International Journal of Management, IT & Engineering
Vol.14 Issue 10, October 2024,
ISSN: 2249-0558 Impact Factor: 7.119
Journal Homepage: http://www.ijmra.us, Email: editorijmie@gmail.com
Double-Blind Peer Reviewed Refereed Open Access International Journal - Included in the International Serial Directories Indexed & Listed at: Ulrich's Periodicals Directory ©, U.S.A., Open J-Gate as well as in Cabell's Directories of Publishing Opportunities, U.S.A

Enhancing System Observability with Machine Learning Techniques for Anomaly Detection

Sanjay Bauskar
Senior Database Administrator, Pharmavite LLC
West Lafayette, IN, USA.
sanjaybauskar@gmail.com

***Abstract*—Anomaly detection is an important process in many domains like industrial control networks, database systems, system level radiation testing, where the detection of abnormal behaviour indicates a problem such as system failure, intrusion or degradation in performance. With increasing levels of complexity of modern databases, it has become increasingly important to ensure that the system can be observed. The purpose of this paper is to review the application of ML in the context of anomaly detection within the context of database systems and analyze its capacity to improve system observability. Anomaly detection is an important process that helps to detect the abnormally behaving instances that signify performance degradation, security intrusion or data corruption. Thus, we classify anomalies as point, contextual, and collective, and discuss the issues with the rule-based approach to monitoring. By leveraging ML algorithms, database administrators can achieve proactive monitoring, predictive maintenance, and real-time detection of subtle deviations in system behavior. This paper underscores the vital role of Senior Database Administrators (DBAs) in implementing these advanced techniques to ensure optimal database performance, security, and reliability.**

Keywords—Anomaly Detection, Machine Learning, Security Breaches,System Observability, Database Systems.

I. INTRODUCTION

One of the main issues of system-level radiation testing is that the identification of the criticality of the failures is often complicated by the fact of dealing with a complex system, where different effects may result in an identical observation. The enhancement of system observability can provide an additional source of information for a more efficient identification of the failures encountered during the system irradiation. The improvement of system observability has been previously demonstrated by the implementation of software-based instrumentation cores in a System-on-a-Chip (SoC)[1], where the performed observations also provided sufficient data to identify the failure root causes.

The challenge of finding patterns in data that deviate from anticipated behavior is known as anomaly detection; these patterns are mostly known as anomalies and outliers. Because abnormalities in data might provide important information about the health state of the product, anomaly identification is crucial in PHM. Most anomaly detection techniques[2], in general, create a profile of typical occurrences before identifying anomalies that deviate from the profile.

Different approaches to anomaly detection fall into one of four main categories: distance-based, clustering-based, classification-based, or statistical. Distance-based approaches take use of outliers' characteristics that are geographically distant from the data gathered via normal nominal products. Normal observations will be assumed to belong to the same cluster or clusters using approaches that are based on clustering. A new observation will be considered anomalous if it is located distant from the cluster centroid (s). Differentiating abnormal from normal data is the job of classification-based algorithms like neural networks, KNN, and one-class SVMs. The statistical characteristics of outliers are also used by statistical approaches[3].

The field of ML has made tremendous strides in recent years. Voice recognition and recommendation systems are two examples of the real-world applications of AI that have emerged from lab research[4]. Anomaly detection in industrial control networks has made extensive use of ML methods, such as supervised learning, unsupervised learning, and reinforcement learning, because of its efficacy and little human inputs. When it comes to industrial control network anomaly detection, there are a few different approaches. One is supervised learning, which involves finding a function that maps the input data with a normal or abnormal target vector. The other is unsupervised learning, which involves making inferences from the unlabeled data to determine if an anomaly has occurred. Finally, reinforcement learning is great for making dynamic decisions in real-world industrial situations in order to maximize specific rewards[5]. The following paper concepts as:

- Defines and categorizes anomalies into point, contextual, and collective types, enhancing understanding of their detection methods.
- Discusses an application of ML techniques for effective anomaly detection, surpassing traditional rule-based approaches.
- Highlights the advantages of real-time monitoring using ML for early detection of potential database issues.
- Demonstrates how machine learning can facilitate predictive maintenance by identifying early signs of system failures.
- Stresses on the principal duties of Senior Database Administrators in live implementation and use of complex anomaly detection solutions.

