© February 2026| IJIRT | Volume 12 Issue 9 | ISSN: 2349-6002

Artificial Intelligence and Data-Driven Analytics in E-Commerce: A Comprehensive Review of Methods, Applications and Research Challenges

C.Subashree[1], Dr.J.Savitha[2]

[1]*Research Scholar, Department of Computer Science, Dr.N.G.P. Arts & Science College, Coimbatore, Tamilnadu, India*

[2]*Professor, Department of Computer Science, Dr.N.G.P. Arts & Science College, Coimbatore, Tamilnadu, India*

Abstract: **The fast growth of e-commerce platforms has generated massive volumes of heterogeneous data, necessitating intelligent and scalable analytical solutions to support decision-making, personalization, and business sustainability. This review paper presents a comprehensive synthesis of recent research on artificial intelligence (AI), machine learning (ML), deep learning (DL), and big data analytics applied to e-commerce environments. The reviewed literature spans key application domains, including customer demand forecasting, profit prediction, churn analysis, recommendation systems, dynamic pricing, product classification, website usability evaluation, and intelligent decision support systems. Studies employing advanced models such as transformer architectures, convolutional and recurrent neural networks, ensemble learning methods, reinforcement learning, and large language models are critically examined alongside traditional data mining and statistical techniques. The review highlights the strengths of AI-driven approaches in capturing complex consumer behavior, improving predictive accuracy, and enabling real-time, data-driven strategies. At the same time, it identifies persistent challenges related to interpretability, data imbalance, scalability, privacy, and real-world deployment. By organizing existing contributions into coherent thematic categories, this review provides a structured overview of methodological trends, datasets, and performance outcomes. The paper concludes by outlining future research directions toward adaptive, explainable, privacy-aware, and business-aligned AI systems for next-generation e-commerce platforms.**

Keywords: ***E-commerce analytics, Artificial intelligence, Machine learning, Deep learning, Customer churn prediction, Demand forecasting, Recommendation systems, Big data analytics, Intelligent decision support and Data mining.***

I.INTRODUCTION

The hasty expansion of e-commerce platforms has fundamentally transformed global retail ecosystems, reshaping how consumers interact with products, services, and businesses. With the proliferation of online marketplaces, vast volumes of heterogeneous data are continuously generated from customer transactions, browsing behavior, reviews, social media interactions, Internet of Things (IoT) devices, and operational logs. This data deluge presents both unprecedented opportunities and significant challenges for e-commerce organizations striving to remain competitive, customer-centric, and operationally efficient. Traditional rule-based and statistical approaches are increasingly inadequate for extracting meaningful insights from such complex, high-dimensional, and dynamic datasets [1].

In this context, artificial intelligence (AI), machine learning (ML), deep learning (DL), and big data analytics have emerged as core enabling technologies for intelligent e-commerce systems. These techniques empower platforms to perform advanced tasks such as customer demand forecasting, churn prediction, profit estimation, recommendation generation, dynamic pricing, product classification, usability evaluation, and intelligent decision support. Recent advances including ensemble learning methods, convolutional and recurrent neural networks, transformer-based architectures, reinforcement learning, and large language models have demonstrated superior performance in modeling nonlinear relationships, temporal dependencies, and contextual patterns inherent in e-commerce data. As a result, AI-driven analytics are now central to data-informed strategic planning, personalization, and automation across digital retail operations [2]. Despite substantial progress, the deployment of AI and data-driven solutions in e-commerce is not without limitations. Key challenges persist related to model interpretability, data imbalance, scalability, privacy preservation, security, concept drift, and real-world integration. Many high-performing deep learning models operate as black-box systems, limiting transparency and trust for business stakeholders. Additionally, the growing reliance on sensitive customer data raises critical ethical and regulatory concerns, particularly in light of data protection frameworks such as GDPR and CCPA. These challenges highlight the need for a comprehensive synthesis of existing research to identify methodological trends, strengths, weaknesses, and open research gaps.

This review paper aims to address this need by presenting a systematic and thematic analysis of contemporary research on AI, ML, DL, and big data analytics in e-commerce. Drawing on an extensive body of literature, the review categorizes studies across major application domains, examines the datasets and methods employed, and evaluates reported outcomes, advantages, and limitations. By consolidating fragmented research into a coherent framework, this paper provides valuable insights for researchers, practitioners, and policymakers seeking to design adaptive, explainable, privacy-aware, and business-aligned intelligent e-commerce systems. Ultimately, the review seeks to guide future research directions and support the development of next-generation data-driven e-commerce platforms capable of sustaining long-term growth and innovation

II. LITERATURE REVIEW

This section presents a structured review of existing research related to the application of artificial intelligence, machine learning, deep learning, and big data analytics in e-commerce systems. The reviewed studies are organized thematically to capture key research trends across domains such as customer behavior analysis, churn prediction, demand and sales forecasting, recommendation systems, dynamic pricing, product classification, website usability evaluation, and intelligent decision support. By synthesizing prior work based on methodologies, datasets, and reported outcomes, this section highlights the evolution of data-driven techniques in e-commerce, identifies their strengths and limitations, and reveals critical research gaps that motivate the need for more adaptive, explainable, and scalable analytical frameworks.

2.1. Deep Learning–Based Demand, Sales, and Profit Forecasting in E-Commerce

Md. Mortuza Ahmmed et al. (2025) [3] contributes to this evolving body of literature by proposing a Conditional Transformer Language Model (CTRL) for customer demand prediction in e-commerce environments. Unlike traditional forecasting models that focus primarily on numerical sales data, this study leverages rich textual information from large-scale customer reviews, specifically the Amazon Reviews 2018 dataset. Through comprehensive data preprocessing, tokenization, and corpus construction, the research demonstrates how unstructured textual data can be effectively transformed into meaningful features for demand prediction. The research further distinguishes itself by conducting an extensive comparative evaluation between the CTRL model and conventional machine learning algorithms. Using standard performance metrics such as accuracy, precision, recall, and F1-score, the authors demonstrate that CTRL consistently outperforms traditional models. This improvement is attributed to the model's enhanced ability to capture contextual dependencies and nuanced consumer sentiment embedded within review text. Overall, this work reflects a paradigm shift from rule-based and shallow learning models toward deep learning driven frameworks for demand forecasting. The findings of this study reinforce existing evidence that transformer-based models significantly enhance predictive accuracy and decision-making capability. Moreover, by linking demand forecasting outcomes to strategic inventory management and proactive market adaptation, the research underscores the practical relevance of advanced deep learning techniques in achieving sustainable growth and operational efficiency in the competitive e-commerce landscape.

Advantages: The use of the Conditional Transformer Language Model enables superior contextual understanding of consumer reviews, leading to significantly improved demand forecasting accuracy compared to traditional machine learning approaches. The deep learning framework effectively handles large-scale datasets and provides actionable insights, supporting proactive inventory planning and informed strategic decision-making in dynamic e-commerce environments. Disadvantages: Transformer-based models like CTRL require substantial computational resources and training time, which may limit their practical adoption for small or resource-constrained e-commerce businesses. While highly accurate, deep learning models operate as black-box systems, making it difficult for stakeholders to interpret how specific features or customer behaviors influence demand predictions.

Norun Nabi et al. (2024) [4] investigates the effectiveness of advanced CNN architectures namely VGG16, ResNet50, and InceptionV3 for predicting e-commerce profits. By emphasizing the role of high-quality datasets, the research demonstrates that CNN-based models significantly outperform conventional algorithms. Among the evaluated architectures, VGG16 achieves the highest prediction accuracy of 92.55%, underscoring its robustness in learning discriminative features relevant to profit estimation. These findings validate the growing consensus in the literature that deep learning models excel in capturing latent patterns embedded within complex e-commerce data streams. Furthermore, the study highlights how CNN-driven analytics can generate actionable insights that directly support revenue optimization and operational efficiency. By leveraging deep feature extraction capabilities, businesses can better understand sales trends, customer behavior, and performance indicators, thereby improving strategic decision-making. However, the authors also stress that successful deployment of CNN models requires substantial computational infrastructure and domain expertise, along with careful consideration of ethical and data privacy issues. Overall, this research contributes meaningfully to the expanding body of knowledge on deep learning applications in e-commerce analytics. It reinforces the paradigm shift toward data-driven, intelligent forecasting systems and provides practical guidance for organizations seeking competitive advantage in a rapidly evolving global digital marketplace. The study also lays a foundation for future research into scalable, interpretable, and ethically responsible deep learning frameworks for profit prediction.

