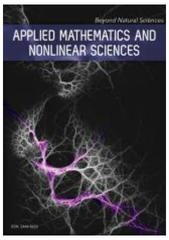
sciendo Applied Mathematics and Nonlinear Sciences, 9(1) (2024) 1-15



Applied Mathematics and Nonlinear Sciences

https://www.sciendo.com

Research on e-commerce user segmentation and customized marketing strategy based on cluster analysis

Yue Zhao[1], Xueyan Niu[1], Shuning Lin[1], Fang Su[1,†]

1. Department of Business Administration, Shandong Labor Vocational and Technical College, Jinan, Shandong, 250300, China.

Submission Info

Communicated by Z. Sabir
Received May 23, 2024
Accepted September 5, 2024
Available online October 4, 2024

Abstract

E-commerce user segmentation is the basis for enterprises to accurately formulate marketing strategies and successfully manage their customer base. With the rapid development of e-commerce, this paper improves the traditional K-means clustering algorithm and proposes a SAPK-means algorithm, which effectively excludes the noisy data and isolated points in the dataset and obtains the high-quality initial clustering center. Company A’s e-commerce platform is used to apply the SAPK-means algorithm for customer segmentation, and the results are analyzed in detail before proposing targeted marketing strategies. The customized marketing strategy’s effect is evaluated through sales and platform user satisfaction. Through experimental testing, this paper concludes that the five types of segmented customer groups account for 9.46%, 18.43%, 36.95%, 21.91%, and 13.24% of the total number of samples, respectively, known as the “platinum customer group”, “gold customer group “Platinum customer group”, “Gold customer group”, “Silver customer group”, “Copper customer group” and “Iron customer group”, respectively.

Keywords: Cluster analysis; E-commerce user segmentation; SAPK-means algorithm; Marketing strategy.
AMS 2010 codes: 97P20

†Corresponding author.
Email address: sufangvv@163.com

ISSN 2444-8656
https://doi.org/10.2478/amns-2024-2668

1 Introduction

Along with the rapid development of economic globalization and the rapid iteration of information technology, not only has the construction of the Internet become more and more perfect, but the Internet of Things has also begun to be established all over the world. While the mobile Internet provides great convenience for people's lives, all kinds of information about users are also emerging. How to mine the value behind the data and realize customized marketing in the context of big data has become the focus of competition in various industries [1-3]. The combination of e-commerce user segmentation and the Internet precision marketing model is sweeping traditional enterprises at an unprecedented speed. Secondly, with the development of e-commerce, online consumers' personal information, consumer behavior, consumption habits, and preferences are also changing. Consumer behavior has also brought a series of changes and reshaping to the company's products and services. The most significant change is that all the user's behavior can be tracked, can be analyzed [4-6]. Therefore, how to efficiently utilize these data, clean, analyze, and evaluate them, discover useful information from a large amount of data, and contribute to the current business development has become a difficult problem that enterprises need to face in the context of digitization [7]. At the same time, how to use big data to realize refined operations and customized marketing is also the focus of enterprises. Taking user data as the basis, mining out the characteristics of users, analyzing and counting these characteristics to find out the potential value information, and abstracting the user portrait is the foundation of enterprises applying big data and also the prerequisite for customized marketing [8-10].

User clustering is a secondary aggregation concept derived from the concept of user profiling and extended to consumer activities. Clustering is a way for organizations to reduce costs and increase the efficiency of user resolution. Literature [11] utilizes methods such as k-mode clustering algorithms and geographic distribution of users, which can categorize consumers according to their purchasing preferences, interests, age groups, and other factors and provide further insights into what potential market opportunities exist in their regions or geographic locations. Literature [12] proved that the K-Prototype clustering algorithm in the field of e-commerce for the effective realization of the task of customer segmentation has a significant advantage. Through the different attributes into account and combined with the appropriate distance measure, the algorithm can more accurately capture the existence of similarity or difference between the internal and external sample points of the clusters, which in turn provides a more accurate, effective, and targeted reference basis for enterprise decision-making. The algorithm can more accurately capture the similarity or difference between the internal and external sample points of each cluster, thus providing a more accurate, effective, and targeted reference for enterprise decision-making. Literature [13] emphasizes that in today's rapidly developing and competitive digital market, customer segmentation using clustering algorithms is a very critical and necessary task, which can help enterprises to deeply understand their target consumer groups and accurately locate the market position, as well as filter out ineffective information and obtain valuable insights from large amounts of data, and finally, to implement targeted marketing strategies to increase user satisfaction and loyalty.