A. Organization of the paper

The following is the overall layout of the paper; Section II provides an overview of anomaly detection. Section III is about the machine learning techniques for anomaly detection. Section IV speaks of increasing the observability of databases using machine learning. Section V senior DBA Section VI literature review Section VII conclusion and future work.

II. ANOMALY DETECTION: AN OVERVIEW

Anomaly detection is a common process for many fields, including database management because it is vital to recognize data that did not behave in a way that was expected. It need to be emphasised, that these anomalies may indicate significant events like, for example, performance decline, security violation or system dysfunction. In this section, the general notion of anomaly detection, its importance in database systems, and categories of anomalies are described [6].Anomaly detection is defined as the ability to find out data points or data patterns in a given dataset that are unusual. Problems like mistakes, fraud, or other strange actions may be signalled by these outliers, which are also known as anomalies. Anomalies can disrupt the smooth functioning of systems, leading to costly downtimes, breaches, or data corruption.

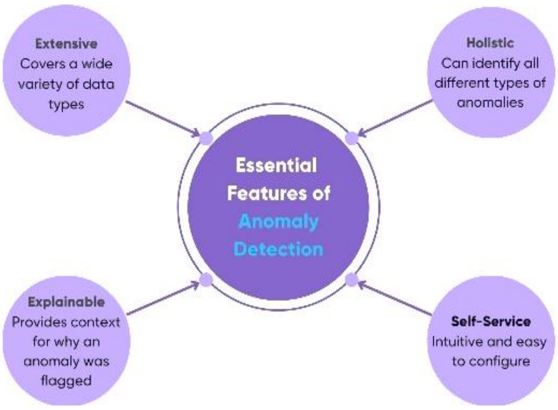
Extensive
Covers a wide variety of data types
Holistic
Can identify all different types of anomalies
Essential Features of Anomaly Detection
Explainable
Provides context for why an anomaly was flagged
Self-Service
Intuitive and easy to configure

Figure 1: Features of Anomaly detection

These features are shown in Figure 1 below, along with a thorough explanation.

- **Extensive:**Users often want a flexible tool capable of handling a large range of data. When working with vastly varied types of data, such as few, many, near-constant, high variance, many, uncommon, and having changepoints (drifts), anomaly detection methods ought to perform rather well.
- **Self-Service:**It is important that the anomaly detection tool be easy enough for a wide range of employees to utilize, regardless of their degree of technical expertise. Anomaly detection services are used by users to lessen their workload, not to increase it. So that users don't become frustrated or sidetracked before even starting to analyze the data, the algorithm's configurability should be modest.
- **Holistic:**Most, if not all, of a user's data flow is often covered by anomaly detection. Anomalies at the record level, transaction data, metadata, or all three might be causing them problems. They could also need other things in the future. A comprehensive solution that can adapt to their evolving needs is required. You must be able to identify abnormalities at high to low levels, regardless of how near or how distant they are from the data.
- **Explainable:**People like AI because it does some of their tasks for them. At the same time, individuals are frequently hesitant to relinquish control and allow AI make choices for them. That is why they want AI proposals to be explainable so that they can see how AI arrived at its conclusions.

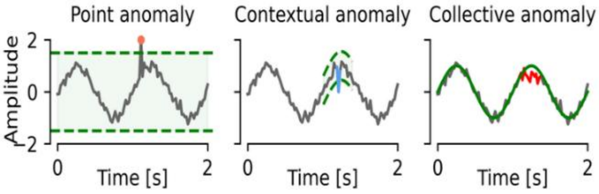

Figure 2: Three time series anomaly types

Types of Anomalies (Point, Contextual, Collective)

Figure 2 depicts the three main forms of anomalies: point anomalies, contextual anomalies, and collective anomalies. Each type represents a different kind of deviation from expected patterns and requires different detection techniques.

