THE TRAJECTORY DATA MINING MODEL FOR COLLEGE STUDENTS IN COMPUS LIFE AND ACADEMIC MANAGEMENT

WUGANG LIU*

Abstract. The main objective of the study is to address the lack of comprehensive management technology in student campus life in universities. Starting from the life trajectory data of students in campus life, a trajectory mining model combining data mining technology and university information system is designed. In addition, an applied clustering algorithm is designed to classify different trajectory feature types. The research results show that in actual trajectory analysis, the categories of action trajectories from dormitories to canteens, from 0 to 4, are 87.64%, 87.86%, 86.97%, 88.63%, and 88.71%, respectively, which are the most matched effective action trajectories. It can be seen that the trajectory analysis model designed in the study is effective and can provide assistance for the comprehensive academic management of college students.

Key words: Data mining, Clustering, Academic management, Trajectory features

1. Introduction. With the increasing attention of the state to social talent cultivation in recent years, college students, an important source of national talent reserve, have gradually received extensive attention from all sectors of society. The traditional talent cultivation system lacks adaptability to personalized and practical talent cultivation, and it is difficult to meet the talent demand of today's society, and it has become a new way of talent cultivation to update the talent cultivation system by using the current wave of social informationization and data development [1-3]. The new practical talent cultivation system not only involves the application of information technology in college talent cultivation, but also involves the definition of modern talent cultivation in college. In the traditional education concept, academic achievement is the most important evaluation index for college students and the ultimate value of students' learning career. However, with the gradual convergence of university talent education and social talent demand, academic performance can no longer form a more comprehensive assessment of students. Students will also develop various campus learning activities such as school-enterprise joint practice, campus activities, part-time entrepreneurship, campus exchanges, etc. Meanwhile, students' learning habits in the process of efficient learning have also become one of the important factors to assess students' comprehensive quality [4-6]. It is difficult to assess the daily learning life of students because they are influenced by social networks and learning life landscape, and they are characterized by both group and diversity. Data mining technology provides a technical grip for this problem. By integrating data mining technology with the information record system of college students, it can track and manage students' campus life and study in a trajectory way, and then achieve adaptive management and efficient management [7-9]. Therefore, this study designs a trajectory mining model combining data mining technology and college information system from the perspective of students' academic trajectory, and achieves academic tracking and analysis by analyzing students' trajectories.

The innovation of this study is to extend data-driven student management from learning management to comprehensive management of campus life, and apply trajectory mining models to information systems. By comprehensively analyzing the life trajectory of students, targeted management is implemented.

2. Related Works. Lee S M's team conducted a follow-up study on the adjustment of dental hygiene students to campus life and proposed appropriate management strategies. The study analyzed students' adjustment to campus activities in terms of their club participation, personality, professional adaptation, and interpersonal relationships. The results of the study showed that the campus life management strategy developed by the study can effectively improve the students' adaptability to campus life [10]. Li W's team developed

---

*School of Arts and Science, Nanning College of Technology, Nanning, 530100, China (`wuganglwg@163.com`)

an intelligent campus management system based on IoT technology from the perspective of smart campus, which uses IoT face recognition technology as long as the data collector and realizes the tracking of students' campus life trajectory and campus life through standardized data analysis. management. The results of the study showed that the system is practical [11]. Kim Y's team analyzed the satisfaction of college students with campus life and analyzed the mediating role between students' social network consistency and satisfaction in campus life based on the survey data. The results of the study showed that the satisfaction of college students with campus life must be reflected in their self-efficacy through participation in campus social life [12]. Purnama S's team proposed a support system for college students' digital weaknesses in the learning process, which is based on the perspective of students' learning life and combines electronic devices and student-centered blockchain to enhance the digital capabilities of cooperative education while meeting the needs of university students' learning lives [13]. Way conducted a study on the important factors influencing students' learning behaviors in their daily learning lives in higher education, which is a combination of qualitative and quantitative analysis from the perspective of students' daily learning lives, emotional content, and temporal dimensions. The results of the study showed that online teaching and learning can fully complement offline teaching and learning and contribute to student outcomes [14].

In the development of data mining technology, its specific application in various fields is its main development situation. Haoxiang team applied data mining technology to online privacy data protection and used a perturbation algorithm to solve similar problems. There is also a significant improvement in the efficiency of the model, compared to other privacy-preserving algorithms [15]. Kuma applied data mining to finance and marketing and designed a data mining-based decision system for financial market information. This system analyzed organizational performance from a practical point of view and determined how the decision solution can be used to help companies balance competitive pressures under external environmental factors such as tax pressure and industrial costs. The results of the study showed the feasibility of this solution [16]. Edastama P's team proposed a data mining tool-based student data analysis system, which is a comprehensive data warehouse in the form of web information reports, and mined the characteristics and patterns of student data through basic data to finally achieve the effect of assessing the status of students' academic and campus life [1,7]. Ageed addressed the issue of combining data mining technology with cloud computing notation. The results of the study showed that the technique designed in the study effectively solves the cloud compatibility problems that arise when data mining is applied in parallel with cloud computing [18]. Mengash designed a data mining model for predicting the performance of college applicants in colleges and universities. Data mining model, which is combined with a reliable standardized admissions system, enables the prediction of possible post-admission learning outcomes of cohort students before they are admitted to colleges and universities. Over two thousand students were selected as the dataset to validate the model proposed in the study, and the predictive accuracy of the model was analyzed by tracking the actual academic performance of students after admission. The results of the study showed that the model designed in the study is able to predict the academic performance of students after enrollment, and such performance prediction can be used as a basis for student admission judgment]1,9].

Garg et al. proposed a decentralized evaluation system to address the issue of tampering in online education evaluation systems, to ensure the integrity of online education evaluation and further achieve the review of online education content. The research results showed that the system is practical [20]. Dutt et al. applied fuzzy set technology to learning neural network classification technology and proposed a digital learning assistance system for people with learning disabilities, achieving intelligent and personalized learning guidance. The research results showed that this method has a more efficient guidance function [21]. Choudhary et al. applied deep learning algorithms to personalized learning recommendation systems based on the extraction and analysis of user preference information, thereby improving the accuracy and personalization of recommendations for different users. The results showed that the system can effectively improve recommendation accuracy and personalization level [22].

It can be seen that data mining technology has unique advantages in the analysis of group data, and it can be better integrated with other systems. There have been some research examples of data mining technology applied in the university system so far. However, it can be found that the current data analysis of students in universities mainly focuses on the analysis of students' academic performance and learning status, but not

from the perspective of students' comprehensive campus life, which is relatively one-sided. Therefore, this study starts from students' life trajectory data in campus life, combines data mining technology with university information system, and designs a trajectory mining model to provide new ideas for students' data analysis.

