International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18 (2026) pp. 1-8

https://doi.org/10.70917/ijcisim-2026-0054
Article

Optimizing Student Development Pathways and Personal Competency Enhancement Strategies in Colleges and Universities Using Cluster Analysis

Shucheng Li *

Law School, Tiangong University, Tianjin 300387, China; lishucheng2001@126.com

Abstract: The development of higher education is inseparable from the cultivation and development planning of students. The current quality of higher education development in universities and the rationality and effectiveness of student development planning directly affect the effective value of higher education in serving society. Through empirical analysis and the application of data-driven methods, this study optimizes students' future development pathways, particularly focusing on the development characteristics of undergraduate students from the 2020-2023 cohorts at a comprehensive higher education institution. The study employs an optimized extended centroid clustering algorithm combined with a Bayesian scoring function to classify the clustering results derived from multivariate data—including student growth pathways, participation in extracurricular activities, and mental health status—into five distinct categories with significant differences. Through research experiments, compared to traditional extended centroid clustering methods, the co-word clustering algorithm developed in this study can more clearly demonstrate the developmental characteristics of student groups. The co-word clustering method utilizes keyword data mining to extract sample characteristics of students' multidimensional development, thereby conducting multi-angle clustering to analyze the basic characteristics of student types and providing recommendations for fostering differentiated developmental capabilities among students.

Keywords: clustering analysis; higher education; student developmental pathways; personal capability enhancement; data-driven

1. Introduction

1.1. Research Background

Following the rapid development of mass higher education, the knowledge base, learning habits, abilities, development aspirations, and development requirements of individual higher education students have exhibited diverse and non-linear characteristics. The traditional approach to talent cultivation, which emphasizes a one-size-fits-all, uniform training model, is no longer adequate to meet the personalized development needs of university students in the new era [1-4]. How to promote students' diverse development through scientifically designed development pathways while respecting and safeguarding individual choices is an important component of higher education reform in the new era. Additionally, educational evaluation models are undergoing a transformation from focusing on evaluation goals to emphasizing the improvement of evaluation processes. Universities should shift away from the mindset of judging students' performance solely based on learning outcomes and instead focus on uncovering students' potential and cultivating their ability for sustained development [5-7].

As a result, how to effectively utilize big data technologies to conduct in-depth analysis of students' learning behaviors and performance, thereby achieving precise education, has become an important means [8-9]. Especially in recent years, the development of educational informatization and intelligent analysis technologies has made data-driven decision-making an increasingly common approach in student management and talent cultivation [10-11]. By collecting a range of student data, including academic performance, course selection, extracurricular activities, and psychological assessments, universities can better construct a comprehensive "student development profile" [12]. However, faced with massive amounts of heterogeneous data, determining which information can be useful for educational decision-making has become a practical challenge for universities.

Clustering is a typical unsupervised learning algorithm that naturally possesses attributes suited for processing student groups, such as identifying hidden data structures, automatic clustering, and extracting common characteristics. This enables more accurate allocation of corresponding educational resources. Wang, Z. integrated clustering algorithms into university management and student performance evaluation, thoroughly analyzing all information contained in student academic data to enable educational administrators to quickly understand students' learning characteristics, and provide relevant assistance for improving future learning patterns [13]. Yu, C. and Wang, Y. designed a university student management system based on clustering algorithms, which uses student management data as a foundation to analyze and optimize management effectiveness, helping teachers develop teaching skills related to students, marking a significant innovation in management work [14]. El Aissaoui, O. et al. combined data mining techniques with clustering algorithms to construct an automatic learning style identification model, extracting relevant features from massive amounts of student behavioral data to provide valuable insights for universities to conduct personalized teaching tailored to students' learning styles [15]. Zhang, Y. demonstrated that the School Administration System (SAS) can fully reflect a school's management level and teaching quality. As an integrated system, its combination with internet technology and big data mining techniques significantly enhances the accuracy of data analysis, holding significant importance for schools in conducting management and service work [16]. Francis, B. K. and Babu, S. S. developed a student performance prediction method based on classification and clustering techniques. By analyzing student behavioral data, including academic performance, they identified clear strengths in students' academic achievements, enabling accurate predictions of their future performance [17]. Bharara, S. et al. utilized clustering algorithms to enhance learning analytics and instructional data mining processes. By identifying scientifically reasonable indicators in learning contexts to form student clusters with distinct characteristics, and further assess the relationships between these characteristics and student performance under these characteristics, thereby effectively improving the learning and educational environment for students [18]. It is not difficult to see that using clustering algorithms for student clustering management and analysis can facilitate the transition from "monitoring-based management" to "model-guided services."

