Customer Relationship Management System with Decision Support

Qiyi Liang[1], Ma. Visitacion Gumabay[2]

1. Yangjiang Vocational and Technical College, Yangjiang 529566 , Guangdong, China
2. St.Paul University Philippines,Tuguegarao City, Cagayan 3500

Abstract: This paper studied customer relationship management system with decision support. At present, the fitness studio management system has accumulated a large amount of customer data, but the studio still has problems in customer management such as inaccurate customer target positioning and imperfect customer relationship management functions. It is still unable to conduct intelligent analysis, resulting in a decline in the overall management ability of customer groups. As such, customer turnover is high, and customer loyalty is reduced. To solve such problems, this research constructed a customer relationship management system with decision support, which can be used to identify customers with high studio and high platform loyalty, and with repeated purchase of GT fitness studio. With the purpose of helping the studio achieve accurate customer relationship management and business decision management, this study built an intelligent customer cluster analysis system through K-means clustering algorithm, which can automatically cluster customers into four (4) categories, indicate the characteristics of consumer behavior under each category, and screen out high-quality fitness customers with large consumption amount, high consumption frequency, high platform evaluation, high loyalty, and heavy consumption. At the same time, it also screens out general customers. Moreover, the intelligent group display of the clustered customers is visualized in the form of reports, and the visual display of each group is also carried out, so that decision makers can conveniently implement accurate group management of different groups of customers.

Keywords: Customer relationship management; Decision support; Data mining; Fitness studio; Visual analysis

1 Introduction

This study mainly developed a fitness studio management system for GT Fitness Studio. GT Fitness Studio is a sports fitness membership private training studio, which had been established for more than ten years, and now has dozens of fitness coaches, and a number of franchise stores. The customer competition of the fitness studio is fierce, and the customer loss is serious, so it is urgent to use data analysis and mining technology for customer relationship management, so as to find high-quality members, improve management level, and win the market competition.

2 Methodology

2.1 Research design

The system was be developed using a rapid prototyping model, which required the rapid construction of a working software prototype to understand and clarify issues, enable developers and users to reach a consensus, and ultimately develop software products that satisfy customers based on identified customer needs. Figure 1 shows that rapid prototyping consists of five (5) basic steps:
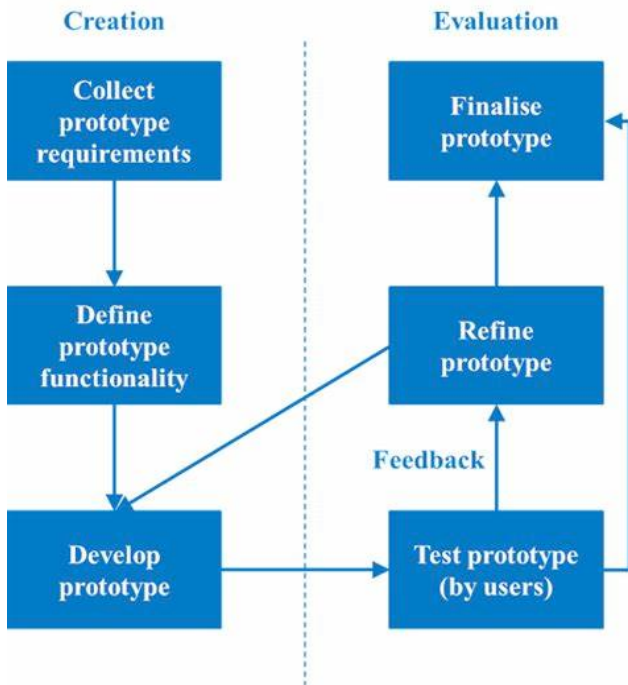
Figure 1 Rapid Prototyping Flowchart

2.2 System architecture

Figure 2 shows that the architecture of the system consists of five (5) layers: data source and data storage tier, data processing tier, data analysis and mining tier, application system tier, and data presentation.

