Migration Letters

Volume: 20, No: S13(2023), pp. 560-565

ISSN: 1741-8984 (Print) ISSN: 1741-8992 (Online)
www.migrationletters.com

Machine Learning Algorithms for Anomaly Detection in IoT Networks

Dr.J. Grace Hannah[1], Dr.A. Sampath Dakshina Murthy[2], Dr. Gauri Kalnoor[3],

M. Vetriselvan[4] and Dr.M.S. Nidhya[5]

Abstract

Anomaly detection in IoT networks is a pivotal task for identifying unexpected events that can yield crucial insights in sectors like healthcare, finance, and security. This paper explores various machine learning algorithms for anomaly detection, including supervised, unsupervised, and semi-supervised approaches. It discusses the challenges in implementing these algorithms in IoT environments, emphasizing the need for lightweight and efficient solutions. Preprocessing techniques, evaluation metrics, and case studies are examined, providing a comprehensive overview of practical applications and performance evaluation methods. The paper presents a case study on anomaly detection in an IoT-based temperature monitoring system using a Gaussian Mixture Model (GMM). The study demonstrates the successful integration of the algorithm, emphasizing benefits such as preventive maintenance, quality assurance, and operational efficiency. Performance metrics such as precision, recall, and F1 score are utilized for evaluation, showcasing the algorithm's effectiveness in identifying anomalies. Future research directions are outlined, emphasizing the need for real-time anomaly detection with limited resources, incorporating human expertise in decision-making, and addressing ethical considerations. The importance of diversity and transparency in algorithm development is highlighted, and suggestions for integration of additional sensors, dynamic threshold adjustments, and optimized edge computing are proposed. In conclusion, the paper emphasizes the significance of machine learning algorithms in anomaly detection for IoT networks, offering insights into their applications, challenges, and future directions. It underscores the need for continuous exploration and adaptation to evolving challenges, ensuring the security, effectiveness, and ethical considerations in the development and deployment of anomaly detection algorithms in IoT networks.

Keywords: *Anomaly Detection, Supervised, Unsupervised, and Semi-supervised.*

Introduction

Anomaly detection, also called outlier detection, is the identification of unexpected events, observations, or items that differ significantly from the norm [1]. The significance of anomaly detection is that anomalies in IoT data, which occur sparsely, can yield crucial actionable information in various sectors such as healthcare, finance, and security [2]. Anomaly detection algorithms act as checkpoints for the incoming traffic at various

[1] Assistant Professor, Department of Information Technology, Vels Institute of Science, Technology and Advanced Studies (VISTAS), Chennai. Grace.scs@velsuniv.ac.in

[2] Associate Professor, Department of ECE, Vignan's Institute of Information Technology (A), Visakhapatnam. sampathdakshinamurthy@gmail.com

[3] Assistant Professor, BMS College of Engineering, Bangalore. Gaurik.cse@bmsce.ac.in

[4] Artificial Intelligent and Data Science, Panimalar Engineering College, Chennai. Vetrinelson7@gmail.com

[5] Associate Professor, Department of CS & IT, Jain Deemed-to-be University, Bangalore. nidhyaphd@gmail.com

stages, ranging from the IoT network level to the data centre [2]. These algorithms process data in real-time, analyze it, and identify any abnormal patterns. Machine learning for anomaly detection is crucial in identifying unusual patterns or outliers within data [3]. Simple statistical techniques such as mean, median and quantiles can be used to detect univariate anomaly feature values in the data set [4]. Machine learning (ML) plays a pivotal role in anomaly detection by providing sophisticated techniques to identify irregular patterns or outliers [5]. Some of the most benefits of Anomaly detection include Preventive Maintenance, Quality Assurance and Operational Efficiency.

- Preventive Maintenance: Early detection of anomalies enables proactive maintenance of temperature control systems, reducing the risk of equipment failures.
- Quality Assurance: Ensures the quality of stored goods by preventing temperature variations that could compromise sensitive products.
- Operational Efficiency: Minimizes downtime by addressing issues promptly, improving the overall efficiency of warehouse operations.

