Sustainable Digital Marketing under Big Data: An AI Random Forest Model Approach

Keyan Jin[1], Zoe Ziqi Zhong[2,*], Elena Yifei Zhao[3]

[1]Department of Quantitative Methods in Economics and Business, University of Granada, 51001 Ceuta, Spain；kjin@correo.ugr.es

[2]Department of Management, London School of Economics and Political Science, London WC2A 2AE, United Kingdom; z.zhong6@lse.ac.uk

[3]Annenberg school for communication and journalism, University of Southern California, Los Angeles, CA 90089, United States; zhaoyife@usc.edu

***Correspondence:** z.zhong6@lse.ac.uk

Abstract: Digital marketing refers to the process of promoting, selling, and delivering products or services through online platforms and channels using the internet and electronic devices in a digital environment. Its aim is to attract and engage target audiences through various strategies and methods, driving brand promotion and sales growth. The primary objective of this scholarly study is to seamlessly integrate advanced big data analytics and artificial intelligence (AI) technology into the realm of digital marketing, thereby fostering the progression and optimization of sustainable digital marketing practices. Firstly, the characteristics and applications of big data involving vast, diverse, and complex datasets are analyzed. Understanding their attributes and scope of application is essential. Subsequently, a comprehensive investigation into AI-driven learning mechanisms is conducted, culminating in the development of an AI random forest model (RFM) tailored for sustainable digital marketing. Subsequent to this, leveraging a real-world case study involving enterprise X, fundamental customer data is collected and subjected to meticulous analysis. The RFM model, ingeniously crafted in this study, is then deployed to prognosticate the anticipated count of prospective customers for said enterprise. The empirical findings spotlight a pronounced prevalence of university-affiliated individuals across diverse age cohorts. In terms of occupational distribution within the customer base, the categories of workers and educators emerge as dominant, constituting 41% and 31% of the demographic, respectively. Furthermore, the price distribution of patrons exhibits a skewed pattern, whereby the price bracket of 0-150 encompasses 17% of the population, while the range of 150-300 captures a notable 52%. These delineated price bands collectively constitute a substantial proportion, whereas the range exceeding 450 embodies a minority, accounting for less than 20%. Notably, the RFM model devised in this scholarly endeavor demonstrates a remarkable proficiency in accurately projecting forthcoming passenger volumes over a seven-day horizon, significantly surpassing the predictive capability of logistic regression. Evidently, the AI-driven RFM model proffered herein excels in the precise anticipation of target customer counts, thereby furnishing a pragmatic foundation for the intelligent evolution of sustainable digital marketing strategies.

Keywords: big data; artificial intelligence; random forest model; social media; sustainable digital marketing

1. Introduction

As society evolves and technology rapidly advances, the field of marketing has transitioned through three historical periods: from traditional marketing to online marketing and further to digital marketing. Technological transformations accompany these stages of development and essentially involve category competition and path competition, each closely linked to consumer purchasing experiences (Diez-Martin et al., 2019; Low et al., 2020). In light of the rapid and pervasive advancement within the realm of information technology and the mobile internet sector, an escalating number of consumers have turned to the internet as a decisive factor in their consumption deliberations, thereby wielding an increasingly profound influence on the operational landscape of diverse industries (Saura et al., 2020). The 2022~2023 Digital Marketing Trends Report provided a comprehensive analysis of the latest trends and future predictions in the field of digital marketing, with a particular emphasis on the application of big data and artificial intelligence (AI) (Eleonora, Tetiana, Ihor, Nataliia & Anastasiia, 2023). Propelled by the pervasive phenomenon of "Internet Plus," a substantial convergence between the traditional sectors and the mobile internet domain is manifest (Kuzior et al., 2021). In the advent of the "Big Data" era, a transformative paradigm is unfolding whereby a growing array of decisions within the prominent spheres of economic and commercial activities are progressively predicated upon rigorous data-driven analysis and exploration. This evolution serves to imbue decision-making processes with scientific rigor and interpretability, effectively supplanting conventional reliance on antecedent experience and intuition (Yousaf et al., 2021). The 2023 Big Data Application Status Research Report in Business Marketing extensively investigated the application of big data in the business sector, with a special focus on practical cases and success stories in the digital marketing field (Liu, Wan & Yu, 2023). Furthermore, the landscape of cloud computing is presently poised within an epochal "golden opportunity era." The continual proliferation of cloud computing-related products and solutions is poised to catalyze an accelerated global developmental trajectory (Dash & Chakraborty, 2021). In this context, big data is defined as a massive, diverse, and intricate collection of data that typically requires highly parallelized computing methods and techniques for processing and analysis (Demirbaga & Aujla, 2022; Nagendra et al., 2022). Big data technology encompasses areas such as data storage, data mining, machine learning, etc., aiding in extracting meaningful insights from vast datasets and supporting digital marketing decisions.

In the field of technology and engineering management, the close integration of digital marketing and big data analytics has had profound implications for enterprise innovation and sustainable development. Paniagua and Delsing explored the impact of the Internet of Things on industrial structure and business models (Paniagua & Delsing, 2020). They emphasized how digital technology created new business value through connecting devices and analyzing data, profoundly affecting enterprise innovation and management. Samara et al. focused on the influence of big data and AI on corporate strategy (Samara, Magnisalis & Peristeras, 2020). They highlighted the crucial role of data-driven decision-making in enterprise management and market competition, especially in the importance of the digital era. Behl discussed how big data affected corporate competitiveness (Behl, 2022). Scholars pointed out that big data analysis allowed businesses to understand market trends, improve operational efficiency, and achieve innovation. Peng et al. delved into the enhancement of enterprise productivity through digital technology (Peng & Tao, 2022). They discussed how digital technology changed business models, particularly its impact on digital marketing and market positioning. Munsch focused on the relationship between social media and digital marketing (Munsch, 2021). They emphasized the importance of social media in shaping brand image and building customer relationships, providing technology and engineering companies with opportunities to stand out in the digital age. These scholars' research offers a deep understanding of digitization, big data, and digital marketing in the field of technology and engineering management, providing valuable insights for businesses seeking innovative and sustainable development paths in a competitive market. Digital marketing, through online platforms and social media, provides technology and engineering companies with precise opportunities for market positioning and product development. The 2023 Innovative Survey Report on AI in Digital Marketing delved into AI's innovation and disruptive changes in digital marketing, providing robust support for future digital marketing strategies (Ziakis & Vlachopoulou, 2023). The 2022 Sustainable Digital Marketing Practices Report focused on sustainable practices in the digital marketing field and the role of big data and AI in promoting green digital marketing (Thangam & Chavadi, 2023). Strategies such as social media engagement, content marketing, and search engine optimization enhance the visibility and reputation of companies in highly competitive markets, helping to establish closer customer relationships. This provides real-time data for enterprises, enabling managers to make more informed decisions and drive market strategies and resource allocation. Simultaneously, the application of big data analytics in the technology sector brings crucial technological advantages. By monitoring and analyzing device data, enterprises implement predictive maintenance, reduce downtime, and improve production efficiency. In supply chain management, big data analytics optimizes demand forecasting and inventory management, reducing costs and enhancing delivery efficiency. Big data analytics contributes to quality control and production optimization in engineering management. The 2023 Digital Trends Analysis Report in the field of Technology and Engineering Management, published by the International Association of Technology and Engineering Management, investigated the application trends of digital technology in the domain of technology and engineering management. It provides strategic insights for businesses in formulating digital marketing strategies (Rakic, Medic, Leoste, Vuckovic & Marjanovic, 2023). By real-time monitoring data during the production process, companies can promptly identify issues and improve product quality. The combination of these two, in the context of sustainable digital marketing, creates a broad prospect. Digital marketing, through big data analytics, achieves more precise target marketing, allowing companies to tailor marketing strategies to meet the needs of target customers. Furthermore, the integration of digital marketing and big data analytics contributes to more effective resource management, reducing resource costs associated with traditional marketing while concurrently fostering environmental sustainability through digital platforms. Sustainable digital marketing not only emphasizes economic benefits but also underscores corporate social responsibility. By using digital channels to communicate social responsibility activities, companies enhance their image and meet consumers' demand for a sense of social responsibility. In conclusion, the mutual integration of digital marketing and big data analytics creates a more intelligent, efficient, and sustainable development path in the field of technology and engineering management.

Digital marketing, operating within a digital environment, utilizes the internet and electronic devices through online platforms and channels to promote, sell, and deliver products or services. It encompasses various strategies and methods, including advertising, promotion, social media marketing, email marketing, etc., with the aim of attracting and engaging target audiences and achieving brand promotion and sales growth (Amiri et al., 2023; Khare et al., 2023; Shankar et al., 2022). The bedrock upon which digital marketing is constructed resides within network marketing, characterized by its expeditiousness, heightened efficacy, and cost-effectiveness. In the scholarly discourse, Lu et al. advanced a paradigm centered on a user interest model grounded in domain ontology, constituting a multi-agent predictive framework for consumer behavior in the domain of e-commerce. Notably, this approach surmounts the constraints inherent in conventional methodologies for prognosticating consumer behavior (Lu et al., 2021). This model encompasses a constellation of agents, including a consumer behavior prediction agent, a user interest model management agent, a monitoring agent, and a source data monitoring agent. By means of inter-agent interactions, these entities collectively contribute to the forecast of consumer behavior. Similarly, Khanbaghi and Zecevic propounded a novel methodology for anticipating individual consumption behavior, predicated upon quadratic clustering and the theory of hidden Markov chains. The process of clustering consumption behavior introduces a quadratic clustering penalty term, concomitant with applying the hidden Markov chain framework to infer state transitions from consumption hierarchies to sequential patterns (Khanbaghi & Zecevic, 2020). In the current era of big data, digital marketing still faces the challenge of deeply integrating big data and AI. Based on the above analysis, current research primarily focuses on predicting digital consumer behavior through machine learning. However, deeper integration for achieving sustainable digital marketing is still lacking.