Figure 2 System architecture

2.3 Use Case Diagram

Users of the system are divided into common users and administrator. The common user use case diagram is shown

in Figure 3. The administrator user usage is shown in Figure 4, and the overall system usage is shown in Figure 5.
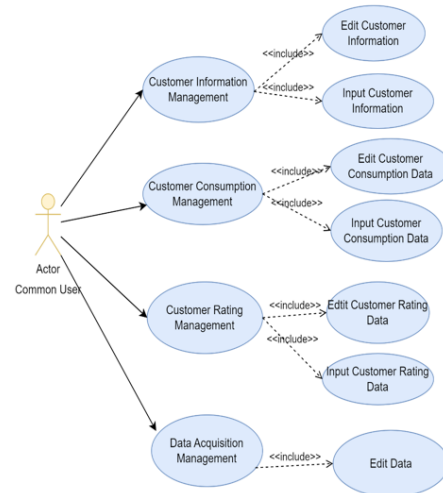

Figure 3 Common user use case
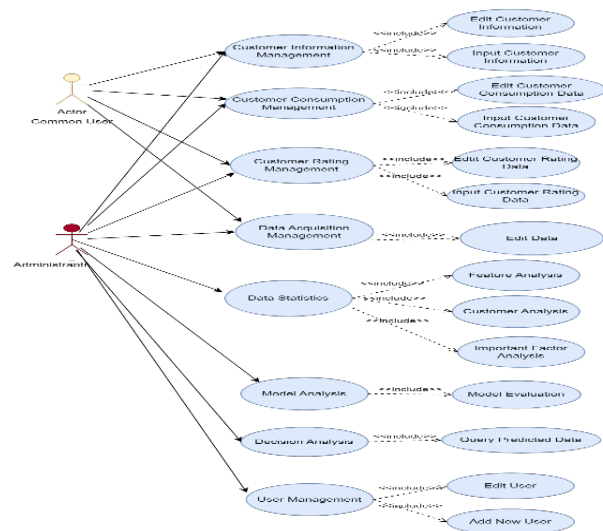

Figure 4 Administrator use case

Figure 5 System use case

3 Results and discussions

3.1 Problems and challenges encountered by GT fitness studios in CRM

Through systematic and thorough interviews, observations, and file scans, the researchers identified some of the problems and challenges GT Fitness Studios encountered in CRM. It was found that the while Studios management system has accumulated a lot of customer data, there are still some problems in the customer management such as inaccurate customer target positioning and imperfect customer relationship management. It still cannot conduct intelligent analysis and mining of customers and cannot intelligently predict high-quality target customers, resulting in a decline in the overall management ability of customer groups, high customer turnover, and low customer loyalty.

3.2 Significant predictors used in the customer relationship management analysis

3.2.1 K-Means algorithm

In this study, the algorithm was used to automatically learn a classification label according to the training data, so as to identify different groups of users, in order to distinguish the advantages and disadvantages of customers.

3.2.2 Significant predictors

Based on the investigation and data cooperation of GT fitness studio, relevant data was extracted from the application business system to compose the sample data.

After data collection, the sample data for model training was taken, but needed to be sorted and cleaned before use. After data collection, 66,759 samples were obtained. Next, data analysis and data modeling are carried out through python language.

The original data set also needed to be cleaned before mining, because there are still some data that do not meet the requirements of classification and mining, such as missing values.

The data features are: UserID, OrderDate, OrderNumber, and OrderAmount. Data characteristics: UserID, FacilityRating, EnvironmentRating, ServiceRating, and OverallRating. Each represents the customer number, and the customer scores the facilities, environment, service,

and comprehensive scores of the studio.

This study intended to use K-means clustering algorithm for research, which is an unsupervised algorithm without predictive factors. The result of clustering was a new data label for each piece of data. The detailed description of each field is detailed in the significant predictors below:OrderDate, OrderNumber, and OrderAmount, which are the customer's sales data factors in the studio. In addition, there are also factors for customers to score studios, such as FacilityRating, EnvironmentRating, ServiceRating, and OverallRating.

3.3 Algorithm and model that can be used to analyze the high-quality customers of fitness studios

3.3.1 Data modeling

In this study, cross-industry Data Mining Standard process (CRISP-DM) was adopted for data analysis and mining modeling. According to the basic process of CRISP-DM model, it starts from business understanding and data understanding, to data preparation, to modeling, then to evaluation, and finally to model deployment.

3.3.2 Consumption data analysis

(1) Import data to the platform

The data is imported into the mining platform through python as a data training set. As shown in Figure 6.

