Precision Marketing Analysis Based On HC K-means Algorithm

Jingwen Li

Hebi Vocational College of Energy and Chemistry, Hebi 458010, China

Corresponding author. E-mail: 15939259107@163.com

Received: Apr. 8, 2026; Accepted: May 25, 2026

In order to improve the sales level of e-commerce platform products and promote the healthy development of e-commerce platform, a construction method of e-commerce affective precision system based on Hierarchical Clustering K-means algorithm (HC K means) is proposed. This method first combines the weighted Euclidean distance with the Recency, Frequency, and Monetary (RFM) model and adopts an entropy-weighted method to improve the evaluation dimension of customer value. Further, the density index and the maximum-minimum distance method are introduced to optimize the selection of initial cluster centers, and a dynamic radius adjustment mechanism is designed to adapt to the data distribution, to solve the defects of the traditional algorithm being sensitive to the initial centers and having unstable clustering results. Experimental results based on 100,000 real transaction records from an e-commerce platform show that the HC K-means algorithm has a customer data classification detection rate of 91.83%, which is 10.86 and 6.54 percentage points higher than GMDA (80.%) and TSCEA (85.29%), respectively. The false detection rate is only 2.91%, which is lower than GMDA (5.86%) and TSCEA (8.92%). In customer value segmentation, the accuracy of potential customer value classification is 93%, the current value classification accuracy of core customers is 91%, and the customer loyalty classification accuracy is 93%. Main contributions: An improved HC K-means clustering framework is proposed, through density-guided initial center selection and dynamic radius adjustment, significantly improving the accuracy of e-commerce customer segmentation and marketing conversion efficiency, providing quantifiable decision support for enterprises' precise marketing.

Keywords: Statistics; K-means clustering algorithm; e-commerce platform; affective precision; RFM system; HC K-means algorithm

http://dx.doi.org/10.6180/jase.202610_33.054

1. Introduction

With the rapid development of the Internet and e-commerce, competition among e-commerce platforms (ECP) is becoming increasingly fierce [1, 2]. To maintain competitive advantage, ECP requires precise marketing analysis to understand customer needs and provide personalized services. Affective precision is a data-driven approach that segments customers into different groups to send targeted marketing messages. However, existing online marketing processes face several challenges. Traditional customer classification methods lack accuracy and efficiency, resulting in wasted marketing resources due to untargeted strategies. Moreover, clustering algorithms like K-means are prone to local optima, leading to unstable and biased customer classification. Additionally, the lack of a scientific weighting method for classification indicators reduces the reliability of customer segmentation, making it difficult to identify high-value customers and their potential needs [3–6].

Current research on affective precision has several shortcomings. Some approaches heavily depend on big data, facing privacy and processing limitations [7]. The AISAS

model may be insufficient in data dimensionality and real-time performance [8]. Sales prediction systems may have limited forecasting accuracy and generalization ability [9]. Marketing strategies may lack rapid response to market dynamics [10]. To address these limitations, Chatterjee et al. proposed an integrated webpage prediction technique using K-means clustering to categorize web log datasets, effectively reducing the impact of user delays in e-commerce [11]. Liu et al. proposed various generalization methods based on the K-means algorithm, including changing data representation, distance measurement, and label allocation, to solve complex clustering problems such as iterative subspace projection and constrained clustering [12]. To improve the intelligent evaluation of e-commerce transaction volumes, a system combining K-means and SOM algorithms was proposed, achieving a better denoising effect and an AUC value of 0.9712 [13]. Wang proposed the Dynamic Update Hash Index (DUHI) clustering method for DNA storage, which reduces sequence redundancy by more than 10% and improves sequence reconstruction rate to over 99% [14].

To compensate for the deficiencies of existing methods, this study proposes an affective precision analysis method combining HC K-means with the RFM model. By introducing weighted Euclidean distance and entropy-based weight allocation, the accuracy and objectivity of customer classification are improved. The dynamic radius adjustment mechanism and density index enhance the algorithm's adaptability and robustness. Different from existing studies that focus only on system integration, the algorithmic innovation lies in density-guided initial center selection and dynamic radius adaptive adjustment. This assumption breaks through the theoretical limitations of traditional K-means, which is sensitive to initial centers and has poor adaptability to uneven density. Experimental evidence demonstrates that HC K-means significantly outperforms existing benchmark methods in classification detection rate and error rate.

2. Materials and methods

To provide data support for affective precision, this study uses cluster analysis of e-commerce customer data to classify customers based on their purchasing behavior, geographic location, age, gender, interests, and other characteristics, to better identify the characteristics and behavioral patterns of different customer groups. By applying the HC K-means algorithm to the user data of ECP, a system that can realize affective precision is constructed. Following the workflow of data preprocessing, feature extraction, cluster analysis and performance evaluation, Pandas is used successively for data cleaning and Z-score standardization. Based on the RFM model, customer feature vectors are extracted. The HC K-means clustering algorithm and comparison algorithm are implemented using Scikit-learn. Finally, the performance is evaluated through indicators such as classification detection rate and error rate. The system can effectively divide users into different groups and develop personalized marketing strategies for the characteristics of different groups.

2.1. modeling

The K-means algorithm is also easy to obtain the local optimal solution. Therefore, the study uses data similarity as a metric in the clustering process for the classification of e-commerce customers. To describe the similarity metric, the study introduces the Euclidean distance to describe the similarity metric of K-means algorithm for e-commerce customer classification [15]. The definition of Euclidean distance can be expressed in Eq. (1).

$$d\left(z_u, z_w\right) = \sqrt{\sum_{j=1}^{N_d}\left(z_{u,j} - z_{w,j}\right)^2} \tag{1}$$

In Eq. (1), z_u and z_w denote the feature vectors of the e-commerce customer data samples, respectively; $z_{u,j}$ denotes the data value corresponding to the j th feature vector in the data samples; $z_{w,j}$ denotes the value of the j th feature vector in the data samples; and N_d denotes the number of feature vectors. In the data processing of different e-commerce customer objects, there are variable differences in the data of different objects. To determine the reliability of the similarity measure calculation, the study performed a weighting operation on the Euclidean distance, and the weighted Euclidean clustering can be calculated by Eq. (2).

$$d\left(z_u, z_w\right) = \sqrt{\sum w_j\left(z_{u,j}, z_{w,j}\right)} \tag{2}$$

In Eq. (2), w_j denotes the weight value of the j th sample in the customer data sample. The method of selecting the initial cluster center based on density index and maximum minimum distance method in this study can be applied multiple times, and then the clustering results obtained each time are compared to select the optimal combination of initial cluster centers. This can further improve the accuracy of the algorithm, reduce clustering result bias caused by improper selection of initial clustering centers, and thus enhance the performance of the algorithm in handling complex data. The RFM model is based on three key indicators - Recency, Frequency, and Monetary. The study utilizes RFM to segment the information of customers that need to be collected, and the model segments e-commerce customers

from the three dimensions of purchase interval, number of purchases and purchase amount [16]. The study combines the consumption characteristics of e-commerce customers and assigns different weights to the segmentation indexes of e-commerce customer groups from the three dimensions, so as to more accurately assess the value of each customer. A schematic diagram of the RFM model is shown in Fig. 1.

