

## Research Article

# User Value Identification Based on Improved RFM Model and $K$ -Means++ Algorithm for Complex Data Analysis

Jun Wu,<sup>1,2</sup> Li Shi,<sup>1</sup> Liping Yang,<sup>2</sup> Xiaxia Niu ,<sup>2</sup> Yuanyuan Li,<sup>2</sup> Xiaodong Cui,<sup>3</sup> Sang-Bing Tsai ,<sup>4</sup> and Yunbo Zhang<sup>5</sup>

<sup>1</sup>College of Information Science and Technology, Beijing University of Chemical Technology, Beijing 100029, China

<sup>2</sup>School of Economic and Management, Beijing University of Chemical Technology, Beijing 100029, China

<sup>3</sup>Industrial and Commercial Bank of China Limited, Beijing 100088, China

<sup>4</sup>Regional Green Economy Development Research Center, School of Business, Wuyi University, Wuyishan 354300, China

<sup>5</sup>Datang Carera (Beijing) Investment Co. Ltd., Beijing 100191, China

Correspondence should be addressed to Xiaxia Niu; 15136212624@163.com and Sang-Bing Tsai; sangbing@hotmail.com

Received 8 March 2021; Revised 6 April 2021; Accepted 16 April 2021; Published 3 May 2021

Academic Editor: Yuanpeng Zhang

Copyright © 2021 Jun Wu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In recent years, with the development of machine learning and big data technology, user data has become an important element in the production process of enterprises. For today's e-commerce platforms, the deep mining of user's purchase behavior is helpful to understand user's purchase preferences and accurately recommend products that meet user expectations, which can not only improve user satisfaction but also reduce platform marketing cost. To accurately identify the user value of online purchasing on an e-commerce platform, this paper uses an improved RFM model to extract user features and uses the  $K$ -means++ clustering algorithm to realize user classification. The indicators of the traditional RFM model characterize user features from three angles: recent purchase time ( $R$ ), purchase frequency ( $F$ ), and total consumption amount ( $M$ ). The user group and scenarios studied in this paper are different from the previous literature: (1) the user group is relatively fixed, (2) the consumer goods are relatively single, and (3) the characteristics of repeated purchase are obvious. Therefore, based on the existing literature, this paper extracts the user characteristics studied and improves and models the traditional indicators. Based on the real purchasing data from September to December 2018, it calculates the indicators that improved RFM, empowers the weight to indicators, and finally classifies the value of users by using the  $K$ -means++ algorithm. The experimental results show that the user classification based on the improved RFM model is more accurate than the user classification based on the traditional RFM model, and the improved RFM model can identify the user value more accurately, which provides a strong support for the e-commerce platform to realize the accurate marketing strategy based on big data.

## 1. Introduction

Big data technology is the product of the development of information processing technology. Today is an era of data explosion; the enterprise operation process will produce massive data. Today's big data technology is widely used in various scenarios of enterprise operations, internet of things (IOT) environment [1–6], etc., providing decision support for various decisions of enterprises. Compared with the traditional marketing methods such as questionnaire surveys, the purchase behavior of users often reflects the user's psychological preferences. In today's business competition, cus-

tomers are the main focus of the company to maintain excellent performance. It is found that the cost of acquiring new customers is much more expensive than retaining existing customers. Thus, what companies care most is how to sell more products to existing customers. Nowadays, the number of e-commerce platforms is increasing rapidly, the operating cost is getting higher and higher, and the marketing expenditure is also expanding accordingly. Using the purchase records of users on the platform to understand the decisions made by users in the real environment has become an urgent problem to be solved in the efficient operation of enterprises. At the same time, as a new business model, a community

e-commerce platform has less research on the classification of community e-commerce customers. Based on the above background, this paper provides an effective method for the e-commerce platform to identify customer value and realize a precision marketing strategy.

The research object of this paper is the T-app community e-commerce platform customers; this platform mainly sells cooked food, pasta, and other goods, a total of 134 kinds. The platform focuses on the community traffic, aiming at the needs of community residents, and its customers' purchase behavior has the characteristics of a low single consumption amount and high consumption frequency. The RFM model has good representativeness in reflecting customer value and customer purchase preference and is widely used in financial industry [7], retail [8], insurance and telecommunication [9], education industry [10], and e-commerce industry [11]. The RFM model can well construct the outline of the research object in this paper.

However, in the original RFM model, there is a large randomness in the  $R$  index of the latest consumption time. New customers who have just come into contact with the platform and loyal customers of the platform may have the same performance in the  $R$  index of the latest consumption time, and the model cannot describe the dependence of customers on a single commodity. In order to help the platform describe the multidimensional attributes of customers and more accurate customer value, an improved RFM model with five indicators is proposed.