This study designs a trajectory mining model that can manage students' comprehensive campus life by combining campus life trajectory data and big data mining strategies. The model collects information using different information endpoints of integrated information systems and processes trajectory characteristics based on this. This is a novel perspective and effective information tool for managing students' campus life. This study further designs an applied clustering algorithm and uses it to classify different trajectory feature types. The application of this method greatly enhances the accuracy and effectiveness of research and implementation of management strategies. This study can predict students whose behavior patterns may change through models, and predict the direction of changes, which has important practical significance for early warning and prevention of student behavior problems. Meanwhile, the applied clustering algorithm in this study is better at clustering data into groups with features. Compared to other clustering algorithms, it processes and classifies data more meticulously and accurately. Overall, this study provides a new perspective and effective means to assist universities in comprehensively managing students' academic lives by designing and implementing a model that excavates their daily academic life trajectories, combined with big data mining technology and existing university information systems.

3. Data Mining Model Design for College Students' Trajectories. These feature points include campus consumption records, campus network usage records, campus access control records, scholarship information, and student information. Through these data, it can roughly depict the campus image of students. When analyzing the characteristics of students' campus life trajectory, it needs to label the campus and its internal functional areas, as universities may have multiple campuses and multiple functional areas with similar functions within the same campus. Therefore, it labels the campus and differentiates its functions into twelve different categories. On this basis, the semantic trajectories of students will be further segmented, and the segmentation operation can appropriately divide the long-term trajectories of students. The main trajectory segmentation method used is time-threshold trajectory segmentation. After extracting the required student activity trajectory feature information, an academic management model based on clustering algorithm is proposed. The model is mainly divided into three modules, namely data clustering analysis, trajectory frequent pattern analysis, and trajectory deviation analysis. The model first uses the k-means algorithm for clustering analysis, and then uses the PrefixSpan strategy to perform frequent pattern analysis on the student trajectories within the cluster based on the clustering results. Finally, based on the output results of frequent trajectory patterns, it calculates the degree of deviation between the trajectories of individual students within the cluster and the trajectories of the cluster center, and provides academic warnings based on the degree of deviation to achieve previous academic management. Overall, this model extracts valuable information by analyzing students' behavioral trajectories, and then identifies common and abnormal patterns of student behavior through clustering and pattern analysis, thereby achieving early warning and management of students' academic performance.

3.1. Design of Campus Life Information Collection Model for College Students. In the campus life of college students, the factors that affect students' academic achievement are mainly divided into two categories, which are personal and impersonal factors. The personal factors include students' cognitive ability, creative ability and other inherent abilities, while the impersonal factors are based on the view of academic life and social network in which students live [23]. In the student trajectory feature analysis model part, the trajectory feature analysis model construction is shown in Fig 3.1.

From Fig 3.1, the trajectory feature analysis model converts and analyzes information feature points of students in campus life and academic management information feature points, respectively. The campus life information feature points include campus consumption records, campus network usage records, and campus access control records. The academic management information feature points include scholarship information and student information, etc. The dimension of information features is shown in Figure 3.2.

From this, three types of data collection contacts are derived. One is the campus information system, i.e., the system containing dormitory access control information, student campus network information, and student consumption information. The second is the basic student information and dormitory assignment information. The third is student academic performance and scholarship data. Starting from these three data
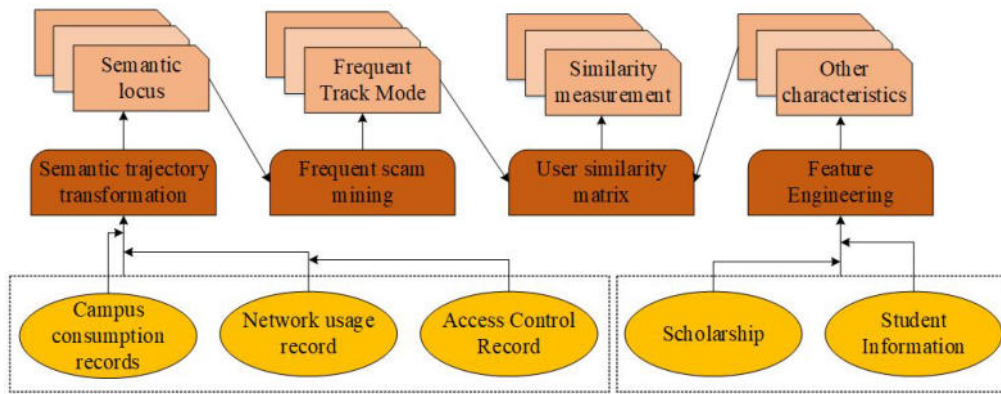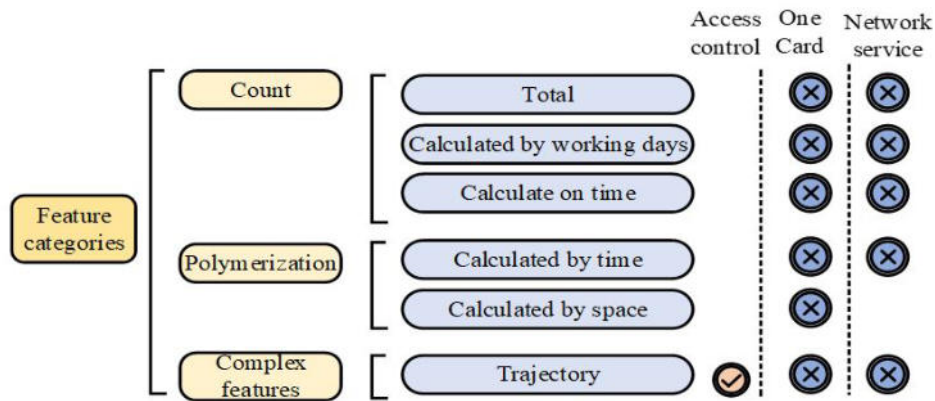
Fig. 3.1: Trajectory feature analysis model

Fig. 3.2: Information feature dimension

dimensions, the model can basically outline the campus image of students when conducting data collection. The data collection process is based on quantitative behavior statistics, time statistics and frequency statistics as the main quantitative measures, and different quantitative statistics are used depending on the nature of the behavior. Quantitative behavior statistics refers to the quantitative count of a behavior or behavior results. Time statistics is used to measure the duration of a student's behavior, while the frequency statistics is used to measure the number of times a student performs a particular behavior. The model quantitation values are collected in the manner shown in Fig 3.3.

