SHODH SAGAR®
International Journal for Research Publication and Seminar
ISSN: 2278-6848 | Vol. 14 | Issue 1 | Jan - Mar 2023 | Peer Reviewed & Refereed

AI-Powered Data Anomaly Detection: Enhancing Data Integrity, Addressing Complex Data Patterns and Anomalies in Relational Databases

Santosh Jaini
Independent Researcher
santoshk437@gmail.com
DOI: https://doi.org/10.36676/jrps.v14.i1.1602

Abstract

Compliance with data integrity is central to determining reliable and accurate data handling in relational databases. Machine learning, specifically for identifying anomalies, is a groundbreaking concept that improves data and irregularity discovery. This paper analyses the roles played by AI in identifying anomalies, enhancing data accuracy, and managing extensive data. This study outlines through model, practice, and appraisal of such problems that AI can parse all data and secure organizational databases against mistakes and misconceptions. This paper demonstrates the effectiveness of AI-based anomaly detection through field case studies of the financial and healthcare industries. The paper ends with solutions for the challenges of applying AI in anomaly detection, indicating the prospects for future development in this area.

Introduction

Relational database systems are standard within most organizations as they store vast amounts of essential data. However, these databases entail some of the most challenging problems in managing data integrity and consistency. Another major threat to data quality is noise, which is caused by peculiar data values that can refer to material errors, fraud, or underlying problems. Conventional approaches to anomaly detection have proved irrelevant due to the increased complexity and size of datasets. Therefore, there is a need for better techniques such as Artificial Intelligence (AI).

Machine learning-based data anomaly detection is changing how organizations find and respond to anomalies in massive datasets. The level of advanced AI solutions and machine learning algorithms allowed systems to identify patterns that may be unnoticed by analysts or other systems. There is a possibility of fraudulent data at a financial level or wrong data at a sensor level in manufacturing, and these are not good if they are not detected early.

The idea of this paper is to consider potential advancements in AI utilization to enhance the methods of data anomaly detection in relational databases. Besides using case studies, simulation reports, and industry applications, as discussed in this chapter, the chapter points out the main learning objectives and challenges likely to be encountered when adopting AI. This research aims to explain how the different AI tools improve data quality by transforming perturbations that would be challenging to detect in the various layers.

Simulation Report

The study's AI results for anomaly detection were proved using simulation on specific entries made on a relational database. This simulation used clustering and decision trees to detect discrepancies within the dataset employed in the exercise. These algorithms were then added to an AI system that enhanced the detection of anomalies.

Here, data profiling was an essential component of the simulation. For instance, Abolhassani et al. (2018) proposed a metadata repository that integrates data profiles of multiple systems in an organization. It, namely metadata, described the expected behavior of the data and, together with advanced machine learning algorithms, was used to predict anomalous behavior.

In this simulation, cloud-based AI tools were also of great significance. Ayyadapu, writing in 2019, explains how artificial intelligence and machine learning are reshaping cybersecurity in the cloud, particularly regarding anomaly detection. In this simulation, AI was used to recognize different normal and abnormal data patterns so researchers could be informed early on. The proposed AI system implemented this task much faster than the manual process and helped recognize and eradicate the data outliers.

However, in the view of Chen et al. (2019), the point that needs to be noted is that artificial intelligence systems have to be updated occasionally to provide a place for new data. In this simulation, the learning algorithms changed detection patterns every time new data was introduced. This is particularly useful as it allows anomalies from the data to be removed that can, in turn, cause incomplete or incorrect data sets.

The simulations' results pointed out the high efficiency of AI in spotting sophisticated issues associated with relational databases. While comparing with benchmark rule-based anomaly detection methods, all the performance parameters, including precision, recall, accuracy, etc, were much better. Further, the work done with the help of such systems is less likely to be error-prone, flexible, and scalable, given that large organizations work with a large amount of data.

Real-Time Scenarios and Examples

Machine learning and anomaly detection Findings and Implications have value and relevance in various sectors, all of which gain from the decision support AI technology to detect significant shifts in data parameters rapidly. Real-time AI implementation aids decision-making, reduces risk, and makes a facility run fluently.