Therefore, based on the improved RFM model and  $K$ -means++ algorithm, this paper proposes a method suitable for T-app community e-commerce platform user classification. According to the real purchase records of users on the T-app community e-commerce platform, the improved RFM model is used to analyze the data of user purchase records. The  $K$ -means++ clustering algorithm is used to classify and calculate the users of the e-commerce platform. Finally, the classification results are explained, and the corresponding customer value analysis is given combined with the specific indicators of each subdivided user. This study uses a quantitative analysis method to segment and cluster platform customers. Customer segmentation with clear value and purchase preference is helpful for the platform to effectively allocate marketing resources for specific customer groups and to establish a healthy long-term relationship with customers.

## 2. Relevant Works

**2.1. RFM Model.** The RFM model was first proposed by Hughes [12]. As a popular tool of customer value analysis, it has been widely used for measuring customer lifetime value [13] and in customer segmentation and behavior analysis [14]. In the following paragraphs, we provide a brief description of the RFM model in the above literature.

RFM is short for recency, frequency, and monetary, which refer to recency of the last purchase, purchase frequency, and monetary value of purchase, respectively.  $R$  (recency) represents the time interval between a customer's last purchase date and the end date of a statistical period. The shorter the interval, the bigger the value of  $R$ .  $F$

(frequency) indicates the number of purchases made by the customer during the statistical period. The larger the value of  $F$  is, the higher the customer loyalty and the stronger the intention to repurchase would be.  $M$  (monetary) represents the total amount the customer spends in purchases during the statistical period. Hughes attached equal importance to these three variables [12], while Stone believed that the importance of the three variables varies among industries due to their different characteristics, suggesting unequal weights of these variables [13]. The RFM model is widely used in customer value analysis, and researchers have extended it according to different aspects.

Cheng and Chen combined RFM analysis with a rough set theory to establish rules for customer classification [13]. Chiang proposed an RFMDR model (based on an RFM/RFMD model), an extended version of RFM analysis, to identify valuable online shopping customers for the industry and to generate fuzzy association rules [15]. Kolarovszki et al. have proposed a novel modeling method for postal services using multidimensional segmentation. This CRM design proves useful in postal service companies [16]. Song et al. proposed a statistic-based approach to evaluate potential users via time series. With this approach, it is possible to segment time intervals of RFM in a large-scale dataset [17]. Aiming at the randomness of recent consumption time  $R$  index in the RFM model, and the colinearity between index  $F$  and index  $M$ , Bao et al. proposed an improved RFM model;  $K$ -means is used to cluster the customer data and proved the effectiveness of the improved RFM model [18]. In order to help the platform describe the multidimensional attributes of customers and more accurate customer value, an improved RFM model with five indicators is proposed. In this study, we use the RFM model as the basis to select the variables for clustering, so as to establish the clustering criteria objectively.

**2.2.  $K$ -Means++ Algorithm.** Clustering is the process of dividing a set of physical or abstract objects into groups of similar objects. The  $K$ -means algorithm, as one of the most popular clustering algorithms, was first used by MacQueen in 1967 [19], and it has been used extensively in various fields including data mining, statistical data analysis, and other business applications.

The literature survey reveals that one of the major applications of  $K$ -means is customer segmentation [20]. The  $K$ -means algorithm is widely used to effectively identify valuable customers and develop pertinent marketing strategies [21]. In particular, Cheng and Chen used the RFM model and  $K$ -means to perform customer relationship management, and experimental results demonstrate that the model they proposed is an effective method in customer value analysis [13].

$K$ -Means is a fast method in clustering analysis, but the accuracy and running time of clustering results largely depend on the location of the initial clustering center [22–24]. In order to solve the problem that  $K$ -means is sensitive to initial points, Arthur et al. proposed the  $K$ -means++ algorithm to improve the method of  $K$ -means randomly selecting initial clustering centers, that is, to make the

distance between the clustering centers as far as possible when selecting initial clustering centers. The results show that  $K$ -means++ can significantly improve the final error of classification results [25]. Because the  $K$ -means++ algorithm is accurate and efficient, this paper uses the  $K$ -means++ algorithm to classify customers.

### 3. Improved RFM Model and $K$ -Means++ Clustering Algorithm

In order to accurately identify the user value of the T-app community e-commerce platform, this study uses an improved RFM model to extract a user's features and uses the  $K$ -means++ clustering algorithm to achieve user classification. The indicators of the traditional RFM model portray customer characteristics from three perspectives: recency of the last purchase ( $R$ ), frequency of the purchases ( $F$ ), and monetary value of the purchases ( $M$ ). However, the user groups and scenarios studied in this article are quite different from the previous literature: (1) the user group is relatively fixed, (2) the consumer goods are relatively singular, and (3) the characteristics of repeated purchases are obvious. Therefore, based on the existing literature, we extract the characteristics of the users studied and improve and model the traditional indicators, as follows.