Advantages: One major advantage of employing CNN-based models for e-commerce profit prediction is their superior predictive performance. The ability of architectures such as VGG16, ResNet50, and InceptionV3 to automatically extract high-level features from complex datasets enables more accurate modeling of sales and profit trends compared to traditional algorithms. This enhanced accuracy supports better forecasting, informed pricing strategies, and optimized inventory management. Disadvantages: Despite their effectiveness, CNN-based profit prediction models entail (need) significant computational and infrastructural demands. Training and deploying deep learning architectures require high-performance hardware, substantial storage, and specialized expertise, which may pose challenges for small and medium-sized enterprises with limited resources.

Jianhao Zhang (2025) [5] investigates the effectiveness of ensemble learning models for forecasting future sales volumes using historical e-commerce data. By continuously feeding past sales records into predictive models, the research identifies temporal patterns such as seasonality, weekly trends, and long-term fluctuations. Three advanced machine learning models Gradient Boosting Decision Trees (GBDT), Light Gradient Boosting Machine (LightGBM), and Extreme Gradient Boosting (XGBoost) are evaluated to determine their suitability for handling complex sales dynamics. Feature engineering techniques, including time-based, lag, and rolling features, are employed to capture temporal dependencies, while Grid Search based hyperparameter tuning enhances model performance. Experimental results demonstrate that XGBoost outperforms the other models, particularly in handling demand surges caused by holidays and special events. The model achieves strong predictive performance, with high R^2 scores on both training and test datasets, indicating good generalization capability. The findings emphasize that accurate sales forecasting enables e-commerce platforms to optimize inventory levels, allocate resources efficiently, and maintain market stability during peak demand periods. Overall, the study reinforces the importance of machine learning driven forecasting frameworks in supporting data-driven decision-making and sustaining competitiveness in dynamic e-commerce environments.

Advantage: A major advantage of this approach is its strong predictive accuracy in capturing nonlinear sales patterns and sudden demand surges. By leveraging XGBoost along with advanced feature engineering and hyperparameter tuning, the model effectively identifies seasonal trends and special-event effects, enabling e-commerce platforms to proactively manage inventory, reduce stockouts, and enhance customer satisfaction during peak periods. Disadvantage: A key limitation of the study lies in data dependency and model optimization constraints. The predictive performance is influenced by the quality and scope of historical data, which may limit applicability across different marketplaces or changing market conditions. Additionally, the model’s effectiveness depends on careful code implementation and parameter tuning, and suboptimal configurations may reduce forecasting reliability in real-world deployments.

2.2. Customer Churn Prediction Using Machine Learning and Deep Learning

Ashish Suresh Awate et al. (2025) [6] advances this line of research by proposing an integrative framework that combines sentiment analysis with customer segmentation to enhance customer churn prediction. Utilizing Amazon customer reviews, the authors develop a Bidirectional Long Short-Term Memory (BiLSTM)-based sentiment prediction model capable of classifying customer sentiments into positive and negative categories with high accuracy and strong AUC performance. This sentiment-driven insight is then combined with a segmentation model that categorizes customers into five loyalty groups Champions, Loyalists, Potential Loyalists, At-Risk, and Detractors based on demographic and behavioral attributes. A key contribution of this work lies in its dynamic integration mechanism, wherein newly arriving customer data are continuously analyzed and used to update loyalty classifications. This adaptability allows businesses to detect early signs of customer dissatisfaction and potential churn in near real time. By blending sentiment-derived insights with segmentation dynamics, the framework offers a scalable and practical methodology for proactive churn management. Overall, the study contributes meaningfully to the literature on customer analytics by demonstrating how hybrid deep learning and segmentation models can improve customer retention strategies and support data-driven decision-making in evolving business environments.

Advantage: A significant advantage of this integrated framework is its ability to dynamically identify churn-prone customers by combining emotional cues from sentiment analysis with behavioral segmentation. The use of a BiLSTM-based sentiment model enhances the accuracy of understanding customer opinions, while loyalty-based segmentation enables actionable insights, allowing organizations to implement timely and targeted retention strategies. Disadvantage: Despite its effectiveness, the framework faces challenges related to scalability and domain-specific customization. As customer bases and data volumes grow, maintaining real-time performance and adapting the sentiment and segmentation models to different industries or regional contexts may require additional computational resources and extensive retraining, potentially limiting widespread adoption.

Ilham Huda et al. (2023) [7] propose a comprehensive data mining framework for predicting customer churn by integrating segmentation and classification techniques. The study employs the Recency, Frequency, and Monetary (RFM) model to segment customers based on their purchasing recency, transaction frequency, and monetary value. This segmentation is combined with classification models to distinguish between churn and non-churn customers using multiple behavioral variables, including session activity, application interaction, purchasing behavior, claims, add-to-cart actions, and discounts. The use of real transactional and interaction data strengthens the practical relevance of the proposed model. To evaluate classification performance, the study compares decision tree and Support Vector Machine (SVM) algorithms. The results demonstrate that the decision tree model achieves the highest accuracy, reaching approximately 87–88%, as validated through confusion matrix and ROC analysis. The findings further identify key factors influencing churn, such as product categories purchased, claim reasons, session behavior, and cart abandonment. Based on these insights, the study proposes actionable retention strategies, including improved stock management, targeted notifications, and incentive mechanisms such as vouchers and loyalty points. Overall, this research contributes both theoretical and practical insights into CRM-oriented data mining models, demonstrating how structured segmentation and interpretable classification can effectively support churn prediction in e-commerce environments.

Advantage: A major advantage of this study lies in its use of an interpretable and practical data mining framework that combines RFM-based segmentation with decision tree classification. The high accuracy achieved by the decision tree model enables businesses to clearly understand the factors influencing customer churn, facilitating actionable strategies such as inventory optimization, targeted promotions, and proactive customer engagement. Disadvantage: A key limitation of the proposed approach is its reliance on traditional data mining algorithms, which may it struggle to capture complex nonlinear relationships present in large-scale and highly dynamic e-commerce data. As customer behavior evolves rapidly, the model may require frequent recalibration and may not achieve the same adaptability and predictive power as advanced deep learning–based churn prediction systems.

Shanthini et al. (2025) [8] explores the application of Artificial Neural Networks to predict customer churn using a dataset of 1,644 customer records from e-commerce and subscription-based contexts. The research employs a Multilayer Perceptron (MLP) ANN architecture with hyperbolic tangent and Softmax activation functions to classify customers as churned or retained. A comprehensive set of independent variables—including delayed deliveries, frequent product returns, high cart abandonment, poor customer service, hidden fees, complicated return processes, negative reviews, and low engagement is used to capture multidimensional aspects of customer dissatisfaction. The model achieves an accuracy of 75.7%, demonstrating the effectiveness of ANN-based approaches in identifying churn patterns. Notably, hidden fees, complex return procedures, and low customer engagement emerge as the most influential churn predictors. These findings provide actionable insights for businesses to design personalized marketing strategies, improve service transparency, and enhance customer support mechanisms. The study aligns with prior research that highlights the superiority of deep learning models over traditional techniques in churn prediction, while also acknowledging challenges related to interpretability and data imbalance. Overall, this work contributes to the literature by reinforcing the applicability of ANN models in customer analytics and outlining directions for future enhancements using explainable AI and hybrid modeling techniques.

Advantage: A key advantage of this study is its ability to capture complex nonlinear relationships among diverse churn-related factors using an ANN-based model. By identifying critical predictors such as hidden fees, complicated return processes, and low engagement, the approach enables businesses to implement targeted retention strategies, personalized interventions, and proactive service improvements that can significantly reduce customer churn. Disadvantage: A notable disadvantage of the proposed approach is the limited interpretability of the ANN model, which makes it difficult for business stakeholders to fully understand how individual variables influence churn decisions. Additionally, data imbalance within the churn dataset may affect prediction reliability, indicating the need for improved preprocessing techniques and explainable AI methods to enhance model transparency and practical adoption.

