

## *Retraction*

# **Retracted: Analysis of E-Commerce Marketing Strategy Based on Xgboost Algorithm**

### **Advances in Multimedia**

Received 12 December 2023; Accepted 12 December 2023; Published 13 December 2023

Copyright © 2023 Advances in Multimedia. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This article has been retracted by Hindawi, as publisher, following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of systematic manipulation of the publication and peer-review process. We cannot, therefore, vouch for the reliability or integrity of this article.

Please note that this notice is intended solely to alert readers that the peer-review process of this article has been compromised.

Wiley and Hindawi regret that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.



The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

### **References**

- [1] H. Chen and W. Wan, "Analysis of E-Commerce Marketing Strategy Based on Xgboost Algorithm," *Advances in Multimedia*, vol. 2023, Article ID 1247890, 9 pages, 2023.

## Research Article

# Analysis of E-Commerce Marketing Strategy Based on Xgboost Algorithm

Hong Chen <sup>1,2</sup> and Wei Wan <sup>3</sup>

<sup>1</sup>*Institute of Petroleum Engineering, Yangtze University, Wuhan 430000, Hubei, China*

<sup>2</sup>*The National University of Malaysia, Bangi, Selangor 43600, Malaysia*

<sup>3</sup>*Nanfeng College Guangzhou, Guangzhou 510000, Guangdong, China*

Correspondence should be addressed to Wei Wan; [wanw@nfu.edu.cn](mailto:wanw@nfu.edu.cn)

Received 16 August 2022; Revised 15 September 2022; Accepted 25 November 2022; Published 29 May 2023

Academic Editor: Tao Zhou

Copyright © 2023 Hong Chen and Wei Wan. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

With the development and popularization of multimedia technology, people's communication and information dissemination become more and more convenient. The use of new media technology in marketing accelerates the enterprise's adaptation to market changes. With the help of new marketing technology, the quality of marketing can be improved and the effect of marketing can be maximized. This paper reviews the current literature on e-commerce marketing and then analyzes the feasibility of precision marketing in e-commerce market in the new media era. In order to screen potential consumers and improve the success rate of precision marketing, this paper establishes a prediction model for precision marketing of bank credit cards. A public dataset selected in this paper contains the basic characteristics of the customer, such as age, gender, monthly income, monthly consumption, and whether the customer has agreed to process the credit card after personalized recommendation. After data cleaning and preprocessing, XGboost algorithm is used to predict whether the customer will process the credit card. The calculation shows that the XGboost algorithm can still maintain a high accuracy when dealing with fewer characteristic variables. This study is helpful to better predict users' consumption intention, further explore potential customers, and improve the success rate of precision marketing.

## 1. Introduction

With the continuous popularity of Internet applications, people's clothing, food, and living have undergone earth-shaking changes. For e-commerce, e-commerce combines the unique convenience of the Internet to change the traditional sales methods to a certain extent and promotes the development of the industry [1]. As of December 2021, the size of Chinese netizens reached 1.032 billion, and the Internet penetration reached 73.0%. According to relevant statistics, the annual e-commerce volume reached 34.81 trillion yuan in 2019, an increase of 6.7% over 2018.

Various data show that the business scale of e-commerce is not only growing rapidly but also has penetrated into many fields such as economy, people's livelihood, and culture. Big data technology is a new technology developed

with the development of computer technology, known as the oil in the information age [2].

The problems faced by precision marketing in the new media era mainly include the following:

- (1) Blind e-commerce marketing mode: The traditional marketing model has a direct relationship with the experience of marketing personnel. It is not targeted enough to analyze the actual needs of consumers, nor to establish a good product demand and audience relationship, which makes e-commerce marketing more blind. According to incomplete statistics, during the development of e-commerce in China, the losses caused by blind sales account for about 30% of the total economic amount, and the profit rate has also decreased by about 40%. It is unable to provide efficient and fast

services for consumers, which makes consumers resist and leads to certain economic losses for e-commerce enterprises[3].

- (2) Low level of e-commerce marketing service: Through the analysis of the traditional e-commerce marketing process, it is found that the marketers cannot fully understand the basic needs of consumers, nor can they analyze the service intentions of consumers. Therefore, it is unable to provide effective and active marketing services for consumers, which makes consumers lack of correct understanding of products, and the purchase rate continues to decrease.
- (3) Lack of accurate analysis of users' consumption habits and behaviors: The data of consumers is huge, and the traditional e-commerce model cannot conduct in-depth analysis and research on these data. Therefore, marketers cannot scientifically judge the basic needs of consumers, nor can they provide more accurate marketing services for consumers, resulting in the continuous decline of their purchase rate [4]. For example, at the node of WeChat sales, it only publishes a circle of friends regularly or irregularly, or transmits product information and preferential activities to customers. Although the coverage is relatively large, the pertinence is not strong. In the long run, it will lead to boredom of most customers and fail to achieve good marketing results.