	ID	UserID	OrderDate	OrderNumber	OrderAmount
0	1	GT115840	2023/5/10	4.0	89.50
1	2	GT109700	2023/2/5	1.0	38.92
2	3	GT121628	2023/12/1	3.0	92.94
3	4	GT114229	2023/6/5	1.0	50.40
4	5	GT122026	2023/3/25	1.0	27.54
5	6	GT105503	2023/2/6	1.0	28.94
6	7	GT118193	2023/3/5	1.0	29.58
7	8	GT120867	2024/4/11	1.0	42.98
8	9	GT118735	2023/8/31	NaN	56.46
9	10	GT100424	2023/6/8	2.0	61.12
10	11	GT108550	2023/2/1	1.0	26.94
11	12	GT116853	2024/5/16	NaN	65.96
12	13	GT104038	2023/1/17	1.0	37.74
13	14	GT120031	2023/3/11	1.0	27.94
14	15	GT101356	2023/1/6	3.0	93.98
15	16	GT120017	2023/9/23	1.0	44.98
16	17	GT117888	2023/8/21	1.0	15.96
17	18	GT111475	2024/4/16	5.0	101.26
18	19	GT111711	2023/7/7	1.0	33.74
19	20	GT107221	2023/1/28	1.0	27.54

Figure 6 Customer spending data set

(2) View data information

Figure 7 shows the basic information of the data set. From the figure, the dimension of the data set, the size of the sample, the name and value of the data feature, and the type of the data feature is evident.

```
df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 67616 entries, 0 to 67615
Data columns (total 5 columns):
 #   Column       Non-Null Count  Dtype
---  ------       --------------  -----
 0   ID           67616 non-null  int64
 1   UserID       67616 non-null  object
 2   OrderDate    67616 non-null  object
 3   OrderNumber  67616 non-null  int64
 4   OrderAmount  67616 non-null  float64
dtypes: float64(1), int64(2), object(2)
memory usage: 2.6+ MB
```

Figure 7 Data informatio

Missing value processing.Missing value refers to the data missing or truncated due to lack of information in the rough data, resulting in incomplete values of one or some attributes in the existing data set. Using python, the missing values for the training set were seen. Here, the method of filling the mean is used to deal with missing values, and the processing results are shown in Figure 8.

```
ID             0
UserID         0
OrderDate      0
OrderNumber    0
OrderAmount    0
dtype: int64
```

Figure 8 Missing value processing result

(3) Duplicate value processing

First, the data set was reviewed repeatedly.As seen from the figure 9, there are no duplicate values, so there is no need to process duplicate values.

```
: print(df[df.duplicated()])

Empty DataFrame
```

Figure 9 Duplicate value viewing

(4) Handle outliers

After the simple data preprocessing, missing values, outliers, and duplicate values have been processed to obtain a clean and complete data set, and the data was re-saved to the table and database. The result of data set cleaning is shown in Figure 10.

	ID	UserID	OrderDate	OrderNumber	OrderAmount
0	1	GT115840	2023/5/10	4	89.50
1	2	GT109700	2023/2/5	1	38.92
2	3	GT121628	2023/12/1	3	92.94
3	4	GT114229	2023/6/5	1	50.40
4	5	GT122026	2023/3/25	1	27.54
5	6	GT105503	2023/2/6	1	28.94
6	7	GT118193	2023/3/5	1	29.58
7	8	GT120867	2024/4/11	1	42.98
8	9	GT118735	2023/8/31	2	56.46
9	10	GT100424	2023/6/8	2	61.12
10	11	GT108550	2023/2/1	1	26.94
11	12	GT116853	2024/5/16	2	65.96
12	13	GT104038	2023/1/17	1	37.74
13	14	GT120031	2023/3/11	1	27.94
14	15	GT101356	2023/1/6	3	93.98
15	16	GT120017	2023/9/23	1	44.98
16	17	GT117888	2023/8/21	1	15.96
17	18	GT111475	2024/4/16	5	101.26
18	19	GT111711	2023/7/7	1	33.74
19	20	GT107221	2023/1/28	1	27.54

Figure 10 The cleaned data set

(5) Descriptive statistics

The following descriptive statistics were carried out, including the analysis of the central trend of the data, the analysis of the degree of data dispersion, the analysis of frequency, and the distribution of the data.