Xiancheng Xiahou et al. (2022) [9] proposes a hybrid churn prediction framework that integrates k-means clustering for customer segmentation with Support Vector Machine (SVM) based prediction. Customers are first segmented into three distinct groups based on shopping behavior, enabling the identification of core customer segments. Subsequently, churn prediction is performed using SVM and compared against logistic regression (LR). Experimental results demonstrate that customer segmentation significantly improves all key prediction metrics, including accuracy, precision, recall, and AUC, confirming the necessity of segmentation prior to prediction. Moreover, the SVM model consistently outperforms the LR model, highlighting the superiority of machine learning approaches over traditional statistical methods for churn prediction. The research validates its framework using a large real-world dataset comprising nearly one million B2C customers, underscoring its practical relevance. By demonstrating how segmentation enhances predictive performance and how SVM delivers superior results compared to logistic regression, the study contributes valuable insights to the literature on data-driven customer retention strategies. However, the authors acknowledge methodological and data-related limitations, suggesting that future studies should explore alternative segmentation methods, incorporate a broader set of behavioral variables, and validate findings across multiple datasets to improve generalizability.

Advantage: A key advantage of this study is its effective combination of customer segmentation and machine learning based prediction. By applying k-means clustering prior to churn prediction, the framework significantly improves predictive performance, and the use of SVM enables more accurate identification of potential churners, supporting targeted retention strategies and informed decision-making in B2C e-commerce environments. Disadvantage: A notable limitation of the proposed approach is its reliance on a single segmentation technique and a limited set of predictive variables. Using only k-means clustering may restrict segmentation quality, and ignoring additional behavioral features available in B2C platforms may limit the model's generalizability and robustness when applied to diverse datasets or different e-commerce contexts.

Pondel Maciej et al. (2021) [10] proposes a deep learning–based model for customer churn prediction using real-world e-commerce data with extreme class imbalance, where approximately 75% of customers are one-time buyers and only 2% are regular purchasers. By incorporating the complete transaction history of each customer, the model aims to better represent purchasing behavior in the retail sector. Despite the challenging data characteristics, the proposed approach achieves promising results, reporting 74% accuracy, 78% precision, and 68% recall. While such metrics may appear modest in balanced datasets, they are considered meaningful within this business context due to the inherent uncertainty in defining churn for one-off customers. The research contributes to the literature by addressing a realistic and underexplored retail scenario and demonstrating that deep learning models can deliver actionable insights even under severe data imbalance. The authors acknowledge the preliminary nature of their work and highlight the need for more advanced feature selection methods, improved dataset partitioning strategies, and clearer temporal definitions of churn. Overall, this study advances understanding of churn prediction in retail e-commerce and lays the groundwork for more refined, behavior-aware predictive frameworks.

Advantage: A key advantage of this study is its use of real-world, highly imbalanced e-commerce data to develop a deep learning based churn prediction model. By leveraging the full transaction history of customers, the approach captures nuanced behavioral patterns and demonstrates that meaningful predictive performance can be achieved even in challenging retail scenarios dominated by one-off buyers, making it practically relevant for customer retention strategies. *Disadvantage:* A notable limitation of the proposed approach is its reliance on basic feature selection and data sampling techniques. Using only filter-based feature selection and random dataset splitting may limit model robustness and fail to preserve customer-level temporal consistency, while the lack of a clear definition of the churn timing point constrains the precision and interpretability of churn predictions in retail contexts.

Md Sakibul Hasan et al. (2024) [11] addresses this challenge by developing high-precision machine learning models for customer churn prediction using a comprehensive dataset that combines demographic attributes, transactional histories, and behavioral patterns. The dataset incorporates static customer information such as age, gender, location, and account tenure, alongside dynamic indicators including purchase frequency, basket size, total spending, payment preferences, discount usage, order completion and return behavior, recency of transactions, and product category preferences. This multidimensional feature design enables a richer representation of evolving customer interests and engagement levels. To evaluate predictive performance, the study compares multiple machine learning models using a systematic training and testing strategy. Experimental results show that XGBoost consistently outperforms other approaches, achieving performance scores above 0.9 across key evaluation metrics, while Random Forest also demonstrates strong results with scores generally exceeding 0.85. Beyond model comparison, the research proposes a machine learning based churn alert system capable of monitoring customer behavior in near real time and assigning churn risk dynamically. The findings reinforce the value of predictive churn analytics for revenue protection and strategic decision-making in e-commerce, while highlighting the importance of incorporating increasingly rich and granular data sources to further enhance model effectiveness.

Advantage: A major advantage of this study is its comprehensive integration of demographic, transactional, and behavioral data to achieve highly accurate churn prediction. By leveraging ensemble learning models such as XGBoost and Random Forest, the proposed approach delivers strong predictive performance and supports the implementation of real-time churn alert systems, enabling e-commerce platforms to proactively deploy targeted retention strategies and safeguard revenue streams. Disadvantage: A key limitation of the proposed approach is its strong dependence on the availability and quality of rich, detailed customer data. Collecting, maintaining, and processing such multidimensional datasets can be resource-intensive and may raise privacy and governance concerns, potentially limiting the scalability and generalizability of the churn prediction framework across smaller platforms or data-constrained environments.

2.3. Customer Segmentation and Behavior Analysis in E-Commerce

Ritu Punhani et al. (2021) [12] investigates customer segmentation in a B2B e-commerce setting using transactional and demographic data from Autofurnish.com across two consecutive years (2018 and 2019). Using the RapidMiner platform, the authors apply the k-means clustering algorithm to segment customers based on multiple attributes and compare behavioral trends across the two periods. The analysis reveals a decline in total users from 2018 to 2019, with the highest customer activity concentrated in a specific geographic region (Agra) and age group (25–34 years). These insights are leveraged to propose targeted strategies aimed at increasing customer influx and avoiding ineffective practices in future operations. Overall, the study demonstrates the practical value of clustering-based customer segmentation for understanding temporal shifts in customer behavior within B2B e-commerce. By linking demographic and geographic patterns to platform performance, the research contributes actionable insights for managerial decision-making and highlights how open-source data mining tools can support strategic planning in competitive digital markets.

Advantage: A key advantage of this study is its practical application of k-means clustering to real-world B2B e-commerce data, enabling clear identification of high-value customer segments based on age and location. The use of an accessible tool such as RapidMiner allows for rapid analysis and actionable insights, supporting targeted marketing strategies and informed decision-making to improve customer acquisition and platform performance. *Disadvantage:* A notable limitation of the approach is its reliance on basic clustering techniques and a limited temporal scope. Using only k-means clustering and data from two years restricts the ability to capture complex behavioral dynamics, revenue-driven outcomes, or evolving customer preferences, suggesting the need for more advanced models and longer-term analysis in future research.

Qing Zhang et al.,(2022) [13] investigates the application of data mining and neural network technologies for customer behavior analysis and purchasing power prediction in e-commerce management systems. The research employs clustering techniques and Naïve Bayes classification to categorize product information and consumer purchase preferences, enabling the extraction of hidden behavioral patterns. Subsequently, neural network models, including convolutional neural networks (CNNs), are utilized to predict future consumer purchasing power and consumption trends, effectively addressing the nonlinear characteristics of e-commerce data. Empirical results demonstrate strong predictive performance, with a high correlation coefficient of 0.9785 between actual and predicted consumption values and a maximum relative average error of only 2.32%. The study further highlights that uncertainty-aware clustering methods yield better classification effectiveness when distinguishing customer groups with significant behavioral differences. By combining clustering-based segmentation with neural network–driven prediction, the proposed framework captures seasonal trends, monthly purchasing behaviors, and varying customer value patterns. Overall, this research contributes valuable insights into the design of intelligent e-commerce management systems, illustrating how integrated data mining and deep learning approaches can support accurate customer value prediction and strategic business planning.