In addition, the advantages of e-commerce precision marketing in big data environment are mainly included:

- (1) Improve the accuracy of the target group of e-commerce marketing: The user will not change greatly in the short term after consumption. Therefore, we can analyze customers' consumption behavior with the help of advanced big data technology, so as to conduct more accurate marketing. Through the analysis of changes in consumer behavior, we can change the strategy of precise marketing to consumers in a timely manner, so that the marketing direction can fully meet the basic needs of customer groups, thereby improving the comprehensive economic benefits of e-commerce enterprises [5].
- (2) Reduce the marketing cost of e-commerce: At the stage of operation and development, e-commerce enterprises can carry out marketing activities more accurately relying on advanced big data technology. This kind of marketing method is different from the traditional marketing mode. It can not only solve the problem of large human cost investment but also significantly reduce the marketing cost. The use of traditional marketing methods may consume a large amount of manpower and material resources and cannot promote the significant growth of turnover in a short time. Therefore, e-commerce precision marketing based on big data environment can divide

users in detail and propose targeted marketing strategies. At this stage, the marketing strategy and scheme can also be continuously adjusted in combination with the results of customer feedback.

- (3) Easy to associate with the audience: At the stage of using big data technology, e-commerce enterprises not only need to analyze the audience but also need to study the specific needs of users. According to the analysis results, it is not difficult to understand that it is very important to do a good job of correlation analysis. For example, we should analyze the types of goods in the user's shopping cart, build relevant correlations, and analyze the user's actual needs, consumption habits, consumption ability, and so on. At the same time, it is also necessary to explore the recent needs of customers. Through the information provided by big data, special marketing personnel will accurately analyze the behavior habits of consumers, so as to make relevant associations with the audience customers and carry out targeted follow-up personalized marketing activities.

Through big data technology, the purchase behavior and consumption habits of customers are analyzed, so as to achieve accurate marketing for e-commerce platforms and improve the reliability of marketing [6]. This paper conducts a research study on the precision marketing strategy of e-commerce in the large data environment in order to tap potential customers for e-commerce enterprises and improve the success rate of precision marketing of e-commerce enterprises.

## 2. Literature Review of E-Commerce Marketing

On the basis of traditional marketing theory and recent related research results, this paper makes a literature review from the following three aspects: customer acquisition, customer value mining, and personalized marketing.

*2.1. Research Based on Acquiring Customers.* The latest research on how to obtain customers has turned to the Internet, social media, and big data. De Vries et al. found that the user behavior on social media and network forums such as Facebook and twitter, as well as the quantity and quality of message transmission among users, have become the main force affecting brand communication. Digital channels represented by social media have become an important way for enterprises to obtain customers [7]. Ponzoa and Erdmann studied various influencing factors of e-commerce websites in attracting customers, and demonstrated that network analysis methods play an important role in acquiring customers [8]. Britt followed how an enterprise used big data and AI technology to acquire and maintain customers in the form of case studies. Through this study, the author found that customers obtained through digital channels are better than those obtained through traditional offline channels in terms of customer value and brand loyalty [9].

*2.2. Research Based on Segmented Customers.* Wendell Smith first put forward the idea of market segmentation. The birth of market segmentation comes from the segmentation and cutting of market user groups according to several factors when enterprises respond to different needs of different users. The classic “decentralized market” theory is developed on the basis of this theory.

Foreign research in this area also shows a trend towards big data and artificial intelligence. Dogan et al. studied the application case of customer data segmentation by using the RFM (recency, frequency, and monetary) model. This study shows that compared with traditional demographic-based customer segmentation, the results of customer segmentation generated by the RFM model can more effectively support enterprises to carry out precision marketing activities [10]. Johansson and Wikstrom studied the application of machine learning technology in customer segmentation. The study compared the performance of logistic regression model and support vector machine in customer clustering. The results showed that the accuracy rate of logistic regression model in dividing users into three groups reached 71%, which was higher than the accuracy rate of manual classification by experts [11]. Wang studied the use of data mining technology to generate consumer attribute tags. This research generates a dynamic “user portrait” model by extracting static tags and dynamic tags and combining them with user data on the network platform. On this basis, it proposes strategies for personalized recommendation such as mastering user psychology, matching user needs, and accurately adjusting prices [12].

Domestic research on this aspect started late, but it has been steadily advancing. Jing comprehensively summarized the strategy of online precision marketing based on big data, defined the concept of big data precision marketing, summarized the common problems of traditional online marketing, and finally put forward the common strategies of precision marketing based on big data [13]. Zhao et al. studied a predictive recommendation algorithm based on Bayesian network, which calculates the posterior probability of possible consumption by constructing the prior probability of user behavior. Through the data test of the real browsing log of the credit card app, the research verified that the algorithm has higher accuracy than the traditional marketing model [14]. Cao and Zhang studied the precision marketing strategy based on the user portrait system from the aspects of the construction of the user portrait, the design of the precision marketing system, and the precise promotion of the product advertisement [15].