Customized marketing views consumers as an important resource for a company, providing them with a personalized experience by understanding their preferences, needs, behaviors, and attitudes in order to build a closer and longer-lasting relationship. Literature [14] systematically regressed the marketing strategies of e-commerce consumers in the past decade and found that in-depth exploration and understanding of the interplay between consumer behavior, market demand, and technological innovation is crucial and that only by continuously tracking the latest scientific advances and strengthening effective communication and cooperation between academia and the practice community can the industry be pushed towards a more sustainable, intelligent direction and achieve greater success. The Literature [15] aims to explore the impact of changes in information dissemination on the strategic direction and opportunities for enterprises, providing a new way of thinking for enterprises to develop differentiated marketing strategies according to the characteristics of different clusters and hoping to help small and medium-sized enterprises better adapt to today's highly competitive and ever-changing market environment. Literature [16] found that through the use of social media platforms and personalized advertising tools, e-commerce firms are better able to interact with consumers and provide customized and precise services, and this resilience allows firms to adapt to the changing and competitive market environment and achieve sustained growth.

The article first analyzes the basic theory of K-means, further improves the K-means algorithm, puts forward the SAPK-means algorithm, and elaborates in detail on the specific process of the algorithm and how to segment the customers of e-commerce websites. Company A e-commerce platform as an empirical research object, through the use of the SAPK-means algorithm proposed in this paper to segment and analyze the platform's customers, put forward the marketing strategy for the analysis results, and finally through the changes in sales and user satisfaction of Company A's e-commerce platform to test the feasibility and effectiveness of the algorithm proposed in this paper.

2 Method

2.1 K-means clustering algorithm

2.1.1 K-means algorithm

The core idea of the K-means algorithm is to classify the data from the set of K cluster classes and classify the training data according to a certain error method. The steps of the algorithm are: (1) Select K from the dataset as the center point of the cluster class. (2) Calculate the distance from the dataset to the center point of each cluster class using an iterative method, construct the objective function to determine the center point of the cluster class using the sum of squares of the errors, and redefine the prime points. (3) Repeat the iteration until the error in constructing the objective function is minimized [17]. The flow of the K-Means algorithm is shown in Fig. 1:

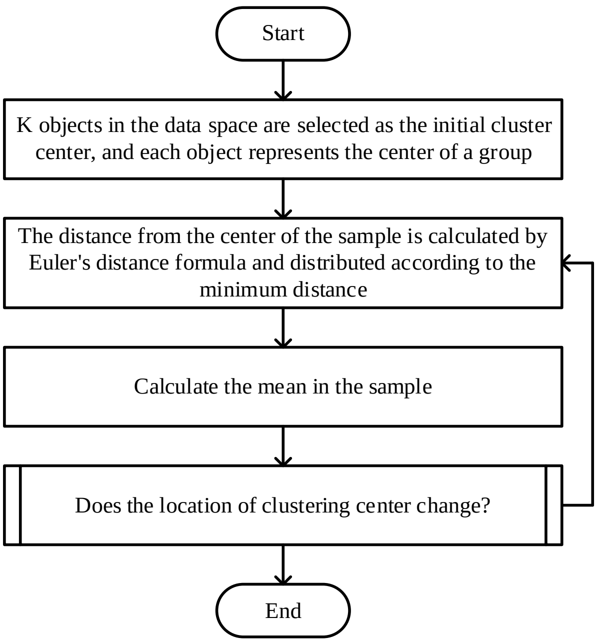


Figure 1. K-means algorithm process

The K-means algorithm is to group similar data. The algorithm needs to define the number of centroids in its operation, followed by calculating the distance from the centroids to each data and reclassifying the centroids. The error function constructed is calculated as follows:

$$P=\sum_{i=1}^{n}\sum_{j=1}^{k}\| x_i-c_j\|^2 \tag{1}$$

Where: x_i denotes the ith data in the jth cluster class: c_j denotes the centroid of the jth cluster class.

2.1.2 K-means cluster analysis method

Clustering is a widely used "unsupervised learning" learning task, mostly used in research and analysis, by learning unknown information markers in the samples, which can be used to explore the intrinsic patterns and properties of the data and thus provide the basis for further analysis of the data. Clustering divides the samples in a dataset into a number of non-intersecting subsets, which is called a "cluster". Each cluster in such a division corresponds to a potentially unknown concept or category, which is automatically formed during the clustering algorithm and whose semantics need to be determined and named by the user. Clustering is a separate process for finding the internal distribution structure of data and a preliminary step for other learning tasks such as classification [18].

