International Journal for Housing Science and Its Applications
Publish September 9, 2025. Volume 47, Issue 1 Pages 511-518

https://doi.org/10.70517/ijhsa47143

A Study on the Effectiveness of Ideological and Political Education Based on Learning Analytics and Behavior Tracking in a Data-Driven Context

Jiani Pan[1,*]
[1] Wuxi City College of Vocational Technology, Wuxi, Jiangsu, 214000, China
Corresponding authors: (e-mail: panjianiyzl@163.com).

Abstract As teaching models increasingly integrate with the big data era, data-driven and technology-enabled approaches have become important tools for evaluating the effectiveness of ideological and political education. This paper explores reliable methods for assessing the effectiveness of ideological and political education primarily through learning analytics and behavior tracking. The research data was preprocessed by merging data, handling missing and abnormal cases, and converting features. The Gaussian mixture model clustering algorithm is selected, and a probabilistic model is introduced to handle complex distributions of various student behavior data, achieving clustering of student behavior data. Based on student learning activity sequences and cognitive styles during the ideological and political education process, a learning behavior model is constructed to analyze the quality of the ideological and political education process based on information from the learning process. Taking second-year students from a certain major at University E as the research subjects, after applying Gaussian mixture clustering to their ideological and political-related learning data, the learning behavior model proposed in this paper achieved a prediction accuracy of 0.8839 for student learning behavior and performance.

Index Terms ideological and political education, learning behavior model, Gaussian mixture clustering, learning activity sequence

I. Introduction

Looking back at the development of ideological and political education, the focus on its effectiveness has been one of the key driving forces behind its continuous development and holds enduring significance in the theoretical exploration of ideological and political education. In today's rapidly evolving landscape of information technology and the internet, the volume of information and data is growing exponentially. Against this backdrop of rapid informatization, traditional ideological and political education has begun to reveal its limitations in the new era, thereby constraining its effectiveness [1]-[3]. Additionally, the infiltration of Western capitalist ideology has become more covert, posing a significant challenge to both the development of ideological and political education and the construction of the national socialist ideology [4]-[6]. Against this backdrop, there is an increasing need for big data to play a role in enhancing the effectiveness of ideological and political education [7]. Educational administrators can uncover the immense value hidden within massive datasets through data mining and processing, thereby enhancing the effectiveness of ideological and political education services [8], [9].

In the process of China's modernization, ideological and political education must adapt to circumstances, evolve with the times, and innovate in response to trends. Therefore, the application of big data technology in ideological and political education and the digitalization of such education will be the primary direction for the foreseeable future [10]. To achieve full integration between the two, leveraging their respective strengths and combining the value advantages of ideological and political education with the technical advantages of big data analysis, it is necessary to position ideological and political education as the value-driven core, utilize data-driven analytical methods as new tools, and adopt a new "big data" perspective to understand and transform the world [11]-[14]. For example, data-driven technologies such as data mining and machine learning can be used to conduct in-depth exploration and analysis of students' learning behaviors and ideological dynamics, uncovering underlying patterns and trends [15]-[17]. Data-driven analytical tools can assist educators in more intuitively understanding data, thereby providing technical support for decision-making and practice in ideological and political education [18], [19]. Ultimately, this will achieve self-reform in ideological and political education, complete its digital transformation, promote the construction of a modernized ideological and political education system, and lead the development of Chinese-style modernization [20], [21].

This paper first combines the characteristics of ideological and political education course teaching to explain the merging, missing, and abnormal handling of student behavior data, as well as the conversion process. It then introduces the Gaussian mixture model clustering algorithm, detailing its basic principles and algorithmic steps. Based on the sequence of students' learning activities and cognitive styles in ideological and political education courses, relevant research parameters and variables are defined, and a model for processing students' activity sequences and determining learning styles is designed. Subsequently, research subjects are selected, research data is collected to form a research sample, and Gaussian mixture clustering analysis is conducted on the sample. Finally, parameter optimization analysis of the proposed learning behavior model is performed, along with a comparative assessment of its predictive performance.

