Journal of Marketing Analytics
https://doi.org/10.1057/s41270-026-00470-6

SOFTWARE REVIEW

Big data in consumer behavior research: a systematic review of data sources, analytical methods, and research questions

Qiankun Liu[1] · Ruigang Wang[2] · Muhabaiti Pareti[1,3] · Alessandra Castellini[1] · Davide Viaggi[1] · Maurizio Canavari[1]

Received: 23 October 2025 / Accepted: 23 February 2026

Abstract
While big data offers unprecedented opportunities for decoding consumer behavior, researchers face significant challenges in navigating the fragmentation of heterogeneous data sources and the complexity of integrating advanced analytical techniques with behavioral theory. To address these gaps, this study conducts a systematic literature review following the PRISMA 2020 protocol, analyzing 127 peer-reviewed articles published between 2012 and 2023. Moving beyond prior reviews that focus on isolated domains or specific data types, this research contributes an integrated framework that maps specific data types (structured vs. unstructured) to analytical methods (e.g., machine learning, text mining) and their application in solving core behavioral research questions. By systematically aligning these elements, the framework demonstrates how big data can be operationalized to address theoretical gaps in consumer research. The analysis reveals that big data is extensively employed to analyze consumption patterns, sentiment, and decision-making pathways, yet challenges regarding data quality, algorithm interpretability, and privacy protection remain prevalent. Practically, this review provides a methodological roadmap for scholars and practitioners to leverage big data for precise behavioral prediction and personalized marketing, while underscoring the urgent need for interdisciplinary collaboration to address ethical compliance in data-driven strategies.

Keywords Big data · Consumer behavior · Systematic review · Machine learning · Text mining

Introduction

In the digital era, the rapid growth of social networks, mobile devices, and Internet of Things (IoT) technologies has generated an unprecedented volume of data, fundamentally transforming the landscape of marketing and consumer research. Big data, characterized by high volume, velocity, and variety, enables companies to capture consumer interactions in real time, transcending the limitations of traditional self-reported surveys (Erevelles et al. 2016; L. Li 2023). Consequently, integrating big data analytics into consumer behavior research has gained prominence as it offers a granular view of decision-making processes that were previously unobservable (Hofacker et al. 2016). Recent systematic reviews confirm that these analytics are now central to understanding consumer behavior in digital marketing, providing high precision in tracking decision pathways (Theodorakopoulos & Theodoropoulou 2024).

The application of big data represents a paradigm shift in how consumer behavior is analyzed. Unlike traditional methodologies that rely heavily on retrospective questionnaires and interviews, which are often subject to recall bias, big data analytics utilizes unstructured data (e.g., social media text, images, clickstreams) to provide dynamic approximations of behavioral patterns. For instance, textual content from Google Trends and user-generated photos allow researchers to infer latent preferences and willingness to pay with greater predictive accuracy than conventional field data (Kanavos et al. 2018). Crucially, while big

Prof. Maurizio Canavari has passed away.

✉ Qiankun Liu
qiankun.liu@studio.unibo.it

[1] Department of Agricultural and Food Sciences, Alma Mater Studiorum-University of Bologna, Viale Giuseppe Fanin 50, 40127 Bologna, Italy

[2] Institute of Food and Nutrition Development, Ministry of Agriculture and Rural Affairs, Beijing 100081, China

[3] College of Economics and Management, Xinjiang Agricultural University, Nongda East Road 311, Urumqi 830052, China

Published online: 26 March 2026

data offers deep insights, it is important to recognize that these datasets represent digital behavioral traces rather than an unmediated reflection of human reality. This transition allows for the connection of disparate data points, from sentiment to transaction logs, to create real-time proxies of consumer psychology and behavior.

Despite these advancements, a critical theoretical and methodological gap remains. While data availability has exploded, our understanding of how to theoretically ground these diverse data sources to answer complex behavioral questions remains superficial. Previous reviews have largely fragmented the field by treating big data as a purely technical tool rather than a theoretical enabler. Some have focused narrowly on technical algorithms and data mining architectures without connecting them to behavioral theory (Babiceanu & Seker 2016; Ierkens et al. 2019). Others have been restricted to specific domains, such as e-commerce recommendation systems or hospitality guest experience, lacking a cross-industry perspective that generalizes consumer insights (Buettner 2017; Xiang et al. 2015). Critically, existing literature fails to explain why specific data types explain specific psychological mechanisms. What is missing is a holistic perspective that systematically links data sources (where the data comes from) with analytical methods (how it is processed) and consumer behavior theories (what questions it answers). Current literature lacks an up-to-date synthesis that addresses how unstructured big data is theoretically interpreted to decode complex decision-making pathways.

To address this fragmentation and provide a distinct conceptual contribution, this study conducts a systematic review of 127 peer-reviewed articles to develop an integrated framework for big data applications in consumer behavior research. This paper goes beyond a descriptive overview; it identifies the specific "theoretical mechanisms" that big data unlocks. Unlike prior works, this study explicitly maps the trajectory from data acquisition to behavioral insight, offering a roadmap for researchers to select the right analytical tools for specific theoretical constructs. Accordingly, the paper addresses the following research questions to guide this conceptual integration.

RQ1: How are heterogeneous big data sources (structured vs. unstructured) aligned with specific analytical techniques (e.g., AI, text mining) to enhance behavioral modeling?

RQ2: What are the evolving research themes in data-driven consumer behavior, and how do they bridge the gap between computational data science and traditional behavioral constructs?

The paper is organized as follows: Sect. "Methodology" outlines the review methodology; Sect. "Bibliometric overview of the literature" presents the bibliometric analysis; Sect. "Data and technical foundations" discusses data sources and analytical models; Sect. "Big data applications for understanding consumer behavior" analyzes the data sources, methods, and research topics; Sect. "Discussion" addresses implications and future directions; and Sect. "Conclusion" draws conclusions. This structure aims to provide a conceptual and methodological roadmap for scholars and practitioners interested in advancing data-driven consumer research.

Table 1 summary of data source. *Source*: Authors' elaboration based on the selection process on Scopus and Web of Science

Category	Specific Standard Requirements
Research database	Web of Science core collection, Scopus
Citation indexes	WOS (SSCI, SCIE), Scopus
Searching period	January 2012 to December 2023
Language	"English"
Searching keywords	WOS TS = (("big data" OR "data mining" OR "machine learning") AND ("consumer behavior" OR "buying behavior" OR "customer behavior")) SCOPUS TITLE-ABS-KEY (("big data" OR "data mining" OR "machine learning") AND ("consumer behavior" OR "buying behavior" OR "customer behavior"))
Subject categories	"Business" "Computer Science Information Systems" "Environmental Sciences" "Management" "Agricultural Economics" "Green Sustainable Science Technology"
Document types	"Articles"
Data extraction	Export with full records and cited references in RIS format
Sample size	371

Methodology

Search strategy and data collection

This paper follows the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 protocol (Moher et al. 2009, 2015) to make the flow chart and keep this paper replicability. PRISMA 2020 reflects recent technological developments, including the increasing use of automation and machine learning assisted tools in systematic reviews, by emphasizing transparency in reporting automated processes involved in study selection and data extraction (Sarkis-Onofre et al. 2021). To ensure the comprehensiveness and accuracy of the retrieval data, the summary of the data source is shown in Table 1, while Fig. 1 represents the selection process followed by the PRISMA protocol (Page et al. 2021).

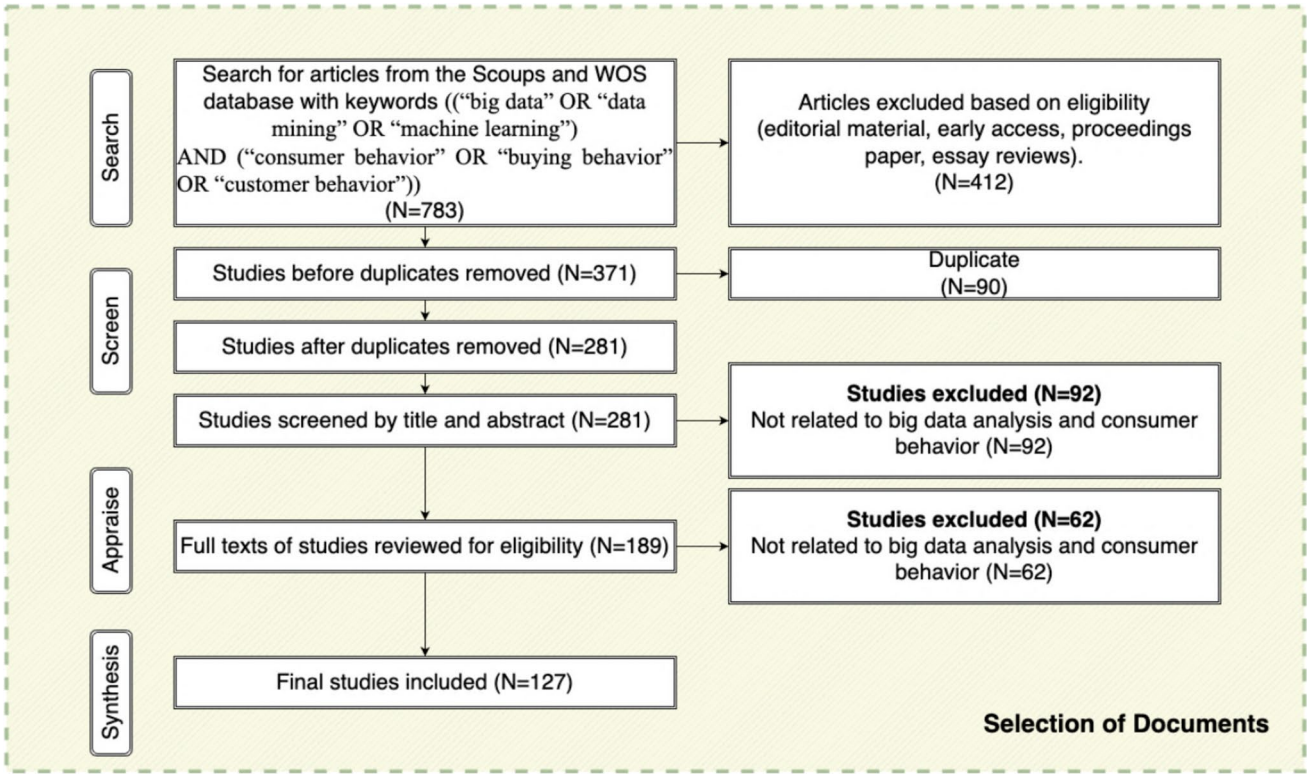

Fig. 1 Process of literature selection. Source: Author's elaboration based on the selection process on Scopus and Web of Science

The literature was collected from January 2012 to December 2023 using the Web of Science and Scopus databases. To ensure high-quality evidence and comparability, the review was restricted to peer-reviewed journal articles published in the English language. This language restriction was applied because English serves as the primary lingua franca for global scientific discourse, ensuring that the selected studies represent accessible and comparable findings within the international research community. Conference proceedings, editorials, and review essays were strictly excluded to prioritize validated findings that have undergone rigorous peer review rather than preliminary or non-empirical discussions.

The Web of Science index is selected as SCI-EXPANDED, SSCI, and the final Web of Science retrieval strategy is [*TS=((*"big data" OR "data mining" OR "machine learning") AND ("consumer behavior" OR "buying behavior" OR "customer behavior"))*]. This same search string was applied to Scopus using the "TITLE-ABS-KEY" field to ensure comprehensive and comparable results across both platforms.

As illustrated in the PRISMA flow (Fig. 1), the selection process consisted of four distinct phases. Initial searching based on keywords and subject category filters yielded 371 articles (N = 371). Duplicate entries across the two databases were identified and removed, resulting in 281 unique articles (N = 281). Titles and abstracts were then screened against predefined inclusion criteria to ensure thematic relevance. This step led to the exclusion of 92 irrelevant articles that focused on technical engineering aspects rather than consumer behavior, leaving 189 articles (N = 189) for full-text evaluation.

To mitigate selection bias and ensure methodological rigor, full-text articles underwent a comprehensive eligibility assessment. This assessment was operationalized through a Quality Assessment checklist consisting of five key criteria. Methodological Rigor: clarity in research design and execution; Data Transparency: explicit description of data sources and collection periods; Theoretical Alignment: direct relevance to consumer behavior constructs or the S–O-R framework; Analytical Clarity: transparent reporting of algorithms or statistical methods; Contribution: providing actionable insights for behavioral research. Each article was scored on these dimensions, and only those demonstrating high compliance across all categories were retained. Based on these rigorous quality and relevance checks, 62 articles were excluded. Specifically, these articles were removed due to insufficient methodological clarity (e.g., unspecified data origins) or a lack of direct relevance to consumer decision-making theories. This resulted in a final sample of 127 articles (N = 127) included in the review.

To minimize subjectivity during the manual screening process, the selection was conducted strictly according to the established protocol. Specifically, quality control was enforced by having two independent researchers cross-verify the selection at every stage. Any ambiguities regarding article inclusion were resolved through an iterative re-evaluation process to ensure consistency.

After data cleansing, quantitative and qualitative analyses were performed on the final dataset of 127 research papers selected. This study employs a two-part methodology for the literature data analysis, starting with a bibliometric analysis to quantitatively assess the landscape of big data research in consumer behavior. In the second part, a thematic analysis is conducted, utilizing theme maps to qualitatively explore key themes and emerging trends within the literature.

Theoretical framework and thematic classification

To provide a solid theoretical underpinning for this review, we adopt the Stimulus-Organism-Response (S–O-R) paradigm as the conceptual backbone for classifying big data applications. While alternative frameworks such as the Technology Acceptance Model (TAM) and the Theory of Planned Behavior (TPB) are widely used to explain consumer actions, they primarily focus on cognitive beliefs about technology utility or rational intentions shaped by social norms (Sujood & Pancy 2024). In contrast, the S–O-R framework is prioritized here because it explicitly accounts for the interplay between external environmental cues and internal emotional states (Song et al. 2024). This alignment closely mirrors the linear flow of big data analytics, from raw digital stimuli to algorithmic processing and behavioral output (Vinkóczi et al. 2026).

This adoption aligns with a growing body of research that employs the S–O-R model as a structural framework for systematic literature synthesis in digital consumer contexts (Erensoy et al. 2024). The S–O-R model posits that environmental cues (Stimulus) influence an individual's internal cognitive and emotional states (Organism), which subsequently drive behavioral outcomes (Response). However, it is important to reflect on the model's assumptions, noting that while it provides a clear classificatory device, its traditionally linear structure may oversimplify the complex, non-linear feedback loops inherent in social influence or perceived behavioral control, which are areas where TPB might offer more depth (Sujood & Pancy 2024). Nevertheless, for the purpose of this review, S–O-R serves as the most effective lens to capture the sensory-driven nature of digital marketing.

