Vol 4, Issue 7, July 2024 www.ijesti.com E-ISSN: 2582-9734
International Journal of Engineering, Science, Technology and Innovation (IJESTI)

Advances in Machine Learning for Consumer Behaviour Analysis: A Systematic Review (2013–2025)

Rupam Kumari [1]

[1] Research Scholar, Department of Computer Science, Capital University, Koderma, Jharkhand.

Prof. (Dr.) Anand Kumar [2]

[2] Supervisor, Department of Computer Science, Capital University, Koderma, Jharkhand.

Dr. Supriya Shree [3]

[3] Co-Supervisor, Assistant Professor, Department of Computer Science,
St. Xavier's College of Management & Technology, Patna.

ABSTRACT

Consumer behaviour analysis has witnessed a major transformation with the advancement of machine learning and data mining techniques, enabling organizations to extract actionable insights from vast and diverse datasets. This study synthesizes recent literature investigating predictive modelling, behavioural segmentation, market basket analysis, sentiment extraction, and purchase forecasting across multiple digital contexts. Reviews of research conducted between 2013 and 2025 including studies by Lin (2025), Stylianou and Pantelidou (2025), Segun-Falade et al. (2024), Gupta et al. (2024), and others highlight that modern algorithms such as CatBoost, XGBoost, Random Forests, SVMs, neural networks, RNNs, and clustering methods significantly enhance behavioural prediction accuracy. These approaches outperform traditional linear models by identifying complex, nonlinear, and temporal patterns in transactional data, online reviews, clickstreams, and social media interactions. Findings across studies consistently demonstrate that machine learning improves precision marketing, customer segmentation, churn prediction, recommendation systems, and credit risk forecasting, while also informing macroeconomic indicators and retail strategies. The review further emphasizes the growing use of unstructured data and advanced methods such as Transformers, deep learning, reinforcement learning, and probabilistic modelling. Overall, the literature indicates that machine learning–driven consumer behaviour analysis has become essential for data-driven decision-making, personalized marketing, and enhanced customer engagement in competitive digital environments.

Keywords: *Consumer Behaviour Analysis, Machine Learning Techniques, Data Mining Applications, Predictive Modelling.*

1. Introduction

Consumer behaviour refers to the way individuals, groups, and organizations select, purchase, use, and dispose of goods, services, ideas, or experiences in order to satisfy their needs, wants, and desires. It is influenced by a combination of cultural, social, personal, and psychological factors, and involves complex decision-making processes that explain not only what consumers buy, but also why, when, where, and how they buy. In the digital era, consumers leave extensive data traces through online shopping, social media activity, mobile applications, loyalty programmes, feedback forms, and browsing histories. This explosion of data has made traditional analytical tools inadequate for capturing the intricate, dynamic, and often non-linear patterns of modern consumer behaviour. In this context, advanced machine learning methods and data mining techniques have emerged as powerful approaches for analysing consumer behaviour patterns. Machine learning models such as decision trees, Random Forests, Support Vector Machines, neural networks, deep learning [1], reinforcement learning, and ensemble methods can automatically learn from historical data to predict future actions like purchase intent, churn, or product preferences. Data mining techniques, including clustering, association rule mining, sequential pattern mining, text mining, and sentiment analysis, help in segmenting customers, discovering co-purchase relationships, identifying behaviour sequences, and extracting opinions from unstructured text [2]. Through integrating these techniques, businesses can achieve more accurate customer segmentation, personalized recommendations, dynamic pricing, targeted promotions, and early churn detection. This leads to improved marketing effectiveness, better customer experience, and enhanced loyalty [3-9]. At the same time, the growing reliance on algorithmic decision-making raises critical concerns related to data privacy, transparency, fairness, and interpretability. Therefore, the study of advanced machine learning and data mining in consumer behaviour not only strengthens data-driven marketing and strategic decision-making, but also calls for responsible, ethical, and consumer-centric use of analytics in an increasingly competitive and digital marketplace.

