International Journal of Scientific Research in Computer Science, Engineering and Information Technology
ISSN : 2456-3307 (www.ijsrcseit.com)
doi : https://doi.org/10.32628/IJSRCSEIT

Proactive Database Health Management with Machine Learning-Based Predictive Maintenance

Sandeep Kumar Dasa
Independent Researcher, USA

ARTICLEINFO

Article History:

Accepted: 01 Feb 2023
Published: 09 Feb 2023

Publication Issue
Volume 10, Issue 1
January-February-2023

Page Number
294-302

ABSTRACT

This assignment focuses on prediction maintenance using machine learning approaches to ensure the good health of the database. Convention approaches to database management are largely followers where problems crop up in the system, and solutions are sought only then. This causes a lot of issues, such as resource wastage. With the help of machine learning, the idea of predictive maintenance is focused on future breakdowns, enabling executable actions that enhance efficiency and reduce downtime. This study employs predictive analysis by checking database statistics (CPU, memory, response time) and using machine learning algorithms to detect signs of approaching problems. The process involves using models to provide the required accuracy and running a real-time application to evaluate the model's usefulness in preventing database failure. Expected outcomes suggest that predictive maintenance can improve database dependability, minimize unavailability, and decrease maintenance expenses. This approach will provide proactive management rather than reactive and change how the database is maintained to a more assertive and economical method.

Keywords : Database Health Management, Predictive Maintenance, Machine Learning, Proactive Monitoring, Reactive Maintenance Limitations, Database Reliability, Anomaly Detection, Operational Efficiency, System Resilience, Data-Driven Maintenance

Introduction

Solutions for maintaining functional, efficient, and responsive databases are important to keep databases healthy. Databases hold crucial business data for various clerical exercises of an association, such as transactional handling and insightful analysis. Conventional organizational maintenance practices tend to be more reactive, where administrators only attend to problems when they happen. Indeed, this "break-fix" approach leads to unscheduled outages and incurs considerable operational costs; teams need to detect and correct issues quickly. It also has the potential of not identifying deeper problems before they occur, for example, loss of data, prolonged stoppage time, or performance problems that affect user feedback.

Copyright: © the author(s), publisher and licensee Technoscience Academy. This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License, which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited

This is where ML comes into play, offering a more exhaustive solution of predictive maintenance, a method that uses trends in data to predict a problem before it happens. Data gathered across a system or organization's resources can be fed to the ML algorithms, and they shall R analyze how usage patterns, error reports, and other performance indicators differ from and describe regular behavior, noting that such differences indicate potential future issues. If predictive maintenance identifies such signs, database teams can prevent them through early interventions such as resource allocation or component replacement. However, that change towards a more proactive approach increases the reliability of databases. It decreases operation expenses while making a system more resilient, painting the picture of a smarter and more progressive style of managing databases.

Simulation Report

Data Collection and Preparation:

As with other forms of machine learning, predictive maintenance models need multiple forms of data to give consistent information about the health of the databases. Untapping, memory, disk occupying, errors detected, and the response time percentage provide the complete picture of the system's performance. Collecting and consolidating these types of data ensures that predictive models contain deep and shallow health of databases. As Liu et al. (2018) correctly pointed out, data fusion becomes more of an issue when several datasets are utilized in developing a single model. In addition, Hussain et al. (2018) also argue that comparing time series data is appropriate for time series analysis of predicting analytics into SLA and its relevance in explaining temporal features to forecast probable health issues in the system. To forecast, enhancing such data using data preprocessing greatly provides better precision, improving the systems' reliability.

Simulation Setup:

The environment for model training and its assessment was designed to assess various strategies used in machine learning. From this simulation, what can be actual database health monitoring in near real-time in a conflict-free environment without the pressure of a live system? The same architecture was described in Mandala (2019) using Kafka Streams and AWS IoT for based monitoring where real-time monitoring is also feasible. However, it resides, in fact, in the database. Oliván (2017) also identifies data-driven prognostics as one of the key challenges for constructing accurate models pertinent to various infrastructures, including databases. Such simulation setups provide a controlled environment log that may be used to improve the models and determine the extent of realism of success in conditions that can be simulated with considerable accuracy in other ways.

