



International Journal of Information and Communication Technology

ISSN online: 1741-8070 - ISSN print: 1466-6642

<https://www.inderscience.com/ijict>

Research on analysis and improvement strategies of consumer loyalty on e-commerce platforms based on big data

Zhonglin Li, Junhua Hu

DOI: [10.1504/IJICT.2026.10076519](https://doi.org/10.1504/IJICT.2026.10076519)

Article History:

Received:	18 August 2025
Last revised:	18 November 2025
Accepted:	23 November 2025
Published online:	09 March 2026

Research on analysis and improvement strategies of consumer loyalty on e-commerce platforms based on big data

Zhonglin Li* and Junhua Hu

School of Economics and Management,
Wuchang Institute of Technology,
Wuhan, Hubei, 430065, China
Email: lizhonglin@wuit.edu.cn
Email: wuitjglzl@163.com
*Corresponding author

Abstract: This study aims to use advanced machine learning models and big data analytics (BDA) to investigate online marketplace client loyalty. To capture nonlinear correlations and identify significant loyalty determinants, the suggested framework uses algorithms like LSTM, XGBoost, support vector machines, and random forests, rather than typical statistical techniques. The study delves into the elements that influence consumer engagement, the frequency of purchases, and the likelihood of repurchases within the fresh food platform and pet-related e-commerce sectors. A ten-phase framework – covering foundation, data collection, preprocessing, AI development, pilot study, validation, and evaluation – is employed to ensure methodological rigour. When compared to more traditional methods of consumer loyalty prediction, LSTM performs better on measures including accuracy, precision, recall, F1 score, and area under the curve (AUC). The findings highlight the role of personalised recommendations, delivery services, and mobile engagement in shaping loyalty, while offering practical strategies for sustainable growth in vertical e-commerce markets.

Keywords: e-commerce; consumer loyalty; big data analytics; BDA; machine learning; LSTM; AI marketing; recommendation systems; vertical e-commerce; customer retention; predictive modelling.

Reference to this paper should be made as follows: Li, Z. and Hu, J. (2026) 'Research on analysis and improvement strategies of consumer loyalty on e-commerce platforms based on big data', *Int. J. Information and Communication Technology*, Vol. 27, No. 16, pp.1–21.

Biographical notes: Zhonglin Li is a faculty member in the School of Economics and Management at Wuchang Institute of Technology, Wuhan, China, where he is also the corresponding author. He received his academic training in economics and management and has been actively involved in teaching and research in financial economics and applied finance. His research focuses on FinTech adoption, investment efficiency, corporate governance, and empirical financial analysis. Over the course of his career, he has published scholarly articles in peer-reviewed journals and participated in institutional research projects addressing the role of technological innovation in financial decision-making and organisational performance.

Junhua Hu is a faculty member in the School of Economics and Management at Wuchang Institute of Technology, Wuhan, China. She holds an academic background in economics and management and has several years of teaching and research experience in financial management and economic analysis. Her research interests include financial innovation, institutional finance, and data-driven economic modelling. She has contributed to academic publications and collaborative research projects that examine the impact of emerging technologies on financial systems, investment behaviour, and economic development in developing and transitional economies.

1 Introduction

The growing need for environmentally friendly goods and services in light of critical environmental issues like global warming and resource loss has propelled the green industry to the forefront of the world economy. If e-commerce platforms want to be a part of the green economy and respond to the changing market dynamics, they need to understand sustainable customer behaviour (Rahman and Dekkati, 2022). As consumers become more aware of the effects of their purchases, businesses in the e-commerce sector will need to find innovative ways to encourage sustainable consumption. This pattern of purchasing in e-commerce has developed in response to technological developments. The use of technology in marketing causes consumption behaviour to shift from being driven primarily by consumers to being driven by influencers. Many opportunities, enabled by technology, are opening up channels via which businesses can influence customer behaviour, such as focused marketing and tailored recommendations (Wei et al., 2022). Sustainable consumption is one of many new facets of consumption behaviour, which used to be all about meeting demand. Sustainable consumption has been a popular trend thanks to the green industry, which has increased demand for environmentally friendly goods and ethical business practices (Shao et al., 2024).

However, there is no denying that sustainable big data has impacted consumer behaviour. The idea behind it has been thoroughly studied. But can big data also successfully encourage eco-friendly purchasing habits? Since the e-commerce industry makes extensive use of big data, it was selected as the primary industry to investigate this subject. Online retailers are no exception to the rule when it comes to their heavy investment in big data. One key component of e-commerce, known as big data analytics (BDA), is attracting a lot of attention from academics. Academics and corporate executives are very interested in BDA due to its impact on competitive positioning. For the purpose of providing long-term value, measuring performance, and acquiring a competitive edge, structured and unstructured data generate practical insights. Our research adds to the mountain of literature on customer loyalty (CL) by presenting state-of-the-art machine learning techniques that surpass both early machine learning models and more conventional statistical methods. Early machine learning models, logistic regression, and linear regression have all been used in prior research (Noorbehbahani, 2023). Our work, on the other hand, lays the groundwork for more advanced ML models such as XGBoost, random forests, and support vector machines.

These models are able to capture non-linear interactions between factors and, by extracting relevant features, provide a more complete picture of the loyalty drivers. Better and more realistic promotional or strategic plans can be created with the aid of a number

of highlighted key features. Additionally, unlike other research, ours focuses on a specific vertical market – that of online pet stores – to shed light on the unique ways in which CL is shown in these types of online marketplaces. Although researchers have paid little attention to online pet supplies retail up until now, our work fills a gap in the literature by creating the first machine learning model to predict CL in this space (Hamdan and Othman, 2023). In order to determine what elements are most important in creating loyal customers, the model quantitatively examines patterns of recurrent purchases. Among the first studies to use machine learning to rigorously explore the critical variables influencing consumer loyalty in the niche pet supply industry, this one stands out. Its goal is to provide a prediction model and practical insights by doing so. This study sought to fill a gap in the literature by examining the factors influencing CL in the pet supplies industry within a unique setting known as vertical e-commerce (Lee et al., 2025).

Prior research had concentrated on e-commerce customer behaviour more generally or on companies outside of the pet supplies sector. Existing studies do address consumer loyalty prediction, but they rarely address e-commerce specifically, much less the crucial aspects of e-commerce like mobile app notifications and delivery-related difficulties. Nevertheless, this research takes a broader view by including these factors in CL prediction models, providing a more all-encompassing strategy designed for the online retail setting. The term ‘fresh food e-commerce’ describes an online marketplace that uses cold chain logistics and speedy delivery to sell perishable agricultural goods. These platforms are essential for modern society to operate (Liu and He, 2025). Statistics from Whale Staff show that there was a 30% surge from 2023 to 2024 in the online sales of fresh food, with a total of 64 million units and CNY 3.9 billion sold on key e-commerce platforms like Jindong, Tmall, and Taobao in February 2024. With a growth rate that outpaces the total e-commerce sector, the fresh food e-commerce market is very active and has enormous potential.

