AG Publishing House

Advanced Research Techniques: Theories, Methods, And Practices (Volume-2)

Volume 2
Year: 2025

AG PH Books

Machine Learning for Data Analysis in Research: A Review

N V S K Vijayalakshmi K [1*], Dr. J Sasikala [2]

[1]*Assistant professor, Sir C R reddy College of Engineering, Eluru, AP, India.*
[2]*Professor,Dept of IT,Annamalai University,Chidambaram,Tamilnadu.*

Abstract

The implementation of machine learning (ML) in data analytics has drawn significant attention in recent years due to its potential to revolutionise a variety of domains. Reviewing the varied literature on the use of machine learning in research data analysis is the aim of this essay. This review underscores the growing necessity of integrating machine learning (ML) techniques with big data (BD) analytics to address challenges in scalability, speed, and complexity. ML offers promising tools for generating valuable insights and driving smarter decision-making in various domains. However, to fully harness its potential, future research must focus on developing adaptable, cost-effective, and robust ML models capable of handling noisy, incomplete, and real-time data. As data volumes grow, the synergy between ML and BD will remain vital in advancing analytical capabilities, particularly for Internet of Things (IoT) applications and business intelligence.

Keywords: Machine Learning, Data Analysis, Deep Learning, Internet of Things (IoT), Data Processing, Decision-Making, etc.

1 Introduction

The commercial value of Big Data (BD) is enormous in a number of industries, such as "healthcare, transportation, e-business, power monitoring, and economic services". However, when confronted with this vast quantity of data, the conventional method is unable to effectively conduct data analysis. “ABI (Advance Business Intelligence)” conducted research that suggests that there will be more than 30 billion interconnected devices available to satisfy informational requirements [1]. Performing data

* ISBN No. - 978-93-49028-02-9

administration, processing, and analysis is a complex and challenging task due to the immense volume of data that these actual systems can generate from a variety of resources. It is a challenging issue for numerous industries and organizations to integrate the IT departments, healthcare companies, government agencies, and research institutions of today [2]. To solve this problem, BD science was given its own domain, and in order to support effective and quick growth, new trends in research and teaching are needed. The effectiveness and efficiency of Machine Learning (ML) in BD analysis are contingent upon the algorithms and the dataset's configuration, which necessitates numerous time-consuming operations. In practice, not every system can guarantee peak performance without changing the module [3]. With the use of new scientific discoveries that can be combined with machine learning (ML) systems to make decisions, BD solutions may provide high-performance outcomes quickly. ML is an influential instrument for managing BD, according to a variety of studies. As seen in, this is similar to the connection between BD and the ML correlation between the sources and individual learning. Based on this viewpoint, individuals are capable of acquiring knowledge from the sources in order to address innovative issues. Similarly, they are capable of resolving novel issues by acquiring knowledge from BD [4].

A. Machine learning

Machine learning (ML), a branch of artificial intelligence, is the study and development of "statistical algorithms" that may gain insight from data and generalise to new data, performing tasks without explicit instructions. Neural networks are a class of statistical algorithms that have outperformed many earlier machine learning techniques in a subdiscipline of machine learning thanks to developments in the area of deep learning [5]. Machine learning is used in many fields, including "computer vision, speech recognition, email filtering, natural language processing, agriculture, and health". Predictive analytics is the use of machine learning to solve business issues. Techniques for mathematical optimisation, or mathematical programming, and statistical analysis form the foundation of machine learning. The study of exploratory data analysis (EDA) using unsupervised learning is the focus of the related discipline of data mining [6].

B. Importance of machine learning

The amount of data being generated is increasing daily. Every day, the globe generates more data than it has ever done before. Without machine learning, it would be very difficult to evaluate and use all that data. Machine learning is therefore creating a whole new range of possibilities for human interaction with computers and other devices [7]. Machine learning has many applications for businesses, such as data analysis, translation and transcription, chatbots for automated customer support, fraud detection, security threat identification, personalisation, and recommendations. Additionally, machine learning is propelling the exciting innovations of the future, like robots, augmented and virtual reality, drones, and autonomous cars [8].

C. Data Analytics

Data analytics is the act of examining, transforming, and evaluating raw data to identify important "trends, patterns, and insights that might guide decisions". This involves using statistical techniques, tools, and algorithms to transform massive volumes of data into insights that may be put to use. Data analytics is employed extensively in a variety of sectors to enhance corporate performance, optimise operations, anticipate trends, and resolve complex problems [9]. The following lists the essential elements of data analytics:

Data Collection: First, information is gathered from a variety of sources, including databases, websites, sensors, and surveys. Data may be either unstructured (text, graphics, etc.) or structured (rows and columns).