Machine Learning Models:

The main aspect related to the machine learning model selection for the predictive maintenance process is important because each type offers a certain amount of accuracy and interpretability. Jahnke (2015) evaluated several models for failure detection and observed that decision trees have rule-based predictions, which can be easily understood. On the other hand, neural networks can easily identify complex and non-linear patterns at the cost of interactivity. Similarly, Rafique & Velasco (2018) emphasize that machine learning in automating network monitoring is a task that is not merely comparable but ought to be and can be simultaneous and unintermitting with the health monitoring of databases for signs of anomalies. In predictive maintenance, by incorporating both interpretable and complex models into a system, the system's transparency and accuracy in detecting possible problems may be achieved where such issues pertain to databases.

Evaluation Metrics:

While assessing the reliability of the developed predictive maintenance models, certain parameters like accuracy, precision-recall and error rate, mean absolute error (MAE) & root mean square error (RMSE) were incorporated. These metrics are, however, very useful when assessing how good a particular model is in creating analytics insights with or with minimal contention to false positives or negatives. Vistisen et al. (2019) detail the process of determining the patterns of vital sign deterioration, which they argue are useful in proving that such signs produce meaningful alarms and timely action. Such metrics will assist DBAs in getting metrics that tell them where problems are and how more effective and practical solutions can enhance the DB system's efficiency.

Real time scenarios

Identifying Signs of the Raised Response Times and Latency

In this case, the machine learning monitoring model never stops looking at database quantitative indications like response time and I/O latency. The model also uncovers an increased response time and an equally increased latency covet, which could indicate a high database load or resource contestation. It has been analyzing current historical patterns and predicting that if no further action is to be taken, the response time will soon go beyond some threshold, thus possibly creating delays for the users. Again, this is an early warning to the database team to find out what has caused this and readjust resources before the system slows down again. Akin to the method mentioned by Henze et al. (2019), which says that predictive consequences can detect faint warning signs, similar things to database systems by key performance indicators are considered early signs that call for preventive action.

Real-Time Anomaly Detection of Single Unusual Error Rates

This scenario includes setting and monitoring real-time error rate levels inside a database, for example, for queries or connection failures. An error rate of one of these Centers is automatically calculated by an ML model that alerts when these rates rise above their normal fluctuation, indicating that a device might have malfunctioned, the network was congested, or the configurations were incorrect. In the later model, if the error frequency increases in current values, it measures the current values against historical performance to estimate the probability of a vital flaw. Such an alert helps administrators look into and solve the problem rather than go for the removal of data or spend hours troubleshooting. This idea aligns with the Newaz et al. (2019) Healthguard framework, where real-time anomaly detection is mentioned as the stable ML source's ability to immediately raise an alarm when particular data metrics begin to display abnormal characteristics.

Real-Time Feedback Loop for Continuous Model Improvement

With changing data conditions, continuation and improvement in matters of real-time feedback become central to issues of predictive accuracy. In this case, a feedback loop in the system updates the machine learning model at once by using the metrics of the current databases. For example, if new information concerning resource utilization, delay, or failure rate is received, the model shifts to these changes to refine its estimate. This approach is similar to Shi (2018), who argued that changes to the model should be made depending on the data situation. Such a model is useful in predicting issues under changing operating conditions as it constantly updates its knowledge from new data and remains relevant for database maintenance.

Graphs

Table 1 : Model Performance Metrics

Model	Accuracy	Precision	Recall	1-Score
Decision Tree	0.85	0.82	0.88	.85
Random Forest	0.90	0.88	0.91	.89
Neural Network	0.92	0.91	0.93	.92