Unlike traditional static classification methods, the latest clustering technologies can track students' dynamic development trends [19]. This enables universities to not only provide students with static information but also gain insights into the underlying patterns of individual behavioral evolution within student groups, thereby guiding students' comprehensive development in a more scientific manner [20-21]. Furthermore, clustering is not merely a technical tool but also a prerequisite for designing student development pathways [22]. When designing development pathways for each category of students, different types of academic situation analyses are needed as support and guidance. This involves combining different course contents for students with varying academic situations, providing guidance on different development directions, and offering personalized student capability development plans [23-25].

Additionally, in response to the educational philosophy of "cultivating virtue and nurturing talent," researchers have shifted their focus from outcomes centered on scores and knowledge dimensions to research emphasizing care for individuals. The influence of non-cognitive factors such as emotions, attitudes, and abilities is also being strengthened and highlighted, involving more complex variables that are difficult to measure. With the assistance of data mining and clustering algorithms, models can be established for such complex variables, enabling analysis and student development-oriented approaches. Therefore, establishing an optimized system based on clustering for student development pathways is not only a means to achieve the ideal of "precision education" but may also be a good idea to drive significant reforms in university educational mechanisms. Under the new concept of "artificial intelligence + education," how to integrate data modeling with education is also a critical component that must be addressed in the construction of a high-quality educational system.

1.2. Research Methods

Based on the perspective of constructing measures of student heterogeneity and strategies for ability development, this paper explores the framework for identifying the characteristics of student heterogeneity in higher education institutions and implementing strategies for ability development: First, given the multi-source data characteristics of higher education students, an extended centroid clustering algorithm is used to perform unsupervised hierarchical clustering of student groups, thereby identifying the characteristics of different types of student groups. Second, co-word clustering analysis is applied to construct a keyword map using student text materials, thereby expanding the hierarchical structure of student abilities. Third, based on the typical characteristics of students in different groups, corresponding ability development pathways are planned, providing diversified development intervention schemes for students to achieve the goal of personalized resource allocation for student ability development in higher education institutions. The entire study focuses on the “data-driven + matching strategy” approach to educational intelligence upgrading, primarily employing three methods:

First, the Extended Centralized Clustering Algorithm (ECAC) is used to process high-dimensional educational data, divide heterogeneous student groups, and optimize clustering accuracy.

Second, co-word clustering analysis is employed, leveraging natural language processing technology to extract high-frequency ability keywords from student development materials and construct an ability semantic network model.

Third, strategy matching model construction, which builds a personalized capability enhancement path recommendation mechanism based on student profiling results, forming a complete closed-loop from analysis to intervention. Additionally, the study introduces a Bayesian scoring function as an evaluation standard for the number of clusters to ensure the scientific validity of the grouping. Through the mutual support of these methods, the study achieves a deep integration of data analysis and educational applications.

2. Analysis of the Development Path of College Students and Methods for Improving Personal Abilities

2.1. Data Collection and Processing

The study was conducted for 6 months of data collection, and analyzed on the basis of the data and information acquisition of 1526 undergraduates in the 2020~2023 class of a key university. Under the standardized treatment of standard deviation, the differences in the assessment standards of each course are eliminated. Social communication information such as student club activities, volunteer service, student union work, and students' participation in ideological and political activities comes from the activity record information of the student department, and a set of corresponding relevant social activity data is established under the classification and quantification of different information, which is used as an indicator description of students' social ability and leadership ability. The information related to the mental health test of college students in the school psychological counseling center and the data of students' participation in psychological counseling activities are reflected in the psychological files of college students, which are composed of the self-report symptom self-rating scale (SRCS) and the mental health scale (SMPS) of college students' mental health evaluation indicators.