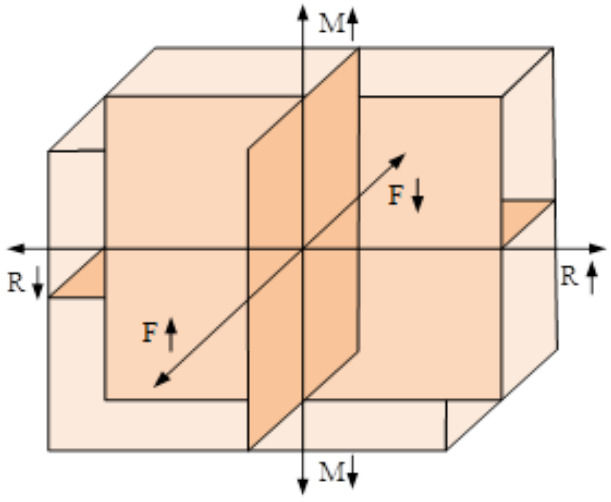

Fig. 1. Schematic diagram of RFM model.

In carrying out the assignment process, the study introduces the entropy value method, this is because the entropy value method is an objective weight determination method. The study utilizes the entropy value method to assign weights to the three indicators in the classification of e-commerce customers [17, 18]. The study assumes that the original data in the e-commerce customers is m and the corresponding index for classification is n, which enables the entropy value method to be utilized for the classification study of e-commerce customer data. In the study of e-commerce customer classification, the relevant original index data needs to be preprocessed first to ensure data consistency. The preprocessing process is divided into two parts: positive processing and negative processing. The positive processing can be represented by Eq. (3).

$$x_{ij}^{*} = \frac{x_{ij} - \min\left(x_j\right)}{\max\left(x_j\right) - \min\left(x_j\right)} \tag{3}$$

In Eq. (3), x_{ij} denotes the j th categorical observation corresponding to the i th studied client in the original data; $\min\left(x_j\right)$ denotes the observed minimum value; and $\max\left(x_j\right)$ denotes the observed maximum value. The negative treatment can be expressed in Eq. (4).

$$-x_{ij}^{*} = \frac{\max\left(x_j\right) - x_{ij}}{\max\left(x_j\right) - \min\left(x_j\right)} \tag{4}$$

After the positive and negative processing of the raw data, it is possible to calculate the information entropy of the customer categorization index, which can be calculated using the Eq. (5).

$$H_j = -K \sum_{i=1}^{m} Y_{ij} \ln Y_{ij} \tag{5}$$

In Eq. (5), K is a constant coefficient, which takes the value of $\frac{1}{\ln(m)}$; Y_{ij} represents the information entropy value of j corresponding to the i customer under study in the customer classification. In this case, j corresponds to the weight of the customer classification index, which can be expressed in Eq. (6).

$$w_j = \left(1 - H_j\right) / \left(n - \sum_{i=1}^{n} H_j\right) \tag{6}$$

The study is carried out on the new data obtained after weighting the indicators using the entropy value method, and the optimal number of clusters is determined by clustering several times. In the process of determining the optimal number of clusters, experiments were conducted with K values ranging from 2 to 10 . When K = 6, the average intra-cluster distance decreased to 0.1588 , and the data generalization ability reached 0.94 . Further increasing K beyond 6 yielded no significant improvement in either metric. Therefore, the number of clusters was set to 6 for subsequent customer classification.

2.2. E-commerce affective precision system based on HC K-means

First, the RFM model is used to subdivide customers from three dimensions: purchase interval, purchase frequency, and purchase amount. Next, Python is used for data processing, obtaining a total of 3,000 samples. The dataset includes the following fields: Order ID (unique identifier), Item Title (product name), Purchase Interval (integer, days), Purchase Frequency (integer), Purchase Amount (float), and optional fields such as Geolocation, Age, Gender, Interest Categories, and User Registration Date. For each customer, statistical measures such as average purchase interval, total purchase frequency, and total purchase amount are calculated to reflect overall consumption characteristics. The data then undergoes cleaning (removing incomplete or canceled orders), conversion (calculating derived features), and Z-score standardization to ensure comparability across

features. The processed data will be used for subsequent cluster analysis [19].

On the basis of e-commerce customer classification, using the density index and maximum minimum distance method, the clustering center is effectively improved, and the sample point with the highest density is selected as the starting center of clustering. Density is defined by Eq. (7).

$$\rho(x) = \frac{1}{\sum_{y \in N_r(x)} \|x - y\|^2} \tag{7}$$

In Eq. (7), $N_r(x)$ is all the points in a neighborhood centered on x and of radius r. The sample-point with the highest density can serve as the initial center of the clustering. In addition, a dynamic radius adjustment mechanism, which can adaptively adjust the radius of a neighborhood according to the distribution characteristics of the data, is introduced. Fig. 2 is a schematic diagram of the cluster analysis based on the density index and the maximum minimum distance clustering.

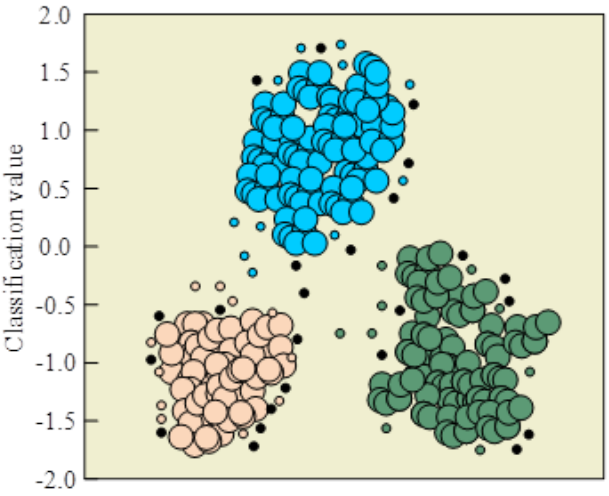

Fig. 2. Analysis diagram based on density index and maximum minimum distance clustering.

