Vol 6, Issue 3, March 2026 www.ijesti.com E-ISSN: 2582-9734
International Journal of Engineering, Science, Technology and Innovation (IJESTI)

Advancements in Predictive Analytics: Leveraging Big Data, AI, and Machine Learning

MD. Asif Ali
M. Tech. in Computer Science Engineering, CBS Group of Institutions, Jhajjar, Haryana.

Amreesh Kumar Yadav
A.P CSE Department, CBS Group of Institutions, Jhajjar, Haryana.

ABSTRACT

The integration of big data, machine learning (ML), and artificial intelligence (AI) has revolutionized decision-making through predictive analytics, enabling the analysis of complex datasets. Predictive analytics, involving statistical techniques and data mining, supports decision-making in sectors like healthcare, finance, and manufacturing. With the need for intelligent systems, AI-driven analytics optimize forecasting and operational efficiency. While challenges in scalability, privacy, and model interpretability persist, the combination of ML and big data enhances accuracy in real-time decision-making across industries. These advancements have transformed traditional decision support systems, providing strategic insights and improving outcomes across various domains.

Keywords: *Predictive Analytics, Big Data, Machine Learning, Artificial Intelligence.*

I. INTRODUCTION

In recent years, the rapid growth of digital technologies has significantly transformed the way organizations collect, process, and utilize data for decision-making. The emergence of big data, machine learning (ML), and artificial intelligence (AI) has enabled the development of advanced predictive analytics frameworks capable of handling large-scale and complex datasets. Predictive analytics refers to the use of statistical techniques, data mining, and machine learning algorithms to analyze historical and real-time data in order to forecast future outcomes and support decision-making processes (Savadatti et al., 2022). With the increasing demand for intelligent systems, predictive analytics has become a core component of modern decision support systems across various domains, including healthcare, finance, manufacturing, energy, and supply chain management. Big data plays a central role in predictive analytics by providing massive volumes of structured, semi-structured, and unstructured data generated from diverse sources such as sensors, social media platforms, IoT devices, and enterprise systems. However, traditional analytical methods are insufficient to process such large-scale and high-velocity datasets effectively (Mittal & Sangwan, 2019). This limitation has led to the integration of machine learning techniques that are capable of learning patterns, identifying correlations, and making predictions with high accuracy. Machine learning models such as decision trees, random forests, support vector machines, artificial neural networks, and deep learning architectures have demonstrated strong performance in predictive tasks due to their ability to adapt to complex and nonlinear data structures (Bokonda et al., 2020). The integration of AI and big data analytics has further enhanced the capabilities of predictive systems by enabling automated reasoning and intelligent decision-making. Husnain (2026) emphasized that the combination of big data, AI, and deep learning has transformed traditional systems into intelligent, data-driven environments that improve forecasting accuracy, operational efficiency, and strategic planning. Similarly, Strielkowski et al. (2025) highlighted that machine learning-based predictive frameworks significantly enhance real-time decision-making in large-scale systems such as electric power grids by improving demand forecasting and fault detection capabilities. In addition to technical advancements, predictive analytics frameworks are increasingly being applied in various industrial and social domains. For instance, Paramesha et al. (2024) reported that the integration of AI, IoT, and big data analytics in business intelligence systems enables organizations to gain real-time insights and improve strategic decision-making. In healthcare, predictive models have been widely used for disease prediction and early diagnosis, improving patient outcomes and reducing healthcare costs (Venkatesh et al., 2019). These applications demonstrate the versatility and importance of predictive analytics in addressing complex real-world problems. Despite these advancements, several challenges remain in implementing large-scale predictive analytics frameworks. Issues such as data heterogeneity, computational complexity, scalability, privacy concerns, and model interpretability continue to limit widespread adoption (Abu-Salih et al., 2021). Additionally, traditional machine learning models often struggle with high-dimensional data and require significant computational resources for training and optimization (Safizadeh, 2025). To overcome these limitations, researchers have proposed hybrid approaches, including deep learning-based architectures and distributed computing frameworks such as Hadoop and Spark, to improve scalability and efficiency. Therefore, the development of a machine learning-based predictive analytics framework for large-scale big data processing and decision support systems is essential to address the growing complexity of modern data environments. Such frameworks aim to integrate efficient data processing techniques with advanced predictive models to enhance accuracy, reduce processing time, and support real-time decision-making. By leveraging AI-driven insights, organizations can improve operational performance, optimize resource utilization, and gain a competitive advantage in dynamic environments.