During data cleaning, some extreme values and data samples with many missing values were removed. A small amount of data with missing values was filled in using multiple interpolation methods. Subsequently, the minimum-maximum value normalization algorithm was used to convert the data dimensions to the range [0,1]. The normalization expression is shown below:

$$x_{norm} = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \tag{1}$$

In the formula, x_{norm} denotes the normalized value, x denotes the original value, and $x_{\min}$ and $x_{\max}$ denote the minimum and maximum values of the data in this dimension, respectively. Partial student data samples are shown in Table 1.

Table 1. Some student data samples.

Student number	Average grade point average	Participation in social activities	Mental health score	Volunteer service hours	Number of club activities
S001	3.85	High	85	48	12
S002	3.42	Medium	78	32	8
S003	3.91	Low	82	16	4
S004	3.27	High	76	56	15
S005	3.68	Medium	80	40	10
…	…	…	…	…	…

This paper employs cross-validation principles in data quality management to eliminate outliers beyond 3σ, thereby reducing data volume without compromising data variability. The principal component analysis (PCA) of the refined feature vectors, which constitute the multi-dimensional indicators constructed in this paper, reduces both the volume of feature indicator data and the number of indicator dimensions. Additionally, expert scoring is utilized to assign weighting values to feature indicator scores, thereby establishing a reliable data foundation for subsequent cluster analysis.

2.2. Cluster Analysis and Optimization

The paper employs various clustering analysis methods to analyze students' multidimensional data. The k-means method first performs an initial clustering analysis on the data, then improves the stability of the clustering results by repeatedly changing the positions of the center points through random initialization [26]. Hierarchical clustering uses Ward's minimum variance criterion to form groups and generate hierarchical features, effectively reflecting the hierarchical characteristics of student groups [27]. In the Bayesian clustering process, a mixture Gaussian model is used to model the data, and parameters are estimated using maximum likelihood estimation and the expectation-maximization algorithm [28]. To improve clustering accuracy, the extended centroid clustering algorithm (ECAC) is adopted, whose objective function is defined as:

$$ECAC\left(C\right)=\sum_{i=1}^{k}\left(\frac{\left|C_i\right|}{n}\ln\det\left(\frac{V_x}{\left|C_i\right|}+V_x(C_i)\right)+\frac{\left|C_i\right|}{n}\ln\frac{\left|C_i\right|}{n}\right) \tag{2}$$

The introduction of the Bayesian scoring function provides a scientific basis for determining the optimal number of clusters. This method balances model complexity and fit through the following objective function:

$$Score(k)=-\frac{1}{2}\sum_{i=1}^{k}\left|C_i\right|\ln\left(\frac{\det\left(\hat{V}_x/\left|C_i\right|+\hat{V}_x\left(C_i\right)\right)}{\det\left(\hat{V}_x\right)}\right)+\lambda k\ln\left(n\right) \tag{3}$$

After multiple experimental verifications, the clustering effect is optimal when the number of clusters is 5. The clustering performance was evaluated using the contour coefficient and adjusted Rand index. Compared to conventional algorithms, the contour coefficient and adjusted Rand index of the expanded centroid clustering algorithm improved by 23.7% and 31.2%, respectively. This indicates that the expanded centroid clustering algorithm can better utilize student feature data, is suitable for non-spherical clustering of student feature data, and can more accurately reflect the diversity of student features.