The logical position of this study problem statement is rooted in recognition of the crucial roles played by digital marketing and big data analytics in the field of technology and engineering management, as well as a profound understanding of the prospects for sustainable digital marketing. Firstly, the separate applications of digital marketing and big data analytics in technology and engineering management have demonstrated significant benefits. Digital marketing, through strategies such as social media and search engine optimization, provides opportunities for market positioning and brand promotion, while big data analytics plays a crucial role in predictive maintenance, supply chain optimization, and more. However, the combination of these two can create more powerful synergies, bringing about a more comprehensive and profound transformation in technology and engineering management. Secondly, the understanding of the prospects for sustainable digital marketing further reinforces the logical position of the problem statement. In today's society, businesses need to focus not only on economic benefits but also on upholding sustainable development principles in their digital marketing practices. The combination of digital marketing and big data analytics offers opportunities for more precise target marketing, resource optimization, and the dissemination of social responsibility, laying the foundation for achieving sustainable digital marketing. In-depth research into the integrated application of digital marketing and big data analytics can reveal new business models and management strategies, helping businesses better adapt to and lead sustainable development trends. Lastly, as global market competition intensifies, businesses need more intelligent and efficient methods to drive innovation and enhance productivity. The combination of digital marketing and big data analytics provides opportunities for businesses to achieve these goals. By revealing the potential value of this integrated application, this study aims to provide innovative solutions for businesses, prompting technology and engineering management practices to better adapt to the changing modern business environment.

The motivation for this study stems from a deep understanding of the critical roles of digital marketing and big data analytics in the field of technology and engineering management and their potential impact on sustainable digital marketing. Firstly, the importance of digital marketing in technology and engineering enterprises is evident not only in market positioning and product development but also in aspects like brand promotion and customer relationship management. With increasing competition, businesses urgently need to utilize digital channels for more precise target marketing while enhancing visibility and reputation through platforms like social media. This not only drives innovation but also emphasizes the indispensable nature of digital marketing for sustainable business development. Secondly, the application of big data analytics in technology and engineering management brings unprecedented technological advantages to businesses. Applications in predictive maintenance, supply chain optimization, and intelligent decision support enable businesses to operate and manage more efficiently. Big data analytics provides real-time, comprehensive data insights for technology and engineering enterprises, prompting decision-makers to formulate more strategic and informed plans. This solid foundation allows businesses to improve production efficiency and reduce costs. Considering the importance of both digital marketing and big data analytics, the motivation for this study is further established. The study contends that the combination of digital marketing and big data analytics not only enhances the competitiveness of businesses but also opens up new possibilities for sustainable digital marketing. Through an in-depth study of their interaction in the field of technology and engineering management, the aim is to provide businesses with a more intelligent, efficient, and sustainable path for business development, making substantive contributions to industry progress and innovation. Therefore, the motivation for this study lies in deepening the understanding of the integrated application of digital marketing and big data analytics in technology and engineering management, providing in-depth theoretical and practical guidance for achieving sustainable digital marketing.

This study aims to incorporate big data and AI deep learning models into sustainable digital marketing strategies for corporate brands. It utilizes big data technology to collect, process, and analyze a vast amount of social media data to extract valuable insights regarding customer behavior, preferences, and trends. The novelty of this endeavor resides in the amalgamation of extensive data analytics and the deep learning paradigm, facilitating a proficient prognostication of the consumption patterns exhibited by patrons of clothing brands. This fusion empowers enterprises and platforms to strategically deploy pertinent advertisements targeted at the specific clientele, thereby fostering an augmentation in both product conversion rates and transaction frequencies. The ultimate consequence manifests in these commercial entities' elevated operational efficiency.

This study is organized into the following sections. Section 1 is the introduction, elucidating the study background, motivation, and objectives. Section 2 is the literature review, analyzing the characteristics and applications of big data and the current study on sustainable marketing. It highlights the shortcomings of existing studies and introduces the study questions of this study. Section 3 is the study methodology, introducing the basic architecture and process of the random forest (RF) algorithm used and leveraging AI to construct an RFM model for analyzing consumer behavioral data and predicting their purchasing behavior. Section 4 covers the experiments, detailing the dataset, environment, parameter settings, and experimental results. Section 5 is the conclusion, summarizing the study contributions, highlighting limitations, and providing prospects for future study.

2. Literature Review

2.1 Characteristics and application analysis of big data

2.1.1 The characteristics of big data

The quintessential attributes of big data encompass five distinctive facets: abundance, rapidity, heterogeneity, diminished value concentration, and veracity. These attributes collectively encapsulate the quintessence that researchers commonly associate with the defining traits of big data, namely, voluminous data volumes, divergent composition, intricate interrelationships, and expeditious refresh cycles, as succinctly depicted in Table 1.

Table 1 Characteristics of big data

Characteristic dimension	Content
A large amount of data	Through the comprehensive acquisition of all pertinent data points, individuals can systematically encapsulate a plethora of intricate particulars, thereby engendering a substantial repository of information (Sun, Shang, Xia, Bhowmick & Nagarajaiah, 2020).
Heterogeneity	The deliberate attenuation of stringent accuracy pursuit on a macroscopic scale can yield precise outcomes at the microscopic level, as postulated by Hancock and Khoshgoftaar (Hancock & Khoshgoftaar, 2020).
Correlation	Owing to the pronounced conflation of interconnected states, there arises a

	transference of informational emphasis from causality to correlation. The concept of relevance facilitates a directed concentration on the 'what' rather than delving into the 'why,' as expounded by Stergiou, Psannis, and Gupta (Stergiou, Psannis & Gupta, 2020).
Quick update	The landscape of unstructured voluminous data is characterized by its dispersed distribution and perpetual fluctuations across diverse domains. The expeditious derivation of consequential insights and the formulation of pragmatic foundations for informed decision-making aimed at result optimization pose significant queries. Enhancing efficiency is anticipated to concomitantly expedite the pace of data updates, as posited by Jablonka, Ongari, Moosavi, and Smit (Jablonka, Ongari, Moosavi & Smit, 2020).

2.1.2 Application analysis of big data

Big data has diverse applications encompassing domains such as defense, governmental administration, critical infrastructure, societal oversight, and economic expansion. Its pervasive integration extends across realms, including smart urban environments, competitive athletics, and electronic commerce (Andronie et al., 2021). The inherent worth of data materializes not merely from its existence but rather from its astute utilization and substantive contributions to commercial endeavors. Currently, the endeavor to extract business value through the deployment of big data can be distilled into four distinct avenues, as succinctly delineated in Table 2 presented herein.

Table 2 Application ways of big data

Path dimension	Specific content
Customer group segmentation	In the context of contemporary market dynamics, the imperative of addressing diverse customer needs across multiple market segments has become increasingly pronounced. Concurrently, the pursuit of tailoring optimally suited products for distinct demographic cohorts emerges as a pivotal strategy in materializing the paradigm of individualized demand servicing (Wang, Yang, Wang, Sherratt & Zhang, 2020).
Real environment simulation	The strategic utilization of big data, endowed with predictive prowess, holds the potential to harness emergent market demands, emulate authentic trend landscapes, identify refined investment domains, and consequently engender heightened returns on investment (Marinakis, 2020).
Strengthen departmental ties	Facilitating interdepartmental communication efficiency and augmenting the efficacy of production and management chains, the internal sharing of data resources and analytical outcomes within an enterprise stands to be a transformative practice (Koo, Kim & Kim, 2022).
Control service cost	Leveraging the analytical capabilities of big data contributes substantively to streamlining the customer comprehension process, curtailing temporal expenditures, unveiling latent dimensions of demand, guiding pioneering endeavors in product and service innovation, and efficaciously managing cost structures (Albahri et al., 2022).

2.2 Sustainable Digital Marketing

In the pursuit of sustainable digital marketing, the procedural stages for implementing integrated big data marketing are delineated herewith. Initially, the compilation and refinement of customer information ensue. Customer data are sourced from diverse information channels, with consolidation and summarization of internal and external organizational data systems (Li & Jia, 2020). Subsequently, the taxonomy comprehension diverges from conventional marketing classification methodologies. Following data cleansing, customer categorization is effectuated, wherein those possessing inherent label attributes are automatically classified as per prescribed selection criteria. Big data judiciously extracts pertinent customer classification information from an array of seemingly unrelated data sources tailored to the requisite objectives (Romero Leguina et al., 2020; Zhou et al., 2022). Thirdly, the customer persona is deciphered. Customers harboring dynamic classification project-associated labels undergo comprehensive analysis across varying dimensions, contingent on the distinctive requisites of application programs. The identification of distinctive attributes and habitual conduct within customer clusters lays the groundwork for precise marketing endeavors (Bracarense et al., 2022; Butuner et al., 2023). Ultimately, the formulation of marketing strategy and execution blueprint transpires. Initiating the process by defining phased marketing objectives and substantiating initial metrics, the framework aims to enhance audience engagement or amplify communication effectiveness. The marketing implementation blueprint undergoes refinement, integrating the outcomes derived from big data analysis into tailored product sales strategies and activity blueprints (Lee et al., 2021; Kim et al., 2022).