In analyzing the trajectory characteristics of students' campus life, it is necessary to mark the campus and the functional areas within the campus repeatedly, because the university may have more than one campus, and there may be several functional areas with similar functions within the same campus for diverting student traffic. Therefore, the campus is labeled as C_γ and the functional areas are classified into twelve different categories according to the functions they perform: classroom, dormitory, cafeteria, library, courtyard building, bathroom, office, hospital, supermarket, water room, multimedia area, and other areas, denoted by F_η. For the first λ location within the campus, it can be expressed by $p_{(\gamma,\lambda)}$, which can be defined by the location and spatial position in the form of equation (3.1).
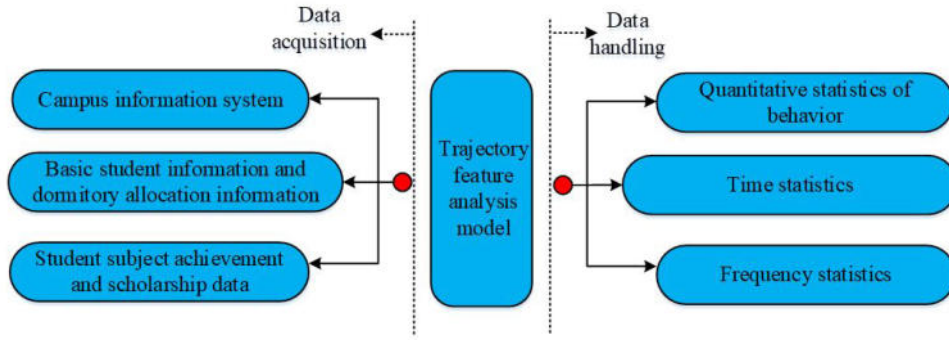
$$P = (p.loc, p.fun) \tag{3.1}$$

Fig. 3.3: Model quantization value acquisition method

In equation (**??**), $p.loc$ denotes the actual spatial location and $p.fun$ denotes the functional area. Based on this, the study defines student activity as a vector dependent on activity behavior, represented by $a.attr$. The student activity sequence is also defined as a spatio-temporal sequence around the individual student, as shown in equation (3.2).

$$Aseq = \{(t_1, p_1), \ldots, (t_k, p_k)\} \tag{3.2}$$

In equation (3.2), t denotes the timestamp, $t_i < t_j \ (i < j)$, assuming the existence of a given active sequence with sequence parameters. p_1 and p_k are different spatio-temporal point locations. When both spatio-temporal point locations satisfy the constraints of Equation (3.3) at the same time, the two point locations can be judged as the same point location.

$$\begin{cases} p_i = p_{i+1} \\ |t_i - t_{i+1}| < \xi \end{cases} \tag{3.3}$$

In equation (3.3), ξ denotes the sequence parameters, t_i denotes the time, and p_i denotes the location. The model performs dwell point detection on the activity sequence, and then uses the dwell point data as the basis for trajectory compression, and finally outputs the semantic trajectory. In the semantic trajectory representation, the activity sequence of individual student and individual is fixed, and the relationship between trajectory Tra and activity sequence is shown in equation (3.4).

$$Tra \subseteq Aseq \tag{3.4}$$

3.2. Analysis Model of Campus Life Information Trajectory for College Students. . Due to the different daily routines of different students, there are periodic differences in their life trajectory information. Therefore, the system needs to effectively distinguish this differential information [24]. On this basis, the model will further segment the semantic trajectories of the students, and the segmentation operation can divide the trajectories of the students for a long time appropriately. The main trajectory segmentation methods can be divided into three types: time-threshold trajectory segmentation, set topology trajectory segmentation, and trajectory semantic trajectory segmentation. Since the student trajectories are based on the campus teaching and activity time as the main axis, the study adopts the time-threshold trajectory segmentation method. According to the time-threshold trajectory segmentation method, students' action trajectories are divided into day-based daily trajectories, and each segment of daily trajectories represents a day's travel of students. When dividing the daily trajectory, a day is not a day divided by a specific time point in physical time, but a complete activity of students in a basic time unit of a day is used as the basis for dividing the trajectory. If a student's activity trajectory exceeds the physical time boundary of a day, but is still within the complete activity of the day, then this part of the trajectory is still slid into the daily trajectory. An example of this classification is shown in Fig 3.4.
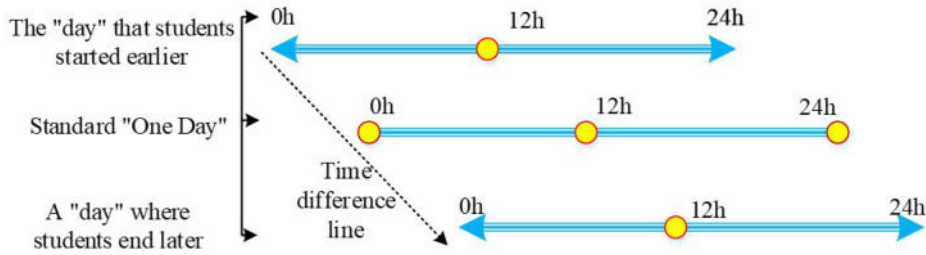
Fig. 3.4: Division example

The daily trajectory can be expressed in the form of equation (3.5).

$$DTra = \{(t_1, p_1), \ldots, (t_s, p_s)\}, DTra \in Tra \tag{3.5}$$

In Tra, there exists one and only (t_j, p_j), such that $(t_1, p_1) = (t_j, p_j)$,$(t_{1+n}, p_{1+n}) = (t_{j+n}, p_{j+n})$. The imputation pattern is then shown in equation (3.6).

$$TraP = P'_1 \overset{\Delta t'_1}{\to} P'_2 \overset{\Delta t'_2}{\to} \overset{\Delta t'_{u-1}}{\underset{\ldots}{\to}} P'_u \tag{3.6}$$

Where $p'_i \in \{p_j\}$. If p'_i corresponds to (t_i, p_i), and p'_{i+1} corresponds to (t_j, p_j), then $\Delta t'_i = t_j - t_i$.

3.3. 3.3 Academic Management Model Design. . After extracting the required student activity trajectory feature information, the study proposes an academic management model based on clustering algorithm. The model is divided into three main modules, which are data clustering analysis, trajectory frequent pattern analysis and trajectory deviation analysis. K-means can classify student trajectory data information based on data features [25]. The model first uses the k-means algorithm for clustering analysis, based on which the frequent pattern analysis of student trajectories within the clustered clusters is performed using the PrefixSpan strategy based on the clustering results. The PrefixSpan method flow is shown in Figure 3.5.

Finally, based on the output results of frequent trajectory patterns, it calculates the degree of deviation between the trajectories of individual students in the clusters and the trajectories of the cluster centers, and carries out academic warning according to the degree of deviation to achieve prior academic management. The specific structure is shown in Fig 3.6.

When the model performs the clustering operation, it assumes that there exists a base data set X and each data has a M dimensional feature vector X_n, then x_{nm} represents the feature value of the m feature of the n data. The clustering approach is shown in Fig 3.7.

The model divides the data instances into clusters as shown in equation (3.7).

$$C = \{C_1, C_2, \ldots, C_k\} \tag{3.7}$$

There is no intersection between clusters, while each cluster has a cluster center, and the similarity between different clusters is relatively low, but the data instances inside the clusters are more similar. The sum of the distance between the data inside the cluster and the cluster center is the objective function, as shown in equation (3.8).

$$P(U, c) = \sum_{k=1}^{K} \sum_{n=1}^{N} u_{nk} \sum_{m=1}^{M} d(x_{nm}, c_{km}) \tag{3.8}$$