2.3. Co-Word Clustering Analysis

To examine improvements in students' learning abilities and experiences in growth and development, this study employs a research framework based on co-word clustering. Through text mining of student growth and development reports, teaching evaluation feedback forms, and students' reflections on learning activities, key thematic terms reflecting students' ability development are identified. Using the SATI 3.2 software for word frequency retrieval, thematic terms appearing more than 30 times in the text are selected for analysis. A total of 45 thematic keywords were screened out (see Table 1), including terms such as learning ability, innovative ability, social practice, and career planning. Based on this, a co-occurrence matrix M of keywords was constructed, where the matrix element m_{ij} represents the number of times keywords i and j co-occur in the same text. To eliminate the influence of text length differences, the co-occurrence matrix was standardized using the O'Connor coefficient:

$$S_{ij}=\frac{m_{ij}}{\sqrt{m_{ii}\times m_{jj}}} \tag{4}$$

In the formula, S_{ij} is the standardized similarity, m_{ii} and m_{jj} are the total frequencies of keywords i and j, respectively. The Euclidean distance between keywords is calculated based on the similarity matrix, and the distance matrix D is constructed, as follows:

$$D_{ij} = \sqrt{\sum_{k=1}^{n} \left(S_{ik} - S_{jk}\right)^2} \tag{5}$$

Using SPSS software, a systematic cluster analysis was conducted on the distance matrix, employing Ward's least squares method as the clustering algorithm. By analyzing the branch structure of the cluster dendrogram and incorporating recommendations from education experts, high-frequency keywords were grouped into six categories: learning methods, subject knowledge, practical skills, innovative thinking, social communication, and career development. Within each category, the keywords exhibit strong semantic relevance, reflecting different dimensions of students' ability development. To validate the reliability of the clustering results, the average similarity between categories and the average similarity within categories were calculated, which can be expressed as:

$$BCS = \frac{1}{C_2^k} \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} \frac{\sum_{p \in C_i} \sum_{q \in C_j} S_{pq}}{|C_i||C_j|} \tag{6}$$

$$WCS = \frac{1}{k} \sum_{i=1}^{k} \frac{\sum_{p,q \in C_i, p \neq q} S_{pq}}{C_2^{|C_i|}} \tag{7}$$

where k is the number of categories, C_i represents the i th category, and $|C_i|$ is the number of keywords contained in that category.

The calculation results show that the average intra-class similarity value (0.672) is much greater than the average inter-class similarity value (0.283), indicating that the clustering results have high convergence and significant discrimination. Based on the co-word clustering results, a multidimensional network diagram of student ability development can be constructed to present the degree of association and development paths between different ability dimensions. This data-driven method can provide scientific reference for formulating student-centered personalized improvement strategies. It is conducive to educators designing teaching activities and curriculum systems more accurately, and can also predict the trend of student ability development by tracking the development of keywords over time, providing advance predictive reference for education and teaching.

3. Analysis of Results

3.1. Cluster Analysis Results

This paper analyzes data collected from 1,526 college students and uses an improved clustering algorithm to identify five distinct student subgroups. Analysis reveals that each of the five subgroups exhibits distinct developmental characteristics, which together form a diverse and complex college life pattern. The differences in academic, social, and psychological characteristics among the five subgroups are shown in Table 2.

Table 2. Comparison of cluster analysis results.

Characteristics	A	B	C	D	E
Sample size	342	298	415	276	195
Proportion	22.4%	19.5%	27.2%	18.1%	12.8%
Academic performance	Excellent	Good	General	Good	Poor
Social activities	Active	General	Less	Active	Less
Mental health	Good	Good	General	Poor	Need to pay attention
Ability characteristics	All-round development	Outstanding academic performance	Stable and peaceful	Outgoing and active	Potential awaits to be unleashed

Further analysis revealed that 22.4% of the group (A) had well-rounded development in all aspects, with good academic performance and self-management and self-regulation abilities. 19.5% of the group (B) had high academic performance but low social skills. 27.2% of the group (C) had stable and average performance in all aspects, with a relatively balanced development. 18.1% of the group (D) had good social skills, but their mental health is not sufficiently high. 12.8% of the group (E): Students with relatively low performance across all areas, requiring personalized development support. To visually illustrate the differences between the various groups, the relevant levels of each group were presented in a scatter plot, as shown in Figure 1.