Amidst the ongoing technological advancements, the transformative potential of big data and AI has redefined conventional paradigms, impacting diverse industries. The application of big data and AI technology has augmented transactional efficiency, considerably shortening the distance from demand to supply, engendering enhanced mutual understanding between enterprises and consumers, thus affording patrons the indulgence of personalized experiences (Shaw et al., 2021; Nica & Stehel, 2021). Manufacturers, retailers, and internet developers alike reap substantial benefits from this paradigm shift. With the proliferation of online commerce and the advent of the AI era, precision marketing determinations will increasingly draw upon the insights derived from big data analytics. Companies will recalibrate their operations predicated on data-driven optimization (Foresti et al., 2020).

Ngiam et al. assert that the advent of big data brings novel insights and value generation and transformative mechanisms influencing market dynamics, organizational structures, and governmental-citizen interactions. Consequently, it is posited that the cardinal role of big data lies in predictive analytics (Ngiam & Khor, 2019). Kastouni et al., in their discourse, underscore the analytical competencies and tools wielded by telecommunications operators leveraging big data, encompassing the creation of behavior classification labels for internet-connected consumers, identification of latent needs, alteration of specific customer preferences pertaining to mobile phone brands, as well as formulation of tailored traffic plans and app recommendations suited for mobile devices (Kastouni & Lahcen, 2022). Big data has engendered heightened returns within online marketing domains by harnessing internet-driven consumer predilections. Ageed et al. contend that the integration of data mining into the corporate marketing ecosystem encounters challenges due to underlying cost structures, culminating in restricted penetration and utilization rates. They advocate for enterprises to persist in augmenting data mining efforts, nurturing personnel proficient in amalgamating marketing and data mining acumen, and underscore the imperative of prioritizing data security during mining endeavors (Ageed et al., 2021). Informed by the classical RFM model within marketing, Li et al. differentiate diverse consumer behavior models, amalgamating these with the grey prediction model to furnish prognostications pertaining to future consumer behavior (Li et al., 2019). Kang contributes by enhancing the attribute reduction algorithm grounded in differential matrix principles, subsequently merging it with rough set theory to derive a novel predictive methodology (Kang, 2020). Indeed, it is notable that within e-commerce platforms, the cohort of targeted customers for esteemed brands is comparably smaller in contrast to the broader non-targeted customer base, thereby predisposing the emergence of categorical imbalance concerns.

Firstly, existing research primarily focuses on utilizing machine learning methods to predict digital consumer behavior, leaving room for further exploration on how to delve deeper into the integration of big data and AI technologies to optimize digital marketing strategies. Secondly, despite some existing cases showcasing the application of big data in marketing, there is relatively limited research in the realm of sustainable digital marketing within the context of social media. Concrete empirical analyses and applications in this area are lacking. Hence, this study addresses the prediction of digital consumer behavior and underscores the integral fusion of big data techniques and deep learning models. This integration aims to comprehensively and accurately interpret consumer behavior, with a specific emphasis on digital marketing within the realm of social media. This approach aids in uncovering user behavior and trends on social media platforms, offering a fresh perspective for enterprises to formulate more precise marketing strategies. The primary objective of this study is to investigate the sustainable digital marketing strategy within the context of social media. The present study analyzes the beneficial ramifications arising from integrating big data and AI technology into digital marketing strategies, thereby augmenting the practical efficacy of these strategies in application.

3. Research Methodology

3.1 Research Model

In order to comprehend the dynamics of sustainable digital marketing, harnessing the potential of big data analysis is essential. In this study, big data encompasses customer information gathered from various information channels, including online platforms, social media, and internal company systems. These data sources provide valuable insights into consumer behavior, preferences, and trends. Rigorous data processing and cleansing extract meaningful patterns and correlations, establishing the foundation for subsequent AI-based sustainable digital marketing model development. The AI-based sustainable digital marketing model aims to predict various marketing indicators, encompassing customer behavior, preferences, and trends. The model can learn to manage the interaction between customers and digital marketing activities through in-depth training on historical data. Once trained, the model can provide valuable insights, enabling customer segmentation, behavior prediction, and preference identification. These insights facilitate tailored resonance-based marketing strategies for specific customer segments, thereby enhancing the overall efficiency and impact of digital marketing activities.

For this purpose, the foundational framework is the random forest model (RFM), an ensemble learning approach that predicts outcomes by constructing multiple decision trees and aggregating their results. Its structure is illustrated in Figure 1 (Balyan et al., 2022; Zhou et al., 2022).

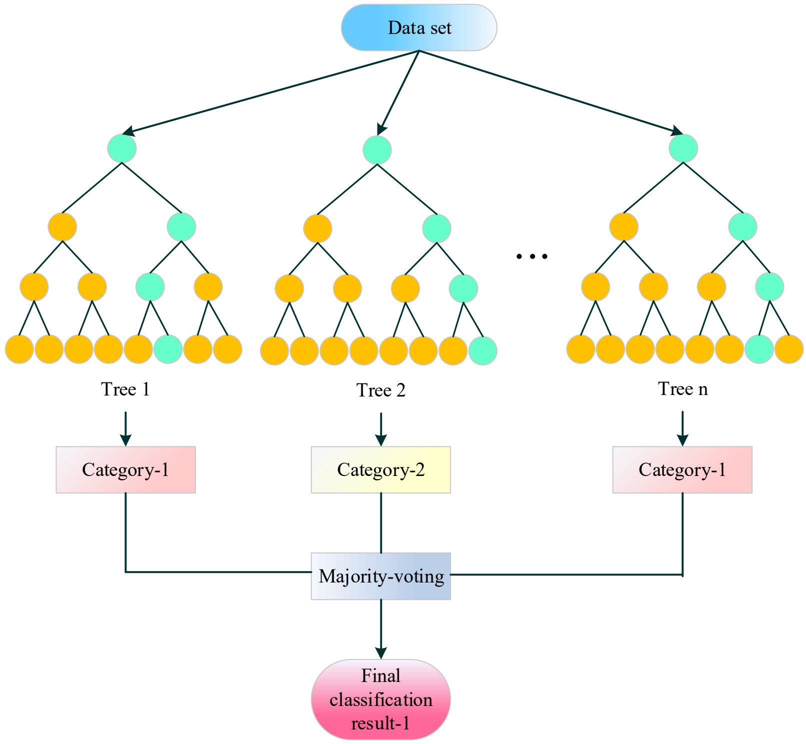

Figure 1 Schematic Representation of RF Structure

In Figure 1, the RF is an ensemble of multiple decision trees, combining the simplicity and interpretability of individual decision trees with the advantages of ensemble learning. It exhibits high accuracy and robustness, making it suitable for both classification and regression problems. The core concept of the RF lies in constructing multiple decision trees by randomly selecting diverse training samples and feature subsets. The predictions of each tree are aggregated to yield the final prediction (Bai et al., 2022; He et al., 2023). The fundamental process is illustrated in Figure 2.

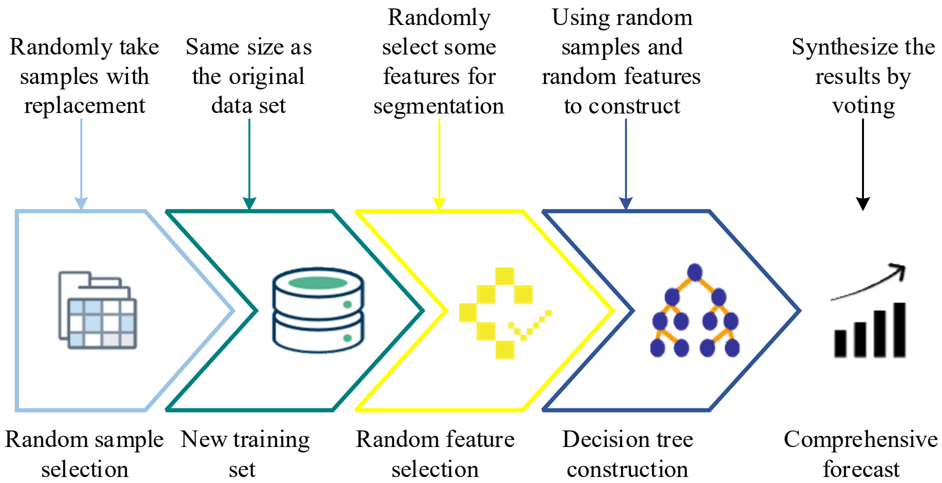

Figure 2 Basic Process of RF

In Figure 2, the RF consists of four fundamental steps: random sample selection, random feature selection, decision tree construction, and comprehensive prediction. Firstly, random samples are drawn with replacements from the original training dataset to form a new training set of the same size, allowing for potential sample repetition. Secondly, for each node of every decision tree, a subset of features is randomly selected from all features for splitting. This reduces the correlation between decision trees and increases diversity within the ensemble model. By harnessing both randomized samples and randomized features, a multitude of decision trees are systematically erected. Each individual tree is nurtured utilizing a subset of samples and features, culminating upon the fulfillment of predetermined halting criteria. These criteria encompass conditions such as achieving a specified depth or the cessation of node division predicated upon sample count. For classification problems, the results of each tree are aggregated through a voting mechanism to derive the final classification result, as shown in Eq. (1).

$$\hat{y} = mode(h_1(x), h_2(x), \cdots, h_n(x)) \tag{1}$$

Eq. (1), where $\hat{y}$ represents the prediction result of the RF. $h_i(x)$ signifies the prediction result of the i-th tree for input sample x, and n denotes the number of decision trees in the RF.

