Journal of Combinatorial Mathematics and Combinatorial Computing
www.combinatorialpress.com/jcmcc

Research on the Transformation and Upgrading Path of Student Management in Colleges and Universities Based on Artificial Intelligence

Mu Mu[1,✉]

[1]*School of Digital Arts, Suzhou Industrial Park Institute of Services Outsourcing College, Suzhou, Jiangsu, 215123, China*

ABSTRACT

Colleges and universities are an important part of higher education, providing a large number of talents for social development. The study optimizes the way of student management in colleges and universities based on artificial intelligence technology. Firstly, the K-means algorithm in cluster analysis is used to classify students' campus behavioral characteristics. Then use Apriori algorithm to correlate students' behavioral characteristics with academic performance. Finally, colleges and universities can take differentiated management measures for different categories of students. The clustering analysis of 12,885 students' consumption behavior, work and rest behavior, and study behavior in college Z, followed by the correlation analysis between the clustering results and academic performance, and a total of 10 correlation rules were found. Colleges and universities can formulate management rules based on the analysis results to improve management efficiency. In addition, the student management work of colleges and universities can be optimized and upgraded in several directions, including the awareness of student management work in colleges and universities, the information platform, the archive management work, the student management team, and the information security work.

Keywords: Artificial intelligence; K-means algorithm; Apriori algorithm; college student management

1. Introduction

In recent years, the rapid development of global artificial intelligence (AI) technology has become an important power source to promote scientific progress and economic acceleration, and AI technology has entered the stage of commercial operation, which is profoundly changing all aspects of people's lives. Artificial Intelligence (AI) is a technology in which computer systems are able to perceive, reason, learn and make decisions to solve problems and complete tasks by imitating human intelligence [1-3]. Its application in image recognition can be achieved through computer vision technology for face recognition,

✉ Corresponding author.
E-mail address: 18651110126@163.com (M. Mu).
Received 20 May 2024; Revised 19 July 2024; Accepted 22 August 2024; Published Online 15 April 2025.
DOI: 10.61091/jcmcc127a-035

object recognition and other functions [4-5]. Its application in natural language processing can recognise and understand human language, making intelligent assistants, intelligent translation, intelligent customer service and so on possible [6-7].

The current artificial intelligence technology also profoundly affects the process of China's higher education reform. Whether it is the content of higher education teaching, teaching methods, teaching tools and teaching management have been embedded with a lot of technical elements of the artificial intelligence era, which has had an impact on the division of disciplines in colleges and universities as well as on students' learning content and learning methods [8-11]. Compared with the stereotype and inefficiency of the single offline management mode under the traditional education model, the widely established academic work system or teaching affairs system enables college students to enjoy the seamless integration of global education resources, scientific analysis and instant evaluation of learning outcomes [12-15]. Of course, student management in colleges and universities will become more humane and personalised with the development of AI technology [16]. Through humane management, colleges and universities can create a warm and inclusive learning atmosphere, allowing students to thrive in a caring environment, while personalised education helps students to give full play to their strengths and maximise their personal value [17-20].

The study first constructs behavioral indicators from students' consumption, work and rest, and study, and uses the improved K-means algorithm to perform cluster analysis to summarize the characteristics of different behavioral habits. Then the improved Apriori algorithm was used to mine the association rules between students' behavioral habits and grades, and analyze the correlation between students' behavioral habits and grades. According to the mined association rules, differentiated management is adopted for different categories of students to improve management efficiency. Finally, the effective paths of rationally optimizing the awareness of student management work in colleges and universities, scientifically constructing the information platform, strengthening the archive management work, reasonably forming the student management team, and improving the information security work are proposed.

2. Method

2.1. Student behavior analysis based on K-means clustering algorithm

2.1.1. K-means clustering algorithm. There are five major types of clustering algorithms, namely hierarchical clustering, density-based clustering, grid-based clustering, model-based clustering, and division-based clustering.The K-means clustering algorithm is also known as the K-means algorithm. It is the classical clustering algorithm commonly used in cluster analysis at present. It mainly divides the data objects in the dataset through an iterative process, using the distance between the data objects as a similarity measure: it makes the similarity between data objects in clusters composed of data objects of the same category higher and the difference lower, and the difference between different clusters higher and the similarity lower [21]. After division each data element belongs to only one cluster and it is guaranteed that each cluster contains one or more data elements. The mean value of the objects in each cluster is taken to represent the center of mass (cluster center) of the cluster. Campus student behavior data such as student meals, consumption, and outings are of high quality and are more suitable for cluster analysis using this algorithm.