#### 3.1. Indicator Definition

- (1)  $R_1$ : average consumption time interval, refers to the average transaction time interval of a certain user in a certain period
- (2)  $F_1$ : frequency of the purchases, refers to the number of order transactions for a certain user within a certain period
- (3)  $M_1$ : monetary value of the purchases, refers to the total amount spent by a user in a certain period
- (4)  $S$ : customer contribution time, refers to the time interval from the first transaction to the last transaction of a certain user
- (5)  $P$ : repeat purchase attributes, refers to the number of purchases of a category of goods purchased by a certain user in the reference time

3.2. *Indicator Modeling.* The calculation formula of the  $R_1$  indicator is:

$$R_1 = \frac{T_{\text{last\_time}} - T_{\text{first\_time}}}{F_1}, \quad (1)$$

where  $T_{\text{last\_time}}$  indicates the time of the customer's last order transaction within the reference time period and  $T_{\text{first\_time}}$  indicates the time of the customer's first order transaction within the reference time period.

The calculation formula of the  $M_1$  indicator is

$$M_1 = \sum_i^n M_i, \quad (2)$$

where  $n$  represents the total number of a customer's consumption in the reference time period and  $M_i$  represents the amount of single consumption by the customer.

The calculation formula of the  $S$  indicator is

$$S = T_{\text{last\_time}} - T_{\text{first\_time}}. \quad (3)$$

In (3),  $T_{\text{first\_time}}$  is the time of the first transaction in the customer's purchase history.

In the improved RFM model, we adopted average consumption time interval ( $R_1$ ) to replace recency of the last purchase ( $R$ ) in the RFM model, thus overcoming the shortcomings of the large randomness of the  $R$  indicator in the traditional RFM model; for regular customers with high transaction frequency, the average order transaction time is more representative. The customer contribution time indicator ( $S$ ) reflects the customer's loyalty to the platform and continuous consumption ability, that is, the time interval from the first transaction in the customer's history to the last transaction at the reference time, while the repurchase indicator ( $P$ ) can describe the degree of customer dependence on a single product.

#### 3.3. Determination of Indicator Weights

3.3.1. *Step1-1: Standardization of Indicator Data.* Due to the different natures of the indicators in this model, the value of the average transaction interval  $R_1$  indicator is a negative indicator of customer value; the smaller the value of this indicator, the greater the customer value. On the contrary, the remaining four indicators are all positive indicators, that is, the greater the value of the customer under this indicator, the greater the customer value. Moreover, the five indicators of this model have inconsistent units and differ largely in value ranges. In order to achieve the purpose of normalization and standardization of heterogeneous indicators, the  $R_1$  indicator is specially standardized by the reverse standardization method, namely, formula (4), and the other indexes are standardized by the forward standardization method, namely, formula (5).

$$x'_{ij} = \frac{\max \{x_{ij}, \dots, x_{nj}\}}{\max \{x_{ij}, \dots, x_{nj}\} - \min \{x_{ij}, \dots, x_{nj}\}}, \quad (4)$$

$$x'_{ij} = \frac{x_{ij} - \min \{x_{ij}, \dots, x_{nj}\}}{\max \{x_{ij}, \dots, x_{nj}\} - \min \{x_{ij}, \dots, x_{nj}\}}. \quad (5)$$

In the above formula,  $x_{ij}$  represents the value of the  $j$ -th index of the  $i$ -th sample.

3.3.2. *Step1-2: Index Weight Calculation Using Entropy Method.* In order to obtain more objective segmentation results, in this study, we use the entropy weight method in

the objective weighting method to calculate the weights of the five indicators in the improved model.

### 3.4. Clustering Based on K-Means++ Algorithm

**3.4.1. Step2-1: Find the Optimal Number of K Clusters Using the Contour Coefficient Method.** Contour coefficient is an evaluation method of clustering effect, which was first proposed by Rousseeuw in 1986 [26]. It considers the agglomeration and separation of clusters comprehensively. Good clustering results should have both a smaller cohesion and a larger separation degree between clusters. In the research of this article, we use the maximum average contour coefficient method to determine the optimal cluster number  $K$ . The calculation process is as follows.