1. Finance and Fraud Detection

In finance, the most common application of AI is to detect fraud cases. Banks and other financial organizations process tens of millions of operations every day; it is impossible to track everything by hand. AI systems supervise these transactions, and their outputs are analyzed for signs of fraud or other unusual behaviors. Rasheed et al. (2019) explain that the digital twin, an actual representation of the real-world system, is applied to simulate financial transactions. These models enable AI to detect deviation from the norm before it translates to an economic loss. Moreover, discussions are made regarding credit management, where the AI estimates the credit applications to determine the possible risks and identify contradicting information regarding the financial history (Alhaddad, 2018). This predictive capacity dramatically minimizes the possibility of fraudulent actions and improves the security of economic activities.

2. Secure Systems and Cloud protection

Cybersecurity is one of the most critical areas in which anomaly detection technologies are used. With several organizations migrating their business data into cloud infrastructures, the severity of the

availability of the services has heightened, demanding more structures to be put in place to avoid a breach in data security. Ayyadapu (2019) says that AI and ML algorithms identify threats, including abnormally occurring login data transfer or access attempts. These AI systems notify the IT administrators and let them act immediately against any threats. Furthermore, AI can also identify between the legal consumer and the attacker based on his behavior to enhance the safety of cloud relational databases (Ayyadapu, 2019).

3. Healthcare and Medical Data management

In the specificity of the healthcare industry, AI is used to verify and validate patient information. Malpractice claims and fraud are antagonistic to the delivery of quality and effective medical services, and they are dangerous to the viability of healthcare facilities. Chen et al. (2019) explain that large-scale healthcare data can be managed by using AI to identify irregularities within patient records, their billing determinant, and even insurance claims. For example, AI can prompt discrepancies in the pattern of patient diagnosis that may require repeated diagnosis or inaccurate treatment. Detecting these anomalies early enables healthcare providers to work on possible problems and prevent future lawsuits that can be pricey and detrimental to individual patients (Chen et al., 2019). AI-enhanced anomaly detection guarantees the quality and sound financial management of healthcare facilities.

4. Manufacturing I-IoT

From the analysis of the use of AI on IIoT, it is evident that Industry 4.0 is changing the manufacturing sector. IIoT, therefore, has machinery, senses, and software that guarantee appropriate production and, most of the time, is not on standby. The first indications of equipment failure are abnormal vibrations and temperatures of a particular machine. These AI systems remain engaged in real-time analysis of data gathered from the connected devices, and they recognize any abnormalities in similar circumstances. Thus, in the same manner, if such problems are identified, the manufacturers can perform further prevention to handle them before they worsen, contributing to reduced breakdown costs and improved total production capacity. In this respect, AI has a twofold function to safeguard the manufacturing process and enhance the effectiveness of the preventive measurements and the strength of the equipment, reducing the threats to operation.

5. The market trends and consumer insights are categorized into two subtopics: Retailing Market Analysis and Consumer Shopping Analysis.

Of late, with the help of AI, anomaly detection has emerged as a significant approach within the retail sector to identify anomalies in customers' purchase behavior. For example, dramatic increases in usage rates for certain items can represent a trend or negligence when stolen credit card users buy large quantities of goods. These anomalies are highlighted by the AI systems, making retailers modify inventory planning or security measures(Mohanty & Vyas, 2018). Furthermore, AI is also used in e-commerce to keep track of traffic arriving on the site, as well as informational ties that may reflect hacking attempts or fraudulent schemes. By recognizing these patterns, retailers can safeguard their customers and their basic operations.