- **Point Anomalies:**Those are specific pieces of information that stand out from the rest of the data collection. For example, in a database monitoring system, an unusually high CPU usage in a single record could be considered a point anomaly. Point anomalies are the simplest to detect and can often be identified using statistical methods or distance-based machine-learning models. Algorithms like Isolation Forests or One-Class SVMs are commonly used for detecting point anomalies.
- **Contextual Anomalies:** These anomalies occur when a data point is anomalous in a specific context but may appear normal otherwise. For instance, high frequency of queries on a database during off-peak use may be an anomalous event, whereas the same may not be during peak usage. Time related contextual anomalies are quite common in time series data collection where the context plays an important role. A number of contextual anomalies are identified through discovering sequential dependencies using learning models such as LSTM networks or HMMs.
- **Collective Anomalies:**These occur when many continuous data points behave in an abnormal manner albeit in synchrony contrarily to other individual points. In database systems, a collective anomaly could be a set of database transactions that are legal by themselves but if performed together show a collapse or an attack in a system. Clustering or sequence analysis approaches such as DBSCAN or sequence-based neural networks are normally used to detect collective anomalies[7].

A. Anomaly Detection in Database Systems

Conventional database monitoring employs set point mechanisms and static rules for detecting anomalies that do not work well in challenging cutting-edge environments. As the amount and complexity of data increase in modern databases, ML has turned into a mandatory instrument to identify such sophisticated trends as the number and degree of variations from traditional static rules [8].

- **Performance Anomalies:**These are aspects like query response delays, abrupt memory swells, or weird disk operations. Many of these performance anomalies can be identified by clustering or regression models since this approach teaches the system about regular performance and identifies variations from it. For example, Netflix employing machine learning to continuously supervise the congestions in their distributed database system.
- **Security Anomalies:**Databases are a perfect avenue for cyberattacks, and recognizing such changes can avoid or minimize cyberattacks. Security anomalies include any access patterns that go against the norm, unauthorized changes in stored data, or attempts at using SQL injection. RF or Neural Networks are supervised machine learning models through which historical security events can be trained in order to identify future anomalies. Further, there is use of unsupervised techniques such as Autoencoders to identify unseen (new) attack patterns.
- **Resource Utilization Anomalies:**Resource usage management such as CPU, memory and disks used is very important in database systems. Sudden changes in resource usage may indicate underlying issues like inefficient queries, hardware failures, or configuration errors. Machine learning models, particularly time-series analysis techniques such as ARIMA or LSTMs, are used to predict resource usage trends and detect deviations from expected behavior.

III. MACHINE LEARNING TECHNIQUES FOR ANOMALY DETECTION

Anomaly detection using ML encompasses a wide array of algorithms, each tailored to different types of anomalies and data.The branch of AI known as ML is crucial because it allows computers to learn and improve at tasks traditionally performed by humans without the need for extensive manual instruction. Farming and gardening are analogous to ML. Programs are the plant, you are the gardener, and algorithms are the seeds and nutrition[9]. The field studying how computers may learn to do tasks without human intervention is known as ML. The last ten years have seen incredible advancements in ML, which have brought us self-driving vehicles, realistic voice recognition, effective online search, and a far better knowledge of the human genome. Nowadays, ML is so commonplace that you likely use it often without even realizing it. It is also seen by many experts as the most effective path towards AI on par with humans. With an increasing number of important applications realising the potential of ML, including data mining, image recognition, NLP, and expert systems, ML is emerging as a breakthrough technology[10].