II. RESEARCH BACKGROUND

Husnain (2026) had reviewed how big data, artificial intelligence (AI), and deep learning were transforming supply chain management into an intelligent and data-driven system. The study had discussed the role of these technologies in enhancing supply chain visibility, improving predictive and prescriptive decision-making, and increasing operational efficiency at the strategic, tactical, and operational levels. It had been reported that big data provided access to large volumes of diverse information, while AI and deep learning extracted meaningful insights from complex datasets to optimize demand forecasting, inventory management, logistics, and risk mitigation. The review had further highlighted that emerging trends such as the Internet of Things (IoT), automation, and explainable AI were contributing to the development of real-time, adaptive, and sustainable supply chains. Overall, the paper had presented the opportunities, challenges, and future directions of smart supply chain analytics, emphasizing its growing importance in modern supply chain systems.

Strielkowski et al. (2025) had examined the rapid advancement of IoT, Artificial Intelligence (AI), cloud computing, and Big Data, and had reported that these technologies significantly accelerated the adoption of predictive analytics in electric power systems. The study had introduced a novel predictive analytics framework that integrated both supervised and unsupervised machine learning techniques, including linear and logistic regression, decision trees, random forests, and clustering algorithms, for short-term power demand forecasting and early fault detection. Using grid load data from the U.S. Department of Energy’s Open Energy Data Initiative (OEDI), the researchers had systematically demonstrated the implementation and optimization of these models for power systems. Their findings had indicated efficiency improvements ranging from 14% to 24%, along with enhanced reliability, economic savings, reduced greenhouse gas emissions, and improved infrastructure utilization. The study had also discussed regulatory barriers, industry adoption challenges, and suggested future directions concerning scalability, renewable integration, and policy-oriented research.

Safizadeh (2025) reported that machine learning models, particularly deep learning approaches, had played an important role in analysing complex and large-scale datasets. The study introduced a hybrid analytical framework that combined Annealed Gradient Descent (AGD) and Hybrid Orthogonal Projection and Estimation (HOPE) to improve prediction accuracy, reduce training time, and enhance model stability. It was observed that AGD had accelerated convergence and reduced the risk of being trapped in local minima, thereby improving training efficiency. HOPE was found to support dimensionality reduction and noise elimination through orthogonal projection. The findings indicated that the proposed AGD–HOPE framework had significantly improved deep learning performance for complex and time-series data. Experimental results showed that training time had been reduced by up to 40%, while prediction accuracy had increased by up to 6% compared to traditional methods such as SGD, SVM, and XGBoost. The framework was also reported to have demonstrated robustness and stable convergence across diverse datasets and architectures.

Paramesha et al. (2024) had examined the transformative role of digital technologies in reshaping business intelligence (BI), emphasizing that the integration of Big Data, Artificial Intelligence (AI), and the Internet of Things (IoT) had created new opportunities for data-driven decision-making and strategic insight generation. The study had reported that the growing volume of data from diverse sources had necessitated advanced analytical approaches to derive meaningful information. Through an extensive literature review and keyword co-occurrence with cluster analysis, the authors had identified key themes and interrelationships within the domain. It had been observed that Big Data analytics uncovered hidden patterns, AI enabled predictive modelling, and IoT provided real-time data from interconnected devices. The study had further proposed a comprehensive BI framework incorporating IoT-based data collection, Big Data processing, AI-driven analysis, edge computing for low-latency decisions, and blockchain for data security, thereby enhancing organizational efficiency, innovation, and competitive advantage.