Anomaly detection algorithms act as checkpoints for the incoming traffic at various stages, ranging from the IoT network level to the data centre. In this paper, we will explore the various machine learning algorithms for anomaly detection in IoT networks.

Literature Survey

Machine learning algorithms can be broadly classified into supervised, unsupervised, and semi-supervised learning algorithms. Anomaly detection modeling is a subset of unsupervised machine learning. It's unsupervised since there's no predetermined target or “ground” [6]. Supervised anomaly detection models are designed to detect anomalies in a dataset using labeled data, where each data point is classified as normal or anomalous [7]. Supervised Anomaly Detection requires a labeled dataset containing both normal and anomalous samples to construct a predictive model, which can then be used to detect anomalies in new data [8]. Supervised algorithms are effective when labeled data is available and the anomalies are well defined.

Unsupervised Anomaly Detection is a technique where the model learns to identify anomalies without prior knowledge of what constitutes normal [9]. In this technique, the model identifies data points that lie outside the normal range of the data set. In this post, we'll be going through the need for anomaly detection and baseline algorithms for anomaly detection with underlying mathematics [10]. Anomaly detection is an important technique in machine learning and is used in a wide range of applications, including fraud detection, system health monitoring, and intrusion detection [9].

Semi-supervised learning (SSL) is a machine learning technique that uses both labeled and unlabeled data for training [11]. SSL works well for clustering and anomaly detection, where labeled data is scarce and the anomalies are not well defined. In anomaly detection, the goal is to detect patterns or observations that are unusual or different from the norm [12]. SSL algorithms are designed to leverage the benefits of both supervised and unsupervised learning techniques and can be used effectively in situations where labeled data is scarce or expensive to obtain [12].

A comparative study of machine learning algorithms for anomaly detection in IoT networks reveals that both supervised and unsupervised learning algorithms have their strengths and weaknesses [18]. In-vehicle can-bus network data anomaly detection is studied using traditional ML algorithms and deep learning algorithms [19]. The study shows that Multi-Layer Perception (MLP) relies more on network configuration for intrusion detection and has dependencies between input and output. Therefore, it is essential to choose the most appropriate algorithm based on the specific requirements and characteristics of the IoT network.

The implementation of machine learning algorithms for anomaly detection in IoT networks is challenging due to several factors [14]. A detailed analysis of five ML algorithms (logistic regression, support vector machine, decision tree, random forest, and artificial neural network) for anomaly detection in IoT networks reveals that several challenges persist [13]. The scarcity of IoT resources and the need for real-time processing of data are some of the challenges in implementing these algorithms. Moreover, the algorithms need to be optimized for the IoT environment, which requires the development of lightweight and efficient algorithms that can operate with limited resources.

Preprocessing techniques play a crucial role in anomaly detection in IoT networks. This paper reviews the data preprocessing techniques used by anomaly-based network intrusion detection systems (NIDS) [15]. These techniques are used to identify relevant features and to reduce the dimensionality of the data, which improves the accuracy and efficiency of the anomaly detection algorithms.

Implementation

We choose a Gaussian Mixture Model (GMM) for anomaly detection due to its ability to capture complex distributions in data. The algorithm is trained on a subset of the dataset representing normal temperature patterns.

Data Collection: Temperature data is collected over several weeks from IoT sensors placed strategically within the warehouse. The dataset includes normal temperature patterns during regular warehouse operations.

Data Preprocessing: Before training the model, data preprocessing steps are applied:

- Feature Selection: Relevant features such as time of day, day of the week, and historical temperature values are considered.
- Normalization: The temperature data is normalized to ensure that all features are on a similar scale.

Training and Testing: The GMM is trained on a labeled dataset containing instances of normal temperature behavior. The testing set includes new temperature data collected in real-time, enabling the model to identify anomalies during operation.

The anomaly detection algorithm is integrated into the existing IoT temperature monitoring system. It continuously analyzes incoming temperature data in real-time and raises alerts when anomalies are detected.

