specifically, it is believed that the analysis of network traffic for anomalies is indispensable for the assurance of integrity and reliability of the enterprise network. There are some methods which have the ability to identify:

DoS attacks:

These are sudden spikes or unusual patterns in traffic that may indicate an ongoing DoS attack, thus enabling organizations to take necessary actions to reduce its impact.

Malware communication:

Anomalous outbound traffic patterns could indicate that infected devices attempt to call back to command and control servers, enabling the detection of malware infection at an early stage.

6. CHALLENGES:

6.1. Overfitting

Definition:

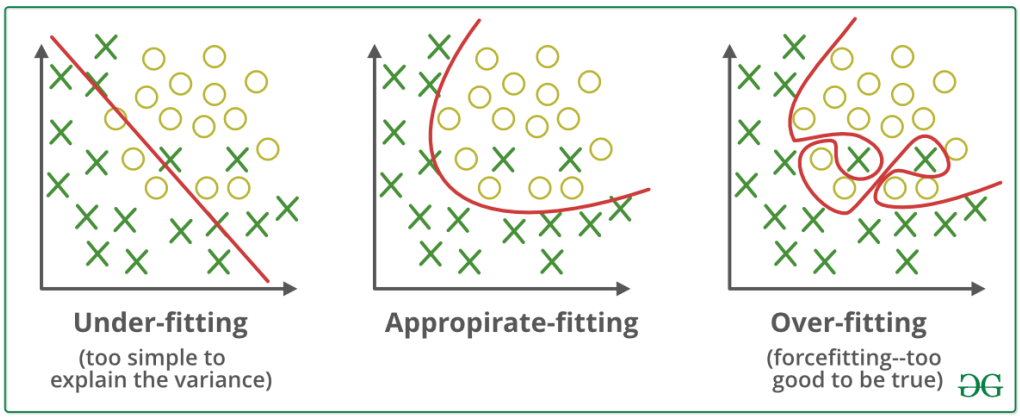
Overfitting is when a model of detection becomes overly complex and starts memorizing the noise of the training data instead of learning the pattern. This results in models which perform well on the training data but completely fail on unseen real-world data.

Impact:

This might result in too many false negatives-overfitting causes the model not being able to generalize from its training, which misses the true anomalies. This might be very critical in dynamic environments where types of new attacks might not be included in the training set.

Mitigation Strategies:

Precious model design and choice of validation, techniques like cross-validation, and methods of regularization will help avoid overfitting. Using simpler models or incorporating ensemble methods will greatly help in improving generalization.



6.2. Scalability:

Definition: All anomaly detection systems must be able to take into consideration the volumes of data generated by ever-increasing networks.

Impact: Considering this, with growing network traffic, it turns to be a bottleneck,

because real-time analysis requires huge computation capability. Inefficient algorithms can make them incapable of processing data in time for the detection of anomalies, thus responding late to any potential threat.

Mitigation Strategies: Ensuring that the algorithms are tuned so that they scale in terms of performance and resource utilization. Large data could be handled with techniques such as distributed computing, data sampling, and dimensionality reduction.

6.3. Irregular Traffic Patterns:

Definition: Networks are a carrier of highly variable traffic, making it hard to establish any consistent baseline of normal behaviour.

Impact: Strange traffic, if set up inaccurately, may result in missed anomalies or false positives that the detection system could have caught. For example, regular spikes during peak hours may be misclassified as anomalies, while subtle attacks might get passed undetected.

Mitigation Strategies: Some of the strategies to handle this challenge include continuous learning and adaptive thresholding. It enables the system to learn from oncoming data and dynamically adjust the baseline. It can thus achieve a better discrimination between true anomalies versus normal fluctuations.

6.4. Evolving Threat Landscape:

Definition: The threat landscape changes from day to day, as every single moment, cyber attackers work on new methods of breaching into networks.

Impact: Anomaly detection systems have to adapt and learn new attack patterns. Traditional rule-based systems usually cannot keep pace with the evolution of threats, hence always creating security gaps.

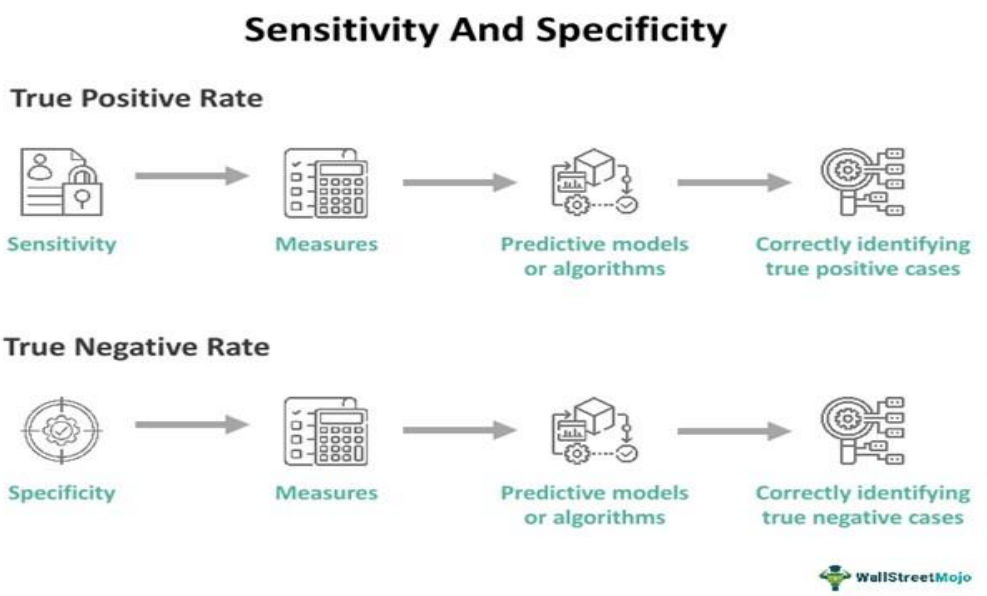
Mitigation Strategies: Meanwhile, the anomaly detection systems have to make use of machine learning techniques that will allow them to continuously update themselves and learn from incoming data. In addition, threat intelligence feeds are very useful for providing context and insight into emerging threats.