2. Related Reviews

Lin (2025) had examined how the growing intensity of market competition and the increasing complexity of consumer behavior had compelled enterprises to find more precise ways of identifying potential customers and improving conversion rates. The study had focused on applying machine learning techniques to predict consumer behavior and enhance precision marketing. Four models Support Vector Machine (SVM), Extreme Gradient Boosting (XGBoost), Categorical Boosting (CatBoost), and Backpropagation Artificial Neural Network (BPANN) were employed to forecast consumers' purchase intentions, and their performance across various scenarios was experimentally verified. The findings had revealed that CatBoost and XGBoost achieved the most effective predictive performance when handling complex and large-scale datasets, with F1-scores of 0.93 and 0.92 respectively, while CatBoost reached the highest ROC-AUC value of 0.985. Although SVM had shown strength in accuracy, it underperformed slightly with larger datasets. Feature importance analysis had further indicated that variables such as page views and residence time significantly influenced purchasing behavior. Based on these results, Lin had proposed practical applications for optimization marketing strategies, including recommendation systems, dynamic pricing, and personalized advertising. The study had also suggested that future research could enhance model performance by incorporating unstructured data such as consumer reviews, images, videos, and social media content, and by exploring deep learning models like Transformers and Self-Attention mechanisms to better capture complex temporal patterns.

Stylianou and Pantelidou (2025) had examined how the rapid progress of Big Data technologies had transformed various sectors, with particular emphasis on the retail industry. Their research had investigated the interrelation between Big Data analytics and consumer behavior by analyzing supermarket transaction datasets to derive macroeconomic insights. Drawing on more than two million transaction records from a multinational supermarket chain, the study had implemented several advanced analytical approaches such as the Apriori algorithm for association rule mining, K-Means clustering for customer segmentation, collaborative filtering for recommendation systems, and the AutoRegressive Integrated Moving Average (ARIMA) model for forecasting time series trends. The findings had indicated distinctive purchasing associations, clear customer segmentations based on shopping tendencies, and predictive patterns in consumer demand, all contributing to improved marketing strategies, inventory control, and policymaking. Moreover, the study had emphasized that transactional data could serve as a reflection of larger economic transitions particularly variations in consumption during periods of financial instability thereby connecting micro-level consumer behaviors to macroeconomic indicators. The originality of this work had been attributed to its integration of multiple machine learning methods into a unified analytical framework that extended the role of Big Data beyond commercial applications toward informing public economic policy.

Segun-Falade et al. (2024) had discussed that machine learning (ML) algorithms had revolutionized the domain of predictive analytics, particularly in understanding and anticipating customer behavior. Their review had examined how these advanced algorithms were employed to strengthen predictive analytics in customer behavior studies, enabling more strategic and data-driven decision-making within businesses. They had emphasized that predictive analytics, which traditionally relied on historical data to forecast future outcomes, became far more powerful and accurate when integrated with ML. The researchers had highlighted that ML algorithms such as decision trees, neural networks, and support vector machines could process vast datasets, uncovering complex and nonlinear patterns beyond human analytical capacity. The review had noted that data for such applications were collected from diverse sources transaction records, social media interactions, and customer feedback and subsequently preprocessed before being input into ML models. Through clustering, classification, and regression techniques, ML had been shown to effectively segment customers, forecast purchasing tendencies, and detect potential churners. The study had further observed that one of the major advantages of ML in predictive analytics was its ability to deliver highly personalized customer experiences. Through predicting individual preferences and needs, businesses had been able to refine marketing strategies, product recommendations, and customer interactions, thereby enhancing satisfaction and loyalty. The authors had provided examples such as e-commerce platforms using ML-based predictive analytics to recommend products aligned with customers' browsing and purchase histories. Additionally, they had pointed out that ML algorithms continually improved with new data, leading to real-time updates and increasingly accurate predictions in dynamic market environments. In conclusion, Segun-Falade et al. had argued that the integration of ML algorithms into predictive analytics had significantly enhanced businesses' capacity to interpret and forecast customer behavior, optimize operations, and strengthen long-term customer engagement, with ongoing technological evolution promising even deeper insights in the future.

Gupta et al. (2024) had examined the evolving landscape of e-commerce, noting a transformative shift in consumer behavior attributed to the proliferation of digital technologies and online platforms. The researchers had emphasized that understanding and predicting such dynamic behavior was essential for businesses aiming to sustain competitiveness in online markets. Their review had explored the application of machine learning (ML) techniques in analyzing and forecasting e-commerce customer behavior, particularly through the lens of customer reviews. They had discussed how the internet had enabled consumers to share opinions online, significantly shaping purchasing decisions. The study had reported that ML models were increasingly utilized to extract meaningful insights from these reviews, thereby helping businesses interpret customer preferences more effectively. The authors had synthesized prior literature covering aspects such as motivations for online shopping, the influence of trust and security, user experience, social impact, personalization, and post-purchase behavior. The review had highlighted the multidimensional factors shaping e-commerce customer behavior and underscored ML's pivotal role in decoding complex consumer sentiments. In conclusion, the authors had suggested that further research on ML methodologies, particularly within the realm of big data, was necessary to improve predictive accuracy and deepen the understanding of online consumer behavior.