Lourens et al. (2023) examined the growing significance of data science and machine learning in shaping the future of the automotive industry. The authors explained that these technologies had become increasingly essential because of their ability to automatically learn from data and optimize industrial processes and products. The study had defined both data science and machine learning and had drawn clear parallels between the two domains. It had also described the concept of automatic optimization and demonstrated its value when integrated with data analytics. Further, the article had illustrated the existing applications of these technologies across key subprocesses within the automotive value chain. Since the industry was still in the early stages of exploring these innovations, the authors had also presented futuristic use cases to highlight their transformative potential. The study concluded that such technological advancements could enhance productivity, strengthen customer focus, and improve the overall product development and customer interaction processes in the automotive sector.

Savadatti et al. (2022) had presented an elementary review on machine learning and predictive analysis as essential tools for extracting meaningful insights from large and complex datasets. The authors had explained that predictive analysis had been considered an advanced analytical approach that utilized algorithms, data mining, and statistical modelling to forecast future outcomes from historical data. It had been reported that many organizations and companies had applied these techniques to identify patterns, opportunities, and potential risks across diverse sectors such as business, medicine, and education. The study had further highlighted that machine learning and predictive analysis had approached problem-solving from different perspectives, while also emphasizing that predictive analysis could be effectively performed through deep learning methods. Moreover, the review had provided deeper insights into the field by discussing important aspects such as algorithms, feature selection, and practical applications. It had served as a useful guide for researchers and practitioners interested in applying predictive analysis in academic and professional domains.

Abu-Salih et al. (2021) had discussed that the continuous increase in both the quality and quantity of data generated from daily business operations, along with the constant integration of related social data, had rendered traditional statistical methods inadequate for handling such large and complex data streams. The authors had emphasized that this challenge had necessitated the development of advanced and sophisticated analytical approaches capable of extracting valuable insights for the business domain. Their chapter had highlighted the foundational aspects of social big data (SBD) analytics, with particular focus on the importance of predictive analytics in this context. They had further presented a structured framework for SBD predictive analytics and introduced various predictive analytical algorithms along with their applications across important domains. In addition, top-tier tools and APIs had been reviewed. A case study based on predictive analytics using social data had also been provided, supported by experiments that had demonstrated its significance and practical utility.

Bokonda et al. (2020) reported that Artificial Intelligence (AI) had grown considerably over the previous decade, with Machine Learning (ML) emerging as one of the most widely used branches for predictive analysis. Their study was intended to review the major trends and methods of ML applied in predictive analytics. For this purpose, research articles were collected from three scientific databases and filtered through defined selection criteria, focusing on studies published during the previous five years, with priority given to peer-reviewed journal papers. This screening process had resulted in the final selection of 30 research articles for detailed review. The study had aimed to assist researchers, companies, and practitioners in selecting appropriate ML techniques according to application domains. It was highlighted that Decision Tree (DT) and Artificial Neural Network (ANN) had been widely used in education, Logistic Regression (LR), Random Forest (RF), and DT in building applications, DT in botany, RF and ANN in social sciences, and RF in medicine.

Venkatesh et al. (2019) reported that health prediction had become highly essential in modern life, particularly due to the growing burden of heart disease worldwide. The authors indicated that big data analytics had played a significant role in forecasting future health status and improving healthcare outcomes. In their study, a Big Data Predictive Analytics Model for Disease Prediction using the Naïve Bayes technique (BPA-NB) had been proposed. It was explained that the Naïve Bayes classifier, based on Bayes' theorem with feature independence assumptions, had been found suitable for large healthcare datasets. The heart disease dataset had been obtained from the UCI Machine Learning Repository, where the model had been trained and later tested for disease classification. Hadoop-Spark had been employed as the big data computing framework for analysis. The findings demonstrated that the proposed BPA-NB model had achieved an accuracy of 97.12%, thereby showing effectiveness in early disease detection and future health condition prediction.