For regression problems, the results of each tree are combined by calculating the mean, as demonstrated in Eq. (2).

$$\hat{y} = \frac{1}{n}\sum_{i=1}^{n} h_i(x) \tag{2}$$

RF, through the integration of outcomes from multiple decision trees, exhibits robust generalization ability, effectively mitigating the risk of overfitting. Additionally, it can handle high-dimensional and large-scale datasets, is less sensitive to feature selection, and possesses some resilience toward outliers and missing values. In summary, by constructing multiple decision trees and aggregating their results, RF demonstrates robust predictive capability, making it suitable for various complex data analysis and forecasting tasks.

3.2 Constructing the RFM Model for Sustainable Digital Marketing with AI Approach

Integrated learning is a prevalent approach wherein multiple homogeneous core learners are employed. Subsequently, these core learners are combined according to specific rules to yield enhanced overall performance compared to a single core learner. Due to its commendable efficacy, integrated learning has emerged as a prominent research focus within the domain of machine learning in recent times. Within ensemble learning, there exist two primary factors influencing the ultimate classification performance (Sisodia & Sisodia, 2022). For an individual basic learner, it is imperative that their classification performance is not subpar, effectively surpassing random estimation. Secondly, the presence of diversity among several core learners is pivotal. This divergence arises from distinct training methodologies inherent to individual core learners, consequently delineating the integrated learning process into two principal categories: the former encompassing core learners displaying limited correlation, allowing the simultaneous generation of multiple core learners. Bagging and RF are representative techniques within this category. In contrast, the latter category entails substantial correlation among basic learners, mandating their generation to be predicated upon the prior category, which necessitates the continuous generation of the former. A representative method in this vein is the Boosting Cluster Algorithm (BCA) (Islam et al., 2022; Ong, 2022).

For the RFM methodology employed in this study, consider the training dataset of the current node denoted as A, encompassing attributes $x_1, x_2, \cdots, x_n$, and the class label $H \in \{0, 1\}$. Here, $f(x_i, A)$ represents the fundamental learner when attribute x_i partitions the dataset A. The foundational classifier employed is a decision tree. During the top-to-bottom construction of a decision tree, the selection of the current partition node necessitates the designation of an attribute. Common criteria for evaluating each potential attribute include information gain, gain rate, and the Gini index. Perturbation within RF stems from sample and attribute perturbations, thereby enhancing the operational efficiency of individual trees by reducing the number of samples and attributes per iteration (Satpathi et al., 2023; Ghosh & Maiti, 2022). Two primary determinants influence the final classification performance of RF: one pertains to the inter-tree correlation. A diminished correlation translates to fewer branches and leaves inserted between each tree, ultimately fostering robust overall classification performance. Conversely, heightened correlation corresponds to relatively diminished overall classification performance.

4. Experimental Design and Performance Evaluation

4.1 Datasets Collection

Offline shopping is less prevalent compared to other consumer destinations such as supermarkets. Upon reaching the payment counter, the assortment of products within each shopper's cart is diverse and captivating. Yet, a common thread unites them: these products are genuine necessities for the consumers, representing a fundamental departure from the dynamic of e-commerce shopping carts. The latter can be modified with ease, allowing the addition or removal of items regardless of temporal or spatial constraints. Consumers enjoy the freedom to purchase desired goods without limitations, even in the absence of immediate need.

As consumers repeatedly engage with shopping carts, a sense of possession and ownership over the products solidifies, reinforcing their transactional commitment. These behavioral patterns and underlying psychological aspects significantly contribute to product transactions, consequently influencing variables such as the quantity and timing of additional purchases. The act of browsing a brand or its associated category fulfills the consumer's consumption demand for that particular product category. Consequently, behavioral variables such as the frequency of trademark browsing, the timing of the last browse, and the duration of trademark interaction are selected for analysis. User-initiated keyword searches containing brand-related terms on the platform indicate an informed understanding of products, with brand-associated search queries serving as markers of consumer expectations. Moreover, several factors exert influence over consumer purchasing behavior, including purchasing power, sensitivity to promotions, and responsiveness to reviews.

This study adopts enterprise X as an illustrative case, representative of a prominent entity in the digital marketing service industry renowned for its commercial and technical solutions. A schematic depiction of its solution is presented in Figure 3:

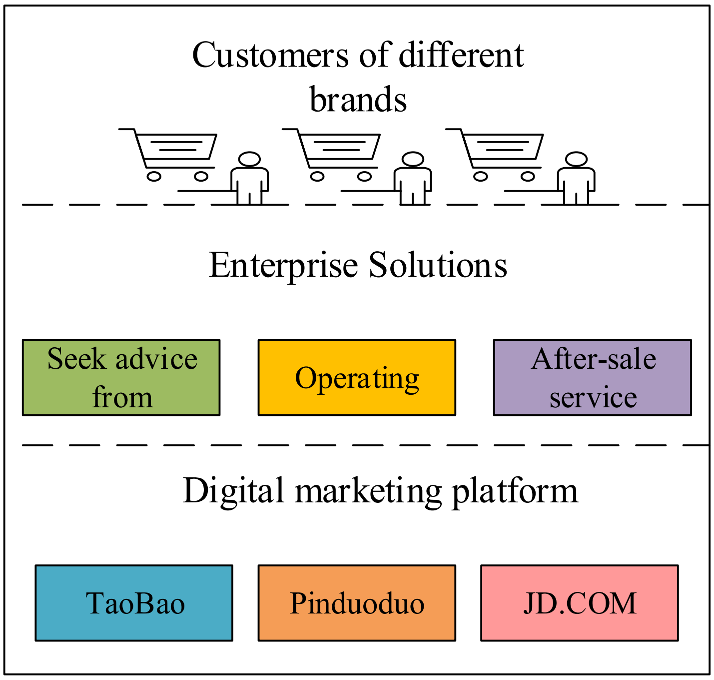

Figure 3 Solution of case enterprise

In Figure 3, the strategic approach undertaken by Enterprise X entails establishing a linkage between the e-commerce platform and the intermediary associated with the brand. This concerted effort facilitates the brand in achieving reduced e-commerce costs and swift integration into the digital marketing arena, particularly in instances where the brand's presence is relatively novel. This study employs the brand consumption data encompassing a one-month period from enterprise X. The selection process involves identifying consumer behavior data from the preceding month and pertinent internal factors. Subsequently, users are classified based on their engagement with brand-related purchases over the ensuing 7-day interval. The inherent attributes of the original dataset are detailed in Table 3:

Table 3 Original dataset properties

User attribute	Character value
Number of visits to the brand	Integer type
Number of times to join the shopping cart	Integer type
Number of payments	Integer type
Last time, to browse the brand	Year/Month/Date
Last time to join the shopping cart	Year/Month/Date
Basic characteristics of users	Character type
User's job characteristics	Character type
Customer unit price in this month	Float type

Purchasing power	Integer type
Social influence ability	Integer type
User sensitivity	Integer type
Whether to buy next week	Factor type

Within the dataset, the features are categorized into user attributes and the likelihood of making a purchase in the following week. These features depict user behavior and attributes on Enterprise X's e-commerce platform, encompassing visit frequency, shopping cart interactions, payment instances, browsing duration, user type, occupational traits, spending capacity, social influence, sensitivity, and more. These features will be employed in subsequent experiments and analyses to investigate user purchasing behaviors and various factors influencing their purchase decisions. A dataset containing data from 50,000 consumers was selected as the raw data for this experiment. Before any analysis takes place, data preprocessing is imperative to ensure data quality and consistency. Initially, timestamp data was transformed into a year/month/day format for subsequent analysis. Subsequently, missing values were handled using appropriate methods for imputation or removal. In order to enhance model accuracy, numerical data underwent normalization to ensure a uniform scale. Additionally, categorical attributes were encoded into numeric forms to facilitate model processing. A random sampling approach was employed to select training data during the training process, as depicted in Figure 4.

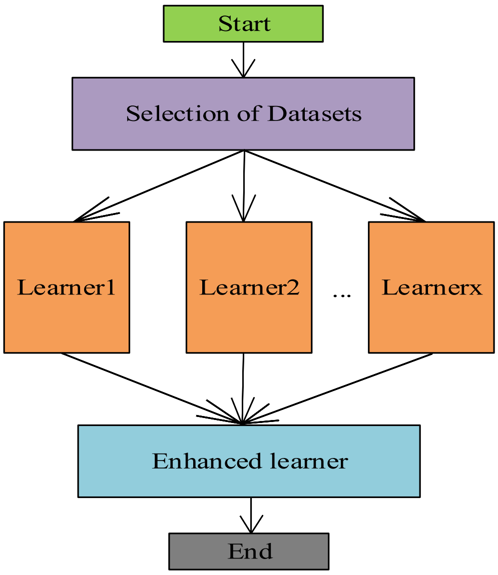

Figure 4 Flow of RF random selection of training data

In Figure 4, the depicted process involves the model's randomized selection of the training dataset, whereby the data is concurrently channeled into multiple learners, subsequently entering enhanced learners that yield the training outcomes.