From Fig. 2, The light blue cluster at the upper left corner shows higher classification values, indicating a dense concentration of points with similar characteristics. The orange cluster at the lower left corner shows lower classification values, with points scattered slightly between them. The green cluster, located in the lower right corner, is in between in terms of classification values, but also shows a degree of variability. Although there is some overlap, especially in the middle region of the classification value space, clusters often reflect the true categories of data points. The clustering method used makes use of the density index and the maximum minimum distance metric. Density index captures local density features of data points, helping to identify arbitrarily shaped clusters. At the same time, the maximum-minimum distance ensures a clear boundary between clusters by measuring the separation between clusters. Although high dimensional data can pose challenges to clustering algorithms, the key is how the data is modeled, not the algorithm itself [20]. This combination enables accurate and robust clustering in high dimensional data spaces. In the affective precision system, it is necessary to first calculate to get the average distance of all customer elements in the system data, which can be expressed by Eq. (8).

$$d_c = \frac{2}{g(g-1)} \sum_{i=1}^{n} \sum_{j=i+1}^{n} d_{ij} \tag{8}$$

In Eq. (8), g denotes the number of sample data in the data set of the system; d_{ij} denotes the Euclidean distance from x_i to x_j of the data in the data set. At this time, the local density of the data in the system can be calculated by Eq. (9).

$$\rho_i = \sum x \left(d_{ij} - d_c \right) \tag{9}$$

In Eq. (9), χ denotes the random variable in the density. The study addresses the limitations of density point selection in the first initial sample by calculating the density point with the largest data set in the system sample. A new HC K-means algorithm is proposed, whose algorithm flow is shown as: Step 1: Input the data set and number of clusters, Step 2: Compute the sample average distance, and Step 3: Compute the sample density value, for each sample object xi in the data set Z, based on the definition, which is the number of sample objects in the area centered on the sample object xi and with the radius of dc. Eq. (7) is quoted to define the density, and Eq. (9) is used to calculate the local density of the data in the system. Step 4: Select the first initial cluster center, find out the point pi with the greatest density in the sample, and use it as the first initial cluster center, ci. Step 5: Select other initial cluster centers, and select the remaining initial cluster centers according to the maximum-minimum distance criterion combined with the density method. Using the minimum-maximum distance method, the point with the farthest distance and highest density from the selected cluster center is selected as the next initial cluster center in the remaining samples. The object with the largest density value in all sample points within the range of sample point x_j mean distance dc is selected as the next initial cluster center, and this process is repeated until the K th initial cluster center, ck, is found. Step 6: construct the sample distribution matrix and calculate the feature weights, construct the sample distribution matrix and calculate the proportion of each sample under

each feature, as shown in Eq. (10).

$$P_j^{(i)} = \frac{x_j^{(i)}}{\sum_{j=1}^{n} x_j^{(i)}} \tag{10}$$

In Eq. (10), $P_j^{(i)}$ denotes the specific gravity of the j th sample at the i th characteristic. This specific weight reflects the relative importance or contribution of the sample to a particular feature. $x_j^{(i)}$ indicates the value of the j th sample on the i th characteristic. In data set Z, each sample consists of multiple features, here specifically referring to the specific value of the j th sample on the i th feature. n represents the total number of samples in the data set. Then, calculate the proportion of each sample under each feature, the information entropy of the attribute, the difference coefficient on the feature dimension, and the weight of the feature. After obtaining the feature weights, the entropy of the customer attribute information can be calculated, as shown in Eq. (11).

$$H(i) = -\sum_{i=1}^{n} P_j^{(i)} \log P_j^{(i)} \tag{11}$$

At this time, the coefficient of variation of the dimension of e-commerce customer characteristics can be expressed by Eq. (12).

$$q_i = 1 - H(i) \tag{12}$$

After calculating the coefficient of variation of the feature dimension, the clustering weights in the e-commerce customer affective precision system can be defined, as shown in Eq. (13).

$$\omega_i = \frac{q_i}{\sum_{i=1}^{m} q_i} \tag{13}$$

Step 7: Compute the weight distance, computing the distance from each sample xi in the data set to the cluster center ci. Step 8: Categorize sample points, classifying individual sample points into individual clusters based on weight distances. Step 9: Compute the mean of objects of the same class, computing the mean of objects within each cluster. Step 10: Update the cluster center with the mean value of the objects within the cluster as the new cluster center. Step 11: Judge if the cluster center changes or reaches the maximum number of runs, repeating steps seven to ten until the resulting cluster center reaches the maximum number of runs or no more changes. Step 12: output the results, output the clusters formed by clustering.

From this study, an e-commerce affective precision system based on HC K-means is constructed. To implement the system, Python is first used in the study combined with relevant data processing and machine learning libraries, such as Pandas for data cleaning and pre-processing, and Scikit-Lean implements K-means and HC-K means clustering algorithms. Flask or Django are used to build web services for user interaction and data presentation. To ensure real-time performance, research was done to increase computational efficiency by using frameworks such as Apache Hadop to distribute data and computing tasks to multiple servers for parallel processing. When implementing this solution, the MEM Caching Distributed Caching System was used to cache commonly used data to reduce double counting and data read time. The front end uses HTML, CSS, and Javascript to implement the user interface, and the back end uses AJAX to exchange data asymmetrically, improving the user experience. The system implementation assumes that the customer consumption behavior maintains a stable pattern in the short term and that the RFM indicators can fully represent the customer value. The key parameters include a clustering number K = 6, a density radius coefficient α = 1.0, a high-density threshold of 1.2, a low-density threshold of 0.8, and a maximum iteration number of 100. The research is based on 100,000 real transaction data from a certain ECP. The customer feature vectors are constructed using the RFM model, and HC K-means is compared with GMDA, TSCEA, K-means and KNN. The implementation process includes data standardization, density-guided initial center selection, and dynamic radius iterative clustering. The performance is evaluated using classification detection rate, error rate, retention rate, complaint rate and customer value classification accuracy. Each group of experiments is repeated 30 times and the mean and standard deviation are calculated.