Mittal and Sangwan (2019) had reported that enormous volumes of data were being generated daily, characterized by features such as high velocity, massive volume, uncertainty, non-stationarity, and real-time processing requirements. They had observed that conventional machine learning techniques were not considered suitable for big data analytics because such methods were unable to efficiently handle these complex characteristics. The study had further indicated that traditional storage and processing approaches had failed to satisfy the computational and scalability demands of big data environments. In their paper, the authors had discussed several challenges associated with applying traditional machine learning techniques to big data analytics and had also highlighted possible solutions to address these issues. Based on their survey, it had been concluded that approaches such as parallel processing, dimensionality reduction, GPU-based computation, MapReduce jobs, deep learning, online learning, and incremental learning had emerged as effective strategies for overcoming the major challenges related to big data analytics.

Vinitha et al. (2018) had examined the growing role of big data in biomedical and healthcare communities and had emphasized that accurate analysis of medical data could support early disease identification, improved patient care, and enhanced community health services. The authors had observed that incomplete medical datasets often reduced analytical accuracy, while regional variations in disease patterns could weaken outbreak prediction. To address these issues, they had proposed a system employing machine learning algorithms for effective prediction of disease occurrences in high-risk communities. Real-life hospital data had been used to evaluate the modified prediction models. A latent factor model had been applied to reconstruct missing data and improve reliability. The study had specifically focused on cerebral infarction as a regional chronic disease. By integrating both structured and unstructured hospital data, Decision Tree and MapReduce algorithms had been utilized. The proposed approach had achieved 94.8% prediction accuracy and had shown faster convergence than the CNN-UDRP model.

Nithya and Ilango (2017) reported that machine learning had emerged as a highly effective approach for predictive analysis and pattern recognition, particularly when large datasets were involved. The authors explained that machine learning, as one of the fastest-growing areas in computer science, had gained significant importance in health informatics due to its ability to develop algorithms that could learn from data and improve over time. They observed that machine learning techniques had been widely applied across multiple domains, with the healthcare sector benefiting substantially from prediction-based applications. The study highlighted that machine learning had provided various alerting systems, risk management tools, and decision-support mechanisms aimed at improving patient safety and healthcare quality. Furthermore, the authors noted that the healthcare industry had faced major challenges related to electronic health record management, data integration, computer-aided diagnosis, and disease prediction. Overall, the paper presented an overview of diverse machine learning prediction techniques, tools, and healthcare applications.

III. KEY FINDINGS FROM STUDY

Author (Year)	Objective	Methodology / Approach	Domain / Dataset	ML Techniques Used	Key Findings	Limitations
Husnain (2026)	To analyze AI, big data, and deep learning in supply chain intelligence	Literature-based review	Supply chain systems	Deep learning, AI analytics	Improved forecasting, visibility, and decision-making efficiency	Limited real-time implementation studies
Strielkowski et al. (2025)	To develop predictive analytics for electric power systems	Experimental framework using OEDI dataset	Energy systems	Regression, decision trees, random forest, clustering	14–24% efficiency improvement, better reliability	Regulatory and scalability issues
Safizadeh (2025)	To improve big data model performance using hybrid optimization	Hybrid analytical framework	Big data systems	AGD + HOPE deep learning framework	40% reduction in training time, 6% accuracy gain	High computational complexity