In the context of this study, big data technologies serve as the lens to capture and analyze these three stages. Specifically, big data captures macro-level events and digital marketing inputs that function as external stimuli triggering consumer reactions. Simultaneously, advanced algorithms, such as sentiment analysis and emotion mining, offer analytical proxies for the organism's internal cognitive states, thereby shedding light on the "black box" of consumer psychology, including attitudes and preferences. This phrasing reflects an analytical contribution toward understanding consumer states rather than an absolute epistemic resolution of human psychology. Finally, transactional and behavioral data capture the response, revealing the final decision-making pathways and actual consumption patterns. By integrating big data analytics into each stage of the S–O-R trajectory, this study provides a conceptual contribution that is distinct from previous meta-analytic and thematic reviews (e.g., Vieira 2013), which primarily focused on traditional offline or structured retail environments.

Based on this theoretical logic, we organized the reviewed literature to systematically analyze the interplay between data sources, analytical methods, and behavioral insights. Specifically, the application of big data in consumer behavior analysis is categorized into several key domains aligned with the S–O-R framework. The primary category encompasses consumer behavior and distinct big data types, with subcategories including consumption patterns (representing the Response), preferences and attitudes (representing the Organism), and decision-making processes. Additionally, various models and algorithms, particularly those involving artificial intelligence (AI), are highlighted as the critical methodological tools enabling this analysis, with some AI algorithms demonstrating exceptional performance in recent years. Other essential topics, such as influencing factors and the broader impacts on consumer behavior, are also explored in the following sections.

Bibliometric overview of the literature

Distribution of articles by year

The 127 papers used in this study were published in 110 journals by 83 authors in 68 countries. Figure 2 presents the number of articles published every year on big data in the consumer behavior field. The first publication was in 2014. Starting from 2018, there was an increase in the number of publications. The probable reason that big data in consumer behavior has increased rapidly after 2018 could be ascribed to the widespread availability of open-source analytical tools and the accumulation of digital consumer footprints. Furthermore, the reviewed articles were predominantly published in high-impact journals (e.g., Journal of Marketing, Electronic Markets, Journal of Business Research, Journal of Cleaner Production, Technological Forecasting

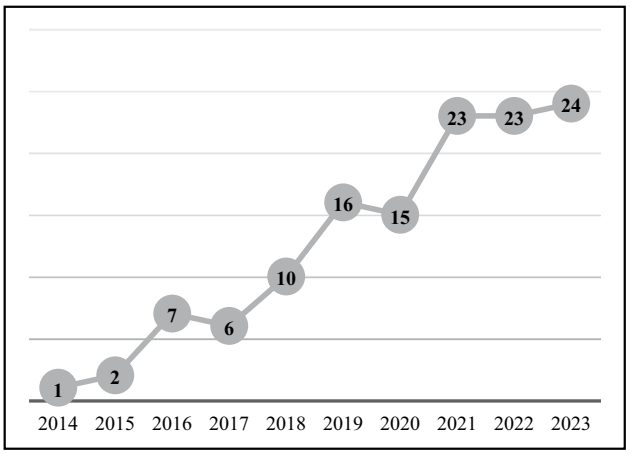

Fig. 2 Distribution of articles by year of publication. Source: Author's elaboration of data collected from Scopus and Web of Science

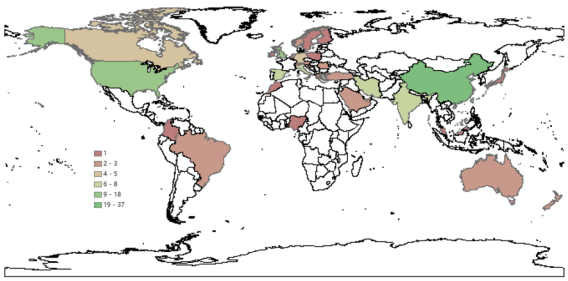

Fig. 3 GIS map of high-yield countries. Source: Author's elaboration of data collected from Scopus and Web of Science. *Notes*: Map generated using Natural Earth data. The boundaries and names shown on this map do not imply official endorsement or acceptance by the authors or the publisher

and Social Change), reflecting the growing academic interest and methodological maturity of this domain.

Distribution of articles by countries

According to Fig. 3, in terms of paper output by country, China, the United States, and South Korea published the most papers in this field, while the differences in paper output among other countries were relatively small. The top 4 countries (China, the United States, the Republic of Korea, and the United Kingdom) account for a significant portion of the total publications. There is a long tail distribution with many countries having fewer publications (2–3 each). European countries like Spain, Italy, Germany, and France have moderate numbers of publications. Other countries have fewer publications, indicating differences in research focus and resources. However, this concentration also suggests underlying structural factors beyond mere resource availability. The dominance of these regions may reflect disparities in digital infrastructure, such as data storage capabilities, as well as the prevalence of English as the lingua franca in high-impact journals, which can create entry barriers for scholars from non-Anglophone or developing regions.

Cooperation map

To better reflect the core authorship and relevance of big data in the field of consumer behavior analysis, the 127 papers were visualized using an author collaboration graph (Fig. 4).

The node sizes in Fig. 4 represent the number of papers published by the authors, and the lines represent the collaboration relationships. There are 11 core authors in the sample literature, and the top authors include Lawson M Cade, Francis Azell, Webb Anne, Asensio Omar Isaac, Bhardwaj Khushi, Hollauer Catharina, Banboukian Aline, Cotsman Ashley, Shalkh Omar, and Li Mimi. In general, the cooperation in the field of consumer big data research is relatively close, the research communities in this field are in a relatively concentrated state, and the scholars are closely connected. This connectivity supports the in-depth exploration of this field, as it facilitates the integration of diverse perspectives and aids the dissemination of innovative methodologies among core researchers.

Keywords analysis and thematic evolution

To map the conceptual structure of the field, we employed CiteSpace (v.6.3.R1) to generate keyword co-occurrence and timeline visualizations (Abramo et al. 2011; Chen 2006; Diem & Wolter 2013; Mayr & Scharnhorst 2015; van Eck & Waltman 2010). The specific parameters were set as follows: time slicing from 2012 to 2023 (1-year per slice), node types selected as "Keyword", and selection criteria based on the g-index (k = 25). No pruning methods were applied to preserve the network's structural integrity.

Figures 5 and 6 illustrate the evolution of key research terms. The timeline view reveals that the research trajectory can be interpreted through two distinct developmental phases, reflecting a notable theoretical shift in the field.

The research landscape between 2012 and 2017 was primarily defined by high-centrality keywords including "marketing", "behavior change" and "customer experience". From a theoretical perspective, this foundational period represented the digitization of traditional marketing paradigms, as researchers sought to translate established offline consumer psychology constructs, specifically satisfaction and attitude, into digital ecosystems.

In contrast, the period from 2018 to 2023 marked a distinct theoretical transition, evidenced by a surge in technical

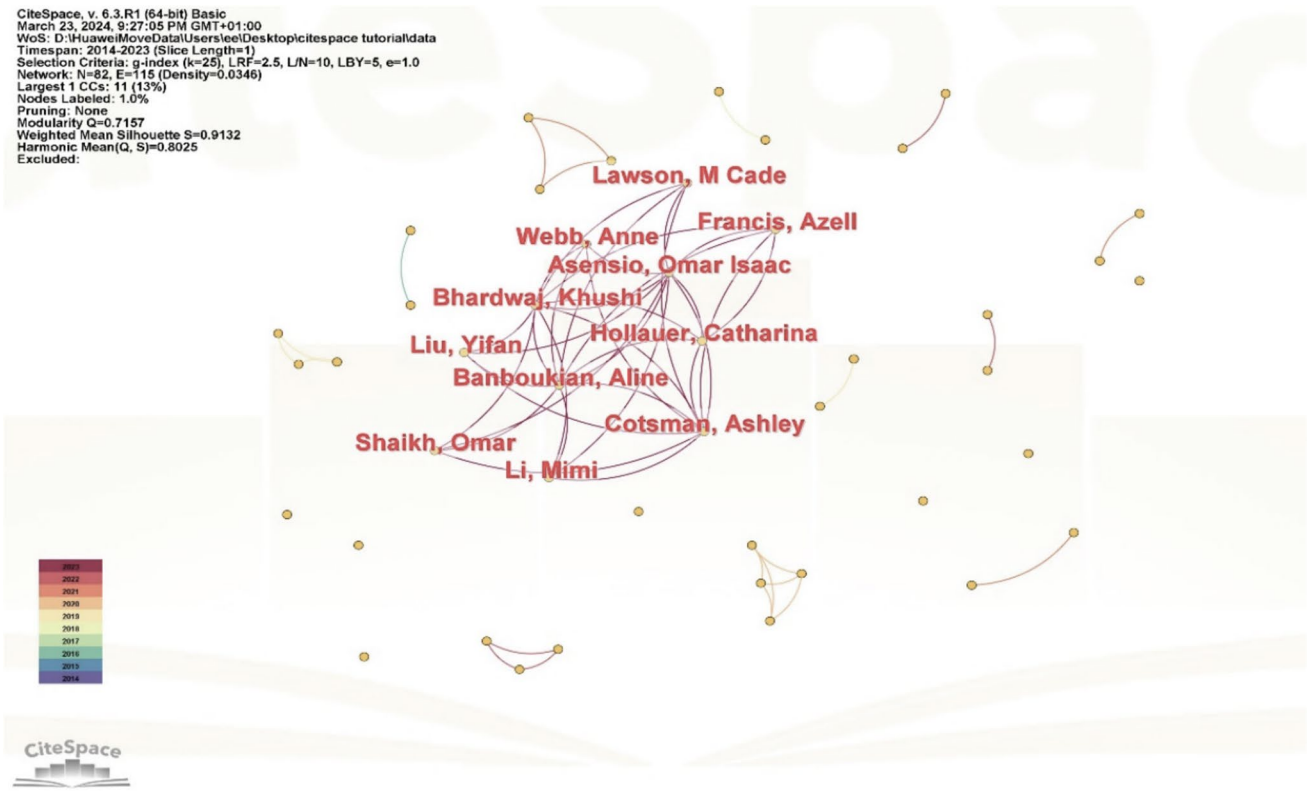

Fig. 4 Cooperation map of high-yield authors. Source of data: Author's elaboration of data collected from Scopus and Web of Science

Fig. 5 Top 10 keywords with the strongest citation bursts. Source of data: Author's elaboration of data collected from Scopus and Web of Science

Top 10 Keywords with the Strongest Citation Bursts

Keywords	Year	Strength	Begin	End	2014 - 2023
marketing	2014	1.56	**2014**	2018	
behavior change	2014	0.97	**2014**	2017	
customer behaviour	2015	1.38	**2015**	2018	
big data analytics	2016	1.85	**2016**	2019	
customer experience	2016	1.14	**2016**	2017	
data analytics	2017	1.95	**2017**	2018	
academic research	2018	0.97	**2018**	2019	
machine learning	2020	1.75	**2020**	2023	
aged	2020	1.3	**2020**	2021	
consumer attitude	2014	1.21	**2020**	2021	

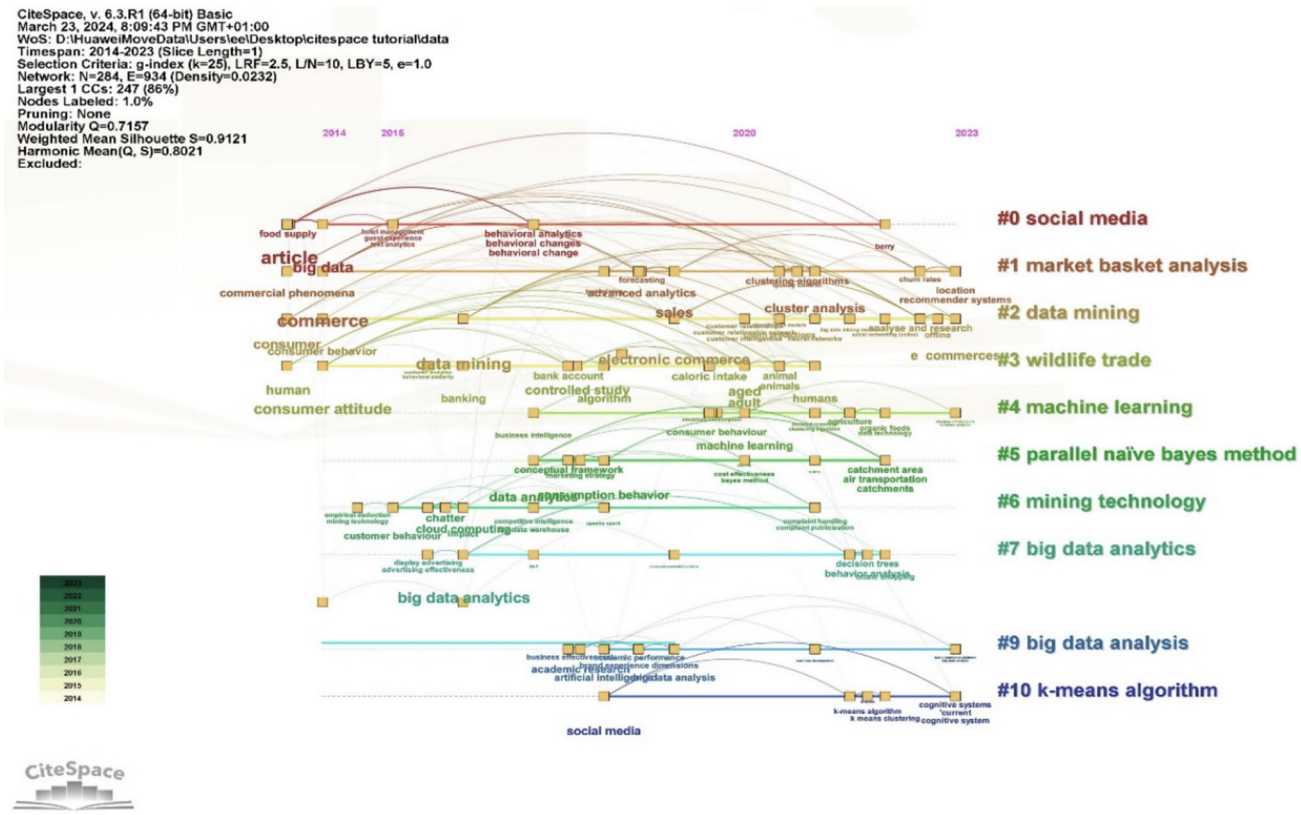

Fig. 6 Keyword timeline chart. Source of data: Author's elaboration of data collected from Scopus and Web of Science

terminology such as "machine learning", "deep learning" and "data mining". This shift highlights a movement from descriptive analytics to predictive modeling. Furthermore, the co-occurrence of terms like "consumer attitude" alongside "neural networks" indicates a modern interdisciplinary integration. In this context, advanced AI extends beyond its role as a mere tool, functioning instead as a methodological lens to examine the complex, non-linear psychological mechanisms underlying consumer behavior.

Data and technical foundations

Data sources in consumer behavior research

Big data in consumer behavior research is drawn from diverse sources, which can be critically mapped onto established theoretical frameworks such as the Stimulus-Organism-Response (SOR) model. Rather than viewing these sources in isolation, they are best understood by their capacity to capture distinct dimensions of the consumer experience.