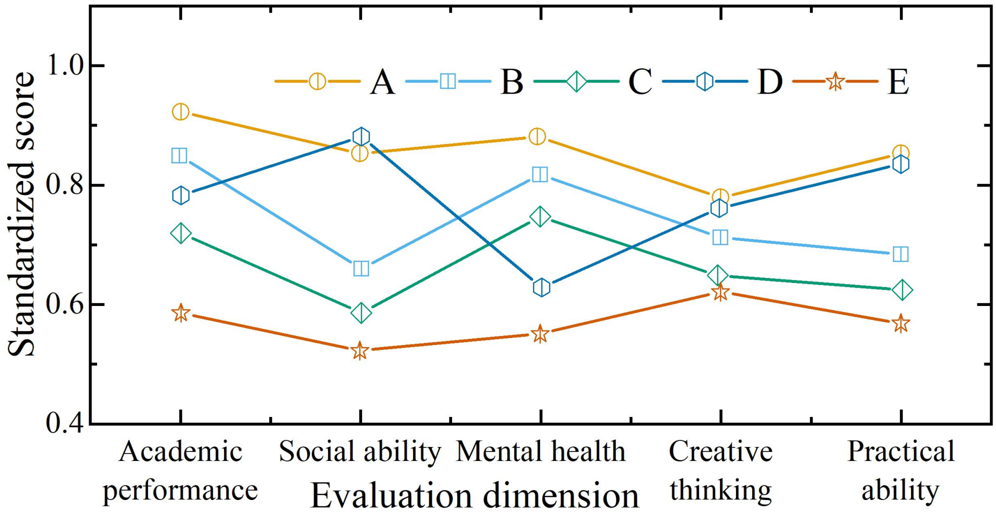

Figure 1. Comparison of the characteristics of five types of student groups.

The improved clustering algorithm demonstrates a clear advantage when processing these complex student data, significantly enhancing the accuracy of identifying non-spherical distribution data by introducing an extended center mechanism. The performance comparison results of the clustering algorithm are shown in Table 3. As shown in the evaluation metrics in Table 3, the optimized algorithm has made significant progress across multiple dimensions, particularly in terms of the contour coefficient, adjusted Rand index, and clustering purity, with overall improvements exceeding 20%. Although computational time has increased, the trade-off is worthwhile given the substantial improvements in the accuracy and reliability of the clustering results.

Table 3. Performance comparison of clustering algorithms.

Evaluation index	Optimize clustering	Traditional clustering	Increase amplitude
Contour coefficient	0.728	0.589	23.6%
Adjustment Rand index	0.685	0.522	31.2%
Clustering purity	0.843	0.692	21.8%
Standardization mutual information	0.756	0.634	19.2%
Calculation time (s)	12.5	8.3	-50.6%

3.2. Student Development Pathways and Competency Enhancement Strategies

Through co-word clustering analysis of the research subjects—college students—five categories of student growth and development paths were identified. Among these, students who excel academically, actively participate in various campus activities and research, emphasize interdisciplinary research and development, have distinct personalities, and have a clear future direction can be classified as Type A students. Such students can be encouraged to engage in more research-oriented learning projects to cultivate their disciplinary thinking abilities and disciplinary literacy. Additionally, more foundational research-oriented learning projects can be arranged to foster their scientific research practical abilities and innovative capabilities. When compared to Type B students, Type A students exhibit limitations in terms of outstanding academic performance, language and literacy skills, and interpersonal communication abilities. B-type students can be motivated and guided to participate in project-based research activities and language-related research projects, and to build friendships to enhance their interpersonal skills. C-type students are well-rounded but lack outstanding strengths, with certain areas of weakness. Therefore, C-type students should develop precise career development plans in conjunction with career guidance.