In equation (3.8), U denotes the matrix describing the affiliation status of the clusters, and u_{nk} denotes the affiliation status of the data instance with the ordinal number n for the ordinal number k. $d(x_{nm}, c_{km})$ denotes the distance between the center of the clustering cluster and the data instance. Since this distance
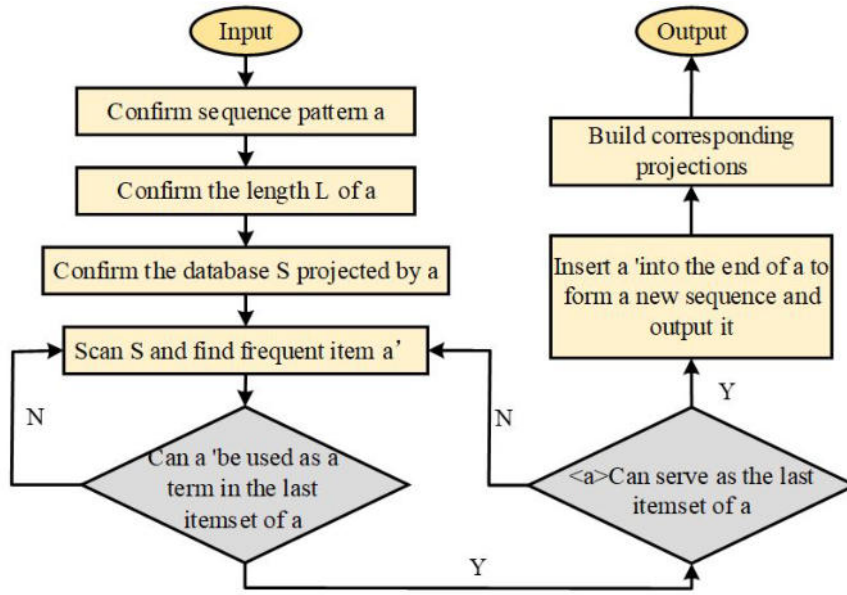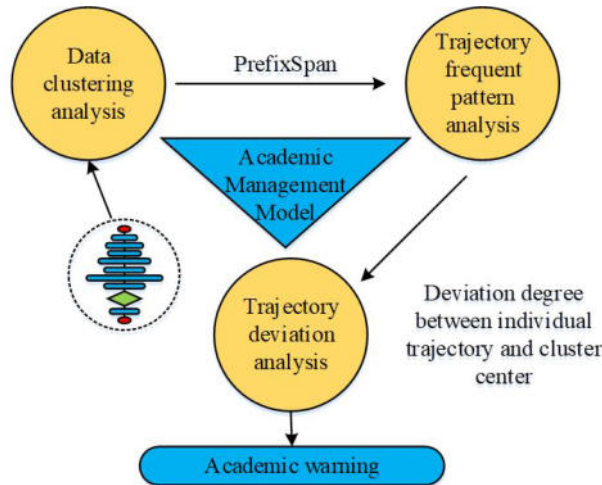
Fig. 3.5: The PrefixSpan method flow

Fig. 3.6: Academic management model structure

yields different results with different metrics, the study requires a choice of metric for the model. The model mainly uses the Euclidean metric as the main metric, as shown in equation (3.9).

$$P\left(U, c\right) = \sum_{k=1}^{K} \sum_{n=1}^{N} u_{nk} \sum_{m=1}^{M} d\left(x_{nm} - c_{km}\right)^{2} \tag{3.9}$$

As can be seen from equation (3.9), the metric treats all data features equally, and differences in data feature differences lead to different clustering results, so a weighting mechanism needs to be added to the
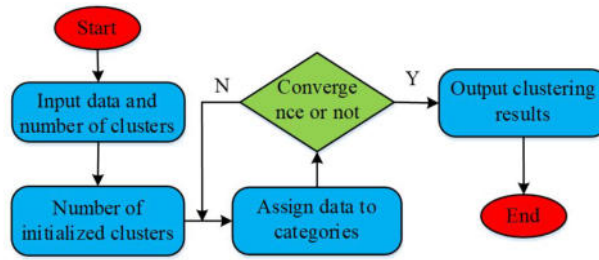
Fig. 3.7: Clustering method

model, as shown in equation (3.10).

$$P\left(U,c\right) = \sum_{k=1}^{K} \sum_{n=1}^{N} u_{nk} \sum_{m=1}^{M} w_m^{\beta}\left(x_{nm} - c_{km}\right)^2 \tag{3.10}$$

In equation (3.9), w_m denotes the data feature weights with the number m and β denotes the custom parameter. Equation (3.10) can be transformed into equation (3.11) since the weights are subject to the naturalness condition of data sum equal to 1.

$$w_m = \frac{1}{\sum_{t \in F}\left[D_m/D_t\right]^{1/(\beta-1)}} \tag{3.11}$$

In equation (3.11), D_m denotes the sum of variances of all features within the clustered clusters. After designing the weighting mechanism, the study introduces the objective and subjective combining weighted k-means (Wosk-means) algorithm to assign values to each data feature, and the assignments are made in two ways: subjective weight assignment and objective weight assignment, and the integrated weights are shown in equation (3.12).

$$a_m = \frac{w_m v_m}{\sum_{i=1}^{M} w_m v_m} \tag{3.12}$$

In equation (3.12), w denotes the subjective weights, v denotes the objective weights, and m denotes the data feature numbers within the data clusters. The flow of the Wosk-means algorithm is shown in Fig 3.8.

In Fig 3.8, the model first standardizes the initial data, after which the data feature weights are initialized to make all data feature weights consistent. After processing the weights, the cluster centers need to be confirmed, and in the face of the given cluster centers and weights, the distance metric of the weights needs to be used to update the division of clusters. Based on this, the mean values of all features within the clusters are divided according to the existing weights and clusters, and the clustering centers are calculated and updated. Finally, the feature weights are updated according to the new clustering centers and clusters, and whether the algorithm converges or not is observed. At present, there are two main algorithms in the field of trajectory frequent pattern analysis, namely, Apriori algorithm and tree algorithm. The study uses the PrefixSpan algorithm which integrates the two algorithms for analysis, and the method can effectively reduce the cost of data mining. In the PrefixSpan algorithm, all sequences are arranged in an ordered manner, while all sequences are composed of item sets, which can be further split into different items. First, the database is input to the model and the minimum support minSup is defined. The length of the sequence pattern α is set to L and the projected database is $S\mid_{\alpha}$. A word scan is performed on $S\mid_{\alpha}$ and frequent items are found that satisfy the qualification.