Human-generated content, such as online reviews, social media posts, and blogs, serves as a primary window into the "Organism" component of the SOR framework. By analyzing spontaneous consumer attitudes and sentiments, researchers can decode internal cognitive and affective states that drive purchase decisions (Brachten et al. 2021). Similarly, transactional and web-based data represent the external "Stimuli". Clickstream data and interaction frequencies reveal browsing trajectories that often prefigure the "Response" phase (purchase or abandonment), offering empirical evidence to validate user experience dimensions, such as perceived usability, based on actual navigation patterns.

Beyond traditional streams, the landscape is rapidly evolving with emerging data modalities. We observe a significant shift towards Voice data and Biometric data (e.g., eye-tracking, heart rate variability), which capture subconscious physiological reactions, offering a deeper layer of implicit behavioral evidence that can triangulate findings from self-reported surveys; however, it is essential to treat these physiological signals as proxies for internal states rather than unmediated reflections of objective psychological truth. Furthermore, the rise of the Metaverse has introduced new forms of AR/VR behavioral trace data, while mobile app analytics now provide high-frequency location and usage context. These emerging sources provide continuous streams of real-time behavioral information that allow for hyper-personalized dynamic modeling (Rajan 2024).

While these diverse data sources offer a more holistic view of the consumer, their integration involves significant trade-offs. For instance, while unstructured user-generated content provides high ecological validity by capturing consumers in naturalistic settings, it is often plagued by "noise", lack of demographic representativeness, and selection bias. Conversely, biometric data offers high precision but often lacks the contextual depth required to explain the "why" behind a behavior. Recognizing these limitations is crucial for a balanced synthesis of big data insights.

Finally, open data from government or public institutions add valuable socioeconomic and demographic context, serving as control variables that enhance the robustness of behavioral models. To harness these diverse data types, organizations employ tools such as web scrapers, Application Programming Interfaces (APIs), and cloud-based infrastructures. Yet, critical challenges remain, such as the risk of re-identifying individuals within anonymized datasets (privacy) (Biswas et al. 2023; Martin et al. 2017; Nie & Han 2019; J. P. Singh et al. 2017; Soria-Comas & Domingo-Ferrer 2016), the vulnerability of large-scale repositories to unauthorized access or breaches (data security), and the complexity of ensuring users fully comprehend how their data is utilized in automated modeling (informed consent). These issues require strong governance to balance the depth of behavioral insight with ethical responsibility (S. Li et al. 2022a, b).

Table 2 Categories of big data used in consumer behavior research from this literature review. *Source of data*: Collected through a literature review compiled by the author

Categories	Subcategories	Example articles
Unstructured data	Marketing research data (e.g., Retailing and Advertising, diet survey data)	(Dubé et al. 2014), (Green et al. 2020)
	User-generated content (e.g., tweets, reviews of social websites, Google keywords)	(Ozturkcan et al. 2019), (Silva et al. 2020), (Pantano et al. 2019), (de Luca et al. 2019)
	Web search data (e.g., Taobao live banding data, TripAdvisor, e-commerce platform, Hadoop cloud computing platform)	(Xu & Chen 2023), (Giglio et al. 2020), (Xue 2023), (J. Wang & Zhang 2021)
Structured data	Enterprise database (e.g., ICT dataset from airplane, EIS)	(Adler et al. 2022), (Upadhyay et al. 2024)
	Industry database (revenues of luxury)	(Volkova & Karpushkin 2023)
	Professional database (e.g., genetic data, American Statistical Association DataFest)	(Daviet et al. 2022), (Y. Lee & Kim 2020)

Types of big data

Based on our review, big data used in consumer behavior research can be classified into structured and unstructured categories (Table 2). Structured data, typically stored in SQL or similar systems, includes enterprise, industry, and professional databases, while unstructured data comprises user-generated content, web search logs, and marketing research data.

Unstructured data have been widely adopted in recent years. User-generated content from platforms such as social media and e-commerce reviews enables the analysis of attitudes, preferences, and willingness to pay (Alyoubi 2019; Barbera et al. 2023; Bello-Orgaz et al. 2016; Du & Lin 2022; Ehsani & Hosseini 2025; Fernandez 2019; Hosseini & Ghalamkari 2018; Labrecque 2014; Lv 2022, 2023). Market research data provide precise insights into specific consumption contexts, such as food affordability and accessibility (Dubé et al. 2014; Green et al. 2020).Web search data, including search queries and clickstreams, reflect consumer intentions in real time and are increasingly used in text mining and personalized recommendation systems (Dutta & Das 2017; J. Wang & Zhang 2021; Xue 2023).

Structured data are also important but less extensively applied. Professional databases such as genetic datasets offer unique applications, for example integrating biological influences into consumer research (Daviet et al. 2022). Industry databases provide sector-specific insights, such as demonstrative consumption in the luxury fashion market (Volkova & Karpushkin 2023). Enterprise databases like ICT datasets and Google Trends are frequently used to analyze consumer trends and market predictions (Adler et al. 2022; Upadhyay et al. 2024). Together, structured and unstructured data offer complementary perspectives on consumer behavior. Nonetheless, the prevailing focus on unstructured data, which is largely driven by the availability of social media and web logs, may inadvertently lead to an algorithmic bias where findings are skewed toward digital-native populations, potentially marginalizing segments of the consumer base with lower digital footprints.

AI and machine learning models in consumer behavior analysis

The advent of digital transformation, social network adoption, cloud infrastructure, and big data technology now enables researchers to develop models to track and store observed customer behavior (Sarasquete 2017). Various predictive frameworks have been established to capitalize on this data. For instance, a Big Data-based Purchase Decisions Prediction Model was established to analyze consumer behavior on cross-border e-commerce platforms (W. Mu

2019), while an Improved Deep Forest model was designed to predict e-commerce consumers' repurchase behavior (W. Zhang & Wang 2021). Beyond the prediction of purchase and repurchase behaviors, a Deep Trust Network Model was developed to understand behavioral characteristics in consumer patterns (Y. Wang 2022). For predicting consumer behavior, a Neural Network-based Precision Marketing Model focuses on data mining to study user churn prediction and user value enhancement, which are the two most critical factors influencing marketing revenue (H. Liu 2021). In the context of product lifecycle management, a multi-scale digital model was developed to support complex decision-making (Udugama et al. 2023).

As big data technology rapidly evolves, new algorithms and models have emerged to handle the extensive use of large datasets. For example, text mining and data mining play significant roles in this field (Praveen Kumar et al. 2019). These algorithms are instrumental in predicting and tracking consumer behavior by structuring unstructured data (El Majdouli et al. 2016; Y. Liu 2016; Rendón et al. 2020). An algorithm for collecting and processing power consumption data, along with a load planning algorithm, was developed to encompass all levels of device interaction (Zhukovskiy et al. 2021). Furthermore, consumer preferences and attitudes (Serrano et al. 2021), online consumption behavior (Evangelin & Vasantha 2022), and energy consumption (Abassi et al. 2023) were all analyzed using text mining and data mining, making these algorithms widely popular in the analysis of big data for consumer behavior. In addition to data mining and text mining, CNN-LSTM (Alikhani et al. 2022) and ANN (Praveen Kumar et al. 2019) models have been applied in demand response and retail management. It is important to note that the application of such complex models often requires substantial computational resources and extensive training datasets, which can present a barrier to entry for smaller-scale studies compared to simpler statistical approaches.

Consumer behavior researchers have also employed AI algorithms in big data analysis for consumer behavior research. Among AI algorithms, machine learning and deep learning are commonly used in this context. An analysis of online consumer behavior data from the Google Merchandise Store found that the ensemble model, eXtreme Gradient Boosting (an AI algorithm), is the most suitable for predicting purchase conversions among online consumers, with oversampling identified as the best method to mitigate data imbalance bias (J. Lee et al. 2021). Additionally, researchers have demonstrated that machine learning is more effective than traditional algorithms in handling non-linear patterns, with larger sample sizes leading to more accurate results. The decision tree model and the arithmetic mean calculation method have proven more effective than conventional algorithms (R. Li et al. 2022a, b). The combination of machine learning and data mining has also been utilized to analyze consumer behavior in the context of energy usage (Abassi et al. 2023). Despite these advantages, the reliance on algorithmic output requires caution. While larger sample sizes improve accuracy, they do not automatically eliminate inherent data biases. Thus, it is likely that new AI algorithms and techniques will continue to be adopted in consumer behavior research in the future.

Big data applications for understanding consumer behavior

Big data has become an essential tool for analyzing various dimensions of consumer behavior, including consumer consumption patterns, sentiment analysis and emotion mining, decision-making pathways, future behavior predictions, identify factors that affect consumer behavior and responses to macro-level events. With the proliferation of social media platforms, e-commerce websites, and advanced algorithms such as artificial intelligence, data collection has become increasingly efficient, enabling more sophisticated analyses (Adler et al. 2022; Y. Liu 2016). Unlike traditional approaches that often relied on single-source datasets and static surveys, modern applications of big data leverage integrated and heterogeneous data sources to provide more comprehensive insights into consumer behavior (Green et al. 2020). These integrated datasets allow for a multi-faceted understanding of consumer activity when used in a complementary manner.

The rise of digital technologies has also transformed how consumers express opinions and interact with products online. Platforms such as Facebook, X (formerly Twitter), and TripAdvisor have facilitated the formation of consumer communities based on shared preferences and behaviors (Adamopoulos et al. 2018). The widespread use of these platforms has significantly reshaped information acquisition and decision-making processes in the consumer landscape (Ghose & Todri-Adamopoulos 2016). Given that big data encompasses both pre-purchase and post-purchase activities, it enables a deeper exploration of how consumption behaviors evolve across digital environments.

In the following sections, we classify and discuss the major applications of big data in consumer behavior research, while critically evaluating how these methods offer superior explanatory power compared to traditional paradigms and acknowledging the methodological challenges and boundary conditions that limit their application.

Analyzing consumption patterns

As shown in Table 3, the literature review spans a wide range of consumer consumption topics, including motivations for consumption (Zhou et al. 2019), sustainable consumption (Chandra & Verma 2023; Guo & Zhang 2019; Oh et al. 2019; Pham et al. 2019; D. Ye et al. 2022a, b), consumption structure (Guo & Zhang 2019), food consumption (Vepsäläinen et al. 2022), repurchase behavior and household electricity consumption (Oprea et al. 2021; Oprea & Bâra 2020; Ushakova & Jankin Mikhaylov 2020), alongside diverse demographic behaviors (Adinolfi et al. 2011; Chiang & Yang 2018; Dąbrowska 2011; X. Liu & Zhao 2021; Morris et al. 2020; Moschis 2012). These studies collectively reflect the richness and diversity of consumption patterns that can be decoded through big data analytics.

From a methodological perspective, two dominant analytical approaches emerge: text mining and data mining, each applied to different types of big data such as social media posts, transaction logs, smart meter data, and user-generated content. Unlike traditional self-reported surveys which often suffer from recall bias, these methods allow for the observation of actual behavioral traces.

In text mining analysis, the process typically involves content filtering and data extraction (J. Li & Hu 2021). For instance, Y. Ye et al. (2022a, b) categorized consumers into green and traditional groups, demonstrating that sustainable consumption behaviors can create positive spillover effects across various domains. Similarly, analysis of tweets has shown that consumer patterns are regionally differentiated, shaped by local cultural contexts and attitudes toward food innovations (Pindado & Barrena 2021). Studies leveraging word-of-mouth (WOM) data have revealed that personality similarity among social media users significantly increases post-exposure purchase likelihood (Adamopoulos et al. 2018). NLP-based psychographic segmentation, using the Big Five Factor (BFF) model and the Schwartz Value Survey (SVS), further enables refined profiling of consumer preferences (H. Liu et al. 2019). Online reviews, particularly negative ones, also provide valuable temporal and emotional cues about consumer sentiment and behavioral cycles (Sun & Zhao 2022).

In data mining analysis, various algorithms and frameworks are employed to extract patterns from structured datasets. For example, F. Zhou et al. (2019) found that "cost-effectiveness" is the most influential factor in vehicle purchasing decisions. Guo & Zhang (2019) used e-commerce transaction data to detect shifting consumption structures over time, indicating a trend toward consumption upgrading. Smart meter data have proven effective in uncovering residential electricity usage patterns (Ushakova & Jankin Mikhaylov 2020), while supermarket purchase logs provide reliable insight into dietary behaviors (Clark et al. 2021).

Advanced machine learning and pattern recognition techniques are also applied. For instance, S. Singh & Yassine (2019) proposed an unsupervised incremental mining mechanism for analyzing energy consumption patterns. L. Zhang et al. (2021) employed segmentation models to identify consumers with similar purchasing behaviors, thereby optimizing cross-selling strategies and enhancing loyalty. Social media and search engine data have been used to design targeted interventions, such as wildlife conservation strategies that address specific consumption motivations (J. Li & Hu 2021).

Overall, the identification and interpretation of consumption patterns through big data relies on a combination of diverse data sources, advanced analytical models, and contextual behavioral theories. This multi-dimensional approach provides a robust foundation for understanding both individual and collective consumption trends. However, it is crucial to acknowledge the boundary conditions of these findings. While big data excels at mapping what consumers buy, it often lacks the contextual depth to explain why specific patterns emerge.

Sentiment analysis and emotion mining

According to Table 3, the emergence of advanced data analysis technologies and algorithms, particularly artificial intelligence (AI), has substantially enhanced the capacity to analyze consumer attitudes, preferences, and satisfaction levels. Researchers are increasingly adopting big data techniques such as text mining and machine learning to capture consumer sentiment across various contexts. Examples include hotel guest satisfaction and public attitudes toward organic food (Filieri et al. 2022; Gan & Ouyang 2022; Y. Kim et al. 2023; Lan 2023; C. K. H. Lee et al. 2019; M. M. Mariani et al. 2019; M. Mariani & Predvoditeleva 2019; Mendieta-Aragón & Garín-Muñoz 2023; Ramos et al. 2017; A. Singh & Glińska-Neweś, 2022; Venkatraman 2017). In consumer identification and sentiment recognition, AI is employed to classify and understand user profiles (Xie 2023), while enterprises apply data mining to enhance customer relationship management.

A notable trend in the reviewed literature is the extensive use of user-generated content (UGC), including online reviews, social media posts, and e-commerce comments, as a primary data source for sentiment analysis. This focus allows researchers to ground their analysis in established theoretical frameworks. For example, sentiment dynamics in reviews can be interpreted through Affective Events Theory, which posits that specific service encounters trigger emotional reactions driving subsequent behaviors.

