Academic Journal of Computing & Information Science
ISSN 2616-5775 Vol. 6, Issue 3: 111-117, DOI: 10.25236/AJCIS.2023.060314

Research on the Evaluation of Financial Aid for Poor Students in Colleges and Universities Based on K-Means Cluster Analysis

Junyan Tian[#,*], Zhijing Wu[#], Xin Qu[#]

School of Electronic and Optical Engineering, Nanjing University of Science and Technology, Nanjing, Jiangsu, 210094, China
[*]Corresponding author: tianjunyan2021@163.com
[#]These authors contributed equally.

Abstract: *At the time when "smart campus" is popular, the warm care of campus is an important element of smart campus, and the big data consumption evaluation model derived from the analysis of big data and data mining can effectively determine the consumption level and family economic status of various students, so as to accurately identify the students in need of help on campus, and provide important data support for the granting of college The article provides an important data support for grant distribution and facilitates warm-hearted assistance to them. The article classifies each student according to his or her consumption amount and gives the top three hundred students with low consumption levels. Considering the existence of values in the data and the existence of helping classmates to bring meals, only those students who have consumption for all three months and whose consumption amount is below 30 yuan each time are analyzed. A K-Means clustering evaluation model was constructed for them to obtain three types of students with different consumption types, while each student was given a positive rating, which was ranked in ascending order to obtain a list of target students. On this basis, the consumption window factor was added, and after eliminating the non-essential consumption types, the consumption window price level was used to classify them, and two types of low consumption windows and high consumption windows were obtained, while scores were assigned to various windows, and positive scores were given to each student window consumption in turn, and the optimized K-Means clustering evaluation model was used to rank them in ascending order to obtain the target student list.*

Keywords: *K-Means Clustering, Data Mining, Financial Aid for Poor Students*

1. Introduction

The management mode of students in China's colleges and universities basically takes students' accommodation in schools and meals in school canteens as the main form[1]. School diet is one of the most important work of the school, run a good school canteen, to provide students with a satisfactory taste, nutrition and health, health and safety, good value for money catering is the basic requirements of school food service. In recent years, with the continuous application of big data technology in various fields of life, more and more problems can be solved by big data[2].

The core value of big data is the process of extracting the potentially valuable information from the massive data that people cannot predict from the surface. Therefore, we analyze the consumption data of students in a university for several months and build a mathematical model to accurately identify students who need help and provide them with warm care. First of all, we have to analyze and model the consumption data records of college students in March, April and May based on different consumption levels by selecting useful and relevant indicators for clustering analysis[3]. First we identified two dimensional attributes: the mean value of meal consumption and the number of times they did not eat in the cafeteria. Next, these two factors were used to cluster all students into three categories using K-Means clustering: high consumption level, medium consumption level, and low consumption level. Finally, the students' consumption levels were scored and evaluated using the two standardized indicators, and the 300 lowest ranked students were selected and judged as low consumption students[4].

Since there are some sugar water, cakes, etc. that are non-essential consumption windows. These non-essential windows should not be taken into account in the classification, so we first eliminate the data from these windows before classification. For the remaining windows, the average value of each meal

ISSN 2616-5775 Vol. 6, Issue 3: 111-117, DOI: 10.25236/AJCIS.2023.060314

spent is obtained, and they are classified into high consumption and economy windows according to the average consumption[5]. Next, the students' window consumption levels and meal consumption averages were selected as new dimensional factors and K-Means clustering was performed to cluster them into three categories: high window level high consumption, medium window level medium consumption and low window level low consumption. Finally, the two standardized indicators were used to score and evaluate the consumption level of students, and the 300 lowest ranked students were selected and judged as low consumption students[6].

2. Materials and Methods

2.1 K-Means clustering meaning

The K-Means algorithm is a common clustering algorithm whose main idea is to assign each point (i.e., the recorded data) to the class cluster represented by the nearest class cluster centroid given a K value and K initial class cluster centroids, and after all points are assigned, the centroids of the class cluster are recalculated (averaged) based on all points within a class cluster, and then the iterative steps of assigning points and updating class cluster then iterate through the steps of assigning points and updating class cluster centroids until the change in class cluster centroids is small or the specified number of iterations is reached[7] . Since the question requires the consumption level of all students, which is equivalent to the record data, to be assigned to each class cluster, K-Means clustering can be used.