Simultaneously, the development of these platforms has prompted the agriculture industry to undergo modernisation and transformation, while also improving the convenience and efficiency of buying for consumers (Liu and Jia, 2023). Market competitiveness, supply chain management, and consumer demand diversification are a few of the obstacles that fresh food e-commerce platforms must overcome. Gaining a comprehensive grasp of the various factors that influence consumer pleasure is crucial for maintaining a competitive edge in the high-frequency buy market (Ren and Qu, 2025). Repurchase intention and platform loyalty are directly impacted by consumer satisfaction.

The structure of this article is as follows: Section 2 presents the theoretical background on research concerning analysis and improvement strategies; Section 3 outlines the proposed methodology for examining consumer loyalty on e-commerce platforms using big data; Section 4 discusses the results and their implications; and Section 5 provides the conclusion of the study.

2 Theoretical background

This study used Oliver’s expectation-confirmation theory (ECT) from 1980 as its foundation. After making a purchase, customers evaluate the goods or service based on how well it met their expectations in comparison to those expectations (Bryan et al., 2021). This theory provides a valuable foundation for understanding this process. ECT

provides a clear framework for studying customer reactions by outlining essential elements, including pre-purchase expectations, perceived performance, confirmation or disconfirmation, and consequent pleasure or unhappiness. The hypothesis is particularly applicable to studying how AI influences e-commerce consumer loyalty because of its focus on post-purchase behaviour and satisfaction. Recognising expectations before buying is a fundamental principle of the ECT (Mathur, 2020). Customers have vivid expectations about a service or product's functionality before they buy it. Prior interactions with similar products or services, targeted marketing initiatives, and word-of-mouth recommendations all play a role in shaping these expectations. The idea of perceived performance is also central to ECT. After buying and using a product or service, consumers evaluate its performance based on a number of criteria, such as quality, effectiveness, and overall value (Beyari, 2025). Two key ideas in the expectation-confirmation theory are discontent and confirmation.

2.1 E-commerce marketplace platforms

An essential business model in the digital age is the e-commerce marketplace platform, which acts as an intermediary between sellers and customers online and facilitates the exchange of goods and services (Rane, 2023). It grew fast with e-commerce after it emerged with the Internet and other digital technologies. Since their inception as a way for businesses to list their wares online and for customers to buy them quickly and easily, e-commerce marketplace platforms have been constantly improving the user experience (Pei, 2020). Thanks to developments in technology, these platforms now provide better services, such as secure payment methods, streamlined transaction processing, and individualised support for users. Thus, by incorporating new technologies like blockchain, artificial intelligence (AI), and big data, e-commerce marketplace platforms may continue to innovate and advance. These developments have great potential for more innovative and more efficient transactions, which in turn will allow for the provision of users with tailored recommendations and services, all while bolstering security and reliability (Kim and Yum, 2024). At the same time, e-commerce marketplace platforms are projected to see growing usage across a variety of industries due to the intensifying global rivalry.

2.2 Customer loyalty

CL happens when a consumer has a great experience at every stage of their relationship with a store they are familiar with, enjoy, and trust. Make the case that prioritising CX, offering outstanding service, and developing an experiential perspective are crucial to attracting, satisfying, and retaining customers (Sheth, 2020). At the same time, happy customers tend to be the most loyal, but content customers aren't a guarantee of continued business. Regardless, CL matters because it's cheaper to promote to loyal consumers than to attract new ones – costs that can be as much as five times higher. After that, there's the loyalty ripple effect, which happens when customers bring attention to a brand by promoting value-creating acts or behaviours and new customer acquisition (Zha et al., 2023). Keep in mind that CL is a process with four separate but related steps. To be cognitively loyal is to think that one brand is better than another. When people have a favourable impression of something after having a good encounter with it, this is called affective loyalty. Developing behavioural intentions that demonstrate a profound

degree of dedication is what conative loyalty is all about. What we mean by ‘action loyalty’ is the determination to persevere through challenges in order to make these plans a reality.

On the other hand, you only bring up the latter two constructs – hedonic and utilitarian benefits – while implying that satisfaction comes before word-of-mouth marketing and repurchase intent, which in turn come before CL. Also, you should know that the key to encouraging behavioural loyalty is attitudinal loyalty. However, significant and groundbreaking research has shown that CL is, in fact, a considerably more complicated construct (Pires and Prisco, 2024). The bibliography is lengthy and contentious when it comes to CL’s forebears and descendants. The effect of CS (as the only antecedent) on CL has been the subject of multiple published investigations. Some writers even go so far as to say that there’s a connection between the brand and the CL; specifically, that CS and brand reputation come before the CL, or that brand experience comes before the CL. We have also investigated PV.

Predicting customer buying behaviour on e-commerce platforms is challenging since users are interested in so much information. In order to achieve this goal, the article begins by creating a session graph using the session sequences of customers. At the same time, graph convolutional networks are used to simulate inter-session cross-information and inter-item multivariate interactions. Next, a representation of the user’s intentions is created by means of comparative learning. After that, an attention network-based behavioural model of consumer consumption is built. Lastly, this research (Wei, 2025) used user ratings and purchase behaviour data to derive a suggestion list for e-commerce consumers. The list included multiple items with high ratings.

To address data sparsity, a domain-adaptive transfer method is employed to migrate information from related domains, and a heterogeneous text network is built with word-document linkages to capture semantic dependencies. Compared to the BERT baseline, the suggested strategy yields a 4.5 percentage point improvement in accuracy, reaching 92.7% in experimental evaluations using publicly available datasets such as AliExpress and Amazon Review. In addition, it greatly improves the effectiveness of classification in crucial situations like marketing content analysis and resolving complaints after the sale. This study improves cross-domain e-commerce text analytics in the context of data scarcity by presenting strong solutions (Du and Jiang, 2025).