Advantage: A key advantage of this approach is its high predictive accuracy in modeling customer purchasing power and behavior. The integration of clustering techniques with neural networks enables the system to uncover hidden relationships in consumer data and effectively capture nonlinear purchasing patterns, resulting in reliable forecasts that support informed decision-making and long-term strategic planning in e-commerce management. Disadvantage: A notable limitation of the study is the complexity of the proposed hybrid framework, which combines multiple data mining and neural network techniques. Such complexity may increase computational cost and implementation difficulty, particularly for small or resource-constrained e-commerce platforms, and may require careful parameter tuning and domain expertise to ensure stable and scalable real-world deployment.

Yasser D. Al-Otaibi (2024) [14] proposes a deep learning–based framework to predict consumer buying behavior using demographic attributes, specifically age and salary. The research introduces a lightweight deep learning architecture with dense layers trained on publicly available datasets to classify whether a customer is likely to make a purchase. The model demonstrates strong predictive performance, achieving high precision, recall, and F1 scores, indicating its effectiveness in identifying potential buyers. By emphasizing demographic-driven prediction, the study highlights the continued relevance of age and income-related features in consumer behavior analysis. Furthermore, the research outlines a structured workflow that integrates deep learning predictions into e-commerce recommendation and marketing personalization strategies. The findings suggest that such models can significantly enhance customer experience and operational performance by enabling targeted marketing interventions. While the study acknowledges its scope limitations, it establishes a solid foundation for future research by recommending the inclusion of additional demographic and behavioral variables, real-time data streams, and advanced scaling and hyperparameter optimization techniques. Overall, this work contributes to the growing literature on AI-enabled consumer behavior prediction, emphasizing the role of deep learning in data-driven e-commerce strategy formulation.

Advantage: A key advantage of the proposed approach is its simplicity combined with strong predictive performance. By utilizing a lightweight deep learning architecture and easily obtainable demographic variables such as age and salary, the model achieves high accuracy and reliable classification outcomes, making it practical for integration into e-commerce recommendation and personalized marketing systems. Disadvantage: A notable limitation of the study is its reliance on a narrow set of input variables, which restricts the model's ability to capture the full complexity of consumer behavior. Factors such as browsing patterns, purchase history, sentiment, and real-time interactions are not considered, potentially limiting the model's generalizability and effectiveness in highly dynamic and diverse e-commerce environments.

2.4. Big Data Analytics and AI-Driven Decision Support in E-Commerce

Priya Chugh et al. (2024) [15] provides a comprehensive conceptual overview of AI research in e-commerce by analyzing 1,458 scholarly articles indexed in the Scopus database between 1995 and 2024. Using R Studio and VOSviewer, the authors employ performance analysis, clustering techniques, and network visualization to uncover the conceptual structure and thematic evolution of AI-driven e-commerce research. The findings reveal a strong concentration on advanced analytics applications, particularly AI-based product recommendation systems and intelligent customer support solutions such as chatbots and feedback analysis tools. The study highlights the interdisciplinary nature of AI in e-commerce, spanning computer science, marketing, psychology, and operations management. It demonstrates how AI-driven personalization, inventory optimization, supply chain efficiency, churn prediction, and security enhancement collectively contribute to competitive differentiation. A notable contribution of this research lies in its use of the TCCM (Theory–Context–Characteristics–Methodology) framework, which offers structured insights for future research directions and practical guidance for organizations seeking to integrate AI into their e-commerce strategies. Overall, the work positions AI as a foundational technology for innovation, adaptability, and sustained competitive advantage in the rapidly evolving digital retail ecosystem.

Advantage: A key advantage of this study is its comprehensive and systematic mapping of AI research in e-commerce over nearly three decades. By combining bibliometric analysis, visualization techniques, and the TCCM framework, the research provides valuable strategic insights for both scholars and practitioners, enabling informed decision-making regarding AI investment, research prioritization, and practical implementation across e-commerce operations. Disadvantage: A notable limitation of the study is its reliance on secondary data sourced exclusively from the Scopus database, which may exclude relevant research published in non-indexed journals or industry reports. As a result, certain practical innovations or emerging AI applications in e-commerce may not be fully captured, potentially limiting the completeness of the conceptual landscape.

Pooja Pande et al. (2025) [16] proposes a comprehensive Big Data Analytics framework designed to enhance understanding of customer behavior and improve decision-making in e-commerce. The framework integrates machine learning, predictive analytics, and customer segmentation to address key challenges such as demand forecasting, customer engagement, and churn prediction. A notable contribution of this research is its emphasis on privacy-preserving techniques and ethical data governance, ensuring compliance with regulations such as GDPR and CCPA while maintaining consumer trust. The study demonstrates that deep learning-based predictive models are particularly effective in identifying purchase intent, preferences, and churn risk, enabling highly personalized recommendations and marketing strategies. Furthermore, the research highlights the importance of scalability and accessibility of Big Data solutions, especially for small and medium-sized enterprises (SMEs). By leveraging cloud-based platforms and AI-driven tools, the proposed framework lowers the barrier to entry for advanced analytics, allowing organizations with limited resources to benefit from data-driven insights. By bridging academic theory and practical implementation, this work contributes to the literature by offering a scalable, ethical, and application-oriented model for deploying Big Data Analytics in e-commerce, supporting real-time decision-making, enhanced customer experience, and long-term business growth.

Advantage: A major advantage of this framework is its holistic integration of advanced analytics with privacy-preserving mechanisms. By combining machine learning–driven customer insights with ethical data governance, the approach enables e-commerce businesses to deliver personalized experiences and accurate demand forecasts while maintaining regulatory compliance and customer trust, thereby supporting sustainable growth. *Disadvantage:* A key limitation of the proposed framework is its reliance on sophisticated data engineering and analytical capabilities, which may still pose implementation challenges despite cloud-based scalability. Organizations with limited data maturity or expertise may struggle to effectively operationalize advanced predictive models and privacy mechanisms, potentially reducing the immediate impact of the framework in real-world settings.

Cheng-Yuan Chen et al. (2025) [17] explores the application of big data mining and analytics algorithms in optimizing portfolio investment decisions within intelligent financial systems. By integrating data mining outputs with big data analysis and visualization, the proposed framework enhances decision support for financial portfolios. Experimental results indicate that incorporating big data technologies improves the operational efficiency of the intelligent financial portfolio decision model by approximately 35.3%, demonstrating tangible performance gains. The research further discusses the role of secure intelligent services, including data encryption and automated monitoring platforms, to support real-time analysis, classification, and basic prediction of financial data. Beyond technical optimization, the study emphasizes organizational transformation, highlighting the need for enterprises and financial practitioners to adapt skills and workflows to leverage intelligent finance effectively. While acknowledging that current theories of security and intelligent services remain imperfect, the authors argue that big data integration compensates for existing shortcomings by improving efficiency, risk management, and competitive vitality. Overall, this work contributes to the literature by evidencing how big data mining and intelligent services can jointly enhance portfolio decision-making and enterprise financial management in evolving digital economies.

Advantage: A key advantage of the proposed approach is its demonstrated improvement in decision-making efficiency through the integration of big data mining and intelligent financial models. By combining data mining, analytics, visualization, and secure intelligent services, the framework significantly enhances portfolio investment decisions and operational efficiency, enabling enterprises to better manage risks and respond proactively to financial challenges. *Disadvantage:* A notable limitation of the study is the immaturity of existing security and intelligent service theories within intelligent finance. Although big data integration mitigates some shortcomings, unresolved issues related to security robustness, model generalizability, and theoretical completeness may constrain broader adoption and require further empirical validation across diverse financial contexts.