4.2 Experimental Environment

The experimental operating system was a 64-bit Windows 10 system with an Intel i5-7500 CPU and 16GB of memory. The software environment utilized was R 3.6.1. Given that this study employs the R programming language for implementation, executing R programs may impose higher demands on computational memory. Therefore, specific data preprocessing measures are requisite. Firstly, character-type attributes (excluding user names) undergo scanning, with time-type attributes subtracted from a fixed date (the inception of the data cycle). Secondly, the removal of rows with missing values is executed. Thirdly, recognizing the notable dissimilarities between users who make brand purchases and those who do not, the study adopts an "under-sampling" strategy to rectify imbalances, thereby aligning the numbers of positive and negative instances.

4.3 Parameters Setting

At the microscopic level of RF construction, a pivotal parameter is the number of nodes within each individual tree. On a broader scale, the size of the tree assumes significance as a crucial parameter. Firstly, the selection of the number of nodes is undertaken. Based on the outcomes derived from executing the R programming language, it becomes evident that the accuracy and efficacy are relatively heightened when the number of nodes is set at 2. Subsequent to ascertaining the optimal count of decision tree node variables within the model, the next phase involves determining the number of decision trees within the model. In order to arrive at this parameter, a visual analysis of the model is conducted, employing a model configuration characterized by two node variables. Observations indicate that once the count of decision trees within the RF surpasses 400, the model's error tends to stabilize. In pursuit of heightened operational efficiency, the decision is made to set the number of decision trees within the RF model at 400, thereby achieving an optimal model configuration. Consequently, a RFM is instantiated, encompassing a total of 400 decision trees, with each decision tree node comprising two variables.

4.4 Performance Evaluation

4.4.1 Analysis of the basic situation of consumers

The statistical information regarding consumer attributes, such as gender, age, educational background, and occupation, obtained from the AI-based RFM model is summarized in Figure 5.

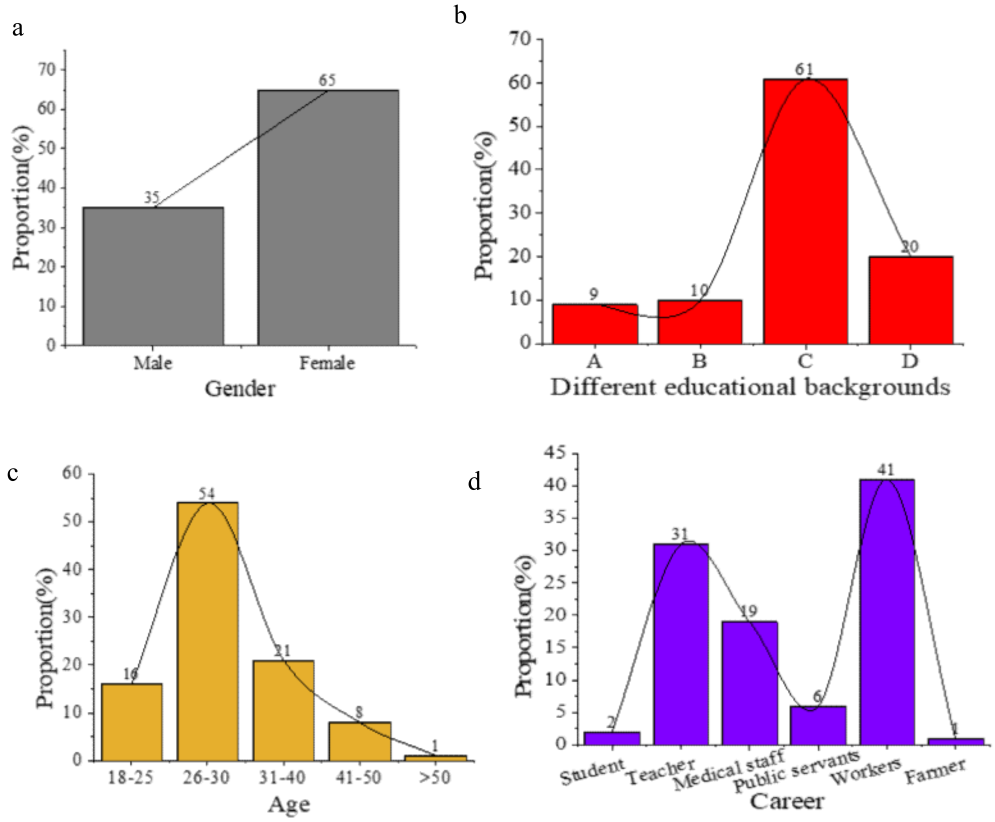

Figure 5 Basic statistics of consumers (a: gender b: educational background c: age d: occupation A: primary and secondary schools B: High School C: Undergraduate D: Postgraduate)

In Figure 5, the essential consumer information derived from the AI-based RFM model is outlined as follows. Over the course of the preceding 30-day period, the clientele associated with this specific brand predominantly consists of female users. Consequently, the brand's classification within the skincare category aligns logically with a notable prevalence of married consumers. Examination of consumer demographics, including age distribution and educational attainment, reveals that a substantial 54% of brand patrons fall within the age range of 26 to 30. In comparison, a mere 1% surpass the age of 50. This pattern underscores the propensity of the primary demographic for online shopping and internet usage within this age cohort, characterized by relative affluence. In contrast, the elderly demographic tends to exhibit a more gradual receptiveness to novel technological trends. Furthermore, a marked prevalence of university-affiliated individuals across all age strata is discernible. Delving into the occupational distribution of consumers, workers and educators contribute significantly, comprising 41% and 31% of the consumer base, respectively. These analytical outcomes strongly illustrate the effectiveness of AI-based RFM in interpreting consumer demographic data. By considering factors like purchase frequency, monetary value, and recency of purchase, the AI-based RFM model generates the aforementioned insights about consumer characteristics. These insights contribute to a better understanding of the brand's audience attributes and shed light on the consumption behaviors of diverse age groups, genders, and occupational segments.

4.4.2 Consumer Behavior Analysis

The study is primarily centered on prospective purchasers within the upcoming seven-day timeframe, necessitating an examination of certain consumer behaviors exhibited on the platform during the preceding month. The initial metrics scrutinized in this investigation encompass the frequency of brand engagement, cart additions, and order placements by consumers over the past month. The outcomes of this analysis are graphically depicted in Figures 6 and 7.

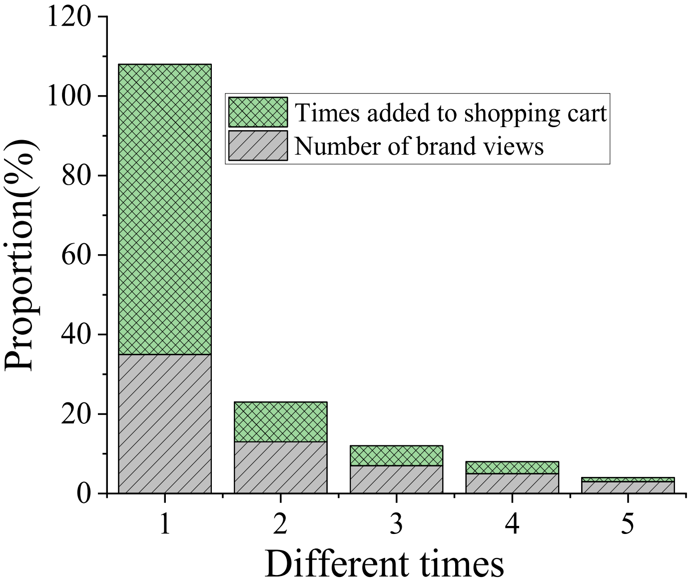

Figure 6 Statistics on the times of checking the brand and joining the shopping cart in the past month (1: 1-15 times, 2: 15-30 times, 3: 30-45 times, 4: 45-60 times, 5: > 60 times)

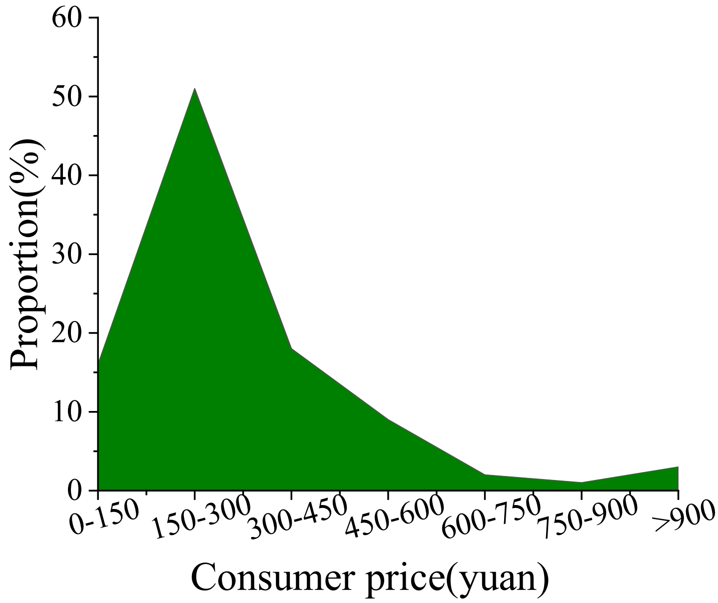

Figure 7 Statistics of customer unit price in the past month

Based on the observations derived from Figures 6 and 7, it is apparent that within the cohort of users who engaged with the category, visited the brand, and made additions to their shopping carts during the previous month, a notable proportion, specifically 34%, exhibited a brand visitation frequency of no more than 15 times. Moreover, a substantial percentage (73%) engaged in shopping cart additions within a frequency of 15 times or less. It is also noteworthy that the distribution of customer unit prices displays a skewed pattern, with a prevalence of 17% corresponding to the price range of 0-150 and a dominant 52% attributed to the price range spanning 150-300. The AI-based RFM model conducted an in-depth consumer behavior analysis over the past month, offering valuable insights to understand consumer behavioral patterns. These findings are instrumental in providing valuable insights for precise marketing strategies.