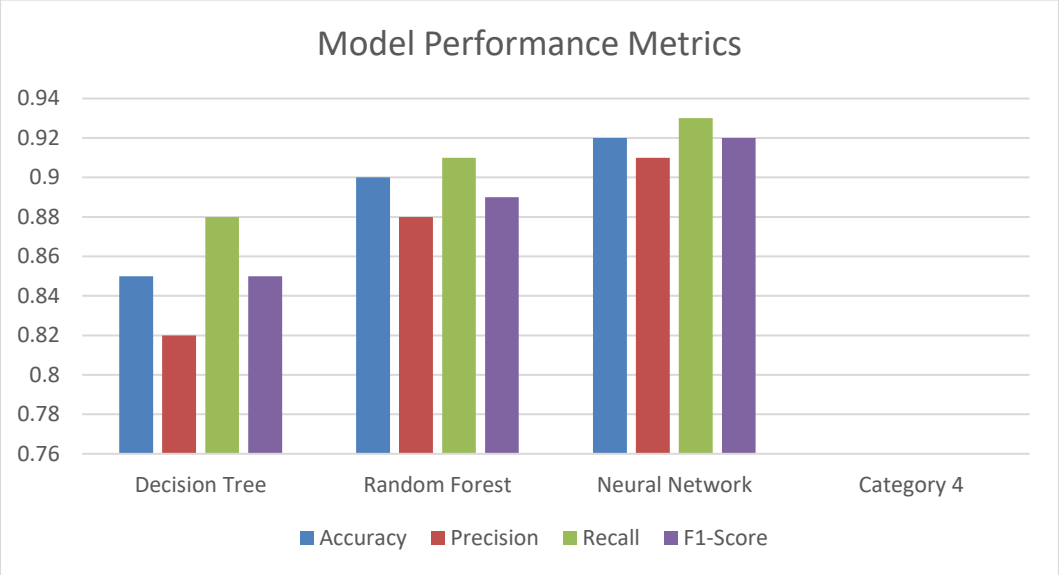

Fig 1 : Model Performance Metrics

Table 2 : Data Trends

Time	CPU Usage (%)	Memory Usage (%)	Disk Usage (%)	Error Rate (%)	Latency (ms)
00:00	35	60	70	0.02	15
01:00	40	62	72	0.03	17
02:00	38	61	71	0.03	16
03:00	45	65	75	0.04	20
04:00	50	70	78	0.05	22

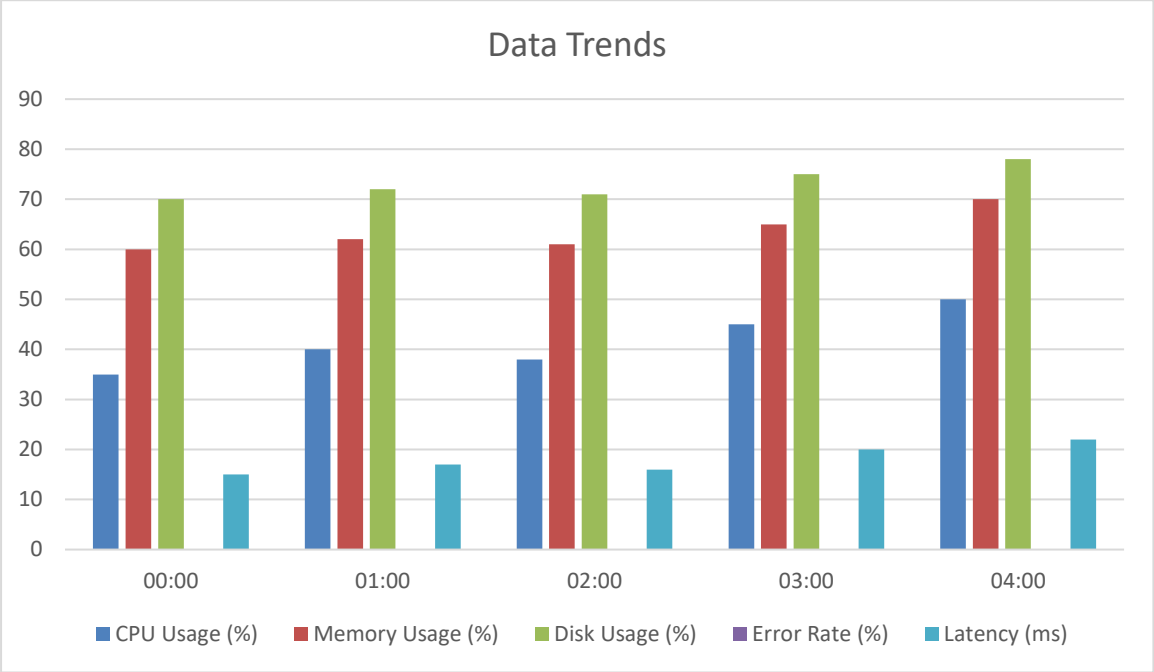

Fig 2 : Data Trends

Challenges and Solutions

Data Quality and Availability:

In its simplest form, PMM relies on high-quality data to develop accurate predictive maintenance models. However, some factors make the data collection process difficult: missing data, inconsistency, and different data types. It turns out that databases must be monitored for many qualitative and quantitative characteristics, such as CPU, memory, disks, and error rates, to predict health accurately. However, there is no common practice regarding data collection. Kasula (2017) pointed out that the specific needs for data differ in industries when applying machine learning. Because of this, appropriate mechanisms such as data validation checks and standard formats should be employed when entering the data. Also, accumulated data can be collected in a single location and may enhance model quality.

