International Journal for Housing Science and Its Applications
Publish August 10, 2025. Volume 46, Issue 4 Pages 639-647

https://doi.org/10.70517/ijhsa46457

Research on e-commerce user group segmentation and precision marketing model based on improved K-means algorithm

Ning Yu[1,*]
[1] Department of Economics and Trade, Maanshan Technical College, Maanshan, Anhui, 243031, China
Corresponding authors: (e-mail: ebwork@163.com).

Abstract In today's rapid development of digital economy, e-commerce platforms are facing the challenges of diversified user demands and fierce competition. Traditional marketing methods have problems such as high cost, poor effect, low conversion rate, and it is difficult to meet the demand for personalized services. For the problems of poor accuracy of user group division and poor marketing effect of e-commerce platform, this paper proposes an improved K-means clustering algorithm integrating genetic algorithm. The method optimizes cluster center selection by genetic algorithm, which solves the local optimal problem caused by the randomness of the initial point selection of traditional K-means algorithm. Based on 4000 user consumption data, the study constructed a user portrait model containing 8 dimensions, such as gender, age, average monthly total consumption, etc., and utilized the improved algorithm to divide the user groups. The experimental results show that the improved algorithm achieves an average accuracy of 90.16% on the Iris dataset and 80.94% on the Wine dataset, which is 13.41% and 14.37% higher than the traditional method, respectively. The user groups were successfully divided into four clusters, with the high-value user group spending an average of 3,500 yuan, accounting for 11.08% of the total. The study formulated differentiated marketing strategies for different user groups, providing an effective solution for e-commerce platforms to achieve precision marketing.

Index Terms K-means algorithm, genetic algorithm, user group division, precision marketing, clustering algorithm, e-commerce platform

I. Introduction

With the popularization of the Internet, e-commerce in the era of big data has been developed rapidly. In today's society, e-commerce is an important transaction method between enterprises and consumers [1]. Traditional marketing focuses on the innovation of goods, traditional marketing strategies and audience coverage, and its marketing concepts, methods and means can not match the rapid expansion of large-scale and rapid growth in volume [2]-[4]. Therefore, the marketing approach of e-commerce is gradually changing, evolving from the past rough and ready to precise and intensive, thus giving rise to precision marketing.

Precision marketing is a new marketing concept proposed by Kotler at the annual meeting of the United Nations in 2005 and has been widely used in various industries [5]. Different from the blind push in the past, precision marketing is customer-centered and pays more attention to the customer's experience and interactivity with the customer, and this method is based on a large amount of data, using data mining to find a specific audience group, and then using relevant technology to personalize the filtered audience group to push the promotional message [6]-[9]. E-commerce precision marketing has the characteristics of accuracy, efficiency, controllability, analyzability and measurability, which can help e-commerce enterprises to reduce marketing costs and improve efficiency, and it is of great significance to e-commerce enterprises [10], [11].

User group division as the basic step of precision marketing, the precision marketing method of e-commerce platform based on clustering analysis divides users into different groups or categories by calculating the distance between different data points, and then realizes personalized recommendation and coupon rewards [12], [13]. Among them, the K-Means algorithm is an important branch of data mining, as well as one of the most classical and widely used division-based clustering algorithms, which is used for user group division [14]. However, the basic K-Means algorithm is only applicable to the static capture of user features, with poor delineation performance for outliers or noisy data, and because it is based on Euclidean distance, the delineation results may not be reasonable when the complex structure or user scale features are large [15]-[17]. In order to guarantee the accurate division of user groups for precise marketing, the K-Means algorithm should be optimized.

This study proposes an improved K-means clustering method incorporating genetic algorithm to address the problems of randomness in the selection of initial clustering centers and easy to fall into local optimization in the division of e-commerce user groups by the traditional K-means algorithm. Firstly, the selection of initial clustering center is optimized by genetic algorithm to improve the stability of the algorithm and the quality of clustering; secondly, based on the consumer behavior data of e-commerce users, a multi-dimensional user portrait model is constructed, and the improved algorithm is used to classify the user groups; finally, for the different characteristics of different user groups, differentiated precision marketing strategies are designed, which provide e-commerce enterprises with scientific decision-making support and practical guidance.