4.4.3 Prediction results of RFM

Upon completing a continuous seven-day recording period, the resultant forecast dataset consists of approximately 18.5 million records. Within this dataset, the total count of users who indeed made purchases during the subsequent seven days amounted to 39,877. Among the users identified in accordance with the extraction criteria defined in this study, a cumulative count of 21,637 users was established as actual purchasers, constituting approximately 54.2% of the entire population of verified purchasers. The predictive model is then applied to the forecasting dataset, and a comparative assessment of predicted outcomes is conducted among Logistic regression, RF, and XGBoost models. These comparative results are depicted in Figure 8:

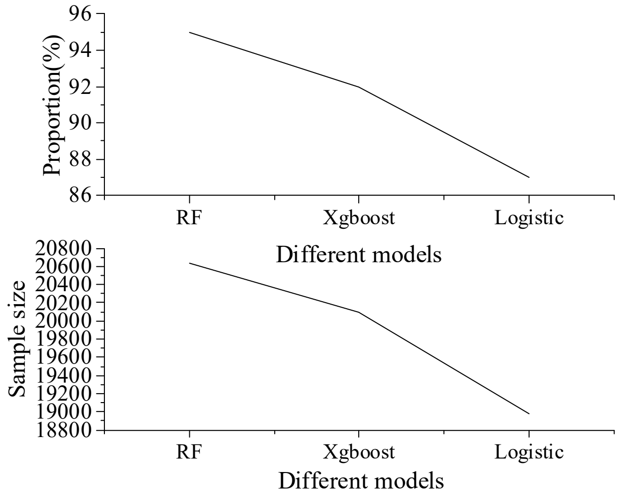

Figure 8 Forecast results of three models for the number of passenger orders in the next seven days

As depicted in Figure 8, a comparative analysis of the projected count of prospective clientele across the three models reveals noteworthy observations. Specifically, the prognosticated figures for RF and Xgboost closely approximate the 20,000 marks, exhibiting minimal disparity between them. Conversely, the anticipated count produced by the Least Squares Regression model is enumerated at 18,976, indicative of discernible advantageous attributes inherent to RF and Xgboost models. The data indicates that the AI-based RFM model accurately predicts passenger order quantities for the next seven days.

Furthermore, a comparison was conducted between the AI-based RFM model and other models, such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs), in forecasting passenger order quantities for the upcoming seven days, as illustrated in Figure 9.

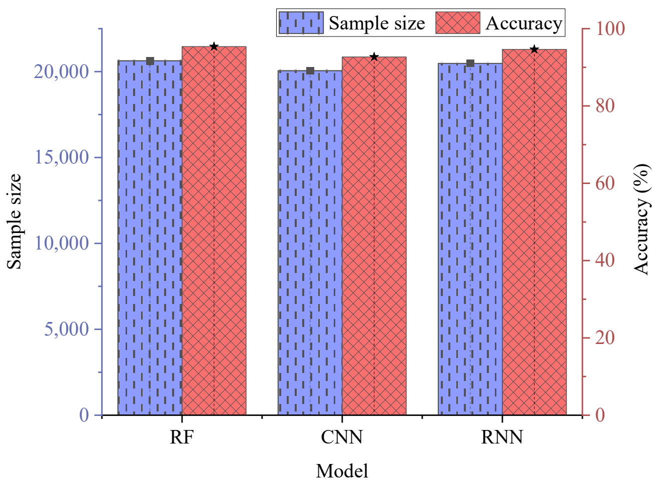

Figure 9 Forecasting Results of Three Models for Consumer Order Quantities

In Figure 9, the AI-based RFM model demonstrates superior predictive performance, achieving an accuracy rate of 95.36%. The accuracy rates for CNN and RNN are 92.6% and 94.64%, respectively, slightly lower than the AI-based RFM model. This data indicates that the RFM model offers relatively precise forecasts for the number of purchasers, closely approximating actual purchase counts, thereby showcasing commendable performance.

4.5 Discussion

This study integrates the optimization of a sustainable digital marketing strategy within the framework of big data and AI-powered RFM. The outcomes effectively scrutinize fundamental customer attributes from the preceding month, encompassing gender, age, educational background, and occupation. In the context of predicting the forthcoming seven-day passenger ticket volumes, the employed RFM exhibits a superior advantage over the Logistic regression approach. Gao and Ding's investigation underscored that the application of refined RF methodologies to digital marketing planning facilitates real-time adjustments of marketing campaigns and product content aligned with shifts in the target market. Moreover, it emphasizes the importance of promptly iteratively updating products based on user feedback to enhance product market competitiveness and optimize user experiences (Gao & Ding, 2022). This study focuses on sustainable digital marketing, with a particular emphasis on the application of big data analytics and AI technology within digital marketing. Compared to the research conducted by Gao and Ding, they highlighted using a refined random forest approach in digital marketing to achieve real-time adjustments to market changes. Sinha et al.'s research contributed an interactive user interface that offers digital marketing professionals real-time access to insights derived from back-end AI systems. This AI-driven model prognosticates potential click-through rates for combined content based on analogous prior activities (Sinha et al., 2020). Compared to the study, the research by Saura and others also underscored how small and medium-sized enterprises applied data science in digital marketing, encompassing areas such as product sales, brand promotion, and market entry. Collectively, these studies underscore the significance of data science in digital marketing, particularly in real-time adjustments and product optimization. In a parallel vein, Kiguchi et al. employed RFM to construct a customer churn prediction model within digital marketing, comparing its performance against logistic regression and decision tree models to identify the optimal prediction model (Kiguchi et al., 2022). This aligns with the findings of the current study. In conclusion, a growing body of research underscores the efficacy of AI-driven sustainable digital marketing strategies within the context of the vast big data landscape.

Saura et al. explored the role and usage of data science in small and medium-sized enterprises (SMEs) in digital marketing, finding that the adoption of data science may involve aspects such as product sales, brand promotion, and market entry (Saura, Palacios-Marqués & Ribeiro-Soriano, 2023). Kongar et al., combining machine learning and data envelopment analysis (DEA), utilized advanced analytical tools to process large-scale Twitter messages and metrics data. Their study, analyzing tweets and brand efficiency, provided industry practitioners with recommendations on how to utilize social media more effectively (Kongar & Adebayo, 2021). While this study does not directly address Twitter messages, it focuses on social media analysis in digital marketing, emphasizing the impact of technological use on societal, economic, and environmental sustainability. Yadav et al., focusing on the IoT-based agri-food supply chain, recognized the emerging paradigm of digitization and IoT in this field and assessed its potential role in enhancing social well-being (Yadav, Choi, Luthra, Kumar & Garg, 2022). In comparison to Yadav et al.'s research, which concentrates on the IoT-based agri-food supply chain, this study does not specifically discuss agricultural supply chains. Instead, the emphasis is on sustainability in digital marketing, highlighting the direct impact of technological use on environmental, economic, and social sustainability. Sharma et al. employed a covariance structural equation model to validate the relationship between external environmental factors and internal environmental factors in SMEs. Their results confirmed a positive correlation between imitative homogenization and normative homogenization in external environmental factors and the intention to adopt agile innovation management. Additionally, internal environmental factors such as top management leadership, flexible culture, family culture, and organizational readiness were also confirmed to have a positive correlation with the adoption of agile innovation management (Sharma, Singh, Jones, Kraus & Dwivedi, 2022). This study, like Sharma et al.'s research, focuses on the impact of advanced technology on sustainable marketing strategies, validating the study findings. In summary, while these studies differ in specific topics, they collectively emphasize the application of advanced technology in digital marketing, particularly data science, social media analysis, IoT, and agile innovation management. These studies provide rich theoretical and empirical support for the fields of digital marketing and sustainability, aiding in driving innovation and sustainable development for businesses in the digital era.

5. Conclusion

5.1 Research Contribution

5.1.1 Practical Implications

The study shares the focus of Al-Emran et al. on the critical role of advanced technology in sustainable development. Al-Emran et al.'s research concentrates on the pivotal role of advanced technology in the environmental, economic, and social pillars of sustainable development. They developed the "Technology-Environmental, Economic, and Social Sustainability Theory" (T-EESST), which surpasses existing information system theories by linking technology use with three dimensions of sustainable development. The study posits that advanced technology may have positive or negative impacts on sustainable development (Al-Emran, 2023). Similar to T-EESST, this study focuses on the application of big data analytics and AI technology in digital marketing, emphasizing the impact of these technologies on social, economic, and environmental sustainability. By delving into the direct effects of technology use on sustainable digital marketing, the study supports the core viewpoint of T-EESST, asserting that technology use can have positive or negative effects on sustainable development. These findings offer empirical support, highlighting how technology applications in digital marketing directly influence the three dimensions of sustainable development.