2.5. IoT-Enabled Intelligent Analytics for E-Commerce Decision-Making

Yasser Filahi et al. (2025) [18] investigates how IoT-enabled data acquisition and machine learning models can jointly improve e-commerce decision-making. Customer behavior data collected via IoT devices are analyzed using a wide range of machine learning and deep learning techniques, including logistic regression, Naïve Bayes, SVM, Random Forest, AdaBoosting, GRU, and LSTM. Beyond recommendation generation, the framework addresses demand forecasting to ensure product availability and reduce stock-related dissatisfaction. Extensive experimentation shows that AdaBoosting consistently outperforms both traditional machine learning and deep learning models, achieving superior accuracy, F1-score, precision, and recall particularly in handling imbalanced sentiment datasets. The study further demonstrates the effectiveness of TF-IDF vectorization for sentiment-based recommendation tasks. Overall, this research advances the literature by illustrating how IoT and machine learning can be combined to deliver data-driven insights that optimize operations, enhance personalization, and improve customer satisfaction in e-commerce. The authors also outline future research directions, including addressing IoT data heterogeneity, strengthening privacy and security through techniques such as federated learning, and exploring advanced deep learning architectures to further enhance predictive performance. This work positions IoT-driven intelligent analytics as a key enabler of scalable, efficient, and adaptive e-commerce strategies.

Advantage: A major advantage of this approach is its effective fusion of IoT-generated real-time data with machine learning analytics to support comprehensive e-commerce decision-making. By leveraging diverse algorithms and demonstrating the strong performance of AdaBoost on imbalanced datasets, the framework enables accurate recommendations, reliable demand forecasting, and improved customer satisfaction, thereby supporting customer-centric and operationally efficient retail strategies. *Disadvantage:* A key limitation of the proposed framework lies in the challenges associated with IoT data heterogeneity and privacy. The integration of large volumes of heterogeneous data from multiple IoT devices requires robust preprocessing, normalization, and fusion techniques, while the collection of sensitive real-time customer data raises security and privacy concerns that must be carefully addressed to ensure trustworthy and scalable real-world deployment.

2.6. Dynamic Pricing Strategies Using Data Mining and Reinforcement Learning

Chunli Yin et al. (2021) [19] investigates intelligent dynamic pricing for e-commerce platforms by combining data mining methods with deep reinforcement learning and game-theoretic modeling. The authors design a two-period dynamic pricing game to analyze optimal pricing strategies under varying market structures (single vs. multiple suppliers), consumer participation levels, and market maturity (mature versus emerging markets). By examining duopoly settings and comparing pricing strategies across different consumer types naive and experienced the study derives Nash and subgame-perfect Nash equilibria to identify profit-maximizing strategies. In addition, the research proposes a data mining–based dynamic pricing model that incorporates auction mechanisms and analyzes substitution effects in multiproduct environments. The findings suggest that intelligent dynamic pricing tools grounded in data mining can improve customer satisfaction and economic efficiency for e-commerce enterprises. While the model demonstrates general applicability across platforms, the study acknowledges limitations in fully integrating production planning with pricing decisions, indicating an important direction for future research. Overall, this work contributes to the literature by linking reinforcement learning, data mining, and game theory to advance adaptive pricing strategies in e-commerce.

Advantage: A major advantage of this study is its comprehensive integration of data mining, reinforcement learning, and game-theoretic analysis to model dynamic pricing in diverse market conditions. By accounting for supplier competition and heterogeneous consumer behavior, the proposed framework offers theoretically grounded and practically relevant insights that can help e-commerce platforms optimize pricing strategies and improve profitability. Disadvantage: A key limitation of the proposed approach is its limited integration of production planning with dynamic pricing decisions. While production constraints are considered, the lack of a tightly coupled pricing–production optimization framework restricts the model's ability to fully reflect real-world operational dependencies, suggesting the need for more holistic models in future research.

Salvatore Carta et al. (2019) [20] introduce *Price Probe*, a software framework designed to forecast future product prices in e-commerce platforms. The approach primarily relies on the Autoregressive Integrated Moving Average (ARIMA) model, enhanced with exogenous variables derived from external data sources. Using Amazon as a large-scale experimental platform, the authors collected extensive historical price data via APIs and web crawlers, alongside popularity indicators from social media and Google Trends. The inclusion of Google Trends as an external feature was shown to significantly improve prediction accuracy, demonstrating the value of incorporating consumer interest signals into traditional time-series models. The research highlights how combining structured price histories with external behavioral indicators enables more precise forecasting of price trends, benefiting both consumers and businesses. The study also emphasizes the scalability of the proposed solution, as it was evaluated on a very large dataset compared to similar works. While the focus remains on ARIMA-based modeling, the authors acknowledge the potential of extending the framework to neural network architectures and financial technology applications, such as stock market prediction, by integrating social media and news-driven signals. Overall, this work contributes to the literature by demonstrating how hybrid time-series and external-feature–based models can enhance predictive analytics in large-scale e-commerce environments.

Advantage: A key advantage of the proposed Price Probe framework is its effective integration of external behavioral signals, particularly Google Trends, into a well-established time-series forecasting model. This hybrid approach significantly improves price prediction accuracy while maintaining interpretability and scalability, making it practical for real-world e-commerce platforms with large volumes of historical pricing data. Disadvantage: A notable limitation of the approach is its strong dependence on the selection and availability of external features. Since the forecasting performance is highly influenced by how exogenous data such as Google Trends are chosen and collected, the model's generalizability may be constrained, and important factors like social media sentiment or product reviews may remain underutilized without further extension.

2.7. Recommendation Systems and Personalized Marketing Using Deep Learning

Minseo Park et al. (2024) [21] addresses this limitation by proposing a transformer-based recommendation framework that explicitly incorporates concurrent purchase information. Using data from a Korean e-commerce platform, the authors observe that concurrent purchases account for approximately 23% of all orders, underscoring their practical importance. The proposed method represents each order as a natural language–like sentence encoding timestamps, product names, attributes, categories, and indicators of concurrent purchases. These sentences are modeled as sequences and used to fine-tune a BERT-based architecture using the Next Sentence Prediction objective. Experimental results demonstrate that integrating concurrent purchase information significantly improves recommendation performance across multiple evaluation metrics, including accuracy, F1-score, and particularly normalized discounted cumulative gain (NDCG). The findings highlight the effectiveness of combining structured transactional data with advanced natural language processing techniques to better capture nuanced consumer purchasing patterns. The study further positions concurrent purchase modeling as a critical yet underexplored feature in recommender system research and outlines future directions involving large-scale generative language models and encoder–decoder transformer architectures to overcome current limitations.

Advantage: A key advantage of this approach is its innovative integration of concurrent purchase behavior into transformer-based recommendation systems. By representing purchase histories as semantically rich sequences and leveraging BERT's contextual modeling capabilities, the framework captures complex co-purchase relationships that are typically overlooked, leading to substantial improvements in recommendation accuracy and ranking quality. *Disadvantage:* A notable limitation of the proposed method is its dependence on BERT's restricted context length, which may constrain the model's ability to fully represent long-term user behavior and extended purchase histories. This limitation can affect scalability and performance in large-scale e-commerce environments, motivating the need for alternative architectures such as long-context generative or encoder–decoder transformer models.

2.8. Product Classification, Attribute Extraction, and Vision–Language Models

Fan Liu et al. (2023) [22] addresses these limitations by introducing MEP-3M, a large-scale multi-modal e-commerce product classification dataset comprising over three million products. Each product is represented using images, textual descriptions, and OCR-extracted text, and annotated with hierarchical, tree-structured labels that include highly fine-grained third-level categories. The dataset supports multiple practical research tasks, including standard product classification, hierarchical classification, fine-grained categorization, and product representation learning. Baseline results using image-only, text-only, and multi-modal models are provided, offering a strong benchmark for future research. A key contribution of this work lies in its alignment with recent advances in vision–language modeling, positioning MEP-3M as a valuable resource for training and evaluating multi-modal foundation models in the e-commerce domain. By capturing real-world challenges such as category imbalance and long-tailed distributions, the dataset facilitates the development of more robust and generalizable classification systems. Despite acknowledged limitations related to category coverage and data noise, the study establishes MEP-3M as a significant step forward in supporting large-scale, multi-modal learning for e-commerce applications.

Advantage: A major advantage of the MEP-3M dataset is its unprecedented scale and rich multi-modality, which enable comprehensive exploration of vision–language learning in e-commerce. The combination of images, textual descriptions, and OCR text with hierarchical and fine-grained labels makes it particularly well suited for training robust product classification models and pre-training e-commerce specific foundation models that reflect real-world data complexity. *Disadvantage:* A key limitation of the dataset is its incomplete and evolving category coverage, as e-commerce product taxonomies change rapidly over time. Additionally, the presence of noisy images and textual descriptions may affect model performance and generalizability, necessitating further data cleaning, label denoising, and continuous dataset updates for sustained real-world applicability.