II. Analytical basics and techniques

II. A.K-Means Algorithm

K-Means algorithm as one of the most widely used clustering algorithms. Due to its strong interpretability and simple algorithmic principle, the K-Means algorithm is also widely used in many practical application scenarios. The algorithmic principle of K-Means is to iteratively search for the class centroids of each class, so as to make the classes highly aggregated within the class and highly dispersed among the classes. The following section introduces the basic algorithmic principle of K-Means algorithm and the superiority and limitations of K-Means algorithm [18], [19].

II. A. 1) Theory of the K-Means algorithm

K-Means algorithm is an unsupervised clustering algorithm, which realizes the principle of simple and easy to understand, the clustering effect is good.

The core idea of the K-Means algorithm is: first of all, randomly select K samples from the dataset as the initial clustering of the clusters of the class center point C_i ($1 \le i \le k$, i for the integer), calculate the rest of the sample points and the clustering center of the Euclidean distance, will be labeled as a sample of the closest similarity distance of the sample center of the class to which the C_i belongs to. And the data objects are assigned to the clusters corresponding to the cluster center C_i. Then the average value of the sample points in each cluster is calculated as the new clustering center, and the next iteration is performed until the clustering center no longer changes or the maximum number of iterations is reached to abort. The Euclidean distance between the sample points in the space and the clustering center is calculated by the formula:

$$d(x, C_i) = \sqrt{\sum_{j=1}^{m} (x_j - C_{ij})^2} \tag{1}$$

where x is the sample points contained in the dataset. C_i is the class center of the i th cluster, and m is the dimension of the sample data. x_j, C_{ij} is the j th attribute value of the sample x and the centroid C_i. The error sum of squares SSE of the data set is calculated as:

$$SSE = \sum_{i=1}^{k} \sum_{x \in C_i} | d(x, C_i) |^2 \tag{2}$$

where the size of SSE indicates how good the clustering result is, and k is the number of clusters.

II. A. 2) K-Means algorithm flow

K-Means algorithm steps:

Step1: Randomly draw K samples from the data set, recorded as the initial class center points of the clusters.

Step2: Calculate the distance from each sample point to the center of the K clusters, and divide the sample points into the clusters belonging to the class center point with the smallest distance, and so on for a number of iterations.

Step3: During each iteration, update the center point (center of mass) of each cluster using a formula such as mean.

Step4: For the K clustering center, the cycle of 2,3 step iteration update, if the location point change is very small (can be developed to allow the offset threshold), it is considered to have reached a smooth state, the end of the iteration, for different clusters and cluster centers can be selected from different shapes of labeling.

II. A. 3) Algorithm advantages and disadvantages

K-Means algorithm advantages:

(1) The principle is simple, easy to process, fast convergence.

(2) Clustering effect is good.
(3) The interpretability of the algorithm is relatively strong.
(4) Less parameter settings, easy to adjust the parameters.
K-Means algorithm limitations:
(1) The selection of K value is not well grasped.
(2) It is difficult to converge to the optimum for non-convex data sets.
(3) For the class imbalance dataset, the clustering effect is not good.
(4) The final result is related to the selection of the initial class center, and it is difficult to jump out of the local optimum if the initial point is not reasonably selected.
(5) Noise points and outliers are easy to make the center of the class shift, which affects the clustering effect.

II. B.Precision marketing

II. B. 1) Market segmentation

Market segmentation refers to the user's consumer behavior, potential demand and other characteristics of the situation of categorization, a product group (market) into a number of sub-markets with strong similarity, so as to develop effective marketing strategies based on the characteristics of the sub-markets. In order to achieve greater marketing results. According to the size of the market, the degree of consumer aggregation, consumer purchasing power, consumer demand variability, consumer behavior and other factors, a product market (product market) is divided into a number of geographic areas or the process of consumer groups.