*2.3. Research on Personalized Recommendation.* In the book *Principles of Marketing*, Kotler et al. put the two contents of “Internet” and “precision” together for research, and pointed out that cutting-edge technology will become a powerful tool to support marketing. They encouraged enterprises to use the Internet, multimedia, and other means to carry out marketing activities: “understand the marketing connotation, advocate personalized communication, take accuracy as the starting point and end point, and connect time, users,

content, and response from beginning to end [16].” Farris and Bendle also emphasized the importance of digital channels and the quantification of marketing communication. They studied how enterprises use technical means to track and quantify the results of communication, so as to facilitate enterprises to comprehensively analyze the effects of marketing activities [17]. Economists Benaziz and Lehrer studied the psychology and behavior of consumers in the “multiscreen” era and pointed out that the accurately used media resources will create new business opportunities for enterprises [18].

Huang studied several common marketing forms based on WeChat official account. He believes that the traditional mass media communication is a kind of “common” and “fuzzy” communication, which can only roughly distinguish the audience and cannot be “accurate.” The advertising communication based on WeChat’s official account can accurately analyze the content of fans in the background of WeChat, so as to achieve personalized and customized information communication [19]. Hu and Lu studied the precise marketing strategy based on social media. In the article, she elaborated the role of social media on sales channels in terms of visits, exposure rate, related information, search engine, and so on, which can better realize customer relationship management and provide purchase decisions [20].

The problem of data imbalance is common, so in order to improve the impact of unbalanced data on the model’s prediction results. SMOTE-Regular, SMOTE-Borderline, and other algorithms are derived according to different starting angles of samples. From the perspective of algorithm processing, the cost-sensitive learning algorithm is mainly processed to adapt to the different characteristics of the cost of misjudgment in different situations, and its implementation usually includes reforming the learning model, adjusting the weight, and so on. Among them, the classical logistic model is often used, and because its logit function is symmetric, it is often used in combination with other models when dealing with unbalanced data. In addition, ensemble learning is often used for improvement. The ensemble learning algorithm combines multiple weak classifiers into strong classifiers in a certain way and then synthesizes the results of the base classifiers to get the final prediction result. The most common ensemble learning algorithms are Random Forest (RF), XGBoost, and Adaboost.

### **3. Empirical Analysis of Precision Marketing Based on Xgboost Algorithm**

As the most core member of China’s financial system, commercial banks have created great impact on the financial industry and even the economic society with the emergence of the banking e-commerce platform. As major state-owned commercial banks and large joint-stock commercial banks have launched their own e-commerce platforms, this has a certain impact on the e-commerce business of other sales enterprises. As more and more commercial banks begin to engage in e-commerce business, the proportion of e-commerce business income in the total business income of

commercial banks is also increasing, and e-commerce business has gradually become the focus of commercial banks [21].

Although the business scale of e-commerce platform built by commercial banks is growing, its overall operation effect is still not ideal. Alibaba, Tencent, and other Internet companies continue to make use of their accumulated customer advantages and use e-commerce platforms to launch alternative products for traditional banking business, which has seriously affected the profitability of commercial banks. Based on this, this paper constructs the accurate marketing prediction model of bank credit card. By analyzing the dataset of customers, this paper combines the XGboost algorithm to predict whether the customers handle credit cards, and calculates the importance of each characteristic variable to screen the most important characteristic variable used in precision marketing, and then excavate potential customers, which helps to improve the success rate of precision marketing.

**3.1. Selection and Analysis of Datasets.** The dataset studied in this paper is selected from the credit card marketing dataset published by a website. As the original data set contains more than 100 million data samples, limited by the equipment conditions, this paper reads 1,000 pieces of customer data for credit card marketing, among which the characteristic variables include the customer's age, gender, monthly income, monthly consumption, and the proportion of monthly income to monthly consumption. The target variable is whether the customer responds after precise marketing, that is whether the customer has handled a credit card. A value of 1 indicates that marketing is successful, and a value of 0 indicates that marketing is unsuccessful.

Some data are shown in Table 1.

Firstly, the correlation analysis is carried out on the indicators in the dataset, and the Spearman's correlation coefficient is used to obtain the correlation coefficient of each indicator. From Table 2, it can be seen that the correlation between the monthly income and monthly consumption is strong, and the correlation between other indicators is weak.

In order to better understand the data, this paper makes a visual analysis of all characteristic variables in the dataset. Figure 1 is a scatter diagram of age distribution, showing the age distribution of consumer customers in the dataset samples.

It can be seen from Figure 1 that in the selected dataset, the age of consumer customers is mostly between 20 and 40 years old, with an average age of 29.6 years.

The second indicator is the monthly income of customers. It can be seen from Figure 2 that the monthly income of all customers is concentrated in the range of 14,844 yuan to 16,644 yuan. The proportion of customers whose monthly income exceeds 30,000 yuan does not exceed 10%.

