http://www.jctjournals.com

ISSN(ONLINE):2278 - 3814

Volume-13, Issue-07, July 2024

JOURNAL OF COMPUTING TECHNOLOGIES (JCT)

International Journal

Page Number: 01-07

A Literature Survey on a New Approach to Data Analysis Using Machine Learning

[1]Pinky Mishra, [2]Prof. Manish Kumar Singhal,
[1]M.Tech Scholar, [2]Associate Professor & H.O.D,
[1,2]Department of Information Technology (IT)
[1,2]NRI Institution of Information Science & Technology (NIIST), BHOPAL (M.P), INDIA,
[1]pinky93.11oct@gmail.com [2]*manishsinghal@gmail.com*

***Abstract*— In this survey paper discuss the machine learning (ML) has emerged as a transformative tool for data analysis across various domains. This survey paper explores the evolution of data analysis techniques, with a focus on how modern machine learning algorithms are reshaping traditional approaches. The paper examines various ML methods, including supervised, unsupervised, and reinforcement learning, and their applications in fields such as healthcare, finance, and marketing. Emphasis is placed on the benefits of automation, scalability, and predictive accuracy that machine learning brings to complex data sets. Additionally, the survey highlights challenges such as data quality, algorithm selection, and interpretability, and discusses potential future directions for research in improving ML-based data analysis. The study aims to provide a comprehensive understanding of the state-of-the-art in data analysis using machine learning and to inspire new approaches to tackling emerging data challenges.In this review paper discuss aims to provide a comprehensive understanding of the state-of-the-art in data analysis using machine learning and to inspire new approaches to tackling emerging data challenges.**

***Keywords*— Analytics; Cyber security; Machine learning, etc.**

I. Introduction

Data analysis has evolved with the integration of machine learning techniques, offering a transformative approach to uncovering insights from large and complex datasets. Traditional data analysis methods, often reliant on manual processes and static models, struggle to handle the increasing volume and variety of data generated in today's digital world. The new approach leverages machine learning algorithms to automatically identify patterns, predict outcomes, and adapt to changing data environments. By automating the analysis process, machine learning enables faster, more accurate decision-making, reduces human bias, and uncovers deeper insights that would otherwise remain hidden. This shift toward machine learning-driven analysis represents a powerful tool for industries looking to gain a competitive edge by making data-driven decisions more efficiently and effectively.

Data analysis has undergone a significant transformation with the introduction of machine learning techniques. Traditional methods of data analysis often relied on statistical models and human interpretation, which, while effective, were limited in their ability to handle large, complex datasets. Machine learning offers a new approach, enabling automated discovery of patterns and insights from vast amounts of data. This method leverages algorithms that learn from the data itself, improving their predictions and decisions over time. By incorporating machine learning into data analysis, businesses and researchers can uncover deeper insights, make more accurate predictions, and enhance decision-making processes. This shift not only accelerates data-driven innovation but also opens up new possibilities for solving complex problems across various fields, from healthcare and finance to marketing and engineering.

New approach to data analysis through machine learning introduces several key advantages. One of the most notable is its ability to handle unstructured data, such as text, images, and videos, which traditional statistical methods struggle to analyze effectively. Additionally, machine learning models are capable of adapting to changing data patterns, making them highly suited for real-time applications like predictive maintenance, fraud detection, and personalized recommendations.

Another important aspect is scalability. Machine learning algorithms can process and analyze vast amounts of data at speeds that far surpass human capabilities. This allows organizations to not only gain insights faster but also apply these insights at scale, leading to more dynamic and data-driven strategies.

Furthermore, the integration of deep learning, a subset of machine learning, has enabled even more advanced forms of data analysis. Deep learning models can discern intricate relationships within datasets, unlocking insights that were previously inaccessible. This is particularly beneficial in fields like image recognition, natural language processing, and genomics, where understanding complex patterns is crucial.

As machine learning continues to evolve, the possibilities for data analysis will expand further, enabling more sophisticated decision-making, improving operational efficiencies, and fostering innovation across industries.