This study provides profound insights for practitioners, decision-makers, and researchers in the field of digital marketing. By emphasizing the innovative and driving role of technology in digital marketing, the study not only underscores the importance of technological advancements but also delves into the direct effects of these technologies on environmental, economic, and social sustainability. This provides practical strategic guidance for businesses aiming at sustainable development in the digital era. Additionally, the study offers decision-makers evidence-based guidance, aiding in more effective planning and managing digital marketing strategies to achieve sustainable development goals.

5.1.2 Theoretical Contributions

The present study undertakes an analysis of the attributes and applications inherent to big data, subsequently constructing an AI-driven RFM tailored for sustainable digital marketing endeavors. By utilizing a chosen brand as an illustrative exemplar, the research employs consumption data spanning a duration of one month. Within this context, the investigation identifies pertinent criteria derived from consumer behavioral patterns during the previous month and pertinent internal factors. Subsequently, users are categorized based on their engagement with the brand's offerings within the ensuing seven-day period, thereby facilitating the prediction of forthcoming customer orders over the same duration. The outcomes underscore the AI RFM's capacity to proficiently prognosticate the volume of target customers, thereby furnishing a pragmatic underpinning for the intelligent evolution of sustainable digital marketing paradigms. This study has made significant contributions to the sustainable development of digital marketing strategies by integrating big data analysis of customer behavior and trends with an AI-based RFM. The application of big data and deep learning models to the realm of digital marketing has shown promising prospects for enhancing the accuracy and effectiveness of marketing strategies. This research provides valuable insights and inspiration for related fields of study and practical applications.

In the field of engineering management, the following impacts are anticipated: Data-Driven Decision Making: the application of big data and deep learning models in sustainable digital marketing can serve as a model for data-driven decision-making in engineering management. This approach could lead to more accurate predictions and analyses of project needs, progress, and risks, thereby optimizing decision-making processes. Precision Marketing Strategies: the emphasis on the AI-based RFM model's predictive capabilities for customer behavior could guide engineering management in refining and personalizing marketing strategies. This precision could be especially effective in project promotion, resource allocation, and partner selection. Enhanced Project Efficiency: similar to optimizing conversion rates and transaction rates in the study, engineering management could utilize data analysis and deep learning to optimize project processes, resource utilization, and project delivery efficiency, ultimately improving overall project efficiency. Real-Time Monitoring and Feedback: the importance of real-time data analysis highlighted in the study can be applied to engineering management to monitor project progress, quality, and costs in real time. This approach enables timely adjustments to project strategies, minimizing risks and delays. Market Insights and Competitive Analysis: through big data analysis, engineering management can gain better insights into market demands and competitive landscapes, providing robust support for project positioning and market strategies. This enables a more targeted approach to project management and promotion. In summary, this study introduces a data-driven and deep-learning mindset to the field of engineering management, potentially offering more accurate, efficient, and flexible approaches and strategies for project decision-making, execution, and monitoring.

This study introduces the theoretical framework of T-EESST, combining technology use with environmental, economic, and social sustainability. This theoretical framework transcends traditional information system theories, providing a more comprehensive understanding of the impact of technology on sustainability in digital marketing. This study introduces a new theoretical perspective to the field of sustainable development, offering valuable discourse space for the academic community. The theoretical framework is poised to have far-reaching implications not only in the field of digital marketing but across the entire information systems domain, expanding the theoretical understanding of the relationship between technology and sustainable development.

5.2 Future Works and Research Limitations

A limitation of this study pertains to its exclusive reference to widely recognized algorithms. In order to attain heightened precision and validation, a comprehensive array of experiments is requisite to ascertain and contrast the efficacy of diverse algorithms. Additionally, an adaptive investigation should be undertaken to delve further into the capabilities of individual algorithms. It is acknowledged that user purchasing behavior is subject to multifarious influences, some of which involve impulsive consumption tendencies. Consequently, future research endeavors will encompass a psychological perspective, thereby facilitating the elucidation of user psychology, user categorization, and tailored communication strategies. Furthermore, endeavors to sustain customer engagement will be augmented through offline interactions and incentivized returns. Furthermore, the data used in this study primarily originates from specific brand consumer behaviors, which may limit the generalizability and applicability of the study to some extent. Future studies could consider expanding the scale and scope of data samples to encompass a broader range of brands and industries, enhancing the study's generalizability and applicability. Additionally, further optimization and improvement of the AI-based RFM model could enhance its predictive accuracy and stability and explore the application of additional deep learning algorithms.

Fundings:

This work is supported by the China Scholarship Council (CSC) under the grant number: 202106980007

References

Ageed, Z. S., Zeebaree, S. R., Sadeeq, M. M., Kak, S. F., Yahia, H. S., Mahmood, M. R., & Ibrahim, I. M. (2021). Comprehensive survey of big data mining approaches in cloud systems. *Qubahan Academic Journal*, *1*(2), 29-38.

Albahri, A. S., Alnoor, A., Zaidan, A. A., Albahri, O. S., Hameed, H., Zaidan, B. B., ... & Yass, A. A. (2022). Hybrid artificial neural network and structural equation modelling techniques: a survey. *Complex & Intelligent Systems*, *8*(2), 1781-1801.

Al-Emran, M. (2023). Beyond technology acceptance: Development and evaluation of technology-environmental, economic, and social sustainability theory. *Technology in Society*, *75*, 102383.

Amiri, A. M., Kushwaha, B. P., & Singh, R. (2023). Visualisation of global research trends and future research directions of digital marketing in small and medium enterprises using bibliometric analysis. *Journal of Small Business and Enterprise Development*, *30*(3), 621-641.

Andronie, M., Lăzăroiu, G., Iatagan, M., Hurloiu, I., & Dijmărescu, I. (2021). Sustainable cyber-physical production systems in big data-driven smart urban economy: a systematic literature review. *Sustainability*, *13*(2), 751.

Bai, J., Li, Y., Li, J., Yang, X., Jiang, Y., & Xia, S. T. (2022). Multinomial random forest. *Pattern Recognition*, *122*, 108331.

Balyan, A. K., Ahuja, S., Lilhore, U. K., Sharma, S. K., Manoharan, P., Algarni, A. D., ... & Raahemifar, K. (2022). A hybrid intrusion detection model using ega-pso and improved random forest method. *Sensors*, *22*(16), 5986.

Behl, A. (2022). Antecedents to firm performance and competitiveness using the lens of big data analytics: a cross-cultural study. *Management Decision*, *60*(2), 368-398.

Bracarense, N., Bawack, R. E., Fosso Wamba, S., & Carillo, K. D. A. (2022). Artificial intelligence and sustainability: A bibliometric analysis and future research directions. *Pacific Asia Journal of the Association for Information Systems*, *14*(2), 9.

Butuner, R., Cinar, I., Taspinar, Y. S., Kursun, R., Calp, M. H., & Koklu, M. (2023). Classification of deep image features of lentil varieties with machine learning techniques. *European Food Research and Technology*, *249*(5), 1303-1316.

Dash, G., & Chakraborty, D. (2021). Digital transformation of marketing strategies during a pandemic: Evidence from an emerging economy during COVID-19. *Sustainability*, *13*(12), 6735.

Demirbaga, U., & Aujla, G. S. (2022). MapChain: A blockchain-based verifiable healthcare service management in IoT-based big data ecosystem. *IEEE Transactions on Network and Service Management*, *19*(4), 3896-3907.

Diez-Martin, F., Blanco-Gonzalez, A., & Prado-Roman, C. (2019). Research challenges in digital marketing: sustainability. *Sustainability*, *11*(10), 2839.

Eleonora, Z., Tetiana, R., Ihor, R., Nataliia, P., & Anastasiia, M. (2023). Analysis of digital marketing trends in 2022-2023 of the global fuel and energy complex (experience for Ukraine). *Calitatea*, *24*(192), 360-367.

Foresti, R., Rossi, S., Magnani, M., Bianco, C. G. L., & Delmonte, N. (2020). Smart society and artificial intelligence: big data scheduling and the global standard method applied to smart maintenance. *Engineering*, *6*(7), 835-846.

Gao, W., & Ding, Z. (2022). Construction of Digital Marketing Recommendation Model Based on Random Forest Algorithm. *Security and Communication Networks*, *2022*.

Ghosh, A., & Maiti, R. (2022). Application of SWAT, Random Forest and artificial neural network models for sediment yield estimation and prediction of gully erosion susceptible zones: study on Mayurakshi River Basin of Eastern India. *Geocarto International*, *37*(25), 9663-9687.

Hancock, J. T., & Khoshgoftaar, T. M. (2020). CatBoost for big data: an interdisciplinary review. *Journal of big data*, *7*(1), 1-45.

He, B., Armaghani, D. J., & Lai, S. H. (2023). Assessment of tunnel blasting-induced overbreak: A novel metaheuristic-based random forest approach. *Tunnelling and Underground Space Technology*, *133*, 104979.

Islam, U., Muhammad, A., Mansoor, R., Hossain, M. S., Ahmad, I., Eldin, E. T., ... & Shafiq, M. (2022). Detection of distributed denial of service (DDoS) attacks in IOT based monitoring system of banking sector using machine learning models. *Sustainability*, *14*(14), 8374.