Evaluation metrics are essential for assessing the performance of machine learning algorithms for anomaly detection in IoT networks. Beyond accuracy, the most commonly used metrics when evaluating anomaly detection solutions are F1, Precision, and Recall [16]. False Positive and False Negative rates are also crucial when deciding on the anomaly detection system [16]. However, evaluating anomaly detection algorithms and selecting the most appropriate one for a particular purpose is not an easy task since different algorithms have different strengths and weaknesses [17]. Therefore, it is essential to use multiple metrics to evaluate the performance of the algorithms accurately.

Case Study

Case studies of anomaly detection using machine learning algorithms in IoT networks provide insight into the practical applications of these algorithms. One such study involved collecting historical data from telemetry sensors, IoT cameras, and control devices in a smart manufacturing plant [20]. The data provided information about the performance of the machines, which was used to identify anomalies and predict machine failures. Another study focused on detecting different kinds of anomalies based on a data set using two Logistic Regression and Artificial neural networks [21]. These case studies demonstrate the effectiveness of machine learning algorithms for anomaly detection in

IoT networks and the potential benefits of implementing these algorithms in various sectors.

In this case study "Anomaly Detection in IoT-Based Temperature Monitoring System", we aim to simulate an anomaly detection algorithm to monitor and identify unusual temperature patterns in an IoT-based temperature monitoring system. The goal is to proactively detect anomalies that could indicate equipment malfunctions or environmental irregularities. Consider a scenario where IoT devices are deployed in a warehouse to monitor temperature conditions. These devices continuously collect temperature data and transmit it to a central server for analysis. The primary concern is to identify anomalies that may lead to equipment failures, potential damage to stored goods, or compromise the quality of sensitive products.

Performance Evaluation: Performance metrics such as precision, recall, and F1 score are used to evaluate the algorithm's effectiveness. False positives and false negatives are carefully considered, as they have implications for operational decisions.

Table 1: Performance metrics

Precision: 1.00	Recall: 0.57	F1 Score: 0.73

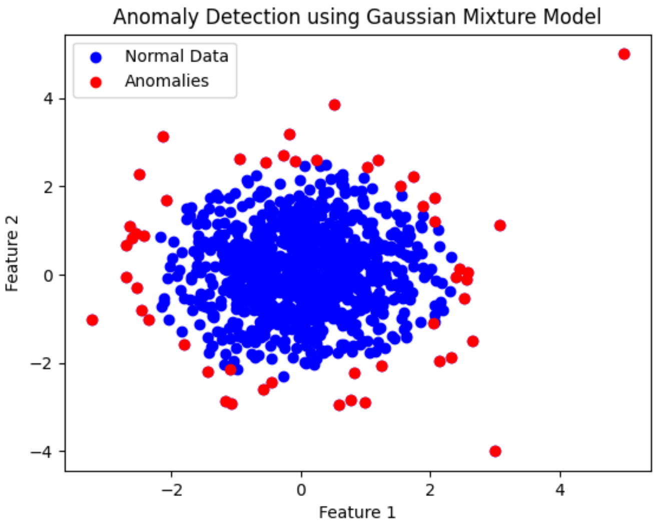

Figure 1: Anomaly Detection

Results: The anomaly detection algorithm successfully identifies anomalies in temperature patterns that deviate significantly from normal conditions. Alerts are generated promptly, allowing the warehouse staff to investigate and address potential issues before they escalate.

This case study demonstrates the practical application of an anomaly detection algorithm, specifically a Gaussian Mixture Model, in an IoT-based temperature monitoring system. The successful implementation provides valuable insights into potential anomalies, contributing to preventive maintenance and ensuring the reliability and quality of warehouse operations. Ongoing research and development are essential to adapt and optimize anomaly detection algorithms for evolving challenges in IoT networks.