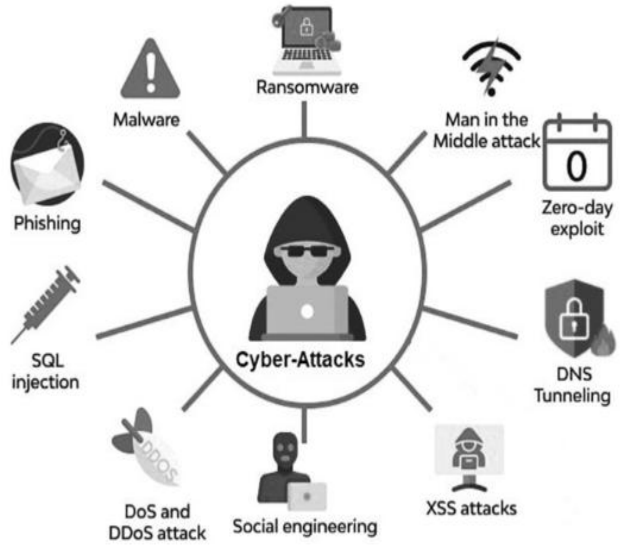

Fig. 1: New Approach to Data Analysis Using Machine Learning

Despite its many advantages, the adoption of machine learning in data analysis is not without challenges. One of the primary hurdles is the need for large, high-quality datasets to train machine learning models effectively. Incomplete or biased data can lead to inaccurate predictions and skewed insights, which may affect critical business or research decisions. Moreover, developing, training, and maintaining machine learning models requires specialized expertise in data science and programming, creating a barrier for some organizations lacking these resources

II. LITERATURE REVIEW

Ugochukwu Ikechukwu Okoli, et.al (2024),Author are presented The use of ML in proactive defense mechanisms marks a paradigm shift in cyber security. By harnessing the power of advanced analytics and pattern recognition, ML contributes to a more intelligent, adaptive, and efficient defence posture. The ability to detect subtle anomalies, automate responses, and continuously learn from evolving threats positions ML as a cornerstone in modern cyber security strategies. Machine learning and artificial intelligence algorithms can be leveraged upon to analyze large volumes of data and identify patterns, anomalies, and previously unknown threats. Machine learning models can be trained on historical data to recognize known threat patterns and improve accuracy over time [01].

Fatima Al-Quayed,et.al,(2024), Author are analysis performance Real-time communication between machines, sensors, devices, and people makes it easier to transmit the data needed to make decisions. Informed decision-making is empowered by the comprehensive insights and analytics made possible by this connectedness in conjunction with information transparency. Industry 4.0-based wireless sensor networks (WSNs) are an integral part of modern industrial operations however, these networks face escalating cyber security threats. These networks are always vulnerable to cyber-attacks as they continuously collect data and optimize processes. Increased connections make people more susceptible to cyber attacks, necessitating the use of strong cyber security measures to protect sensitive data. This study proposes a predictive framework intended to intelligently prioritize and prevent cyber security intrusions on WSNs in Industry 4.0. The proposed framework enhances the cyber security of WSNs in Industry 4.0 using a multi-criteria approach. It implements machine-learning and deep-learning algorithms for cyber security intrusion detection in WSNs of Industry 4.0 and provides prevention by assigning priorities to the threats based on the situation and nature of the attacks[02].

Priya Thapa, et.al, (2024),the use of machine learning techniques for anomaly detection to enhance cyber security in cloud computing. We provide background on cloud computing architectures, cyber threats, and anomaly detection. We then comprehensively survey state-of-the-art machine learning methods for anomaly detection in cloud environments, including supervised, unsupervised, and hybrid approaches. Specific techniques covered include neural networks, support vector machines, clustering, and ensemble methods. We analyze the strengths and limitations of these techniques, and provide recommendations for selecting appropriate algorithms based on factors like labelled data availability and detection goals. Challenges and open research questions in deploying machine learning for cloud security are discussed. We argue that AI-enhanced anomaly detection has excellent potential to identify novel attack patterns and improve resilience against continually evolving threats. Our analysis aims to provide guidance for researchers and practitioners developing the next generation of intelligent cyber defense systems[03].