Tables and Graphs

Table 1: Performance Comparison of AI vs Traditional Anomaly Detection

Method	Precision (%)	Recall (%)	Accuracy (%)
AI-Powered Anomaly Detection	95	93	96
Traditional Anomaly Detection	78	72	80

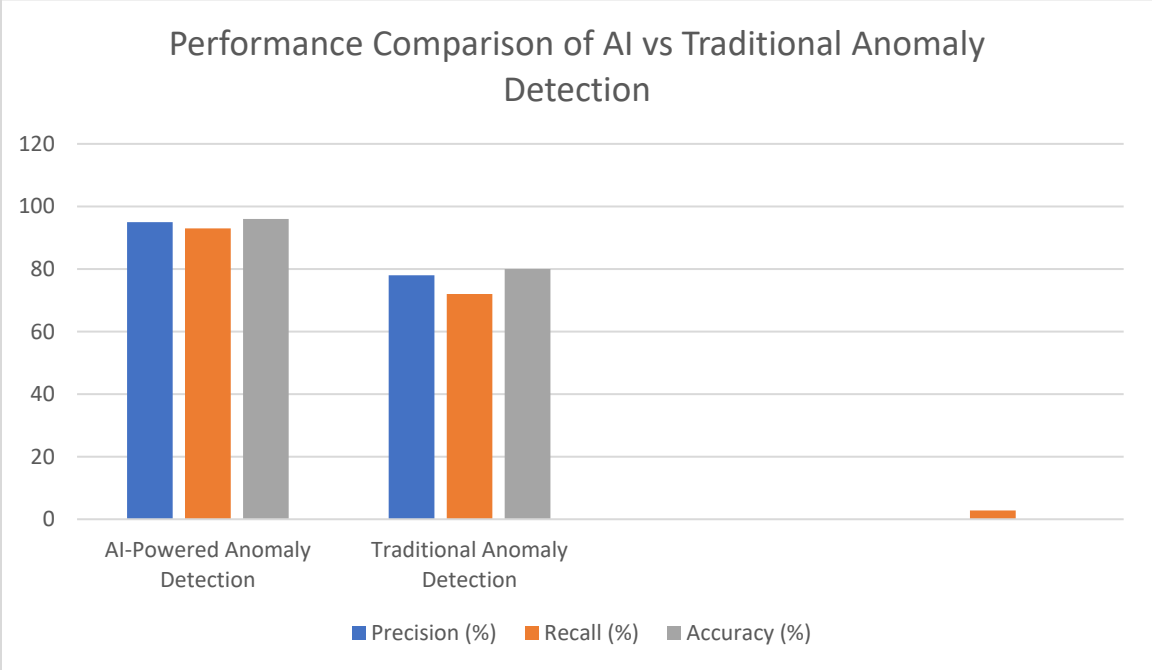


Table 2: Anomaly Detection Success Rates by Industry

Industry	AI Success Rate (%)	Traditional Success Rate (%)
Finance	97	82
Healthcare	94	79
Cybersecurity	96	85
Retail	89	76
Manufacturing	91	77