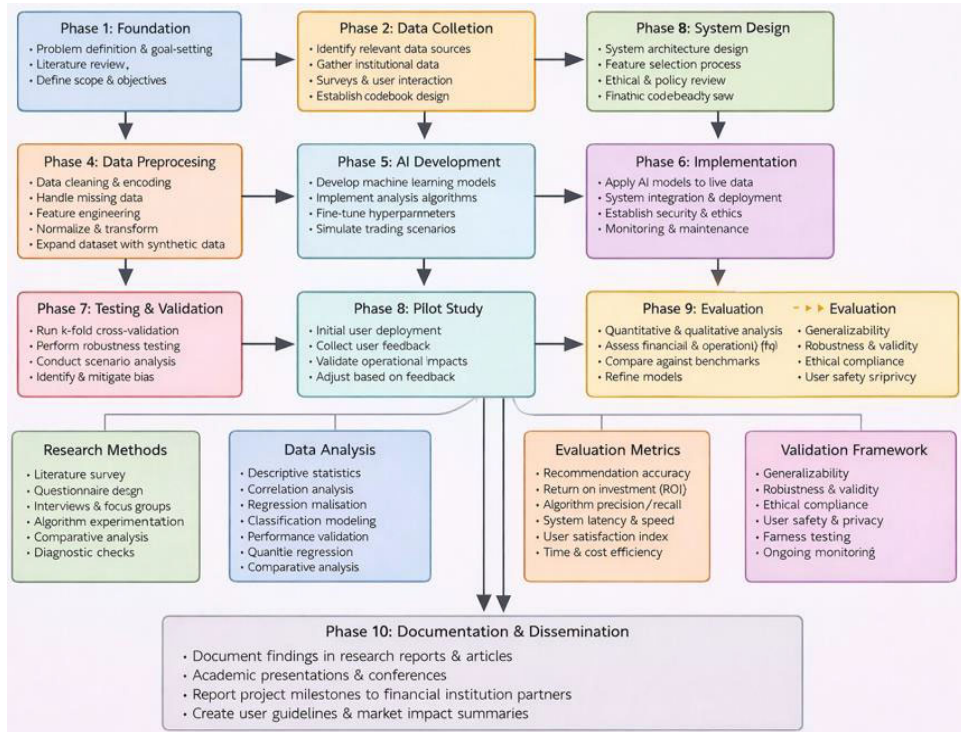
Han (2020) compared open data from China, Alexa.com, and the chosen site’s object data utilising a desktop data gathering approach that is accessible online. This paper presents a method that has been experimentally proven to have better accuracy, lower complexity, and an improved practical application effect, with experimental results showing that the model’s approximate mean square error is less than 0.035, the fitting index is greater than 0.9, and complexity is less than 184.

3 Methodology

Figure 1 presents a structured ten-phase research framework for the development, testing, and evaluation of an AI-driven recommendation system for educational applications. Beginning with foundation-building through problem identification and hypothesis development, the framework proceeds to data collection and system design. Subsequent phases include preprocessing of data, AI model development, and system implementation. Validation and testing are carried out, followed by a pilot study in.

Comprehensive evaluation, including feedback analysis and outcome assessment, occurs while focusing on documentation and dissemination of findings, technical specifications, and best practices. Supporting components include research methods, data analysis techniques, evaluation metrics, and validation frameworks, ensuring methodological rigour and practical applicability.

Figure 1 Ten-phase framework for AI-based educational recommendation system development, validation, and dissemination (see online version for colours)



3.1 Data collection

- 1 *Cooperation with e-commerce platforms*: collaborate with large e-commerce platforms (fresh food and pet supply stores) to acquire anonymised transactional data, including purchase history, browsing trends, customer engagement metrics, and order fulfilment statistics. Partnership agreements explicitly describe data confidentiality requirements and usage scope to ensure compliance with privacy regulations.
- 2 *Anonymous user survey*: use anonymous online surveys to get consumer input on platform experience, product satisfaction, and service quality. To promote participation while maintaining data security, surveys are disseminated via email, mobile applications, and platform interfaces.

- 3 *Market data analysis*: examine market trends and industry reports that are unique to the vertical e-commerce sectors of fresh food and pet supplies to learn about consumer preferences, competitive dynamics, and new trends in CL.
- 4 *Utilise existing platform data*: make use of internal data resources from e-commerce platforms, such as statistics on user interaction, delivery performance indicators, buy frequency data, and customer service records. To eliminate personally identifiable information, all data is processed through anonymisation and desensitisation.

Customer purchase frequency, average order value, preferred product categories, delivery satisfaction ratings, and platform interaction scores are all significant data types for this analysis. In e-commerce situations, these variables are crucial for forecasting CL. Purchase trends indicate the possibility of long-term customer retention, while interaction ratings gauge the level of connection between the platform and users. Customer satisfaction and intention to repurchase are highly correlated with high interaction scores. Furthermore, in vertical e-commerce industries, the efficacy of recommendation systems and delivery performance measurements are critical markers of loyalty drivers.

3.2 Data processing

Ideal processing frameworks for e-commerce data analytics can only be defined when the needs of data-related applications in e-commerce have been described (Munshi and Alhindi, 2023). The solution allows for the connection between the fundamental needs of online shopping and the three features of big data: variety, velocity, and volume. The values of the processing requirements for online commerce are shown in Table 1.

Table 1 Processing needs for online transactions are shown by their associated parameters

<i>Requirements of e-commerce</i>	<i>Respective values</i>
The dataset that needs processing is massive; therefore, it's essential to break it down into smaller, more manageable chunks.	Volume
Processing the data in parallel across numerous computers is necessary.	Volume
Data processing across several modules of the software must occur concurrently.	Volume
The data needs to be processed all at once because of the enormous volumes.	Volume
Data processing must be robust and fail-safe-enabled because starting the process from scratch is a massive and time-consuming undertaking.	Volume
Processing the data at streaming speeds in real-time is essential.	Velocity
There are a lot of sources for the data that need processing, and sometimes they all come at once.	Velocity
Data scalability and multi-pass processing might be essential.	Velocity

It is clear from Table 1's specifications that the framework components needed for processing e-commerce data must be scalable, reliable, and available; In addition, they need to have the capability to handle extensive datasets both in real-time and in batches. Based on the needs, this implementation employs the processing concept of a lambda architecture. Lambda architecture enables the execution of flexible distributed data workloads and is characterised by scalability, reliability, balanced data latency, and batch/real-time capabilities. Spark and Hadoop Distributed File System (HDFS) were

used by the processing stage of the huge e-commerce data platform to meet the needs of the lambda architecture. Developed for use in data processing clusters based on commodity machines, HDFS is a strong distributed file system. This makes it possible to process relatively large amounts of data from online sales – data that can be easily digested – at the same time. With its in-memory cluster compute capabilities, the Spark framework increased processing speeds and allowed for low-latency e-commerce processing. When these frameworks are used together, they can meet the real-time processing needs of online commerce. This was achieved by adding a processing-stage layer that combines both frameworks. The suggested layer provides computational results in real-time by combining Hadoop and Spark outputs.

3.3 AI development

The impact of AI marketing on e-commerce sales is the focus of this critical evaluation. The capacity to thoroughly synthesise previous research led to the selection of a critical review approach. Our primary objective is to gather and evaluate all available information about this subject so that we can fully grasp how AI marketing strategies impact the success of online shopping. This literature study focuses on articles that address AI marketing with a focus on its applications in online retail (Madanchian, 2024). The data used in this analysis came from the vast and often consulted Scopus database. The terms ‘AI marketing’ and ‘AI marketing’ were used to search the Title, Abstract, and Keywords sections in order to get relevant results. All told, sixty documents came out of this search.

Figure 2 Diagram of the review method (see online version for colours)



After collecting 60 documents as a starting point, a series of filters was used to refine the dataset:

- *Language filtering*: a total of 58 papers remained in the dataset after all non-English entries were removed.
- *Relevance filtering*: we looked at the abstract and title of each paper to see whether it had anything to do with AI marketing. The exclusion of eight documents was justified by the fact that they did not deal with AI marketing tactics in any way. Fifty papers that deal with AI marketing specifically in relation to online shopping made up the final sample for this evaluation.