Table 3 Synthesis of big data applications in consumer behavior research. *Source of data*: Collected through a literature review compiled by the author

Research Domain	Data Sources	Analytical Methodologies	Key Research Inquiries and Findings	Representative Studies
5.1 Consumption Patterns	E-commerce transaction records	Data mining; Trend analysis	Detected a structural evolution in consumption towards higher quality and "consumption upgrading" over time	Guo and Zhang (2019)
	Smart meter data (Electricity)	Pattern recognition; Unsupervised mining	Identified distinct household energy usage profiles, enabling granular segmentation beyond demographics	Ushakova and Jankin Mikhaylov (2020)
	Supermarket purchase logs	Frequency analysis; Behavior modeling	Revealed discrepancies between actual dietary habits (purchase data) and perceived healthy eating	Clark et al. (2021)
5.2 Sentiment and Emotion	Online reviews; Social media (UGC)	Text mining; Sentiment analysis	Demonstrated that naturally occurring UGC captures dynamic sentiment fluctuations more authentically than static surveys	Lv et al. (2022)
	Hotel reviews; Food media	AI-based classification; Satisfaction modeling	Validated that specific service encounters trigger emotional reactions that drive satisfaction (Affective Events Theory)	Singh and Glińska-Neweś (2022)
	IoT behavioral data	Real-time analytics	Enabled the capture of real-time emotional states and immediate reactions to environmental stimuli	Brachten et al. (2021)
5.3 Decision Pathways	Social media UGC	Purchase intention modeling	Found that UGC acts as a critical "social proof" mechanism that directly reshapes purchase intentions	Kiran and Vasantha (2016)
	E-commerce interaction logs	User journey analysis	Showed that simplifying digital touchpoints reduces cognitive load, significantly improving conversion rates	Rajan (2024)
	IoT-enabled environments	Decision network modeling	Illustrated how IoT integration streamlines decision networks by providing real-time information feedback loops	Yan et al. (2020)
5.4 Predicting Behavior	Web search data (e.g., Google Trends)	Trend forecasting; NLP; AI	Proved that search query volume is a leading indicator for predicting consumer demand and preferences	Buettner (2017); Silva et al. (2019)
	Structured and Unstructured data	Deep learning; Hybrid models	Confirmed that combining diverse data types (Hybrid approach) significantly outperforms single-source models in predictive accuracy	Ryu et al. (2020)
	CRM and Customer databases	Predictive analytics; Segmentation	Utilized historical interaction data to accurately predict user churn and optimize retention strategies	Šimović et al. (2023)
5.5 Influencing Factors	E-commerce behavior data	Multi-stage decision modeling	Revealed that different factors influence consumers at different stages (e.g., browsing vs. purchasing) of the decision journey	Xiao and Piao (2022)
	Social media profiles	Psychographic profiling	Found that personality similarity between users significantly amplifies the impact of Word-of-Mouth (WOM) on decisions	Adamopoulos et al. (2018)
	Sociodemographic data	Factor analysis	Quantified the impact of income, social dynamics, and product attributes on green purchasing decisions	L. Dong (2022)
5.6 Macro-events Response	Online transaction platforms	Digital behavior modeling	Captured a rapid, non-linear shift to digital purchasing channels during the COVID-19 pandemic	Sakas et al. (2021)
	Supply chain transaction logs	Panic buying modeling	Identified that panic buying is primarily driven by scarcity anxiety and uncertainty rather than actual supply shortages	Prentice et al. (2020)
	Social media (Disaster periods)	Social signal analysis	Showed positive correlations between disaster recovery discussions and demand for specific goods (e.g., used vehicles)	Shibuya and Tanaka (2018)

Similarly, comparisons between pre-consumption expectations and post-consumption sentiment in UGC align with Expectancy-Disconfirmation Theory, offering a dynamic view of satisfaction formation.

Unlike traditional survey-based research that relies on stated preferences (Lv et al. 2022), big data approaches leverage naturally occurring UGC to extract insights from larger, more diverse samples while mitigating self-reporting bias. Although this approach reduces certain biases, it is important to acknowledge others, such as selection bias where the data may only represent digitally active populations, and algorithmic bias. This shift enables a more realistic understanding of consumer attitudes and emotional responses. Studies have identified price, subsidies, and after-sales service as influential factors shaping consumer preferences, with price often playing the most dominant role (Jung et al. 2021). Other studies suggest that psychological depreciation and perceived value significantly impact attitudes toward product repair decisions, such as with smartphones (Makov & Fitzpatrick 2021). Despite these advantages, the reliability of sentiment analysis is often challenged by data quality issues. A significant portion of UGC may contain "noise", such as sarcasm or fake reviews, which standard text mining algorithms struggle to interpret accurately. Furthermore, distinct inconsistencies often arise between sentiment scores and actual behavior; high online engagement does not always translate into purchase actions, highlighting a gap between digital expression and physical consumption.

The comparison between traditional and big data-based sentiment analysis reveals methodological differences: while conventional methods depend heavily on subjective questionnaires or interviews, big data analytics processes are grounded in the objective evaluation of behavioral data using pre-defined algorithms and classification rules (Makov & Fitzpatrick 2021). However, it is critical to note a limitation: while AI approaches excel at processing volume and identifying correlation, they sometimes lack the nuance of qualitative coding in detecting sarcasm or complex emotional context without advanced Natural Language Processing (NLP) fine-tuning. For instance, AI-powered tools allow firms to develop detailed consumer preference profiles, which in turn support the design of targeted, personalized marketing strategies. Furthermore, the integration of Internet of Things (IoT) technologies enables real-time data collection from various devices, providing an even richer base for understanding consumer sentiment.

In the area of personalized management, especially within the energy sector, big data tools help identify key energy consumption patterns to develop user-specific solutions. An energy-aware IT ecosystem was proposed to aggregate behavioral data via IoT, combined with personalized recommendation mechanisms and predictive modeling (Fotopoulou et al. 2017). Similarly, spatial–temporal and financial data have been proven effective in predicting consumer attitudes (Urkup et al. 2018).

For product recommendation systems, large-scale data mining on consumer behavior has been applied to improve accuracy and scalability. One example is the use of Amazon behavioral data to develop a recommendation system supported by cloud infrastructure and the MapReduce programming model (Kanavos et al. 2018). In parallel, intelligent recommendation frameworks have been developed and deployed in B2C e-commerce platforms, enhancing both targeting accuracy and advertising effectiveness (Fotopoulou et al. 2017; Jiménez-Marín et al. 2020).

In conclusion, sentiment mining through big data empowers organizations to move beyond surface-level consumer insights and instead derive emotionally rich, behavior-driven intelligence. This enables more effective personalization, higher customer engagement, and stronger predictive capabilities.

Modeling decision-making pathways

As indicated in Table 3, understanding consumer decision-making processes is a central objective in the analysis of behavioral mechanisms, particularly through the modeling of decision pathways. Conceptually, big data has reshaped the traditional funnel model of decision-making into a non-linear, network-based framework where feedback loops occur in real-time. A key topic in this area is consumer willingness to purchase, which has increasingly been investigated using user-generated content (UGC). UGC serves a vital function in uncovering consumers' intentions and underlying thought processes. For example, existing research demonstrates that UGC from social media platforms, when analyzed through big data methodologies, can significantly influence consumers' purchase intentions (Kiran & Vasantha 2016).

Another important aspect of decision-making concerns online shopping behavior. Studies have shown that minimizing the number of touchpoints in e-commerce platforms can lead to higher sales conversion rates, as simplified decision journeys reduce cognitive load. The availability of detailed online information also improves consumers' ability to evaluate products and services before making decisions (Xiao & Piao 2022). In this context, the Internet of Things (IoT) has fundamentally reshaped consumer decision networks by enabling real-time data flows. The integration of IoT and big data analytics has proven effective in clarifying decision-making mechanisms and mitigating decision fatigue (Yan et al. 2020). This mirrors the current trajectory of the field, where AI and data-driven frameworks are increasingly recognized as central to modernizing

decision-making paradigms (Petrescu & Krishen 2024). However, the effectiveness of such streamlined pathways is not absolute. While automation reduces cognitive load for routine purchases, it can be less effective for high-involvement decisions where consumers actively seek detailed information. In these cases, over-simplified algorithmic recommendations may fail to address the consumer's need for comprehensive evaluation, suggesting that the value of big data interventions depends heavily on the product category.

From a technical modeling perspective, decision tree algorithms and artificial intelligence (AI) have been employed to simulate and analyze consumer behavior patterns. For example, an AI-driven analysis of Expedia.com user data revealed that longer-stay hotel guests are more likely to choose bundled packages, with clear regional differences in preferences (Y. Lee & Kim 2020). These insights demonstrate how predictive modeling techniques can be leveraged to uncover latent structures in consumer choices and optimize targeting strategies.

In essence, the modeling of consumer decision pathways benefits significantly from combining behavioral data sources such as UGC with algorithmic methods like AI and decision trees. This approach offers a robust framework for understanding how consumers evaluate options and make purchase decisions in increasingly digital environments.

Predicting future consumer behavior

As presented in Table 3, after identifying existing consumption patterns, sentiment analysis and decision-making, the next critical step is the prediction of future consumer behavior, which plays a vital role in informing strategic marketing decisions. Accurate forecasting not only enhances consumer engagement and resource allocation but also contributes to long-term business growth (Buettner 2017).

A variety of big data sources are used in predictive modeling (Chang et al. 2023; Fang et al. 2020; Golmohammadi et al. 2021; Izang et al. 2019; D. Kim et al. 2019; C. Liu et al. 2023; Maeng et al. 2023; Mendieta-Aragón & Garín-Muñoz 2023; Molitor et al. 2023; Naraine et al. 2020; Seo & Yoo 2023; Shokouhyar et al. 2021; Y. Tian 2022; Tupikovskaja-Omovie & Tyler 2021; Wei et al. 2022; Yang et al. 2020), ranging from web search data (e.g., Google Trends, XING social network datasets) (Buettner 2017; Silva et al. 2019), to payment and transactional data (Martens et al. 2016), customer relationship management (CRM) records (Šimović et al. 2023), and data extracted from cross-border e-commerce platforms (Mu 2019). Many of these sources include both structured (e.g., transaction logs) and unstructured data (e.g., user-generated content, search queries), which together enrich the predictive accuracy of behavioral models (Ryu et al. 2020).

A critical methodological comparison arises here between AI-based predictive analytics and traditional econometric models. While traditional econometrics excels at identifying the statistical significance of relationships, AI and machine learning models, such as deep learning and Random Forests, often demonstrate superior performance in prediction accuracy and scalability by capturing complex, non-linear relationships that linear models might miss. For instance, recent studies applying machine learning approaches to the Indian market demonstrated significantly higher accuracy in predicting niche behaviors, such as green purchase intention, compared to conventional linear models (Choudhury et al. 2024). Among the 127 studies collected on behavior forecasting with big data, the earliest efforts date back to 2016 (Martens et al. 2016). Key research themes include forecasting personality-driven product preferences (Buettner 2017), variety-seeking behavior (Tian et al. 2018), and fashion consumption trends (Silva et al. 2019). Across these studies, researchers have increasingly employed artificial intelligence (AI) and machine learning (ML) techniques, including deep learning and natural language processing (NLP). These advanced approaches have demonstrated superior performance in terms of both accuracy and scalability when compared to traditional methods such as linear regression and decision trees (Jackson & Ivanov 2023). However, this superiority is not absolute. Researchers have observed instances of overfitting, where models perform exceptionally well on historical training data but fail to generalize to new, unseen market conditions.

In particular, Google Trends has emerged as a powerful tool for forecasting macro-level consumer behavior. For instance, Havranek & Zeynalov (2021) demonstrated that search data significantly improves the reliability of tourism demand predictions, offering actionable insights for destination marketing organizations. More broadly, studies validate the effectiveness of algorithmic models including text mining and ML frameworks in identifying evolving consumer preferences and anticipating behavioral shifts.

In essence, the predictive modeling of consumer behavior through big data integrates heterogeneous datasets and advanced analytical methods. This enables organizations to proactively respond to market dynamics, personalize offerings, and optimize strategic planning based on data-driven foresight.

Identifying influencing factors

According to Table 3, while relatively few studies have directly analyzed influencing factors through big data, those available offer valuable insights into the multi-stage and multi-dimensional nature of consumer decision-making. A key conclusion from this literature is that consumer

decision-making is a multi-stage process, with different influencing variables shaping each stage (Xiao & Piao 2022). Major categories of influencing factors include product attributes, psychological factors, income levels, social dynamics, and cultural influences (L. Dong 2022). Additionally, organizational factors such as privacy management, strategic alignment, structural design, and functionality affect how big data is adopted and applied within enterprises (Félix et al. 2018). More specifically, recent research highlights that the quality of analytics and the strategic role of decision-makers are pivotal in translating these data capabilities into firm-level marketing agility (Haverila et al. 2025).

Beyond these studies, broader research shows that consumer behavior is shaped by a complex interplay of individual, social, and cultural factors. For example, cultural background significantly influences consumption patterns. A study conducted in Korea found that although social class characteristics do affect organic food consumption, individual lifestyle is a more decisive factor in actual purchasing behavior (Han & Lee 2022). Similarly, regional cultural preferences were found to influence the growing popularity of Chinese barbecue, reflecting localized convergence in food choices (B. Wang et al. 2023).

Sustainability orientation is another important driver. Consumers who prefer socially responsible products are more likely to engage in sustainable consumption behaviors (Y. Ye et al. 2022a, b). On the psychological level, personality traits also play a critical role. For instance, consumers who display similar personality profiles on social media platforms are more likely to respond positively to word-of-mouth messages and make subsequent purchases (Adamopoulos et al. 2018).

Finally, exposure to digital advertising influences consumer behavior in nuanced ways. According to Ghose & Todri-Adamopoulos (2016), consumers with limited exposure to display ads tend to initiate more exploratory brand searches, whereas those with high ad exposure are more prone to direct brand search behavior.

From an overall perspective, while empirical research using big data to study influencing factors remains limited, available studies point to a multi-dimensional framework in which cultural, psychological, social, and technological factors intersect to shape consumer decision-making. Future work should expand on these findings by leveraging more diverse big data sources and advanced modeling techniques to bridge the gap between identifying correlation and establishing causation. Despite these granular insights, a recurring critique in the literature addresses the epistemological limitation of big data, the difficulty in distinguishing correlation from causation. While factor analysis can identify strong associations (e.g., between income and green purchasing), it often fails to account for confounding variables that are not captured in the digital footprint.

Assessing the impact of macro-level events

As illustrated in Table 3, there is a notable relationship between major societal disruptions and changes in consumption patterns. Such events, which include global pandemics and natural disasters, have presented valuable opportunities for behavioral analysis through big data. The strength of big data here lies in its immediacy; unlike traditional retrospective studies, big data allows for the monitoring of behavioral adaptations in real-time as events unfold.

The COVID-19 pandemic has been the most extensively studied macro-event in recent years. Researchers have explored its impact on multiple facets of consumer behavior, including changes in Airbnb booking patterns (Filieri et al. 2023), shifts in online purchasing behavior (Sakas et al. 2021), and consumer satisfaction at point-of-sale locations (Brandtner et al. 2021). The crisis also triggered observable differences in consumer preferences between mass-market and luxury goods (Pang et al. 2022), and significantly influenced panic buying behavior during supply chain uncertainty (Barnes et al. 2021; Prentice et al. 2020). In addition, studies found that perceptions of service quality were notably affected by the pandemic context, especially in sectors where in-person interaction was limited (Nilashi et al. 2021).