Alizadeh et al. (2023) had emphasized that artificial intelligence (AI) played a vital role in marketing and forecasting customer behaviour. Their study had aimed to assess the prediction of customer behaviour through AI techniques. It was concluded from various studies that AI technologies had significantly contributed to analysing large datasets concerning consumer performance, which included examining client behaviour, conducting surveys, and reviewing purchase histories. The researchers had pointed out that machine learning methods could effectively forecast customer behaviour in marketing contexts, and such predictive insights were beneficial in formulating marketing decisions and advertising strategies. Moreover, previous research had suggested that AI could assist organizations in optimizing their marketing operations by enabling the creation of more targeted advertisements and ensuring the efficient utilization of marketing budgets. Within this framework, AI was found to enhance customer experience by providing personalized product recommendations aligned with individual preferences, thus improving shopping satisfaction. Additionally, AI had been used to optimize and manage online advertising campaigns, assisting retail managers in delivering more effective and targeted marketing to suitable audiences. Overall, Alizadeh and colleagues had summarized that AI supported business managers in data analysis, consumer behaviour forecasting, and the design of personalized experiences, ultimately leading to optimized marketing strategies and increased customer engagement.

Jobanputra et al. (2023, September) had explored how Artificial Intelligence (AI) had revolutionized business operations, with marketing emerging as a primary area of transformation. The study had emphasized that analyzing and predicting consumer behavior was a major driver in marketing, enabling marketers to understand buying patterns and anticipate future purchasing decisions. It had defined consumer behavior as encompassing all decisions related to searching, purchasing, evaluating, selecting, using, and disposing of products or services to meet expectations. The researchers had noted that marketers analyzed consumer data to identify relationships between browsing activities and purchasing actions through data on searches, views, and browsing patterns. The article had examined various modern analytical techniques, tools, and models used to predict consumer behavior. Through employing AI, vendors had been able to interpret consumer reactions toward specific products or services and analyze related online posts. The study had further highlighted that AI applications in marketing led to improved search accuracy, enhanced user experience tracking, personalized content curation, and increased sales—facilitating strategic decision-making. It had also discussed how data mining, clustering methods, and the AISAS model could be effectively applied to analyze consumer behavior and formulate marketing strategies.

Akter et al. (2023) had conducted a study highlighting that the exponential growth of e-commerce in the United States had significantly reshaped the retail landscape, creating both opportunities and challenges for online retailers. The research had applied machine learning techniques to formulate a strategic online sales approach grounded in deep consumer behavior analysis. Using comprehensive U.S.-based datasets, the researchers had examined unique traits and preferences of American consumers. The dataset had included detailed information on customer purchase histories, such as items bought, transaction values, and purchase frequencies, along with browsing history data capturing user interaction patterns like visited pages, time spent on each page, and product views. To analyze these patterns, the study had implemented established machine learning models, including Random Forests, Logistic Regression, and Gradient Boosting Classifiers. Among these, the Random Forest model had demonstrated superior accuracy, effectively identifying complex behavioral patterns within the data. The findings had emphasized that integrating machine learning algorithms could transform the online retail sector in the U.S. by enabling large-scale, data-driven, and personalized decision-making processes, ultimately allowing retailers to tailor shopping experiences to individual consumer preferences and behaviors.

Mathur et al. (2022, November) had reviewed the emerging trends in modeling consumer behaviour through machine learning and data mining techniques applied to real customer data. The researchers had noted that most behavioural models were designed to address specific marketing questions over defined time periods, making consumer forecasting a complex and uncertain task. They had emphasized that selecting an appropriate method and strategy was essential for building accurate client behaviour models. It had been observed that marketers often found it difficult to manipulate predictive models for personalized marketing actions, although most models remained relatively straightforward in structure. Due to this simplicity, many existing models tended to ignore numerous influential factors, resulting in less reliable predictions. The study had synthesized prior literature on consumer behaviour analysis using diverse machine learning and data mining approaches, concluding that implementation in Python was both practical and effective owing to its ease of use and its capacity to ensure precision, minimize error rates, and enhance overall model accuracy.