II. B. 2) Precision marketing methods

Precision marketing requires an in-depth understanding of the target audience and the characteristics of the product or service, as well as the selection of appropriate marketing channels and the development of a detailed marketing plan. At the same time, it is important to continuously optimize the marketing campaigns and measure their effectiveness to ensure that they are always effective. The following are several common precision marketing methods:

(1) Database-based marketing methods:

The establishment of a comprehensive consumer database is the cornerstone of the enterprise to achieve precision marketing, while the creation of a potential consumer database is a long and complex task that requires enterprises to continue to accumulate and improve.

Currently, based on the potential consumer database, there are mainly some precision marketing methods as follows: email marketing, social media marketing, SMS marketing, telemarketing, content marketing

(2) Internet-based marketing methods

Internet-based marketing methods refer to a marketing approach that utilizes Internet technologies and platforms to promote products or services, increase brand exposure, and achieve sales and business growth through online channels. Here are some common Internet-based marketing methods:

Currently, there are the following ways of Internet-based precision marketing methods:

Data-driven personalized marketing: through the collection and analysis of big data, to understand the user's interests, behavioral preferences and purchasing history and other information, based on which to tailor the push to the user personalized ads and marketing content, to improve the user's click-through rate and the conversion rate of purchase.

Social media marketing: operate fan and follower groups through social media platforms, set up precise target groups and place advertisements, utilize the communication characteristics of social media to rapidly expand the influence of the brand, and directly interact with users to increase user participation and brand loyalty.

Content marketing: By providing valuable and interesting content, such as articles, videos, pictures, etc., we attract users' interest in the brand or product and gradually lead them to make purchases or other desired behaviors. With the help of social media and search engines and other channels for wide dissemination, to increase brand awareness and user attention.

Mobile marketing: With the wide application of mobile Internet and smart devices, various means such as push notification, SMS marketing and APP promotion are adopted to realize precise marketing strategies for users. Targeted placement and personalized services are provided based on a variety of information such as the user's geographic environment, personal interests and consumption behavior.

III. Application of improved K-means algorithm in precision marketing

III. A. Improved K-means algorithm incorporating genetic algorithm

E-commerce in operation will have a large number of users to browse the system, so that there will be a large

amount of data storage, the use of traditional data analysis methods can no longer meet the e-commerce operation.

This paper proposes the use of data mining technology for e-commerce data analysis. At present, a variety of algorithms are used in e-commerce data mining, in which K-means algorithm is widely used in data mining because of its easy implementation. However, through a large number of practical applications, it is found that the algorithm has drawbacks in terms of the quality of data mining in the application, which affects the effect of data mining. In order to solve this problem, the traditional K-means algorithm is improved in the article, and the improvement is mainly realized based on genetic algorithm.

III. A. 1) Genetic algorithms

Genetic algorithm is an algorithm proposed based on the theory of biological evolution, which is currently used in many fields such as machine learning, adaptive control, etc., and occupies an important significance in the field of artificial intelligence. The classical genetic algorithm component consists of biological individual, biological population, biological genetic approach, biological fitness calculation, biological reproduction approach, biological selection method, biological hybridization approach, and biological mutation approach.

Biological individuals are the chromosomes in organisms using digital simulation. Chromosome encoding methods are binary and numerical encoding methods, due to the complexity of data in e-commerce, binary encoding method is used in this paper. Namely:

$$g(x_l^t,k)=u_k+\frac{u_k-v_k}{2l-1}\left(\sum_{j=1}^{t}x_l^{i(kl+j)}\times 2^{j-1}\right) \tag{3}$$

where u_k denotes the upper real k limit and v_k denotes the lower real k limit.

The floating point numbering is: suppose there are n data, with x_l^t denoting the l th generation of the t th chromosome, $t\in\{1,2,.....n\}$, and the number of loci per chromosome is denoted by $L=m$. For floating-point encoding this can be represented by $X_0=(x_0^1x_0^2\ldots x_0^n)^T$.