Formally, suppose that the sample set $D=\{x_1,x_2,\dots,x_m\}$, which contains m unlabeled samples, and each sample $x_i=(x_{i1},x_{i2},...,x_{im})$ is an n-dimensional feature vector, then the clustering algorithm divides the sample set D into k disjoint clusters $\{C_l|1=1,2,...,k\}$, where $C_l,\cap_{l'\neq l}C_l=\varnothing$ and $D=U_{l=1}^{k}C_l$ correspondingly, denote the "cluster labeling" of a sample x by $\lambda_j\in\{1,2,\dots,k\}$, i.e., $x_j\in C_{\lambda_j}$ so that the clustering can be represented by the cluster labeling vector $\lambda=(\lambda_1;\lambda_2;...;\lambda_m)$, which contains m elements. The result of clustering can be represented by a cluster labeling vector 8 containing m elements 1491.

For a given set of samples $D=\{x_1,x_2,\dots,x_m\}$, the clusters resulting from clustering are divided $C=\{C_1,C_2,......,C_k\}$ into clusters whose minimized squared errors are given by:

$$E=\sum_{i=1}^{k}\sum_{x\in\square_i}\| x-u_i\|_2^2 \tag{2}$$

Where μ_i is the mean vector of cluster C_i, sometimes called the center of mass, Eq:

$$\mu_i=\frac{1}{|C_i|}\sum_{x\in C_i}x \tag{3}$$

Equation (3) describes how closely the samples in the cluster are organized around the cluster mean vector. The smaller the value of E, the more similar the samples in the cluster are. It isn't easy to get the depreciation of E through Eq. Finding its optimal solution requires examining all possible cluster divisions of the sample set D. It is an NP problem, and the only way to optimize to get the approximate solution formula is through iteration [19].

2.2 Flow of SPAK-means algorithm

The SAPK-means algorithm consists of the following operational procedures:

1) Input the set of sample elements and initialize the parameters.

   Given a set with N sample elements, bias parameter $p = p_m$,descent step $step = p_m / 10$, the technical mark H and supervisory mark H are both zero, the convergence condition is no change in the cluster center cycle 30 times, and the termination condition is no change in the cluster center cycle 300 times.

2) Perform iterative operations.

   (1) Carry out iterative operations to obtain the number of K clusters. If $H = 1$, proceed to (3), otherwise continue to the next execution.

   (2) Observe the convergence of the clustering effect, if convergence, analyze and generate $Sil(K)$ and record $H = 1$, then go to (4); otherwise, go back to the previous step.

   (3) Verify whether the center of the cluster meets the convergence requirement, if so, obtain K clusters and analyze the generation of $Sil_{\max}$, if $Sil(K) < Sil(K-1)$, then $H = H + 1$. When $Sil(K) > Sil_{\max}, H = 0$.

   (4) verifies whether it is $H > K / 2$. whether the value of K is 2. and verifies whether the number of loops meets the end requirement, if so, go to (5), otherwise go to (1).

   (5) Verify the number of optimal clusters specified in $Sil_{\max}$. If it is 2, analyze and generate Hartigan metrics for comparison.

   (6) Output the number of clusters and the centroids of the clusters.

3) Initialize the K-means algorithm with the value K of the output clusters and the centroids of the clusters.

4) Execute the K-means algorithm to obtain the final clustering effect. The flow of the SAPK-means algorithm is shown in Fig.2:

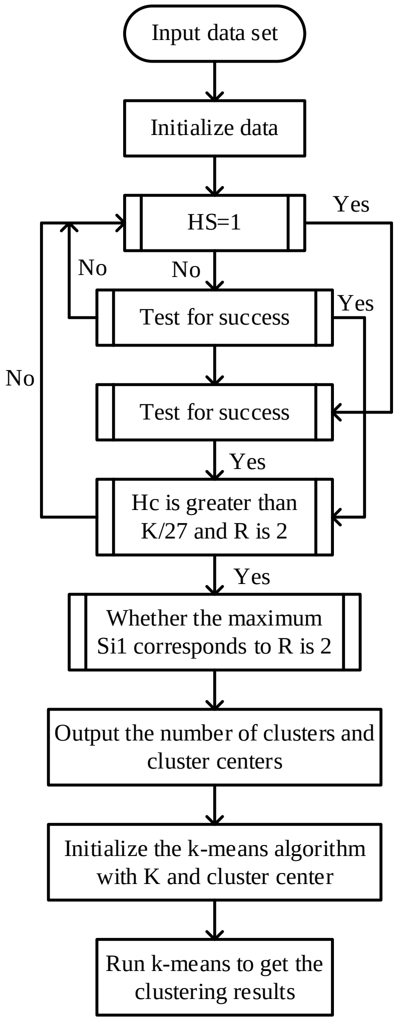


Figure 2. The process of the SPAK-means algorithm

2.3 SAPK-means algorithm applied to e-commerce customer segmentation

2.3.1 Steps in Customer Segmentation

With the increasing development of e-commerce websites and data mining techniques, it is becoming more and more a trend to accomplish customer segmentation using data mining techniques. E-commerce sites do not know the needs of customers and provide the same services to customers. This is obviously not the optimal strategy from an economic and marketing point of view. Suppose we can distinguish the different preferences of different customers by using data mining technology, so that customers with the same preferences form a customer group, and provide different services to different customer groups. In that case, it will largely improve customer satisfaction and loyalty. A detailed explanation of customer segmentation is shown in Figure 3. Different five-point stars represent customers with different needs, and in the absence of customer segmentation, all customers are treated in the same way. After customer segmentation, the organization offers various services to different customer groups [20].

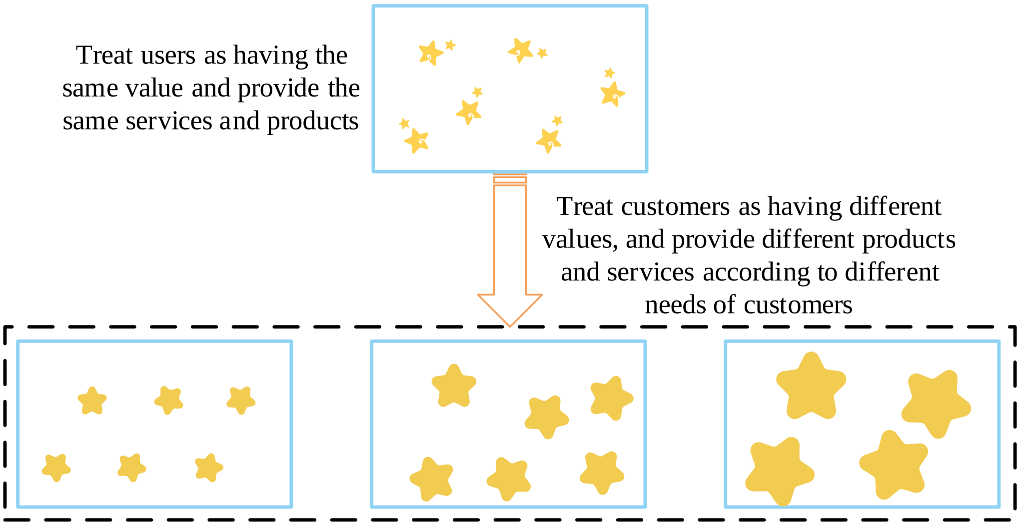


Figure 3. Detail of customer segmentation

Usually, we base the visitor logs in the e-commerce site or CRM on the relevant information, the first data preprocessing, and then build the relevant model, the use of clustering methods to segment the customer, and provide a basis for the enterprise to make decisions.

SAPK-means algorithm applied to customer segmentation of the specific steps are as follows:

1) Obtain the required data table from the relevant data table of the e-commerce website.

2) Judge whether there is a clustering trend in the acquired data table. If there is a clustering trend, then carry out clustering. Otherwise, cancel the following clustering steps.

3) The improved k-means algorithm is applied to the acquired customer dataset D. D is categorized into $c1, c2, c3......$ equal classes by the clustering algorithm.

4) For each class, one or several rules corresponding to the characteristics of each class are generalized according to the characteristics of the data objects in the class.