Ebrahimi et al. (2022) had conducted a study to examine how social network marketing (SNM) influenced consumers' purchase behavior (CPB). The researchers had combined structural equation modeling (SEM) with unsupervised machine learning techniques as an innovative methodological approach. Their study population had comprised Facebook Marketplace users residing in Hungary, and convenience sampling had been employed to minimize bias. Out of 475 questionnaires distributed, 466 were fully completed, yielding a high response rate of 98.1%. The findings had revealed that all dimensions of social network marketing entertainment, customization, interaction, word-of-mouth (WoM), and trend had exerted significant positive effects on consumer purchase behavior within Facebook Marketplace. Moreover, hierarchical clustering and K-means algorithms had been utilized to classify consumers into nine distinct behavioral groups based on demographic characteristics, suggesting that differentiated marketing strategies could be tailored for each cluster. The study had further highlighted the managerial implications of offering diverse products and services to suit various consumer segments. Additionally, the research had demonstrated methodological rigor by employing *plspm* and *Matrixpls* packages in R to assess model predictive power, while the application of unsupervised algorithms had provided novel insights into consumer behavioral patterns.

Chaudhary et al. (2021) had conducted a study emphasizing the growing popularity of social media in society and its significant role in influencing consumer purchasing behavior. The researchers had collected data from various social media platforms, including Facebook, Twitter, LinkedIn, YouTube, Instagram, and Pinterest, to analyze consumer behavior patterns. Given the diverse, high-speed, and voluminous nature of data emerging from these platforms, they had employed predictive big data analytics to process and interpret the information. The study had utilized big data technologies to manage and analyze the datasets for predicting consumer tendencies on social media. Consumer perception and attitudes toward these platforms had been examined based on multiple parameters and criteria. To ensure the reliability and quality of results, the data had undergone extensive preprocessing to remove outliers, noise, errors, and duplicate records. A mathematical model based on machine learning techniques had been developed for predicting consumer behavior, with 80% of the data used for training and 20% reserved for testing, resulting in a robust predictive framework for social media consumer analysis.

Liashenko et al. (2021) had conducted a study focusing on the analysis of economic agent behavior, which was considered a central theme in microeconomics. The researchers had emphasized that with the exponential growth of data and computing capabilities, there arose a need to apply behavioral economics methods to study human behavior more effectively. In their investigation, they had developed a survey designed to identify behavioral patterns among modern consumers based on their store selection criteria and responses framed around behavioral economics theorems. The collected data had been clustered using machine learning algorithms, and the Random Forest classifier had been trained for predictive modeling. Silhouette analysis results had indicated that K-means clustering produced the most suitable clusters for further modeling, while additional visualizations had been performed using TSNE, hierarchical, and spectral analysis. The study had ultimately provided a tool for classifying customer preferences and analyzing contemporary industry trends. Specifically, it had aimed to enhance the understanding of consumers in food retail chains, thereby improving their "buyer's journey." The authors had suggested that the developed machine-learning-based clustering and classification tool could be integrated into business processes for better consumer analysis. However, they had acknowledged the limitation of the study's sample, which primarily consisted of rational and knowledgeable individuals, not fully representative of the general population. Hence, they had proposed future research directions such as exploring behavioral trends in other industries, utilizing geodata for refined analysis, and assessing the influence of online network advertising on consumer behavior through semantic and image recognition analysis.

Shrirame et al. (2020, July) had conducted a study emphasizing the analytical potential of user-generated content such as reviews, ratings, and comments for enterprise use. The researchers had explained that the examination of such consumer behavior provided deeper insights into customers' needs and enabled the prediction of their future intentions toward services. It had been suggested that by applying cognitive and data-driven approaches, e-commerce organizations could effectively monitor product usage and consumer sentiments, thereby formulating appropriate marketing strategies to enhance personalized shopping experiences and organizational profitability. The paper had aimed to implement advanced tools including data visualization, natural language processing, and machine learning to understand organizational demographics more effectively. Moreover, the authors had developed recommender systems using collaborative filtering, neural networks, and sentiment analysis to improve decision-making and optimize consumer engagement.

Orogun and Onyekwelu (2019) had discussed that in recent years, customer behaviour models were generally developed using data mining techniques applied to customer datasets, each designed to answer a specific question at a particular point in time. They noted that predicting customer behaviour had remained a challenging and uncertain task, requiring appropriate methodologies and analytical approaches. Once a predictive model had been established, it became difficult for marketers to manipulate it effectively to determine optimal marketing actions for individual customers or groups. The researchers observed that despite the complexity inherent in customer behaviour analysis, most models remained relatively simple and often excluded numerous relevant factors, thereby reducing the reliability of their predictions. To address this limitation, the study aimed to develop an association rule mining model for predicting customer behaviour, utilizing data from a typical online retail store to extract significant behavioural trends and actionable insights from the customer data.