On top of the trajectory pattern, the study transforms the trajectory data in a certain way and thus forms the distance between the trajectory features. The distance is expressed in the form of similarity. Similarity is essentially a comparison of the percentage of similar nodes with similar matches. Since there may be reading trajectory matches with different lengths between two trajectories, it is necessary to first find the frequent match
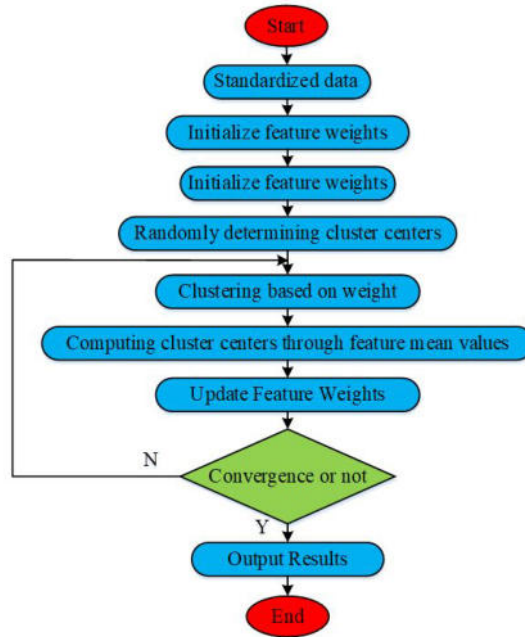
Fig. 3.8: Wosk-means algorithm flow

pattern, and then calculate the different match lengths to finally obtain the combined similarity of different length trajectory matches. It supposes that there exists a trajectory pattern $TraP_1$, as shown in equation (3.13).

$$TraP_1 = p'_{11} \overset{\Delta t'_{11}}{\to} p'_{12} \overset{\Delta t'_{12}}{\to} \cdots \overset{\Delta t'_{1[u-1]}}{\to} p'_{1u} \tag{3.13}$$

In equation (3.14), t indicates the time and p indicates the location. A trajectory pattern $TraP_2$ is also presented.

$$TraP_2 = p'_{21} \overset{\Delta t'_{21}}{\to} p'_{22} \overset{\Delta t'_{22}}{\to} \cdots \overset{\Delta t'_{2[u-1]}}{\to} p'_{2v} \tag{3.14}$$

Then the similarity of the two trajectories is shown in equation (3.15).

$$S\left(TraP_1, TraP_2\right) = \sum_{k=1}^{K} f_w\left(k\right) Sl\left(FT_1^k, FT_2^k\right) \tag{3.15}$$

In equation (3.15), k represents the trajectory matching length, $f_w\left(\right)$ is the weight ratio representation, l is the trajectory matching pattern representation, and FT_1^k and FT_2^k represent the pattern matching subset, respectively.

4. College Student Trajectory Data Mining Model Trajectory Analysis Results.

4.1. Elbow Method Test . In the study of trajectory analysis of college students' trajectory data mining model, student information was first collected from various information segments within the university. The main information collection ends were five types of campus student information system, all-in-one card consumption record, network service record, access control record and action trajectory record. The dataset settings used in the experiment are shown in Table 4.1.

The experimental setup is shown in Table 4.2.

Table 4.1: Data type and data scale

Data source	Data properties	Data size
Campus student information system	Number of students	6714
	Location type and quantity	23
	Time interval	2021.12.01-2022.12.01
All-in-one card consumption record	Number of data records	185298
Network service record	Number of data records	187592
Access control record	Number of data records	167817
Action trajectory record	Maximum action trajectory length	181

Table 4.2: The experimental setup

Simulation settings	Detailed description
Experimental software	Use Python programming language for algorithm development and data processing.
	The applied clustering algorithm was implemented using the Scikit learn library.
	Use Jupyter Notebook for interactive calculations.
Experimental hardware	Desktop computer with Intel Core i7 processor and 16GB of memory.
	Using solid-state drives as storage devices.
Experimental condition	Experimental data collection period: December 1, 2021 to December 1, 2022.
	Use internal data sources within universities for analysis.
	Use elbow analysis and homogeneity analysis to evaluate performance.

Based on Table 4.1 and 4.2, the study first tested the performance of the applied clustering algorithm designed for the study, in which the elbow analysis method and the homogeneity analysis method were used to analyze the data, as shown in Fig 4.1.

In Fig 4.1, in the elbow method test, the overall error sum of squares of the applied clustering algorithm designed in the study showed a significant decreasing trend when the number of clusters rose. The decrease shrank significantly after the number of clusters was greater than 5, while the decrease almost disappeared when the number of clusters reached about 8, which shows that the number of clusters in the interval of 5 to 8 is the optimal setting range. The homogeneity test showed that the homogeneity of the algorithm was significantly improved at the number of clusters 5 and 8. In the comparison test, the clustering of data features in the traditional k-mean algorithm was more evenly distributed and did not provide effective information. In contrast, the applied clustering algorithm designed in the study first clustered the data into two large clusters of 0 and 1, where the feature distribution was still balanced. Then the algorithm further divided the two large clusters into five small clusters, where the feature differences between the clusters were already obvious.

4.2. Trajectory Feature Analysis. The results of trajectory features for different student types are shown in Table 4.3.

From Table 4.3, the model can classify trajectories in more detail for different disciplines and genders, from which the trajectory classification can reveal the characteristic patterns of action trajectories of different types of students in campus actions. The matching statistics of action trajectories between two two locations are shown in Fig 4.2.

In Fig 4.2, the action trajectory category from dormitory to dormitory received the highest number of matches within each cluster, with 98.92%, 98.98%, 96.15%, 98.37%, and 97.61% from category 1 to category 4, respectively. This was followed by the category from dormitory to canteen, with 87.64%, 87.86%, 86.97%, 88.63%, and 88.71% from category 1 to category 4, respectively. However, the category from dormitory to dormitory was somewhat invalid, so a trajectory similarity analysis was also needed, as shown in Fig 4.3.