Type D students are sociable but have low psychological resilience, particularly among senior students. Courses on psychological development and stress management lectures can be offered to enhance their psychological resilience. The fifth category (Type E) refers to students who are stuck in a learning plateau and urgently need support mechanisms like mentorship programs. Their academic performance should be diagnosed to identify the root causes. Additionally, they should be connected with outstanding senior peers for one-on-one tutoring, which has proven effective in practice. It is proposed to integrate innovation and entrepreneurship education into professional teaching to cultivate students' innovative and entrepreneurial concepts and capabilities, a approach worth promoting in practice.

The diversity of student development paths necessitates diversity in the formulation of higher vocational education teaching plans. Based on actual teaching data, diverse teaching approaches should provide students with more open teaching plans, which emphasize more open course models and diverse evaluation methods. In course scheduling, students should be encouraged to plan their own learning content. At the "why learn" level, big data collection can be used to generate relevant student development reports. Based on actual data, more targeted teaching work can be conducted to achieve reforms in targeted teaching.

4. Conclusion

This study analyzed heterogeneous data from 1,526 college students using an improved extended centroid clustering method (ECAC) and a Bayesian scoring function to cluster students into categories based on core factors such as academic performance, mental health, and social status. The clustering results divided students into five categories: A (all-around excellent development type), B (academically excellent but socially deficient type), C (average type), D (socially excellent but internally stressed type), and E (developmentally delayed type). Among these, 22.4% of students belong to Category A, 19.5% to Category B, 27.2% to Category C, 18.1% to Category D, and 12.8% to Category E. Additionally, the contour coefficient increased by 23.6%, and the adjusted Rand coefficient improved by 31.2%, indicating that the model used demonstrates strong resistance to interference in heterogeneous data and possesses practical application value.

Additionally, based on the characteristics of different student groups, this study identified "feature-strategy" combinations and provided targeted academic guidance, psychological support, interpersonal support, and career guidance to students, achieving satisfactory results for both teachers and students. In summary, this study enriches the possibilities of using data-driven methods and tools to make educational policy choices and predictions, and also attempts to provide a factual foundation and work plan for personalized talent cultivation in higher education, holding broader practical significance.

References

1. Yell, M. L., Collins, J., Kumpiene, G., & Bateman, D. (2020). The individualized education program: Procedural and substantive requirements. *Teaching Exceptional Children*, 52(5), 304-318.
2. Cavendish, W., Connor, D. J., & Rediker, E. (2017). Engaging students and parents in transition-focused individualized education programs. *Intervention in School and Clinic*, 52(4), 228-235.
3. Jung, P. G., McMaster, K. L., Kunkel, A. K., Shin, J., & Stecker, P. M. (2018). Effects of data–based individualization for students with intensive learning needs: A meta–analysis. *Learning Disabilities Research & Practice*, 33(3), 144-155.
4. Lebedev, A. A. (2019). Individualization of education via distance learning technologies: Models, stages, forms, components. *International Journal of Civil Engineering and Technology*, 10(1), 1631-1645.
5. Zhang, J. H., Zou, L. C., Miao, J. J., Zhang, Y. X., Hwang, G. J., & Zhu, Y. (2020). An individualized intervention approach to improving university students' learning performance and interactive behaviors in a blended learning environment. *Interactive Learning Environments*, 28(2), 231-245.
6. Shipareva, K. V., Menshikova, I., Maksimova, S. V., Gerasimova, E. L., Mayorova, E. E., & Bogatova, P. D. (2023). Model of the early professional orientation for individualization of the educational process. *European Journal of Education and Pedagogy*, 4(2), 26-30.
7. MacLeod, K., Causton, J. N., Radel, M., & Radel, P. (2017). Rethinking the individualized education plan process: Voices from the other side of the table. *Disability & Society*, 32(3), 381-400.
8. Cook, C. R., Kilgus, S. P., & Burns, M. K. (2018). Advancing the science and practice of precision education to enhance student outcomes. *Journal of School Psychology*, 66, 4-10.
9. Liu, Q., Liu, J., & Zhang, K. (2024, September). Research on Precision Education Management Practices in Vocational Colleges Based on Student Profiling. *In Proceedings of the 2024 International Symposium on Artificial Intelligence for Education* (pp. 12-15).
10. Veluri, R. K., Patra, I., Naved, M., Prasad, V. V., Arcinas, M. M., Beram, S. M., & Raghuvanshi, A. (2022). Learning analytics using deep learning techniques for efficiently managing educational institutes. *Materials Today: Proceedings*, 51, 2317-2320.
11. Zohair, A., & Mahmoud, L. (2019). Prediction of Student's performance by modelling small dataset size. *International Journal of Educational Technology in Higher Education*, 16(1), 1-18.