Varghese et al. (2018, May) had examined how power consumption datasets could reveal significant insights into consumer behavior. The researchers had noted that behavioral models could be developed to identify consumer groups with similar usage patterns, which were valuable for a range of utility applications. Although the same dataset had been used, different applications extracted different behavioral models depending on the selection of relevant features, the rationale for which had been clearly discussed. The study had considered two specific applications demand response and rational tariff design—and had pointed out that clustering large and volatile datasets posed major challenges. To address dimensionality reduction, a representative curve had been proposed that encapsulated all relevant analytical features, including seasonal variations. The team had implemented machine learning algorithms such as k-means, expectation maximization, and self-organizing maps, and tested the proposed approach on a real-world dataset comprising 789 consumers. A performance index had been employed to assess the relative efficiency of various algorithms, and the resulting outcomes had been systematically analyzed and discussed.

Lang and Rettenmeier (2017, April) had explained that consumer behavior in e-commerce could be characterized by a series of interactions with an online shop, and they had demonstrated that recurrent neural networks (RNNs) were particularly suitable for modeling and predicting such behavior. Their study had emphasized that RNNs offered several advantages over traditional approaches used in real-world production systems. When RNNs were applied directly to sequences of consumer actions, they had produced prediction accuracies equal to or greater than those of vector-based models such as logistic regression, without requiring extensive feature engineering. Moreover, the researchers had shown that RNNs enabled an intuitive linkage between individual actions and their corresponding predictive outcomes, thereby clarifying how consumer activities influenced the evolving probability estimates over time. The empirical validation of these findings had been conducted using data from a large European online fashion platform, where the effectiveness and interpretability of RNN-based models in forecasting consumer behavior had been clearly demonstrated.

Zuo et al. (2016, December) had examined how predicting consumers' purchasing behavior had become a central theme in consumer behavior research over the past decades. The authors had emphasized that most existing business models relied heavily on linear equations to evaluate the influence of contextual factors such as age, gender, income, product price, and promotional activities, primarily because linear models were both easily interpretable for academics and practical for business professionals. However, they had pointed out that, unlike fields such as pattern recognition or text classification, studies on purchase behavior had been overly dependent on linear approaches particularly linear discriminant analysis and logistic regression. With the exponential growth of business data enabled by information and communication technologies such as POS systems and sensors, the researchers had argued that traditional linear models were no longer adequate for extracting meaningful insights. Consequently, they had applied two machine learning methods—Bayes classifier and Support Vector Machine (SVM) to real-world datasets to assess their effectiveness in predicting customer purchase behavior, thereby demonstrating the growing importance of machine learning for data-driven marketing and decision-making.

Kim et al. (2016) had examined the predictive potential of big data analytics in understanding consumer behavior by utilizing detailed customer online activity records. The researchers noted that earlier purchase prediction models based solely on clickstream data suffered from several limitations. Hence, they had proposed a novel approach integrating information theory with machine learning to enhance purchase prediction accuracy. The study had analyzed clickstreams from 5,000 panel members along with their purchase data across electronics, fashion, and cosmetics categories. Using the concept of entropy from information theory to summarize clickstreams and the random forest algorithm to construct predictive models, the research had demonstrated improved prediction accuracy ranging between 0.56 and 0.83. This represented a significant advancement over previous clickstream-based models. Furthermore, the findings had indicated that consumers' information search behaviors varied notably across different product categories, highlighting the diverse nature of online purchasing patterns.

Yuan (2015) had discussed that traditional credit bureau analytics, such as credit scores, were primarily dependent on slowly changing consumer characteristics, making them less adaptable to evolving customer behaviors and market dynamics. The study had aimed to employ machine learning techniques to develop forecasting models for consumer credit risk. By integrating data from credit accounts, credit bureau records, and customer information provided by a major commercial bank (referred to as "the Bank" for confidentiality), the researcher had sought to construct out-of-sample predictive models. These models were reported to address practical challenges commonly faced by chief risk officers and policymakers, including determining appropriate times and amounts for reducing individual credit lines, assessing creditworthiness of existing and potential clients, and predicting overall credit defaults and delinquencies to support enterprise-wide and macroprudential risk management.