Beyond the pandemic, other macro-events have also been shown to reshape consumer decision-making. For example, in the aftermath of the Great East Japan Earthquake and Tsunami, social media activity related to recovery and rebuilding efforts was found to be positively correlated with increased demand for used vehicles, suggesting a rapid behavioral adaptation driven by necessity (Shibuya & Tanaka 2018).

These findings underscore the potential of big data analytics to assess the real-time behavioral consequences of macro-level disruptions. By capturing large-scale behavioral shifts across digital platforms, researchers can offer timely insights into how consumers respond under crisis conditions. Such insights are increasingly valuable for crisis management, supply planning, and adaptive marketing strategies. However, data derived from these events is subject to representativeness bias.

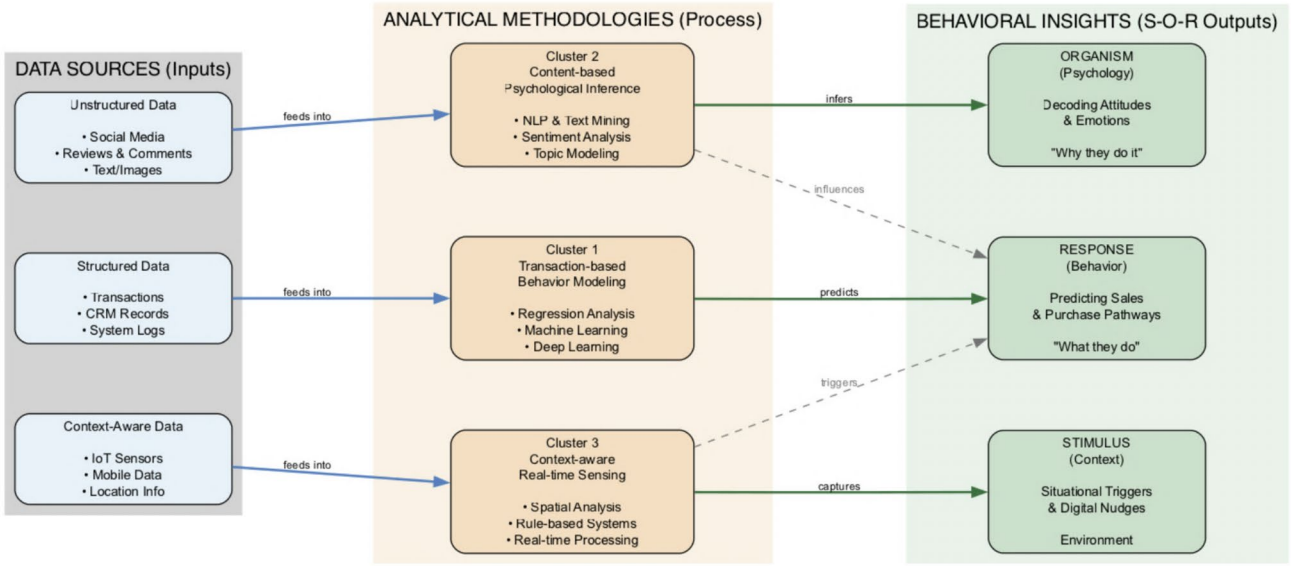

Fig. 7 The "Data-Method-Theory" integrative framework. Source of data: Author's elaboration

Discussion

Integrative framework and answers to research questions

This study provides a systematic review of big data applications in consumer behavior research. To answer our research questions regarding the evolution of methodologies (RQ1) and the derivation of behavioral insights (RQ2), we propose an integrative framework that connects data sources, analytical methods, and theoretical outcomes. Specifically, to provide a complete answer to RQ2 regarding evolving research themes, we synthesize these findings into a "Data-Method-Theory" integrative framework (see Fig. 7). It is important to clarify that this framework is synthetic in nature, as it is derived from mapping existing literature patterns rather than being a product of formal theory building. It serves as a heuristic device to organize the field into three distinct clusters: (1) Transaction-based behavior modeling (using structured data and regression/ML to predict sales); (2) Content-based psychological inference (using unstructured text/images and NLP to decode attitudes); and (3) Context-aware real-time sensing (using IoT/mobile data to capture situational triggers).

This evolutionary view directly addresses RQ2 by showing how the integration of data science has shifted from validating traditional economic outcomes to discovering latent psychological mechanisms. Compared with earlier reviews, this framework offers a broader and more up-to-date synthesis that spans structured and unstructured data as well as multiple behavioral dimensions, including attitudes, preferences, and decision-making processes. It highlights that the choice of data source is no longer merely technical but theoretically consequential, as it dictates which aspect of the "black box" (cognitive vs. behavioral) can be illuminated.

Theoretical contributions

Our findings directly address the research question concerning how big data reshapes theoretical understanding. Unlike earlier reviews that mainly concentrated on specific domains such as e-commerce or recommendation systems (Buettner 2017; Xiang et al. 2015), this study demonstrates that big data challenges the traditional "rational agent" assumption by capturing impulsive, context-dependent behaviors that surveys often miss.

A significant observation emerging from this review is the substantial divergence between "stated preferences" (derived from self-reported surveys) and "revealed preferences" (evidenced by actual digital footprints). While traditional literature assumes a high correlation between intention and action, our review of big data applications reveals significant discrepancies, suggesting that contextual triggers (captured via IoT or clickstreams) often override stable preferences. This aligns with recent findings on consumer readiness, where factors like green awareness act as critical moderators that bridge the gap between abstract green attitudes and actual purchase intentions (Shehawy & Ali Khan 2024). Our synthesis suggests that big data does not just validate existing theories but exposes the limitations of traditional rational choice frameworks, showing that digital environments often induce non-linear decision pathways. However, a skeptical perspective is necessary, Big Data may predominantly capture impulsive reactions to

platform architecture, thereby complicating the theoretical interpretation of consumer autonomy.

Methodological transition in big data-driven consumer research

The review reveals a methodological transition from descriptive analytics to predictive and prescriptive modeling, enabled by advances in machine learning and AI-driven approaches such as deep learning, natural language processing, and ensemble models (J. Lee et al. 2021; Xue 2023). This evolution reflects a shift from analyzing structured historical records toward decoding complex, real-time behavioral streams. However, a critical evaluation of the reviewed literature exposes a significant limitation regarding the trade-off between prediction and explanation. This limitation fuels a growing skeptical perspective in the literature, which suggests that the pursuit of algorithmic accuracy may come at the cost of theoretical hollowness. While big data proponents argue for a theory-free approach, critics contend that without the explanatory power provided by traditional theory, big data models remain fragile against structural market changes. This reveals a significant discrepancy between predictive accuracy and theoretical interpretability, thereby reducing our ability to derive robust generalizations.

Moreover, while the literature highlights the integration of psychological constructs with algorithmic modeling (Buettner 2017; Makov & Fitzpatrick 2021), many studies still treat psychological theories as secondary to algorithmic accuracy. We systematically address these methodological gaps, noting that current research often overlooks ethical challenges, including generalizability, the reliability of automated data (veracity), and user privacy (Barbado et al. 2019; Zhao et al. 2021).

Future research directions

Based on the identified gaps, future research could advance along three specific and analytical directions.

First, moving beyond simply collecting "more data", researchers could focus on causal AI and the integration of heterogeneous and multimodal data. Current studies often confuse correlation with causation. The development of AI-driven data fusion should aim to solve this by combining behavioral big data with field experiments (A/B testing). Crucially, this integration requires rigorous protocols for examining the quality of input data. Future scholars could develop standardized "data auditing" frameworks to detect noise, bot traffic, and algorithmic bias in unstructured datasets before ensuring they are suitable for behavioral inference. This hybrid approach is particularly promising when combined with traditional methods such as surveys to validate the "why" behind the "what" (Anderson et al. 2016; Chaudhary et al. 2021; Y. Wang 2022).

Second, expanding the research context is essential to test the boundary conditions of consumer theories. Future research should apply data-driven insights to specific value-based consumption contexts. For example, scholars could investigate how consumption values interact with green behavior in developed markets (Chwialkowska et al. 2024) or examine the role of green consumer values in the willingness to buy specific categories like regional agricultural products (Eastman et al. 2025). Beyond individual-level analyses, future studies could consider group dynamics, organizational contexts such as nonprofit institutions, and industry-level applications in areas like tourism, transportation, and crisis management (Grandhi et al. 2021; Vepsäläinen et al. 2022). For instance, how does the "herding effect" in social commerce differ between crisis periods and normal periods? Such an expansion would allow a more nuanced understanding of how big data influences behaviors across multiple levels.

Third, advocate for cross-cultural validation and advanced privacy protection mechanisms. Behavioral models developed in single cultural settings may lack external validity in mobile-first markets like China or Africa, making cross-national comparisons crucial. Simultaneously, ethical inquiries must transcend general privacy discussions to address the specific dynamics of the "privacy paradox", investigating why consumers, despite professing to value privacy, frequently trade it for minor conveniences. Furthermore, potential psychological impacts must be systematically addressed through fairness audits, transparency mechanisms, and stakeholder engagement. This aligns with the emerging call for marketing analytics research to transcend commercial metrics and actively contribute to the "better good" and societal well-being (Petrescu & Krishen 2025).

Limitations

Several methodological, theoretical, and epistemic limitations must be acknowledged. Regarding methodological constraints, this review is restricted to English-language journal articles, excluding industry reports and non-English literature that could enrich cross-cultural diversity. We also acknowledge that the data collection relied on specific keywords and database criteria tailored to the study's objectives. Consequently, the final sample of 127 articles represents a purposive selection rather than an exhaustive representation of the entire field. The manual screening process may also introduce selection bias despite quality control measures. From a theoretical perspective, the proposed integrative framework is synthetic and descriptive, meaning it serves to

organize current trends but requires further empirical validation to establish its predictive power. Epistemically, the reliance on digital behavioral traces may disproportionately reflect the activities of digitally active populations, potentially overlooking consumers with limited digital footprints and thus creating a "digital shadow" in the derived insights. Finally, given the rapid evolution of big data and AI, the findings are time sensitive.

Conclusion

This systematic review of 127 peer-reviewed journal articles published between 2012 and 2023 provides an integrated perspective on how big data is utilized to decode consumer behavior. By aligning data sources, analytical methods, and research objectives, this study offers a comprehensive framework that captures the evolving dynamics of consumer behavior research in the era of big data. Our findings show that big data enables large-scale, real-time analysis of consumer preferences, decision-making processes, sentiment, and responses to macro-level events. Structured data sources (e.g., enterprise databases) and unstructured content (e.g., online reviews, social media) are widely applied, with analytical techniques such as machine learning, deep learning, and text mining driving the transition from descriptive to predictive and prescriptive modeling.

The implications of this review extend to three distinct domains. First, for academic research, this study contributes to theoretical advancement by highlighting the shift towards data-driven psychological inference. It provides a timely synthesis of key themes, analytical trends, and research gaps, suggesting that future scholarship could integrate computational methods with established behavioral theories to avoid "atheoretical" data mining. Second, for practitioners and managers, it offers strategic insights on leveraging big data for personalized marketing, behavioral prediction, and consumer engagement. Our review suggests that managers could move beyond static demographic segmentation to dynamic, context-aware targeting based on real-time digital footprints, thereby enhancing conversion rates through hyper-personalization. Third, for policymakers, this study highlights the growing need for consumer data literacy and robust ethical frameworks. As individuals increasingly interact with algorithm-driven platforms in everyday decision-making, regulators could enforce compliance with frameworks like the General Data Protection Regulation (GDPR) to address specific concerns about the authenticity of user-generated content, unequal access to proprietary data, and ensure strict transparency regarding how consumer data is collected and utilized. Furthermore, establishing robust mechanisms for trust-building is essential, particularly in platform-based environments like the sharing economy, where trust dictates user engagement and data sharing willingness (Hartl et al. 2025). Finally, we reflect on how the limitations identified in this review offer a roadmap for future researchers. Despite the progress in integrating big data analytics with behavioral modeling and the transition toward predictive insights, several critical challenges persist, specifically regarding ensuring data quality, addressing algorithmic bias, improving interpretability, and protecting user privacy, especially in applications involving sensitive personal data. To address these issues, future research should prioritize hybrid methodological designs that combine observational big data with controlled experimental settings. Such an approach would allow researchers to bridge the gap between observed digital behaviors and the underlying psychological drivers, providing a more comprehensive explanation of the discrepancies between consumer intention and actual action. Ultimately, fostering interdisciplinary collaboration between data scientists and behavioral researchers will be essential to ensuring the sustainable success of these applications while maintaining a balance between data utility and privacy protection.

Acknowledgments Qiankun Liu: Data collection, visualization, drafting, and revision. Ruigang Wang: Paper revisions. Muhabaiti Pareti: Paper review. Alessandra Castellini: Paper review. Davide Viaggi: Paper review. Maurizio Canavari: Paper guidance and review.

Funding Open access funding provided by Alma Mater Studiorum - Università di Bologna within the CRUI-CARE Agreement. This work was supported by the Fundamental Research Funds of the University of Bologna and the Chinese Scholarship Council.

Declarations The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Abassi, A., A. Arid, and H. Benazza. 2023. Moroccan consumer energy consumption itemsets and inter-appliance associations using machine learning algorithms and data mining techniques.

ASME Journal of Engineering for Sustainable Buildings and Cities. https://doi.org/10.1115/1.4062113.

Abramo, G., C. A. D'Angelo, and F. Viel. 2011. The field-standardized average impact of national research systems compared to world average: The case of Italy. *Scientometrics* 88 (2): 599–615. https://doi.org/10.1007/s11192-011-0406-x.

Adamopoulos, P., A. Ghose, and V. Todri. 2018. The impact of user personality traits on word of mouth: Text-mining social media platforms. *Information Systems Research* 29 (3): 612–640. https://doi.org/10.1287/isre.2017.0768.

Adinolfi, F., M. De Rosa, and F. Trabalzi. 2011. Dedicated and generic marketing strategies: The disconnection between geographical indications and consumer behavior in Italy. *British Food Journal* 113 (3): 419–435. https://doi.org/10.1108/00070701111116473.

Adler, N., A. Brudner, R. Gallotti, F. Privitera, and J. J. Ramasco. 2022. Does big data help answer big questions? The case of airport catchment areas & competition. *Transportation Research Part B: Methodological* 166:444–467. https://doi.org/10.1016/j.trb.2022.10.013.

Alikhani, M., M. P. Moghaddam, and F. Moazzen. 2022. Optimal demand response programs selection using CNN-LSTM algorithm with big data analysis of load curves. *IET Generation, Transmission and Distribution* 16 (24): 4980–5001. https://doi.org/10.1049/gtd2.12650.

Alyoubi, B. A. 2019. The impact of big data on electronic commerce in profit organisations in Saudi Arabia. *Research in World Economy* 10 (4): 106. https://doi.org/10.5430/rwe.v10n4p106.