5) Evaluate the clustering results. If the clustering results are credible, it is confirmed that they will be used in practical applications. Otherwise, they will be re-clustered and analyzed with other clustering algorithms. The specific steps of customer segmentation are shown in Figure 4:

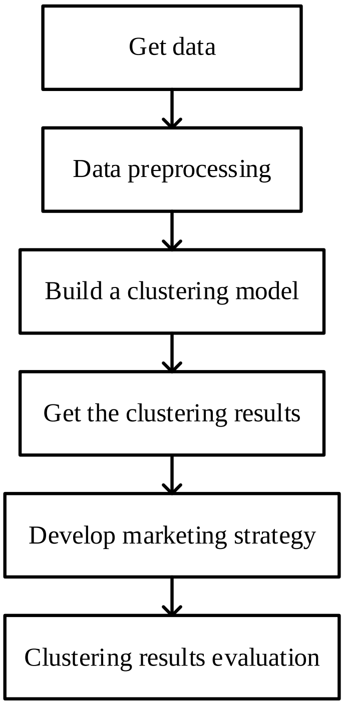


Figure 4. Specific steps for customer segmentation

2.3.2 Data acquisition

This paper utilizes data from a cosmetic e-commerce website. Due to the existence of massive data with many attributes and complexity in the data source, we need to extract the used data, including customer information tables, product information tables, and customer order tables. Derived from the database.

The customer information is shown in Table 1. The customer information table stores the basic information of the customer: customer ID, customer name, registration date, all points, age, gender, phone number, and mailing address. This experiment only uses the customer ID, age, and mailing address fields. The table is the field-filtered customer information table.

Table 1. Customer information

Attribute Name	Data Type	Length
Customer ID	Char	18
Gender	Char	1
Age	Int	2
Education	Char	1
Locus	Char	1

Commodity information is shown in Table 2. The commodity information table stores the basic information of the commodity: commodity ID, commodity name, commodity type, commodity sales price, commodity discount promotional price, commodity shelf time, commodity inventory, and other information. This experiment is only used in the commodity ID, commodity name, commodity type, and commodity sales price. The table is the field-filtered commodity information table.

Table 2. Commodity information

Attribute Name	Data Type	Length
Customer ID	int	4
product_name	char	20
product_class_id	int	4
product_class_id	char	5

Commodity categories are shown in Table 3. The commodity category table stores information about the category to which the commodity belongs: category ID, category name, parent category ID, and so on. Only the category ID and category name fields are utilized in this experiment, and the table is the commodity category information table after field filtering.

Table 3. Commodity information

Attribute Name	Data Type	Length
product_class_id	int	4
product_category	char	20
parent_id	int	4

The customer order information is shown in Table 4. The Customer Order Information table stores the history of the customer's purchase of goods, including customer ID, product ID, quantity of goods purchased, price of goods, discount rate, total price of the order, and date of purchase. The table displays the customer order information table.

Table 4. Customer order information

Attribute Name	Data Type	Length
customer_id	char	18
product_id	int	4
price	money	8
number	int	4
unit_price	money	88
deal_date	datetime	4

3 Results and discussion

In order to further explore the specific application and effectiveness of the SAPK-means algorithm proposed in the previous section in e-commerce user segmentation, the e-commerce platform of Company A is selected for experimental testing in this section, and from the experimental results, we explore the customized marketing strategies for different e-commerce customer groups.

3.1 Analysis of clustering results

K-means clustering of customer groups of Company A's e-commerce platform using SPSS software to obtain five types of segmented customer groups, which are defined in this paper as A, B, C, D, and E. The number of samples included in each type of segmented customer group is shown in Table 5.

Table 5. The sample size of various segmentation customer groups

Subdivision	A	32512
	B	63354
	C	127000
	D	75312
	E	45520
In Effect		332152
Deletion		0

Company A's e-commerce platform of five types of segmented customer groups of customer basic information and consumer behavior characteristics of in-depth mining analysis to get the company's e-commerce platform segmented customer groups of customer characteristics and product preferences, segmented customer groups of customer characteristics and product preferences as shown in Table 6.

Table 6. Customer characteristics and product preferences of customer base

Segmentation customer base number	Customer ratio (%)	Average purchase frequency (times)	Average purchase gold (yuan)	Average purchase type (species)	Last purchase time (day)	Time (year)
A	9.46	25.19	27414	10.63	15	7.18
B	18.43	18.7	15371	20.47	6	4.88
C	36.95	36.11	3467	19.2	3	5.18
D	21.91	11.53	1207	15.03	22	1.59
E	13.24	4.52	219	6.44	31	1.59

Analysis of the characteristics of class A segmented customer groups: most of the customer groups in this category are men aged 35 to 55, most of whom have a bachelor's degree or above, and their units are mainly state organs and state-owned enterprises, with an annual income of more than 150,000 yuan, and the number of customer groups accounts for 9.46% of the total number of samples. The biggest consumption behavior of Class A segmented customer groups is that the consumption frequency is high, the amount of single consumption is large, and the types of products consumed are mainly high-end goods or services such as large-amount bank wealth management, automobiles, and precious metals. The A-type segment customer group belongs to the customer group with the largest value contribution to the e-commerce platform of Company A, which can be called the "platinum customer group".