The third indicator is the monthly consumption of customers. From Figure 3, it can be seen that the monthly consumption of all customers is concentrated at 10,000 yuan, the minimum is not less than 4,000 yuan, and the maximum is not more than 15,000 yuan.

The fourth indicator is the age data distribution of customers. It can be seen from Figure 4 that the gender distribution is relatively balanced and there is not much difference.

It can be seen from the distribution of the proportion of monthly consumption to monthly income in Figure 5 that the proportion of monthly consumption to monthly income is concentrated between 0.51 and 0.74. There are many sample data in this range, and the highest proportion is not more than 0.86. The selected sample data does not include moonlight family customers.

Figure 6 shows whether the customer has handled a credit card after precise marketing of the 1,000 pieces of customer data for credit card marketing. The number 1 indicates that the customer successfully handled the credit card after the precision marketing, while the number 0 indicates that the customer still did not handle the credit card after the precision marketing. It can be seen from Figure 6 that the number of customers who handle credit cards after precision marketing is less than that of those who do not handle credit cards.

### 3.2. Data Preprocessing

**3.2.1. Standardization of Data.** In the data shown in Table 1, there is a large difference in the value of each column of data. In order to avoid the influence of different units and values on the machine learning algorithm, the data of the above five indicators are standardized. The formula for standardization processing is the min-max formula, which converts the values of all characteristic indicators to between 0 and 1 [22].

$$Z = \frac{X - \min}{\max - \min}. \quad (1)$$

**3.2.2. Balance of Sample Data.** It can be seen from Figure 6 that the number of customers who have processed credit cards and those who have not processed credit cards after precise marketing is different, among which the number of customers who have not processed credit cards is large and accounts for a high proportion. If the data is directly brought into the model for calculation, there will be errors in the results. When the dichotomous data in the project is not balanced, the output of the model may be affected. The output categories of many models are based on the threshold. For example, in the logistic regression, those less than 0.5 are negative examples, and those greater than 0.5 are positive examples. When the data is unbalanced, the default threshold will cause the model output to tend toward the category with more data. Therefore, it is necessary to balance the number of the two types of samples.

Sampling is a common method for processing unbalanced data, including undersampling and oversampling [23]. Undersampling discards a large amount of data, resulting in overfitting. NearMiss is essentially a prototype selection method, which selects the most representative samples from most class samples for training, mainly to alleviate the problem of information loss in random

TABLE 1: Customer data of credit card marketing.

	Age	Monthly income	Monthly consumption	Gender	Consumption/income	Respond
0	30	7275	6062	0	0.83	1
1	25	17739	13648	0	0.77	1
2	29	25736	14311	0	0.56	1
3	23	14162	7596	0	0.54	1
4	27	15563	12849	0	0.83	1
5	30	15765	8217	0	0.52	1
6	23	14286	9379	0	0.66	1
7	29	18770	11059	0	0.59	1
.....						

TABLE 2: Correlation coefficient of each index.

	Age	Monthly income	Monthly consumption	Gender	Consumption/income	Respond
Age	1.000	-0.017	-0.103	0.047	-0.153	-0.281
Monthly income	-0.017	1.000	0.851	-0.032	-0.368	0.100
Monthly consumption	-0.103	0.851	1.000	-0.091	0.149	0.369
Gender	0.047	-0.032	-0.091	1.000	-0.106	-0.202
Consumption/income	-0.153	-0.368	0.149	-0.106	1.000	0.447
Respond	-0.281	0.100	0.369	-0.202	0.447	1.000

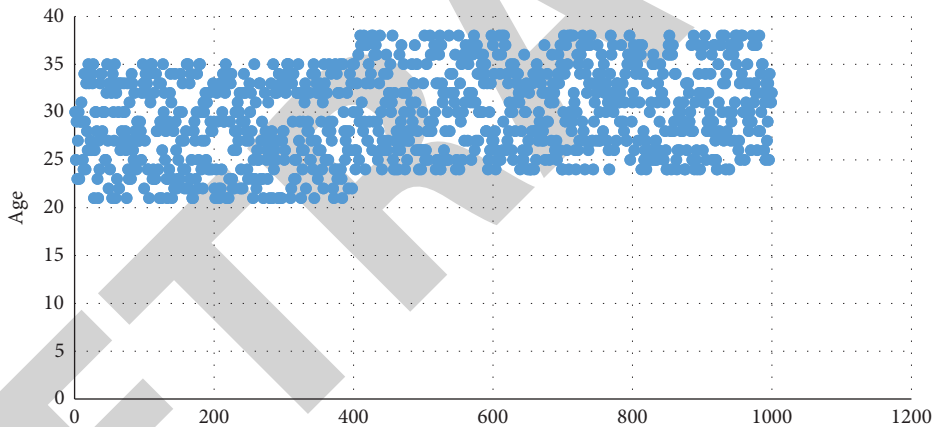


FIGURE 1: Scatter diagram of age distribution.