3. Results and discussion

To validate the performance of the affective precision system for ECP, the study obtained relevant public data of customers in an e-commerce company, which include registration, order, comment, browsing, activity and conversion data. This information was used in the construction of the dataset for validating the performance of the affective precision system. The following experiments were designed to verify the performance of the proposed algorithm. The main purposes included evaluating the detection rate and error rate of customer classification, analyzing the ability to segment customer value, testing the classification effect of consumption habits, assessing the enhancement effect on product sales after system application, and examining the stability of the algorithm through parameter sensitivity analysis. The study used K-Nearest Neighbor algorithm and K-means algorithm to compare with HC K-means in the system.

3.1. Performance analysis of ECP affective precision system

To verify the effectiveness of the ECP precision marketing system constructed in this study in the process of classifying and processing customer data in e-commerce, the GMDA clustering algorithm proposed in the latest literature [21], the two-stage clustering ensemble algorithm (TSCEA) proposed in literature [22], and the HCK-means method (random scan clustering efficiency ERSC) used in literature [23] were adopted for comparative experiments. Comparison of the detection rate and detection error rate of data during customer classification using three methods is shown in Fig. 4.

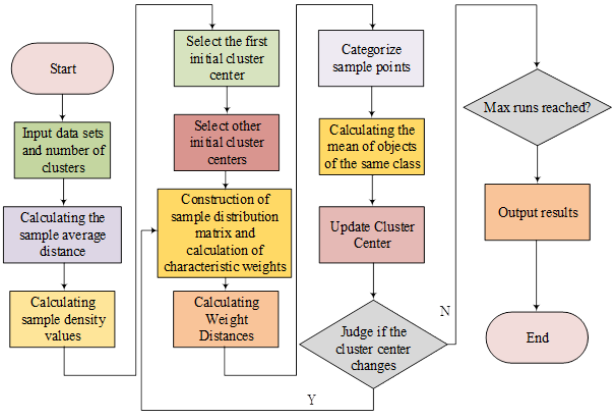

Fig. 3. Location of the KVMRT project

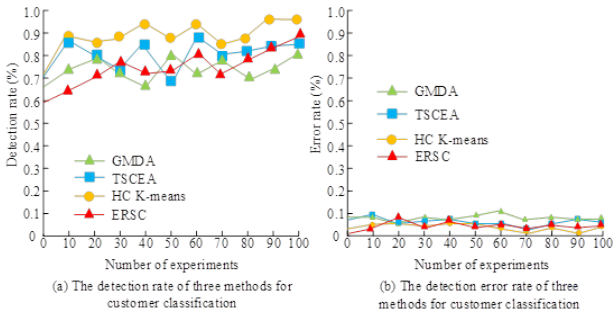

Fig. 4. Comparison results of detection rate and detection error rate of data during customer classification process.

From Fig. 4(a), the classification of e-commerce customer data fluctuated as the number of detections increased. The detection rate of HC K-means was 91.83% for customer data, and 80.97% for GMDA and 85.29% for TSCEA. From Fig. 4(b), the false detection rates for customer data by all three methods were low, all less than 10%, indicating that all three methods had some accuracy. HC K-means had an error rate of 2.91% for detecting customer data, and 5.86% for GMDA and 8.92% for TSCEA. ERSC performed the worst. This indicates that the HC K-means method used to construct the model has high reliability. To verify the ability of the affective precision system to classify the value of customers on the ECP, the study took the potential value, current value and loyalty of customers as indicators and analyzed them using the system. The study analyzed a 2021 sales dataset from a store provided by Alibaba's Tianchi platform. This dataset contains over 100,000 transaction records, covering 11 key fields including order number, transaction time, payment platform, and order amount, and includes known customer classification tags. This test set can be a subset of the entire dataset, where each customer has a real classification label. As shown in Eq. (5), it is the result of analyzing customer value through a affective precision system.

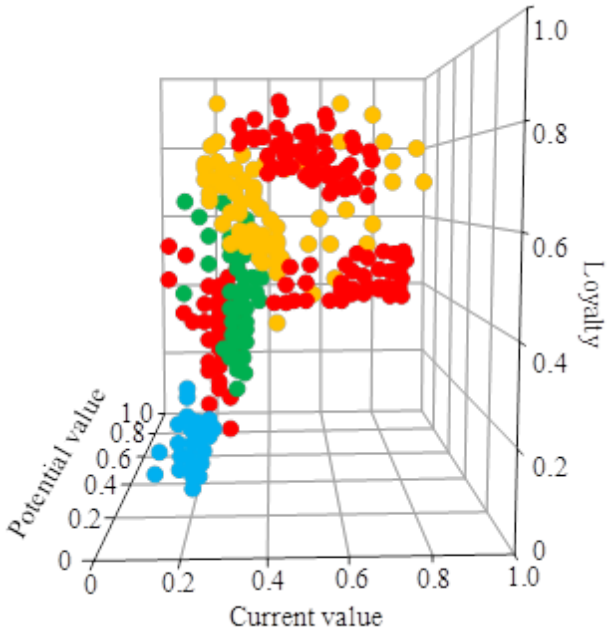

Fig. 5. Analysis results of customer value in affective precision systems.

In Fig. 5, blue represents potential customers, red represents core customers, green represents important customers, and orange represents general customers. Through the analysis of the customer types in the figure, it was found that the potential value of potential customers was the highest, 0.93; the current value of the highest core customers was 0.91; the loyalty of the highest general customers and core customers, the two were 0.93 and 0.92 respectively. The customer value system could effectively classify existing customers in ECP. The categorization of customers was conducive to the formulation of corresponding marketing strategies for different customer groups, so as to better

provide strong support for the development of the ECP. The above customer value segmentation results not only verified the effectiveness of the HC K-means algorithm in differentiating various value groups, but also revealed the essential differences in value drivers between potential customers and core customers: The high value of potential customers mainly stems from the improvement of their recent activity levels, while the high value of core customers is maintained by both high frequency and high amount of transactions. This provides a direct basis for the subsequent formulation of group incentive strategies. To verify the ability of the affective precision system to classify different customer consumption habits in the clustered data, the study also utilized the three methods mentioned above for comparison. The density peak clustering proposed by [24] and [25] was introduced. As shown in Fig. 6, the results of the classification of customer spending habits in the clustered dataset by three methods are shown.