Future Directions

Future research directions in anomaly detection using machine learning algorithms in IoT networks are crucial to address the challenges and limitations of the existing algorithms [22][14]. One potential research direction is to develop novel algorithms that can detect anomalies in real-time with limited resources, such as edge computing [22]. Another direction is to incorporate human expertise in the decision-making process of anomaly detection algorithms to improve their accuracy and reduce false positives [22]. Finally, research can be done to develop algorithms that can detect attacks that are not yet known or even imagined, such as zero-day attacks [14]. It is essential to continue exploring new directions in anomaly detection using machine learning algorithms to improve the security and effectiveness of IoT networks. While machine learning algorithms can improve the security and effectiveness of IoT networks, ethical considerations must be taken into account [23][24][25]. Biases in AI algorithms can lead to unfair and discriminatory outcomes [24][25]. Therefore, it is essential to ensure that the data used to train the algorithms is diverse and unbiased. Additionally, it is important to ensure that the algorithms are transparent, explainable, and accountable, so that they can be audited and evaluated for fairness and accuracy [23]. In addition to the above future directions, explore the integration of other IoT sensors, such as humidity sensors or motion detectors, to enhance anomaly detection capabilities. Implement algorithms that dynamically adjust anomaly detection thresholds based on seasonal or operational changes. Develop algorithms optimized for edge computing to address resource constraints in IoT devices. Ethical considerations must be taken into account when developing anomaly detection algorithms to ensure that they do not harm individuals or society.

Conclusion

In conclusion, machine learning algorithms stand out as powerful tools for anomaly detection in IoT networks, holding significant potential to enhance the security and effectiveness of these networks. The selection of supervised, unsupervised, or semi-supervised algorithms should be guided by the specific requirements of the IoT network, considering the strengths and weaknesses of each approach. Preprocessing techniques, evaluation metrics, and case studies offer valuable insights into the practical applications of these algorithms. Future research endeavors should prioritize the development of novel algorithms capable of real-time anomaly detection with limited resources. This entails incorporating human expertise into the decision-making process and addressing ethical considerations. Ongoing exploration of new directions in anomaly detection using machine learning algorithms is crucial for continually improving the security and effectiveness of IoT networks. In summary, anomaly detection in IoT networks is a critical task with applications in various sectors like healthcare, finance, and security. Machine learning algorithms have demonstrated high accuracy in detecting anomalies, showcasing their potential. Understanding the nuances of different algorithms, employing effective preprocessing techniques, and utilizing appropriate evaluation metrics are essential for practical applications. Future research should focus on overcoming challenges and limitations, with a commitment to ethical considerations, ensuring that machine learning algorithms continue to be instrumental in enhancing the security and efficiency of IoT networks.

References

Avi Networks. (n.d.). What is Anomaly Detection? Definition & FAQs. https://avinetworks.com/glossary/anomaly-detection/#:~:text=Anomaly%20detection%20is%20the%20identification.

Chatterjee, A. and Ahmed, B.S. (2022). IoT anomaly detection methods and applications: A survey. Internet of Things, 19, p.100568.

www.knowledgehut.com. (n.d.). Anomaly Detection with Machine Learning Overview. https://www.knowledgehut.com/blog/data-science/machine-learning-for-anomaly-detection.

builtin.com. (n.d.). 5 Anomaly Detection Algorithms to Know | Built In. [online] https://builtin.com/machine-learning/anomaly-detection-algorithms.

www.linkedin.com. (n.d.). 4 Machine Learning Algorithms For Anomaly Detection. https://www.linkedin.com/pulse/4-machine-learning-algorithms-anomaly-detection-hope-edet-b8ehf?trk=article-ssr-frontend-pulse_more-articles_related-content-card [Accessed 9 Jan. 2024].

Swansburg, J. (2023). Supervised Anomaly Detection: A Better Way to Model Anomalies. https://medium.com/@swansburg.justin/supervised-anomaly-detection-a-better-way-to-model-anomalies-bf39f67158ee.

Venujkvenk (2023). Anomaly Detection Techniques: A Comprehensive Guide with Supervised and Unsupervised Learning. https://medium.com/@ venujkvenk/anomaly-detection-techniques-a-comprehensive-guide-with-supervised-and-unsupervised-learning-67671cdc9680.