Model Selection and Complexity:

To be fit for predictive maintenance, choosing the right machine learning model has to be simple but effective. At the same time, decision trees are easily understandable and include. The algorithm's rules for making a decision and other models can be depth-wise sophisticated, including neural networks with a higher ability to reveal computationally intensive patterns. Xiang & Abouelyazid (2018) insist that the simpler models are more interpretable and amenable to practical use, while the models in question can discover fine-grained dependencies in the data. The approach to this problem is to employ the hybrid of simple and complex models as the final ensemble model. Readily, this blend can help reduce the errors within the predictions and properly restrict the usage of computational prowess.

False Positives and False Negatives:

A good predictive maintenance model should have low false positives, which will call for unnecessary maintenance, and low false negatives, which will omit important faults. According to Vistisen et al. (2019), it is crucial to remember that thresholds need to be adjusted one way or another to achieve better prediction without harming the balance. Adjusting these using real-time feedback leads to lower error levels and enhanced model performance and robustness. Even more, a comparison with historical data can be performed occasionally to get more accurate predictions and exclude superfluous and uninteresting alerts from being displayed.

Real-Time Processing Constraints:

Real-time predictive maintenance models need architectures to process large data inflows with little or no latencies. As Mandala (2019) shows, Kafka Streams can be used to deal with predictive workloads, which is crucial for dealing with big data in real time. The availability of scalable cloud solutions is maintained, along with the use of advanced data streaming capabilities such as Kafka, to provide prediction and alerts promptly to reduce potential delays in problem identification.

Privacy and Security:

Models for predictive maintenance have the problem of data being sensitive information. In their article, Newaz et al. (2019) recommended that guarding database information use anonymization plus higher access controls. Encryption, audit trails, and controlled access show that students' data is secure and adheres to the privacy policy act.

CONCLUSION

In this paper, Mobile Ad-Hoc Networks (MANETs) have revealed a dynamic and versatile field with immense potential for various applications. As we have explored throughout this review, MANETs offer a decentralized communication paradigm that enables wireless devices to collaborate and communicate without relying on a fixed infrastructure. This unique feature makes MANETs suitable for a wide range of scenarios, including military operations, disaster recovery, sensor networks, and even everyday communication in challenging environments.

However, it is crucial to acknowledge that MANETs also come with their fair share of challenges. The issues related to network security, routing protocols, scalability, and quality of service remain active areas of research and development. Furthermore, the ever-evolving landscape of wireless technologies and the increasing demand for reliable and efficient mobile communication solutions continue to drive innovation in the MANET domain.

As technology advances, we can anticipate the emergence of new MANET architectures and protocols that address these challenges more effectively. The future of MANETs holds promise for improved reliability, security, and scalability, making them an even more viable option for a broader spectrum of applications.

REFERENCES

[1]. Henze, D., Gorishti, K., Bruegge, B., & Simen, J. P. (2019, December). Audioforesight: A process model for audio predictive maintenance in industrial environments. In 2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA) (pp. 352-357). IEEE.

[2]. Kilaru, N., Cheemakurthi, S. K. M., & Gunnam, V. (2022). Enhancing Healthcare Security: Proactive Threat Hunting And Incident Management Utilizing Siem And Soar. International Journal of Computer Science and Mechatronics, 8(6), 20–25.

[3]. Mallreddy, S.R., Nunnaguppala, L.S.C., & Padamati, J.R. (2022). Ensuring Data Privacy with CRM AI: Investigating Customer Data Handling and Privacy Regulations. ResMilitaris. Vol.12(6). 3789-3799

[4]. Kilaru, N. B., & Cheemakurthi, S. K. M. (2023). Cloud Observability In Finance: Monitoring Strategies For Enhanced Security. NVEO-NATURAL VOLATILES & ESSENTIAL OILS Journal| NVEO, 10(1), 220-226.