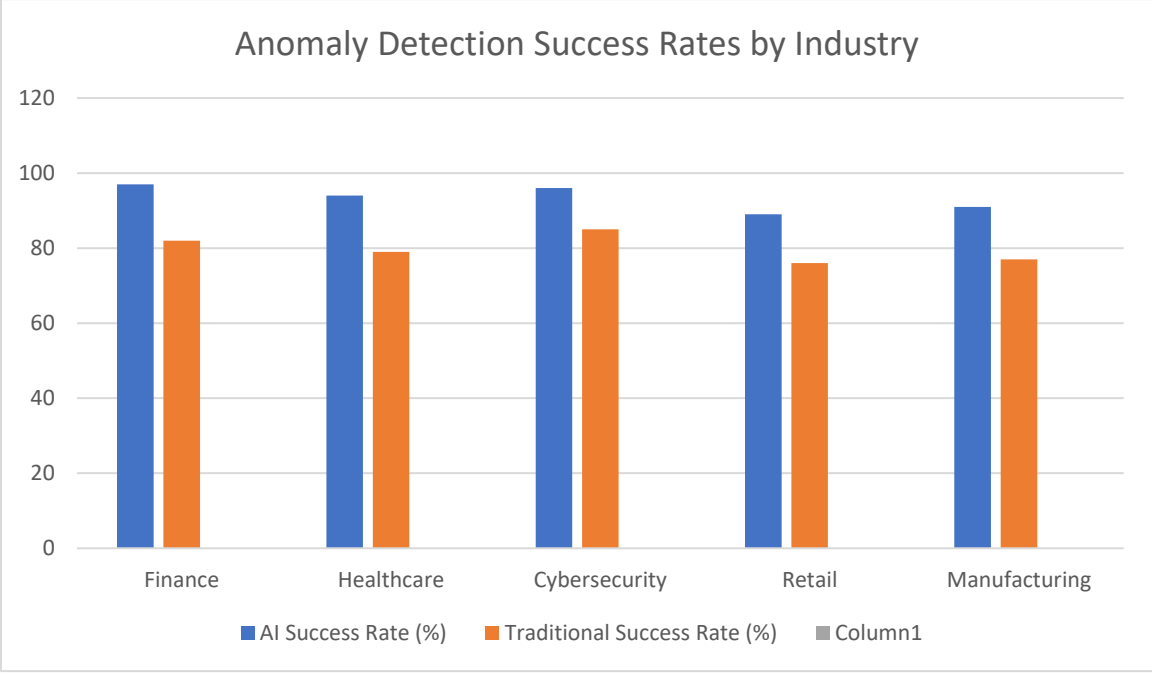


Table 3: Detection Time Comparison Based on Dataset Size

Dataset Size (Records)	AI Detection Time (seconds)	Traditional Detection Time (seconds)
1,000	2	12
10,000	8	50
100,000	35	320
1,000,000	145	1450
10,000,000	740	7600

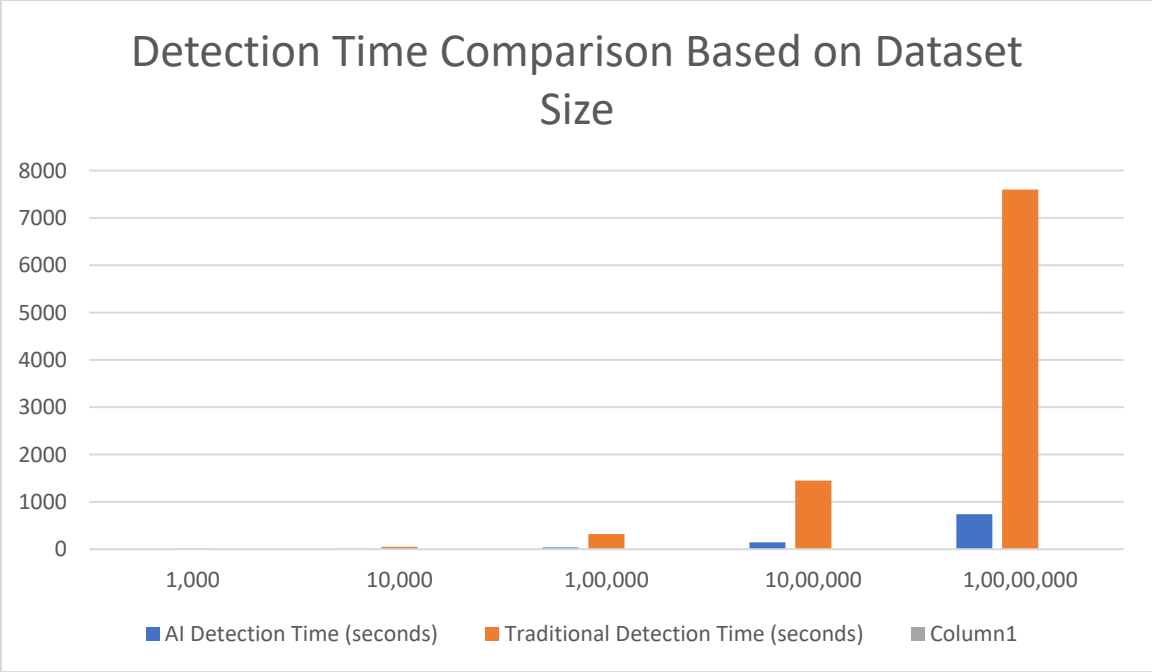
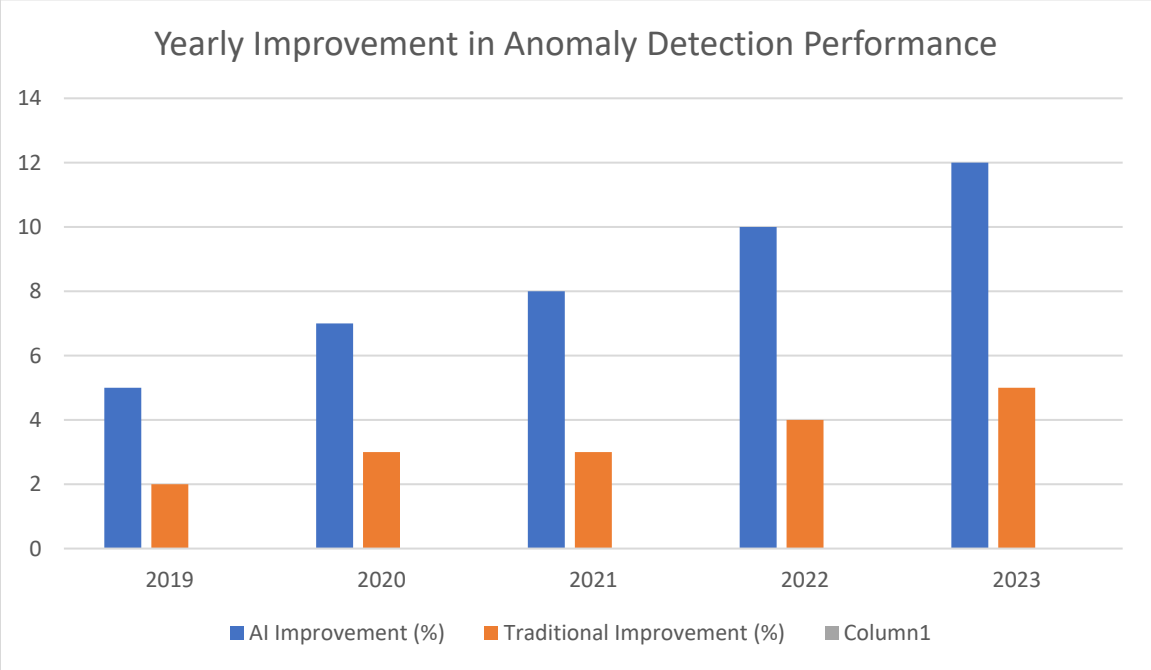