The main idea is that given a dataset $M=\{m_1,m_2,m_3\cdots m_n\}(m_i\in R^d)$ containing n d-dimensional data, the number of clusters to be generated is K, and each cluster can be represented as $S=S_i,i=1,2,3...K$ each cluster S_i has a cluster center u_i, and the Euclidean distance is used as a criterion to judge the formula (1) as follows:

$$V=\sum_{i=1}^{k}\sum_{x_j\in s_i}(x_j-u_i)^2 \tag{1}$$

K-means algorithm minimizes the value of V in the above equation as the final goal through an iterative process.

K-means clustering algorithm can be used in learning behavior analysis to classify students into different categories and suggest appropriate teaching suggestions according to the characteristics of different categories of students.

2.1.2. Optimization of clustering center selection. This subsection optimizes the selection of K-means clustering centers, mainly based on the density of the clustering is improved to improve the efficiency of selecting the clustering centers, here two main work is done, respectively, the center optimization selection and the calculation of the radius [22]. Simply put, it is to select high-density points as the clustering center, so that the density within the calculated radius can be maximized. The meaning and description of the symbols involved are shown below:

X_{ij} - The set dataset $X=\{x_1,x_2,\ldots,x_n\}$.

x_i - contains p features $X=\{x_{i1},x_{i2},\ldots,x_{ip}\}$.

C_j - Divided cluster classes, where $j=k$.

k_i - $i=1,2,\ldots,n$.

In order to solve the difficulty in defining density points due to the large amount of student behavior data in colleges and universities, Euclidean distance is introduced for the metric. The Euclidean distance between any two data objects x_i and x_j is defined as shown in equation (2):

$$d(x_i,x_j)=\sqrt{(x_{i1}-x_{j1})^2+(x_{i2}-x_{j2})^2+\ldots+(x_{ip}-x_{jp})^2} \tag{2}$$

The Euclidean distance density function density of data object x_i in the dataset of this paper is defined as shown in equation (3):

$$density(x_i)=\sum_{j=1}^{n}(d(x_i,x_j)\sum_{l=1}^{n}d(x_l,x_j)^{-1}) \tag{3}$$

Where: the domain radius of the data object $x_i(1\le i\le n)$ in the data set is R, as shown in equation (4), and it is verified that the best effect can be obtained when the adjustment factor is $r=0.16$:

$$R_i=n^r\times\sum_{i=1}^{n}e^{-density(x_i)}\times n^{-1} \tag{4}$$

From Eq. (2), Eq. (3) and Eq. (4), when x_i is the center and R is the radius, then the density of data points in the circle at this time, defined as $Y=(x_i)$, leads to Eq. (5):

$$Y(x_i)=|\{p\,|\,d(x_1,p)\le R_i,p\in X\}| \tag{5}$$

From this, the average density of data points within the dataset can be introduced, as shown in equation (6):

$$PY(x)=\frac{1}{n}\sum_{x\in X}Y(x) \tag{6}$$

This subsection improves the traditional K-means algorithm on the algorithm for selecting cluster centers by changing the traditional pattern of randomly selecting cluster centers, using the Euclidean distance between data objects to measure the similarity, and using the density to optimize the selection of cluster centers.

2.2. Achievement correlation analysis based on Apriori algorithm

2.2.1. Apriori algorithm. In order to study out the connection between the behavioral characteristics of college students and their grades, the Apriori algorithm is used to mine out the association rules between

the behavioral characteristics of college students and their grades, which provides decision-making support for the relevant managers of colleges and universities.

Apriori algorithm is the classic algorithm that searches and filters through layers in order to generate frequent itemsets, and then mines the association rules between the data [23]. It adopts an iterative form of layer-by-layer search, using L_{k-1} frequent itemsets to determine C_k candidate itemsets until all frequent itemsets are found. Firstly, the thresholds of support and confidence are formulated, the support of candidate 1-itemset C_1 is calculated, and compared with the predefined support threshold, frequent 1-itemset L_1 is found, then candidate 2-itemset C_2 is searched according to L_1-frequent itemsets, and frequent 2-itemset L_2 is found according to the support, according to the recursive method, the candidate C_k-itemsets and frequent itemsets L_k are obtained, and then, the confidence of the frequent itemsets is calculated and compared with the confidence threshold, if frequent itemsets are found, the confidence threshold of frequent itemsets is calculated and the confidence threshold of frequent itemsets is compared, if frequent itemsets are not found, the confidence threshold of frequent itemsets is calculated, the confidence threshold of frequent itemsets is compared. Threshold, and if the confidence of the frequent itemset is less than the threshold, it is cut off. Finally, strong association rules are generated from the searched frequent sets, which are required to be larger than the support and confidence thresholds, and the candidate rules smaller than the thresholds are deleted to obtain the final association rules.

The specific flow of traditional Apriori algorithm is shown in Figure 1.