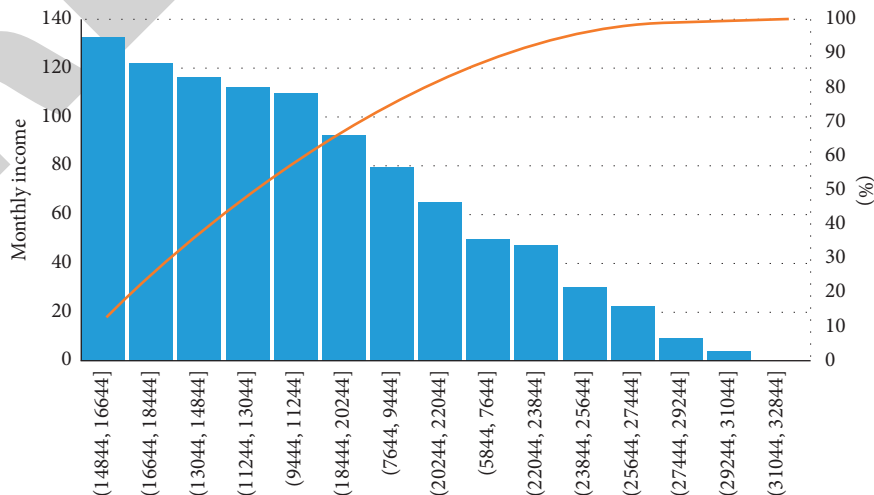


FIGURE 2: Distribution of monthly income.

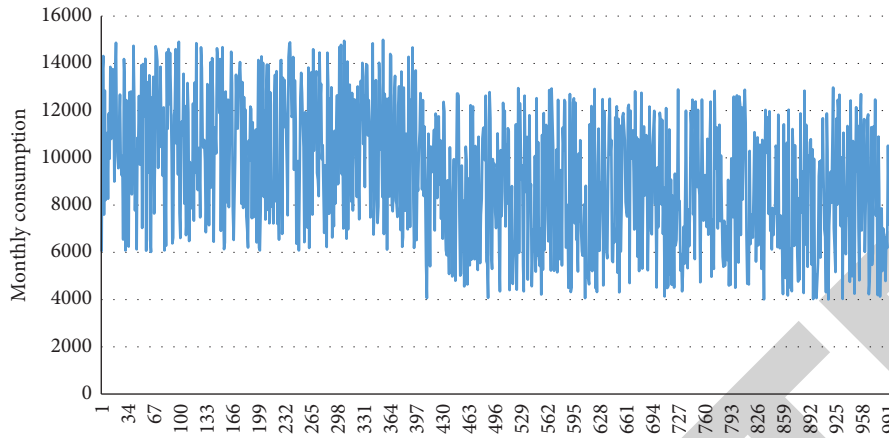


FIGURE 3: Monthly consumption distribution.

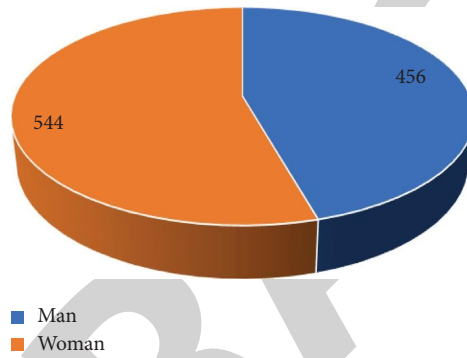


FIGURE 4: Proportion distribution of gender data.

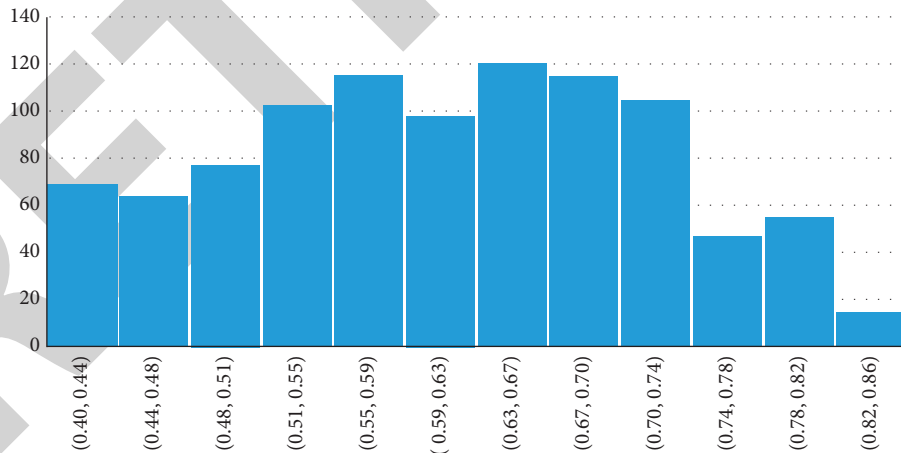


FIGURE 5: Proportion distribution of monthly consumption to monthly income.