2.2 K-Means clustering implementation process

Assume that given a data sample X, which contains n objects$X = \{X_1, X_2, X_3, \cdots, X_n\}$, where each object has m dimensional attributes, and for this paper, we select two dimensional attributes: the mean value of meal consumption and the number of meals not eaten in the canteen.

Initialize k clustering centers$\{C_1, C_2, C_3, \cdots, C_k\}$, where$1 < k \leq n$, and then by calculating the Euclidean distance formula for each object to each cluster center, the formula is as follows[8].

$$dis(X_i, C_j) = \sqrt{\sum_{t=1}^{m} (X_{it} - C_{jt})^2} \tag{1}$$

In the above equation, theX_i denotes the i-th object$1 \leq i \leq n$, andC_j denotes the jth clustering center of$1 \leq j \leq k$, andX_{it} denotes the tth attribute of the ith object, the$1 \leq t \leq m$, C_{jt} denotes the tth attribute of the jth clustering center.

Compare the distance of each object to each cluster center in turn, assign the object to the rayon of the nearest cluster center, and get k class clusters$\{S_1, S_2, S_3, \cdots, S_k\}$[9].

The K-Means algorithm defines the prototype of class clusters in terms of centers, and the class cluster center is the mean value of all objects within the class cluster in each dimension, which is calculated as follows[10].

$$C_t = \frac{\sum X_i \in s_l \ X_i}{|S_l|} \tag{2}$$

whereC_l denotes the center of the 1st cluster, the$1 \leq l \leq k$, and$|S_l|$ denotes the number of objects in the 1st class cluster, andX_i denotes the i-th object in the 1st class cluster, and$1 \leq i \leq |S_l|$.

3. Model construction and solution

3.1 Classification modeling solution based on consumption level

The In order to better implement the campus big data modeling and to accurately identify students in need of help on campus, we used the data of students' cafeteria consumption records for three months from March to May 2020 given in the question to model and analyze the clustering of different consumption levels.

ISSN 2616-5775 Vol. 6, Issue 3: 111-117, DOI: 10.25236/AJCIS.2023.060314

For the canteen consumption record data given in the attachment, we first sorted the student number column to group all consumption records of the same student together. It is easy to find that the only data that have an impact on a student's consumption level are the consumption time and the consumption amount. To facilitate the processing of the data, we divided the day into three time periods according to the consumption time, as shown in Table 1.

Table 1: Meal time division

Time Period	Meals
6:00~11:00	Breakfast
11:00~14:00	Chinese food
16:00~20:00	Dinner

According to the needs of the study, we classify the consumption levels of students, which are clustered into 3 main categories. The mean values of these 3 categories of consumption data, corresponding to different consumption levels, can be derived as the type of consumption of each student, which are.

(1) High consumption level: Students who have not eaten in the cafeteria more often and also have a higher average consumption value when they eat in the cafeteria are considered to have a high consumption level.

(2) Moderate consumption level: without considering the average of consumption in the cafeteria, only the number of times they did not eat in the cafeteria was considered. If the number of times they did not eat in the cafeteria was at a moderate level, then it means they ordered take-out more often, and then such students were considered to be at a moderate consumption level.

(3) Low consumption level: The mean value of meal consumption is low, while the number of meals not eaten in the cafeteria is low, and we consider this category of students as low consumption level.

3.2 Modeling solutions based on window consumption habits

The article classifies the nature of the cafeteria window column: into high consumption window and economic window. We believe that the criterion to distinguish a window as high consumption or economic is the average value of consumption of each meal, and a high average value is a high consumption window; a low average value is a low consumption window. In order to calculate the consumption average of each window, we first classify the window column in the data, take the average of the consumption of the same group, and arrange all the averages in descending order, where the first 50% of the windows we set them as high consumption windows, and the last 50% of the windows we set them as economic windows. The calculated results are shown in Table 2: where the top 10 high-spending windows are given on the left and the top 10 economical windows are given on the right.