Analysis of the characteristics of the customer group of category B segment: most of the customer groups of this type of customer segment are women aged 26 to 40, most of whom have a bachelor's degree or above, and their units are mainly state organs and state-owned enterprises, and the number of customer groups accounts for 18.43% of the total number of samples. The biggest consumption behavior of the B segment customer group is that the consumption frequency is higher. Still, the average single consumption amount is slightly lower than that of the A segment of the customer group. The main types of goods consumed are mid-to-high-end digital products, mid-end brand cosmetics, bank wealth management products, high-end clothing, etc., and the types of consumer goods are also relatively rich. The contribution of the B-type segment customer group to the value of Company A's e-commerce platform is slightly lower than that of the A-type segment customer group, which can be called the "golden customer group".

Analysis of the characteristics of the C segmented customer groups: the majority of the segmented customer groups the 26 to 40-year-old men, most of their educational level in a bachelor's degree and above, the unit is mainly in the private sector or self-employed businesses, the number of customer groups accounted for the total number of samples of 36.95%. Segmenting customer groups' consumption behavior is characterized by the high frequency of consumption, average amount of single consumption, and average level of total sample. The average single consumption amount is basically the same as the average level of the total sample. The goods consumed by the customers are in the low and medium grades. The types of goods consumed are mainly concentrated in electronic products, household appliances, clothing, shoes, hats, food and beverages, etc. The types of goods consumed are also relatively rich. The contribution of the C-class customer group to the value of the e-commerce platform of Company A is not high, and there exists a larger space for improvement, which can be called the "silver customer group". It can be called the "silver customer group".

Analysis of the characteristics of the customer group of the D segment: most of the customer groups of this type of segment are women aged 18 to 30, most of their educational level is at the bachelor's degree or above, and the units are mainly private enterprises, and the number of customer groups accounts for 21.91% of the total number of samples. The consumption behavior of the D segment customer group is characterized by low consumption frequency. The average single consumption amount is also low, but the types of consumer goods are relatively rich. The main level of goods consumed by this type of segmented customer group is low-grade goods, and the types of goods consumed are mainly clothing, shoes and hats, food and beverages, cosmetics, mother and child, etc. The Type D segment customer group has a low retention time and a low contribution to the value of Company A's e-commerce platform, which can be called a "copper customer group".

Analysis of the characteristics of the customer group of the E segment: most of the customer groups of this type of segment are men aged 26 to 40, most of whom have an educational level below high school, and their units are mainly private enterprises or freelancers, and the number of customer groups accounts for 13.24% of the total number of samples. The average consumption amount and consumption frequency of the E segment customer group are much lower than the average consumption amount and average consumption frequency of the total sample, and the retention time is also much lower than the average retention time of the total sample and the types of goods consumed by users include food and beverage, clothing, recharge, and payment products. The E-segment customer group has strong liquidity, easy to churn, and has a low contribution to the value of Company A's e-commerce platform, which can be called an "iron customer group".

3.2 Presentation of customized marketing strategies

3.2.1 Orientation from user needs

A company in the big brands and genuine goods is the old users of overseas goods the most important demand, at present A company to the global direct mode, and Japan, South Korea, Australia and other well-known suppliers at home and abroad to cooperate, in-depth product origin direct mining high-quality and meet the needs of Chinese consumption of goods, to eliminate counterfeits from the source, 100% genuine commitment, but at present most of the goods are domestic non-famous brands, A company can Company A can continue to broaden the cooperation with brand-name companies to screen more brand-name goods from the source to meet customer demand.

3.2.2 Positioning from user pain points

As a cross-border e-commerce company, company A's ultra-cost-effective genuine products are the main needs and pain points of users choosing goods, and at present, company A's commodity prices do not have much competitive advantage with large e-commerce platforms.

3.2.3 Finding Differentiation from Competitor Positioning

Company A, as a membership e-commerce platform that can refund the membership fee, relatively speaking, has certain differences and advantages over the general membership e-commerce company, but for new users, even if the membership fee can be refunded to a certain period, due to the lack of trust in the platform, the operation of the recharge to become a member has been intercepted by the majority of customers. Company A can do two sets of membership systems. One is the recharge membership system, once the user can become a member to enjoy the benefits of membership. The other is the growth value membership system. Through a simple task mode, slowly cultivate the user's trust. You can become a member of the value of the growth of the two membership systems to form a good match. You can better emphasize the differences in the brand of the A company.