Paramesha et al. (2024)	To study AI, IoT, and big data in business intelligence	Literature review + cluster analysis	Business intelligence systems	AI, IoT analytics, big data processing	Improved real-time decision-making and insights	Integration complexity and security issues
Lourens et al. (2023)	To explore ML in automotive industry optimization	Conceptual + application review	Automotive sector	Machine learning, data science models	Enhanced productivity and product optimization	Early-stage adoption
Savadatti et al. (2022)	To review predictive analytics techniques	Review study	Multi-domain applications	ML algorithms, data mining, deep learning	ML effective for forecasting and risk analysis	Lack of standardized frameworks
Abu-Salih et al. (2021)	To propose predictive analytics framework for social big data	Framework development + case study	Social media & business data	Predictive ML algorithms	Effective insight extraction from social data	Privacy and data quality issues
Bokonda et al. (2020)	To review ML trends in predictive analytics	Systematic literature review	Multi-domain studies	DT, ANN, RF, LR	DT & RF widely used across sectors	Limited comparative performance analysis
Venkatesh et al. (2019)	To predict heart disease using big data analytics	Experimental model using Hadoop-Spark	Healthcare (UCI dataset)	Naïve Bayes classifier	97.12% prediction accuracy	Limited generalization across diseases
Mittal & Sangwan (2019)	To analyze ML challenges in big data	Theoretical review	Big data systems	Parallel processing, DL, MapReduce	ML requires scalable computing frameworks	High computational cost
Vinitha et al. (2018)	To improve disease prediction accuracy	Latent factor model + ML	Healthcare hospital data	Decision Tree, MapReduce	94.8% accuracy in disease prediction	Missing data challenges
Nithya & Ilango (2017)	To study ML in healthcare predictive analytics	Review study	Healthcare systems	ML prediction models	Improved diagnosis and decision support	Data integration issues

IV. CONCLUSION

The review of literature on machine learning-based predictive analytics frameworks for large-scale big data processing clearly indicates that predictive systems have become a fundamental component of modern decision support environments. The integration of big data, machine learning, and artificial intelligence has significantly enhanced the ability of organizations to extract meaningful insights from massive and complex datasets. Studies such as Husnain (2026) and Strielkowski et al. (2025) demonstrate that AI-driven predictive models improve forecasting accuracy, operational efficiency, and real-time decision-making in domains such as supply chain management and energy systems. Furthermore, research findings highlight that machine learning algorithms including decision trees, random forests, naïve Bayes, and deep learning models have been widely applied across multiple sectors such as healthcare, business intelligence, automotive systems, and social analytics. For instance, Venkatesh et al. (2019) achieved high prediction accuracy in disease detection using big data frameworks, while Safizadeh (2025) showed that hybrid optimization techniques significantly enhance model efficiency and reduce training time. These advancements confirm that predictive analytics has evolved into a powerful tool for addressing complex, data-intensive problems. However, despite these advancements, several challenges remain unresolved. Issues such as data heterogeneity, scalability constraints, computational complexity, privacy concerns, and lack of model interpretability continue to limit the full-scale implementation of predictive analytics systems (Mittal & Sangwan, 2019; Abu-Salih et al., 2021). Therefore, there is a strong need for more robust, scalable, and interpretable machine learning frameworks capable of handling real-time big data environments effectively. Overall, the literature confirms that machine learning-based predictive analytics plays a crucial role in transforming raw data into actionable intelligence, thereby supporting strategic and operational decision-making across industries.

V. FUTURE SCOPE

- **Development of Real-Time Predictive Systems:** Future research can focus on building real-time analytics frameworks capable of processing streaming big data for instant decision-making in dynamic environments.
- **Integration of Explainable AI (XAI):** Improving model interpretability is essential to increase trust and transparency in predictive analytics systems, especially in critical domains like healthcare and finance.
- **Edge and Cloud Computing Hybrid Models:** Combining edge computing with cloud-based big data frameworks can reduce latency and improve processing efficiency for large-scale applications.
- **Scalable Deep Learning Architectures:** Future studies can explore distributed deep learning models that efficiently handle high-dimensional datasets with reduced computational cost.
- **Enhanced Data Privacy and Security Mechanisms:** With increasing data sensitivity, secure machine learning frameworks incorporating encryption and privacy-preserving techniques will be crucial.
- **Automated Machine Learning (AutoML) Systems:** Development of automated model selection and optimization techniques can simplify predictive analytics for non-expert users.
- **Cross-Domain Predictive Applications:** Future research can expand predictive analytics applications into smart cities, agriculture, transportation, and environmental monitoring.