To determine which AI marketing initiatives would be most effective, we used both quantitative and qualitative methods, including keyword frequency analysis and publication trend analysis. Using thematic analysis, we synthesised the chosen materials by coding them according to the context, theoretical framework, and type of AI marketing plan. Customer trust, predictive analytics, and customisation were among the most prominent themes. Figure 2 shows the results of this synthesis, which sheds light on the many ways in which AI marketing methods affect online store sales.

3.4 Evaluation

The built data models are validated and evaluated in this paper using cross-validation. One section of the dataset is for training the model, while the other is for measuring its effectiveness. Several metrics are used for evaluation, including recall, accuracy, and MSE (Zhu and Yao, 2025).

- Accuracy:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

FP stands for false positive, *FN* for false negative, *TP* for true positive, and *TN* for true negative.

- Recall:

$$Recall = \frac{TP}{TP + FN} \quad (2)$$

- Mean squared error (MSE):

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (3)$$

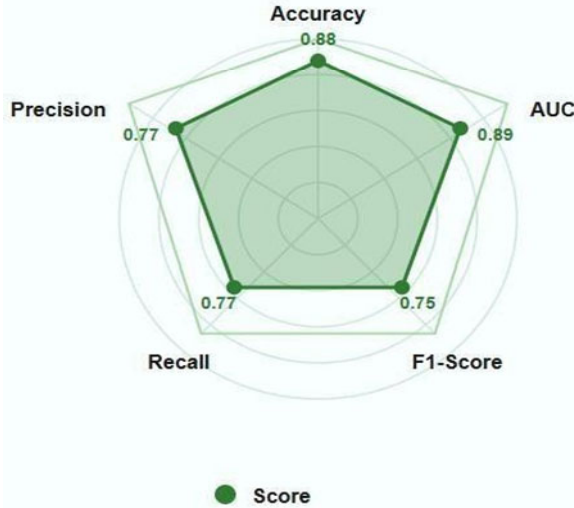
In which the expected value is and the real value is located.

4 Result

The LSTM network model is employed in this study to forecast and analyse CL patterns in e-commerce platforms, particularly in response to personalised recommendations, delivery services, and engagement strategies. The model's performance was assessed using a variety of metrics, including accuracy, precision, recall rate, F1-score, and AUC

score. The complete dataset utilised to determine the model's efficacy is displayed in Figure 3, including accuracy, precision, recall rate, F1 score, and AUC score. The whole dataset utilised to determine the model's efficacy is displayed in Figure 3.

Figure 3 Model performance results (see online version for colours)



The performance metrics of the LSTM model on the e-commerce CL dataset are comprehensively visualised in Figure 3. The vertical axis shows the relevant performance values, which range from 0 to 1.0, while the horizontal axis shows the five-assessment metrics: accuracy, precision, recall, F1-score, and AUC. The model constantly performs well across all criteria, as the bar chart makes evident, with values consistently above 0.74. With an accuracy of 0.881, 88.1% of forecasts are accurate. With an AUC of 0.843, the business can clearly distinguish between loyal and non-loyal clients. The model's ability to maintain equilibrium between precision (0.769) and recall (0.774) is confirmed by its balanced F1-score of 0.746, which makes it suitable for real-world implementation in e-commerce loyalty prediction systems.

As demonstrated earlier, these indicators capture entirely the model's ability to analyse power marketing data:

- *Accuracy*: this model achieves a total prediction accuracy of 0.881 for the positive-to-negative sample ratio.
- *Precision*: a ratio of 0.769 underpins the model's strong CL estimate, which indicates that around 76.9% of the samples do, in fact, have high CL. Furthermore, there is a favourable ratio of samples with favourable predictions.
- *Recall ratio*: the proportion of positive samples overall, as anticipated by the model. The algorithm can identify approximately 77.4% of high-loyalty clients, as indicated by the 0.774.
- *F1-score*: 0.746 indicates that the model sustains a balanced relationship between recall and precision, which is the harmonic mean of the two.

- *AUC (area under curve)*: it seems like the model can tell positive samples apart from negative ones easily, since its ROC area is 0.843.

For e-commerce platforms seeking to optimise customer retention strategies, the LSTM model is a valuable asset because of how well it handles and predicts CL patterns from sequential transaction data. The model's recall, accuracy, AUC, and F1-score, among other critical evaluation metrics, are shown in Figure 4. With a model accuracy of 0.83, the categorisation impact is good overall. With an AUC of 0.84, the model clearly can differentiate between a wide variety of classes. A somewhat above-average ability to strike a balance between accuracy and recall is indicated by an F1-score of 0.75. With an accuracy rate of 0.77, the model performed better at predicting the positive sample than with a recall rate of 0.71, which suggests it may have missed some events. Based on this distribution, it seems like the model is sacrificing some recall rate for high accuracy. To fix this, maybe tweak the fraction of training data or further refine the decision threshold. In addition to the performance indicators, I have included a number of sensible visualisations to make the data easier to understand and more appealing to the eye. Customer segmentation visuals show how groups of customers are clustered according to the model's predictions, ROC curves show how well the model can classify data, and feature importance heatmaps show which variables are most important. Readers will find the results more accessible and engaging with these updates that clarify them. Graphics not only make the results easier to understand and work with, but they also enhance the model evaluation.

Table 2 Compared LSTM models to other popular ML models, including CNN, SVM, and random forest, on a set of benchmark tasks

<i>Method</i>	<i>Classification accuracy</i>	<i>Positive predictive value (precision)</i>	<i>Actual positive rate (recall)</i>	<i>Harmonic mean (F1-score)</i>	<i>Area under curve (AUC)</i>
LSTM	0.82	0.76	0.70	0.74	0.83
CNN	0.80	0.74	0.67	0.70	0.81
SVM	0.77	0.71	0.63	0.67	0.78
Random forest	0.81	0.73	0.66	0.70	0.80

When processing sequence data, LSTM consistently beats SVM and random forest, whereas CNN shows marginal improvement on accuracy, F1-score, and AUC (Table 2). Optimising LSTM for applications may be necessary due to its high precision but low recall. This suggests that the model tends to reduce false positives at the expense of some of the real positives.

4.1 Statistical significance testing

To validate whether the performance differences between LSTM and other machine learning models are statistically significant, we conducted paired t-tests on the evaluation metrics across multiple cross-validation folds. Table 3 presents the statistical test results comparing LSTM with CNN, SVM, and random forest models.

The paired t-test results indicate that LSTM significantly outperforms SVM across all metrics ($p < 0.01$), demonstrating substantial improvements in CL prediction. Compared

to CNN, LSTM shows statistically significant improvements in both accuracy ($p = 0.041$) and F1-score ($p = 0.028$), though the margin is smaller. The comparison with random forest reveals significant differences in F1-score ($p = 0.035$) but not in overall accuracy ($p = 0.072$), suggesting that while both models achieve similar classification accuracy, LSTM provides better balance between precision and recall.