Gupta, A. (2019). Machine Learning for Anomaly Detection. [online] GeeksforGeeks. Available at: https://www.geeksforgeeks.org/machine-learning-for-anomaly-detection/.

Desai, U. (2023). Unsupervised Machine Learning with Anomaly Detection. [online] Medium. https://utsavdesai26.medium.com/unsupervised-machine-learning-with-anomaly-detection-5fae4fd2c957.

Bajaj, V. (2020). Unsupervised Learning For Anomaly Detection. [online] Medium. Available at: https://towardsdatascience.com/unsupervised-learning-for-anomaly-detection-44c55a96b8c1.

AltexSoft. (n.d.). Semi-Supervised Learning, Explained. https://www.altexsoft.com/blog/semi-supervised-learning/#:~:text=In%20a%20nutshell%2C%20semi%2Dsupervised.

Gupta, A. (2019). ML | Semi-Supervised Learning. [online] GeeksforGeeks. Available at: https://www.geeksforgeeks.org/ml-semi-supervised-learning/.

Huč, A., Šalej, J. and Trebar, M. (2021). Analysis of Machine Learning Algorithms for Anomaly Detection on Edge Devices. Sensors, 21(14), p.4946.

Diro, A., Chilamkurti, N., Nguyen, V.-D. and Heyne, W. (2021). A Comprehensive Study of Anomaly Detection Schemes in IoT Networks Using Machine Learning Algorithms. Sensors, 21(24), p.8320.

Landauer, M., Onder, S., Skopik, F. and Wurzenberger, M. (2023). Deep learning for anomaly detection in log data: A survey. Machine Learning with Applications, [online] 12, p.100470.

Bohutska, J. (2021). Anomaly Detection — How to Tell Good Performance from Bad. [online] https://towardsdatascience.com/anomaly-detection-how-to-tell-good-performance-from-bad-b57116d71a10.

Katser, I. (2023). A Review of Anomaly Detection Metrics (with a lot of related information). https://medium.com/@katser/a-review-of-anomaly-detection-metrics-with-a-lot-of-related-information-736d88774712.

Dong, Y., Chen, K., Peng, Y. and Ma, Z.Y. (2022). Comparative Study on Supervised versus Semi-supervised Machine Learning for Anomaly Detection of In-vehicle CAN Network. arXiv (Cornell University).

Sheth, V., Tripathi, U. and Sharma, A. (2022). A Comparative Analysis of Machine Learning Algorithms for Classification Purpose. Procedia Computer Science, 215, pp.422–431.

Ángel Jaramillo-Alcázar, Govea, J. and Villegas-Ch, W. (2023). Anomaly Detection in a Smart Industrial Machinery Plant Using IoT and Machine Learning. Sensors, 23(19), pp.8286–8286.

Harini Bilakanti, Sreevani Pasam, Varshini Palakollu and Sairam Utukuru (2024). Anomaly detection in IoT environment using machine learning. Security and Privacy.

Alwahedi, F., AlyaziaAldhaheri, Mohamed Amine Ferrag, Ammar Battah and Tihanyi, N. (2024). Machine learning techniques for IoT security: Current research and future vision with generative AI and large language models. Internet of Things and Cyber-Physical Systems.

Zhang, Y., Suleiman, B., Muhammad Johan Alibasa and Farid, F. (2023). Privacy-Aware Anomaly Detection in IoT Environments using FedGroup: A Group-Based Federated Learning Approach. Research Square (Research Square).

ISACA. (n.d.). Bias and Ethical Concerns in Machine Learning. [online] Available at: https://www.isaca.org/resources/isaca-journal/issues/2022/volume-4/bias-and-ethical-concerns-in-machine-learning#:~:text=One%20of%20the%20main%20areas.

www.linkedin.com. (n.d.). Ethical Considerations: Privacy, Security, and Accountability in the Age of AI-Assisted Mental Health Care. https://www.linkedin.com/pulse/ethical-considerations-privacy-security-accountability-age-ai-assisted.