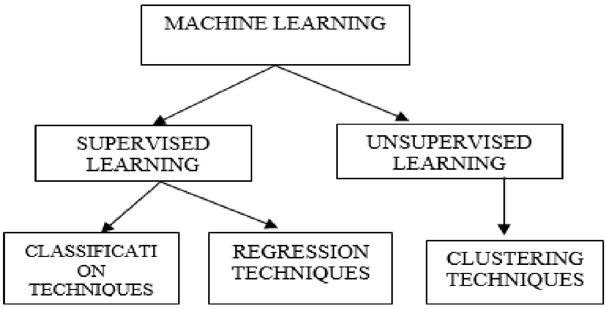

Figure 3: Types of machine learning

Figure 3 displays the many kinds of ML models. With its potential answers in all these areas and more, ML is poised to play a crucial role in our future society. There are several subfields within ML, such as supervised and unsupervised algorithms, and subfields within that, such as classification, regression, and clustering, which each aim to provide different kinds of outputs.

A. Supervised Learning

Supervised learning uses tagged training data to derive a function that is then utilised for verification and classification[11].Classification and regression are the two main branches of supervised learning.

- **Classification:**The goal of classification, a supervised learning approach, is to assign a value to a variable or make predictions about its future behavior. Typically, a classification method will have two phases: one for building the training model and another for verifying it. Image recognition, voice recognition, text categorisation, spam filtering, and fraud detection are just a few of the many applications of classification.
- **Regression**: Another supervised learning approach that is used for making predictions is regression. When working with categorical data, classification is more appropriate than regression, which uses continuous data to make value predictions. The one that has to be forecasted is called the dependent variable. In regression, there is a singular dependent variable. Independent variables are those that are used for training or modelling purposes. Regression analysis is referred to as linear regression when there is a single independent variable and as multiple regression when there are many independent variables.

B. Unsupervised Learning

Unsupervised learning uncovers latent structure in data that is not labelled. One of the key components of unsupervised learning is clustering. By using clustering, datasets are divided into groups or clusters where there is a minimal amount of intercluster similarity and a maximum amount of intracluster similarity between data points. Applications for clustering include picture segmentation, document retrieval, customer segmentation, and pattern classification. Several clustering methods.

- Regarding clustering techniques, K-means has been among the most used. suggested PK-means, a parallel implementation of k-means. Since calculating distance is the most resource-intensive part of the k-means algorithm, it is possible to run separate calculations for different data points in parallel. The PK-means algorithm employs the map, combine, and reduce operations. Every data point is assigned to the nearest centre using the map function. After the map function's intermediate outputs have been aggregated, the combined function provides the aggregated data to the reducer. The centroids are updated using the reduce function.
- The goal of hierarchical clustering is to create a hierarchy of clusters using either the top-down method of divisive clustering or the bottom-up method of agglomerative clustering. One kind of agglomerative clustering is SHC, or single-linkage hierarchical clustering. Because of the algorithm's intrinsic data reliance, achieving parallelism is challenging.

IV. ENHANCING DATABASE OBSERVABILITY WITH MACHINE LEARNING

Database observability, focusing on their integration into observability systems, real-time monitoring, predictive maintenance, and visualization of detected anomalies. As database systems become more complex, traditional monitoring methods often fall short in identifying hidden patterns and anomalies, leading to performance issues or system failures. ML concepts on the other hand offer a much more cumbersome and automatic solution to these challenges compared to simply monitoring a system or checking for symptoms of an anomaly.

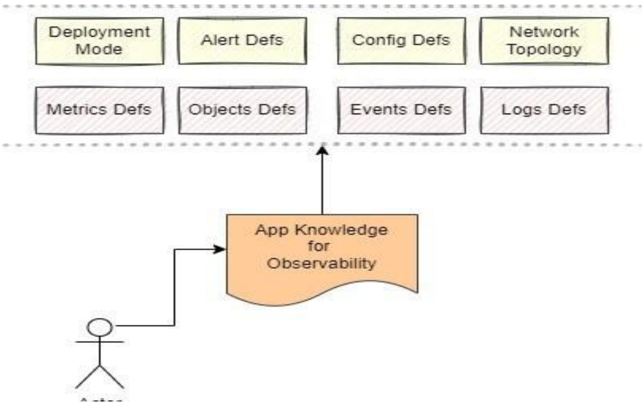

Figure 4: Database Observability and Storage Insights

Awareness of the application is crucial to achieve effective database observability as depicted in the show sin Figure 4. This includes the different deployment modes, defined metrics, well known logs & event list, configurations, the database object, the network, visualization dashboard as well as the possibility alerts/issues.