Suppose the data to be classified is divided into  $K$  clusters. For each vector in the cluster, calculate their contour coefficients separately. For one of the points  $i$ , the contour coefficient of the  $i$  vector is

$$S(i) = \frac{b(i) - a(i)}{\max \{a(i), b(i)\}}. \quad (6)$$

In formula (6),  $a(i)$  = average (the distance from the  $i$  vector to all other points in the cluster to which it belongs), expressed as the average value of the dissimilar degree from the  $i$  vector to other points in the same cluster;  $b(i)$  = min (the average distance from the  $i$  vector to all points in the nearest cluster), expressed as the minimum value of the average dissimilar degree from the  $i$  vector to other clusters. Average the contour coefficients of all vectors, and the result is the total contour coefficient of the clustering result. The larger the total contour coefficient value, the more ideal the clustering result.

The calculation result of the total profile coefficient is shown in Figure 1.

When  $K$  takes 4, the total contour coefficient of the cluster is 0.4022, which is the largest in the range of (3, 9). It can be concluded that the optimal cluster number  $K_{opt}$  is 4 in this data set.

**3.4.2. Step 2-2: Clustering.** The clustering algorithm process is shown in Figure 2.

In formula (a),  $D(x')^2 / \sum_{x \in X} D(x)^2$ , of Figure 2, value  $D(x)^2$  represents the distance from the data point  $x$  to the nearest cluster center (select  $S_i$  ( $S_i = x' \in X$ )). For formula (b) of Figure 2, the average error is equal to  $\sum_{i=1}^n [\min_{r=1, \dots, k} d(x_i, c_r)^2]$ .

## 4. Case Study

**4.1. Numerical Experiment.** The observation period is from September 1, 2018, to December 30, 2018, for three months. There are 3558 customer purchase records and 580 customers. The community shopping platform mainly sells 134 kinds of commodities such as cooked food and pasta. The data is processed as follows.

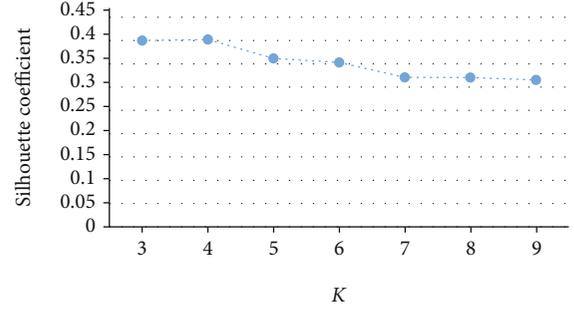


FIGURE 1: Comparison of contour coefficients when  $K$  is (3, 9).

**4.1.1. Raw Data Cleaning and Index Calculation.** The initial data consists of 12 dimensions such as user ID, product ID, purchase quantity, and consumption date. Four dimensions of user ID, consumption amount, consumption time, and product ID are selected from them, and the corresponding five indicators of each customer are calculated to form the initial data, as in Table 1.

In Table 2, the customer whose user ID is YH181102000001 has spent 52 days on the platform. A total of 5 purchases occurred during the reference time period, with a cumulative consumption of 64 yuan, and an average purchase was made on the platform every 10.40 days. The purchase of goods was relatively casual, and no repurchase of goods occurred (the value of  $P$  indicator is 1).

The  $R_1$  indicator adopts the reverse standardization method, namely, formula (4), and the other indicators adopt the forward standardization method, namely, formula (5) for standardization.

**4.1.2. Empower Indicators.** The entropy method is used to get the weight of each index. The weight of each index is shown in Table 2. At the same time, the weight of each index is multiplied by its corresponding weight to get the weighted data set.  $R_1'$ ,  $F_1'$ ,  $M_1'$ ,  $S'$ , and  $P'$  are the weighted indexes.

**4.2. Clustering Using K-Means++ Method.** After using the  $K$ -means algorithm to cluster the weighted data set, we get four different customer groups.

Exporting the standard data of four types of customer groups and calculating the average value of five indicators for each type of customer along with the number of users of each type of user, we get Table 3. The data of each user group in Table 3 is the average of the standardized data which has not been weighted.

**4.3. Customer Value Ranking and Value Analysis.** Since the user data after clustering at this time is weighted data, the total value of users of this type can be obtained by adding up the values of the indicators of each user category, calculated as follows:

$$\text{Value} = S' + F_1' + M_1' + R_1' + P'. \quad (7)$$

The calculated results are shown in Table 4.