III. A. 2) K-means algorithm based on genetic algorithm

Due to the large amount of data in e-commerce directly using the traditional K-means algorithm leads to the problem of long running time and poor mining results. In order to solve this problem, it is proposed to combine genetic algorithm and K-means algorithm together to form K-means algorithm based on genetic algorithm.

Cluster class center selection in K-means algorithm is related to the effect of the algorithm, in order to solve this problem genetic algorithm is introduced in the cluster class center for center selection. The first step is to encode the center point, using floating point encoding method for encoding, and according to the genetic algorithm fitness function for optimal selection. This improves the algorithm stability.

In this paper, K-means algorithm is used for e-commerce data classification, where each group can be considered as a cluster class. Each group is clustering similar ones together.K-means algorithm is used to get cluster classes K from e-commerce database and iterative method is used until the error result reaches the satisfactory result. About K-means algorithm is applied in e-commerce system as:

Obtain dataset $D=\{d_j,|\ j=1,2,...n\}$ from the e-commerce system database, assuming that the number of centers of classification clusters in dataset D is $N_k=\{C_i\mid i=1,2,...,k\}$. The sum of standard deviations is used as a judgment method in the calculation process, and the judgment method is calculated as:

$$f(P_k)=\sum_{i=1}^{k}\sum_{P\subset C_i}|P-m_i|^2 \tag{4}$$

where the criterion function $f(P_k)$ denotes the sum of errors between clusters and c_i denotes the i th cluster. m_i denotes the mean of c_i clusters as $m_i=\frac{1}{t_i}\sum_{p\subset C_i}p$.

The similarity between the data was calculated using Euler's formula:

$$d(x_i,m_j)=\sqrt{\sum_{i=1}^{d}(x_{it}-m_{jt})^2} \tag{5}$$

Through the analysis it is not difficult to find that there are problems in the classical K-means algorithm are:

(1) Collecting data from e-commerce databases and selecting cluster class centers for the dataset, the common

selection method is the random selection method. Random selection of cluster class points causes a relatively large impact on the clustering results of the K-means algorithm.

(2) The amount of data in the business recommendation is relatively large, the use of the above formula as a criterion function is bound to affect the mining efficiency.

For the above problems, this paper proposes to use genetic algorithm for business recommendation K-means data mining. The traditional clustering algorithm K-means algorithm is analyzed and found to have slow running efficiency when dealing with large amounts of data. Genetic algorithm can effectively solve the problem when dealing with complex computing environment.

The classical K-means algorithm is improved because it is more affected by the selection of cluster class centroids. The key aspect of the improved K-means algorithm is that when the clustering process has a large distance calculation data can be introduced into the genetic algorithm variant operation for cluster class center point selection. The key technique for improving K-means algorithm is:

(a) Chromosome and cluster components. In the e-commerce management system using floating-point coding, and K-means algorithm is the core idea of the cluster class center selection, this paper constructs the chromosome definition method expressed in $Si(j,k)$.

(b) Algorithm adaptation calculation. Namely:

$$f(P_k)=\sum_{i=1}^{k}\sum_{P\subset C_i}|P-m_i|^2 \tag{6}$$

According to the above equation it can be obtained that the clustering effect is very obvious when the function is obtained when the function is small. In this way to get the fitness calculation method is:

$$f(R_i)=E_{\max}-E(R_i) \tag{7}$$

where $E_{\max}$ denotes the maximum chromosomal variance of population R.

(c) Data selection. Selection can be done according to the following formula. The selection probability is obtained using a roulette wheel algorithm in the calculation process. The calculation method is:

$$P_i=\frac{f(R_i)}{\sum_{i=1}^{n}f(R_i)}(i=1,2,..n) \tag{8}$$

(d) Mutation operation. Set the mutation rate to $P=0.05$.

III. A. 3) Algorithm time complexity analysis

Algorithmic time complexity analysis of the improved K-means algorithm yields a time complexity of $O(t)$ in step 2, a K-means clustering time complexity of $O(n)$ in step 1, and a time complexity of $O(t^2)$ for the problem of sorting the data involved, and since the genetic algorithm crossover and mutation are done in a shorter period of time, the program running time and clustering time complexity is consistent with $O(n)$. Therefore the total time complexity of the improved algorithm is $O(T\times(t^2+t+2n))$.