(6) Feature analysis

Data visualization technology was used to analyze two continuous feature variables, OrderNumber and OrderAmount.The histogram of each feature is shown in Figures 11 and 12.
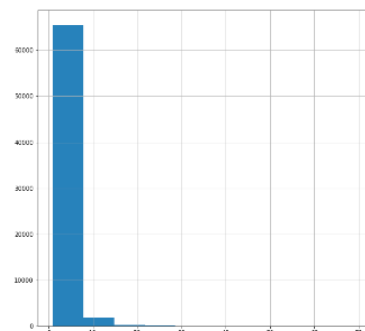

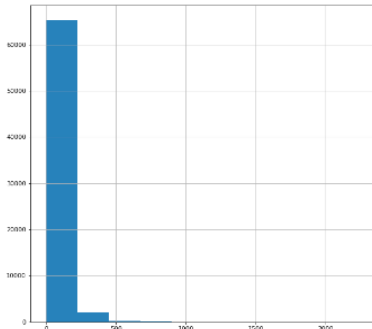Figure 11 Consumption item data distribution

Figure 12 Monthly consumption data

At the same time, the following data are also analyzed: Analysis of Consumption Frequency in Different Months. Analysis of Consumption Amount in Different Months. Relationship between Order Quantity and Consumption Amount. User Consumption Amount Analysis. Data Cluster Analysis and Modeling. Score Data Distribution Analysis. Analysis of the Relationship Between Service Rating and Sales. Relationship between Studio Facility Rating and Sales. Relationship between Studio Environment Rating and Sales. Relationship between the Overall Score of Fitness Studio and the Sales Amount.

3.3.3 Building model

(1) System model training

Through data preprocessing, the data have been preliminarily cleaned and analyzed, and the important characteristics of the predicted variables have been clarified. In order to model the data, it was divided into training sets, test sets, and validation sets. A clustering algorithm is used to model the data, and python machine learning library Scikit-learn was used for machine learning training.

(2) K-means determination of K value of clustering parameter

As seen from Figure 13, when the parameter value is 4, the clustering number and the K value is the same.
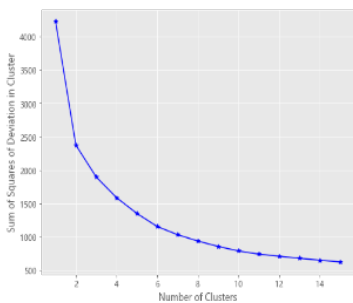

Figure 13 Elbow chart

(3) Modeling results, visualization, and evaluation

After modeling, a new clustering label was generated on the basis of the training set. Since this group is only labeled, in order to better explain the group, further aggregate analysis is carried out on the data of each group (cluster) of the cluster, as shown in Figure 14.

Cluster	OrderNumber	OrderAmount	FacilityRating	EnvironmentRating	ServiceRating	OverallRating
0	22.438841	709.231024	5.793187	5.755364	5.777040	5.801371
1	3.005368	95.596483	4.298408	1.926138	3.151796	3.135690
2	3.523424	112.197032	3.005481	4.380399	3.114968	3.505546
3	2.993483	95.249856	1.710561	2.096591	2.613803	2.110795

Figure 14 Aggregate analysis of each data cluster

The figure above is an aggregate analysis of the data characteristics of each data group. As seen from the figure, each grouping (cluster) has the following characteristics:

Group 0: All kinds of high scores, high orders, and the highest consumption amount.

Group 1: Medium score, medium consumption and order, and important retention customer of the company.

Group 2: The score is low, but the consumption and orders are medium. They are customers that the company needs to further develop.

Group 3: The score is very low; the consumption is also very low. As such, they are company's general customers.

(4) Save the data to the database for the use of the system

After the data clustering has been completed on the surface, and the clustering results were saved in batches and stored in the database for the use of the system to provide decision support for the system.

3.4 System that can be developed to address the problems and challenges encountered

Customer Relationship Management System with Decision Support is divided into 9 functional modules for design. The main functions of the system are as follows:

Users can click "Customer Management" to view all customer information, including adding, deleting and modifying customer information. As shown in Figure 15.

Figure 15 Customer management module

The consumption management module achieves the effect shown in Figure 16.Users can view, add, delete and modify customer consumption information by clicking "Consumption Management".