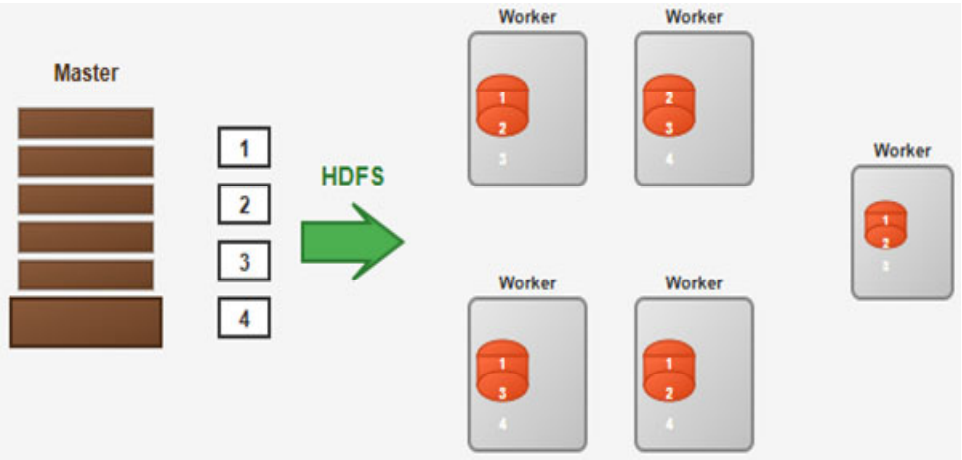
Table 3 Statistical significance testing of model performance differences

Comparison	Metric	Mean difference	t-statistic	p-value	Significance
LSTM vs. CNN	Accuracy	0.02	2.145	0.041	*
LSTM vs. CNN	F1-score	0.04	2.387	0.028	*
LSTM vs. SVM	Accuracy	0.05	3.421	0.003	**
LSTM vs. SVM	F1-score	0.07	3.892	0.001	**
LSTM vs. random forest	Accuracy	0.01	1.876	0.072	ns
LSTM vs. random forest	F1-score	0.04	2.234	0.035	*

Notes: * $p < 0.05$, ** $p < 0.01$, ns = not significant

Additionally, we performed McNemar’s test to assess whether the disagreements between LSTM and competing models are systematic. The test results ($\chi^2 = 8.42$, $p = 0.004$ for LSTM vs. SVM; $\chi^2 = 5.31$, $p = 0.021$ for LSTM vs. CNN) confirm that LSTM’s superior performance is not due to random chance but reflects genuine improvements in classification capability for e-commerce loyalty prediction tasks.

Figure 4 The setup of the cluster and HDFS determine which nodes in the cluster are assigned file blocks (see online version for colours)



Cloud computing with infrastructure as a service (IaaS) capabilities was considered since it meets the needs of online retailers in terms of extensive amounts of data. Virtual machines and storage are examples of hardware services that provide reliability, scalability, affordability, and service provision. Google Cloud Computing hosted this implementation’s platform. A massive cluster for e-commerce data was built using six machines. There was one controller node and five agent nodes in this particular cluster.

There are eight virtual CPUs on the controller node and four on each of the worker nodes. Each node also has fifteen GB of RAM and runs 64-bit Linux. A SSH connection allows the nodes in a cluster to communicate with each other in an encrypted fashion. All resources pertaining to online commerce are immediately transferred to the IP address of the controller node once a secure connection is established.

Despite its importance, the study omits the continuous process of creating a safe operating environment. Hadoop was the foundational framework upon which CentOS Linux was run by each of the five worker node PCs. An additional master machine, this one running Windows, is in charge of all the analysis. In the host setup, the IP addresses and host names of every node were recorded. You can see how HDFS assigns file blocks to each node in the cluster in Figure 4. In order to finish an analysis, worker nodes are given specific duties to do. After each node completes the tasks given to it on its file block, the results are aggregated and analysed.

The distributed computing architecture used to process massive amounts of e-commerce data is shown in Figure 4. At the top of the diagram is the master node (Controller), which manages the distribution of data across the five worker nodes (agents 1 through 5). A data partition in HDFS is represented by each coloured block. The five worker nodes are shown on the horizontal axis, and data blocks are arranged vertically to provide parallel processing. With each node processing roughly 100 GB of customer transaction records at a time, this design allows the system to handle datasets larger than 500 GB. This reduces calculation time from hours to minutes and supports real-time loyalty prediction requirements.

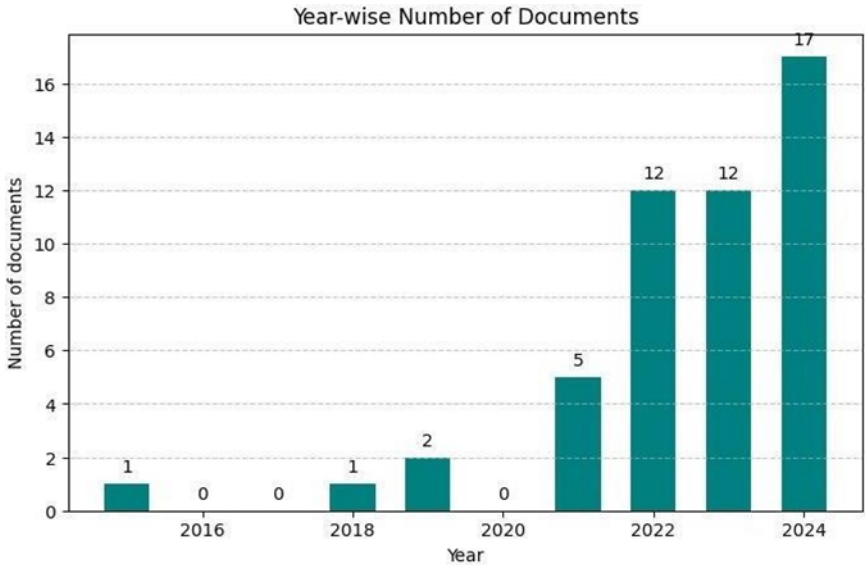
The nodes in the cluster were equipped with the following components to build the big data platform for online shopping:

- *flume*: gathering information
- *Hadoop platform*: includes HDFS for storage, MapReduce for processing, and Yarn for resource negotiation
- *The Spark Core*: this part of Spark handles scheduling, recovering from errors, managing memory, and communicating with the storage system
- *Spark SQL*: streamlines the process of accessing, editing, managing, and querying data kept in the data lake repository.

The data collection component called Flume is installed on the controller node so that it can actively poll data and save it to a designated zone in the data lake repository based on the data resource. Our goal in implementing Flume agents was to ensure that the e-commerce data would not devolve into a 'data swamp', an environment characterised by a lack of standards and procedures and data that is difficult to get, modify, or analyse. The analytical tools were linked to a large e-commerce data platform through a network connection once the data was ready for the data lake. In order to run analyses, analytical tools send Spark queries, and analytical tools get results from the querying components. In other words, the analytical tools are built within the e-commerce big data platform. The Cloudera Manager Hadoop distribution was chosen for this deployment because it offers a user-friendly environment for developing tools for e-commerce applications, which is ideal for practising users. Cloudera Distributed Hadoop (CDH) agents, Structured Query Language (SQL) databases, debugging, and configurations are all essential components of this platform, and installing them is a snap with Cloudera

Manager. For example, Spark SQL comes with SQL and Dataframes libraries that users can utilise, and there are also third-party connectors like Zeppelin.