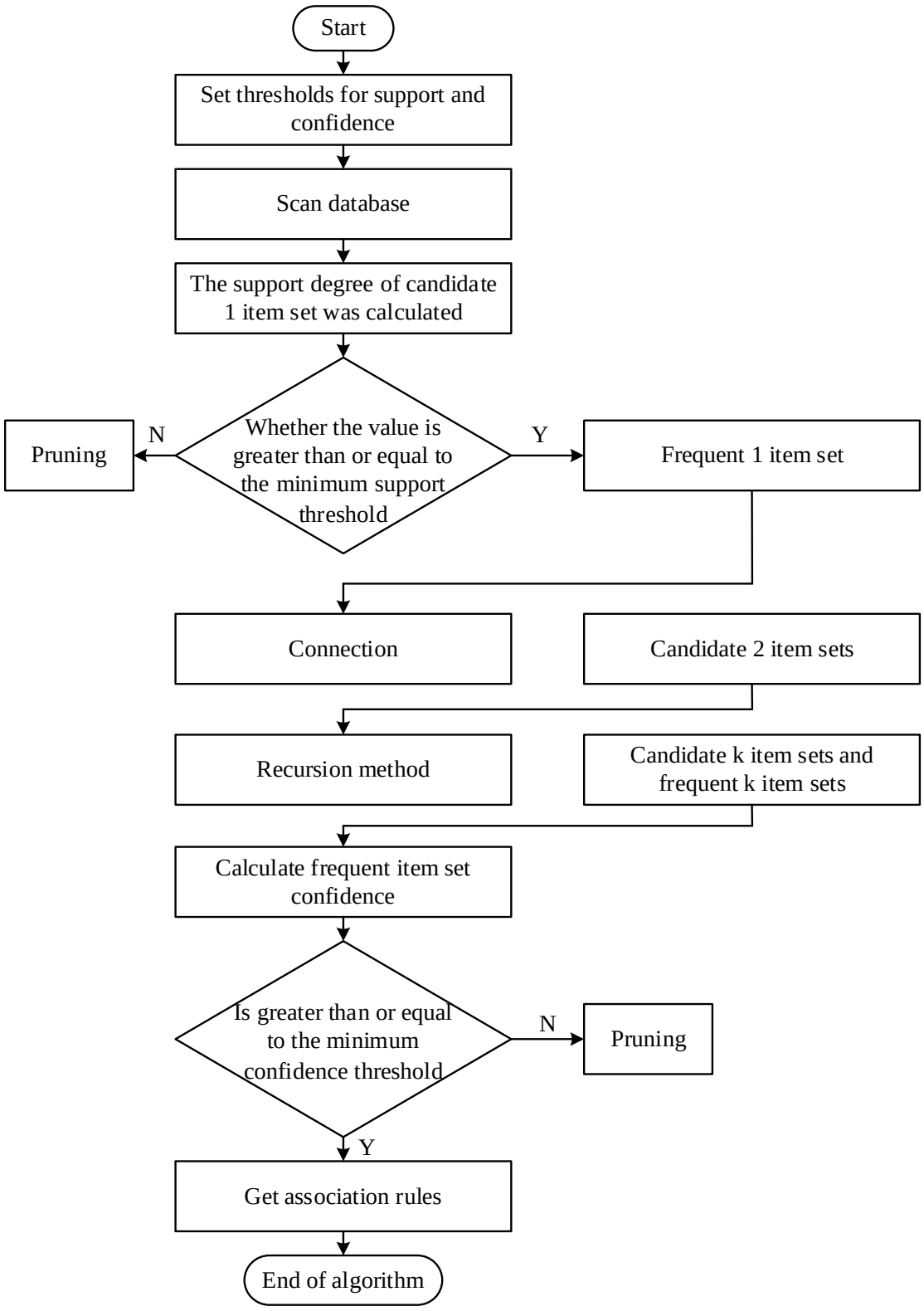

Fig. 1. Flow of traditional Apiori algorithm

The implementation process of Apriori algorithm mainly consists of connection step and pruning step, which in turn generates to get the association rules.

1) Connection step. In traditional Apriori, the self-join step is the process of generating candidate k itemsets C_k by self-joining the frequent itemsets L_{k-1}. In order to obtain the frequent k-item set L_k, it is necessary to obtain the candidate C_k-item set, in which all the connectable items of the frequent $k-1$-item set are selected and connected to facilitate the acquisition of the C_k. In Apriori algorithm, it is assumed that all the items in the itemset are arranged according to the dictionary order, and only the last item in the connectable itemset, i.e., the two frequent itemsets, is different, in other words, the first $k-2$

items of the two frequent itemsets in L_{k-1} are the same. Suppose I_1, I_2 is the two items in L_{k-1}, $I_1[k-1]$ is the $k-1$ th item of I_i, if $(I_1[1]=I_2[1] \wedge I_1[2]=I_2[2] \wedge \cdots \wedge I_1[k-1]=I_2[k-1])$, and $(I_1[k-1]=I_2[k-1])$, then I_1, I_2 can be connected, I_1, I_2 connected to generate the item set for $I_1[1], I_1[2], I_2[1], I_2[2], \ldots, I_1[k-1], I_2[k-1]$, which can be obtained by combining the two items of the frequent $k-1$ item set, if they can generate k items, then it can be generated as a candidate k item set C_k.