- **Improved Handling of Data Quality Issues:** Advanced preprocessing techniques are required to manage missing, noisy, and imbalanced datasets more effectively.
- **Integration with IoT and Sensor Networks:** Combining predictive analytics with IoT ecosystems will enable smarter and more responsive industrial and urban systems.
- **Energy-Efficient Machine Learning Models:** Research can focus on reducing energy consumption in large-scale AI systems to support sustainable computing practices.

REFERENCES

1. Husnain, A. (2026). Integrating Big Data, AI, and Deep Learning for Intelligent Supply Chain Decisions. *Global Trends in Science and Technology*, *2*(1), 1-18.
2. Strielkowski, W., Vlasov, A., Selivanov, K., Rasuk, A., & Smutka, L. (2025). Predictive demand analytics and machine learning in electric power systems for enhancing resilience and efficiency. *Sustainable Energy, Grids and Networks*, *42*, 101722.
3. Safizadeh, B. (2025). Design and Evaluation of Analytical Models for Accelerating Decision-Making Processes in Big Data Analysis Using Machine Learning Techniques. *Scientific Journal of Research Studies in Future Computer Sciences*, *3*(1), 1-10.
4. 11.Paramesha, M., Rane, N., & Rane, J. (2024). Big data analytics, artificial intelligence, machine learning, internet of things, and blockchain for enhanced business intelligence. *Artificial Intelligence, Machine Learning, Internet of Things, and Blockchain for Enhanced Business Intelligence (June 6, 2024)*.
5. Lourens, M., Sharma, S., Pulugu, R., Gehlot, A., Manoharan, G., & Kapila, D. (2023, May). Machine learning-based predictive analytics and big data in the automotive sector. In *2023 3rd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE)* (pp. 1043-1048). IEEE.
6. Savadatti, M. B., Dhivya, M., Meghanashree, C., Navya, M. K., Lokesh, Y., & Kawri, N. (2022, January). An overview of predictive analysis based on machine learning techniques. In *2022 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI)* (pp. 1-6). IEEE.
7. Abu-Salih, B., Wongthongtham, P., Zhu, D., Chan, K. Y., & Rudra, A. (2021). Predictive analytics using Social Big Data and machine learning. In *Social Big Data Analytics: Practices, Techniques, and Applications* (pp. 113-143). Singapore: Springer Singapore.
8. Bokonda, P. L., Ouazzani-Touhami, K., & Souissi, N. (2020, November). Predictive analysis using machine learning: Review of trends and methods. In *2020 International Symposium on Advanced Electrical and Communication Technologies (ISAECT)* (pp. 1-6). IEEE.
9. Venkatesh, R., Balasubramanian, C., & Kaliappan, M. (2019). Development of big data predictive analytics model for disease prediction using machine learning technique. *Journal of medical systems*, *43*(8), 272.
10. Mittal, S., & Sangwan, O. P. (2019, January). Big data analytics using machine learning techniques. In *2019 9th international conference on cloud computing, data science & engineering (Confluence)* (pp. 203-207). IEEE.
11. Vinitha, S., Sweetlin, S., Vinusha, H., & Sajini, S. (2018). Disease prediction using machine learning over big data. *Computer Science & Engineering: An International Journal (CSEIJ)*, *8*(1), 1-8.
12. Nithya, B., & Ilango, V. (2017, June). Predictive analytics in health care using machine learning tools and techniques. In *2017 International Conference on Intelligent Computing and Control Systems (ICICCS)* (pp. 492-499). IEEE.