undersampling. NearMiss uses some heuristic rules to select samples, which can be divided into three categories according to different rules:

NearMiss-1: We select the majority class sample with the closest average distance to the nearest  $K$  minority class samples

NearMiss-2: We select the  $k$  most distant minority class samples and the majority class samples with the nearest average distance

NearMiss-3: We select  $k$  nearest majority class samples for each minority class sample to ensure that each minority class sample is surrounded by the majority class sample

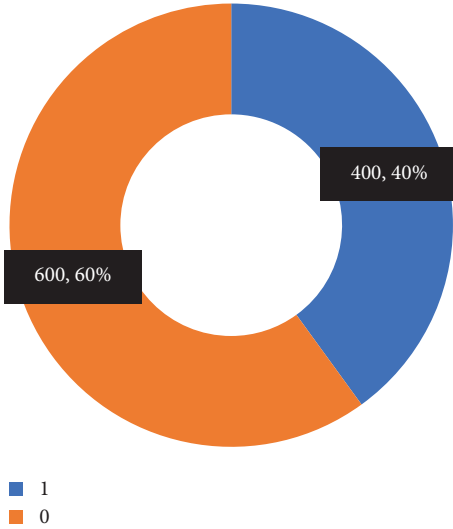


FIGURE 6: Proportion of customers' response after precise marketing.

Due to the large amount of calculation for NearMiss-1 and NearMiss-2, it is necessary to calculate the  $k$  nearest neighbors of each multicategory sample, so this paper adopts the NearMiss-3 sampling method.

**3.3. Framework Model of XGboost Algorithm.** XGboost outperforms GBDT, RGF, and other algorithms. It is an integrated learning algorithm based on boosting [24–26]. The base learner of XGboost can be a decision tree or a linear model. For datasets with a small number of missing values, the model has good fault tolerance and can automatically learn the splitting direction of the decision tree through a sparse sensing algorithm. In the process of decision tree splitting, the approximate algorithm of the greedy algorithm is used to find the most likely splitting point. As distributed computing is adopted, all datasets need to be traversed, so XGboost will consume more computer memory.

The advantages of XGboost algorithm are as follows: firstly, by performing second-order expansion on the loss function, it has higher accuracy than the GBDT algorithm, which only uses first-order expansion. Secondly, the model reduces the complexity of the model by introducing the regular term, which can effectively prevent the model from overfitting. Thirdly, since XGboost automatically learns the splitting direction of the tree through sparse sensing algorithm, in the preprocessing stage, a small number of missing values cannot be processed. However, the disadvantage of XGboost is that it needs to traverse the dataset during the node splitting process of the tree, which greatly consumes computer memory.

XGboost sorts the data according to the features, stores the sorted features in the block structure, and reduces the amount of calculation through the sparse matrix storage format. In the process of feature segmentation and sorting, the sorted feature values are accessed in order to facilitate the search for segmentation points. The model uses parallel processing for features, selects the feature with the largest

information gain as the splitting direction, and calculates the gain of multiple features at the same time. The specific principle of the XGboost algorithm is as follows.

### 3.3.1. Constructing Loss Function and Objective Function.

The deviation and variance of the model are reduced by constructing the loss function. To reduce the deviation, one must reduce the error between the model prediction result and the real value. Reducing variance can be seen as preventing overfitting, which is generally achieved by introducing regular terms into the model and reducing the complexity of the model. The objective function is composed of the loss function and the regular term, and the calculation formula is as follows:

$$\text{Obj} = \sum_{i=1}^n l(y_i, \hat{y}_i) + \sum_{i=1}^t \Omega(f_i). \quad (2)$$

Since boosting follows forward distribution addition, the predicted value of each step is determined by the predicted value of the previous step. Therefore, the final expression of the objective function can be obtained by superposition:

$$\text{Obj} = \sum_{i=1}^n l\left(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)\right) + \Omega(f_t) + \text{constant}. \quad (3)$$

### 3.3.2. Taylor Expansion Approximation of Objective Function.

Because the objective function of XGboost is too complex to be solved directly, Taylor polynomials are used here to approximate the objective function. Taylor's formula is a method of approximating a function  $f(x)$  with  $n$ -th derivative at  $x = x_0$  by  $n$ -th polynomial about  $(x - x_0)$ . According to Taylor's formula, the second-order expansion of Taylor's formula at point  $x$  can obtain the following equation:

$$f(x + \Delta x) \approx f(x) + f'(x)\Delta x + \frac{1}{2}f''(x)\Delta x^2. \quad (4)$$

Therefore, for the objective function of XGboost, the approximate function of the objective function can be obtained by taking  $f(x)$  of the previous formula as the loss function. Taking  $x$  as the predicted value of the first  $(t-1)$  tree and taking  $\Delta x$  as the  $t$  th tree under training. After removing the constant term that has no effect on function optimization, the approximate objective function can be written as

$$\text{Obj}^{(t)} \approx \sum_{i=1}^n \left( g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i) \right) + \Omega(f_t), \quad (5)$$

where  $\Omega(f_i)$  is the regular term.