2) Pruning step. The pruning step reduces the number of itemsets in a candidate itemset, due to the Apriori based a priori, any infrequent $k-1$ itemset cannot be a subset of a frequent k itemset, therefore, if there exists a $k-1$ subset of itemsets in a candidate k itemset that are infrequent, the candidate itemset itself is infrequent, and it needs to be pruned out of C_k it. When scanning the transaction dataset, the number of items in each candidate set C_k of the initially obtained k candidate set is computed to obtain the frequent k itemset, and the computed number is compared with the pre-set support threshold, if the computed number of candidate sets is greater than or equal to the support threshold, the set is called a frequent itemset, and if the computed number of subsets of any of the candidate sets is not frequent, then this k candidate itemset is infrequent and it is deleted from the candidate itemset C_k to reduce the amount of computation to a certain extent.

3) Generate association rules based on frequent itemsets. After searching the frequent itemsets according to the thing database, we can directly utilize them to generate strong association rules. Firstly, the minimum support threshold is set, then the non-empty subset P of the frequent itemset L is generated, and the complementary set Q of P is generated, then P and Q are regarded as the front and back items of the association rule, and then the confidence of this rule is calculated, and if the confidence of this association rule exceeds the minimum support threshold, it is called a strong association rule, and additionally, if the enhancement of this rule is greater than 1, then it is a valid association rule, and only then it can be put into the practice as valuable guidance.

2.2.2. Apriori algorithm optimization. During the Apriori algorithm, there are three definitions:
1) Assume that $T=\{T_1, T_2, \ldots T_m\}(m \geq 1)$ is a set of transactions, $T_i=\{I_1, I_2, \ldots I_n\}(n \geq 1)$ is a set of items, and $k-$ frequent itemsets $=\{I_1, I_2, \ldots I_k\}(k \geq 1)$ is a set of k items.
2) Assume that A (itemset) is the support count of the itemset or the frequency of occurrence of the itemset in the transaction.
3) Assume that C_k is a candidate set of size k and L_k is a frequent set of size k.
Improvement Steps:
1) First scan all the transactions in the database to generate L_1 which contains the items, the support of the items and the transaction D in which all the items are located.
2) Use L_1 as an auxiliary to generate $L_2, L_3, \ldots, L_k$, and obtain C_2 by concatenation construction.
3) Before scanning all transaction records to compute the support of each candidate, use L_1 to obtain the transaction ID with the smallest support between x and y to achieve scanning only C_2 in a particular transaction.
4) In much the same way as constructing C_2, a third itemset $C_3(x, y, z)$ is constructed where x, y and z are the transactions of C_3 and the minimum support between the transaction IDs of L_1 is used for the target transaction labeling, and then C_3 is scanned in the particular transaction until no new frequent itemsets are generated and the execution of this algorithm ends.

3. Results and discussion

The study takes undergraduate students of Z University as the research object, selects the behavioral data of undergraduate students of four grades for one academic year, conducts cluster analysis on students' consumption behavior, work and rest behavior, and study behavior, and carries out correlation analysis on

the basis of the clustering results and the students' academic performance. Among them, the consumption behavior data mainly comes from the one-card system. The data of work and rest behavior mainly come from the Sunshine Attendance System, the One-Card System and the Internet Authentication System. Learning behavior data mainly come from the academic attendance system, the book lending system and the Internet authentication system. The raw data extracted include 8566393 consumption flows of One Card, 753906 sunshine attendance flows, and 930145 book lending flows. Due to the redundancy and structural inconsistency in the data, the student samples with serious missing information and some outliers were excluded after data processing, and finally 12,885 student sample data were obtained.