Mehmet Serhan Çiftlikçi et al. (2025) [23] proposes an LLM-based attribute extraction framework for Trendyol's Turkish-language product catalog and compares it against a transformer-based deep learning NER model. Built on the Mistral architecture, the LLM demonstrates superior performance across precision, recall, and F1-score, particularly in handling complex linguistic structures and diverse product descriptions. A notable contribution is the multi-level prediction capability, enabling not only attribute extraction but also hierarchical classification across product categories and category–attribute combinations, with reported F1-score improvements of up to 48.27% over the DL baseline in hierarchical tasks. Beyond accuracy, the study validates real-world feasibility through deployment within Trendyol's production environment, using Kubernetes and Nvidia Triton Inference Server to support both bulk processing and real-time attribute suggestions during product listing. This integration reduces manual errors, improves catalog consistency, and scales to millions of products. While performance is strong overall, the authors acknowledge challenges in recall for highly implicit or ambiguous attributes, indicating opportunities for further domain-specific fine-tuning. Overall, the work advances automated product classification and information retrieval by demonstrating the practical advantages and trade-offs of LLMs in large-scale e-commerce systems.

Advantage: A key advantage of the proposed approach is its superior contextual understanding enabled by large language models, allowing accurate extraction of both explicit and implicit product attributes and robust hierarchical classification. This capability significantly improves catalog quality, searchability, and operational efficiency, while the successful production deployment demonstrates scalability and real-time applicability in a large e-commerce marketplace. Disadvantage: A notable limitation is the variability in recall for highly implicit or ambiguous attributes, which indicates sensitivity to domain-specific nuances in product descriptions. Addressing these cases may require additional fine-tuning, curated training data, and increased computational resources, potentially raising deployment and maintenance costs for sustained high performance.

Ling Yang et al. (2024) [24] proposes an enhanced e-commerce data processing model that combines data mining techniques with an improved K-Nearest Neighbors (KNN) classification algorithm. The framework employs a dimensional control mechanism alongside the Spark distributed computing platform to efficiently mine massive e-commerce datasets. Following data mining, a KNN algorithm enhanced with an adaptive K-value selection strategy is used to classify the processed data, thereby improving classification accuracy and reducing computational overhead. Experimental results demonstrate that the proposed data mining algorithm significantly reduces mining time to 4.6 minutes and achieves a low error rate of 4.2%, representing nearly a 50% improvement over comparative algorithms. Furthermore, the improved KNN classifier attains a recognition rate of approximately 97.3% while reducing classification time by over 90% compared to traditional KNN and KNN-based clustering approaches. These findings indicate that the proposed model effectively balances accuracy and efficiency, enabling precise marketing and strategic decision-making for e-commerce platforms. However, the study also acknowledges that the dataset used is limited in scope, suggesting the need for further validation on large-scale real-world datasets.

Advantage: A key advantage of the proposed approach is its substantial improvement in both classification accuracy and processing efficiency. By integrating distributed data mining through Spark with an optimized KNN algorithm, the model effectively handles large e-commerce datasets, reduces computational time, and delivers highly accurate classification results, making it well suited for precision marketing and real-time decision support. *Disadvantage:* A notable disadvantage of the study is its reliance on a relatively limited dataset for experimental evaluation. While the results demonstrate strong performance, the generalizability of the proposed algorithms to diverse, real-world e-commerce environments remains uncertain, necessitating further testing on larger and more heterogeneous datasets.

2.9. Website Usability and Web Analytics Using Machine Learning

Biresh Kumar et al. (2023) [25] investigates the application of advanced web metrics derived from Google Analytics and log-based transactional data to evaluate e-commerce website usability. The authors propose a machine learning–based evaluation framework that integrates web analytics algorithms with a KMP algorithm–based multivariate pruning method to efficiently mine patterns from large-scale web data. Multiple supervised learning models, including random forest, logistic regression, and naïve Bayes, are employed alongside multiple linear regression to predict overall usability based on factors such as engagement, effectiveness, sorting behavior, visit frequency, and user interaction patterns. The results demonstrate high predictive performance, with some supervised learning models achieving accuracy rates as high as 98.9%, outperforming conventional web-based mining approaches. The study further emphasizes that usability assessment should extend beyond basic metrics such as page views and visitor counts, incorporating conversion behavior, motivation, and engagement depth. By combining web analytics with machine learning and visualization techniques, the research contributes a novel, data-driven approach to usability evaluation, while also highlighting the need for continuous model adaptation as e-commerce data characteristics evolve over time.

Advantage: A major advantage of this approach is its ability to objectively and accurately evaluate e-commerce website usability using large-scale web and transactional data. By leveraging machine learning models and advanced web analytics, the framework achieves high prediction accuracy and provides actionable insights into user behavior, enabling vendors to tailor website design, content, and functionality to better meet customer needs and improve overall platform performance. *Disadvantage:* A key limitation of the proposed method is the limited interpretability and deployment readiness of complex machine learning models. Despite their high accuracy, concerns related to explainability, integration into reusable systems, and reliance on evolving and heterogeneous web data may hinder practical adoption, necessitating further research into transparent models and robust real-world implementation strategies.

2.10. Systematic Reviews and Research Trends in E-Commerce Analytics

Mehdi Imani et al. (2025) [26] addresses this gap by synthesizing research on ML and DL–based churn prediction published between 2020 and 2024. Following PRISMA 2020 guidelines, the authors conducted a comprehensive review across six major scientific databases, resulting in 240 studies for bibliometric analysis and 61 studies for in-depth qualitative synthesis. The review reveals that ensemble learning methods such as XGBoost, LightGBM, and CatBoost dominate ML-based churn prediction due to their robustness and performance, while DL models including LSTM, CNN, and attention-based architectures are increasingly applied to capture temporal and high-dimensional customer data. The study highlights important methodological trends and challenges, including class imbalance, concept drift, limited interpretability, and an overreliance on accuracy-based evaluation metrics. Although explainable AI techniques such as SHAP and LIME show promise in improving model transparency, their real-world adoption remains limited. Furthermore, the review identifies a lack of business-aligned performance metrics and fairness-aware modeling as critical research gaps. By combining bibliometric mapping with structured qualitative synthesis, this work provides a comprehensive roadmap for advancing churn prediction research toward adaptive, interpretable, and business-relevant solutions.

Advantage: A major advantage of this study is its comprehensive and methodologically rigorous synthesis of recent churn prediction research. By systematically reviewing a large body of peer-reviewed studies and identifying dominant models, emerging trends, and unresolved challenges, the review offers valuable guidance for researchers and practitioners seeking to design robust, adaptive, and business-aligned churn prediction systems. Disadvantage: A key limitation of the review is the absence of formal risk-of-bias assessment and meta-analysis due to study heterogeneity. This limits the ability to quantitatively compare model effectiveness across studies and may reduce the strength of generalized conclusions regarding the superiority of specific ML or DL techniques in real-world churn prediction scenarios.