Data Cleaning: Prior to analysis, the data is cleaned to remove errors, duplicates, missing values, and inconsistencies. This guarantees the data's accuracy and dependability.

Data Processing: The data is cleaned and then converted into an analysis-ready format. This might include aggregating data, classifying variables, or normalizing values.

Data Analysis: To find patterns, correlations, and trends in the data, analytical methods including machine learning algorithms, regression analysis, and descriptive statistics are used.

Data Visualization: Understanding the analysis's findings and sharing insights with stakeholders is facilitated by the visual representation of data provided by charts, graphs, and dashboards.

Interpretation and Decision-Making: The analysis's conclusions are interpreted to help guide choices. Businesses might use these insights to improve productivity, address particular issues, or optimize their plans.

D. Roles of Machine Learning in Data Analytics

Automating Data Processing and Analysis: Data processing and analysis automation is one of machine learning's main functions in data analytics. Cleaning, organizing, and processing data using traditional data analysis techniques involves a lot of human labor and often entails tedious job completion. By automating these processes, machine learning saves time and effort needed to prepare data for analysis, allowing data analysts and scientists to focus on more complex and strategic tasks [10].

Identifying Hidden Patterns and Insights: One of the most significant benefits of machine learning in data analytics is its ability to uncover "hidden patterns, trends, and correlations in data" that may be too complex for traditional methods to detect. Large datasets may be rapidly analyzed by ML algorithms, which can provide insights that would otherwise go unnoticed. Machine learning improves businesses' capacity to make data-driven choices based on more profound and nuanced insights by spotting these patterns [11].

Enhancing Predictive Analytics: Predictive analytics is one of machine learning's most well-known applications in data analytics. Machine learning greatly increases the precision and effectiveness of

predictive analytics' forecasts, which are based on previous data and forecast future patterns or occurrences. By applying machine learning to improve prediction accuracy, organisations may make better decisions by anticipating changes in "the market, customer behaviour, and operational risks" [12].

Enabling Real-Time Analytics and Decision-Making: In today's fast-paced business environment, organisations need to make decisions promptly. Traditional data analysis methods often cannot keep up with the volume and speed of incoming data. As a result of machine learning's capacity to assess vast quantities of data in real time, organisations can make data-driven decisions more promptly. Enabling real-time analytics, machine learning provides companies with a competitive edge by allowing them to respond promptly to shifts in "the market, consumer behaviour, or operational circumstances" [13].

Enhancing Data Visualization and Interpretability: The output of "machine learning algorithms" must be actionable and understandable to humans, regardless of their capabilities.. Machine learning has made a substantial contribution to enhancing data visualisation and interpretability, which helps decision-makers better understand complex data and insights. By making data simpler to comprehend and explain, machine learning helps bridge the gap among "technical data analysis and practical business insights" [13].

Driving Prescriptive Analytics: In addition to predictive analytics, machine learning is also playing a critical role in prescriptive analytics, which involves the recommendation of the most effective course of action to achieve desired results. Prescriptive analytics, which is facilitated by machine learning, offers organizations actionable insights that assist them in optimizing their operations and achieving superior results.

2 Literature Review

(Banu et al., 2024) [14] The integration of machine learning algorithms into data-driven projects is examined in this paper, with a focus on a methodical approach to project conception, data collection, preprocessing, model building, assessment, deployment, and monitoring. The research moves forward by choosing and training suitable machine learning algorithms, using techniques like hyperparameter tweaking and cross-validation to maximize model accuracy and generalizability. Along with techniques for real-time prediction and analysis, the shift from model building to deployment in a production setting is covered. In order to ease stakeholder comprehension, the study's conclusion entails producing actionable findings and creating clear visualizations. For openness and future reference, thorough documenting of methods and model setups is recommended. By using machine learning (ML) in data analytics, this all-encompassing strategy seeks to promote operational efficiency and well-informed decision-making. (Chakraborty et al., 2024) [15] This article reviews some of the important data-driven features of "artificial intelligence" in the medical domain. The article showcased the latest developments in "data-driven medical science" through the use of machine learning (ML) and deep learning (DL) in two different areas: first, the use of chabot technologies in medicine and healthcare, specifically ChatGPT, and second, the latest developments in data science in healthcare through the use of ML and DL. DL-enabled ChatGPT technology was recently developed and illustrated in the article. In

conclusion, we provide a succinct summary of the wide-ranging applications of machine learning and deep learning in medicine, as well as the significant challenges that need to be addressed to incorporate the latest "ML to DL technologies into healthcare". Researchers would greatly benefit from the article's thorough summary of "the data-driven paradigm" change in medicine, which is presented using ML to DL technologies.