3.1. Cluster analysis of student behavior using K-means algorithm

3.1.1. Cluster analysis of student consumption behavior. K-means algorithm cluster analysis was carried out on students' consumption data. According to the evaluation criteria of the clustering algorithm, the best clustering effect is obtained when the number of clusters is set to 5. The clustered student classification results were analyzed, according to the actual situation of student consumption. Compare the average value of each cluster with the average value of the students' overall index, and record H as the average value higher than the student's overall index, and L as the average value lower than the student's overall index (the same below). Table 1 shows the results of cluster analysis of students' consumption behavior in various clusters. The characteristics of students' consumption behavior are divided into five categories. The first group of students had the lowest monthly consumption level (468.16 yuan) and the lowest peak monthly consumption (534.3 yuan), but they consumed frequently (104.17 times), which belonged to the group with a low consumption level. The second group of students had a medium monthly consumption level (664.41 yuan), but a high monthly consumption peak (1106.2 yuan) and a small number of consumption times (68.37 times). The third group of students has a medium to high monthly consumption level (723.85 yuan), a large number of monthly consumption times (84.28 times), and a peak monthly consumption of 876.28 yuan. The fourth group of students had the highest monthly consumption level (891.97 yuan), frequent monthly consumption (97.38 times), and the highest peak monthly consumption (1151.63 yuan). The fifth group of students had a lower monthly consumption level (495.23 yuan), the least number of consumption times (39.7 times), and a lower peak monthly consumption (625.29 yuan).

Table 1. Results of the student consumption behavior clustering analysis

Categories	Student ratio (%)	Monthly consumption (Yuan)	Monthly consumption frequency (Times)	Monthly consumption peak (Yuan)	Category label
1	10.5	468.16	104.17	534.3	LHL
2	17.66	664.41	68.37	1106.2	LLH
3	33.82	723.85	84.28	876.28	HHL
4	24.64	891.97	97.38	1151.63	HHH
5	13.38	495.23	39.7	625.29	LLL
Mean value	-	693.12	81.38	896.32	-

3.1.2. Cluster Analysis of Students' Rest and Recreation Behavior. The K-means algorithm was used to cluster and analyze the students' regular eating habits, Internet habits, early rising habits, and exercise habits in the indicators of work and rest behavior. According to the cluster average criterion, when the number of clusters is 3, the percentage of students in each type of cluster and the average value of indicators are shown in Table 2. Students in the first category often woke up early every month (20.1 times), ate in the school cafeteria more regularly (24.1 times), spent less time on the Internet (174.32 hours), and often participated in physical exercise (53.4 times). Students in the second category stayed in bed more often (6.4 times), ate irregularly in the school cafeteria (9.2 times), spent the most time on the Internet (211.36 hours), and participated in physical activity very little (6.6 times) per month. Students in the third category often

woke up early each month (18.1 times), but went to the school cafeteria for meals irregularly (13.2 times), spent the longest time on the Internet (205.36 hours), and participated in physical exercise less often (17.9 times).

Table 2. Student routine behavior cluster analysis results

Categories	Student ratio (%)	Regular diet (Times)	Get up early (Times)	Internet length (Hours)	Physical activity (Times)	Category label
1	36.33	24.1	20.1	174.32	53.4	HHLH
2	10.39	9.2	6.4	211.36	6.6	LLHL
3	53.28	13.2	18.1	205.36	17.9	LHHL
Mean value	-	16.3	18.7	199.65	29.6	-

3.1.3. Cluster analysis of student learning behavior. K-means algorithm clustering analysis was performed on four indicators of students' study behavior: class attendance, book borrowing, number of visits to the library, and study hours. The percentage of students in each category of clusters and the mean values of the indicators are shown in Table 3. Students in the first category had the highest class attendance rate (95.36%), borrowed fewer books (7.4), went to the library the most times (28.9 times), and studied the longest (154.9 hours). Students in the second category had the lowest class attendance (86.14%), borrowed the fewest books (3.2), visited the library the fewest times (11.7), and studied the shortest amount of time (85.1 hours). Students in the third category had a higher class attendance rate (92.34%), borrowed the most books (15.1), visited the library more often (27.1), and spent more time studying (133.7 hours). Students in category 4 had average class attendance (90.07%), borrowed fewer books (8.6), made an average number of trips to the library (15.8), and spent less time studying (117.7 hours).

Table 3. Student learning behavior clustering analysis results

Categories	Student ratio (%)	Attendance rate (%)	Book borrowing (Copy)	Library number (Times)	Learning length (Hours)	Category label
1	31.05	95.36	7.4	28.9	154.9	HLHH
2	11.46	86.14	3.2	11.7	85.1	LLLL
3	12.69	92.34	15.1	27.1	133.7	HHHH
4	44.8	90.07	8.6	15.8	117.7	LLLH
Mean value	-	91.67	8.7	22.7	125.9	-

3.2. Analysis of the correlation between students' behavior and achievement

The above clustering analysis has classified students' school behaviors into three categories: consumption behaviors, work and rest behaviors, and study behaviors. In order to further study the relationship between students' behavioral characteristics and academic performance, and to explore whether there is an inevitable connection between college students' behavioral characteristics and their academic performance, the Apriori algorithm is selected for the correlation analysis, and the correlation and its law are mined from the big data.