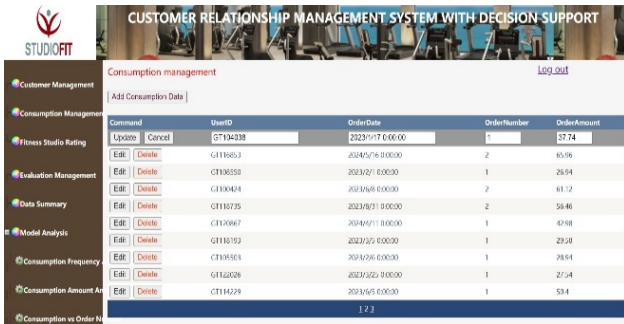
Figure 16 Consumption management module

(1)Rating Management Module.

Customers can evaluate the environment, equipment and services of the fitness studio. Administrator review, add, delete modify evaluation, etc. As shown in Figure 17.

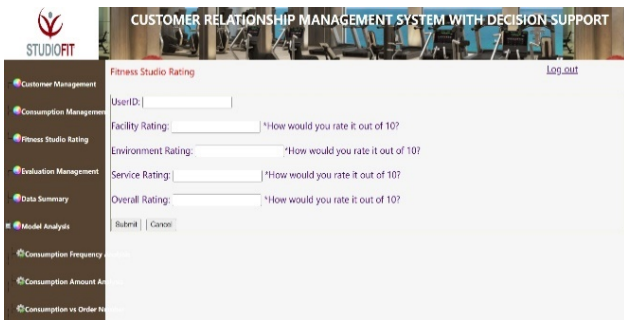(2)Consumption frequency analysis module

Figure 17 Fitness studio rating module

The system makes an in-depth analysis of the consumption frequency of fitness studio customers.

The consumption frequency is the highest from January to March in 2023, reaching a peak in March, and then, the consumption has a trend of decline month by month, which may be related to the closure of a number of franchised chain stores. There was another rebound from January to March 2024, peaking in March 2024. Show in Figure 18.
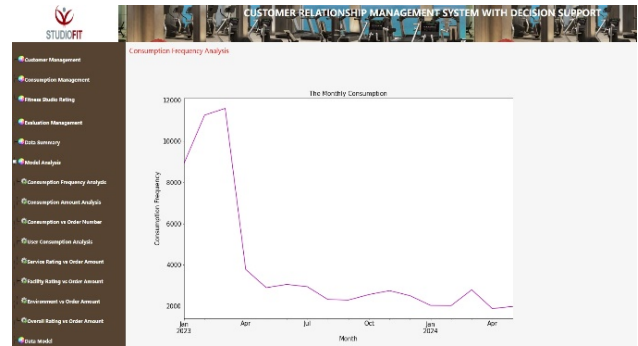

Figure 18 Consumption frequency analysis

(3)Consumption vs order number module

Systematically analyze the relationship between order quantity and consumption amount. Upon analysis, it was found that the amount of consumption and the number of orders are basically linear, indicating that the conversion rate of order volume into sales is high, and the more users place orders (or become members), the greater the total amount of consumption. Show in Figure 19.

Figure 19 Consumption vs order number module

(4)Environment Rating vs Order Amount Analysis module

The higher the fitness studio environment score, the larger the sales range, the larger the sales range, the scores are above 6, which indicates that sales range significantly increased, especially in the scores of 8 and 9 points. The largest sales range, and the higher the environmental score, the more likely to generate large sales - among them, the score of 5 has a largest sale, which may be a random event. Show in Figure 20.
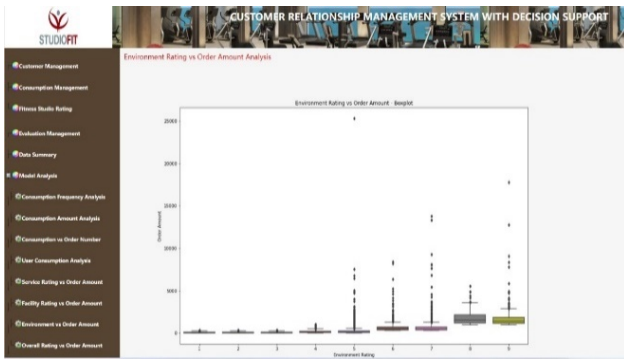
Figure 20 Environment rating vs order amount analysis module