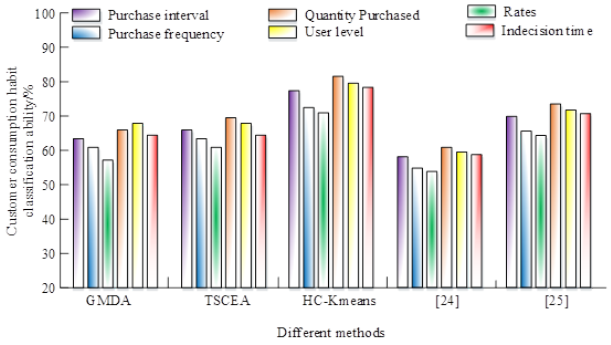

Fig. 6. Classification results of three methods for customer consumption habits in clustered datasets.

From Fig. 6, the classification results of GMDA and TSCEA were closer in terms of purchase spacing, at about 65% and 68% respectively, while HC K-means showed higher classification capability, reaching about 75%. In terms of purchase frequency, the classification result of GMDA was slightly lower at about 62%, while TSCEA and HC K-means were about 66% and 68% respectively. At user level, the HC K-means method showed a significant advantage, with classification results close to 80%, far higher than 70% for GMDA and 73% for TSCEA. This indicates that HC K-means is more efficient at handling user level data. In terms of uncertain time, the classification result of GMDA was slightly lower at about 63%, while TSCEA and HC K-means achieved 67% and 72% respectively, showing the advantages of the latter two methods in processing uncertain time data. As for the rate classification, GMDA and TSCEA performed similarly at about 64% and 65% respectively, while HC K-means was slightly lower at about 63%. However, literature [24] and literature [25] performed poorly, with an average of 61% and 65%, respectively. Overall, the HC K-means method shows a higher ability to categorize customer spending habits on most indicators, especially in terms of purchase quantity and decision time, followed by the TSCEA method and the GMDA method is relatively weak. This indicates that the classification of the affective precision system on the customer consumption system has a more reliable classification ability and can provide reliable data for the development of affective precision strategies. To further validate the performance of the affective precision system, the study took the retention rate and complaint rate of e-commerce customers as validation indicators for performance testing. The classification ability of the three methods on customer retention rate and complaint rate is shown in Fig. 7.

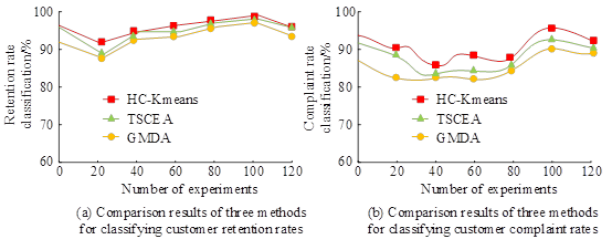

Fig. 7. The classification ability of three methods for customer retention rate and complaint rate.

From Fig. 7(a), the three approaches did not differ significantly in the classification of customer retention rates, but HC K-means' still had some advantages. Of these, HC K-means classified customer retention rates as 96.08%, and GMDA and TSCEA had 93.26% and 91.84% capability to categorize customer retention rates, respectively. From Fig. 7(b), HC K-means classified the customer complaint rate as 90.13%, and GMDA and TSCEA had 87.24% and 85.47% respectively.

3.2. Application performance analysis of ECP affective precision system

To verify the performance of the application of the affective precision system of the ECP, the study applied the system to the classification statistics of the platform's utilization rate and potential customers as a means of verifying its performance in the platform and its predictive ability in the future development. The results of low user utilization and potential customer classification are shown in Fig. 8.

From Fig. 8(a), the study categorized the reasons why users do not use a platform into six points, and the system's statistical probability distribution of the reasons for not using the platform was 28.04%, 21.01%, 10.96%, 25.91%,

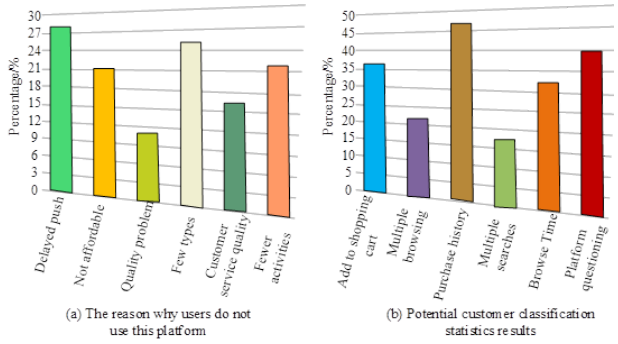

Fig. 8. Low user usage and potential customer classification results.

16.83%, and 22.37%. From Fig. 8(b), the system's statistics on being able to become a potential customer were 36.95%, 22.19%, 47.83%, 18.29%, 33.42%, and 39.92%. From the data, the system is able to differentiate and analyze the low utilization rate caused by different causes in detail, while accurately evaluating the possibility of conversion of various cause groups into potential customers. This shows that the system achieves accurate grasp and prediction of user behavior through advanced data analysis and machine learning mechanisms, providing strong support for increasing user activity and conversion rates. To effectively analyze the effectiveness of the platform's daily use and satisfaction, the study combined the affective precision system to evaluate the feasibility of the ease of use and satisfaction. As shown in Fig. 8, the comparison results of retrieval ease and satisfaction survey are shown.

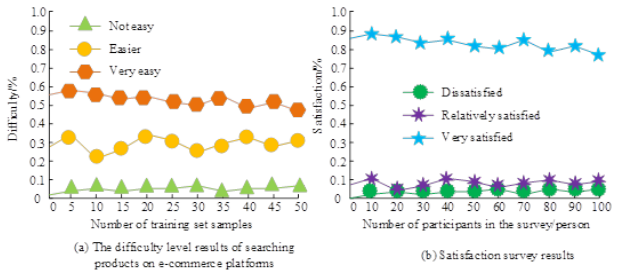

Fig. 9. Comparison of retrieval difficulty and satisfaction survey results.

As shown in Fig. 9(a), the system's analysis of search difficulty revealed that 56.39% of users found the search very easy, 39.57% found it relatively easy, and 3.53% found it difficult to find the desired products. From Fig. 9(b), among the customers who participated in the satisfaction survey, the system judged that 89.08% of the customers were very satisfied with the use of the platform, 9.17% of the customers were more satisfied, and only 1.75% of the customers were dissatisfied with the use of the process. This indicates that the affective precision system is able to effectively categorize and count the information of the existing ECP, providing assistance in formulating appropriate strategies. To further validate the feasibility of the affective precision system, the study took the hot-selling plush hat in winter and the hot-selling sunscreen clothes in summer as the validation commodities, and compares the sales of the commodities after applying the affective precision system with the sales of the commodities without using the system. As shown in Fig. 10, the results of the comparison of the sales of the commodities before and after the use of the affective precision system are shown.