Summary of Literature Review,

Author(s)/ Year	Method / Technique	Dataset	Key Findings	Pros	Cons
Md. Mortuza Ahmed et al.(2025)	Conditional Transformer Language Model (CTRL)	Amazon Reviews 2018	CTRL significantly improved demand forecasting accuracy over traditional ML models	High contextual understanding; improved predictive accuracy	High computational cost; low interpretability
Norun Nabi et al.(2024)	CNN (VGG16, ResNet50, InceptionV3)	E-commerce profit dataset	VGG16 achieved highest accuracy (92.55%) in	Strong feature extraction; high accuracy	Requires high-quality datasets and infrastructure

			profit prediction		
Ashish Suresh Awate et al. (2025)	BiLSTM + Customer Segmentation	Amazon Reviews	Dynamic churn prediction with sentiment-driven loyalty segmentation	Adaptive and scalable churn detection	Scalability and domain customization challenges
Ilham Huda et al. (2023)	RFM + Decision Tree, SVM	B2C Transaction Data	Decision Tree achieved ~87–88% churn prediction accuracy	Interpretable and actionable results	Limited scalability; ignores deep behavioral patterns
Shanthini et al. (2025)	ANN (MLP)	1,644 customer records	Identified hidden fees and low engagement as main churn drivers	Captures nonlinear relationships	Interpretability and data imbalance issues
Ling Yang et al. (2024)	Improved KNN + Spark	E-commerce platform data	97.3% classification accuracy with reduced time	Fast processing; high accuracy	Limited dataset; real-world generalization unclear
Ron Kohavi et al. (2024)	Integrated Data Mining Architecture	Blue Martini / Amazon-like systems	Application-layer data logging improves analytics quality	Eliminates preprocessing bottlenecks	High infrastructure and setup cost
Priya Chugh et al. (2024)	Bibliometric & TCCM Analysis	Scopus (1995–2024)	AI research in e-commerce focuses on recommendation & support	Comprehensive research mapping	Excludes non-Scopus literature
Qing Zhang et al. (2023)	Clustering + CNN + Neural Networks	Consumer purchase data	Achieved correlation of 0.9785 in purchase	High predictive accuracy	Complex hybrid framework
			prediction		
Yasser D. Al-Otaibi (2024)	Deep Learning (Dense Layers)	Public demographic datasets	Age and salary effectively predict buying behavior	Simple and effective model	Limited behavioral features
Salvatore Carta et al. (2019)	ARIMA + Google Trends	Amazon price data	External signals improve price prediction accuracy	Interpretable time-series forecasting	Heavily dependent on external data quality
Jianhao Zhang et al,. (2023)	XGBoost, LightGBM, GBDT Sales Forecasting Study (UCI)	UCI ML Repository	XGBoost handled holiday demand spikes best	Strong generalization and robustness	Data dependency and tuning complexity
Pooja Pande et al. (2025)	Big Data + ML + Privacy-Preserving Analytics	Multi-source e-commerce data	Improved personalization and trust via ethical analytics	Scalable and GDPR-compliant	Requires advanced data engineering
Yasser Filahi et al. (2025)	IoT + ML + DL (AdaBoost, LSTM)	IoT-based customer data	AdaBoost outperformed DL models (88% accuracy)	Effective on imbalanced datasets	IoT data heterogeneity and privacy risks
Xiancheng Xiaohou et al. (2022)	K-means + SVM	B2C behavioral data	Segmentation improved churn prediction significantly	Simple and effective hybrid model	Limited segmentation techniques
Cheng-Yuan Chen et al. (2025)	Big Data Mining + Intelligent Finance	Financial transaction data	Portfolio decision efficiency improved by 35.3%	Enhances financial decision-making	Security theory still evolving
Mehdi Imani et al. (2025)	Systematic Review (ML & DL)	240 studies (2020–2024)	Identified gaps in interpretability and real-	Comprehensive and structured synthesis	No meta-analysis performed

			world use		
Biresh Kumar et al. (2023)	ML + Web Analytics + KMP	Google Analytics logs	98.9% usability prediction accuracy	Objective usability evaluation	Explainability and deployment challenges
Md Sakibul Hasan et al. (2024)	XGBoost, Random Forest	US e-commerce churn dataset	XGBoost achieved >0.9 performance scores	Near real-time churn alerts	Data collection and privacy concerns
Ritu Punhani et al. (2021)	K-means Clustering	Autofurnish.com (2018–2019)	Identified key customer segments by age/location	Easy implementation	Limited temporal and revenue analysis
Chunli Yin et al. (2021)	Reinforcement Learning + Game Theory	Simulated e-commerce markets	Optimized dynamic pricing strategies	Theoretically grounded pricing model	Weak integration with production planning
Pondel Maciej et al. (2021)	Deep Learning	Retail e-commerce data	Promising churn results despite imbalance	Realistic retail scenario	Weak feature selection methods
Minseo Park et al. (2024)	BERT-based Recommendation Model	Katcher's (Korea)	NDCG improved significantly with concurrent purchases	Captures co-purchase behavior	Limited BERT context length
Mehmet Serhan Çiftlikçi et al. (2025)	LLM (Mistral) vs DL-NER	Trendyol product catalog	LLM improved F1-score up to 48%	Extracts implicit attributes	High computational cost
Fan Liu et al. (2023)	Multi-modal Learning (Vision + Text)	MEP-3M (3M products)	Supports fine-grained product classification	Large-scale, multi-modal dataset	Noisy data; evolving categories

III. CONCLUSION

This review paper presented a comprehensive synthesis of contemporary research on artificial intelligence, machine learning, deep learning, and big data analytics in the e-commerce domain. By systematically analyzing studies across diverse application areas including demand and sales forecasting, profit prediction, customer churn analysis, segmentation and behavior modeling, recommendation systems, dynamic pricing, product classification, website usability evaluation, and intelligent decision support the paper highlighted how data-driven methodologies have become central to modern e-commerce ecosystems. The reviewed literature collectively demonstrates a clear paradigm shift from traditional statistical and rule-based techniques toward advanced AI-driven models capable of handling large-scale, heterogeneous, and dynamic data. The findings reveal that deep learning and ensemble learning techniques, such as transformers, CNNs, RNNs, XGBoost, and large language models, consistently outperform conventional methods in predictive accuracy and pattern discovery. These models enable e-commerce platforms to better capture complex consumer behavior, temporal dynamics, sentiment signals, and contextual dependencies, thereby supporting more informed decision-making, personalization, and operational optimization. Furthermore, emerging trends such as multi-modal learning, IoT-enabled analytics, and privacy-preserving big data frameworks highlight the growing maturity and diversification of AI applications in digital commerce. Despite these advances, the review also identifies several persistent challenges that limit large-scale and real-world deployment. Issues related to model interpretability, data imbalance, scalability, computational cost, data quality, privacy, and regulatory compliance remain significant barriers, particularly for small and medium-sized enterprises. Many high-performing deep learning models function as black-box systems, reducing transparency and trust, while the increasing reliance on sensitive customer data raises ethical and governance concerns. Additionally, the lack of standardized datasets, business-aligned evaluation metrics, and adaptive mechanisms to handle concept drift constrains the generalizability of existing solutions.

REFERENCES

[1] Dwivedi, Y. K., Hughes, L., Ismagilova, E., Aarts, G., Coombs, C., Crick, T., Williams, M. D. (2021), "Artificial intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and polic", International Journal of Information Management, 57, 101994. https://doi.org/10.1016/j.ijinfomgt.2019.08.002.

[2] Verma, S., Sharma, R., Deb, S., & Maitra, D. (2021), "Artificial intelligence in marketing: Systematic review and future research direction", International Journal of Information Management Data Insights, 1(1), 100002. https://doi.org/10.1016/j.jjimei.2020.100002.

[3] Md. Mortuza Ahmmed, M. Mostafizur Rahman, Mufti Mahmud, "Enhancing E-commerce Dynamics through Deep Learning-Based Customer Demand Prediction", Transportation Research Procedia, Volume 84, 2025, Pages 145-152, ISSN 2352-1465, https://doi.org/10.1016/j.trpro.2025.03.057.

[4] Norun Nabi, Md Amran Hossen Pabel, Mohammad Anisur Rahman, Md Abu Sufian Mozumder, Md Al-Imran, Md Murshid Reja Sweet, Md Zahidul Islam, Mohammed Nazmul Islam Miah, Refat Naznin and Mohammad Kawsur Sharif, "Unleashing Deep Learning: Transforming E-commerce Profit Prediction with CNNs", Journal of Business and Management Studies, ISSN: 2709-0876, DOI: 10.32996/jbms.2024.6.2.12.

[5] Jianhao Zhang, "Enhancing Predictive Models in E-Commerce: A Comparative Study Using XGBoost Across Diverse Scenarios", ITM Web Conf. Volume 70, 2025 , 2nd International Conference on Data Science, Advanced Algorithm and Intelligent Computing (DAI 2024),https://doi.org/10.1051/itmconf/20257002014

[6] Ashish Suresh Awate, Dr. Sanjay Kumar Sharma, "Deep Learning-Driven Dynamic Segmentation and Sentiment Prediction to Enhance Customer Retention in Online Platforms", Journal of Information Systems Engineering and Management 2025, 10(2) e-ISSN: 2468-4376.