Figure 4.3 illustrates the discrepancy in similarity between students' behavioral trajectories and the centroid trajectories of the cluster category to which they belong. A lower similarity indicates a greater divergence in students' behavioral trajectories from those of other students within the same cluster category. Moreover, a
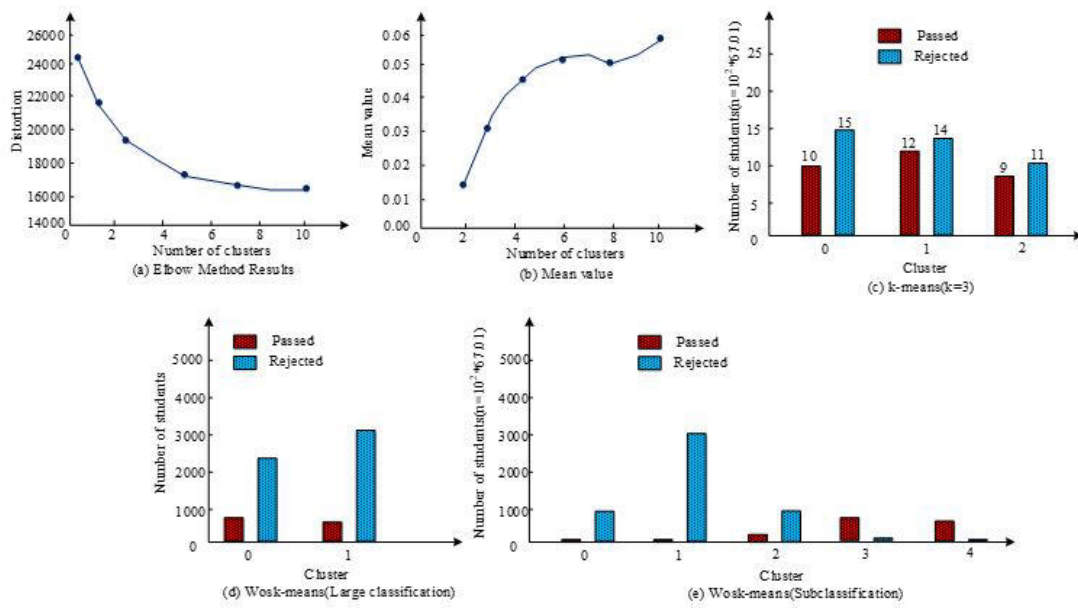
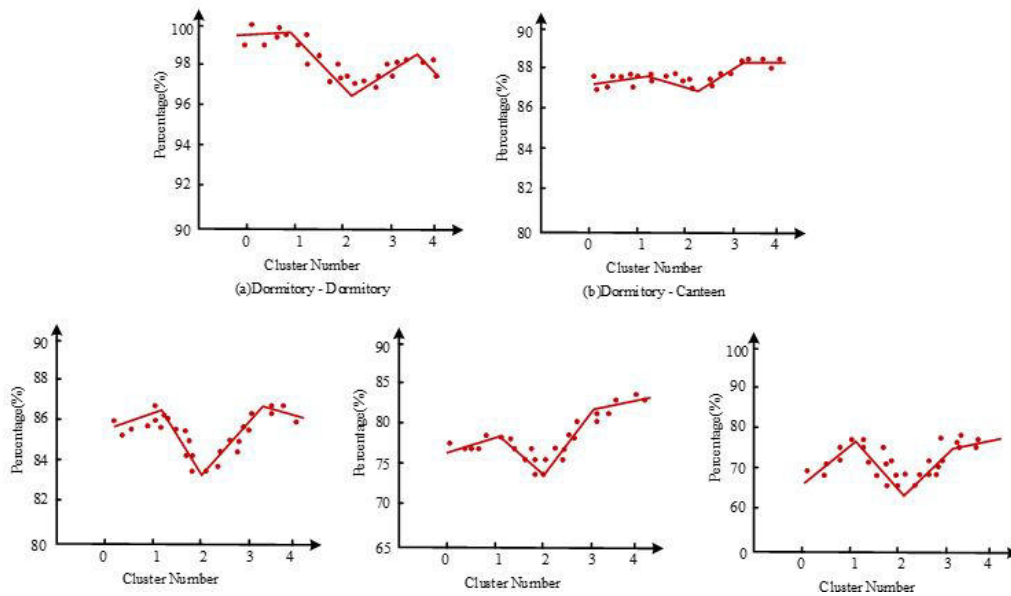Fig. 4.1: Elbow analysis and homogeneity analysis

Fig. 4.2: Matching of action trajectories between two locations

higher percentage of this group of students indicates a greater number of students within that cluster category whose action patterns have changed. In this study, the similarity threshold was set at 10%, where the percentages of students with less than 10% similarity in categories 0, 1, 2, 3, and 4 were 1.18%, 1.17%, 0.66%, 0.42%, and 1.45%, respectively. The behavior patterns of this group of students were likely to change dramatically. The deviation directions of students whose trajectories deviated from the centroid cluster category are shown

Table 4.3: Trajectory features results for different student types

Cluster number	Subdivision	Campus card consumption data						Network usage		
		Breakfast shop	Print shop	Dining room	Bath	Playing field	School hospital	Rate of flow	Duration	Number of connections
0	The male sex	-34.61	39.71	21.06	-1.82	8.96	-42.65	0.76	3.76	6.71
	Femininity	-10.43	51.94	21.05	39.72	-12.52	8.19	3.06	14.77	13.96
	Science	-23.58	48.41	-5.14	16.22	9.77	-27.39	2.48	4.75	15.19
	Liberal arts	-7.55	35.72	10.69	31.87	-17.35	4.33	2.02	15.88	6.78
1	The male sex	-15.31	-49.11	-1.05	-22.07	28.86	-15.92	-7.45	-5.97	-3.77
	Femininity	6.13	-19.05	1.84	29.86	-0.58	28.83	-12.17	7.48	6.95
	Science	2.07	-5.48	-16.18	-12.75	43.48	-5.68	-6.38	-3.84	-2.13
	Liberal arts	-12.55	-49.05	-1.98	13.51	-38.44	8.66	-15.37	4.01	4.28
2	The male sex	2.02	-29.03	-9.02	-57.82	41.57	-11.33	-31.53	-45.43	-48.46
	Femininity	20.66	24.51	7.56	-29.95	-39.98	-16.28	-34.52	-42.65	-45.52
	Science	-1.46	-40.53	-13.46	-42.1	1.32	-28.37	-27.75	-43.92	-45.62
	Liberal arts	29.11	-12.27	2.21	-39.05	-20.44	-0.81	-38.94	-43.46	-47.63
3	The male sex	-10.68	1.92	-13.01	7.25	-5.15	-48.17	10.42	7.82	10.78
	Femininity	-21.81	43.91	14.05	37.65	-40.78	-37.78	4.32	7.75	9.38
	Science	-13.98	-23.05	-2.97	9.52	-29.08	-66.24	11.86	8.43	10.79
	Liberal arts	-17.99	16.41	8.63	40.08	-25.54	-33.18	-0.28	6.46	9.01
4	The male sex	40.43	15.36	3.97	-4.68	77.19	29.07	84.97	54.08	46.39
	Femininity	40.44	35.54	28.12	32.18	-20.62	46.96	47.12	39.27	31.67
	Science	56.66	-0.12	-13.48	10.31	39.43	10.11	63.61	34.31	28.31
	Liberal arts	19.72	71.58	2.33	34.77	-29.93	81.43	52.07	55.72	46.03

Table 4.4: Deviation direction

Percentage (%)	0	1	2	3	4
0	\	41.64	33.34	8.37	16.65
1	25.12	\	58.35	13.82	2.72
2	12.61	37.53	\	50.00	0.00
3	0.00	33.34	66.68	\	0.00
4	10.01	20.00	40.00	30.00	\

in Table 4.4.

In Table 4.4, the vertical direction represents the cluster in which the student is located, and the horizontal direction represents the cluster that the student is deviating towards. The bias of category 0 toward category 1 was higher, with a bias value of 25.12%. the bias of category 1 toward category 0 was higher, with a bias value of 41.64%. The bias of category 2 toward category 3 was higher, with a bias value of 66.68%. The bias of category 3 toward category 2 was higher, with a bias value of 50.00%. The bias of category 4 toward category 1 was higher, with a bias value of 16.65%. This showed that the model designed in the study can not only characterize the trajectory of students' campus actions, but also predict the possible changes of students' action patterns, providing new ideas for the management of students' campus life and academics.