Anderson, K., O. Burford, and L. Emmerton. 2016. App chronic disease checklist: Protocol to evaluate mobile apps for chronic disease self-management. *JMIR Research Protocols* 5 (4): e6194. https://doi.org/10.2196/resprot.6194.

Babiceanu, R. F., and R. Seker. 2016. Big data and virtualization for manufacturing cyber-physical systems: A survey of the current status and future outlook. *Computers in Industry* 81:128–137. https://doi.org/10.1016/j.compind.2016.02.004.

Barbado, R., O. Araque, and C. A. Iglesias. 2019. A framework for fake review detection in online consumer electronics retailers. *Information Processing & Management* 56 (4): 1234–1244. https://doi.org/10.1016/j.ipm.2019.03.002.

Barbera, G., L. Araujo, and S. Fernandes. 2023. The value of web data scraping: An application to TripAdvisor. *Big Data and Cognitive Computing* 7 (3): 121. https://doi.org/10.3390/bdcc7030121.

Barnes, S. J., M. Diaz, and M. Arnaboldi. 2021. Understanding panic buying during COVID-19: A text analytics approach. *Expert Systems with Applications*. https://doi.org/10.1016/j.eswa.2020.114360.

Bello-Orgaz, G., J. J. Jung, and D. Camacho. 2016. Social big data: Recent achievements and new challenges. *Information Fusion* 28:45–59. https://doi.org/10.1016/j.inffus.2015.08.005.

Biswas, S., A. Fole, N. Khare, and P. Agrawal. 2023. Enhancing correlated big data privacy using differential privacy and machine learning. *Journal of Big Data* 10 (1): 30. https://doi.org/10.1186/s40537-023-00705-8.

Brachten, F., T. Kissmer, and S. Stieglitz. 2021. The acceptance of chatbots in an enterprise context – A survey study. *International Journal of Information Management* 60:102375. https://doi.org/10.1016/j.ijinfomgt.2021.102375.

Brandtner, P., F. Darbanian, T. Falatouri, and C. Udokwu. 2021. Impact of COVID-19 on the customer end of retail supply chains: A big data analysis of consumer satisfaction. *Sustainability (Switzerland)* 13 (3): 1–18. https://doi.org/10.3390/su13031464.

Buettner, R. 2017. Predicting user behavior in electronic markets based on personality-mining in large online social networks. *Electronic Markets* 27 (3): 247–265. https://doi.org/10.1007/s12525-016-0228-z. (**(WOS:000407849300007)**).

Chandra, S., and S. Verma. 2023. Big data and sustainable consumption: A review and research agenda. *Vision: THe Journal of Business Perspective* 27 (1): 11–23. https://doi.org/10.1177/09722629211022520.

Chang, S.-Y., V. Bodolica, H.-H. Hsu, and H.-P. Lu. 2023. What people talk about online and what they intend to do: Related perspectives from text mining and path analysis. *Eurasian Business Review* 13 (4): 931–956. https://doi.org/10.1007/s40821-022-00221-4.

Chaudhary, K., M. Alam, M. S. Al-Rakhami, and A. Gumaei. 2021. Machine learning-based mathematical modelling for prediction of social media consumer behavior using big data analytics. *Journal of Big Data* 8 (1): 73. https://doi.org/10.1186/s40537-021-00466-2.

Chen, C. 2006. CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature. *Journal of the American Society for Information Science and Technology* 57 (3): 359–377. https://doi.org/10.1002/asi.20317.

Chiang, L. L. L., and C.-S. Yang. 2018. Does country-of-origin brand personality generate retail customer lifetime value? A big data analytics approach. *Technological Forecasting and Social Change* 130:177–187. https://doi.org/10.1016/j.techfore.2017.06.034.

Choudhury, N., R. Mukherjee, R. Yadav, Y. Liu, and W. Wang. 2024. Can machine learning approaches predict green purchase intention? -A study from Indian consumer perspective. *Journal of Cleaner Production* 456:142218. https://doi.org/10.1016/j.jclepro.2024.142218.

Chwialkowska, A., W. A. Bhatti, A. Bujac, and S. Abid. 2024. An interplay of the consumption values and green behavior in developed markets: A sustainable development viewpoint. *Sustainable Development* 32 (4): 3771–3785. https://doi.org/10.1002/sd.2867.

Clark, S. D., B. Shute, V. Jenneson, T. Rains, M. Birkin, and M. A. Morris. 2021. Dietary patterns derived from UK supermarket transaction data with nutrient and socioeconomic profiles. *Nutrients*. https://doi.org/10.3390/nu13051481.

Dąbrowska, A. 2011. Consumer behaviour in the market of catering services in selected countries of Central-Eastern Europe. *British Food Journal* 113 (1): 96–108. https://doi.org/10.1108/00070701111097367.

Daviet, R., G. Nave, and J. Wind. 2022. Genetic data: Potential uses and misuses in marketing. *Journal of Marketing* 86 (1): 7–26. https://doi.org/10.1177/0022242920980767.

de Luca, P., G. Pegan, and C. Gonzalo-Penela. 2019. Insights from a Google keywords analysis about Italian wine in the US market. *Micro and Macro Marketing* 1:93–116. https://doi.org/10.1431/93018.

Diem, A., and S. C. Wolter. 2013. The use of bibliometrics to measure research performance in education sciences. *Research in Higher Education* 54 (1): 86–114. https://doi.org/10.1007/s11162-012-9264-5.

Dong, L. 2022. Analysis on influencing factors of consumer trust in e-commerce marketing of green agricultural products based on big data analysis. *Mathematical Problems in Engineering*. https://doi.org/10.1155/2022/8221657.

Du, G., and Y. Lin. 2022. Brand connection and entry in the shopping mall ecological chain: Evidence from consumer behavior big data analysis based on two-sided markets. *Journal of Cleaner Production* 364:132663. https://doi.org/10.1016/j.jclepro.2022.132663.

Dubé, L., A. Labban, J.-C. Moubarac, G. Heslop, Y. Ma, and C. Paquet. 2014. A nutrition/health mindset on commercial big data and drivers of food demand in modern and traditional systems. *Annals of the New York Academy of Sciences* 1331 (1): 278–295. https://doi.org/10.1111/nyas.12595.

Dutta, C. B., and D. K. Das. 2017. What drives consumers' online information search behavior? Evidence from England. *Journal of*

Retailing and Consumer Services 35:36–45. https://doi.org/10.1016/j.jretconser.2016.10.015.

Eastman, J. K., R. Iyer, A. Magnier, F. M. Roka, and B. K. Goodwin. 2025. The role of green consumer values on willingness to buy regional agricultural products. *Journal of Strategic Marketing* 0 (0): 1–23. https://doi.org/10.1080/0965254X.2025.2545868.

Ehsani, F., and M. Hosseini. 2025. Consumer segmentation based on location and timing dimensions using big data from business-to-customer retailing marketplaces. *Big Data* 13 (2): 111–126. https://doi.org/10.1089/big.2022.0307.

El Majdouli, M. A., I. Rbouh, S. Bougrine, B. El Benani, and A. A. El Imrani. 2016. Fireworks algorithm framework for Big Data optimization. *Memetic Computing* 8 (4): 333–347. https://doi.org/10.1007/s12293-016-0201-6.

Erensoy, A., A. Mathrani, A. Schnack, J. Elms, and N. Baghaei. 2024. Consumer behavior in immersive virtual reality retail environments: A systematic literature review using the stimuli-organisms-responses (S-O-r) model. *Journal of Consumer Behaviour* 23 (6): 2781–2811. https://doi.org/10.1002/cb.2374.

Erevelles, S., N. Fukawa, and L. Swayne. 2016. Big data consumer analytics and the transformation of marketing. *Journal of Business Research* 69 (2): 897–904. https://doi.org/10.1016/j.jbusres.2015.07.001.

Evangelin, M., and S. Vasantha. 2022. Mechanism of big data analytics in consumer behavior on online shopping. *International Journal of Early Childhood Special Education* 14 (3): 1938–1942. https://doi.org/10.9756/INT-JECSE/V14I3.226. (**(WOS:000791906700020)**).

Fang, Y., H. Wang, L. Zhao, F. Yu, and C. Wang. 2020. Dynamic knowledge graph based fake-review detection. *Applied Intelligence* 50 (12): 4281–4295. https://doi.org/10.1007/s10489-020-01761-w.

Félix, B. M., E. Tavares, and N. W. F. Cavalcante. 2018. Critical success factors for big data adoption in the virtual retail: Magazine Luiza case study. *Review of Business Management* 20 (1): 112–126. https://doi.org/10.7819/rbgn.v20i1.3627.

Fernandez, K. V. 2019. Critically reviewing literature: A tutorial for new researchers. *Australasian Marketing Journal* 27 (3): 187–196. https://doi.org/10.1016/j.ausmj.2019.05.001.

Filieri, R., Z. Lin, Y. Li, X. Lu, and X. Yang. 2022. Customer emotions in service robot encounters: A hybrid machine-human intelligence approach. *Journal of Service Research* 25 (4): 614–629. https://doi.org/10.1177/10946705221103937. (**(WOS:000805270400001)**).

Filieri, R., F. L. Milone, E. Paolucci, and E. Raguseo. 2023. A big data analysis of COVID-19 impacts on Airbnbs' bookings behavior applying construal level and signaling theories. *International Journal of Hospitality Management* 111:103461. https://doi.org/10.1016/j.ijhm.2023.103461.

Fotopoulou, E., A. Zafeiropoulos, F. Terroso-Sáenz, U. Şimşek, A. González-Vidal, G. Tsiolis, P. Gouvas, P. Liapis, A. Fensel, and A. Skarmeta. 2017. Providing personalized energy management and awareness services for energy efficiency in smart buildings. *Sensors* 17 (9): 2054. https://doi.org/10.3390/s17092054.

Gan, M., and Y. Ouyang. 2022. Study on tourism consumer behavior characteristics based on big data analysis. *Frontiers in Psychology*. https://doi.org/10.3389/fpsyg.2022.876993. (**(WOS:000796905900001)**).

Ghose, A., and V. Todri-Adamopoulos. 2016. Toward a digital attribution model: Measuring the impact of display advertising on online consumer behavior. *MIS Quarterly* 40 (4): 889. https://doi.org/10.25300/MISQ/2016/40.4.05. (**(WOS:000397046800006)**).

Giglio, S., E. Pantano, E. Bilotta, and T. C. Melewar. 2020. Branding luxury hotels: Evidence from the analysis of consumers' "big" visual data on TripAdvisor. *Journal of Business Research* 119:495–501. https://doi.org/10.1016/j.jbusres.2019.10.053.

Golmohammadi, A., T. Havakhor, D. K. Gauri, and J. Comprix. 2021. Complaint publicization in social media. *Journal of Marketing* 85 (6): 1–23. https://doi.org/10.1177/00222429211002183. (**(WOS:000680524100001)**).

Grandhi, B., N. Patwa, and K. Saleem. 2021. Data-driven marketing for growth and profitability. *EuroMed Journal of Business* 16 (4): 381–398. https://doi.org/10.1108/EMJB-09-2018-0054.

Green, M. A., A. W. Watson, J. M. Brunstrom, B. M. Corfe, A. M. Johnstone, E. A. Williams, and E. Stevenson. 2020. Comparing supermarket loyalty card data with traditional diet survey data for understanding how protein is purchased and consumed in older adults for the UK, 2014-16. *Nutrition Journal* 19 (1): 83. https://doi.org/10.1186/s12937-020-00602-3.

Guo, L., and D. Zhang. 2019. EC-structure: Establishing consumption structure through mining E-commerce data to discover consumption upgrade. *Complexity* 2019 (1): 6543590. https://doi.org/10.1155/2019/6543590.

Han, S., and Y. Lee. 2022. Analysis of the impacts of social class and lifestyle on consumption of organic foods in South Korea. *Heliyon*. https://doi.org/10.1016/j.heliyon.2022.e10998.

Hartl, B., E. Penz, and E. Schuessler. 2025. Creating a trusting environment in the sharing economy: Unpacking mechanisms for trust-building used by peer-to-peer carpooling platforms. *Journal of Cleaner Production* 489:144661. https://doi.org/10.1016/j.jclepro.2025.144661.

Haverila, M., K. Haverila, M. O. Gani, and M. Mohiuddin. 2025. The relationship between the quality of big data marketing analytics and marketing agility of firms: The impact of the decision-making role. *Journal of Marketing Analytics* 13 (1): 162–179. https://doi.org/10.1057/s41270-024-00301-6.

Havranek, T., and A. Zeynalov. 2021. Forecasting tourist arrivals: Google Trends meets mixed-frequency data. *Tourism Economics* 27 (1): 129–148. https://doi.org/10.1177/1354816619879584.

Hofacker, C. F., E. C. Malthouse, and F. Sultan. 2016. Big data and consumer behavior: Imminent opportunities. *Journal of Consumer Marketing* 33 (2): 89–97. https://doi.org/10.1108/JCM-04-2015-1399.

Hosseini, M., and A. Ghalamkari. 2018. Analysis social media based brand communities and consumer behavior: A netnographic approach. *International Journal of E-Business Research (IJEBR)* 14 (1): 37–53. https://doi.org/10.4018/IJEBR.2018010103.

Ierkens, J. B., P. Fearnhead, and G. Roberts. 2019. The zig-zag process and super-efficient sampling for Bayesian analysis of big data. *The Annals of Statistics* 47 (3): 1288–1320. https://doi.org/10.1214/18-AOS1715.

Izang, A. A., N. Goga, S. O. Kuyoro, O. D. Alao, A. A. Omotunde, and A. K. Adio. 2019. Scalable data analytics market basket model for transactional data streams. *International Journal of Advanced Computer Science and Applications (IJACSA)* 10 (10): 61–68. https://doi.org/10.14569/IJACSA.2019.0101010.

Jackson, I., and D. Ivanov. 2023. A beautiful shock? Exploring the impact of pandemic shocks on the accuracy of AI forecasting in the beauty care industry. *Transportation Research Part E-Logistics and Transportation Review* 180:103360. https://doi.org/10.1016/j.tre.2023.103360.

Jiménez-Marín, G., P. Sanz-Marcos, I. G. Medina, and P. M. F. Coelho. 2020. How big data collected via point of sale devices in textile stores in Spain resulted in effective online advertising targeting. *International Journal of Interactive Mobile Technologies* 14 (13): 65–77. https://doi.org/10.3991/ijim.v14i13.14359.

Jung, J., S. Yeo, Y. Lee, S. Moon, and D.-J. Lee. 2021. Factors affecting consumers' preferences for electric vehicle: A Korean case. *Research in Transportation Business and Management* 41:100666. https://doi.org/10.1016/j.rtbm.2021.100666.

Kanavos, A., S. A. Iakovou, S. Sioutas, and V. Tampakas. 2018. Large scale product recommendation of supermarket ware based on

customer behaviour analysis. *Big Data and Cognitive Computing* 2 (2): 1–19. https://doi.org/10.3390/bdcc2020011.