İsa Avcı et.al, (2023),a significant increase in cyber-attacks, targeting organizations, institutions and even individuals. With the tremendous technological developments, the skill used by attackers, has increased and traditional IDS, can no longer detect sophisticated cyber-attacks. This required finding new, advanced tools to detect these destructive and expensive attacks. After the great successes of ML and DL techniques, in various fields, there have been many studies that use ML techniques in building IDS systems. This study presents an IDS based on feature selection techniques and ML techniques, for intrusion detection. The proposed model achieved promising results, as the RF technique achieved an accuracy of 99.72%, superior to other techniques in this work and related works. Having an intelligent system capable of detecting intrusion, helps significantly in maintaining the privacy and security of users. In this work, the focus is only on whether the traffic is malicious or benign [04].

Asad Yaseen, et. al (2023),a theoretical framework for network anomaly detection in cyber security, emphasizing the integration of adaptive machine learning models, ensemble techniques, and advanced feature engineering. The adaptability of machine learning models enables dynamic responsiveness to emerging cyber threats, forming a foundation for a resilient anomaly detection system. Ensemble techniques, particularly the incorporation of Random Forests, enhance the framework's robustness by amalgamating strengths from diverse models, mitigating false positives and negatives. Advanced feature engineering, coupled with deep learning architectures, contributes to a nuanced understanding of intricate patterns within network traffic. The theoretical exploration encounters challenges in quantifying performance gains, integration complexities, and data privacy concerns. Addressing these challenges is critical for refining and fortifying the proposed framework, ensuring its applicability and effectiveness in real-world cyber security scenarios. The significance of the framework lies in addressing existing gaps in network anomaly detection theories and advancing the understanding of machine learning's role in cyber security[05].

Sungwook Ryu, et.al, (2023), Author are study Each cyber security technology is related to cyber security managers, and system operation and developers, both directly and indirectly. The highlight of this study is that understanding the changes and trends in the cyber security convergence fields is the most effective way to respond to cyber security threats and dangers in a preemptive manner and it also plays an important role in inducing new paradigms in the next generation cyber security technology development. Inducing a new paradigm of cyber security technology development can prevent the availability of computer systems from being compromised by cyber attack [06].

Mostofa Ahsan, et.al, (2022),Machine learning is of rising importance in cyber security. The primary objective of applying machine learning in cyber security is to make the process of malware detection more actionable, scalable and effective than traditional approaches, which require human intervention. The cyber security domain involves machine learning challenges that require efficient methodical and theoretical handling. Several machine learning and statistical methods, such as deep learning, support vector machines and Bayesian classification, among others, have proven effective in mitigating cyber-attacks. The detection of hidden trends and insights from network data and building of a corresponding data-driven machine learning model to prevent these attacks is vital to design intelligent security systems. In this survey, the focus is on the machine learning techniques that have been implemented on cyber security data to make these systems secure. Existing cyber security threats and how machine learning techniques have been used to mitigate these threats have been discussed. The shortcomings of these state-of-the-art models and how attack patterns have evolved over the past decade have also been presented. Our goal is to assess how effective these machine learning techniques are against the ever-increasing threat of malware that plagues our online community[07].

Ziaul Hasan, (2022) – This paper reviews has presented the Data mining and machine learning (ML) methods are used more than ever in cyber security. The use of machine learning (ML) is one of the potential solutions that may be successful against zero-day attacks, starting with categorising IP traffic and filtering harmful traffic for intrusion detection. In this field, certain published systematic reviews were taken into consideration. Recent systematic reviews may incorporate older and more recent works in the topic of investigation.. Both security professionals and hackers use data mining capabilities. Applications for data mining may be used to analyze programme activity, surfing patterns, and other factors to identify potential cyber-attacks in the future. The new study uses statistical traffic features, ML, and data mining approaches. This research performs a concentrated literature review on machine learning and its usage in cyber analytics for email filtering, traffic categorization, and intrusion detection. Each approach was identified, and a summary was provided based on the relevancy and quantity of citations. Some well known datasets are also discussed since they are a crucial component of ML techniques. On when to utilize a certain algorithm is also offered some advice. Four ML algorithms have been evaluated on MODBUS data gathered from a gas pipeline**[08]**