In the benchmark comparison, the total sum of squared errors of the trajectory analysis clustering algorithm designed in the study was slightly lower than that of the benchmark K-means clustering algorithm. From the
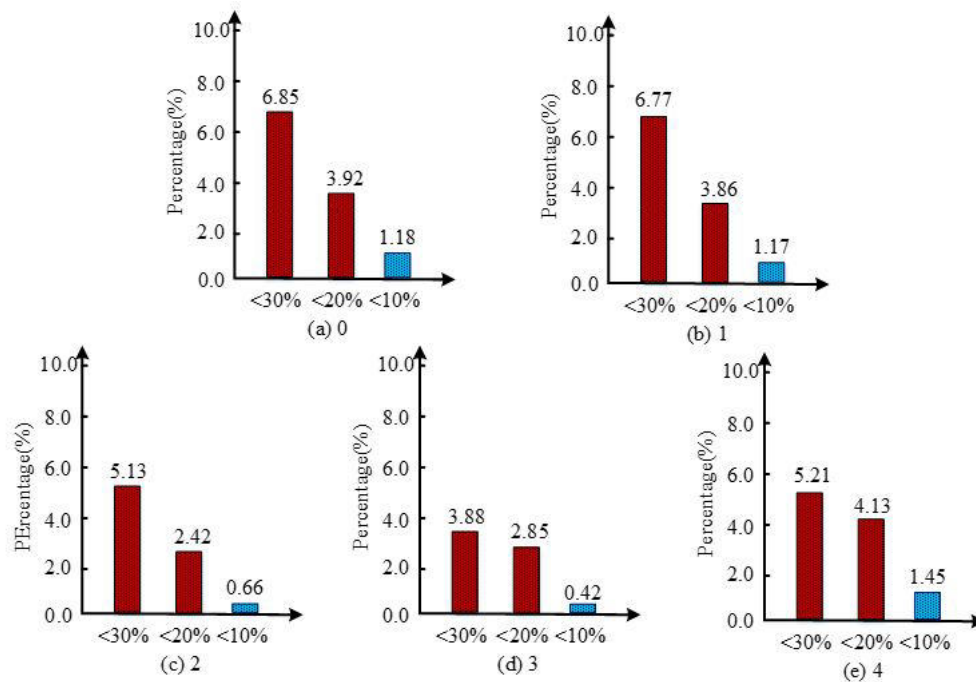
Fig. 4.3: Trajectory similarity analysis

Table 4.5: Benchmark comparison of different dataset sizes

Dataset size	Algorithm	Number of clusters	Total sum of squared errors	Cluster homogeneity
Sample size: 100 cases	Benchmark K-means	5	10251	0.83
		8	7562	0.88
	Applied clustering algorithm	5	8674	0.83
		8	7346	0.89
Sample size: 1000 cases	Benchmark K-means	5	102302	0.77
		8	75525	0.83
	Applied clustering algorithm	5	86007	0.82
		8	73220	0.89

perspective of clustering homogeneity, the total sum of squared errors of the trajectory analysis clustering algorithm was higher. Therefore, the trajectory analysis clustering algorithm designed in the study had more advantages in data feature extraction and analysis.

By increasing the size of the dataset by 10 times, the trajectory analysis clustering algorithm designed in the study still had advantages in the total sum of squared errors, indicating that the designed algorithm had stronger processing power when facing large-scale datasets. In the comparison of clustering homogeneity, the trajectory analysis clustering algorithm had higher clustering homogeneity, indicating that as the dataset size increased, the designed algorithm had more performance advantages.

With the increasing emphasis on talent cultivation in society, college students, as an important source of national talent reserves, are gradually receiving widespread attention from all sectors of society. The traditional talent cultivation system lacks adaptability to personalized and practical talent cultivation, making it difficult to meet the talent needs of today's society. Therefore, utilizing the current wave of social informatization and data development to update the talent cultivation system has become a new way of talent cultivation [26-27]. However, how to effectively utilize data mining technology to track and manage students' campus

life and learning, in order to achieve adaptive and efficient management, is an urgent problem to be solved [28]. In the research results, it can be seen that the model first collected student information from various sub sources within the university. After clustering analysis of these data, the applied clustering algorithms demonstrated significant performance in effectively distinguishing students' behavioral characteristics, while traditional K-means clustering algorithms could not provide this effective information. In addition, the applied clustering algorithm designed in the study had the optimal setting range within the range of 5 to 8 clusters. After analyzing the trajectory characteristics of different types of students, it can be found that the model could classify trajectories in more detail for students of different disciplines and genders. This indicated that through trajectory classification, characteristic patterns of different types of students' school behavior can be discovered. At the same time, it can be observed that the matching rate of behavior trajectories from dormitories to canteens was relatively high, but further trajectory similarity analysis is needed to confirm. In trajectory similarity analysis, there were significant differences between the behavior trajectories of most students and the centroid trajectories of their clustering categories. This difference may indicate a change in students' behavioral patterns.

By analyzing the similarity of trajectories, it was found that the similarity between students' action trajectories and the centroid trajectories of their cluster category was low, indicating that their behavior trajectories were different from those of other students within the cluster category. The more students in this situation, the greater the likelihood of changes in student behavior patterns within the cluster category. Finally, the study also found that students' action trajectories deviate from the direction of centroid clustering categories to a certain extent. For example, the deviation degree from category 0 to category 1 was relatively high, with a deviation value of 25.12%. The deviation degree from category 1 to category 0 was relatively high, with a deviation value of 41.64%. The deviation degree from category 2 to category 3 was relatively high, with a deviation value of 66.68%. The deviation degree from category 3 to category 2 was relatively high, with a deviation value of 50.00%. The deviation from category 4 to category 1 was relatively high, with a deviation value of 16.65%. These results indicated that the model designed in this study can not only describe students' campus action trajectories, but also predict possible changes in student behavior patterns, providing new ideas for students' campus life and academic management. Meanwhile, in the research, it is also possible to predict the possible changes in student behavior patterns, providing new ideas for students' campus life and academic management. Overall, through data mining technology, it is possible to gain a deeper understanding and analysis of the learning behavior and life trajectory of college students, thereby providing more effective support and methods for personalized education of students and talent cultivation in universities. So far, data mining technology has provided people with a novel and efficient method for student management and educational reform.

5. Conclusion. The research addressed the problem that the academic management of students in universities lacks daily campus life management, and proposed a model for mining students' daily academic life trajectories that combines the existing university information system and data mining technology. The model extracted and analyzed the information of students' daily activity, and designed an applied clustering algorithm to classify different trajectory types on this basis. The research results showed that in the elbow test and the homogeneity test, the applied clustering algorithm had obvious variation characteristics at the cluster number 5-8, and this interval was the best cluster number interval. The applied clustering algorithm in the comparison test was better at clustering the data into clusters with features than the ordinary clustering algorithm. In the trajectory analysis, the model could classify trajectories in more detail for different disciplines and genders, in which the categories of action trajectories from dormitory to canteen were 87.64%, 87.86%, 86.97%, 88.63%, 88.71% from category 0 to 4 respectively, which were the most effective action trajectories with the highest number of matches. The percentage of people with similarity less than 10% in categories 0 to 4 were 1.18%, 1.17%, 0.66%, 0.42%, and 1.45%. The behavior patterns of this group of students were likely to change dramatically. In addition, the model was able to predict the direction of trajectory change, with the main directions being category 0 to category 1, category 1 to category 0, category 2 to category 3, category 3 to category 2, and category 4 to category 1. This showed that the model designed in the study can effectively analyze student estimation and provide assistance for comprehensive academic life management in universities.

However, the drawback is that this study mainly relied on data from university information systems, but students' daily life trajectories may be influenced by more elements, such as social media activities, health,

and psychological conditions. At the same time, this study mainly focused on the analysis of student behavior trajectories, but did not involve how to effectively intervene based on these analysis results. Therefore, in future research, it is possible to explore how to use these analysis results to design and implement effective student life management strategies. At the same time, future research can consider integrating more types of data sources to provide a more comprehensive trajectory of student life.