Table 2: Window class classification

Windows	Total average	Windows	Total average
Economic Stir Fry	15.78581	Sauce group	4.86363
Fujian Flavor Group	15.81665	Snack group	4.94174
Hot and spicy pot group	15.96491	Sauce cake set	5.121877
Rice in a barrel	16.58858	Big Bowl Dishes	6.597946
Spicy Chicken	16.99821	Large pancake group	6.602045
Stew group	17.22857	tile pot	6.709219
Aroma Pot Group	17.73656	Nanjing Snacks	6.936093
Spicy Hot Pot	18.84471	Staple food	6.985663
Stir-fry group	19.26961	Large group	7.757182
Five Flavors Theme Restaurant	20.98839	Soup bun group	7.795664

On the basis of normalizing the window data, the data were classified according to the student number and the student's consumption window was replaced with the normalized value found to fill in, and the amount spent within the cafeteria during these three months was multiplied by the normalized value of that window to finally find the sum, and this consumption sum can be taken as the student's window consumption habits. In order to build the consumption habit model, we also introduce the consumption

ISSN 2616-5775 Vol. 6, Issue 3: 111-117, DOI: 10.25236/AJCIS.2023.060314

level ability in question 1 into this model, and use these two dimensional factors as the measurement criteria to perform K-Means clustering analysis on students again. The clustering is mainly implemented by dividing the clusters into the following three categories.

(1) High window class with high consumption: Students in this category go to the high consumption window more often than the economy window and spend a higher amount per meal on average.

(2) Medium window level medium consumption: This category of students go to the economy window and the high consumption window almost equally when dining, and at the same time the consumption level at the window is medium.

(3) Low window class and low consumption: The main consumption window for students in this category is the economy window, and they spend less on each meal.

4. Results and Analysis

4.1 Analysis of clustering results based on consumption level

The consumption data of all students were processed through SPSSPRO according to the above clustering method, and the processed results are shown in the figure below, with the X-axis being the standardized mean value of students' meal consumption and the Y-axis being the standardized number of meals eaten in the cafeteria. In the graph, 1 represents high consumption level, 2 represents medium consumption level, and 3 represents low consumption level. The clustering results of students' consumption habits based on the consumption level of dining out are shown in Figure 1.

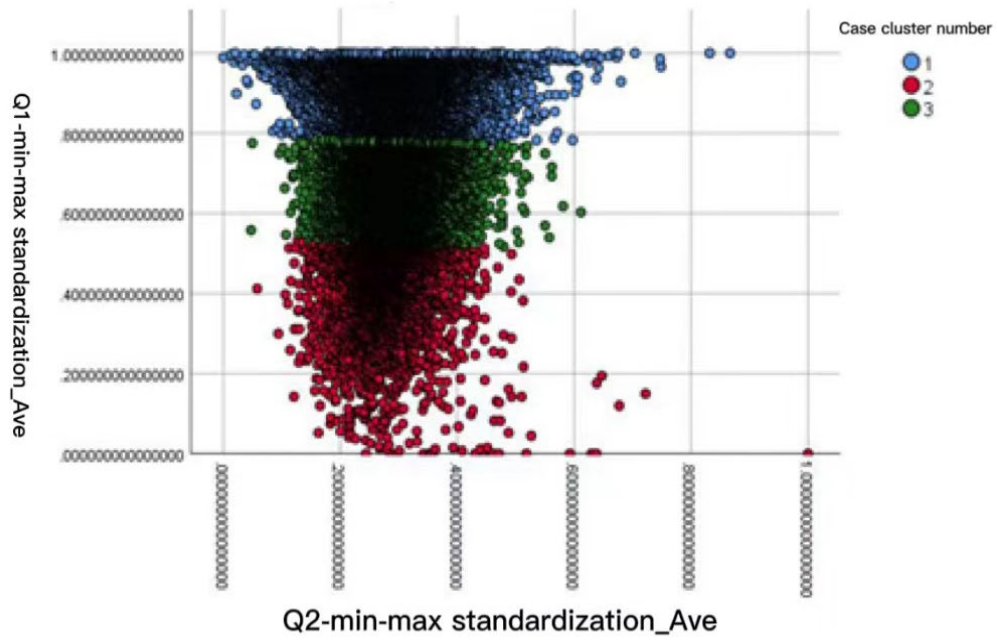

Figure 1: Clustering chart of students' consumption level

The final values and results of the three clustering center coordinates for high consumption level, medium consumption level and low consumption level are shown in Table 3.

Table 3: Clustering results of three types of consumption levels

Clustering	**1-High consumption level**	**2-Medium consumption level**	**3 - Low consumption level**
Number of meals not eaten in the cafeteria	0.8997714184	0.3928037802	0.6600245924
Meal consumption average	0.2996490234	0.2769792338	0.2931408989

The distribution of the number of cases in the three clusters of high consumption level, medium consumption level and low consumption level is shown in Table 4.