(Gupta & Gupta, 2023) [16] In this paper, we examine the state-of-the-art in big data analytics and the learning model for healthcare machine learning decision support systems, and we talk about how they could affect patient outcomes and healthcare delivery. To define the big data field, a preliminary survey was conducted. Many hospitals have embraced Hadoop technologies for data analysis because they provide an effective distributed data storage solution that lowers the need for expensive storage devices like disks and others. For the intelligent behavior of the proposed model, this method suggests supervised learning using a hidden Markov model.

(Towler et al., 2023) [17] Evaluated MATA's ability to assist healthcare treatments by contrasting it with "human-only" thematic analysis methods on the same dataset (1,472 user answers from a behavioral intervention for COVID 2019). Findings that varied somewhat were the result of human judgments or subtleties in the study. Conversation: MATA's results were comparable to those of theme analysis done by humans, although it took a lot less time. MATA can help qualitative researchers swiftly comprehend and analyze huge datasets for basic studies that don't need a deep or nuanced comprehension of the data. During public health crises, for example, this method may provide quick optimization and aid in the creation and execution of interventions.

(Alsharif et al., 2020) [18] Network systems are equipped to analyze the data, gain knowledge, and make judgments based on the available dataset. They can also access a variety of experimental symmetric data from a wide range of network devices. This research only looks at supervised and unsupervised machine learning (ML) approaches, which are thought to be the foundation of smart data analysis for the Internet of Things. In addition to discussing the benefits and drawbacks of each algorithm, this paper covers and analyzes important topics pertaining to supervised and unsupervised machine learning approaches. It also looks at research trends and suggests areas for further investigation.

(Kolisetty & Rajput, 2020) [4] Understanding the significance of machine learning in large-scale data analysis is the goal of this research. It helps to comprehend the ramifications and difficulties associated with data heterogeneity, classification error, and computational complexity in big data. The capacity to use data transformation and knowledge extraction to extract value from massive amounts of data for predictive analysis and decision-making is covered. In addition to discussing the degree to which machine learning may be used to big data analysis, it will provide suggestions on how big data may affect real-time data analysis. Additionally, it will highlight the significance and potential from the perspective of promoting feature research advancement in the big data ML sector. (Qiu et al., 2016) [19] summarise the most recent discoveries in the field of machine learning that pertain to the processing of substantial quantities of data. First, examine the various machine learning approaches. Subsequently, identify some of the most intriguing approaches from recent research, such as deep learning,

representation learning, distributed and parallel learning, transfer learning, active learning, and kernel-based learning". The focus should then shift to the analysis and discussion of the challenges and possible solutions related to machine learning for huge data. Then, for large data processing, look at how machine learning and signal processing methods are closely related. Give a summary of a few unresolved problems and current research directions.

3 Conclusion

In conclusion, this review underscores the growing significance of machine learning (ML) in addressing the challenges of big data (BD) analysis across diverse research and industry domains. As data volumes expand and diversify, traditional ML techniques often fall short due to limitations in scalability, speed, and adaptability. Integrating ML with BD analytics enables more accurate predictions, real-time processing, and efficient decision-making, fostering improved business outcomes and research insights. The fusion of ML with moderate-cost systems offers practical alternatives to high-end computational setups, making advanced analytics more accessible. However, challenges such as data noise, incompleteness, and algorithm sensitivity persist, especially in smart data and IoT contexts. This highlights the need for developing more robust, scalable, and noise-resilient ML models. Future research should focus on creating adaptive frameworks, exploring semi-supervised learning, and enhancing ML's ability to process heterogeneous and high-velocity data. Ultimately, the synergy between ML and BD analytics holds transformative potential, but achieving its full benefits requires continuous innovation, ethical considerations, and contextual understanding of the data environments. By evolving to meet the demands of big data, ML can serve as a powerful enabler of operational excellence, informed decision-making, and scientific discovery in the digital age.

References

[1] A. E. Ezugwu, J. Greeff, and Y. S. Ho, "A comprehensive study of groundbreaking machine learning research: Analyzing highly cited and impactful publications across six decades," J. Eng. Res., vol. 13, no. 1, pp. 371–383, 2025, doi: 10.1016/j.jer.2023.10.027.

[2] R. Pugliese, S. Regondi, and R. Marini, "Machine learning-based approach: Global trends, research directions, and regulatory standpoints," Data Sci. Manag., vol. 4, no. December, pp. 19–29, 2021, doi: 10.1016/j.dsm.2021.12.002.