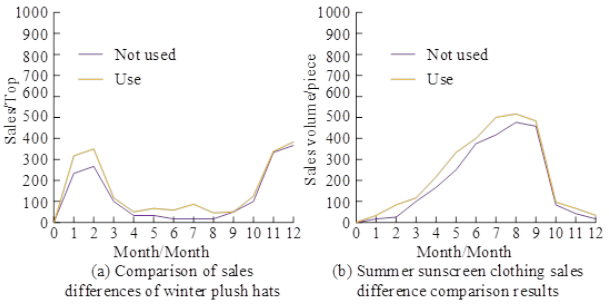

Fig. 10. Comparison results of sales volume of products before and after using affective precision systems.

In Fig. 10(a), there was a difference in the sales of plush hats before and after using the marketing system. Since the plush hat was a best-selling item in winter, its sales in winter were significantly higher than in summer. The average monthly sales volume without using the system was 159, and the average monthly sales volume after using the system was 201. From Fig. 10(b), sunscreen clothing was a best-selling product in summer, and its average monthly sales volume was 239 pieces without using the system and 271 pieces after using the system. The reason and mechanism for the difference in sales of the plush hat and sunscreen clothing in Fig. 10 before and after using the marketing system is mainly the effective use of the affective precision system. The system subdivided the consumers through K-means algorithm combined with RFM model, and deeply understood the consumption habits and demands of different consumers. To compare the performance of various models in practical application, the research implemented the marketing models based on HC K-means, GMDA, TSCEA, KNN, and K-means, which were applied to different industries to explore the sales increase rate of 100 enterprises. The experiment was conducted for 3 years, and the experimental results are shown in Table 1.

From the data in Table 1, the HC K-means model had

Table 1. Comparison of experiment results of marketing system.

Industry	HC K-means	GMDA	TSCEA	KNN	K-means
Fashion & Apparel	25%	18%	20%	15%	17%
Electronics & Gadgets	22%	16%	19%	14%	15%
Home & Garden	20%	15%	17%	13%	14%
Health & Beauty Products	28%	21%	23%	17%	20%

the most outstanding performance in four industries, with higher sales growth rates than other models. Specifically, in the fashion clothing industry, HC K-means achieved a 25% improvement rate, seven percentage points higher than the second ranked GMDA and eight percentage points higher than K-means. In the electronics and accessories industry, HC K-means had a 22% improvement rate, six percentage points higher than GMDA and seven percentage points higher than K-means. In the home horticulture industry, HC K-means had a 20% improvement rate, five percentage points higher than GMDA and six percentage points higher than K-means. In the health and beauty products industry, HC K-means had the highest rate of improvement, at 28%, seven percentage points higher than GMDA and eight percentage points higher than K-means. The GMDA model performed second only to HC K-means in the health and beauty products industry, with a 21% improvement rate, while performing more evenly in other industries. The TSCEA model performed more closely across all industries, with an improvement rate between 17% and 23%. The KNN model had the weakest performance of any industry, rising between 13% and 17%. K-means models performed between KNN and TSCEA, with a rate of improvement between 14% and 20%. The overall comparison results of the experiment are shown in Table 2.

In Table 2, HC K-means algorithm performed better in several key indicators. HC K-means's customer data classification detection rate reached 91.83%, which was significantly higher than GMDA's 80.97% and TSCEA's 85.29%. In terms of error rates, HC K-means led the way with a low error rate of 2.91 percent, compared to 5.86 percent for GMDA and 8.92 percent for TSCEA. In terms of customer retention and complaint rate classification capabilities, HC K-means also showed higher accuracy. In addition, HC K-means also showed good results in the evaluation of platform usability and satisfaction, with a high proportion of customers believing that retrieval is easy and satisfied with the use of the platform.

HC K-means was compared with four benchmark algorithms (GMDA, TSCEA, K-means, and KNN) on the same e-commerce customer dataset (with 100,000 records). Seven indicators were used to evaluate the classification and prediction performance. The results are shown in Table 3.

Table 3 shows that HC K-means outperformed the comparison algorithms in all indicators. Compared with the second-best performing TSCEA, the detection rate of HC K-means increased by 6.54 percentage points, and the error rate decreased by 6.01 percentage points; the classification accuracy of retention rate and complaint rate was 4.24 and 4.66 percentage points higher respectively; the classification accuracy of potential value, current value and loyalty was 4.80, 5.90 and 5.50 percentage points higher respectively, verifying the superiority of this method in precise marketing of e-commerce.

To assess the impact of parameters on the algorithm's performance, Table 4 presents the mean, standard deviation, and the 5th percentile of the detection rate under different parameter settings.

To evaluate the sensitivity of the HC K-means algorithm to key parameters and the stability of clustering, the density radius coefficient (α) and high/low density thresholds were adjusted. Each set of parameters was run 30 times, and the mean detection rate, standard deviation (SD), and the 5th percentile (P5) were recorded. The optimization objective was to maximize the detection rate and minimize the fluctuation. The detection rate was the highest and the fluctuation was the smallest under the benchmark parameter combination, verifying the rationality of the parameter settings. When there was no dynamic radius adjustment or random initialization, the performance significantly declined, indicating that these two improvements are crucial for stability.

3.3. Discussion

By combining K-means algorithm and RFM model, this study aims to accurately classify e-commerce customers and optimize marketing strategies. Research has found that using HC K-means algorithm can significantly improve the generalization ability of customer data and the accuracy of intra-class average distance. In terms of selecting the number of clusters, when the cluster value is set to 6, the data generalization ability reaches 0.94, and the average distance within the class drops to 0.1588, indicating the best classification performance at this time. In addition, the entropy method is used to allocate weights to the three dimensions of the RFM model, further enhancing the ac-

curacy of customer classification. In performance testing, HC K-means algorithm outperforms KNN and traditional K-means algorithm in customer detection rate, error rate, customer value classification, consumption habit classification, as well as customer retention rate and complaint rate classification. Specifically, the customer detection rate of HC K-means is 91.83%, with an error rate of 2.91%, significantly higher than KNN and K-means. In customer value analysis, potential customers have the highest value, reaching 0.93, while the current value of core customers is 0.91. The most loyal general customers and core customers are 0.93 and 0.92, respectively. In terms of product sales, after using a affective precision system, the sales of winter plush hats and summer sunscreen clothing increase from an average of 159 and 239 pieces per month to 201 and 271 pieces, respectively, verifying the feasibility and effectiveness of the system.