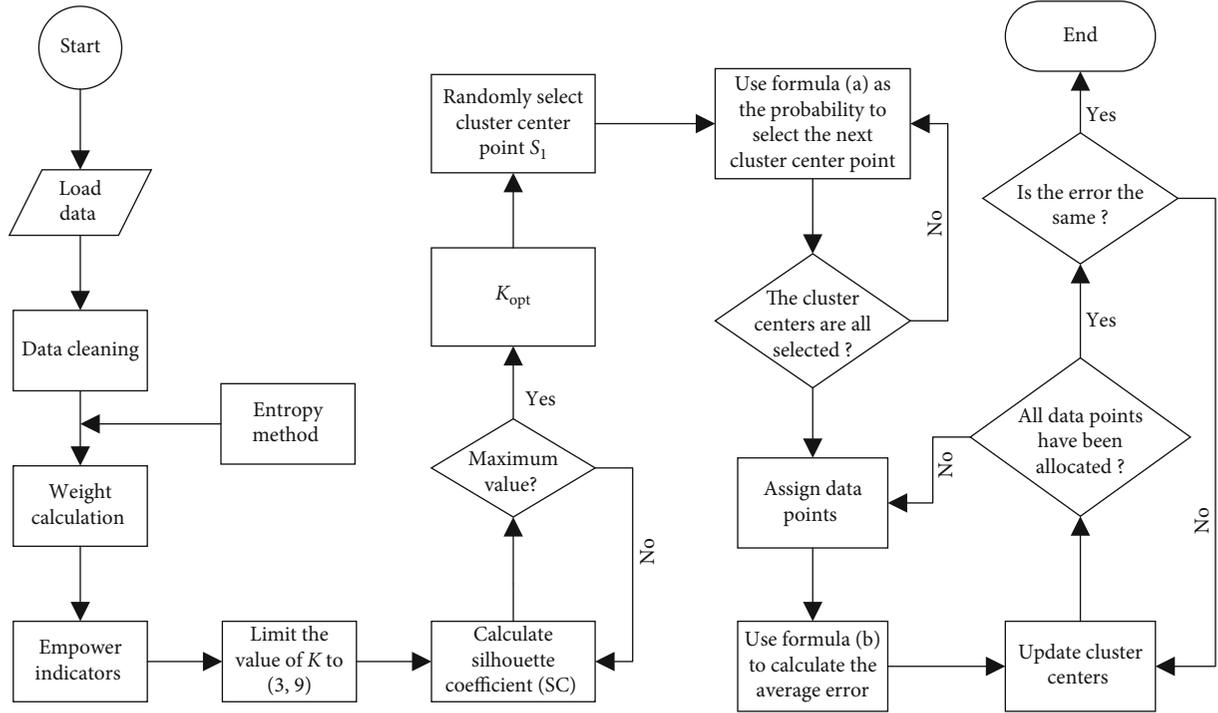


FIGURE 2: Algorithm flow chart.

TABLE 1: Part of customer data.

User ID	S (day)	$F_1$ (frequency)	$M_1$ (yuan)	$R_1$ (day)	P (frequency)
YH181102000001	52.0	5	64.0	10.40	1
YH170621000001	532.0	11	541.0	8.18	6

TABLE 2: Index weight.

Index	$R_1$	$F_1$	$M_1$	S	P
Weight	0.0044	0.2522	0.2752	0.1776	0.2903

TABLE 3: Customer clustering results.

Cluster	$R_1'$	$F_1'$	$M_1'$	$S'$	$P'$	Count
Cluster 3	0.00406	0.16857	0.09913	0.07957	0.1246	39
Cluster 2	0.00396	0.06675	0.04121	0.04815	0.0607	127
Cluster 4	0.00393	0.02145	0.01489	0.10469	0.0142	124
Cluster 1	0.00413	0.01295	0.00848	0.01855	0.0127	290

It can be seen from the model that the higher the amount of value, the greater the profit contribution of this type of customer to the platform. Therefore, according to the value ranking result and the different characteristics of the customer, we divided customers into important retention customers, development customers, loyal customers, and general customers. From the data, the characteristics of each type of customer are as follows.

4.3.1. *Type 1: Important Retention Customers.* Customer group 1 has the greatest value to the platform and is an important maintaining customer of the platform. Its S indicator is large among all customer groups, indicating that this customer group has maintained a long-term consumer relationship with the platform; the  $M_1$  indicator is much higher than other customer groups, indicating that this customer group has made a lot of consumption on the platform and is an important source of profit. At the same time, the value of the P indicator is also much higher than other customer groups, indicating that this customer group is more likely to repurchase a single product. The  $F_1$  indicator and the reverse-standardized  $R_1$  indicator indicate that this type of customer has greater consumer stickiness on the platform and purchases more frequently. As a result, the user characteristics of the platform's high-value customers can be characterized as follows: high consumption frequency, small consumption time interval, high degree of product specificity, and have their own fixed purchase product.