According to the results of clustering processing and analysis of student behavior data, different types of student groups and category characteristics are summarized, providing data output for the association rule algorithm. The total collation of different types of labels and category labels is summarized in Table 4. Students' consumption behavior is divided into low consumption type, frugal type, moderate type, high consumption type, and campus low consumption type. Students' work and rest behavior is divided into

regular, irregular, and less regular. Students' study behavior is divided into more effortful type, not effortful type, effortful type, and not effortful enough type.

Table 4. Student behavior index label

Behavior category	Categories	Category label	Type label
Student consumption behavior	1	LHL	Low extinction
	2	LLH	Thrifty
	3	HHL	Moderate type
	4	HHH	High extinction
	5	LLL	Campus low extinction
Student routine	1	HHLH	Regular type
	2	LLHL	Irregular type
	3	LHHL	Less regular type
Student learning behavior	1	HLHH	Harder type
	2	LLLL	Unworking type
	3	HHHH	Hard type
	4	LLLH	Hard enough type

The main purpose of a student's schooling is to learn, and grades are an important indicator of learning outcomes. The overall grade of a single subject reflects the learning situation of the course, while the comprehensive GPA reflects the overall learning situation of all subjects, the comprehensive GPA is based on the credits taken in each course in accordance with the weighted average grade, the comprehensive GPA is defined to be in the range of 0-5, which is divided into different grades, 4-5 is defined as excellent, 3-4 is defined as good, 2-3 is defined as moderate, 1-2 is defined as passing, and 0-1 is defined as failing. The student behavioral data and achievement data were combined and processed, and the collated results are shown in Table 5.

Table 5. Behavior data and achievement data of students

Student number	Student consumption behavior	Student routine	Student learning behavior	Composite point
001	High extinction	Less regular type	Hard enough type	Medium
002	Moderate type	Less regular type	Hard type	Good
003	Thrifty	Regular type	Harder type	Good
004	High extinction	Irregular type	Unworking type	Medium
005	Campus low extinction	Irregular type	Hard enough type	Medium
006	Moderate type	Regular type	Hard type	Excellent
007	High extinction	Irregular type	Unworking type	Poor
008	Low extinction	Regular type	Hard enough type	Good
009	Campus low extinction	Regular type	Hard type	Excellent
010	Moderate type	Irregular type	Hard enough type	Medium

The collated student behavioral data and achievement data are mined through the improved Apriori algorithm to explore the correlation between students' behaviors and achievements, setting the minimum support threshold at 0.2 and the minimum confidence threshold at 0.4, and digging out 10 correlation rules to obtain the results of mining the correlation rules between students' behaviors and achievements, as shown in Table 6. Taking the first rule as an example, 44.9% of the student group is characterized by waking up early, eating regularly, spending less time on the Internet, and exercising regularly, and 85.6% of this group of students are likely to have "excellent" academic performance. It can be found that students with good grades live relatively regular lives, have the habit of waking up early, surf the Internet appropriately, eat regularly in the cafeteria, have a higher rate of attendance in class, spend a lot of time studying in the library, and check out the most books. The 2nd rule indicates that students with good grades are underperforming in some aspects compared to students with good grades, but have a relatively high level of study effort to ensure that they are not left behind in their academic performance. There are 5 rules, numbered 3 to 7, which are filtered out as "medium", and it can be seen from these 5 rules that the main reason for the poor performance is the lack of effort in studying, and the secondary reason is the irregularity of life, and the consumption behavior does not have much influence on the performance. These students should pay more attention to developing good study habits and should learn from their classmates with good grades. There is no postulate rule for poor grades, but there are antecedent rules for poor grades, such as Rule No. 8: poor => not trying hard enough type, Rule No. 9: poor => irregular type. These rules indicate that students with poor grades generally have chaotic lives and do not work hard in their studies. Also according to the clustering results, it can be seen that students with irregular life tend to have higher expenses.

Table 6. Mining results of student behavior and grade association rules

Numbering	Rule	Support	Confidence
1	Regular type, Hard type=>Excellent	0.449	0.856
2	Regular type, Harder type=>Good	0.508	0.731
3	Regular type, Hard enough type=>Medium	0.353	0.762
4	Less regular type, Hard enough type=>Medium	0.429	0.913
5	Moderate type, Less regular type, Hard enough type=>Medium	0.237	0.775
6	Campus low extinction, Hard enough type=>Medium	0.234	0.77
7	Thrifty, Regular type, Hard enough type=>Medium	0.429	0.808
8	Poor=>Unworking type	0.527	0.917
9	Poor=>Irregular type	0.557	0.904
10	Poor=>Unworking type, Irregular type	0.536	0.889

3.3. Management recommendations

Based on the above analysis, different management measures should be adopted for different types of students.