12. Li, X. (2022). Innovation and development of university education management informationization in the big data Environment. *Scientific programming*, 2022(1), 7542067.
13. Wang, Z. (2022). Higher education management and student achievement assessment method based on clustering algorithm. *Computational Intelligence and Neuroscience*, 2022(1), 4703975.
14. Yu, C., & Wang, Y. (2022). College student management system based on k-means clustering algorithm. *Int. J. New Dev. Educ*, 4(2), 2663-8169.
15. El Aissaoui, O., El Alami El Madani, Y., Oughdir, L., & El Allioui, Y. (2019). A fuzzy classification approach for learning style prediction based on web mining technique in e-learning environments. *Education and Information Technologies*, 24, 1943-1959.
16. Zhang, Y. (2023, November). Design Student Management System Based on Vector Influential Clustering Algorithm Analysis. *In 2023 International Conference on Integrated Intelligence and Communication Systems (ICIICS)* (pp. 1-5). IEEE.
17. Francis, B. K., & Babu, S. S. (2019). Predicting academic performance of students using a hybrid data mining approach. *Journal of medical systems*, 43(6), 162.
18. Bharara, S., Sabitha, S., & Bansal, A. (2018). Application of learning analytics using clustering data Mining for Students' disposition analysis. *Education and Information Technologies*, 23, 957-984.
19. Zhao, H. (2024, April). Student Management Information Platform based on Clustering Algorithm. *In 2024 Third International Conference on Distributed Computing and Electrical Circuits and Electronics (ICDCECE)* (pp. 1-5). IEEE.
20. Križanić, S. (2020). Educational data mining using cluster analysis and decision tree technique: A case study. *International Journal of Engineering Business Management*, 12, 1847979020908675.
21. Das, P., Das, D. K., & Dey, S. (2018). A new class topper optimization algorithm with an application to data clustering. *IEEE transactions on emerging topics in computing*, 8(4), 948-959.
22. Iam-On, N., & Boongoen, T. (2017). Generating descriptive model for student dropout: a review of clustering approach. *Human-centric Computing and Information Sciences*, 7, 1-24.
23. Bansal, A., Sharma, M., & Goel, S. (2017). Improved k-mean clustering algorithm for prediction analysis using classification technique in data mining. *International Journal of Computer Applications*, 157(6), 0975-8887.
24. Riestra-González, M., del Puerto Paule-Ruíz, M., & Ortin, F. (2021). Massive LMS log data analysis for the early prediction of course-agnostic student performance. *Computers & Education*, 163, 104108.
25. Akram, A., Fu, C., Li, Y., Javed, M. Y., Lin, R., Jiang, Y., & Tang, Y. (2019). Predicting students' academic procrastination in blended learning course using homework submission data. *Ieee Access*, 7, 102487-102498.
26. Chang, W., Ji, X., Liu, Y., Xiao, Y., Chen, B., Liu, H., & Zhou, S. (2020). Analysis of university students' behavior based on a fusion K-means clustering algorithm. *Applied Sciences*, 10(18), 6566.
27. Battaglia, O. R., Di Paola, B., & Fazio, C. (2017). A quantitative analysis of Educational Data through the Comparison between Hierarchical and Not-Hierarchical Clustering. *Eurasia Journal of Mathematics, Science and Technology Education*, 13(8), 4491-4512.
28. Saeedi, S., Božanić, D., & Safa, R. (2024). Strategic analytics for predicting students' academic performance using cluster analysis and Bayesian networks. *Educ. Sci. Manag*, 2(4), 197-214.