III. A. 4) Improved K-means algorithm analysis

The software operating system of the experimental platform is Window10, and the compilation software is MATLAB software.

The criteria for evaluating the results of the clustering algorithm uses the correctness criterion, which specifies that the class label of the i th cluster is determined by the class label of the most objects in the cluster, and counts the number of these objects as n, with a total of K clusters. The evaluation criterion used in the experiments of the algorithm in this paper is accuracy, which is able to analyze the degree of accurate classification of virus classes, and the calculation process is shown in the following equation:

$$P(T)=\frac{\sum_c A_1(c,T)}{\sum_c A_1(c,T)+A_2(c,T)} \tag{9}$$

where T in Eq. can describe the delineated data clusters. C denotes the real class labeling in the data set. $A_1(c,T)$ denotes the number of data belonging to the real class C accurately classified into the cluster T. $A_2(c,T)$ indicates the number of data objects belonging to the true category C but incorrectly classified into

cluster T. The higher the correct rate of the clustering result, the better the quality of the clustering is represented.

In order to be able to better verify the clustering effect of the algorithm proposed in this paper, 2 standard datasets from the UCI database are selected for simulation experiments. The datasets are Iris and Wine.

During the experiments, the results are compared using traditional K-means algorithm, density K-means algorithm, K-means clustering method with variable grid optimization, rough K-means clustering algorithm with local density adaptive metrics and the algorithm of this paper. Where the average accuracy is the sum of the accuracy after the algorithm has run 8 experiments divided by 8. Density K-means algorithm is a spatial density K-means algorithm, this algorithm is proposed on the basis of classical K-means algorithm, which is representative and classical in nature. The K-means clustering method for variable grid optimization is a novel K-means algorithm constructed according to the grid type. The rough K-means algorithm with local density adaptive metric is a further improved K-means algorithm based on the density algorithm.

Firstly, five algorithms are experimented in the dataset Iris, and the algorithms run the results according to the initialization center. Because the traditional K-means algorithm is manually specified initial clustering center, so you can choose different initialization center point according to their experience, to ensure that the algorithm runs 8 times, so that you can ensure the fairness and impartiality of the algorithm running results as much as possible.

The experimental results of the Iris experimental dataset are shown in Table 1. In this paper, the genetic algorithm is used to improve the K-means algorithm, which has achieved an average accuracy of 90.16% in the Iris experimental dataset, which is higher than other algorithms.

Table 1: Iris experimental data set experimental results

Algorithm name	Initial center	Accuracy	Mean accuracy
Traditional k-means algorithm	40, 106, 115	86.97%	76.75%
	8, 65, 3	78.54%	
	127, 49, 94	60.42%	
	97, 81, 66	59.56%	
	41, 108, 119	86.72%	
	74, 79, 125	80.40%	
	15, 43, 121	77.97%	
	25, 26, 94	83.45%	
Density k-means algorithm	25, 61, 124	83.15%	83.15%
Variable grid optimization of the k-means clustering method	46, 67, 28	86.07%	86.07%
A rough k-means clustering algorithm for local density adaptive fitness	26, 75, 48	86.92%	86.92%
This algorithm	82, 103, 142	90.16%	90.16%

The experimental results of the Wine experimental dataset are shown in Table 2. In the Wine experimental dataset, the average accuracy of this paper's algorithm is 80.94%, which is 5.72% higher than the rough K-means algorithm with local density adaptive metrics, and 5.27% higher than the K-means clustering method with variable grid optimization. The algorithm in this paper has more performance advantages.