A. *Integration of ML Models in Observability Systems*

Incorporation of the said ML models into observability systems calls for behavior-based analysis instead of the rule-based systems. Conventional methods of monitoring involve setting up alerts that specify specific levels to trigger the alarm to the administrator. However, these thresholds cannot take into consideration the ever changing and complex database landscapes of today and even beneath the cloud.

B. *Real-Time Monitoring and Anomaly Detection*

Real-time is possible with the assistance of machine learning, which also helps avoid possible problems from happening in the first place. While the rule-based system would pre-define rules to be executed, the same with ML solutions can analyze live data streams and recognize slight changes indicating decline in performance, fault in hardware or intrusion.

C. *Predictive Maintenance through Anomaly Detection*

Another use case which is a proven advantage of ML-driven anomaly detection is predictive maintenance. Consequently, in database systems, breakdown or slow running can be expensive. Use of machine learning is able to predict the time it is due for maintenance so that the administrators can address the issues before they worsen.

V. SENIOR DATABASE ADMINISTRATOR (DBA)

Business databases are thus very important structures in today's organizations as they provide the framework for controlling, archiving and accessing some of the most crucial organizational information. As the amount of data continues to grow in complexity, it is almost compulsory to have the services of a Senior Database Administrator (DBA). Anomaly detection is one of the critical components in database management since the identification of irregular patterns is crucial in detecting performance problems, security threats, or a data integrity problem. Over the past few years the application of machine learning (ML) techniques has become well established to improve the ability of the system to detect anomalies, in other words revealing a potential problem[12].

In database systems, anomaly detection plays the crucial role of system health and security check. Anomaly detection in database systems may relate to problems such as high traffic on specific application, slow response to queries, or even intrusion attempts. These deviations, if identified early, allow firms to manage risks, enhance the efficient functioning of an organization, and avoid major system crashes or cyber compromises.

A. *Role of a Senior Database Administrator (DBA)*

A Senior DBA's duty is to be in charge of a company's databases and or to perform the whole process of database administration. Their duty entails the setting up, implementation and management of the databases for optimal functionality, integrity and security. The DBAs have a pivotal responsibility of, guaranteeing high availability of the data, and prevent their corruption, and ensuring efficiency of the system to deal with voluminous loads of transaction and queries.

B. Key Areas for Anomaly Detection in Databases

Anomaly detection in databases typically focuses on four key areas:

- **Security Breaches:**Data breaches, SQL injections, or privilege escalations are the main threats that affect the database security. It is crucial to identify these types of anomalies and investigate them in order to avoid leakage of data as well as conforming to data protection policies.
- **Performance Degradation:**DBAs are therefore required to monitor the performance of the database as regular as it may be. Abnormal activities like slow queries, high locking, or high CPU/Memory usage impacts the overall system performance and needs to be detected and corrected on priority basis.
- **Data Integrity**: It is important to sustain data quality and credibility. Other defects in this domain may be omitted or twofold records, impaired data, or erratic changes in a record, which may signify poor data quality.
- **System Monitoring:**System resources such as CPU, memory or disk space may show abnormal utilization patterns that needs to be controlled by observing the databases. These anomalies could be precursors to system failures or performance bottlenecks.

VI. LITERATURE REVIEW

In this section provide the related work on anomaly detection using machine learning based on data administrator with system observability.

Thisarticle,Sharma, Sharma and Lal, (2019) presents an overview of MLand DLtechniques for anomaly detection in IoT applications. DL and ML are effective methods for examining both typical and anomalous IoT component and device behaviour. The main research questions and difficulties with deep anomaly detection methods for devices with limited resources in real-world IoT situations are outlined in this study. In order to address some network security concerns and calculation delays, fog computing shifts processing to the device or edge[13].