**3.3.3. Generation of Decision Tree.** The complexity of the decision tree is determined by the number  $T$  of leaves and the weight  $W$  of the leaf nodes. By introducing the  $L_2$  normal form as the regular term, the calculation formula of complexity can be obtained by using the following formula:



$$\Omega(f_t) = \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T w_j^2. \quad (6)$$

Next, by scoring the structure of the tree, the greedy algorithm is used to find the splitting income of the tree, and the objective function after tree splitting is obtained as follows:

$$\text{Obj} = -\frac{1}{2} \left[ \frac{G_L^2}{H_L + \lambda} + \frac{G_R^2}{H_R + \lambda} \right] + 2\gamma. \quad (7)$$

### 3.4. Establishment and Solution of Credit Card Prediction Model

**3.4.1. Cross Validation.** In order to ensure the training effect and accuracy of the model, this paper uses the method of 50% cross-validation. Cross-validation is a method used to verify the performance of the model. The basic idea is to divide one part of the data into training sets and the other part of the data into validation sets. By dividing the data into five groups and making a validation set for each group of data, a total of five models will be obtained. The accuracy of these five models is averaged to obtain the evaluation results of the final model. The advantage of the five-fold cross-validation is that the division of the dataset is more detailed, and the error caused by the pure random division of samples is avoided [27].

**3.4.2. Analysis of Results.** Seventy percent of the samples were used for model training, and the rest were used for training. The model evaluation results shown in Table 3 are obtained.

The accuracy rate indicates the proportion of samples with correct predictions in the total sample. The greater the accuracy rate, the better. The recall rate refers to the proportion of individuals who are actually positive samples and are also positive samples after prediction. F1 is the harmonic average of the accuracy rate and the recall rate. Table 3 shows the evaluation results of the credit card precision marketing model established based on the XGboost algorithm. In the training set data, the accuracy rate reached 92.7%, and in the cross-validation set and the test set, the accuracy rate reached more than 80%, indicating that the prediction effect of the model established in this paper is good.

In order to carry out accurate marketing for customers, we can select the most important characteristic variables in accurate marketing by calculating the characteristic importance of each characteristic variable. Table 4 summarizes the names and importance of each feature.

It can be seen from Table 4 that among all characteristics, the highest importance of characteristics is “monthly consumption” and “age” and the lowest importance of characteristics is “gender.” If e-commerce enterprises want to achieve precision marketing, they need to define their basic market positioning. The findings demonstrate that the proposed model has the ability to perform scientific analyses on consumer data, thereby reducing marketing costs and

TABLE 3: Evaluation results of the model.

	Accuracy	Recall	F1
Training set	0.927	0.927	0.926
Cross-validation set	0.84	0.84	0.835
Test set	0.81	0.81	0.803

TABLE 4: Importance distribution of each indicator.

Features	Characteristic importance (%)
Age	20.90
Monthly income (yuan)	9.70
Monthly consumption (yuan)	26.80
Gender	2.70

facilitating the accumulation and analysis of sales data. Moreover, the model has the potential to enhance the analysis of customer behavior and data, leading to improved sales and profits for electricity enterprises.

## 4. Conclusion

With the arrival of the big data era, e-commerce network marketing has ushered in a new development platform, which is conducive to implementing the strategy of precision marketing, saving the expenditure of human costs, and conducting a comprehensive analysis of users’ purchasing habits, so as to carry out marketing activities and promote the sustainable development of e-commerce. The public dataset selected in this paper contains the basic characteristics of customers, such as age, gender, monthly income, monthly consumption, and whether the customer agrees to process the credit card after a personalized recommendation. After cleaning and preprocessing the data, the XGboost algorithm is used to predict whether the customer will process the credit card. In the training set data, the accuracy rate reached 92.7%, and in the cross-validation set and the test set, the accuracy rate reached more than 80%, indicating that the prediction effect of the model established in this paper is good. The most important characteristics are “monthly consumption/monthly income,” followed by “monthly consumption” and “age,” and the lowest importance of characteristics is “gender.” This study is helpful to better predict users’ consumption intentions, further explore potential customers, and improve the success rate of precision marketing.

### Data Availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

### Conflicts of Interest

The author declares that there are no conflicts of interest.