4.3.2. *Type 2: Developmental Customers.* Customer group 3 has a considerable performance in total consumption amount and consumption frequency. However, by observing its consumption time relationship with the platform,

TABLE 4: Customer value ranking.

Cluster	$R_1'$	$F_1'$	$M_1'$	$S'$	$P'$	Count	Quantity of value	Value ranking
Cluster 1	0.00406	0.16857	0.09913	0.07957	0.1246	39	0.475905	1
Cluster 3	0.00396	0.06675	0.04121	0.04815	0.0607	127	0.220726	2
Cluster 4	0.00393	0.02145	0.01489	0.10469	0.0142	124	0.159173	3
Cluster 2	0.00413	0.01295	0.00848	0.01855	0.0127	290	0.056788	4

it can be concluded that the users' time to contact the platform is average, and there is still room for tapping the value of its consumption. The platform should devote itself to transforming its development into important retention customers.

*4.3.3. Type 3: Loyal Customers.* Customer group 3 maintains the longest consumption relationship with the platform among all customers, indicating that this type of customer contacted the platform earlier and maintained a long-term consumption record. Although the consumption amount is not large, they give the platform's cash flow a greater guarantee, being loyal customers of the platform.

*4.3.4. Type 4: General Customers.* Customer group 2 does not have a long-term consumer relationship with the platform. The consumption frequency and the total amount of consumption are very low and so also is the repurchase of a single product, indicating that they had not made considerable consumption on the platform and had not contributed to the profit growth of the platform; thus, they belong to the general customers of the platform.

*4.4. Experimental Result Verification.* After analyzing the value of the customer, we return to the original data of the customer order transaction to verify the T-app customer value analysis result of the improved RFM model. The classification process of the traditional RFM model is similar with the model used in this article, that is, data standardization-calculation of weight-weighting-clustering-value ranking. Among them,  $F$  and  $M$  indicators use forward standardization and  $R$  indicators use reverse standardization. Later, according to the value ranking, the customers are also divided into important retention customers, loyal customers, development customers, and general customers. Now compare the classification results of the original RFM model with the improved RFM model classification results, as shown in Tables 5 and 6.

The analysis shows that under the RFM model, 34 important retention customers have been identified, all indicators of which are better than the rest of the customers, but no purchase behavior has been made recently. The consumer group 3 on the platform has a higher frequency of consumption and a larger amount of consumption, indicating that it has greater consumer stickiness and is a loyal customer of the platform. All indicators of customer group 4 consumption on the platform need to be improved, and they belong to developmental customers. Customer group 2 consumption amount and consumption frequency are both low, but the recent purchase

TABLE 5: The weight of each indicator of the RFM model.

Index	$R$	$F$	$M$
Weight	0.1601	0.4016	0.4382

time indicator has performed well, indicating that it may be due to the customers attracted by the platform's recent marketing strategy, which means that they did not bring considerable profits and cash flow to the platform. So this group belongs to general customers.

The comparison shows that the original RFM model cannot identify the new and old attributes of customers and can only determine the customer's loyalty index based on the customer's consumption frequency and consumption amount, but this has a certain contingency. Customers who make high-frequency consumption in the short-term may terminate their purchases after the platform's promotional activities are over. Such customers should not be considered as having high long-term value. At the same time, improving the RFM model can intuitively feel the differences in various indicators of different user groups. For example, users who spend high on the platform have high repurchase of goods, which can guide the platform to carry out differentiated marketing and improve the customers' experience and ultimately bring platform profit growth.

Now take the customer value indicators and customer classification results of some customers through the RFM model and the improved RFM model as an example to illustrate the credibility of the improved RFM model, as shown in Tables 7 and 8.

In Tables 7 and 8, we can see that the user with the ID of YH171201000030 has the same classification results under the two models. Analysis of the RFM model shows that the user of YH171201000030 made a total of 529 yuan worth of purchases during the observation period, with a frequency of 24 purchases, and the recent purchase time  $R$  index is relatively close, indicating that the customer has a greater probability of repurchasing next. But at the same time, it can also be found that the RFM model cannot accurately identify whether the user has a greater degree of dependence on a single product and the average time interval for placing orders nor can it define loyalty attributes based on the relationship time between the customer and the platform. For this type of important retention customers, improving the RFM model can provide a more detailed description based on T-app user data, which facilitates the platform to conduct one-to-one precision marketing, which is conducive to the retention of such customers.

TABLE 6: RFM model customer segmentation value analysis.

Cluster	$R'$	$F'$	$M'$	Count	Value ranking
Cluster 1	0.1474	0.1622	0.2849	34	1
Cluster 3	0.1135	0.0792	0.1202	106	2
Cluster 4	0.1165	0.0192	0.0324	188	3
Cluster 2	0.0293	0.0152	0.0225	252	4

TABLE 7: RFM model user classification results (partial).