Ravnik et al. (2014, August) had explored the emerging concept of audience-adaptive digital signage, wherein public display systems dynamically adjusted their content based on audience demographics and temporal characteristics. The researchers had emphasized that audience measurement data could serve as a valuable foundation for statistical analysis of viewing behaviors, development of interactive display applications, and future research in observer modeling. Using machine learning techniques on real-world digital signage viewership data, they had aimed to predict consumer behavior within a retail setting—particularly focusing on purchasing decisions and the roles individuals played during such processes. Their case study, conducted in a small retail shop, involved the manual verification and analysis of demographic and audience data from 1,294 customers, of whom 246 had engaged in transactions leading to purchases. Upon comparing different machine learning models, they had found that support vector machines achieved an 88.6% classification accuracy in predicting actual purchase behavior, surpassing the baseline classifier by 7.5%. Furthermore, by incorporating heuristic features into the dataset, the SVM classifier had improved prediction accuracy regarding customer roles in purchase decisions by an average of 15% over the baseline, demonstrating the significant potential of machine learning for adaptive marketing and retail analytics.

Kruppa et al. (2013) had discussed that consumer credit scoring was generally treated as a classification problem in which clients were categorized as having either good or bad credit status. They had pointed out that default probabilities offered more detailed insights into consumers' creditworthiness and were commonly estimated using logistic regression. The researchers had proposed a general framework for estimating individual consumer credit risks through machine learning approaches, emphasizing that since probability represented an expected value, all nonparametric regression methods consistent for the mean were also consistent for probability estimation. They had included random forests (RF), k-nearest neighbors (kNN), and bagged k-nearest neighbors (bNN) as part of this class of nonparametric regressors. Their study had applied these machine learning techniques, along with an optimized logistic regression model, to a large dataset containing complete payment histories of short-term installment credits. Furthermore, they had implemented probability estimation using *Random Jungle*, an RF package developed in C++ that provided a generalized framework for efficient tree growth, probability estimation, and classification. The authors had also described an algorithm for tuning terminal node size to improve probability estimation accuracy. Their experimental findings had revealed that regression RF performed better than optimized logistic regression, kNN, and bNN in predicting default probabilities on the test data of short-term installment credits.

3. Findings from Study on Machine Learning, Big Data, and Consumer Behaviour

Author(s) & Year	Purpose / Focus of Study	Methods / Algorithms Used	Key Findings / Results	Implications / Contributions
Lin (2025)	Predict consumer behaviour & enhance precision marketing.	SVM, XGBoost, CatBoost, BPANN	CatBoost F1 = 0.93; XGBoost F1 = 0.92; CatBoost ROC-AUC = 0.985; SVM slightly weaker on large datasets; Page views & residence time most influential.	Supports precision marketing, recommendation systems, dynamic pricing; suggests future integration of unstructured data & deep learning (Transformers).

Stylianou & Pantelidou (2025)	Analyze supermarket Big Data to link consumer behaviour with macroeconomic trends.	Apriori, K-Means, Collaborative Filtering, ARIMA	Identified strong purchase associations; clear customer clusters; time-series demand patterns; links micro-purchases with macroeconomic shifts.	Provides unified Big Data analytical framework; valuable for marketing, inventory management & public economic policy.
Segun-Falade et al. (2024)	Review ML algorithms in predictive customer behaviour analytics.	Decision Trees, Neural Networks, SVM	ML handles large/complex data; improves segmentation, churn prediction & personalization; models improve continuously with data.	Shows ML boosts prediction accuracy, strategic decision-making & customer engagement.
Gupta et al. (2024)	Review the use of ML to analyze e-commerce consumer behaviour through online reviews.	ML, sentiment analytics	ML extracts insights from reviews; identifies motivations, trust factors, UX & personalization drivers.	Highlights multidimensionality of online consumer behaviour; stresses ML's role in sentiment interpretation.
Alizadeh et al. (2023)	Assess AI-based prediction of customer behaviour in marketing.	AI, ML predictive models	AI enhances analysis of large datasets; optimizes marketing budgets; improves personalized recommendations & advertising.	Demonstrates role of AI in consumer forecasting and personalized targeting.
Jobanputra et al. (2023)	Analyze & predict consumer behaviour using AI techniques.	Data mining, clustering, AISAS model	AI interprets online reactions; improves search accuracy, personalized content & sales.	Supports strategic marketing, consumer trend interpretation & targeted promotions.
Akter et al. (2023)	Predict U.S. online consumer behaviour using ML.	Random Forest, Logistic Regression, Gradient Boosting	Random Forest best performer; accurately identifies complex behavioural patterns from browsing & purchase data.	Shows ML enables large-scale personalized online retail strategies.
Mathur et al. (2022)	Review ML and data mining for consumer behaviour prediction.	Data Mining, Python ML	Many models oversimplify behaviours; Python ensures precision & low errors; reliable for behavioural forecasting.	Calls for more comprehensive models incorporating more variables.
Ebrahimi et al. (2022)	Study Social Network Marketing's effect on purchase behaviour.	SEM, Hierarchical Clustering, K-Means	All SNM dimensions significantly affect purchase behaviour; 9 consumer segments identified.	Shows SNM strongly influences buying; segmentation supports customized marketing.
Chaudhary et al. (2021)	Predict consumer behaviour using social media Big Data.	Big Data Analytics, ML model	Accurate predictions after preprocessing; model trained on 80% data & tested on 20%.	Framework for big-data-driven marketing strategy.