6.5. Balancing Sensitivity:

Definition: In anomaly detection, the right sensitivity balance is a must to identify a threat.

Impact: These systems need to be sensitive enough to identify the real threats without reporting too many false alarms. A high rate of false positives can lead to alert fatigue among security analysts, who may soon begin to ignore genuine threats.

Mitigation Strategies: Detection algorithms should be fine-tuned, with the implementation of mechanisms for feedback that will go a long way in adjusting the sensitivity of the system based on its performance. Additionally, the use of layered security approaches will imply that multiple modes of detection are performed in conjunction with one another to increase overall detection capability.



6.6. Data Quality and Imbalance:

Definition: The quality of the data that anomaly detection models are being trained on and tested with is by far the most important factor in their performance. Another challenge also is that, usually, the datasets are imbalanced with the number of normal instances far outnumbering anomalies.

Impact: Poor data quality may result in mistakenly trained models, whereas imbalanced datasets tend to yield models biased toward the majority class, hence detecting only a few anomalies.

Mitigation Strategies: Data cleaning and preprocessing are very important and are done quite frequently for the assurance of high-quality data. This will help in overcoming class imbalance using techniques like oversampling and under sampling, or generating synthetic data.

7.CONCLUSION:

Anomaly detection is part and parcel of any fraud prevention strategy in the modern day within the financial industry. This will help in segregating anomalies from the regular pattern of transactions and user behavior-which could indicate fraudulent activities like unauthorized access to an account and suspicious transaction behavior. This feature of real-time detection of anomalies within transactions allows organizations to take appropriate action on time, against any suspicious activity for reducing the threat of loss and reputational risk. Anomaly detection systems have to be robust, considering that fraud tactics change day in and day out, since fraudulent activities threaten financial integrity and consumer protection. Additional research and development of anomaly detection techniques would further enhance these systems to stay ahead of the cybercriminals at this point in the cat-and-mouse game of fraud.

REFERENCES:

1) Sarasamma, S. T., Zhu, Q. A., & Huff, J. (2005). Hierarchical Kohonenen Net for Anomaly Detection in Network Security. IEEE Transactions on Systems Man and Cybernetics Part B (Cybernetics), 35(2), 302–312. https://doi.org/10.1109/tsmcb.2005.843274

2) Sarasamma, S. T., Zhu, Q. A., & Huff, J. (2005). Hierarchical Kohonenen Net for Anomaly Detection in Network Security. IEEE Transactions on Systems Man and Cybernetics Part B (Cybernetics), 35(2), 302–312. https://doi.org/10.1109/tsmcb.2005.843274
3) Sarasamma, Suseela T., et al. "Hierarchical Kohonenen Net for Anomaly Detection in Network Security." IEEE Transactions on Systems Man and Cybernetics Part B (Cybernetics), vol. 35, no. 2, Apr. 2005, pp. 302–12. https://doi.org/10.1109/tsmcb.2005.843274.
4) Sarasamma, Suseela T., et al. "Hierarchical Kohonenen Net for Anomaly Detection in Network Security." IEEE Transactions on Systems Man and Cybernetics Part B (Cybernetics), vol. 35, no. 2, Apr. 2005, pp. 302–12, https://doi:10.1109/tsmcb.2005.843274.
5) Sarasamma, Suseela T, Qiuming A Zhu, and Julie Huff. "Hierarchical Kohonenen Net for Anomaly Detection in Network Security." IEEE Transactions on Systems Man and Cybernetics Part B (Cybernetics) 35, no. 2 (April 1, 2005): 302–12. https://doi.org/10.1109/tsmcb.2005.843274.
6) "Hierarchical Kohonenen Net for Anomaly Detection in Network Security." IEEE Transactions on Systems Man and Cybernetics Part B (Cybernetics) 35, no. 2 (April 1, 2005): 302–12. https://doi.org/10.1109/tsmcb.2005.843274
7) VANERIO, J. and CASAS, P. (2017) Ensemble-learning Approaches for Network Security and Anomaly Detection. [Online] Available from: doi.org/10.1145/3098593.3098594.
8) Juan Vanerio and Pedro Casas. 2017. Ensemble-learning Approaches for Network Security and Anomaly Detection.
9) Sebyala AA, Olukemi T, Sacks L. Active Platform Security through Intrusion Detection Using Naïve Bayesian Network for Anomaly Detection. 2002. Available at: http://www.ee.ucl.ac.uk/lcs/previous/LCS2002/LCS116.pdf?origin=publication_det
10) Kosek AM. Contextual anomaly detection for cyber-physical security in Smart Grids based on an artificial neural network model. 2016. Available at: https://doi.org/10.1109/cpsrsg.2016.7684103.