For students with poor economic conditions but good academic performance and good behavioral habits, we can focus on the living conditions of such students, and we can also use this data as an auxiliary support for the assessment of financial assistance for poor students.

For students with poor academic performance, counselors and other administrators should pay more attention to their school life status, whether they are under too much pressure and have negative and depraved emotions, and timely understanding and help to regulate.

For individual students who fail multiple courses without special circumstances, they will be given academic warnings and other treatments, and for students with other abnormalities, they will be given special treatments.

3.4. Effective Paths to Strengthen Student Management in Colleges and Universities

3.4.1. Reasonably optimize the awareness of student management in colleges and universities. When student management personnel in colleges and universities manage students, they must scientifically establish the correct management consciousness, innovate the student management mode, effectively improve the working mechanism of student management in colleges and universities, and guarantee the main position of students in learning. Through the cluster analysis of student behavior, timely grasp of student learning and life dynamics, based on the characteristics of different students for scientific management.

3.4.2. Scientific construction of information platforms. Institutions of higher education should rationally integrate enterprise management mode on the management information platform, scientifically cultivate modern management talents, grasp the specific situation of student management and the actual performance of students in school through big data, scientifically adjust the relevant strategies, and provide students with more effective guidance, so that college students can adapt to the needs of modern enterprise development to a higher degree, and further improve the comprehensive level and personal quality of students in order to The students' comprehensive level and personal quality can be further improved, so as to carry out more effective management of college students.

3.4.3. Strengthening of archive management. Colleges and universities should effectively utilize and scientifically process the relevant data in the student files to ensure that the staff related to file management can identify the student data more scientifically. The archive management department of higher education institutions should make full use of the archive management system, scientifically improve the original

archive system, ensure that there is a close connection between the archive information and the students, to a higher degree to ensure the accuracy of the archive information, more efficient and scientific management of the student archives, to further strengthen the management of students in colleges and universities, and to promote the development of the management of student archives.

3.4.4. Rationalization of the student management team. Institutions of higher education must scientifically build a student management team in the era of big data. It is necessary to further improve the informatization level of student management personnel in colleges and universities, and at the same time, strengthen the informatization construction capacity of the management team. K-means clustering algorithm and association rules can be used to understand the specific situation of students to adjust the student management mode to ensure more efficient management of students. The scientific application of student management mode can not only ensure that institutions of higher education can carry out student management work more conveniently, but also realize the effective operation of student management mode, so that student managers in institutions of higher education can carry out student management work more scientifically.

3.4.5. Improved information security. In the application of information technology, information security is of great importance, and it is an indispensable part of the scientific and modernized transformation of student management in schools. Colleges and universities should build a scientific security management system for the information management platform to ensure that modern science and technology can be applied more safely in the process of student management. According to the specific needs of student management, the student management department of colleges and universities should set up an information security management team, scientifically design and develop the relevant management system, and invest a certain amount of money to build up a scientific and safe management system, to ensure that the management system is strict and perfect, and to guarantee the efficient and orderly operation of the informationization platform.

4. Conclusion

The study uses artificial intelligence technology to promote the level of student management in colleges and universities, and takes 12,885 students in college A as the research object, and draws the following conclusions from the clustering analysis of students' behavioral data, and the correlation analysis of behavioral data and achievement data:

Students' consumption behavior is divided into low consumption type, thrifty type, moderate type, high consumption type, and campus low consumption type through clustering analysis. Students' work and rest behaviors were divided into regular, irregular, and less regular through cluster analysis. Students' study behaviors were classified as more effortful, less effortful, effortful, and less effortful through cluster analysis.

The clustering results were correlated with academic performance, and a total of 10 correlation rules were found. Colleges and universities adopt different management measures for different categories of students to improve management efficiency.

The study proposes to optimize the awareness of student management work in colleges and universities, scientifically construct the information platform, strengthen the file management work, reasonably set up the student management team, and improve the information security work to strengthen the student management work in colleges and universities.

References

[1] Jiayu, Y. (2023). Challenges and Opportunities of Generative Artificial Intelligence in Higher Education Student Educational Management. Advances in Educational Technology and Psychology, 7(9), 92-96.

[2] Qiu, Z., & Han, J. (2022). Artificial Intelligence of Internet of Things Based on Machine Learning and College

Student Management. Mobile Information Systems, 2022(1), 8620277.

[3] Zheng, G. (2024). Research on the Integration of Institutional Management and Emotional Management in College Student Management in the Era of Artificial Intelligence. Adult and Higher Education, 6(1), 34-43.

[4] Li, Y., & Zhu, H. (2022, July). Student Management Online Education Platform Based on Artificial Intelligence Technology. In International Conference on Frontier Computing (pp. 1672-1679). Singapore: Springer Nature Singapore.