Rahbar Ahsan., (2022) - Cyber security has become a significant issue. Machine learning algorithms are known to help identify cyber attacks such as network intrusion. However, common network intrusion datasets are negatively affected by class imbalance: the normal traffic behaviour constitutes most of the dataset, whereas intrusion traffic behaviour forms a significantly smaller portion. A comparative evaluation of the performance is conducted of several classical machine learning algorithms, as well as deep learning algorithms, on the well- known National Security Lab Knowledge Discovery and Data Mining dataset for intrusion detection. More specifically, two variants of a fully connected neural network, one with an auto-encoder and one without, have been implemented to compare their performance against seven classical machine learning algorithms. A voting classifier is also proposed to combine the decisions of these nine machine learning algorithms. All of the models are tested in combination with three different resampling techniques: oversampling, under sampling, and hybrid sampling **[09]**

Dinesh Kalla., (2021)-In this paper, this research comprehensively validates the potential of supervised machine learning models for accurate and reliable malware classification, as demonstrated by the high performance of the Random Forest ensemble technique. Trained on a dataset of over 10,000 portable executable samples engineered with informative static features, the classifier achieves an exceptional accuracy of 99% in predicting malware threats. Augmented by explanatory feature importance plots, the Random Forest model indicates the most distinguishing file properties that steer its predictions, enhancing interpretability. The fine-grained classification report further substantiates precision and recall exceeding 99% across malware and benign app categories. Operationalization is achieved by persisting the model using serialization to enable real-time warnings against new unforeseen threats. While this study focuses solely on static analysis, future work can fuse dynamic behavioural traces for even more resilient defence **[10]**

III. CHALLENGES OF DATA ANALYSIS USING MACHINE LEARNING

Adopting a new approach to data analysis through machine learning presents several challenges. Firstly, the complexity of machine learning algorithms often requires specialized knowledge and skills, which can be a barrier for teams accustomed to traditional data analysis methods. Additionally, the quality and quantity of data play a crucial role; incomplete, biased, or noisy data can significantly impact the performance and reliability of machine learning models. Another challenge is the interpretability of machine learning models, as many are considered "black boxes" that provide little insight into their decision-making process, making it difficult to understand and trust their outcomes. Moreover, integrating machine learning models into existing systems and workflows can be resource-intensive and may require significant changes to infrastructure. Lastly, ongoing maintenance and tuning of models are necessary to ensure they remain effective as new data emerges and underlying patterns evolve, which can be both time-consuming and costly.

Ethical and privacy concerns pose additional hurdles when implementing machine learning in data analysis. Ensuring that models do not inadvertently perpetuate biases or discriminate against certain groups is a critical issue that requires careful consideration and proactive measures. Additionally, safeguarding sensitive data from breaches or misuse is paramount, necessitating robust security protocols and compliance with data protection regulations. The evolving nature of machine learning technologies also means that organizations must stay abreast of rapid advancements and emerging best practices, which can be challenging and resource-draining. Finally, aligning machine learning initiatives with strategic business objectives demands a clear understanding of how these technologies can create value and drive insights, necessitating careful planning and alignment between technical and business teams. Addressing these challenges effectively requires a combination of technical expertise, strategic foresight, and a commitment to ethical practices.

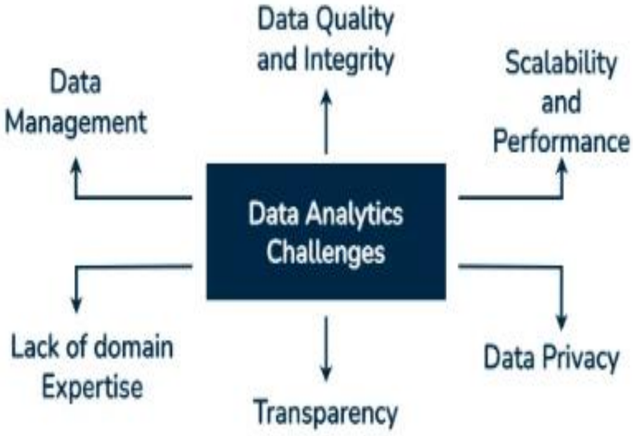
Data Quality and Integrity
Data Management
Scalability and Performance
Data Analytics Challenges
Lack of domain Expertise
Data Privacy
Transparency