### References

- [1] M. Z. Iskandarbekovna, Y. Bolshakova, M. V. Akulich et al., “Internet marketing as an important factor in developing

- business,” *Academicia: An International Multidisciplinary Research Journal*, vol. 11, no. 12, pp. 60–68, 2021.
- [2] J. Ling, “Research on accurate marketing strategy of e-commerce enterprises based on big data,” *Industry and Technology Forum*, vol. 21, no. 13, pp. 231–232, 2022.
  - [3] S. Peng, “Analysis on the impact of e-commerce on traditional business model,” *Modern trade industry*, vol. 40, no. 13, pp. 37–38, 2019.
  - [4] S. Fu, “E-commerce precise marketing strategy under big data environment,” *National circulation economy*, vol. 24, no. 29, pp. 11–13, 2021.
  - [5] Lu. Wang, “E-commerce precise marketing strategy under big data environment,” *Cooperative economy and technology*, vol. 20, no. 03, pp. 82–83, 2019.
  - [6] Y. Zhao, Y. Wu, and L. Pang, “On marketing opportunities and Countermeasures under the background of big data,” *Time honored brand marketing*, vol. 33, no. 14, pp. 27–29, 2022.
  - [7] L. De Vries, S. Gensler, and P. S. H. Leeflang, “Effects of traditional advertising and social messages on brand-building metrics and customer acquisition,” *Journal of Marketing*, vol. 81, no. 5, pp. 1–15, 2017.
  - [8] J. M. Ponzoa and A. Erdmann, “E-commerce customer attraction: digital marketing techniques, evolution and dynamics across firms,” *Journal of Promotion Management*, vol. 27, pp. 1–19, 2021.
  - [9] P. Britt, “Marketing’s turning to AI for customer acquisition,” *Customer Relationship Management*, vol. 25, no. 3, pp. 3–5, 2021.
  - [10] O. Dogan, E. Ayçin, and Z. Bulut, “Customer segmentation by using RFM model and clustering methods: a case study in retail industry,” *International Journal of Contemporary Economics and Administrative Sciences*, vol. 8, no. 1, pp. 1–19, 2018.
  - [11] A. Johansson and J. Wikström, “Customer segmentation using machine learning,” *Journal of Applied Data Sciences*, vol. 2, no. 1, pp. 19–25, 2021.
  - [12] Y. Wang, “Precision marketing strategy based on the “user portrait” model,” *International Conference Education and Management*, vol. 2021, no. 5, pp. 92–96, 2021.
  - [13] L. Jing, “Network marketing strategy based on big data precision marketing,” *Business economics research*, vol. 15, no. 11, pp. 46–47, 2017.
  - [14] H. Zhao, Z. Li, F. Guo, and S. Mu, “Research on precision marketing based on K2 algorithm,” *Computer applications and software*, vol. 36, no. 07, pp. 43–49, 2019.
  - [15] F. Cao and H. Zhang, “Research on accurate marketing based on user portrait,” *Marketing*, vol. 16, no. 44, pp. 33–34, 2020.
  - [16] P. Kotler, G. Armstrong, and J. Saunders, *Principles of Marketing*, Economic Science Press, Beijing, China, 4th edition, 2005.
  - [17] P. Farris, “Marketing metrics: 50+ metrics every executive should master,” *International Commerce Review -- ECR Journal*, vol. 2, pp. 59–66, 2006.
  - [18] S. benaziz and J. Lehrer, *Smart Decision on Screen*, Zhejiang People’s publishing house, Hangzhou, China, 2017.
  - [19] H. Huang, “Analysis on the advertising marketing strategy of WeChat official account in the era of Micro Communication,” *Media*, vol. 18, no. 3, pp. 68–70, 2017.
  - [20] Y. Hu and Y. Lu, “Research on precision marketing strategy based on social media,” *Science and technology think tank*, vol. 9, no. 12, pp. 12–18, 2020.
  - [21] S. Zhang, “Transformation and practice of marketing mode of commercial banks in the era of big data,” *Marketing*, vol. 16, no. 10, pp. 50–52, 2022.
  - [22] X. Zhao, “Research on Application of wind power data governance system based on data standardization,” *Distributed energy*, vol. 7, no. 03, pp. 62–71, 2022.
  - [23] C. Jiang, *Research on Sampling Methods for Unbalanced Data*, Harbin University of technology, Harbin, China, 2022.
  - [24] X. Deng, A. Ye, J. Zhong et al., “Bagging–XGBoost algorithm based extreme weather identification and short-term load forecasting model,” *Energy Reports*, vol. 8, pp. 8661–8674, 2022.
  - [25] S. Song, Li. Rui, X. Wang, H. Gao, Y. Hong, and X. F. Qin, “Predicting the 7th day efficacy of acupoint application of Chinese herbs (Xiao Zhong Zhi Tong Tie) in patients with diarrhea – a machine-learning model based on XGBoost algorithm,” *World Journal of Traditional Chinese Medicine*, vol. 8, no. 2, 2022.
  - [26] H. Su, K. Qi, L. Fu, T. Guo, and L. Zhang, “Anomaly detection method of cloud platform timing monitoring data based on wavelet packet and xgboost,” *Digital technology and application*, vol. 40, no. 07, pp. 6–8, 2022.
  - [27] Z. Zhou, *Machine Learning*, Tsinghua University Press, Beijing, China, 2016.