[5] Zhang, T. (2022, September). Student Management Information Security System Based on Artificial Intelligence and Cloud Computing. In International Conference on Cognitive based Information Processing and Applications (pp. 233-242). Singapore: Springer Nature Singapore.

[6] Jeonghyun, Y. (2023). An Analysis of University Students' Needs for learning support functions of learning management system augmented with artificial intelligence technology. KSII Transactions on Internet and Information Systems (TIIS), 17(1), 1-15.

[7] Colchester, K., Hagras, H., Alghazzawi, D., & Aldabbagh, G. (2017). A survey of artificial intelligence techniques employed for adaptive educational systems within e-learning platforms. Journal of Artificial Intelligence and Soft Computing Research, 7(1), 47-64.

[8] Guan, Y. (2022). Student education management strategy based on artificial intelligence information model under the support of 5G wireless network. Computational intelligence and neuroscience, 2022(1), 4709146.

[9] Xin, L., Wang, F., Sun, X., Ding, G., & Wang, H. (2022). Motivation of student management subjects in universities based on network intelligence. Mobile Information Systems, 2022(1), 5623067.

[10] Khan, I., Ahmad, A. R., Jabeur, N., & Mahdi, M. N. (2021). An artificial intelligence approach to monitor student performance and devise preventive measures. Smart Learning Environments, 8, 1-18.

[11] Lv, J. (2024, July). Construction of Intelligent College Student Management Evaluation System Based on Machine Learning Algorithm. In 2024 3rd International Conference on Artificial Intelligence and Autonomous Robot Systems (AIARS) (pp. 488-493). IEEE.

[12] Jokhan, A., Chand, A. A., Singh, V., & Mamun, K. A. (2022). Increased digital resource consumption in higher educational institutions and the artificial intelligence role in informing decisions related to student performance. Sustainability, 14(4), 2377.

[13] Gajra, V., Lakdawala, K., Bhanushali, R., & Patil, S. (2020, April). Automating student management system using ChatBot and RPA technology. In Proceedings of the 3rd International Conference on Advances in Science & Technology (ICAST).

[14] Chen, W., Samuel, R., & Krishnamoorthy, S. (2021). Computer Vision for Dynamic Student Data Management in Higher Education Platform. Journal of Multiple-Valued Logic & Soft Computing, 36.

[15] Wang, Y., & Sun, Y. (2023, October). Construction of an Artificial Intelligence Student Management Information System Under Big Data Analysis Algorithms. In 2023 2nd International Conference on Artificial Intelligence and Intelligent Information Processing (AIIIP) (pp. 61-65). IEEE.

[16] Li, Z. (2023). Research on the innovation of Intelligent Student management Assessment Information System based on python software design. Advances in Education, Humanities and Social Science Research, 5(1), 75-75.

[17] Abdulhadi, N. M., Ibraheem, N. A., & Hasan, M. M. (2022). Blockchain-Based Framework for Secure and Reliable Student Information Management System Using Artificial Intelligence. In Mobile Computing and Sustainable Informatics: Proceedings of ICMCSI 2022 (pp. 753-762). Singapore: Springer Nature Singapore.

[18] Sufyan Ghaleb, M. M., & Alshiha, A. A. (2023). Empowering Self-Management: Unveiling the Impact of Artificial Intelligence in Learning on Student Self-Efficacy and Self-Monitoring. Eurasian Journal of

Educational Research (EJER), (107).

[19] Sajja, R., Sermet, Y., Cikmaz, M., Cwiertny, D., & Demir, I. (2024). Artificial intelligence-enabled intelligent assistant for personalized and adaptive learning in higher education. Information, 15(10), 596.

[20] Liu, X., & Teng, Q. (2024, July). Design of Intelligent Integrated Student Management Terminal Based on K-Means and ANN Algorithm. In 2024 3rd International Conference on Artificial Intelligence and Autonomous Robot Systems (AIARS) (pp. 934-938). IEEE.

[21] Chang Wenbing,Ji Xinpeng,Liu Yinglai,Xiao Yiyong,Chen Bang,Liu Houxiang & Zhou Shenghan. (2020). Analysis of University Students' Behavior Based on a Fusion K-Means Clustering Algorithm. APPLIED SCIENCES-BASEL(18),6566-.

[22] Wu D. & Ren J.D. (2014). K-means sequence clustering center selection algorithm based on sequence neighborhoods optimization. ICIC Express Letters(8),2245-2250.

[23] Tiantian Li,Fang Liu,Xiaobin Chen & Chao Ma. (2024). Web log mining techniques to optimize Apriori association rule algorithm in sports data information management. Scientific Reports(1),24099-24099.