In this paper,Dawoud, Shahristani and Raun, (2019)returning to the topic of network anomaly detection in order to investigate the possibilities of DL in detecting dangers to networks. They will be examining DL methods that do not rely on supervised learning. The research suggests using Unsupervised DL algorithms as the basis for a semi-supervised detection system. The study delves into the pros and cons of using DL to spot outliers, with a focus on autoencoders as a non-probabilistic approach. In order to discover abnormalities, we provide an in-depth examination of AE. Our findings demonstrate that the USDL would improve detection accuracy by more than 99%[14].

In this paper,Pwint and Shwe, (2019)instead of focussing on identifying outliers, they explore the possibility of using Apache Spark, a big data technology, to categorise various assaults. Using the gold standard dataset created by MAWILab, they categorise 15 distinct kinds of attacks using typical ML methods such as MLR, DT, RF, MLP, and NB. Our findings show that compared to conventional ML, the use of big data technology improves the performance and accuracy of the network traffic anomaly detector[15].

In this paper,Wang et al., (2019)explore the most up-to-date research on using ML to identify anomalies in industrial control networks. They begin by outlining the key distinctions between OT and ICT, and then they compare and contrast their respective benefits and drawbacks in terms of anomaly detection. They weigh the benefits and drawbacks of these two approaches to education. In conclusion, they highlight the encouraging directions this field of study is headed[5].

This research,Nixon, Sedky and Hassan, (2019) aimed in order to assess the real-world uses of online ML techniques for detecting intrusions in IoT networks. To address the challenges of the distributed and resource-constrained IoT perception layer, online learning offers a memory-and time-efficient architecture that can adapt to idea drift and perform anomaly detection[16].

This paper, Portela, Almenares Mendoza and Benavides, (2019)uses the UNSW-NB12 dataset to assess the efficacy of several intrusion detection methods, including supervised ones (KNN and SVM) and unsupervised ones (Isolation Forest and K-Means). The supervised method SVM gaussiana fine achieved an accuracy of 92%, which means it can accurately distinguish between normal and abnormal data, according to the findings. The K-Means approach neatly clusters the data and lets you choose the right amount of groups when it comes to unsupervised algorithms; nonetheless, this dataset is quite agglomerated[17].

This paper,Dutt, Borah and Maitra, (2018)suggests a method for creating an IDS in the WEKA environment using PCA and ML methods. Because it improves detection efficiency, the method improves performance. When compared to the current approaches, the findings demonstrate a decrease in false positive rates and an increase in real positive rates[18].

This paper,Mehmood and Rais, (2016)uses the anomaly-based detection method to evaluate several supervised algorithms. On the KDD99 dataset, the methods have been implemented. Every class in the KDD99 dataset was found to have a low detection rate by at least one algorithm. Precision, true positive rate, and false positive rate are the performance metrics used for this comparison. This Table1 summarizes the contributions of the related works, focusing on methods and key findings relevant to enhancing system observability in the context of anomaly detection[19].