User ID	$R$	$F$	$M$	Customer segmentation results
YH171201000030	2.0	24	529.0	Important retention customers
YH181207000001	19.0	2	14.0	Development client
YH181018000002	72.0	8	138.5	Loyal customers

TABLE 8: User classification results of the improved RFM model (partial).

User ID	$R_1$	$F_1$	$M_1$	$S$	$P$	Customer segmentation results
YH171201000030	4.75	24	529.0	380.0	8	Important retention customers
YH181207000001	1.5	2	14.0	3.0	1	General customer
YH181018000002	0	8	138.5	0	5	General customer

For the user with the ID “YH181207000001” under the RFM model, we only know that he had made a total amount of 14 yuan and a frequency of 2, but because his latest purchase happened to be closer to the end of the latest purchase period, he was classified as a development customer. While under the improved RFM model, it can be known that the user has only been exposed to the platform for 3 days and has only made two low-value consumptions on the platform, and his potential commercial value needs to be further developed, so he is correctly divided into general customers. Therefore, the improved RFM model can overcome the randomness of the  $R$  indicator in the original RFM model and can accurately locate such new customers.

The user with the ID “YH181018000002” made 8 purchases with a total amount of 138.5 yuan on the platform, and the last purchase was 72 days away from the end of the reference period. Because of his relatively considerable consumption data in the overall sample, he was classified as a loyal customer in the original RFM model. However, through the analysis of the improved RFM model, it can be known that the user has considerable data on the total amount of consumption and consumption frequency. But he was concentrated on a certain product 5 times in a day and had not contacted the platform, indicating that he may be attracted by the platform’s recent promotion activities, so the improved RFM model has correctly divide him into general customers, waiting for further observation.

The experimental results given above can show that the improved RFM model can provide a more accurate user description than the original RFM model under the same classification. Moreover, it overcomes the randomness of the  $R$  indicator in the original RFM model and perfected

the RFM model’s shortcomings of only describing the customer stickiness and loyalty through observation of consumption amount and frequency. The improved model can better complete the user value evaluation of T-app.

## 5. Conclusions

In order to accurately analyze the customer value of a T-app community e-commerce platform, a customer value analysis method based on an improved RFM model is proposed. The average order transaction time interval, customer transaction times in a certain period, customer total consumption amount in a certain period, customer relationship duration, five repurchase indicators, use of forward and reverse standardization methods to standardize the indicators, and use of the objective assignment method of information theory entropy method to calculate the weight of the five indicators are included. The concept of the silhouette coefficient is used to determine the best  $K$  value, and the  $K$ -means++ clustering algorithm is used to cluster the weighted indicators; finally, customers are divided into different value customer groups. This paper takes the T-app order transaction data from September 1, 2018, to December 30, 2018, to analyze the customer value, and the results show that the improved RFM model can more accurately analyze the customer value of T-app.

## Data Availability

All information is within the article.

## Conflicts of Interest

The authors have declared that no competing interests exist concerning this study.

## Acknowledgments

This work was supported, in part, by the Funds for the Beijing University of Chemical Technology First-Class Discipline Construction (XK1802-5) and Beijing University of Chemical Technology (GJD202002).