Fig 2 Data Analysis Challenges

Data analysis using machine learning presents several challenges that can impact the effectiveness and accuracy of the results. One primary challenge is the quality and quantity of data; machine learning models require large volumes of high-quality data to perform well, and poor-quality or insufficient data can lead to inaccurate or biased outcomes. Another issue is the selection of appropriate algorithms and models; choosing the wrong approach or failing to properly tune parameters can result in suboptimal performance. Additionally, data privacy and security concerns are critical, as handling sensitive information requires strict adherence to regulations and safeguards. Moreover, interpretability and explain ability of machine learning models can be problematic, as complex models may produce results that are difficult to understand or explain to stakeholders. Lastly, the need for ongoing maintenance and updating of models to adapt to new data or changing conditions is essential but can be resource-intensive. Addressing these challenges effectively is crucial for leveraging machine learning in data analysis.

VI. CONCLUSION

In this survey paper discuss the integration of machine learning into data analysis represents a transformative shift in how we interpret and utilize data. This new approach enhances our ability to uncover intricate patterns and insights that traditional methods might overlook. By leveraging advanced algorithms and models, we can achieve greater accuracy, efficiency, and scalability in data analysis. Furthermore, machine learning enables predictive analytics, allowing organizations to make data-driven decisions with higher confidence. As technology continues to evolve, the synergy between data analysis and machine learning promises to unlock new opportunities and drive innovation across various fields. Embracing this approach not only optimizes current analytical processes but also paves the way for future advancements in data science.

REFERENCES

[1] Ugochukwu Ikechukwu Okoli , Ogugua Chimezie Obi, Adebunmi Okechukwu Adewusi and Temitayo Oluwaseun Abrahams. "Machine learning in cybersecurity: A review of threat detection and defense mechanisms." Vol. 23, Issue 3, 25 January 2024.

[2] Fatima Al-Quayed,Zulfiqar AhmadAnd Mamoona Humayun. "A Situation Based Predictive Approach for Cybersecurity Intrusion Detection and Prevention Using Machine Learning and Deep Learning Algorithms in Wireless Sensor Networks of Industry 4.0."Volume 128 March 2024.

[3] Priya Thapa, Tamilselvan Arjunan. "AI-Enhanced Cybersecurity: Machine Learning for Anomaly Detection in Cloud Computing." Volume: 09 (2024).

[4] İsa Avcı and Murat Koca. "Cybersecurity Attack Detection Model, Using Machine Learning Techniques." Vol. 20, No. 7, 2023.

[5] Asad Yaseen. "The Role of Machine Learning in Network Anomaly Detection for Cybersecurity." 1(1), 1–15 (2023).

[6] Sungwook Ryu,Jinsu Kim, Namje Park. "Study on Trends and Predictions of Convergence in Cybersecurity Technology Using Machine Learning." Vol. 24 No. 3, May 2023.

[7] Mostofa Ahsan, Kendall E. Nygard, Rahul Gomes, Md Minhaz Chowdhury, Nafiz Rifat and Jayden F Connolly, “Cybersecurity Threats and Their Mitigation Approaches Using Machine Learning”, Volume 2 , Issue 3,2022.

[8] Ziaul Hasan, Hassan r. Mohammad, Maka Jishkarian. "Machine Learning and Data Mining Methods for Cyber Security" pp. 47–56,ISSN: 2958-6542Volume 14, Issue 1, 2021, Pages 560 - 571.

[9] Rahbar Ahsan,Wei Shi,Jean-Pierre Corriveau. "Network intrusion detection using machine learning approaches: Addressing data imbalance”Appl. 2022;7:30–39.

[10] Sabry, Ahmad H., et al. "Single-phase grid-tied transformerless inverter of zero leakage current for PV system." IEEE Access 8 (2019): 4361-4371.

[11] S. Chakraborty, P. Kumar, and B. Sinha, "A study on DDoS attacks, danger and its prevention," Int. J. Res. Anal. Rev., vol. 6, no. 2, pp. 10- 15, 2019.

[12] K. H. Zaboon and A. A. Abdullah, "A Review of the Common DDoS Attack: Types and Protection Approaches Based on Artificial Intelligence," Fusion: Practice and Applications, vol. 7, no. 1, pp. 08-08, Dec. 2021.

[13] L. E. Jaramillo, "Malware detection and mitigation techniques: Lessons learned from Mirai DDOS attack," Journal of Information Systems Engineering & Management, vol. 3, no. 3, pp. 19, Jul. 16, 2018.