The superior performance of HC K-means algorithm in e-commerce affective precision system mainly benefits from its scientific data processing methods and innovative algorithm design. First, HC K-means algorithm combines K-means clustering algorithm and RFM model, and classifies customers into groups with similar consumption characteristics and behavior patterns by multidimensional analysis of customer purchase history, viewing behavior and other data. This classification method not only improves the accuracy of classification, but also provides strong support for enterprises to develop personalized marketing strategies. Secondly, the entropy method is introduced into the clustering process to weight the classification indexes, which ensures the objectivity and rationality of the weights. This weighted processing makes the clustering results more realistic, improving the reliability of classification. In addition, the HC K-means algorithm employs a dynamic radius adjustment mechanism, which adaptively adjusts the neighborhood radius according to the distribution characteristics of the data, thus identifying the cluster centers more accurately. This mechanism allows the algorithm to show greater flexibility and robustness when dealing with complex data. The limitations of the proposed method lie in the fact that the density radius and dynamic threshold still require empirical setting, and the computational cost is relatively high when dealing with large-scale data. The research results show that the HC K-means algorithm can provide a highly stable solution for customer segmentation in fields such as e-commerce and finance, and offers a modeling reference for density-guided initialization in improving unsupervised clustering. From the data such as detection rate (91.83%), error rate (2.91%), and standard deviation (1.24%), it can be directly observed that HC K-

Table 2. Studies the main findings.

Comparison Items	Specific Contents	HC K-means	GMDA	TSCEA
Detection Rate of Customer Data Classification	Detection Rate of E-commerce Customer Data	91.83%	80.97%	85.29%
Error Rate of Customer Data Classification	Error Rate of E-commerce Customer Data	2.91%	5.86%	8.92%
Classification of Customer Retention Rate	Classification Ability of Customer Retention Rate	96.08%	93.26%	91.84%
Classification of Customer Complaint Rate	Classification Ability of Customer Complaint Rate	90.13%	87.24%	85.47%
Evaluation of Platform Ease of Use and Satisfaction - Judgment of Retrieval Difficulty	Proportion of Customers Who Think Retrieval is Very Easy	56.39%	1	1
Evaluation of Platform Ease of Use and Satisfaction - Judgment of Retrieval Difficulty	Proportion of Customers Who Think Retrieval is Relatively Easy	39.57%	1	1
Evaluation of Platform Ease of Use and Satisfaction - Judgment of Retrieval Difficulty	Proportion of Customers Who Think Retrieval is Not Easy	3.53%	/	/
Evaluation of Platform Ease of Use and Satisfaction - Judgment of Satisfaction	Proportion of Very Satisfied Customers	89.08%	/	/
Evaluation of Platform Ease of Use and Satisfaction - Judgment of Satisfaction	Proportion of Dissatisfied Customers	1.75%	/	1

Table 3. Comprehensive performance comparison of the proposed HC K-means algorithm and the benchmark algorithm.

Evaluation indicators	HC K-means	GMDA	TSCEA	K-means	KNN
Customer data classification detection rate (%)	91.83	80.97	85.29	82.14	78.56
Error rate of customer data classification (%)	2.91	5.86	8.92	6.23	9.47
Accuracy rate of customer retention classification (%)	96.08	93.26	91.84	90.15	88.32
Accuracy rate of customer complaint classification (%)	90.13	87.24	85.47	84.63	81.79
Accuracy rate of potential value classification (%)	93	87.5	88.2	85.6	83.1
Current classification accuracy of value categories (%)	91	84.3	85.1	82.4	79.8
Accuracy rate of loyalty classification (%)	93	86.7	87.5	84.9	82.3

Table 4. Comprehensive performance comparison of the proposed HC K-means algorithm and the benchmark algorithm.

Parameter settings	Mean detection rate (%)	SD (%)	P5 (%)
Reference parameters ($\alpha = 1.0$, high-density threshold = 1.2, low-density threshold $= 0.8$)	91.83	1.24	89.52
$\alpha = 0.8$	89.76	2.03	86.31
$\alpha = 1.2$	90.41	1.78	87.65
High-density threshold $= 1.5$, Low-density threshold $= 0.5$	90.15	2.21	86.92
High-density threshold = 1.1, Low-density threshold $= 0.9$	90.83	1.56	88.14
No dynamic radius adjustment	88.94	2.45	85.03
Random initialization	85.62	4.12	79.48

means outperforms all the comparison algorithms. Thus, it can be inferred that the density-guided initial centers and the dynamic radius adjustment mechanism are the core factors for improving the stability and accuracy of the algorithm.

4. Conclusions

This paper proposes an e-commerce affective precision system based on the HC K-means algorithm. The main conclusions are as follows. First, the HC K-means algorithm effectively optimizes the selection of clustering centers by introducing density index and maximum-minimum distance methods, while the dynamic radius adjustment mechanism enhances adaptability to data distribution. Second, experimental results based on 100,000 real transaction records from an e-commerce platform show that the algorithm achieves a customer data classification detection rate of 91.83% and a false detection rate of only 2.91%. Third, in customer value segmentation, the accuracy of potential customer value classification reaches 93%, the current value classification accuracy of core customers is 91%, and the customer loyalty classification accuracy is 93%. These results demonstrate that the proposed system provides reliable and quantifiable decision support for precise marketing in e-commerce. Fourth, the system effectively improves product sales, as verified by the increased monthly sales of seasonal products after system application.

The HC K-means algorithm exhibits strong robustness and stability, benefiting from density-guided initial center selection and dynamic radius adjustment. The weighted Euclidean distance combined with the RFM model and entropy-based weight allocation further improves the objectivity of customer classification.

Limitations of this study include the empirical setting of density radius and dynamic thresholds, as well as the relatively high computational cost when processing large-scale data. Future work will focus on three directions: developing adaptive methods for density radius and dynamic threshold selection to reduce reliance on empirical tuning; introducing distributed computing frameworks to improve computational efficiency for large-scale data; and extending the HC K-means algorithm to cross-platform user behavior analysis to verify its generalization ability in multi-source data fusion scenarios.

References

[1] Y. Wang, R. Zou, F. Liu, L. Zhang, and Q. Liu, (2021) *"A Review of Wind Speed and Wind Power Forecasting with Deep Neural Networks"* **Applied Energy *304***: 117766. DOI: 10.1016/j.apenergy.2021.117766.

[2] A. Sarkar, A. Biswas, and M. Kundu, (2022) *"Development of q-rung orthopair trapezoidal fuzzy Einstein aggregation operators and their application in MCGDM problems"* **Journal of Computational and Cognitive Engineering *1***(3): 109–121. DOI: 10 . 47852 / bonviewJCCE2202162.