Table 1: Comparative analysis on related work

Reference	Focus Area	Methods	Key Findings	Limitations	Future Work
[13]	Anomaly detection in IoT applications	Machine Learning, Deep Learning	Discusses challenges of deep anomaly detection in resource-constrained devices; highlights fog computing as a solution for network security and latency issues.	Limited scalability for large-scale IoT deployments; computational overhead on devices.	Investigate more efficient algorithms for real-time processing in constrained environments.
[14]	Network anomaly detection	Unsupervised Deep Learning (Autoencoders)	Proposes a semi-supervised framework, demonstrating accuracy over 99% with unsupervised deep learning algorithms.	Dependency on sufficient labeled data for effective semi-supervised learning.	Explore hybrid models combining supervised and unsupervised techniques for better performance.
[15]	Attack classification	Traditional Machine Learning (Logistic Regression, Decision Trees, etc.)	Shows that using Apache Spark improves prediction accuracy and execution time for classifying network attacks compared to traditional methods.	Limited to the specific dataset; may not generalize well to other environments.	Validate results across diverse datasets and real-world scenarios to enhance robustness.
[5]	Anomaly detection in industrial control networks	Data-based learning, Network-specified learning	Analyzes advantages/disadvantages of operational technology (OT) vs. information and communication technology (ICT) methods, suggesting future work areas.	Lack of comprehensive datasets for industrial control networks; challenges in feature selection.	Develop more representative datasets and refine learning methods tailored for OT environments.
[16]	Online machine learning for IoT intrusion detection	Online Learning Architecture	Introduces an online learning architecture for anomaly detection, emphasizing resource efficiency and adaptability to concept drift.	Challenges with class imbalance in data streams; potential for outdated models.	Focus on methods for addressing class imbalance and model updates in real-time.
[17]	Intrusion detection performance evaluation	Supervised (KNN, SVM) & Unsupervised (Isolation Forest, K-Means)	SVM achieved 92% accuracy; highlights challenges of various algorithms on different datasets.	Performance varies significantly with dataset characteristics; requires extensive parameter tuning.	Investigate automated parameter tuning techniques and evaluate on more diverse datasets.
[18]	Development of IDS	Principal Component Analysis, Machine Learning	Reports improved detection rates with PCA and machine learning in the WEKA environment, achieving higher true positives and lower false positives.	Limited scope of datasets used; potential for overfitting.	Test the approach on a broader range of datasets to ensure generalizability.
[19]	Anomaly-based detection	Supervised Learning Algorithms	Evaluates multiple supervised algorithms on KDD99 dataset, finding no single algorithm excels across all classes, emphasizing the need for tailored approaches.	KDD99 dataset may not reflect current network traffic patterns; outdated attack types.	Explore newer datasets and incorporate ensemble methods for improved classification across classes.

VII. CONCLUSION AND FUTURE SCOPE

Nowadays, software technology is crucial to contemporary life. The rise in demand for software applications has brought with its new security problems. When it comes to safeguarding their systems from vulnerabilities, application security has emerged as a top priority for enterprises. In conclusion, the integration of machine learning techniques for anomaly detection provide a paradigm shift in regard to the observability of databases. This way, instead of relying on rule-based methods of monitoring, organizational management subject their systems to the application of advanced ML solutions for better identification of anomalies with corresponding occurrences. Apart from reducing risks posed by declining performance and vulnerabilities, this approach ensures data accuracy and business productivity. Senior Database Administrators play a central part in this shift since they are charged with the responsibility of implementing and supporting such superior systems. This is especially vital as data environments progress, the implementation of ML for anomaly detection will be crucial for overall database systems, organizational health, and integrated security.

Further studies can build on improving the accuracy of machine learning algorithms for anomaly detection in the context of extensive and decentralized DBMSs. More efforts should be made to propose more refined hybrid learning models integrated with the merits of both supervised and unsupervised learning for more accurate anomaly detection. Further, adopting deeper learning and reinforcement learning into monitoring systems for real-time databases can improve the accuracy and effectiveness of the database observability systems. Database cooperative work with artificial intelligence professionals will also be important when developing these solutions for different sectors.

REFERENCES

[1] Y. Yuan, K. Dehghanpour, F. Bu, and Z. Wang, "A Multi-Timescale Data-Driven Approach to Enhance Distribution System Observability," *IEEE Trans. Power Syst.*, 2019, doi: 10.1109/TPWRS.2019.2893821.

[2] S. B. Wankhede, "Anomaly Detection using Machine Learning Techniques," in *2019 IEEE 5th International Conference for Convergence in Technology, I2CT 2019*, 2019. doi: 10.1109/I2CT45611.2019.9033532.

[3] D. Kwon, H. Kim, J. Kim, S. C. Suh, I. Kim, and K. J. Kim, "A survey of deep learning-based network anomaly detection," *Cluster Comput.*, 2019, doi: 10.1007/s10586-017-1117-8.

[4] A. Kulkarni, Y. Pino, M. French, and T. Mohsenin, "Real-time anomaly detection framework for many-core router through machine-learning techniques," *ACM J. Emerg. Technol. Comput. Syst.*, 2016, doi: 10.1145/2827699.