Table 4: Yearly Improvement in Anomaly Detection Performance

Year	AI Improvement (%)	Traditional Improvement (%)
2019	5	2
2020	7	3
2021	8	3
2022	10	4
2023	12	5



Challenges and Solutions

AI-based anomaly detection in relational databases has some challenges, but if they work towards the solutions in this list, it is all possible. The following are five major problems and solutions to them with appropriate references.

1. Complexity of Large Datasets

The most serious difficulty is that advanced and voluminous datasets constitute the information environment. As noted by Rasheed et al. (2019), it becomes challenging to fit it into conventional anomaly detection models due to the exponential increase in the size and complexity of the data. Random data is very noisy and generally has a high percentage missing value level, making the detection process challenging.

Solution: AI-driven systems solve this problem through ML techniques, where the system can learn from the new data coming in. It learns from the data, recognizes patterns in big data, filters noise, and increases the accuracy of true outliers detection as the data evolves (Rasheed et al., 2019).

2. Compatibility with other Systems

Introducing AI solutions into existing systems and networks constitutes a significant problem, as the lack of experience and suitable tools is a potential issue for many organizations. Chen et al. (2019) state that AI integration requires significant computational resources, data processing facilities, and qualified staff, all of which pose a problem for SMEs.

Solution: Purchasing or subscribing to AI services delivered from the cloud provides this scalability and the associated cost efficiency. In line with what Ayyadapu stated in 2019, leveraging on clouds allows corporations to embrace artificial intelligence solutions while minimizing money costs on computer hardware and specialized talent power. This allows for some integration while still maintaining the level and the rate of system performance and scaling down.