REFERENCES

[1] P. Radanliev, D. De Roure, and R. Walton, "Diabetes & Metabolic Syndrome: Clinical Research & Reviews," Diabetes & Metabolic Syndrome: Clinical Research & Reviews, vol. 14, no. 5, pp. 1121-1132, 2020.

[2] A. Ishaq, S. Sadiq, M. Umer, S. Ullah, S. Mirjalili, V. Rupapara, and M. Nappi, "Improving the prediction of heart failure patients' survival using SMOTE and effective data mining techniques," IEEE Access, vol. 9, pp. 39707-39716, 2021.

[3] H. Lou, "Design of college English process evaluation system based on data mining technology and internet of things," International Journal of Data Warehousing and Mining (IJDWM), vol. 16, no. 2, pp. 18-33, 2020.

[4] Y. Zeng, "Evaluation of physical education teaching quality in colleges based on the hybrid technology of data mining and hidden Markov model," International Journal of Emerging Technologies in Learning (iJET), vol. 15, no. 1, pp. 4-15, 2020.

[5] K. Yu, T. Yuan, and Y. Li, "Application of Data Mining Technology in Sports Data Analysis in Colleges and Universities," in 2021 International Conference on Information Technology and Contemporary Sports (TCS), IEEE, 2021, pp. 329-332.

[6] I. S. P. James, P. Ramasubramanian, and D. M. D. Angeline, "Student absenteeism in engineering college using rough set and data mining approach," International Journal of Advanced Intelligence Paradigms, vol. 23, no. 3-4, pp. 423-433, 2022.

[7] Z. Zhang, Z. Chen, and C. Xu, "Recognition method of diversified teaching mode of college physical training based on data mining," Ann. For. Res, vol. 65, no. 1, pp. 9420-9432, 2022.

[8] L. D. Yulianto, A. Triayudi, and I. D. Sholihati, "Implementation Educational Data Mining for Analysis of Student Performance Prediction with Comparison of K-Nearest Neighbor Data Mining Method and Decision Tree C4.5: Implementation Educational Data Mining for Analysis of Student Performance Prediction with Comparison of K-Nearest Neighbor Data Mining Method and Decision Tree C4.5," Jurnal Mantik, vol. 4, no. 1, pp. 441-451, 2020.

[9] H. Zhang and M. Fang, "Research on the integration of heterogeneous information resources in university management informatization based on data mining algorithms," Computational Intelligence, vol. 37, no. 3, pp. 1254-1267, 2021.

[10] S. M. Lee and J. H. Lee, "The influence of adaptation of dental hygiene students to campus life on satisfaction in major," Journal of Korean Society of Dental Hygiene, vol. 21, no. 3, pp. 281-290, 2021.

[11] W. Li, "Design of smart campus management system based on internet of things technology," Journal of Intelligent & Fuzzy Systems, vol. 40, no. 2, pp. 3159-3168, 2021.

[12] Y. Kim, B. Kim, H. S. Hwang, and D. Lee, "Social media and life satisfaction among college students: A moderated mediation model of SNS communication network heterogeneity and social self-efficacy on satisfaction with campus life," The Social Science Journal, vol. 57, no. 1, pp. 85-100, 2020.

[13] S. Purnama, Q. Aini, U. Rahardja, N. P. L. Santoso, and S. Millah, "Design of Educational Learning Management Cloud Process with Blockchain 4.0 based E-Portfolio," Journal of Education Technology, vol. 5, no. 4, pp. 628-635, 2021.

[14] K. A. Way, L. Burrell, L. D'Allura, and K. Ashford-Rowe, "Empirical investigation of authentic assessment theory: An application in online courses using mimetic simulation created in university learning," Assessment & Evaluation in Higher Education, vol. 46, no. 1, pp. 17-35, 2021.

[15] W. Haoxiang and S. Smys, "Big data analysis and perturbation using data mining algorithm," Journal of Soft Computing Paradigm (JSCP), vol. 3, no. 01, pp. 19-28, 2021.

[16] T. S. Kumar, "Data mining based marketing decision support system using hybrid machine learning algorithm," Journal of Artificial Intelligence, vol. 2, no. 03, pp. 185-193, 2020.

[17] P. Edastama, A. Dudhat, and G. Maulani, "Use of Data Warehouse and Data Mining for Academic Data: a Case Study at a National University," International Journal of Cyber and IT Service Management, vol. 1, no. 2, pp. 206-215, 2021.

[18] Z. S. Ageed, S. R. M. Zeebaree, M. M. Sadeeq, F. F. Kak, H. S. Yahia, M. R. Mahmood, and I. M. Ibrahim, "Comprehensive survey of big data mining approaches in cloud systems," Qubahan Academic Journal, vol. 1, no. 2, pp. 29-38, 2021.

[19] H. A. Mengash, "Using data mining techniques to predict student performance to support decision making in university admission systems," IEEE Access, vol. 8, pp. 55462-55470, 2020.

[20] A. Garg, P. Kumar, M. Madhukar, O. Loyola-González, and M. Kumar, "Blockchain-based online education content ranking," Education and Information Technologies, vol. 27, no. 4, pp. 4793-4815, 2022.

[21] S. Dutt, N. J. Ahuja, and M. Kumar, "An intelligent tutoring system architecture based on fuzzy neural network (FNN) for special education of learning disabled learners," Education and Information Technologies, vol. 27, no. 2, pp. 2613-2633, 2022.

[22] C. Choudhary, I. Singh, and M. Kumar, "SARWAS: Deep ensemble learning techniques for sentiment based recommendation system," Expert Systems with Applications, vol. 216, pp. 119420, 2023.

[23] K. Y. Siu, "The Effect of Working Memory on Bilingual Learning Ability," Journal of Education, Humanities and Social Sciences, vol. 8, pp. 2118-2123, 2023.

[24] R. Darmayanti, "Gema Cow-Pu: Development of Mathematical Crossword Puzzle Learning Media on Geometry Material on Middle School Students' Critical Thinking Ability," Assyfa Learning Journal, vol. 1, no. 1, pp. 37-48, 2023.

[25] Y. Du, "Study on Cultivating College Students' English Autonomous Learning Ability under the Flipped Classroom Model," English Language Teaching, vol. 13, no. 6, pp. 13-19, 2020.

[26] G. HR and P. S. Aithal, "Sales Personnel Training–An Integrated Framework for Indian Brick-and-Mortar Retailers," International Journal of Case Studies in Business, IT, and Education (IJCSBE), vol. 4, no. 1, pp. 172-187, 2020.

[27] E. F. Zeer, V. S. Tretyakova, and V. I. Miroshnichenko, "Strategic directions of pedagogical personnel training for the system of continuing vocational education," The Education and Science Journal, vol. 21, no. 6, pp. 93-121, 2019.

[28] H. Zhang, X. He, and H. Mitri, "Fuzzy comprehensive evaluation of virtual reality mine safety training system," Safety Science, vol. 120, pp. 341-351, 2019.