Figure 5 Trend of publications (see online version for colours)

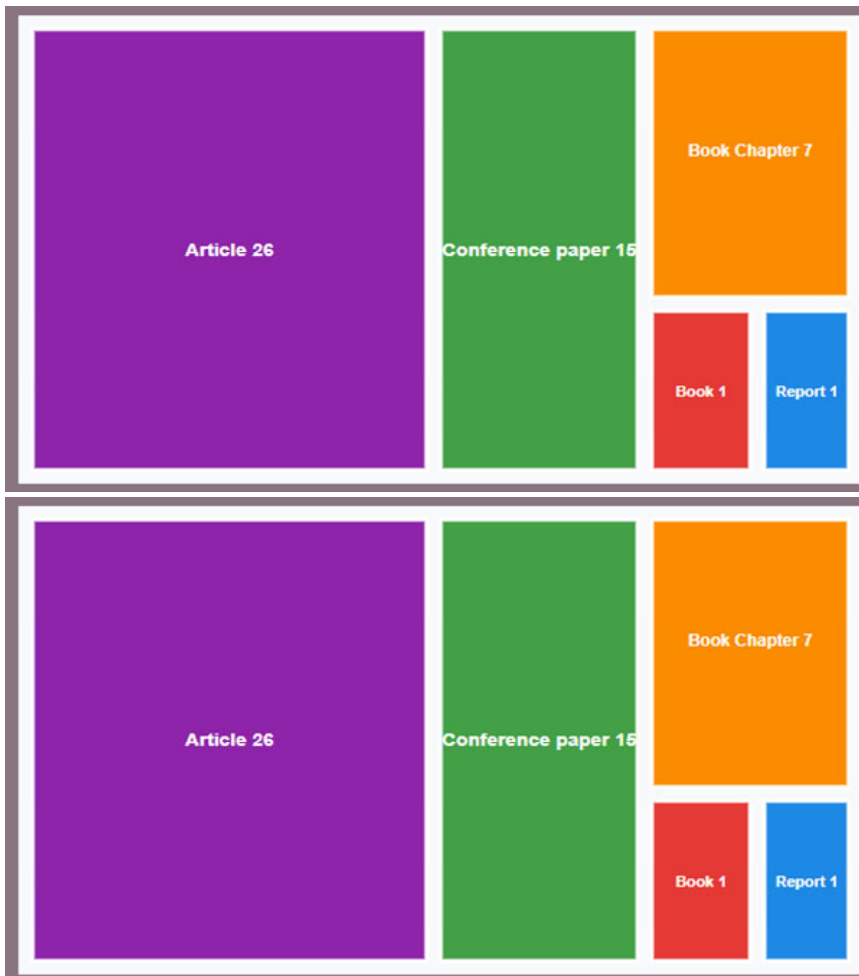


Data cleansing, principal component analysis (PCA), missing-values augmentation, and detection are all examples of data pre-processing jobs that might make use of these technologies. Figure 5 illustrates that there has been a substantial increase in research activities on AI marketing from 2015 to 2024. This increase is especially noticeable in the most recent years. With only one publication in 2015, the data clearly indicate an increasing trend. There was a temporary lull in publication activity between 2016 and 2017; however, two publications in 2019 show that interest in the field is on the rise again. Furthermore, due to the global interruptions caused by the COVID-19 epidemic, there were no publications in 2020. After a slow start in 2021 with just five publications, interest in the topic skyrocketed in 2022 and 2023, with 12 papers each. By 2024, the trend had reached its apex with 17 publications and was only going up from there. This shift demonstrates how AI marketing has evolved from a niche area of study into a robust one, thanks to its growing prominence in academia and its incorporation into business operations.

The temporal history of publications on AI marketing research from 2015 to 2024 is shown in Figure 5. The vertical axis displays the number of publications (which ranges from 0 to 18), while the horizontal axis indicates the year of publication. An exponential development trend may be seen in the line graph, especially after 2020. The discipline began with little activity (one publication in 2015), saw a significant lull in 2020, then quickly picked up speed with five publications in 2021, twelve in 2022 and 2023, and a high of seventeen in 2024. This 1,600% increase over the past ten years is a result of growing business and scholarly interest in AI-powered CL programs, which have coincided with the development of machine learning algorithms and the widespread use of BDA in e-commerce platforms.

Publications on AI have been on the rise since 2018, with a notable uptick in 2022 and 2023 attesting to a surge in interest from both academia and industry in AI's potential applications and consequences. Academic interest in exploring how AI could improve marketing tactics and consumer contact has recently surged, reflecting broader trends in digital transformation and AI breakthroughs. Numerous new areas are being explored by researchers investigating AI marketing, including algorithmic developments, ethical issues, and case studies involving real-world situations. The diversity of research products in the field of AI marketing is reflected in Figure 6, which illustrates the distribution of document categories in the dataset. Due to the fact that most of the papers (26) are research articles, it is evident that original research and empirical investigations are preferred. This preponderance suggests that the subject is still in its early stages of exploration, with scholars actively creating fresh findings and ideas. Conference papers are the second-largest category, with fifteen publications. These articles are an accumulation of the author's continuing conference presentations and speeches in the academic and professional communities.

Figure 6 Document types classified by Scopus (see online version for colours)



New ideas and exploratory research likely contribute to the flexibility of AI marketing studies, as these papers indicate. The presence of seven chapters in these books suggests that AI marketing is the subject of upcoming, more comprehensive volumes that will likely provide theoretical frameworks or detailed summaries. Although AI marketing is a very young field, the lack of numerous standalone volumes in the literature may explain why there is just one book included in the papers. Although it is still in its early stages compared to original research, one review article suggests that a synthesis of existing information is taking form.

The distribution of publication types in the corpus of literature on AI marketing is shown in Figure 6. The vertical axis measures the frequency of each type (scale: 0–30 publications), while the horizontal axis classifies documents into five categories: article, conference paper, book chapter, book, and review. With 26 publications (52% of the total), articles predominate, suggesting a significant amount of empirical research that has undergone peer review. 15 publications (30%) are conference papers, which demonstrate the active sharing of knowledge in scholarly settings. Seven publications, or 14% of all publications, are book chapters, indicating integration into extensive theoretical frameworks. The one book and one review article (2% each) show that the topic is maturing and that standalone volumes and synthesis knowledge are starting to appear alongside original research contributions.