Table 2: Wine experiment data set experimental results

Algorithm name	Initial center	Accuracy	Mean accuracy
Traditional k-means algorithm	25, 105, 48	65.80%	66.57%
	17, 121, 67	72.57%	
	65, 174, 43	89.61%	
	68, 138, 45	60.33%	
	107, 33, 91	70.37%	
	25, 104, 147	59.68%	
	141, 56, 38	57.42%	
	97, 41, 62	52.69%	
Density k-means algorithm	62, 24, 38	70.42%	70.42%
Variable grid optimization of the k-means clustering method	105, 31, 45	75.67%	75.67%
A rough k-means clustering algorithm for local density adaptive fitness	105, 97, 42	75.22%	75.22%
This algorithm	82, 137, 25	80.94%	80.94%

III. B. User group segmentation based on e-commerce users' consumption behavior

The data for the study comes from the user consumption data provided by a shopping APP in 2024.The dataset contains a total of 4,000 user data, including 20 fields such as user attributes, consumption habits and preferences.

The user data in terms of gender, the ratio of male and female users is about 3:7. The age distribution of the sample is realistic, concentrating on 18-25 years old, followed by 25-45 years old, the older the age the lower the proportion of the sample size.

III. B. 1) Selection of clustering indicators

The study of consumer behavior refers to the study of how individuals, groups, and organizations actually select, purchase, use, and dispose of goods, services, ideas, or experiences to satisfy their needs and desires.

The study of consumer behavior by companies should not only focus on the immediate sale of products and services, but also on the long-term exchange relationship. To this end, it is necessary to understand not only how consumers select their preferred products and services, but also to study the consumer's user experience during the consumption process, which will affect the consumer's future purchasing behavior and will have a significant impact on the company's ability to build a long-lasting relationship.

In this section, we will further segment users with their live streaming consumption behavior data, using demographic characteristics as well as certain user consumption behavior characteristics as user segmentation dimensions.

The user segmentation based on users' live streaming consumption behavior in this section will enable enterprises to have a clearer understanding of users' preferences and demands, and formulate targeted marketing strategies for each user group based on users' consumption preferences, in order to achieve efficient marketing and effective retention.

Based on the original data, this paper selects eight indicators representing users' consumption behavior preferences as shown in Table 3 to construct a segmentation model and carry out empirical research.

Consumption habits include age, gender, average monthly total consumption, number of purchases, average interval, number of purchases, user level, and hesitation time.

Table 3: Clustering index

Cluster dimension	Data index
User basic attribute	Gender
	Age
Consumption habits	Total monthly consumption
	Purchase frequency
	Mean interval
	Purchase number
	User level
	Hesitation

III. B. 2) Cluster analysis process

(1) Implementation of K-means algorithm based on genetic algorithm

First, the K value is selected according to the genetic algorithm. The maximum number of iterations of K-means is set to 200 times, and the variation value of genetic algorithm is calculated when K value is taken from 4 to 15 respectively, and the number of clusters when the index is maximum is the final K value.

The results of K value selection are shown in Table 4, from which it can be seen that when K=8, the GA value is the largest, so K takes 8 is the most appropriate.

Table 4: K value selection

K	4	5	6	7	8	9
GA	7.51	6.12	8.45	8.67	9.52	8.24
K	10	11	12	13	14	15
GA	6.28	7.85	9.04	9.15	8.94	8.79

The maximum number of iterations of the genetic algorithm is taken to be 200, and the velocity weight ω is selected to be 0.6 by cross-validation, the acceleration constant factor is set to be $c_1 = c_2 = 2$, and the maximum value of the velocity is the same as that of the taking of the V_{max} to be 0.3. The weight in the fitness function is

selected by cross validation to be $\omega_1 = \omega_2 = 0.9$. Each iteration corresponds to a set of center solutions of K-means, and the optimal iteration obtained is the optimal set of centroids, and then K-means clustering is performed on the dataset, and finally K datasets are obtained, i.e., the final user group.

(2) Clustering center

The data is pre-processed, while the processed data is later clustered by the improved K-means algorithm to obtain four classes. The data of these four classes are processed and the average value is taken for processing. The average value of the data in each class is shown in Table 5.