Kim, D., J. Woo, J. Shin, J. Lee, and Y. Kim. 2019. Can search engine data improve accuracy of demand forecasting for new products? Evidence from automotive market. *Industrial Management & Data Systems* 119 (5): 1089–1103. https://doi.org/10.1108/IMDS-08-2018-0347.

Kim, Y., M. Srite, and H. Zhao. 2023. Trust antecedents in online reviews across national cultures. *Decision Support Systems* 173:113998. https://doi.org/10.1016/j.dss.2023.113998.

Kiran, P., and S. Vasantha. 2016. Analysing the role of user generated content on consumer purchase intention in the new era of social media and big data. *Indian Journal of Science and Technology* 9 (43): 1–7. https://doi.org/10.17485/ijst/2016/v9i43/104754.

Labrecque, L. I. 2014. Fostering consumer-brand relationships in social media environments: The role of parasocial interaction. *Journal of Interactive Marketing* 28 (2): 134–148. https://doi.org/10.1016/j.intmar.2013.12.003.

Lan, M. 2023. Big data and cloud computing-integrated tourism decision-making in smart logistics technologies. *International Journal of e-Collaboration* 19 (7): 1–20. https://doi.org/10.4018/ijec.316880. (**(WOS:001074838300001)**).

Lee, Y., and D.-Y. Kim. 2020. The decision tree for longer-stay hotel guest: The relationship between hotel booking determinants and geographical distance. *International Journal of Contemporary Hospitality Management* 33 (6): 2264–2282. https://doi.org/10.1108/IJCHM-06-2020-0594.

Lee, C. K. H., Y. K. Tse, M. Zhang, and J. Ma. 2019. Analysing online reviews to investigate customer behaviour in the sharing economy: The case of Airbnb. *Information Technology & People* 33 (3): 945–961. https://doi.org/10.1108/ITP-10-2018-0475. (**(WOS:000544071000006)**).

Lee, J., O. Jung, Y. Lee, O. Kim, and C. Park. 2021. A comparison and interpretation of machine learning algorithm for the prediction of online purchase conversion. *Journal of Theoretical and Applied Electronic Commerce Research* 16 (5): 1472–1491. https://doi.org/10.3390/jtaer16050083.

Li, L. 2023. Analysis of e-commerce customers' shopping behavior based on data mining and machine learning. *Soft Computing*. https://doi.org/10.1007/s00500-023-08903-5.

Li, J., and Q. Hu. 2021. Using culturomics and social media data to characterize wildlife consumption. *Conservation Biology* 35 (2): 452–459. https://doi.org/10.1111/cobi.13703.

Li, R., X. Xu, and S. Dong. 2022a. Construction of precision sales model for luxury market based on machine learning. *Mobile Information Systems*. https://doi.org/10.1155/2022/7920628.

Li, S., S. Zhao, P. Gope, and L. Da Xu. 2022b. Data privacy enhancing in the IoT user/device behavior analytics. *ACM Transactions on Sensor Networks* 19 (2): 32:1-32:13. https://doi.org/10.1145/3534648.

Liu, Y. 2016. Design and implementation of Hadoop-based customer marketing big data processing system. *International Journal of Database Theory and Application* 9 (12): 331–340. https://doi.org/10.14257/ijdta.2016.9.12.29.

Liu, H. 2021. Big data precision marketing and consumer behavior analysis based on fuzzy clustering and PCA model. *Journal of Intelligent and Fuzzy Systems* 40 (4): 6529–6539. https://doi.org/10.3233/JIFS-189491.

Liu, X., and H. Zhao. 2021. Dairy brand loyalty measurement model based on machine learning clustering algorithm. *Journal of Intelligent & Fuzzy Systems* 40 (4): 7601–7612. https://doi.org/10.3233/JIFS-189580.

Liu, H., Y. Huang, Z. Wang, K. Liu, X. Hu, and W. Wang. 2019. Personality or value: A comparative study of psychographic segmentation based on an online review enhanced recommender system. *Applied Sciences*. https://doi.org/10.3390/app9101992. (**(WOS:000473748100027)**).

Liu, C., J. Horng, S. Chou, T. Yu, and M. Lee. 2023. Digital capability, digital learning, and sustainable behaviour among university students in Taiwan: A comparison design of integrated mediation-moderation models. *The International Journal of Management Education*. https://doi.org/10.1016/j.ijme.2023.100835. (**(WOS:001035614200001)**).

Lv, H. 2022. Intelligent e-commerce framework for consumer behavior analysis using big data analytics. *Advances in Data Science and Adaptive Analysis* 14 (03n04): 2250007. https://doi.org/10.1142/S2424922X22500073.

Lv, H. 2023. E-commerce consumer behavior analysis based on big data. *Journal of Computational Methods in Sciences and Engineering* 23 (2): 651–661. https://doi.org/10.3233/JCM-226628.

Lv, H., S. Shi, and D. Gursoy. 2022. A look back and a leap forward: A review and synthesis of big data and artificial intelligence literature in hospitality and tourism. *Journal of Hospitality Marketing & Management* 31 (2): 145–175. https://doi.org/10.1080/19368623.2021.1937434.

Maeng, Y., C. C. Lee, and H. Yun. 2023. Understanding antecedents that affect customer evaluations of head-mounted display VR devices through text mining and deep neural network. *Journal of Theoretical and Applied Electronic Commerce Research* 18 (3): 1238–1256. https://doi.org/10.3390/jtaer18030063.

Makov, T., and C. Fitzpatrick. 2021. Is repairability enough? Big data insights into smartphone obsolescence and consumer interest in repair. *Journal of Cleaner Production* 313:127561. https://doi.org/10.1016/j.jclepro.2021.127561.

Mariani, M., and M. Predvoditeleva. 2019. How do online reviewers' cultural traits and perceived experience influence hotel online ratings?: An empirical analysis of the Muscovite hotel sector. *International Journal of Contemporary Hospitality Management* 31 (12): 4543–4573. https://doi.org/10.1108/IJCHM-11-2018-0927.

Mariani, M. M., M. Borghi, and S. Kazakov. 2019. The role of language in the online evaluation of hospitality service encounters: An empirical study. *International Journal of Hospitality Management* 78:50–58. https://doi.org/10.1016/j.ijhm.2018.11.012.

Martens, D., F. Provost, J. Clark, and E. J. de Fortuny. 2016. Mining massive fine-grained behavior data to improve predictive analytics. *MIS Quarterly Management Information Systems* 40 (4): 869–888. https://doi.org/10.25300/MISQ/2016/40.4.04.

Martin, K. D., A. Borah, and R. W. Palmatier. 2017. Data privacy: Effects on customer and firm performance. *Journal of Marketing* 81 (1): 36–58. https://doi.org/10.1509/jm.15.0497.

Mayr, P., and A. Scharnhorst. 2015. Scientometrics and information retrieval: Weak-links revitalized. *Scientometrics* 102 (3): 2193–2199. https://doi.org/10.1007/s11192-014-1484-3.

Mendieta-Aragón, A., and T. Garín-Muñoz. 2023. Consumer behaviour in e-tourism: Exploring new applications of machine learning in tourism studies. *Investigaciones Turísticas* 26:350–374. https://doi.org/10.14198/INTURI.24629.

Moher, D., A. Liberati, J. Tetzlaff, and D. G. Altman. 2009. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *Annals of Internal Medicine* 151 (4): 264–269. https://doi.org/10.7326/0003-4819-151-4-200908180-00135.

Moher, D., L. Shamseer, M. Clarke, D. Ghersi, A. Liberati, M. Petticrew, P. Shekelle, and L. A. Stewart. 2015. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Systematic Reviews* 4 (1): 1. https://doi.org/10.1186/2046-4053-4-1.

Molitor, D., S. Daurer, M. Spann, and P. Manchanda. 2023. Digitizing local search: An empirical analysis of mobile search behavior in offline shopping. *Decision Support Systems* 174:114018. https://doi.org/10.1016/j.dss.2023.114018.

Morris, M. A., E. L. Wilkins, M. Galazoula, S. D. Clark, and M. Birkin. 2020. Assessing diet in a university student population: A longitudinal food card transaction data approach. *British Journal of Nutrition* 123 (12): 1406–1414. https://doi.org/10.1017/S0007114520000823.

Moschis, G. P. 2012. Consumer behavior in later life: Current knowledge, issues, and new directions for research. *Psychology & Marketing* 29 (2): 57–75. https://doi.org/10.1002/mar.20504.

Mu, W. 2019. A big data-based prediction model for purchase decisions of consumers on cross-border e-commerce platforms. *Journal Européen des Systèmes Automatisés* 52 (4): 363–368. https://doi.org/10.18280/jesa.520405.

Naraine, M. L., N. O'Reilly, N. Levallet, and L. Wanless. 2020. If you build it, will they log on? Wi–fi usage and behavior while attending National Basketball Association games. *Sport, Business and Management* 10 (2): 207–226. https://doi.org/10.1108/SBM-02-2019-0016.

Nie, Y., and X. Han. 2019. Research on consumers' protection in advantageous operation of big data brokers. *Cluster Computing* 22 (4): 8387–8400. https://doi.org/10.1007/s10586-018-1837-4.

Nilashi, M., R. A. Abumalloh, A. Alghamdi, B. Minaei-Bidgoli, A. A. Alsulami, M. Thanoon, S. Asadi, and S. Samad. 2021. What is the impact of service quality on customers' satisfaction during COVID-19 outbreak? New findings from online reviews analysis. *Telematics and Informatics*. https://doi.org/10.1016/j.tele.2021.101693.

Oh, J., T. P. Connerton, and H.-J. Kim. 2019. The rediscovery of brand experience dimensions with big data analysis: Building for a sustainable brand. *Sustainability* 11 (19): 5438. https://doi.org/10.3390/su11195438.

Oprea, S.-V., and A. Bâra. 2020. Setting the time-of-use tariff rates with NoSQL and machine learning to a sustainable environment. *IEEE Access* 8:25521–25530. https://doi.org/10.1109/ACCESS.2020.2969728.

Oprea, S.-V., A. Bâra, B. G. Tudorică, M. I. Călinoiu, and M. A. Botezatu. 2021. Insights into demand-side management with big data analytics in electricity consumers' behaviour. *Computers & Electrical Engineering* 89:106902. https://doi.org/10.1016/j.compeleceng.2020.106902.

Ozturkcan, S., N. Kasap, A. Tanaltay, and M. Ozdinc. 2019. Analysis of tweets about football: 2013 and 2018 leagues in Turkey. *Behaviour & Information Technology* 38 (9): 887–899. https://doi.org/10.1080/0144929X.2019.1583284. (**(WOS:000481837100003).**).

Page, M. J., J. E. McKenzie, P. M. Bossuyt, I. Boutron, T. C. Hoffmann, C. D. Mulrow, L. Shamseer, J. M. Tetzlaff, E. A. Akl, S. E. Brennan, R. Chou, J. Glanville, J. M. Grimshaw, A. Hróbjartsson, M. M. Lalu, T. Li, E. W. Loder, E. Mayo-Wilson, S. McDonald, and D. Moher. 2021. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*. https://doi.org/10.1136/bmj.n71.

Pang, W., J. Ko, S. Kim, and E. Ko. 2022. Impact of COVID-19 pandemic upon fashion consumer behavior: Focus on mass and luxury products. *Asia Pacific Journal of Marketing and Logistics* 34 (10): 2149–2164. https://doi.org/10.1108/APJML-03-2021-0189. (**(WOS:000729453900001).**).

Pantano, E., S. Giglio, and C. Dennis. 2019. Making sense of consumers' tweets: Sentiment outcomes for fast fashion retailers through big data analytics. *International Journal of Retail and Distribution Management* 47 (9): 915–927. https://doi.org/10.1108/IJRDM-07-2018-0127.

Petrescu, M., and A. S. Krishen. 2024. Marketing analytics in 2024 conferences: AI and data-driven decision-making. *Journal of Marketing Analytics* 12 (4): 743–745. https://doi.org/10.1057/s41270-024-00357-4.

Petrescu, M., and A. S. Krishen. 2025. Marketing analytics research for the better good: Thoughts and ideas. *Journal of Marketing Analytics* 13 (1): 1–3. https://doi.org/10.1057/s41270-025-00386-7.

Pham, T. H., T. N. Nguyen, T. T. H. Phan, and N. T. Nguyen. 2019. Evaluating the purchase behaviour of organic food by young consumers in an emerging market economy. *Journal of Strategic Marketing* 27 (6): 540–556. https://doi.org/10.1080/0965254X.2018.1447984.

Pindado, E., and R. Barrena. 2021. Using Twitter to explore consumers' sentiments and their social representations towards new food trends. *British Food Journal* 123 (3): 1060–1082. https://doi.org/10.1108/BFJ-03-2020-0192.

Praveen Kumar, Y., R. Suguna, S. Ghosh, and K. Neha. 2019. Semantic web mining in retail management system using ANN. *International Journal of Recent Technology and Engineering* 8 (2 Special Issue 11): 3547–3554. https://doi.org/10.35940/ijrte.B1439.0982S1119.

Prentice, C., J. Chen, and B. Stantic. 2020. Timed intervention in COVID-19 and panic buying. *Journal of Retailing and Consumer Services*. https://doi.org/10.1016/j.jretconser.2020.102203.

Rajan, P. 2024. Integrating IoT analytics into marketing decision making: A smart data-driven approach. *International Journal of Data Informatics and Intelligent Computing* 3 (1): 1. https://doi.org/10.59461/ijdiic.v3i1.92.

Ramos, C. M. Q., D. J. Martins, F. Serra, R. Lam, P. J. S. Cardoso, M. B. Correia, and J. M. F. Rodrigues. 2017. Framework for a hospitality big data warehouse: The implementation of an efficient hospitality business intelligence system. *International Journal of Information Systems in the Service Sector (Ijisss)* 9 (2): 27–45. https://doi.org/10.4018/IJISSS.2017040102.

Rendón, E., R. Alejo, C. Castorena, F. J. Isidro-Ortega, and E. E. Granda-Gutiérrez. 2020. Data sampling methods to deal with the big data multi-class imbalance problem. *Applied Sciences* 10 (4) : 1276. https://doi.org/10.3390/app10041276.

Ryu, G.-A., A. Nasridinov, H. Rah, and K.-H. Yoo. 2020. Forecasts of the amount purchase pork meat by using structured and unstructured big data. *Agriculture*. https://doi.org/10.3390/agriculture10010021.

Sakas, D., I. Kamperos, and P. Reklitis. 2021. Estimating risk perception effects on courier companies' online customer behavior during a crisis, using crowdsourced data. *Sustainability*. https://doi.org/10.3390/su132212725. (**(WOS:000727838400001)**).

Sarasquete, N. 2017. A common data representation model for customer behavior tracking. *Revista Icono 14-Revista Cientifica De Comunicacion y Tecnologias* 15 (2): 55–91. https://doi.org/10.7195/ri14.1078v15i2.1078. (**(WOS:000419280800013)**).