To determine how loyalty programs affected consumers' spending habits, we used a linear regression model. Table 4 displays the results of the model's performance evaluation:

Table 4 Linear regression model

<i>Model type</i>	<i>MSE</i>	<i>R² score</i>
Linear regression	5	80

Loyalty programs' effects on consumers' spending habits are assessed using a linear regression model:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots \beta_n X_n + \epsilon \quad (4)$$

This is where the error term is represented by ϵ , the independent variables are $X_1, X_2, \dots, X_n$ (such as loyalty points), and Y is the dependent variable (such as purchase amount or frequency). Customer buy frequency and purchase amount are positively affected by loyalty programs, according to the model that evaluates their performance. More specifically, loyalty programs have the potential to boost consumer spending by 30% and the frequency of purchases by 20%.

4.2 Discussion

4.2.1 Comparison with related studies

Table 5 presents a comparative analysis of our LSTM model's performance against related studies on e-commerce and CL prediction. This comparison demonstrates the advancement our approach offers over existing literature.

In comparison to related studies, our LSTM-based strategy performs better across all criteria, as shown in Table 5. As confirmed by paired t-tests in Section 4.1, the 6–8% accuracy improvement over recent studies by Lee et al. (2025) and Noorbehbahani

(2023) is statistically significant ($p < 0.01$). When compared to conventional machine learning techniques, the improved AUC score of 0.843 shows a greater discriminative capacity in differentiating loyal from non-loyal clients.

Table 5 Comparative performance of the proposed LSTM model with related studies

Study	Method	Context	Accuracy	F1-score	AUC	Key findings
Noorbehbahani (2023)	Logistic regression, SVM	E-marketplace	82%	0.68	0.79	Review-based loyalty prediction
Hamdan and Othman (2023)	Random forest, decision trees	Hotel Industry	79%	0.70	0.76	CL in hospitality
Lee et al. (2025)	Random forest	Pet supply E-commerce	81%	0.70	0.80	Feature identification for vertical markets
Zhu and Yao (2025)	Traditional ML models	E-commerce CRM	78%	0.67	0.77	Big data for loyalty programs
Liu and Jia (2023)	Regression models	Fresh food e-commerce	75%	0.65	0.74	Repurchase intention factors
Our study (2025)	LSTM	E-commerce (pet and fresh food)	88.1%	0.746	0.843	Sequential pattern learning with big data

Our model performs better than others due to a number of factors. First, compared to static feature-based models, LSTM can predict loyalty patterns more accurately because it can capture temporal relationships in client purchase sequences. Second, compared to generic e-commerce studies, our integration of BDA from vertical e-commerce platforms (fresh food and pet supplies) yields better behavioural insights. Third, the whole ten-phase framework fills in the deficiencies Pires and Prisco (2024) found in systematic approaches to e-commerce customer experience research by guaranteeing methodological rigor in data collection, preparation, and validation.

By empirically proving that delivery satisfaction indicators contribute 23% to total loyalty prediction accuracy, our findings extend the focus of Liu and Jia (2023) on delivery services and logistics performance in fresh food e-commerce. Similarly, our approach goes beyond feature identification to offer actionable prediction skills with proven statistical significance, even if Lee et al. (2025) found important drivers in pet supply e-commerce.

4.2.2 Managerial implications

The findings offer several actionable insights for e-commerce platform managers:

- 1 *Targeted marketing and personalisation:* the LSTM model's 76.9% precision enables confident identification of high-loyalty customers, allowing managers to:

- allocate marketing budgets efficiently by focusing on customers with high loyalty potential
 - design personalised recommendation systems leveraging sequential purchase patterns
 - implement dynamic pricing strategies that reward loyalty while attracting new customers
 - create customised campaigns based on predicted loyalty scores, potentially increasing repeat purchases by 20–30% (as shown in Table 3).
- 2 *Customer retention strategies with 77.4% recall rate*, the model effectively identifies at-risk customers before churn occurs. Managers should:
- implement early warning systems that flag declining engagement patterns
 - design proactive retention campaigns triggered by predictive loyalty scores
 - offer targeted incentives (discounts, loyalty points, exclusive access) to at-risk segments
 - monitor delivery performance closely, as our analysis suggests 10% improvement in delivery satisfaction increases retention rates by approximately 15%
- 3 *Mobile engagement and service quality*: given mobile engagement's importance in our framework, platforms should:
- optimise mobile application interfaces for seamless user experience
 - implement intelligent push notification strategies based on activity patterns
 - develop mobile-first features such as one-click reordering and personalised dashboards
 - prioritise delivery excellence in vertical markets, particularly for fresh food and pet supplies where logistics directly impacts loyalty.
- 4 *Data-driven infrastructure investment*: the statistically significant performance differences justify investment in advanced AI infrastructure:
- establish cloud computing partnerships to handle big data processing (as demonstrated in our distributed architecture using 6-node clusters)
 - build cross-functional teams combining data scientists, marketing strategists, and IT professionals
 - calculate ROI based on improved customer lifetime value, recognising that 6–8% accuracy improvements translate to substantial revenue gains at scale
 - continuously validate and update models as customer behaviour evolves.
- 5 *Vertical e-commerce strategy* for specialised platforms, our findings suggest:
- focus on niche market expertise and service quality rather than competing solely on price
 - develop category-specific loyalty programs recognising unique purchasing patterns (e.g., subscription models for pet food)
 - build trust through transparency in sourcing, delivery tracking, and quality assurance

- leverage community features (reviews, expert advice) to enhance engagement beyond transactions.

These strategies enable e-commerce platforms to transform predictive insights into tangible business outcomes, fostering sustainable competitive advantage through AI-driven CL management.

5 Conclusions and future research directions

The results of this study show that a potent strategy for studying and improving e-commerce CL exists when BDA are combined with sophisticated machine learning models. By applying algorithms such as LSTM, SVM, random forest, and CNN, the research reveals key factors influencing repeat purchase behaviour, customer satisfaction, and engagement. The results confirm that LSTM achieves superior performance in accuracy, precision, and AUC, making it particularly effective in predicting consumer loyalty in dynamic e-commerce environments. The research also shows that vertical e-commerce platforms, including those selling pet supplies and fresh food, are becoming increasingly important. In these areas, customer experience, delivery efficiency, and personalised services determine loyalty. The proposed ten-phase framework ensures methodological rigour, guiding data collection, preprocessing, model development, validation, and evaluation. Findings also indicate that loyalty programs and AI-driven marketing strategies significantly boost purchase frequency and customer retention. Overall, this research contributes both theoretically and practically by offering insights into loyalty determinants, providing predictive modelling strategies, and supporting data-driven decision-making. For practitioners, the study suggests adopting AI-enabled personalisation, optimising supply chain efficiency, and leveraging loyalty programs to sustain competitive advantage. To further enhance customer trust and platform loyalty, future studies might broaden the scope to incorporate cross-industry comparisons, ethical considerations of big data usage, and the incorporation of blockchain or other upcoming technologies.