ISSN 2616-5775 Vol. 6, Issue 3: 111-117, DOI: 10.25236/AJCIS.2023.060314

Table 4: Number of cases in the three consumption levels

Clustering	1-High consumption level	7796
	2-Medium consumption level	3394
	3 - Low consumption level	6369
Number of active cases		17559
Number of cases indeed		0

In the clustering result graph, the closer the coordinate point is to the lower left means the dining consumption level and the number of times not eating in the cafeteria are both low, and the closer the coordinate point is to the upper right means the dining consumption level and the number of times not eating in the cafeteria are both high. The list of the 300 students with low consumption after weighted calculation is put in excel and the top ten numbers are given in the text shown in Table 5.

Table 5: Top ten numbers of students whose meals were judged to be of low consumption level

Student Number	Standardization of the number of meals not taken to the cafeteria	Standardization of student consumption levels	Consumption score
Z0788178	0.052434457	0.162612968	0.107523713
Z0517149	0	0.244112144	0.122056072
Z9787178	0.041198502	0.215028991	0.128113747
Z0517448	0.056179775	0.200399672	0.128289724
Z1757205	0.074906367	0.183118467	0.129012417
Z2547488	0.142322097	0.12074973	0.131535914
Z1597041	0.037453184	0.225911284	0.131682234
Z8827875	0.052434457	0.216851318	0.134642887
Z0517903	0.086142322	0.186665068	0.136403695
Z0027880	0.08988764	0.184042237	0.136964939

4.2 Analysis of clustering results based on window consumption habits

The clustering results are shown below, with the X-axis being the average of the standardized student meal consumption and the Y-axis being the window consumption level of the student after standardization. In the graph, 1 represents high consumption in high window level, 2 represents medium consumption in medium window level, and 3 represents low consumption in low window level. The results of window-based clustering of students' consumption habits are shown in Figure 2.

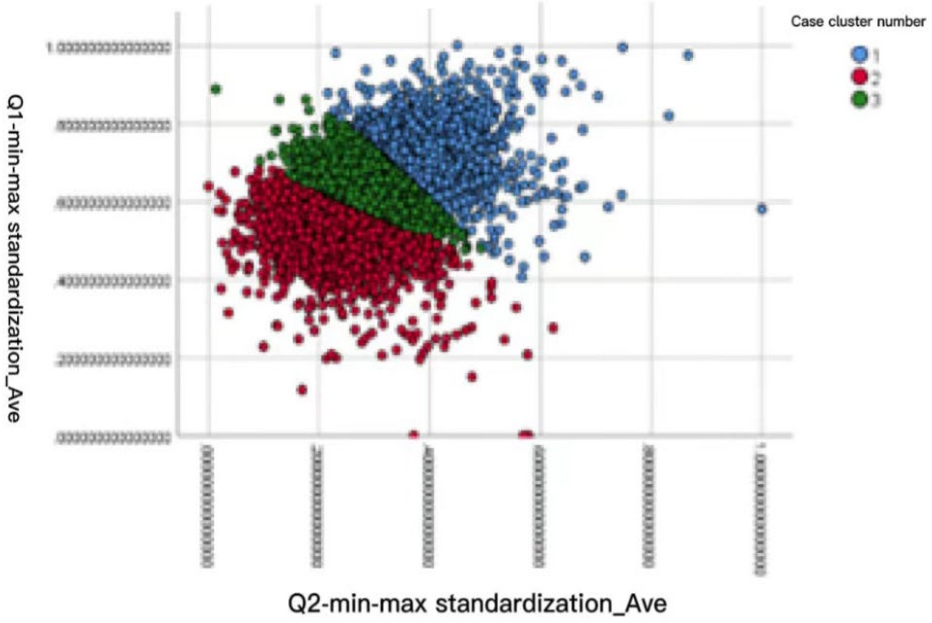

Figure 2: Clustering of students' consumption habits

The final cluster center coordinate values for the three classes of high window class, medium window class and low window class are shown in Table 6.

ISSN 2616-5775 Vol. 6, Issue 3: 111-117, DOI: 10.25236/AJCIS.2023.060314

Table 6: Three types of window level clustering results show

Clustering	1-High window level	2-Medium window level	3 - Low window level
Window consumption level	0.7214472613	0.5476419526	0.6473333411
Meal consumption average	0.3841605620	0.2258811894	0.2877727707

The distribution of the number of cases in the three clusters of high window level, medium window level and low window level is shown in Table 7.