Sarkis-Onofre, R., F. Catalá-López, E. Aromataris, and C. Lockwood. 2021. How to properly use the PRISMA Statement. *Systematic Reviews* 10 (1) : 117. https://doi.org/10.1186/s13643-021-01671-z.

Seo, D., and Y. Yoo. 2023. Improving shopping mall revenue by real-time customized digital coupon issuance. *IEEE Access* 11:7924–7932. https://doi.org/10.1109/ACCESS.2023.3239425.

Serrano, L., A. Ariza-Montes, M. Nader, A. Sianes, and R. Law. 2021. Exploring preferences and sustainable attitudes of Airbnb green users in the review comments and ratings: A text mining approach. *Journal of Sustainable Tourism* 29 (7): 1134–1152. https://doi.org/10.1080/09669582.2020.1838529.

Shehawy, Y. M., and S. M. F. Ali Khan. 2024. Consumer readiness for green consumption: The role of green awareness as a moderator of the relationship between green attitudes and purchase intentions. *Journal of Retailing and Consumer Services* 78 : 103739. https://doi.org/10.1016/j.jretconser.2024.103739.

Shibuya, Y., and H. Tanaka. 2018. A statistical analysis between consumer behavior and a social network service: A case study of used-car demand following the Great East Japan Earthquake and Tsunami of 2011. *The Review of Socionetwork Strategies*

12 (2): 205–236. https://doi.org/10.1007/s12626-018-0025-6. (**(WOS:000451237900006)**).

Shokouhyar, S., A. Dehkhodaei, and B. Amiri. 2021. Toward customer-centric mobile phone reverse logistics: Using the DEMATEL approach and social media data. *Kybernetes* 51 (11): 3236–3279. https://doi.org/10.1108/K-11-2020-0831.

Silva, E. S., H. Hassani, D. Ø. Madsen, and L. Gee. 2019. Googling fashion: Forecasting fashion consumer behaviour using Google trends. *Social Sciences* 8 (4): 111. https://doi.org/10.3390/socsci8040111.

Silva, E. S., H. Hassani, and D. Ø. Madsen. 2020. Big data in fashion: Transforming the retail sector. *Journal of Business Strategy* 41 (4): 21–27. https://doi.org/10.1108/JBS-04-2019-0062.

Šimović, P. P., C. Y. T. Chen, and E. W. Sun. 2023. Classifying the variety of customers' online engagement for churn prediction with a mixed-penalty logistic regression. *Computational Economics* 61 (1): 451–485. https://doi.org/10.1007/s10614-022-10275-1.

Singh, A., and A. Glińska-Neweś. 2022. Modeling the public attitude towards organic foods: A big data and text mining approach. *Journal of Big Data* 9 (1) : 2. https://doi.org/10.1186/s40537-021-00551-6. (**Scopus (WOS:000739967500001)**).

Singh, S., and A. Yassine. 2019. Mining energy consumption behavior patterns for households in smart grid. *IEEE Transactions on Emerging Topics in Computing* 7 (3): 404–419. https://doi.org/10.1109/TETC.2017.2692098.

Singh, J. P., S. Irani, N. P. Rana, Y. K. Dwivedi, S. Saumya, and P. K. Roy. 2017. Predicting the 'helpfulness' of online consumer reviews. *Journal of Business Research* 70:346–355. https://doi.org/10.1016/j.jbusres.2016.08.008.

Song, S., X. Yao, Y. C. Zhao, and Z. Ba. 2024. Get inspired and pay for the goods: An investigation of customer inspiration and purchase intention in livestream shopping. *Journal of Retailing and Consumer Services* 78:103750. https://doi.org/10.1016/j.jretconser.2024.103750.

Soria-Comas, J., and J. Domingo-Ferrer. 2016. Big data privacy: Challenges to privacy principles and models. *Data Science and Engineering* 1 (1): 21–28. https://doi.org/10.1007/s41019-015-0001-x.

Sujood & Pancy. 2024. Travelling with open eyes! A study to measure consumers' intention towards experiencing immersive technologies at tourism destinations by using an integrated model of TPB, TAM captured through the lens of S-O-R. *International Journal of Contemporary Hospitality Management* 36 (11): 3906–3929. https://doi.org/10.1108/IJCHM-11-2023-1771.

Sun, M., and J. Zhao. 2022. Behavioral patterns beyond posting negative reviews online: An empirical view. *Journal of Theoretical and Applied Electronic Commerce Research* 17 (3): 949–983. https://doi.org/10.3390/jtaer17030049.

Theodorakopoulos, L., and A. Theodoropoulou. 2024. Leveraging big data analytics for understanding consumer behavior in digital marketing: A systematic review. *Human Behavior and Emerging Technologies* 2024 (1): 3641502. https://doi.org/10.1155/2024/3641502.

Tian, Y. 2022. An effective model for consumer need prediction using big data analytics. *Journal of Interconnection Networks* 22 (Supp02): 2143008. https://doi.org/10.1142/S0219265921430088. (**Scopus (WOS:000867593900012)**).

Tian, J., Y. Zhang, and C. Zhang. 2018. Predicting consumer variety-seeking through weather data analytics. *Electronic Commerce Research and Applications* 28:194–207. https://doi.org/10.1016/j.elerap.2018.02.001.

Tupikovskaja-Omovie, Z., and D. Tyler. 2021. Eye tracking technology to audit Google Analytics: Analysing digital consumer shopping journey in fashion m-retail. *International Journal of Information Management* 59:102294. https://doi.org/10.1016/j.ijinfomgt.2020.102294.

Udugama, I. A., W. Kelton, and C. Bayer. 2023. Digital twins in food processing: A conceptual approach to developing multi-layer digital models. *Digital Chemical Engineering*. https://doi.org/10.1016/j.dche.2023.100087.

Upadhyay, U., A. Kumar, G. Sharma, S. Sharma, V. Arya, P. K. Panigrahi, and B. B. Gupta. 2024. A systematic data-driven approach for targeted marketing in enterprise information system. *Enterprise Information Systems* 18 (8): 2356770. https://doi.org/10.1080/17517575.2024.2356770.

Urkup, C., B. Bozkaya, and F. Sibel Salman. 2018. Customer mobility signatures and financial indicators as predictors in product recommendation. *PLoS ONE* 13 (7): 0201197. https://doi.org/10.1371/journal.pone.0201197.

Ushakova, A., and S. Jankin Mikhaylov. 2020. Big data to the rescue? Challenges in analysing granular household electricity consumption in the United Kingdom. *Energy Research & Social Science* 64:101428. https://doi.org/10.1016/j.erss.2020.101428.

van Eck, N. J., and L. Waltman. 2010. Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics* 84 (2): 523–538. https://doi.org/10.1007/s11192-009-0146-3.

Venkatraman, S. 2017. A proposed business intelligent framework for recommender systems. *Informatics* 4 (4): 40. https://doi.org/10.3390/informatics4040040. (**Scopus (WOS:000423673600008)**).

Vepsäläinen, H., J. Nevalainen, S. Kinnunen, S. T. Itkonen, J. Meinilä, S. Männistö, L. Uusitalo, M. Fogelholm, and M. Erkkola. 2022. Do we eat what we buy? Relative validity of grocery purchase data as an indicator of food consumption in the LoCard study. *British Journal of Nutrition* 128 (9): 1780–1788. https://doi.org/10.1017/S0007114521004177.

Vieira, V. A. 2013. Stimuli–Organism-Response framework: A meta-analytic review in the store environment. *Journal of Business Research* 66 (9): 1420–1426. https://doi.org/10.1016/j.jbusres.2012.05.009.

Vinkóczi, T., I. M. Végvári, M. Németh, and J. P. Koltai. 2026. From scroll to sale: Predicting of customer engagement in impulse buying cycle on TikTok through the S-O-R model. *Acta Psychologica* 262 : 106006. https://doi.org/10.1016/j.actpsy.2025.106006.

Volkova, E., and G. Karpushkin. 2023. Marketing communications in luxury fashion retail in the era of big data. *Electronic Commerce Research*. https://doi.org/10.1007/s10660-023-09732-2.

Wang, Y. 2022. Big data mining method of marketing management based on deep trust network model. *International Journal of Circuits, Systems and Signal Processing* 16:578–584. https://doi.org/10.46300/9106.2022.16.72.

Wang, J., and Y. Zhang. 2021. Using cloud computing platform of 6G IoT in e-commerce personalized recommendation. *International Journal of System Assurance Engineering and Management* 12 (4): 654–666. https://doi.org/10.1007/s13198-021-01059-1. (**(WOS:000613057500001)**).

Wang, B., C. Shen, Y. Cai, L. Dai, S. Gai, and D. Liu. 2023. Consumer culture in traditional food market: The influence of Chinese consumers to the cultural construction of Chinese barbecue. *Food Control* 143:109311. https://doi.org/10.1016/j.foodcont.2022.109311.

Wei, W., C. B. Sivaparthipan, and P. M. Kumar. 2022. Online shopping behavior analysis for smart business using big data analytics and blockchain security. *International Journal of Modeling, Simulation, and Scientific Computing* 13 (04): 2250053. https://doi.org/10.1142/S1793962322500532.

Xiang, Z., Z. Schwartz, J. H. Gerdes, and M. Uysal. 2015. What can big data and text analytics tell us about hotel guest experience and satisfaction? *International Journal of Hospitality Management* 44:120–130. https://doi.org/10.1016/j.ijhm.2014.10.013. (**Scopus (WOS:000347365100013)**).

Xiao, B., and G. Piao. 2022. Analysis of influencing factors and enterprise strategy of online consumer behavior decision based on

association rules and mobile computing. *Wireless Communications and Mobile Computing*. https://doi.org/10.1155/2022/6849017.

Xie, T. 2023. Artificial intelligence and automatic recognition application in B2C e-commerce platform consumer behavior recognition. *Soft Computing* 27 (11): 7627–7637. https://doi.org/10.1007/s00500-023-08147-3.

Xu, M., and P. Chen. 2023. Big data analysis based on the correlation between live-streaming with goods, perceived value and consumer repurchase. *Applied Mathematics and Nonlinear Sciences*. https://doi.org/10.2478/amns.2023.1.00134.

Xue, L. 2023. Cluster analysis-based big data mining method of e-commerce consumer behavior. *International Journal of Web Based Communities* 19 (1): 53–63. https://doi.org/10.1504/IJWBC.2023.10048319.

Yan, Y., C. Huang, Q. Wang, and B. Hu. 2020. Data mining of customer choice behavior in internet of things within relationship network. *International Journal of Information Management* 50:566–574. https://doi.org/10.1016/j.ijinfomgt.2018.11.013.

Yang, J., D. Min, and J. Kim. 2020. The use of big data and its effects in a diffusion forecasting model for Korean reverse mortgage subscribers. *Sustainability* 12 (3): 979. https://doi.org/10.3390/su12030979.

Ye, D., B. Muthu, and P. Kumar. 2022a. Identifying buying patterns from consumer purchase history using big data and cloud computing. *International Journal of Distributed Systems and Technologies* 13 (7): 1. https://doi.org/10.4018/IJDST.307957.

Ye, Y., X. Lu, and T. Lu. 2022b. Examining the spillover effect of sustainable consumption on microloan repayment: A big data-based research. *Information and Management*. https://doi.org/10.1016/j.im.2020.103288.

Zhang, W., and M. Wang. 2021. An improved deep forest model for prediction of e-commerce consumers' repurchase behavior. *PLoS ONE*. https://doi.org/10.1371/journal.pone.0255906. (**(WOS:000707078200010)**).

Zhang, L., J. Priestley, J. Demaio, S. Ni, and X. Tian. 2021. Measuring customer similarity and identifying cross-selling products by community detection. *Big Data* 9 (2): 132–143. https://doi.org/10.1089/big.2020.0044.

Zhao, X., T. Lu, and Y. Dai. 2021. Individual driver crash risk classification based on IoV data and offline consumer behavior data. *Mobile Information Systems*. https://doi.org/10.1155/2021/6784026.

Zhou, F., M. K. Lim, Y. He, and S. Pratap. 2019. What attracts vehicle consumers' buying: A Saaty scale-based VIKOR (SSC-VIKOR) approach from after-sales textual perspective? *Industrial Management & Data Systems* 120 (1): 57–78. https://doi.org/10.1108/IMDS-01-2019-0034.

Zhukovskiy, Y. L., M. S. Kovalchuk, D. E. Batueva, and N. D. Senchilo. 2021. Development of an algorithm for regulating the load schedule of educational institutions based on the forecast of electric consumption within the framework of application of the demand response. *Sustainability*. https://doi.org/10.3390/su132413801.

Qiankun Liu Qiankun Liu is a PhD candidate at the University of Bologna, Department of Agricultural and Food Sciences (DISTAL). His research focuses on food economics, consumer behavior, and machine learning. He has published papers analyzing consumer behavior.

Ruigang Wang : Dr. Ruigang Wang is currently working at the Institute of Food and Nutrition Development, Ministry of Agriculture and Rural Affairs of China. His research focuses on food loss and waste reduction and green consumption. He has participated in a number of major research projects, including the National Key R&D Program, the National Social Science Fund, and strategic consulting projects of the Chinese Academy of Engineering. As the first author, he has published multiple SCI and SSCI papers in journals such as Resources, Conservation and Recycling, and Sustainable Production and Consumption.

Muhabaiti Pareti : Dr. Muhabaiti Pareti is an Associate Professor at Xinjiang Agricultural University and a visiting professor at the University of Bologna. His research interests focus on food economics and meat science. He has published multiple SCI and SSCI papers relevant with food economics and consumer behavior.

Alessandra Castellini : Dr. Alessandra Castellini is an Associate Professor at the University of Bologna, Department of Agricultural and Food Sciences (DISTAL), specializing in agricultural and food economics, consumer preferences, value chain management, and sustainability. She has led national and international research projects, published 100+papers, teaches agribusiness and international markets, and is active in professional associations including AAWE, SIDEA, SIEA, and EAAE.

Davide Viaggi : Dr. Davide Viaggi is a Full Professor at the University of Bologna, Department of Agricultural and Food Sciences (DISTAL). His research focuses on the Common Agricultural Policy, bioeconomy, water resource management, agro-environmental measures, and farm management. He has coordinated several EU-funded projects, including CAP-IRE, CLAIM, PROVIDE, CONSOLE, and BIOBEC, and has contributed to over 300 scientific publications and 10 monographs, such as The Bioeconomy: Delivering Sustainable Green Growth (CABI, 2018). He is also editor of multiple international volumes and active in major academic associations in agricultural and environmental economics.

Maurizio Canavari : Dr. Maurizio Canavari is a Full Professor at the Department of Agricultural and Food Sciences, University of Bologna. His research focuses on agri-food marketing, consumer behavior, innovation adoption, quality and organic foods, e-commerce, and the wine industry. He is a member of professional associations including IAAE, EAAE, SIDEA, and SIEA. He serves as Editor-in-Chief of Economia agro-alimentare/Food Economy and is on the editorial boards of several international journals in food and agribusiness marketing.