[5]. Jaini, S., & Katikireddi, P. M. (2022). Applications of Generative AI in Healthcare. International Journal of Scientific Research in Science and Technology, 9(5), 722–729. https://doi.org/ https://doi.org/10.32628/IJSRST52211299

[6]. Vasa, Y., Mallreddy, S. R., & Jami, V. S. (2022). AUTOMATED MACHINE LEARNING FRAMEWORK USING LARGE LANGUAGE MODELS FOR FINANCIAL SECURITY IN CLOUD OBSERVABILITY. International Journal of Research and Analytical Reviews , 9(3), 183–190.

[7]. Belidhe, S. (2022). AI-Driven Governance for DevOps Compliance. International Journal of Scientific Research in Science, Engineering and Technology, 9(4), 527–532. https://doi.org/ https://doi.org/10.32628/IJSRSET221654

[8]. Gunnam, V. G., Kilaru, N. B., & Cheemakurthi, S. K. M. (2022). Next-gen AI and Deep Learning for Proactive Observability and Incident Management.Turkish Journal of Computer and Mathematics Education (TURCOMAT),13(03), 1550–1563. https://doi.org/10.61841/turcomat.v13i03.14765

[9]. Gunnam, V. G., Kilaru, N. B., & Cheemakurthi, S. K. M. (2022). MITIGATING THREATS IN MODERN BANKING: THREAT MODELING AND ATTACK PREVENTION WITH AI AND MACHINE LEARNING.Turkish Journal of Computer and Mathematics Education (TURCOMAT),13(03), 1564–1575. https://doi.org/10.61841/turcomat.v13i03.14766

[10]. Vasa, Y., & Singirikonda, P. (2022). Proactive Cyber Threat Hunting With AI: Predictive And Preventive Strategies. International Journal of Computer Science and Mechatronics, 8(3), 30–36.

[11]. Vasa, Y., Cheemakurthi, S. K. M., & Kilaru, N. B. (2022). Deep Learning Models For Fraud Detection In Modernized Banking Systems Cloud Computing Paradigm. International Journal of Advances in Engineering and Management, 4(6), 2774–2783. https://doi.org/10.35629/5252-040627742783

[12]. Katikireddi, P. M. (2022). Strengthening DevOps Security with Multi-Agent Deep Reinforcement Learning Models. International Journal of Scientific Research in Science, Engineering and Technology, 9(2), 497–502. https://doi.org/https://doi.org/10.32628/IJSRSET2411159

[13]. Mallreddy, S. R., & Vasa, Y. (2022). Autonomous Systems In Software Engineering: Reducing Human Error In Continuous Deployment Through Robotics And AI. NVEO - Natural Volatiles & Essential Oils, 9(1), 13653–13660. https://doi.org/https://doi.org/10.53555/nveo.v11i01.5765

[14]. Belidhe, S. (2022b). Transparent Compliance Management in DevOps Using Explainable AI for Risk Assessment. International Journal of Scientific Research in Computer Science, Engineering and Information Technology, 8(2), 547–552.

https://doi.org/https://doi.org/10.32628/CSEIT2541326

[15]. Gunnam, V. G., Kilaru, N. B., & Cheemakurthi, S. K. M. . (2022). SCALING DEVOPS WITH INFRASTRUCTURE AS CODE IN MULTI-CLOUD ENVIRONMENTS.Turkish Journal of Computer and Mathematics Education (TURCOMAT),13(2), 1189–1200. https://doi.org/10.61841/turcomat.v13i2.14764

[16]. Vasa, Y., & Mallreddy, S. R. (2022). Biotechnological Approaches To Software Health: Applying Bioinformatics And Machine Learning To Predict And Mitigate System Failures. Natural Volatiles & Essential Oils, 9(1), 13645–13652. https://doi.org/https://doi.org/10.53555/nveo.v9i2.5764

[17]. Katikireddi, P. M., & Jaini, S. (2022). IN GENERATIVE AI: ZERO-SHOT AND FEW-SHOT. International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT) , 8(1), 391–397. https://doi.org/https://doi.org/10.32628/CSEIT2390668

[18]. Nunnagupala, L. S. C. ., Mallreddy, S. R., & Padamati, J. R. . (2022). Achieving PCI Compliance with CRM Systems. Turkish Journal of Computer and Mathematics Education (TURCOMAT), 13(1), 529–535.