3.3 Changes in Company Sales Growth

Company A developed marketing strategies based on the e-commerce user subgroups. The implementation of marketing strategies was implemented after company A's annual sales achieved a certain growth change. 2014-2023 sales change in company A is shown in Figure 5. Company A's 2014-2023 annual sales showed a steady increase in the trend year by year. The annual sales in 2023 are to reach the peak since the company was founded, achieving 44,261,500 yuan of total sales, but the growth rate is also gradually decreasing each year, with the end of 2020 year-on-year growth of only 10.26%, reaching a new low in the company's history.

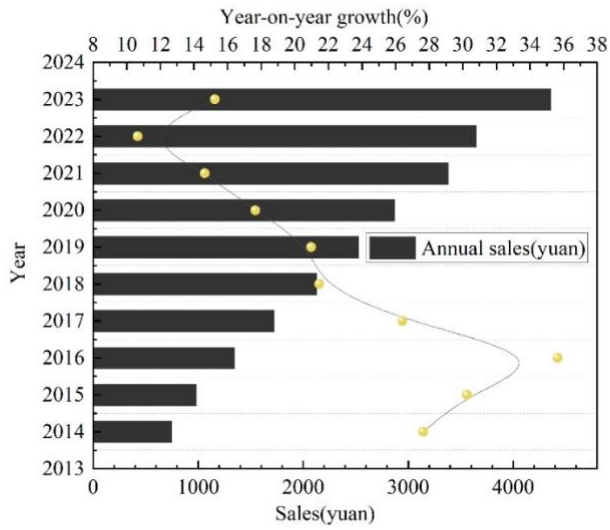


Figure 5. Sales change

3.4 Consumer Satisfaction Analysis

In the context of Company A's marketing strategy development based on e-commerce customer subgroups, this section continues to test consumer satisfaction with 30 daily-use products on Company A's e-commerce platform in order to more intuitively see the distribution of consumer satisfaction and the percentage of good and bad reviews of each product, a satisfaction distribution chart of the 30 daily-use products is drawn, and the distribution of user satisfaction is shown in Figure 6. It is found that among all the comment texts, positive sentiment accounts for 92.142%, and negative sentiment accounts for 7.858%, indicating that most consumers are more satisfied with these 30 daily-use products, and the overall attitude is positive. Four of the products (product numbers 7, 11, 23, and 25) had positive sentiment percentages of 100%, 98.88%, 98.128%, and 100%, respectively, indicating that consumers who purchased these four products were very satisfied with them. The other product, product number 2, has lower consumer satisfaction with a 41.206% negative sentiment, which indicates that there are more problems with this product and there is great scope for improvement.

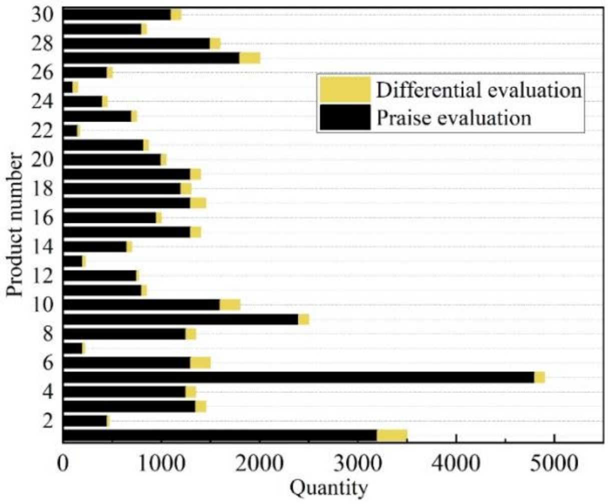


Figure 6. User satisfaction distribution

4 Conclusion

The article explores the specific application and effectiveness of the SAPK-means algorithm in e-commerce user clustering, which is derived from the experimental study of this paper:

1) In the analysis of the clustering results, the customer groups of Company A's e-commerce platform can be divided into five categories: A, B, C, D, and E. The number of customer groups in category A accounts for 9.46% of the total number of samples, which has a high consumption frequency and a large amount of single consumption and is known as the "platinum customer group".

2) In the context of the marketing strategy relying on e-commerce user subgroups, the change in the annual sales growth of Company A from 2014 to 2023 shows a trend of steady growth year by year, and in 2023, the total sales of 44,261,500 yuan were realized.