[5] Q. Wang, H. Chen, Y. Li, and B. Vucetic, "Recent advances in machine learning-based anomaly detection for industrial control networks," in *1st International Conference on Industrial Artificial Intelligence, IAI 2019*, 2019. doi: 10.1109/ICIAI.2019.8850828.

[6] L. Akoglu, H. Tong, and D. Koutra, "Graph based anomaly detection and description: A survey," *Data Min. Knowl. Discov.*, 2015, doi: 10.1007/s10618-014-0365-y.

[7] P. Malhotra, L. Vig, G. Shroff, and P. Agarwal, "Long Short Term Memory networks for anomaly detection in time series," in *23rd European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning, ESANN 2015 - Proceedings*, 2015.

[8] R. A. Ariyaluran Habeeb, F. Nasaruddin, A. Gani, I. A. Targio Hashem, E. Ahmed, and M. Imran, "Real-time big data processing for anomaly detection: A Survey," *Int. J. Inf. Manage.*, vol. 45, no. February, pp. 289–307, 2019, doi: 10.1016/j.ijinfomgt.2018.08.006.

[9] K. W. DeGregory *et al.*, "A review of machine learning in obesity," *Obesity Reviews*. 2018. doi: 10.1111/obr.12667.

[10] S. Ray, "A Quick Review of Machine Learning Algorithms," in *Proceedings of the International Conference on Machine Learning, Big Data, Cloud and Parallel Computing: Trends, Prespectives and Prospects, COMITCon 2019*, 2019. doi: 10.1109/COMITCon.2019.8862451.

[11] S. B. Kotsiantis, "Supervised machine learning: A review of classification techniques," *Informatica (Ljubljana)*. 2007.

[12] M. Ahmed, A. Naser Mahmood, and J. Hu, "A survey of network anomaly detection techniques," *Journal of Network and Computer Applications*. 2016. doi: 10.1016/j.jnca.2015.11.016.

[13] B. Sharma, L. Sharma, and C. Lal, "Anomaly Detection Techniques using Deep Learning in IoT: A Survey," in *Proceedings of 2019 International Conference on Computational Intelligence and Knowledge Economy, ICCIKE 2019*, 2019. doi: 10.1109/ICCIKE47802.2019.9004362.

[14] A. Dawoud, S. Shahristani, and C. Raun, "Deep learning for network anomalies detection," in *Proceedings - International Conference on Machine Learning and Data Engineering, iCMLDE 2018*, 2019. doi: 10.1109/iCMLDE.2018.00035.

[15] P. H. Pwint and T. Shwe, "Network Traffic Anomaly Detection based on Apache Spark," in *2019 International Conference on Advanced Information Technologies, ICAIT 2019*, 2019. doi: 10.1109/AITC.2019.8920897.

[16] C. Nixon, M. Sedky, and M. Hassan, "Practical Application of Machine Learning based Online Intrusion Detection to Internet of Things Networks," in *2019 IEEE Global Conference on Internet of Things, GCIoT 2019*, 2019. doi: 10.1109/GCIoT47977.2019.9058410.

[17] F. G. Portela, F. Almenares Mendoza, and L. C. Benavides, "Evaluation of the performance of supervised and unsupervised Machine learning techniques for intrusion detection," in *2019 IEEE International Conference on Applied Science and Advanced Technology, iCASAT 2019*, 2019. doi: 10.1109/iCASAT48251.2019.9069538.

[18] I. Dutt, S. Borah, and I. Maitra, "A Proposed Machine Learning based Scheme for Intrusion Detection," in *Proceedings of the 2nd International Conference on Electronics, Communication and Aerospace Technology, ICECA 2018*, 2018. doi: 10.1109/ICECA.2018.8474803.

[19] T. Mehmood and H. B. M. Rais, "Machine learning algorithms in context of intrusion detection," in *2016 3rd International Conference on Computer and Information Sciences, ICCOINS 2016 - Proceedings*, 2016. doi: 10.1109/ICCOINS.2016.7783243.