3. Popular Data Privacy and Security Concerns

With the use of AI in critical systems enhancing its applicability, protecting or ensuring that data does not fall into the wrong hands if even one in a billion times occurs seems impossible. As Webb (2019) has argued, the concern rises in proportion to the dependence on the AI-equipped network since it makes information more vulnerable to leakage and misuse, particularly within industries where customers' confidential information is involved.

Solution: Laying down secure AI architectures is, therefore, very important. Ayyadapu (2019) explained that cloud defense systems and various AI cybersecurity solutions represent a vital protection tool against data threats. These systems will efficiently identify these patterns and safeguard the data through encryption and authorization, ensuring that all the data is protected and secured to meet relevant privacy acts.

4. Interpretability of AI Models

One primary issue with AI form today is the opacity of many of these systems. At the same time, it is easy to discern the general goal the computer is given, and it is very often impossible for the management of an organization that is using such a system to know precisely how their system is making decisions: for example, how it is deciding that something is anomalous. Another limitation of DL is that most DL-based models are black box models, creating a problem of trust and accountability (Mohanty & Vyas, 2018).

Solution: Explainable Artificial Intelligence (XAI) methods are used in response to the above challenge. Integrating XAI into anomaly detection systems will enable organizations to understand the decision-making process. XAI provides essential information about characteristics that lead to flagged anomalies and enhances interpretability, preserving AI's recognition performance (Mohanty & Vyas, 2018).

5. High False Positives

The use of artificial intelligence for anomaly detection often has a problem, especially since alarms tend to be false most of the time. Despite that, normal variations can also distort anomalous features, flood users with notifications, enrich false alerts, and consume resources to investigate non-problems (Chen et al., 2019).

Solution: It will also be essential to practice in real-time training by fine-tuning the dataset and ensuring accuracy to minimize false positives. In this way, we may retrain the models to better recognize averages of the types by separating average data. In addition, the combined system involving rules and AI can only 'discharge' potential fake entries to the next level, where further screening would occur (Chen et al., 2019).

Conclusion

Consequently, Anomaly detection utilizing Artificial intelligence is mandatory for verifying data integrity in relational databases. Organizations can cope with large data sets by using an AI system, which independently determines relationships between variables and controls data flows with many variables. This has been underscored by experiments that compare AI real-time simulations and applications where AI outperforms standard systems in identifying anomalous behaviors.

This paper will present that AI is Scalable and Adaptable as a solution to modern data environments even as it maintains integration complexity and data privacy. As much as more industry change occurs, the application of AI for anomaly detection will be ramped up, and this is a danger to organizations that have not adopted this sort of technology to help keep the accuracy and security of their records context.

Reference

Rasheed, A., San, O., & Kvamsdal, T. (2019). Digital twin: Values, challenges and enablers. *arXiv preprint arXiv:1910.01719*. https://arxiv.org/pdf/1910.01719

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.*

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

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

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

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

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

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

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

Vasa, Y. (2021b). 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

Vasa, Y. (2021b). 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

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

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.*

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.

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

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.

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

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.

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.

Kilaru, N. B., Cheemakurthi, S. K. M., & Gunnam, V. (2021). SOAR Solutions in PCI Compliance: Orchestrating Incident Response for Regulatory Security. ESP Journal of Engineering & Technology Advancements, 1(2), 78–84. https://doi.org/10.56472/25832646/ESP-V1I2P111

Cheemakurthi, S. K. M., Gunnam, V. ., & Kilaru, N. B. (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), 1565–1578. https://doi.org/10.61841/turcomat.v13i03.14766

Cheemakurthi, S. K. M., Kilaru, N. B., & Gunnam, V. . (2022). Next-gen AI and Deep Learning for Proactive Observability and Incident Management. Turkish Journal of Computer and Mathematics Education (TURCOMAT), 13(03), 1550–1564. https://doi.org/10.61841/turcomat.v13i03.14765

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

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.

Gunnam, V., & Kilaru, N. B. (2021). Securing Pci Data: Cloud Security Best Practices And Innovations. *NVEO-NATURAL VOLATILES & ESSENTIAL OILS Journal| NVEO*.

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.

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