Liashenko et al. (2021)	Analyze modern consumer decision patterns using behavioural economics + ML.	K-Means, Random Forest, TSNE, Spectral & Hierarchical	K-Means best clustering; RF effective for classification; model reveals behavioural segments.	Useful for food retail analysis & buyer journey optimization.
Shrirame et al. (2020)	Predict behaviour using user-generated content for marketing.	NLP, ML, Collaborative Filtering, Neural Networks	Review, rating & comment analysis improves recommender systems and insights on consumer needs.	Strengthens personalization & e-commerce profitability.
Orogun & Onyekwelu (2019)	Develop customer behaviour prediction model using association rules.	Association Rule Mining	Extracted major behavioural trends from retail data.	Offers explainable insights for marketing decision-making.
Varghese et al. (2018)	Study consumer behaviour using electricity usage data.	K-Means, EM, Self-Organizing Maps	Identified usage-based behavioural groups; representative curve improved clustering.	Useful for tariff design & demand-response strategies.
Lang & Rettenmeier (2017)	Predict e-commerce behaviour using sequential models.	RNNs	RNNs outperform logistic regression, require less feature engineering; strong interpretability.	Introduces sequence-based modelling for online behaviour.
Zuo et al. (2016)	Compare linear model's vs ML for predicting purchase behaviour.	Bayes Classifier, SVM	ML performs better than linear models; linear methods outdated for large business data.	Advocates ML for marketing analytics.
Kim et al. (2016)	Improve purchase prediction accuracy using clickstream + information theory.	Entropy-based features + Random Forest	Accuracy increased to 0.56–0.83; category-specific behavioural differences observed.	Demonstrates improved modelling using entropy-enhanced features.
Yuan (2015)	Apply ML to predict consumer credit risk.	ML credit risk models	Improved predictions for credit defaults & risk management.	Enhances credit scoring & macroprudential planning.
Ravnik et al. (2014)	Predict purchase behaviour using digital signage audience data.	SVM, ML classifiers	SVM accuracy = 88.6%; heuristic features improved predictions by 15%.	Shows ML potential for adaptive retail media & behaviour prediction.
Kruppa et al. (2013)	Compare ML with logistic regression for credit scoring.	RF, kNN, bNN, Logistic Regression	Regression RF outperformed all other models.	Provides ML-based probability estimation framework for credit defaults.

4. Conclusion

The literature collectively demonstrates that the integration of advanced machine learning and data mining techniques has reshaped the study and prediction of consumer behaviour, making traditional analytical approaches insufficient in modern data-rich environments. Across multiple studies, algorithms such as CatBoost, XGBoost, Random Forests, SVMs, neural networks, RNNs, and clustering frameworks [10-17] consistently outperform linear and rule-based models by capturing the multidimensional, dynamic, and nonlinear relationships underlying consumer actions. These technological advances have enabled more accurate forecasting of purchase intentions, segmentation of customer groups, detection of behavioural sequences, assessment of credit risk, and extraction of sentiment from user-generated content [18]. The reviewed studies highlight that machine learning supports a wide range of applications, including precision marketing, dynamic pricing, personalized recommendations, churn prevention, fraud detection, demand forecasting, and macroeconomic analysis [19]. The increasing use of unstructured data reviews, images, clickstreams, and social media content further enhances model richness and predictive capability. At the same time, emerging models such as deep learning, reinforcement learning, and Transformer-based architectures [20] offer promising opportunities for capturing temporal and contextual nuances in consumer behaviour. However, the literature also recognizes certain challenges, including data quality issues, model interpretability, privacy concerns, and biases embedded in algorithmic systems. Addressing these challenges will be crucial for responsible and ethical implementation. Overall, the evidence strongly suggests that machine learning-driven consumer behaviour analytics provides organizations with powerful tools to improve decision-making, optimize marketing strategies, and sustain competitive advantage in increasingly complex digital markets.