Table 7: Number of cases showing the three types of window levels

Clustering	1-High window level	4070
	2-Medium window level	4882
	3 - Low window level	8607
Number of active cases		17559
Number of cases indeed		0

In the clustering result graph, the closer the coordinate point is to the lower left means the dining consumption level and the window consumption level are both low, and the closer the coordinate point is to the upper right means the dining consumption level and the window consumption level are both high. The list of the 300 students with low consumption was weighted and placed in excel, and the top ten numbers were given in the text as follows shown in Table 8.

Table 8: Top ten numbers of students judged as low consumption level by the window

Student Number	Window consumption level standardization	Standardization of dining consumption levels	Score
Z7788814	0.117264	0.169238	0.143251
U1444230	0.2278	0.099487	0.163643
Z2744577	0.314814	0.035932	0.175373
G1597041	0	0.370831	0.185416
Z0517880	0.376318	0.022503	0.199411
G9787951	0.282149	0.123196	0.202673
G9788483	0.280484	0.125218	0.202851
G8757808	0.247122	0.162892	0.205007
G9727048	0.19725	0.212834	0.205042
Z0788352	0.207441	0.222125	0.214783

5. Conclusion and Discussion

5.1 Advantages of the model

This paper mainly uses the K-Means clustering method to cluster and analyze the situation of poor students based on the information of daily food consumption data of college students, so as to lay the foundation for more accurate, wiser and more efficient financial support for students in need in colleges and universities. The K-Means clustering method has the advantages of relatively simple principle, easy implementation, fast convergence, better clustering effect, better interpretability of the algorithm, and the main parameter to be tuned is only the number of clusters k.

5.2 Model Deficiencies and Improvements

The article suffers from the problems of poor selection of K-values, difficult convergence for data sets that are not convex, poor clustering effect and the use of iterative methods that yield only locally optimal results in the process of modeling.

The clustering centers can be obtained once by the K-means algorithm by giving a suitable value to k at the beginning. For the obtained clustering centers, the nearest classes are merged according to the distance of the obtained k clusters, so the number of clustering centers decreases, and when it is used for the next clustering, the corresponding number of clusters also decreases, and finally a suitable number of

ISSN 2616-5775 Vol. 6, Issue 3: 111-117, DOI: 10.25236/AJCIS.2023.060314

clusters is obtained. The number of clusters can be determined by a judging value E to get a suitable position to stop without continuing to merge the clustering centers. Repeat the above cycle until the judging function converges, and finally get the clustering result with the better number of clusters. Or choose density-based clustering algorithm is more suitable, such as DESCAN algorithm.

References

[1] https://blog.csdn.net/qq_32892383/article/details/80107795. K-Means clustering algorithm detailed 2022.5.14
[2] Lin Chongde. Dictionary of Psychology: Shanghai Education Press, December 2003
[3] Sun YH, Zhang MD. Hierarchical DEMATEL analysis method based on R-type clustering [J]. Practice and understanding of mathematics, 2019, 49(06): 42-51.
[4] Li Weichun. Application of Q-type clustering method based on IBM-SPSS in teacher evaluation [J]. Fujian Computer, 2017, 33(01): 150-151. DOI: 10.16707/j.cnki.fjpc. 2017.01.079.
[5] Yi Guijiao. Using big data to build a precise financial aid system for poor students in colleges and universities [J]. School Party Construction and Thought Education, 2020(22): 28-30.
[6] Jiang Dongxing, Fu Xiaolong, Yuan Fang, Wu Haiyan, Liu Qixin. Discussion on the construction of university wisdom campus under the background of big data [J]. Journal of East China Normal University (Natural Science Edition), 2015(S1): 119-125+131.
[7] Yang Junbang, Zhao Chao. A review of research on K-Means clustering algorithm [J]. Computer Engineering and Applications, 2019, 55(23): 7-14+63.
[8] Zhang Yuanhang. On the "precise financial assistance" for students with family economic difficulties in colleges and universities [J]. Thought Theory Education, 2016(01): 108-111. DOI: 10. 16075/ j. cnki. cn31-1220/g4. 2016.01.020.
[9] Tang Shenwei, Jia Ruiyu. A k-mean clustering algorithm based on improved particle swarm algorithm [J]. Computer Engineering and Applications, 2019, 55(18): 140-145.
[10] Tang Dongkai, Wang Hongmei, Hu Ming, Liu Gang. An improved K-means algorithm for optimizing initial clustering centres [J]. Small microcomputer systems, 2018, 39(08): 1819-1823.