The four classes of consumers into which the clustering is divided are characterized significantly, in which the consumers in Cluster 1 do not make many purchases. Consumers in Cluster 2 purchased items at medium average prices but most frequently. Consumers in Cluster 1 and Cluster 2 have the strongest purchasing power and are loyal consumers of the e-commerce platform.

Table 5: The average of all kinds of data

	Cluster 1	Cluster 2	Cluster 3	Cluster 4
The number of people in the cluster	443	979	1024	1554
Average purchase interval	4.8	3.1	8.7	11.4
Average purchase times	10.2	25.6	13.5	7.3
Average purchase amount	3500.0	1985.2	560.87	200.5
The average number of purchases	13.4	40.2	16.1	7.2
Average user level	3	2	1	0
Mean hesitation	4.6	6.8	12.3	13.9

III. C. Application Analysis of Precision Marketing Strategy

After analyzing the data, different clusters were obtained. In order to get a better characterization of the users in each class, five user samples were randomly selected from each class to be analyzed and the dimensions of the selected samples were extended for precise analysis.

Cluster 1 user sampling data is shown in Table 6. Therefore, for high-end luxury consumers, the behavior of stimulating them to make purchases with gifts and coupons does not achieve great results; such users do not care about the amount spent on the membership level, so there is no need to offer them discounts on the price of membership. Their shopping interval is about 7 days, and they are likely to buy new luxury goods, so we can push new luxury goods priced at about 5000 RMB to them when the time interval is 7 days.

Table 6: Cluster 1 user sampling data

	Consumer 1	Consumer 2	Consumer 3	Consumer 4
Mean interval	5days	7days	12days	15days
Purchase frequency	9	12	14	16
Total purchase	2450	3800	4700	7000
Average cost per cost	253	412	286	475
Purchase number	10	13	16	12
User level	3	3	2	1
Mean hesitation	175s	89s	200s	120s
Gift number	5	4	3	1

The analysis results of the user consumption behavior of cluster 1, cluster 2, cluster 3, and cluster 4 are synthesized to get the marketing strategies of each cluster. Detailed statistics of marketing strategies for different clusters are shown in Table 7. For the users of Cluster 1, the marketing method of no push, no coupon, no inducement to order, and no free gift is adopted, and the marketing time is suggested to be 9 days to sell branded products to the users of Cluster 1.

Table 7: Detailed statistics on different clustering marketing strategies

	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Interval time	9days	4days	6days	7days
Brand price	5000	2000	500	300
Member	Non-push	Non-push	Appropriate push	Push

Collateral gift	No	Yes	Yes	No
Coupons	No	No	Push	Push
Excellent order	No	No	Yes	Yes

IV. Conclusion

In this study, the traditional K-means clustering algorithm is improved by incorporating genetic algorithm, which successfully solves the problem of randomness in the selection of initial clustering centers in the division of e-commerce user groups. The algorithm validation results show that the improved method performs well on standard datasets, with an average accuracy of 90.16% on the Iris dataset and 80.94% on the Wine dataset, which is significantly better than the traditional clustering method. In practical application, the 8-dimensional user portrait model constructed based on 4000 e-commerce user data can effectively distinguish different types of consumer groups and successfully classify users into 4 clusters with distinctive characteristics. Among them, cluster 1 represents high-end consumers, with an average consumption amount of 3,500 yuan, lower purchase frequency but higher unit price; cluster 2 is active consumers, with a high number of purchases of 25.6 times, which is the core user group of the platform. Through the in-depth analysis of the behavioral characteristics of different clusters of users, targeted marketing strategies are formulated, including personalized recommendation timing, coupon placement strategy, member service differentiation and other aspects. The study provides a scientific theoretical basis and practical technical solutions for e-commerce platforms to realize precision marketing, which helps to improve user satisfaction and enterprise operational efficiency.

References

[1] Jain, V., Malviya, B. I. N. D. O. O., & Arya, S. A. T. Y. E. N. D. R. A. (2021). An overview of electronic commerce (e-Commerce). Journal of Contemporary Issues in Business and Government, 27(3), 666.