[19]. Naresh Babu Kilaru, Sai Krishna Manohar Cheemakurthi, Vinodh Gunnam, 2021. "SOAR Solutions in PCI Compliance: Orchestrating Incident Response for Regulatory Security"ESP Journal of Engineering & Technology Advancements 1(2): 78-84.

[20]. Jangampeta, S., Mallreddy, S.R., & Padamati, J.R. (2021). Anomaly Detection for Data Security in SIEM: Identifying Malicious Activity in Security Logs and User Sessions. 10(12), 295-298

[21]. Vasa, Y. (2021). Robustness and adversarial attacks on generative models. International Journal for Research Publication and Seminar, 12(3), 462–471. https://doi.org/10.36676/jrps.v12.i3.1537

[22]. Kilaru, N. B., Cheemakurthi, S. K. M., & Gunnam, V. (n.d.). Advanced Anomaly Detection In Banking: Detecting Emerging Threats Using Siem. International Journal of Computer Science and Mechatronics, 7(4), 28–33.

[23]. Sukender Reddy Mallreddy(2020).Cloud Data Security: Identifying Challenges and Implementing Solutions.JournalforEducators,TeachersandTrainers,Vol.11(1).96 -102.

[24]. Naresh Babu Kilaru. (2021). AUTOMATE DATA SCIENCE WORKFLOWS USING DATA ENGINEERING TECHNIQUES. International Journal for Research Publication and Seminar, 12(3), 521–530. https://doi.org/10.36676/jrps.v12.i3.1543

[25]. Gunnam, V., & Kilaru, N. B. (2021). Securing Pci Data: Cloud Security Best Practices And Innovations. Nveo, 8(3), 418–424. https://doi.org/https://doi.org/10.53555/nveo.v8i3.5760

[26]. Katikireddi, P. M., Singirikonda, P., & Vasa, Y. (2021). Revolutionizing DEVOPS with Quantum Computing: Accelerating CI/CD pipelines through Advanced Computational Techniques. Innovative Research Thoughts, 7(2), 97–103. https://doi.org/10.36676/irt.v7.i2.1482

[27]. Vasa, Y. (2021). Quantum Information Technologies in cybersecurity: Developing unbreakable encryption for continuous integration environments. International Journal for Research Publication and Seminar, 12(2), 482–490. https://doi.org/10.36676/jrps.v12.i2.1539

[28]. Jangampeta, S., Mallreddy, S. R., & Padamati, J. R. (2021). Data Security: Safeguarding the Digital Lifeline in an Era of Growing Threats. International Journal for Innovative Engineering and Management Research, 10(4), 630-632.

[29]. Vasa, Y., Jaini, S., & Singirikonda, P. (2021). Design Scalable Data Pipelines For Ai Applications. NVEO - Natural Volatiles & Essential Oils, 8(1), 215–221. https://doi.org/https://doi.org/10.53555/nveo.v8i1.5772

[30]. Singirikonda, P., Jaini, S., & Vasa, Y. (2021). Develop Solutions To Detect And Mitigate Data Quality Issues In ML Models. NVEO - Natural Volatiles & Essential Oils, 8(4), 16968–16973. https://doi.org/https://doi.org/10.53555/nveo.v8i4.5771

[31]. Vasa, Y. (2021). Develop Explainable AI (XAI) Solutions For Data Engineers. NVEO - Natural Volatiles & Essential Oils, 8(3), 425–432. https://doi.org/https://doi.org/10.53555/nveo.v8i3.5769

[32]. Kilaru, N. B., & Cheemakurthi, S. K. M. (2021). Techniques For Feature Engineering To Improve Ml Model Accuracy. NVEO-NATURAL VOLATILES & ESSENTIAL OILS Journal| NVEO, 194-200.

[33]. Singirikonda, P., Katikireddi, P. M., & Jaini, S. (2021). Cybersecurity In Devops: Integrating Data Privacy And Ai-Powered Threat Detection For Continuous Delivery. NVEO - Natural Volatiles & Essential Oils, 8(2), 215–216. https://doi.org/https://doi.org/10.53555/nveo.v8i2.5770