[14] A. I. Jony and S. A. Hamim, “Navigating the Cyber Threat Landscape: A Comprehensive Analysis of Attacks and Security in the Digital Age”, Journal of Information Technology and Cyber Security, vol. 1, no. 2, pp. 53-67, 2023.

[15] I. V. Kotenko and A. V. Ulanov, "Agent-based simulation of DDoS attacks and defense mechanisms," Journal of Computing, vol. 4, no. 2, pp. 16-37, 2005.

[16] Q. Yan, F. R. Yu, Q. Gong, and J. Li, "Software-defined networking (SDN) and distributed denial of service (DDoS) attacks in cloud computing environments: A survey, some research issues, and challenges," IEEE communications surveys & tutorials, vol. 18, no. 1, pp. 602-622, Oct. 5, 2015.

[17] Cisco, "Annual Internet Report (2018–2023) White Paper," Accessed June 11, 2020. [Online]. Available: https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internetreport/white-paper-c11-741490.html.

[18] A. I. Jony and A. K. B. Arnob, “Securing the Internet of ThingsEvaluating Machine Learning Algorithms for Detecting IoT Cyberattacks using CIC-IoT2023 Dataset”, International Journal of Information Technology and Computer Science, 2024. (In Press).

[19] S. S. Shanto, Z. Ahmed and A. I. Jony, “Mining User Opinions: A Balanced Bangla Sentiment Analysis Dataset for E-Commerce”, Malaysian Journal of Science and Advanced Technology, vol. 3, no. 4, pp.272-279, 2023.

[20] Z. Chao-Yang, "DOS attack analysis and study of new measures to prevent," in 2011 International Conference on Intelligence Science and Information Engineering, IEEE, Aug. 2011, pp. 426-429.

[21] M. Idhammad, K. Afdel, and M. Belouch, "Semi-supervised machine learning approach for DDoS detection," Applied Intelligence, vol. 48, pp. 3193-3208, Oct. 2018.

[22] D. Tang and X. Kuang, "Distributed denial of service attacks and defense mechanisms," in IOP Conference Series: Materials Science and Engineering, vol. 612, no. 5, p. 052046, Oct. 2019.

[23] N. Tripathi, "DoS and DDoS Attacks: Impact, Analysis and Countermeasures."

[24] M. Hariharan, H.K. Abhishek, and B.G. Prasad, "DDoS attack detection using C5.0 machine learning algorithm," IJ Wireless and Microwave Technologies, vol. 1, pp. 52-59, 2019.

[25] K. Narasimha Mallikarjunan, A. Bhuvaneshwaran, K. Sundarakantham, and S. Mercy Shalinie, "DDAM: detecting DDoS attacks using machine learning approach," in Computational Intelligence: Theories, Applications and Future Directions-Volume I: ICCI-2017, pp. 261-273, Singapore, Aug. 2018.

[26] I. Sharafaldin, A.H. Lashkari, S. Hakak, and A.A. Ghorbani, "Developing realistic distributed denial of service (DDoS) attack dataset and taxonomy," in 2019 International Carnahan Conference on Security Technology (ICCST), pp. 1-8, Oct. 2019.

[27] S. Pande, A. Khamparia, D. Gupta, and D.N. Thanh, "DDOS detection using machine learning technique," in Recent Studies on Computational Intelligence: Doctoral Symposium on Computational Intelligence (DoSCI 2020), pp. 59-68, Springer Singapore, 2021.

[28] K. Kumari and M. Mrunalini, "Detecting Denial of Service attacks using machine learning algorithms," Journal of Big Data, vol. 9, no. 1, pp. 1-7, Dec. 2022.

[29] M. Zekri, S. El Kafhali, N. Aboutabit, and Y. Saadi, "DDoS attack detection using machine learning techniques in cloud computing environments," in 2017 3rd international conference of cloud computing technologies and applications (CloudTech), pp. 1-7, Oct. 2017.

[30] R. Wazirali and R. Ahmad, "Machine Learning Approaches to Detect DoS and Their Effect on WSNs Lifetime," Computers, Materials & Continua, vol. 70, no. 3, Mar. 2022.