[2] Devita, M., Nawawi, Z. M., & Aslami, N. (2021). Shopee's e-commerce marketing strategy in international business. Journal of Social Research, 1(1), 27-31.

[3] Al-Husban, M. (2017). E-commerce adoption model for traditional retailers in developing countries. International Journal of Infonomics, 10(2).

[4] Johansson, T., & Kask, J. (2017). Configurations of business strategy and marketing channels for e-commerce and traditional retail formats: A Qualitative Comparison Analysis (QCA) in sporting goods retailing. Journal of Retailing and Consumer Services, 34, 326-333.

[5] Qingyou, Y., Guangyu, Q., Qifeng, W., Yang, Y., & Yunfeng, Z. (2021, August). Research on the construction and application strategy of precision marketing model for industrial and commercial customers. In IOP conference series: earth and environmental science (Vol. 831, No. 1, p. 012037). IOP Publishing.

[6] Jiangting, T. (2022). Research on the Precision Marketing and Optimization of Pinduoduo in the Context of Big Data. Pacific International Journal, 5(4), 51-55.

[7] Yu, C., Zhang, Z., Lin, C., & Wu, Y. J. (2020). Can data-driven precision marketing promote user AD clicks? Evidence from advertising in WeChat moments. Industrial Marketing Management, 90, 481-492.

[8] Xu, Z., Nie, H., & Chen, W. (2025). A precision marketing method for e-commerce considering the hidden behavioural characteristics of user online shopping. International Journal of Business Intelligence and Data Mining, 26(3-4), 346-361.

[9] Zhou, Z., & Li, B. (2025). An accurate and rapid pushing of marketing information based on multidimensional data mining. International Journal of Business Intelligence and Data Mining, 26(3-4), 396-412.

[10] Jun, W., Li, S., Yanzhou, Y., Gonzalezc, E. D. S., Weiyi, H., Litao, S., & Zhang, Y. (2021). Evaluation of precision marketing effectiveness of community e-commerce–An AISAS based model. Sustainable Operations and Computers, 2, 200-205.

[11] Zhu, G., & Gao, X. (2019). Precision retail marketing strategy based on digital marketing model. Science Journal of Business and Management, 7(1), 33-37.

[12] Wang, H., Wang, J., & Zhong, Z. (2020, April). Research on precision marketing strategy based on cluster analysis algorithm. In 2020 International Conference on E-Commerce and Internet Technology (ECIT) (pp. 208-211). IEEE.

[13] Wang, K., Zhang, T., Xue, T., Lu, Y., & Na, S. G. (2020). E-commerce personalized recommendation analysis by deeply-learned clustering. Journal of Visual Communication and Image Representation, 71, 102735.

[14] Tabianan, K., Velu, S., & Ravi, V. (2022). K-means clustering approach for intelligent customer segmentation using customer purchase behavior data. Sustainability, 14(12), 7243.

[15] Laccetti, G., Lapegna, M., Mele, V., Romano, D., & Szustak, L. (2020). Performance enhancement of a dynamic K-means algorithm through a parallel adaptive strategy on multicore CPUs. Journal of Parallel and Distributed Computing, 145, 34-41.

[16] Grunau, C., & Rozhoň, V. (2022, June). Adapting k-means algorithms for outliers. In International Conference on Machine Learning (pp. 7845-7886). PMLR.

[17] Ghazal, T. M. (2021). Performances of k-means clustering algorithm with different distance metrics. Intelligent Automation & Soft Computing, 30(2), 735-742.

[18] Wukui Sun,Lulu Jia & Yingyu Shao. (2025). Short-circuit Current Protection Setting Algorithm of Doubly-fed Power Supply Based on K-means Clustering. Smart Grids and Sustainable Energy,10(2),35-35.

[19] Wenqi Yang,Chang Tang,Xinwang Liu,Guanghui Yue,Yuanyuan Liu,Changqing Zhang & En Zhu. (2025). Smooth Multiple Kernel k-Means via Underlying Graph Filtering.. IEEE transactions on neural networks and learning systems,PP,