3) In the context of marketing strategy relying on e-commerce user subgroups, in the analysis of consumer satisfaction, the positive sentiment accounted for 92.142%, and the negative sentiment accounted for 7.858%, which can be obtained that most consumers are more satisfied with the products of Company A as a whole.

References

[1] Deepak, G., & Kasaraneni, D. (2019). OntoCommerce: an ontology focused semantic framework for personalised product recommendation for user targeted e-commerce. International Journal of Computer Aided Engineering and Technology, 11(4-5), 449-466.

[2] Alrumiah, S. S., & Hadwan, M. (2021). Implementing big data analytics in e-commerce: Vendor and customer view. Ieee Access, 9, 37281-37286.

[3] Rao, H. K., Zeng, Z., & Liu, A. P. (2018, May). Research on personalized referral service and big data mining for e-commerce with machine learning. In 2018 4th International Conference on Computer and Technology Applications (ICCTA) (pp. 35-38). IEEE.

[4] Huseynov, F., & Yıldırım, S. Ö. (2017). Behavioural segmentation analysis of online consumer audience in Turkey by using real e-commerce transaction data. International Journal of Economics and Business Research, 14(1), 12-28.

[5] Shen, B. (2021, January). E-commerce customer segmentation via unsupervised machine learning. In The 2nd international conference on computing and data science (pp. 1-7).

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

[7] Alves Gomes, M., & Meisen, T. (2023). A review on customer segmentation methods for personalized customer targeting in e-commerce use cases. Information Systems and e-Business Management, 21(3), 527-570.

[8] Wang, X., Wong, Y. D., Teo, C. C., Yuen, K. F., & Feng, X. (2020). The four facets of self-collection service for e-commerce delivery: Conceptualisation and latent class analysis of user segments. Electronic Commerce Research and Applications, 39, 100896.

[9] Puspitasari, N. B., Pramono, S. W., Rinawati, D. I., & Fidiyanti, F. (2020). Online consumer segmentation study based on factors affecting e-commerce selection. In IOP Conference Series: Materials Science and Engineering (Vol. 722, No. 1, p. 012036). IOP Publishing.

[10] 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.

[11] Kamthania, D., Pawa, A., & Madhavan, S. S. (2018). Market segmentation analysis and visualization using K-mode clustering algorithm for E-commerce business. Journal of computing and information technology, 26(1), 57-68.

[12] SUNIL, S., & Mahendira, T. P. M. (2022, May). E-COMMERCE USER SEGMENTATION. In 2nd intemational Conference on innovative Research in Engineering and Technology (ICIRET-22), Vietnam (pp. 15-16).

[13] Punhani, R., Arora, V. S., Sabitha, S., & Shukla, V. K. (2021, March). Application of clustering algorithm for effective customer segmentation in E-commerce. In 2021 International Conference on Computational Intelligence and Knowledge Economy (ICCIKE) (pp. 149-154). IEEE.

[14] Rosário, A., & Raimundo, R. (2021). Consumer marketing strategy and E-commerce in the last decade: a literature review. Journal of theoretical and applied electronic commerce research, 16(7), 3003-3024.

[15] Gyenge, B., Máté, Z., Vida, I., Bilan, Y., & Vasa, L. (2021). A new strategic marketing management model for the specificities of E-commerce in the supply chain. Journal of Theoretical and Applied Electronic Commerce Research, 16(4), 1136-1149.

[16] Raza, R., Narvel, F., & Shetti, G. (2022). A Study of Attaining Resilience through Digital Marketing Strategies and Customised Marketing Approach in the E-Commerce Sector. Quest Journal of Management, 13(1), 1-9.

[17] Zhang Hongwei,Xia Yuanyou,Lin Manqing,Huang Jian & Yan Yaofeng. (2024). A three-step rockburst prediction model based on data preprocessing combined with clustering and classification algorithms. Bulletin of Engineering Geology and the Environment(7).

[18] Lingzhi Jiang,Qiwu Wu & Zhuoxu Xie. (2024). Risk Assessment of Terrorism in Asia based on K-Means Clustering Analysis. Frontiers in Computing and Intelligent Systems(2),43-46.

[19] Jiyong Liao,Xingjiao Wu,Yaxin Wu & Juelin Shu. (2024). K-NNDP: K-means algorithm based on nearest neighbor density peak optimization and outlier removal. Knowledge-Based Systems111742-.

[20] Wan Benting,Huang Weikang,Pierre Bilivogui,Cheng Youyu & Zhou Shufen. (2024). K-Means algorithm based on multi-feature-induced order. Granular Computing(2).