Jablonka, K. M., Ongari, D., Moosavi, S. M., & Smit, B. (2020). Big-data science in porous materials: materials genomics and machine learning. *Chemical reviews*, *120*(16), 8066-8129.

Kang, X. (2020). Aesthetic product design combining with rough set theory and fuzzy quality function deployment. *Journal of Intelligent & Fuzzy Systems*, *39*(1), 1131-1146.

Kastouni, M. Z., & Lahcen, A. A. (2022). Big data analytics in telecommunications: Governance, architecture and use cases. *Journal of King Saud University-Computer and Information Sciences*, *34*(6), 2758-2770.

Khanbaghi, M., & Zecevic, A. (2020). Jump linear quadratic control for microgrids with commercial loads. *Energies*, *13*(19), 4997.

Khare, V. K., Raghuwanshi, S., Verma, P., & Shrivastava, A. (2023). A bibliometric analysis of the effectiveness of digital marketing on smes in the digital era. *Anvesak. UGC Care Group 1 Journal*, *53*(1), 107-117.

Kiguchi, M., Saeed, W., & Medi, I. (2022). Churn prediction in digital game-based learning using data mining techniques: logistic regression, decision tree, and random forest. *Applied Soft Computing*, *118*, 108491.

Kim, J., Lee, Y., & Song, I. (2022). From intuition to intelligence: a text mining–based approach for movies' green-lighting process. *Internet Research*, *32*(3), 1003-1022.

Kongar, E., & Adebayo, O. (2021). Impact of social media marketing on business performance: A hybrid performance measurement approach using data analytics and machine learning. *IEEE Engineering Management Review*, *49*(1), 133-147.

Koo, D. M., Kim, J., & Kim, T. (2022). Guest editorial: Digital transformation and consumer experience. *Internet Research*, *32*(3), 967-970.

Kuzior, A., Lyulyov, O., Pimonenko, T., Kwilinski, A., & Krawczyk, D. (2021). Post-industrial tourism as a driver of sustainable development. *Sustainability*, *13*(15), 8145.

Lee, M., Kwon, W., & Back, K. J. (2021). Artificial intelligence for hospitality big data analytics: developing a prediction model of restaurant review helpfulness for customer decision-making. *International Journal of Contemporary Hospitality Management*, *33*(6), 2117-2136.

Li, J., Pan, S., & Huang, L. (2019). A machine learning based method for customer behavior prediction. *Tehnički vjesnik*, *26*(6), 1670-1676.

Li, Z., & Jia, W. (2020). The study on preventing click fraud in internet advertising. *J. Comput.*, *31*(3), 256-265.

Liu, Q., Wan, H., & Yu, H. (2023). Application and Influence of Big data Analysis in Marketing Strategy. *Frontiers in Business, Economics and Management*, *9*(3), 168-171.

Low, S., Ullah, F., Shirowzhan, S., Sepasgozar, S. M., & Lin Lee, C. (2020). Smart digital marketing capabilities for sustainable property development: A case of Malaysia. *Sustainability*, *12*(13), 5402.

Lu, J., Ma, J., Zheng, X., Wang, G., Li, H., & Kiritsis, D. (2021). Design ontology supporting model-based systems engineering formalisms. *IEEE Systems Journal*, *16*(4), 5465-5476.

Marinakis, V. (2020). Big data for energy management and energy-efficient buildings. *Energies*, *13*(7), 1555.

Munsch, A. (2021). Millennial and generation Z digital marketing communication and advertising effectiveness: A qualitative exploration. *Journal of Global Scholars of Marketing Science*, *31*(1), 10-29.

Nagendra, N. P., Narayanamurthy, G., Moser, R., Hartmann, E., & Sengupta, T. (2022). Technology Assessment Using Satellite Big Data Analytics for India's Agri-Insurance Sector. *IEEE Transactions on Engineering Management*, *70*(3), 1099-1113.

Ngiam, K. Y., & Khor, W. (2019). Big data and machine learning algorithms for health-care delivery. *The Lancet Oncology*, *20*(5), e262-e273.

Nica, E., & Stehel, V. (2021). Internet of things sensing networks, artificial intelligence-based decision-making algorithms, and real-time process monitoring in sustainable industry 4.0. *Journal of Self-Governance and Management Economics*, *9*(3), 35-47.

Ong, A. K. S. (2022). A machine learning ensemble approach for predicting factors affecting STEM students' future intention to enroll in chemistry-related courses. *Sustainability*, *14*(23), 16041.

Paniagua, C., & Delsing, J. (2020). Industrial frameworks for internet of things: A survey. *IEEE Systems Journal*, *15*(1), 1149-1159.

Peng, Y., & Tao, C. (2022). Can digital transformation promote enterprise performance?—From the perspective of public policy and innovation. *Journal of Innovation & Knowledge*, *7*(3), 100198.

Rakic, S., Medic, N., Leoste, J., Vuckovic, T., & Marjanovic, U. (2023). Development and future trends of digital product-service systems: A bibliometric analysis approach. *Applied System Innovation*, *6*(5), 89.

Romero Leguina, J., Cuevas Rumín, Á., & Cuevas Rumín, R. (2020). Digital marketing attribution: Understanding the user path. *Electronics*, *9*(11), 1822.

Samara, D., Magnisalis, I., & Peristeras, V. (2020). Artificial intelligence and big data in tourism: a systematic literature review. *Journal of Hospitality and Tourism Technology*, *11*(2), 343-367.

Satpathi, A., Setiya, P., Das, B., Nain, A. S., Jha, P. K., Singh, S., & Singh, S. (2023). Comparative Analysis of Statistical and Machine Learning Techniques for Rice Yield Forecasting for Chhattisgarh, India. *Sustainability*, *15*(3), 2786.

Saura, J. R., Palacios-Marqués, D., & Ribeiro-Soriano, D. (2023). Digital marketing in SMEs via data-driven strategies: Reviewing the current state of research. *Journal of Small Business Management*, *61*(3), 1278-1313.

Saura, J. R., Palos-Sanchez, P., & Rodríguez Herráez, B. (2020). Digital marketing for sustainable growth: Business models and online campaigns using sustainable strategies. *Sustainability*, *12*(3), 1003.

Shankar, V., Grewal, D., Sunder, S., Fossen, B., Peters, K., & Agarwal, A. (2022). Digital marketing communication in global marketplaces: A review of extant research, future directions, and potential approaches. *International Journal of research in Marketing*, *39*(2), 541-565.

Sharma, S., Singh, G., Jones, P., Kraus, S., & Dwivedi, Y. K. (2022). Understanding agile innovation management adoption for SMEs. *IEEE Transactions on Engineering Management*, *69*(6), 3546-3557.

Shaw, S., Rowland, Z., & Machova, V. (2021). Internet of Things smart devices, sustainable industrial big data, and artificial intelligence-based decision-making algorithms in cyber-physical system-based manufacturing. *Economics, Management and Financial Markets*, *16*(2), 106-116.

Sinha, M., Healey, J., & Sengupta, T. (2020, July). Designing with AI for digital marketing. In *Adjunct Publication of the 28th ACM Conference on User Modeling, Adaptation and Personalization* (pp. 65-70).

Sisodia, D., & Sisodia, D. S. (2022). Data sampling strategies for click fraud detection using imbalanced user click data of online advertising: an empirical review. *IETE Technical Review*, *39*(4), 789-798.

Stergiou, C. L., Psannis, K. E., & Gupta, B. B. (2020). IoT-based big data secure management in the fog over a 6G wireless network. *IEEE Internet of Things Journal*, *8*(7), 5164-5171.

Sun, L., Shang, Z., Xia, Y., Bhowmick, S., & Nagarajaiah, S. (2020). Review of bridge structural health monitoring aided by big data and artificial intelligence: From condition assessment to damage detection. *Journal of Structural Engineering*, *146*(5), 04020073.

Thangam, D., & Chavadi, C. (2023). Impact of Digital Marketing Practices on Energy Consumption, Climate Change, and Sustainability. *Climate and Energy*, *39*(7), 11-19.

Wang, J., Yang, Y., Wang, T., Sherratt, R. S., & Zhang, J. (2020). Big data service architecture: a survey. *Journal of Internet Technology*, *21*(2), 393-405.

Yadav, S., Choi, T. M., Luthra, S., Kumar, A., & Garg, D. (2022). Using Internet of Things (IoT) in agri-food supply chains: a research framework for social good with network clustering analysis. *IEEE Transactions on Engineering Management*, *70*(3), 1215-1224.

Yousaf, Z., Radulescu, M., Sinisi, C. I., Serbanescu, L., & Păunescu, L. M. (2021). Towards sustainable digital innovation of SMEs from the developing countries in the context of the digital economy and frugal environment. *Sustainability*, *13*(10), 5715.

Zhou, J., Huang, S., & Qiu, Y. (2022). Optimization of random forest through the use of MVO, GWO and MFO in evaluating the stability of underground entry-type excavations. *Tunnelling and Underground Space Technology*, *124*, 104494.

Zhou, Y., Xia, Q., Zhang, Z., Quan, M., & Li, H. (2022). Artificial intelligence and machine learning for the green development of agriculture in the emerging manufacturing industry in the IoT platform. *Acta Agriculturae Scandinavica, Section B—Soil & Plant Science*, *72*(1), 284-299.

Ziakis, C., & Vlachopoulou, M. (2023). Artificial Intelligence in Digital Marketing: Insights from a Comprehensive Review. *Information*, *14*(12), 664.