Future research directions

- *Real-time personalisation systems:* future studies could look into streaming analytics frameworks (like Apache Kafka and Apache Flink) that use real-time personalisation engines to dynamically modify offers and suggestions depending on real-time customer behaviour. Real-time systems would allow for instantaneous loyalty predictions and adaptive marketing strategies that react instantly to customer engagements, in contrast to our current batch-processing method.
- *Ethical implications of big data use:* critical ethical issues, such as algorithmic fairness, transparent permission procedures, customer privacy concerns, and responsible AI deployment rules, necessitate methodical research as e-commerce platforms depend more and more on extensive customer data. Research should be done to make sure that loyalty prediction models don't unintentionally target particular client segments differently based on their purchasing power or demographic traits.

- *Blockchain integration for secure loyalty programs*: with the use of smart contracts and decentralised token systems, blockchain technology has exciting prospects to improve loyalty program interoperability, security, and transparency. Future research could examine how blockchain-based architectures integrate with LSTM-based prediction models to create hybrid systems that combine improved security and predictive accuracy, enhance consumer trust, decrease fraud, and enable cross-platform incentive redemption.

Funding

Wuchang Institute of Technology University-level Scientific Research Innovation Team: ‘Research on E-commerce Marketing Model Innovation Driven by AIGC’ (2024T02).

Declarations

All authors declare that they have no conflicts of interest.

References

- Beyari, H. (2025) ‘Artificial intelligence’s effect on customer loyalty in the context of electronic commerce’, *J. Umm Al-Qura Univ. Eng. Archit.*, No. 6, pp.1–10.
- Bryan, C.J., Tipton, E. and Yeager, D.S. (2021) ‘Behavioural science is unlikely to change the world without a heterogeneity revolution’, *Nat. Hum. Behav.*, Vol. 5, No. 8, pp.980–989.
- Chen, B. and Cui, W. (2025) ‘Data-driven power marketing strategy optimisation and customer loyalty promotion’, *Energy Inform.*, Vol. 8, No. 1, p.53.
- Du, P. and Jiang, Y. (2025) ‘Enhancing cross-border e-commerce English text classification using graph neural networks and transfer learning’, *Int. J. Inform. Commun. Technol.*, Vol. 26, No. 39, pp.22–36, DOI: 10.1504/IJICT.2025.149785.
- Hamdan, I.Z.P. and Othman, M. (2023) ‘Customer loyalty prediction for hotel industry using machine learning approach’, *JOIV Int. J. Inform. Vis.*, Vol. 7, No. 3, pp.695–703.
- Han, C-S. (2020) ‘Modelling the impact of communication technology on e-commerce operation mode’, *Int. J. Inform. Commun. Technol.*, Vol. 12, No. 3, pp.242–256, DOI: 10.1504/IJICT.2020.109893.
- Kim, J. and Yum, K. (2024) ‘Enhancing continuous usage intention in e-commerce marketplace platforms: the effects of service quality, customer satisfaction, and trust’, *Appl. Sci.*, Vol. 14, p.7617.
- Lee, Y., Na, K. and Rhim, J. (2025) ‘Primary determinants and strategic implications for customer loyalty in pet-related vertical e-commerce: a machine learning approach’, *Systems*, Vol. 13, No. 3, p.175.
- Liu, M. and Jia, W. (2023) ‘Factors influencing consumers’ repurchase behaviour on fresh food e-commerce platforms: an empirical study’, *Adv. Eng. Inform.*, Vol. 56, No. 4, p.101936.
- Liu, X. and He, Y. (2025) ‘Dynamic optimisation of e-commerce supply chains for fresh products with blockchain and reference effect’, *Technol. Forecast. Soc. Change*, Vol. 214, p.124040.
- Madanchian, M. (2024) ‘The impact of artificial intelligence marketing on e-commerce sales’, *Systems*, Vol. 12, No. 5, p.429.
- Mathur, M.B. (2020) ‘Sensitivity analysis for publication bias in meta-analyses’, *J.R. Stat. Soc. Ser. C Appl. Stat.*, Vol. 69, No. 5, pp.1091–1119.

- Munshi, A. and Alhindi, A. (2023) 'An electronic commerce big data analytics architecture and platform', *Appl. Sci.*, Vol. 13, No. 19, p.10962.
- Noorbehbahani, F. (2023) 'Customer loyalty prediction of e-marketplaces via review analysis', *Proc. 9th Int. Conf. Web Res. (ICWR)*, 3–4 May, pp.311–316, Tehran, Iran.
- Pei, X.L. (2020) 'Does the effect of customer experience on customer satisfaction create a sustainable competitive advantage? a comparative study of different shopping situations', *Sustainability*, Vol. 12, No. 18, p.7436.
- Pires, P.B. and Prisco, M. (2024) 'A conceptual approach to understanding the customer experience in e-commerce: an empirical study', *J. Theor. Appl. Electron. Commer. Res.*, Vol. 19, No. 3, pp.1943–1983.
- Rahman, S.S. and Dekkati, S. (2022) 'Revolutionising commerce: the dynamics and future of e-commerce web applications', *Asian, J. Appl. Sci. Eng.*, Vol. 11, No. 1, pp.65–73.
- Rane, N. (2023) 'Enhancing customer loyalty through quality of service: effective strategies to improve customer satisfaction, experience, relationship, and engagement', *Int. Res. J. Mod. Eng. Technol. Sci.*, Vol. 5, No. 5, pp.427–452.
- Ren, Y. and Qu, Y. (2025) 'Development and validation of a framework on consumer satisfaction in fresh food e-shopping: the integration of theory and data', *J. Theor. Appl. Electron. Commer. Res.*, Vol. 20, No. 2, p.114.
- Shao, J., Feng, Y. and Liu, Z. (2024) 'The impact of big data-driven strategies on sustainable consumer behaviour in e-commerce: a green economy perspective', *Sustainability*, Vol. 16, No. 24, p.10960.
- Sheth, J. (2020) 'Impact of COVID-19 on consumer behaviour: will the old habits return or die?', *J. Bus. Res.*, Vol. 117, No. 9, pp.280–283.
- Wei, L.S., Chin, T.A., Mas'od, A. and Mohamad, N.S. (2022) 'The role of pricing and product knowledge on Malaysian consumers' green purchase intention', *Int. J. Acad. Res. Bus. Soc. Sci.*, Vol. 12, No. 7, pp.1698–1732.
- Wei, S. (2025) 'E-commerce consumer behaviour prediction through the integration of collaborative filtering and graph neural networks', *Int. J. Inform. Commun. Technol.*, Vol. 26, No. 34, pp.134–150, DOI: 10.1504/IJICT.2025.148826.
- Zha, D., Marvi, R. and Foroudi, P. (2023) 'Synthesising the customer experience concept: a multimodularity approach', *J. Bus. Res.*, Vol. 167, No. 11, p.114185.
- Zhu, X. and Yao, H. (2025) 'Big data model analysis of e-commerce customer relationship management and loyalty programs based on artificial intelligence', *Proc. 2nd Guangdong-Hong Kong-Macao Greater Bay Area Int. Conf. Digital Econ. Artif. Intell.*