## References

- [1] A. M. Al-Momani, M. A. Mahmoud, and M. S. Ahmad, "Factors that influence the acceptance of internet of things services by customers of telecommunication companies in Jordan," *Journal of Organizational and End User Computing*, vol. 30, no. 4, pp. 51–63, 2018.
- [2] I. Kitouni, D. Benmerzoug, and F. Lezzar, "Smart agricultural enterprise system based on integration of internet of things and agent technology," *Journal of Organizational and End User Computing*, vol. 30, no. 4, pp. 64–82, 2018.
- [3] K. G. Srinivasa, B. J. Sowmya, A. Shikhar, R. Utkarsha, and A. Singh, "Data analytics assisted internet of things towards building intelligent healthcare monitoring systems," *Journal of Organizational and End User Computing*, vol. 30, no. 4, pp. 83–103, 2018.
- [4] S. K. Biswas, D. Devi, and M. Chakraborty, "A hybrid case based reasoning model for classification in internet of things (IoT) environment," *Journal of Organizational and End User Computing*, vol. 30, no. 4, pp. 104–122, 2018.
- [5] M. R. Reddy, K. G. Srinivasa, and B. E. Reddy, "Smart vehicular system based on the internet of things," *Journal of Organizational and End User Computing*, vol. 30, no. 3, pp. 45–62, 2018.
- [6] R. Parada, J. Melià-Seguí, and R. Pous, "Anomaly detection using RFID-based information management in an IoT context," *Journal of Organizational and End User Computing*, vol. 30, no. 3, pp. 1–23, 2018.
- [7] R. Heldt, C. S. Silveira, and F. B. Luce, "Predicting customer value per product: from RFM to RFM/P," *Journal of Business Research*, vol. 3, 2019.
- [8] A. Griva, C. Bardaki, K. Pramatar, and D. Papakiriakopoulos, "Retail business analytics: customer visit segmentation using market basket data," *Expert Systems with Applications*, vol. 100, pp. 1–16, 2018.
- [9] K. Khalili-Damghani, F. Abdi, and S. Abolmakarem, "Hybrid soft computing approach based on clustering, rule mining, and decision tree analysis for customer segmentation problem: real case of customer-centric industries," *Applied Soft Computing*, vol. 73, pp. 816–828, 2018.
- [10] H.-C. Chang, "Developing EL-RFM model for quantification learner's learning behavior in distance learning," in *2010 2nd International Conference on Education Technology and Computer*, Shanghai, China, 2010.
- [11] Y. S. Cho, S. C. Moon, S. C. Noh, and K. H. Ryu, "Implementation of personalized recommendation system using  $k$ -means clustering of item category based on RFM," in *In 2012 IEEE International Conference on Management of Innovation & Technology*, pp. 378–383, Bali, Indonesia, 2012.
- [12] A. M. Hughes, *Strategic Database Marketing*, Probus Publishing Company, Chicago, USA, 1994.
- [13] C. H. Cheng and Y.-S. Chen, "Classifying the segmentation of customer value via RFM model and RS theory," *Expert Systems with Applications*, vol. 36, no. 3, pp. 4176–4184, 2009.
- [14] D. Chen, S. L. Sain, and K. Guo, "Data mining for the online retail industry: a case study of RFM model-based customer segmentation using data mining," *Journal of Database Marketing & Customer Strategy Management*, vol. 19, no. 3, pp. 197–208, 2012.
- [15] W. Y. Chiang, "To mine association rules of customer values via a data mining procedure with improved model: an empirical case study," *Expert Systems with Applications*, vol. 38, no. 3, pp. 1716–1722, 2011.
- [16] P. Kolarovszki, J. Tengler, and M. Majerčáková, "The new model of customer segmentation in postal enterprises," *Procedia - Social and Behavioral Sciences*, vol. 230, pp. 121–127, 2016.
- [17] M. Song, X. Zhao, E. Haihong, and Z. Ou, "Statistics-based CRM approach via time series segmenting RFM on large scale data," *Knowledge-Based Systems*, vol. 132, no. 15, pp. 282–291, 2017.
- [18] Z. Bao, Y. Zhao, and Y. Li, "Analysis on Baidu take-way customer value based on improved RFM model," *Journal of Xi'an University of Posts and Telecommunications*, vol. 24, no. 1, pp. 105–110, 2019.
- [19] J. MacQueen, "Some methods for classification and analysis of multivariate observations," in *Proceedings of the fifth Berkeley Symposium on Mathematical Statistics and Probability*, vol. 1, no. 14, pp. 281–297, 1967.
- [20] D. Arunachalam and N. Kumar, "Benefit-based consumer segmentation and performance evaluation of clustering approaches: an evidence of data-driven decision-making," *Expert Systems with Applications*, vol. 111, pp. 11–34, 2018.
- [21] A. Mesforoush and M. J. Tarokh, "Customer profitability segmentation for SMEs case study: network equipment company," *International Journal of Research in Industrial Engineering*, vol. 2, no. 1, pp. 30–44, 2013.
- [22] A. K. Jain, "Data clustering: 50 years beyond  $K$ -means," *Pattern Recognition Letters*, vol. 31, no. 8, pp. 651–666, 2010.
- [23] D. Steinley, "K-means clustering: a half-century synthesis," *British Journal of Mathematical and Statistical Psychology*, vol. 59, no. 1, pp. 1–34, 2006.
- [24] J. Wu, L. Shi, W. P. Lin et al., "An empirical study on customer segmentation by purchase behaviors using a RFM model and  $K$ -means algorithm," *Mathematical Problems in Engineering*, vol. 2020, Article ID 8884227, 7 pages, 2020.
- [25] D. Arthur and S. Vassilvitskii, *k-Means++: the advantages of careful seeding*, Stanford, 2006.
- [26] P. J. Rousseeuw, "Silhouettes: a graphical aid to the interpretation and validation of cluster analysis," *Journal of computational and applied mathematics*, vol. 20, pp. 53–65, 1987.