[3] H. S. Sanil et al., "Role of machine learning in changing social and business eco-system – a qualitative study to explore the factors contributing to competitive advantage during COVID pandemic," World J. Eng., vol. 19, no. 2, pp. 238–243, 2021, doi: 10.1108/WJE-06-2021-0357.

[4] V. V. Kolisetty and D. S. Rajput, "A review on the significance of machine learning for data analysis in big data," Jordanian J. Comput. Inf. Technol., vol. 6, no. 1, pp. 41–57, 2020, doi: 10.5455/jjcit.71-1564729835.

[5] J. Kumari, E. Kumar, and D. Kumar, "A Structured Analysis to study the Role of Machine Learning and Deep Learning in The Healthcare Sector with Big Data Analytics," Arch. Comput. Methods Eng., vol. 30, no. 6, pp. 3673–3701, 2023, doi: 10.1007/s11831-023-09915-y.

[6] G. Gouthami, N. Siddartha, and D. P. Borugadda, "Data Science: the Impact of Machine Learning," Futur. Trends Artif. Intell. Vol. 3 B. 8, no. May, pp. 153–170, 2024, doi: 10.58532/v3bgai8p2ch7.

[7] S. Tufail, H. Riggs, M. Tariq, and A. I. Sarwat, "Advancements and Challenges in Machine Learning: A Comprehensive Review of Models, Libraries, Applications, and Algorithms, Electron" vol. 12, no. 8, 2023, doi: 10.3390/electronics12081789.

[8] S. Uddin, S. Yan, and H. Lu, "Machine learning and deep learning in project analytics: methods, applications and research trends," Prod. Plan. Control, vol. 0, no. 0, pp. 1–20, 2024, doi: 10.1080/09537287.2024.2320790.

[9] B. Khosravi et al., "Demystifying Statistics and Machine Learning in Analysis of Structured Tabular Data," J. Arthroplasty, vol. 38, no. 10, pp. 1943–1947, 2023, doi: 10.1016/j.arth.2023.08.045.

[10] A. Brnabic and L. M. Hess, "Systematic literature review of machine learning methods used in the analysis of real-world data for patient-provider decision making," BMC Med. Inform. Decis. Mak., vol. 21, no. 1, 2021, doi: 10.1186/s12911-021-01403-2.

[11] S. Uddin, S. Ong, and H. Lu, "Machine learning in project analytics: a data-driven framework and case study," Sci. Rep., vol. 12, no. 1, pp. 1–13, 2022, doi: 10.1038/s41598-022-19728-x.

[12] I. H. Sarker, "Machine Learning: Algorithms, Real-World Applications and Research Directions," SN Comput. Sci., vol. 2, no. 3, pp. 1–21, 2021, doi: 10.1007/s42979-021-00592-x.

[13] N. Ben and S. K. Yadav, "Women Empowerment - A Study of SHGS in Satna & Maihar District of Madhya Pradesh," Int. J. Innov. Sci. Eng. Manag., pp. 324–332, 2025, doi: 10.69968/ijisem.2025v4i1324-332.

[14] A. Banu, J. Sravanthi, S. Bolugoddu, L. Panjala, and S. Hasanoddin, "Data-Driven Insights : Applying Machine Learning in Data Analytics," Intell. Syst. Appl. Eng., 2024.

[15] C. Chakraborty, M. Bhattacharya, S. Pal, and S. S. Lee, "From machine learning to deep learning: Advances of the recent data-driven paradigm shift in medicine and healthcare," Curr. Res. Biotechnol., vol. 7, no. August 2023, p. 100164, 2024, doi: 10.1016/j.crbiot.2023.100164.

[16] D. Gupta and S. Gupta, "Big Data Analysis using Machine Learning for Health Systems," Int. J. Comput. Appl., vol. 185, no. 9, pp. 8–13, 2023, doi: 10.5120/ijca2023922745.

[17] L. Towler et al., "Applying machine-learning to rapidly analyze large qualitative text datasets to inform the COVID-19 pandemic response: comparing human and machine-assisted topic analysis techniques," Front. Public Heal. vol. 11, no. October, pp. 1–11, 2023, doi: 10.3389/fpubh.2023.1268223.

[18] M. H. Alsharif, A. H. Kelechi, K. Yahya, and S. A. Chaudhry, "Machine learning algorithms for smart data analysis in internet of things environment: Taxonomies and research trends," Symmetry (Basel), vol. 12, no. 1, 2020, doi: 10.3390/SYM12010088.

[19] J. Qiu, Q. Wu, G. Ding, Y. Xu, and S. Feng, "A survey of machine learning for big data processing," EURASIP J. Adv. Signal Process., vol. 2016, no. 1, 2016, doi: 10.1186/s13634-016-0355-x.