[7] Ilham Huda, Agus Achmad Suhendra, Moch Arif Bijaksana, "Design of Prediction Model using Data Mining for Segmentation and Classification Customer Churn in E-Commerce Mall in Mall", December 2023,JOIV International Journal on Informatics Visualization 7(4):2280, DOI:10.30630/joiv.7.4.02414.

[8] Shanthini B, Dr.Subhashini S, Kiruthika S, Yamuna S, Jothi Francina V, Ramesh V, "Predicting Customer Churn Using Artificial Neural Networks: A Data-Driven Approach to Enhance Customer Retention in E-Commerce", Journal of Information Systems Engineering and Management, 2025, 10(41s), e-ISSN: 2468-4376.

[9] Xiancheng Xiahou and Yoshio Harada,"B2C E-Commerce Customer Churn Prediction Based on K-Means and SVM", J. Theor. Appl. Electron. Commer. Res. 2022, 17, 458–475. https://doi.org/10.3390/jtaer17020024.

[10] Pondel Maciej., Wuczyński, M., Gryncewicz, W., Łysik, Łukasz ., Hernes, M. ., Rot, A., & Kozina, A. (2021), "Deep Learning for Customer Churn Prediction in E-Commerce Decision Support", Business Information Systems, 1, 3–12. https://doi.org/10.52825/bis.v1i.42.

[11] Md Sakibul Hasan, Md Abubokor Siam , Md Abdul Ahad , Mohammad Nazmul Hossain, Mehedi Hasan Ridoy, MD. Nazmul Shakir Rabbi, Arat Hossain, and Tanaya Jakir, "Predictive Analytics for Customer Retention: Machine Learning Models to Analyze and Mitigate Churn in E-Commerce Platforms", Journal of Business and Management Studies, ISSN: 2709-0876, DOI: 10.32996/jbms.2024.6.4.22.

[12] Ritu Punhani, V.P.S Arora, A. Sai Sabitha, Vinod Kumar Shukla,"Segmenting e-Commerce Customer through Data Mining Techniques", CONSILIO 2020, Journal of Physics: Conference Series 1714 (2021) 012026, IOP Publishing, doi:10.1088/1742-6596/1714/1/012026.

[13] Qing Zhang , Abdul Rashid Abdullah, Choo Wei Chong , and Mass Hareeza Ali, "E-Commerce Information System Management Based on Data Mining and Neural Network Algorithms", Hindawi, Computational Intelligence and Neuroscience, Volume 2022, Article ID 1499801, 11 pages, https://doi.org/10.1155/2022/1499801.

[14] Yasser D. Al-Otaibi, "Enhancing e-Commerce Strategies: A Deep Learning Framework for Customer Behavior Prediction", Engineering,

Technology & Applied Science Research, ISSN-2241-4487,Vol. 14, No. 4, 2024,pp- 15656-15664.

[15] Priya Chugh and Vishu Jain, "Artificial Intelligence (AI) Empowerment in E-Commerce: A Bibliometric Voyage", NMIMS Management Review, 32(3) 159–173, 2024,journals.sagepub.com/home/nrm, DOI: 10.1177/09711023241303621

[16] Pooja Pande, Adokshaja Krishnarao Kulkarni, Balamurugan P, Balakrishnan S, Vanaja Ramalingam and Revathi R, "Big Data Analytics in E-commerce Driving Business Decisions Through Customer Behavior Insights", ITM Web of Conferences 76, 05001 (2025),ICSICE-2025, https://doi.org/10.1051/itmconf/20257605001.

[17] Cheng-Yuan Chen, Wei-Yu Lin,b, Chunwei Lu, "Big Data Mining Algorithm for Optimal Generation of Intelligent Financial Business Portfolio Investment Decision" Academic Journal of Business & Management, ISSN 2616-5902 Vol. 7, Issue 5: 124-131, DOI: 10.25236/AJBM.2025.070516.

[18] Yasser Filahi, Omer Melih Gul, Ali Elghirani, Erkut Arican, Ismail Burak Parlak,Seifedine Kadry, Kostas Karpouzis, "Enhanced E-commerce decision-making through sentiment analysis using machine learningbased approaches and IoT", (2025) PLoS One 20(6): e0326744. https://doi.org/10.1371/journal.pone.0326744.

[19] Chunli Yin and Jinglong Han, "Dynamic Pricing Model of E-Commerce Platforms Based on Deep Reinforcement Learning", Computer Modeling in Engineering & Sciences,CMES, 2021, vol.127, no.1,Tech Science Press, DOI: 10.32604/cmes.2021.014347.

[20] Salvatore Carta, Andrea Medda, Alessio Pili , Diego Reforgiato Recupero and Roberto Saia, "Forecasting E-Commerce Products Prices by Combining an Autoregressive Integrated Moving Average (ARIMA) Model and Google Trends Data", *Future Internet*, *11*(1), 5. https://doi.org/10.3390/fi11010005.

[21] Minseo Park and Jangmin Oh, "Enhancing E-Commerce Recommendation Systems with Multiple Item Purchase Data: A Bidirectional Encoder Representations from Transformers-Based Approach", Applied Sciences 2024, 14, 7255. https://doi.org/10.3390/app14167255.

[22] Fan Liu,Delong Chen,Xiaoyu Du,Ruizhuo Gao,Feng Xu,"MEP-3M: A large-scale multi-modal E-commerce product dataset", Pattern Recognition, Vol 140, 2023, 109519, ISSN 0031-3203, https://doi.org/10.1016/j.patcog.2023.109519.

[23] Mehmet Serhan Çiftlikçi, Yusuf Çakmak , Tolga Ahmet Kalaycı , Fatih Abut , Mehmet Fatih Akay and Mehmet Kızıldag, "A New Large Language Model for Attribute Extraction in E-Commerce Product Categorization", Electronics 2025, 14, 1930. https://doi.org/10.3390/electronics14101930.

[24] Ling Yang, Fuli Qi, "Enhanced E-Commerce Data Processing Using Dimensional Control and Optimized KNN Algorithm", Informatica 48 (2024) 17–32, https://doi.org/10.31449/inf.v48i17.6288

[25] Biresh Kumar,Sharmistha Roy, Anurag Sinha, Celestine Iwendi, and L'ubomíra Strážovská, "E-Commerce Website Usability Analysis Using the Association Rule Mining and Machine Learning Algorithm", Mathematics 2023, 11, 25. https://doi.org/10.3390/math11010025.

[26] Mehdi Imani, Majid Joudaki, Ali Beikmohammadi, and Hamid Reza Arabnia," Customer Churn Prediction: A Systematic Review of Recent Advances, Trends, and Challenges in Machine Learning and Deep Learning", Mach. Learn. Knowl. Extr. 2025, 7, 105. https://doi.org/10.3390/make7030105.

Authors' Profile

C.SUBASHREE received her MCA (Computer Applications) Degree in the year 2025 in KGISL-IIM (Institute of Information Management) College for affiliated to Bharathiar University, Tamilnadu, India. She is pursuing her Ph.D Degree (Full-Time) in Dr. N.G.P Arts and Science College, Coimbatore, Tamilnadu ,India. Her current research of interests includes Data Mining, Big data, Machine learning and Deep Learning.

Dr.J.Savitha received her Ph.D Degree from Karpagam University, Coimbatore in the year 2017. She received her M.Phil

Degree from Annamalai University, in the year 2009.She received her M.Sc., Degree from Annamalai University, in the year 2006. She is working as Professor, in Department of Computer Science, Dr.N.G.P. Arts & Science College, Coimbatore, Tamilnadu, India. She has above 19 years of experience in academic field. She has published 8 books, more than 34 papers in International Journals, National & International Conferences so far. Her areas of interest include Image Processing, Cyber Security, Artificial Intelligence, Machine Learning, Networks and Web Development.