[3] J. Zoroja, I. Klopotan, and S. Ana-Marija, (2020) *"Quality of e-commerce practices in European enterprises: Cluster analysis approach"* **Interdisciplinary Descrip-**

tion of Complex Systems: INDECS *18*(2-B): 312–326. DOI: 10.7906/indecs.18.2.17.

[4] A. Mukhayaroh, S. Marlina, S. Hartini, S. Muryani, A. Sinnun, S. Nurajizah, C. Adiwihardja, et al., (2020) *"Implementation of clustering algorithm method for customer segmentation"* **Journal of Computational and Theoretical Nanoscience *17***(2-3): 1388–1395. DOI: 10.1166/jctn.2020.8815.

[5] K. H. Leung, D. Y. Mo, G. T. Ho, C.-H. Wu, and G. Q. Huang, (2020) *"Modelling near-real-time order arrival demand in e-commerce context: a machine learning predictive methodology"* **Industrial Management & Data Systems *120***(6): 1149–1174. DOI: 10.1108/IMDS-12-2019-0646.

[6] S. Sieranoja, (2019) *"How much k-means can be improved by using better initialization and repeats?"* **Pattern Recogn *93***(04): DOI: 10.1016/j.patcog.2019.04.014.

[7] Y. Huang. "Research on e-commerce precision marketing strategy based on big data technology". In: *2021 2nd International Conference on E-Commerce and Internet Technology (ECIT)*. IEEE. 2021, 87–90. DOI: 10.1109/ECIT52743.2021.00026.

[8] W. Jun, S. Li, Y. Yanzhou, E. D. S. Gonzalezc, H. Weiyi, S. Litao, and Y. Zhang, (2021) *"Evaluation of precision marketing effectiveness of community e-commerce–An AISAS based model"* **Sustainable Operations and Computers *2***: 200–205. DOI: 10.1016/j.susoc.2021.07.007.

[9] S. Wang and Y. Yang, (2021) *"M-GAN-XGBOOST model for sales prediction and precision marketing strategy making of each product in online stores"* **Data Technologies and Applications *55***(5): 749–770. DOI: 10.1108/DTA-11-2020-0286.

[10] X. Yang, P. Gong, et al., (2020) *"Marketing strategy of green agricultural products based on consumption intention"* **Agricultural & Forestry Economics and Management *3***: 16–24. DOI: 10.23977/agrfem.2020.030103.

[11] R. P. Chatterjee, K. Deb, S. Banerjee, A. Das, and R. Bag, (2019) *"Web mining using k-means clustering and latest substring association rule for e-commerce"* **Journal of Mechanics of Continua and Mathematical Sciences *14***(6): 28–44. DOI: 10.26782/jmcms.2019.12.00003.

[12] H. Liu, J. Chen, J. Dy, and Y. Fu, (2023) *"Transforming complex problems into K-means solutions"* **IEEE transactions on pattern analysis and machine intelligence *45***(7): 9149–9168. DOI: 10.1109/TPAMI.2023.3237667.

[13] J. Niu, (2021) *"Intelligent evaluation model of e-commerce transaction volume based on the combination of k-means and SOM algorithms"* **International Journal of Information and Communication Technology *18***(2): 189–206. DOI: 10.1504/ijict.2021.113043.

[14] P. Wang, B. Cao, T. Ma, B. Wang, Q. Zhang, and P. Zheng, (2023) *"DUHI: Dynamically updated hash index clustering method for DNA storage"* **Computers in Biology and Medicine *164***: 107244. DOI: 10.1016/j.compbiomed.2023.107244.

[15] D. K. Sharma, S. Lohana, S. Arora, A. Dixit, M. Tiwari, and T. Tiwari, (2022) *"E-Commerce product comparison portal for classification of customer data based on data mining"* **Materials Today: Proceedings *51***: 166–171. DOI: 10.1016/j.matpr.2021.05.068.

[16] H. P. Patra and K. Nikitha, (2022) *"Analyse k-means algorithm and implementing a new clustering algorithm"* **International Journal of Computational Systems Engineering *6***(4): 178–181. DOI: 10.1504/IJCSYSE.2021.120288.

[17] S. S. Prasetyo, M. Mustafid, and A. R. Hakim, (2020) *"Penerapan fuzzy c-means kluster untuk segmentasi pelanggan e-commerce dengan metode recency frequency monetary (rfm)"* **Jurnal Gaussian *9***(4): 421–433. DOI: 10.14710/j.gauss.v9i4.29445.

[18] S. N. Sharma and P. Sadagopan, (2022) *"Influence of conditional holoentropy-based feature selection on automatic recommendation system in E-commerce sector"* **Journal of King Saud University-Computer and Information Sciences *34***(8): 5564–5577. DOI: 10.1016/j.jksuci.2020.12.022.

[19] M. Rezaei and P. Fränti, (2023) *"K-sets and k-swaps algorithms for clustering sets"* **Pattern Recognition *139***: 109454. DOI: 10.1016/j.patcog.2023.109454.

[20] P. Fränti and S. Sieranoja, (2023) *"K-means properties on six clustering benchmark datasets: P. Fränti, S. Sieranoja"* **Applied intelligence *48***(12): 4743–4759. DOI: 10.1007/s10489-018-1238-7.

[21] W. Haibin, G. Xin, Y. Xiao, L. Shuangming, and S. Guidong, (2024) *"A GMDA clustering algorithm based on evidential reasoning architecture"* **Chinese Journal of Aeronautics *37***(1): 300–311. DOI: 10.1016/j.cja.2023.09.015.

[22] C. Liu and S. Yang, (2024) *"A two-stage clustering ensemble algorithm applicable to risk assessment of railway signaling faults"* **Expert Systems with Applications *249***: 123500. DOI: 10.1016/j.eswa.2024.123500.

[23] P. Fränti, (2018) *"Efficiency of random swap clustering"* **Journal of big data *5***(1): 1–29. DOI: 10.1186/s40537-018-0122-y.

[24] A. Rodriguez and A. Laio, (2014) *"Clustering by fast search and find of density peaks"* **science *344***(6191): 1492–1496. DOI: 10.1126/science.1242072.

[25] S. Sieranoja and P. Fränti, (2010) *"Fast and general density peaks clustering"* **Pattern recognition letters *128***: 551–558. DOI: 10.1016/j.patrec.2019.10.019.