References

1. Lin, J. (2025). Application of machine learning in predicting consumer behavior and precision marketing. *PLoS One*, *20*(5), e0321854.
2. Stylianou, T., & Pantelidou, A. (2025). A machine learning approach to consumer behavior in supermarket analytics. *Decision Analytics Journal*, 100600.
3. Segun-Falade, O. D., Osundare, O. S., Kedi, W. E., Okeleke, P. A., Ijomah, T. I., & Abdul-Azeez, O. Y. (2024). Utilizing machine learning algorithms to enhance predictive analytics in customer behavior studies. *International Journal of Scholarly Research in Engineering and Technology*, *4*(1), 001-018.
4. Gupta, K., Chaturvedi, A., Singh, A., Rai, A., Asrani, D., & Asrani, K. (2024). E-Commerce customer behavior using machine learning. *International Journal of Innovative Research in Computer Science & Technology (IJIRCST)*, *12*(1), 324-330.
5. Alizadeh, H., Kashani, H. N., Filshoor, M. J., & Khameneh, A. P. (2023). Evaluation of consumer behavior prediction based on artificial intelligence in marketing. In *Conference Paper*.
6. Jobanputra, R., Kulkarni, P., Marhoon, H. A., & Palloan, H. (2023, September). Role of artificial intelligence in analyzing and predicting consumer behavior. In *AIP Conference Proceedings* (Vol. 2736, No. 1, p. 060016). AIP Publishing LLC.
7. Akter, R., Nasiruddin, M., Anonna, F. R., Mohaimin, M. R., Nayeem, M. B., Ahmed, A., & Alam, S. (2023). Optimizing Online Sales Strategies in the USA Using Machine Learning: Insights from Consumer Behavior. *Journal of Business and Management Studies*, *5*(4), 167-183.
8. Mathur, N., Kumar, S., Joshi, T., & Dhuliya, P. (2022, November). Analyzing consumer behavior predictions: a review of machine learning techniques. In *2022 International Conference on Advances in Computing, Communication and Materials (ICACCM)* (pp. 1-5). IEEE.
9. Ebrahimi, P., Basirat, M., Yousefi, A., Nekmahmud, M., Gholampour, A., & Fekete-Farkas, M. (2022). Social networks marketing and consumer purchase behavior: the combination of SEM and unsupervised machine learning approaches. *Big Data and Cognitive Computing*, *6*(2), 35.
10. Chaudhary, K., Alam, M., Al-Rakhami, M. S., & Gumaei, A. (2021). Machine learning-based mathematical modelling for prediction of social media consumer behavior using big data analytics. *Journal of Big data*, *8*(1), 73.

11. Liashenko, O., Kravets, T., & Prokopenko, M. (2021). Consumer behavior clustering of food retail chains by machine learning algorithms. *Access to science, business, innovation in digital economy*, 234-251.
12. Shrirame, V., Sabade, J., Soneta, H., & Vijayalakshmi, M. (2020, July). Consumer behavior analytics using machine learning algorithms. In *2020 IEEE international conference on electronics, computing and communication technologies (CONECCT)* (pp. 1-6). IEEE.
13. Orogun, A., & Onyekwelu, B. (2019). Predicting consumer behaviour in digital market: a machine learning approach.
14. Varghese, A. C., Padmini, V., Kumar, G., & Khaparde, S. A. (2018, May). Smart grid consumer behavioral model using machine learning. In *2018 IEEE innovative smart grid technologies-asia (ISGT asia)* (pp. 734-739). IEEE.
15. Lang, T., & Rettenmeier, M. (2017, April). Understanding consumer behavior with recurrent neural networks. In *Workshop on Machine Learning Methods for Recommender Systems*.
16. Zuo, Y., Yada, K., & Ali, A. S. (2016, December). Prediction of consumer purchasing in a grocery store using machine learning techniques. In *2016 3rd Asia-Pacific World Congress on Computer Science and Engineering (APWC on CSE)* (pp. 18-25). IEEE.
17. Kim, M., Im, I., & Han, S. (2016). Purchase prediction by analyzing users' online behaviors using machine learning and information theory approaches. *Asia pacific journal of information systems*, *26*(1), 66-79.
18. Yuan, D. (2015). *Applications of machine learning: consumer credit risk analysis* (Doctoral dissertation, Massachusetts Institute of Technology).
19. Ravnik, R., Solina, F., & Zabkar, V. (2014, August). Modelling in-store consumer behaviour using machine learning and digital signage audience measurement data. In *International Workshop on Video Analytics for Audience Measurement in Retail and Digital Signage* (pp. 123-133). Cham: Springer International Publishing.
20. Kruppa, J., Schwarz, A., Arminger, G., & Ziegler, A. (2013). Consumer credit risk: Individual probability estimates using machine learning. *Expert systems with applications*, *40*(13), 5125-5131.