(5)Decision analysis module

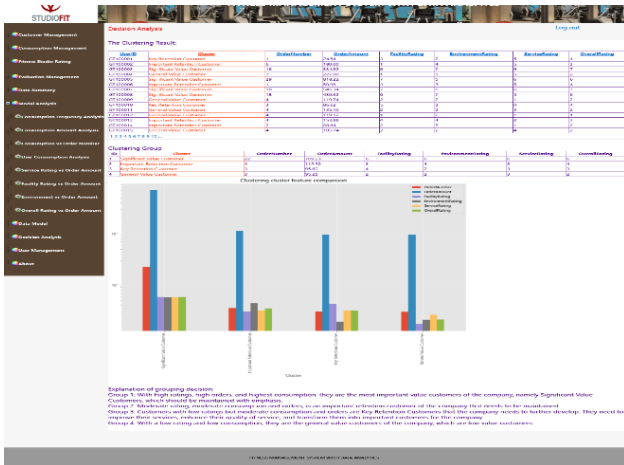The Decision Analysis module presents the value of each customer in detail. Show in Figure 21.

Figure 21 Decision analysis module

4 Extent of compliance of the developed application with ISO 25010 software quality standards

Table 1 Summary Table on the Extent of Compliance of the Developed System to the ISO/IEC 25010 Software Quality Standards

	Criteria	Mean	Descriptive Interpretation
1.	Functionality	4.57	Very Great Extent
2.	Performance Efficiency	4.50	Very Great Extent
3.	Compatibility	4.50	Very Great Extent
4.	Usability	4.43	Very Great Extent
5.	Reliability	4.53	Very Great Extent
6.	Security	4.54	Very Great Extent
7.	Maintainability	4.48	Very Great Extent
8.	Portability	4.50	Very Great Extent
	Overall Category Mean	4.51	Very Great Extent

Table 1 shows the overall result of the system evaluation in compliance with ISO 25010 Software Quality Standards. It shows that the developed system obtained an overall mean of 4.51 with a descriptive interpretation of "Very Great Extent". This implies the developed system complies with the ISO 25010 software quality standard as assessed by IT experts in terms of functionality, performance efficiency, compatibility, usability, reliability, security, maintainability, and portability.

5 Enhancements that can be done to improve the developed system

Based on the assessment of IT experts, measures could be taken improve the existing functions of the system, which include improving the operation efficiency of the system so as to collect more data and provide basic support for further data analysis. Furthermore, fitness studios should assign dedicated administrators and data analysts to use the system to improve the efficiency of decision making and the efficiency of data analysis and marketing analysis. Finally, fitness studios should have dedicated data collection and analysis personnel to collect data regularly to assist decision-makers in analyzing and mining data.

6 Conclusion

Based on the results of this study, the researchers came to the following conclusions:

This paper studied the customer relationship management system with decision support. A customer relationship management system with decision support was built to tap high-quality customers with high studio and platform loyalty, and repeated purchase of GT Fitness Studios.

The intelligent customer cluster analysis system was built by K-means clustering algorithm, which automatically clusters customers into four (4) categories, points out the consumer behavior characteristics of each category of customers, and selects high-quality fitness customers with large consumption amount, high consumption frequency, high platform evaluation, high loyalty and

heavy consumption. At the same time, screen out general customers.

References

[1] Wang Xin-Yan, Zhang Xiu-Mei, and Chen Cui-ping.(2024). Behavioral characteristics of smart classroom learning based on educational data mining technology. Education measurement and evaluation (06), 31-46. Doi: 10.16518 / j.carol carroll nki emae. 2024.06.003.

[2] Nana Wang, Wenyi Li & Bruce Lee.(2024). Fault diagnosis method of oil-immersed transformer based on unbalanced small sample DGA data and improved CatBoost decision tree. Power system protection and control (23), 167-176. Doi: 10.19783 / j.carol carroll nki PSPC. 240478.

[3] Zhang Junkai, Li Xin, Han Junxian, Zhao Juan & Cheng Longxue.(2024). Research on picking robot system optimization based on deep learning algorithm. Agricultural mechanization research (04), 58-62. Doi: 10.13427 / j.carol carroll nki njyi. 20230728.001.

[4] Jiang Qingling & Zhang Fan.(2023). Web image crawler program design based on Python and Requests module. (6) computer programming skills and maintenance,.prg 59-61 +95. Doi:10.16184/j.cnki.com. 2023.06.014.