II. Data preprocessing

II. A. Data Merging

The primary purpose of data merging is to integrate and summarize various types of data from different sources or tables to form a more comprehensive and accurate data view, laying the foundation for better understanding the data and identifying correlations and trends. Using functions such as merge, join, and concat in the pandas library, we merged the key content from four tables compiled from different platforms and channels. Among these, the final exam results include all questions from the exam and students' responses. Based on research foundations in the field of ideological and political education, the 55 questions were categorized into four groups based on content, resulting in 12 dimensions of student learning behavior and exam performance data. The dimensions of student learning behavior and exam performance data for ideological and political education courses are as follows:

(1) Course overall evaluation: average homework score.
(2) Final exam score: Post-class video learning duration.
(3) Exam duration: Performance on the first historical period.
(4) Midterm exam score: Performance on the second historical period.
(5) Attendance performance: Performance on the third historical period.
(6) Classroom performance: Performance on the fourth historical period.

II. B. Missing and Abnormal Handling

Since the selected data comes from an online teaching platform used by students to aid their learning, it is necessary to conduct a thorough check after merging the data to identify and correct various issues such as missing information, type errors, data anomalies, and duplicate content. For example, the midterm scores of students who did not participate in the midterm exam should be set to zero, and the data of students who did not take the final exam should be deleted.

II. C. Data Conversion

Data with different features vary in length due to differences in their meaning and representation. To further improve the performance and stability of machine learning models, it is necessary to use normalization methods to scale the feature values of the data to a uniform scale. All data in this paper have relatively fixed maximum and minimum values, so it is appropriate to use Min-Max normalization, as shown in Equation (1), to linearly scale all data to the range [0,1].

$$X' = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \tag{1}$$

III. Clustering and prediction of learning behavior

III. A. Gaussian Mixture Clustering

Gaussian Mixture Model (GMM) clustering is a clustering algorithm based on a probabilistic model, which assumes that all data points are generated from a mixture of a finite number of Gaussian distributions. Unlike hard clustering algorithms such as K-Means, GMM is a typical soft clustering method, providing each data point with the probability of belonging to each cluster.

III. A. 1) Basic Principles

The basic principle is to treat each sample in the dataset as a mixture of multiple Gaussian distributions, with each Gaussian distribution corresponding to a cluster center. The Gaussian mixture model fits the Gaussian mixture model by maximizing the expected value of the log-likelihood function between the sample data and the model prediction. The Expectation-Maximization (EM) algorithm is used to iteratively update the model parameters, including the mean, covariance matrix, and weights of the Gaussian distributions.

The goal of the Gaussian mixture clustering algorithm is to maximize the log-likelihood function, which is the sum of the logarithms of the probabilities of assigning all data points to clusters. The log-likelihood function is expressed as in Equation (2):

$$\log P\left(X \mid \theta\right)=\sum_{i=1}^{N} \log \sum_{k=1}^{K} P\left(x_i \mid z_i=k\right) P\left(z_i=k\right) \tag{2}$$

The parameter optimization process for the Gaussian mixture clustering algorithm typically employs the expectation-maximization (EM) algorithm, which consists of E steps and M steps:

In the E step, for each data point x_i, the posterior probability $P(z_i=k \mid x_i)$ of it belonging to each cluster is calculated, i.e., the probability of data point x_i coming from each cluster is calculated as shown in Equation (3):

$$P(z_i=k \mid x_i)=\frac{P(x_i \mid z_i=k) P(z_i=k)}{\sum_{j=1}^{K} P(x_i \mid z_i=j) P(z_i=j)} \tag{3}$$

Among these, $P(x_i \mid z_i=k)$ denotes the probability density function of the Gaussian distribution when x_i belongs to cluster k, and $P(z_i=k)$ denotes the weight of cluster k. By calculating the posterior probability of each data point belonging to each cluster, we can determine the cluster with the highest probability and assign the data point to the corresponding cluster.

In step M, re-estimate the mean vector, covariance matrix, and weight of each cluster. Specifically, for cluster k, calculate its mean vector μ_k, covariance matrix Σ_k, and weight α_k as shown in equations (4)-(6):

$$\mu_k=\frac{\sum_{i=1}^{N} P(z_i=k \mid x_i) x_i}{\sum_{i=1}^{N} P(z_i=k \mid x_i)} \tag{4}$$

$$\Sigma_k=\frac{\sum_{i=1}^{N} P(z_i=k \mid x_i)(x_i-\mu_k)(x_i-\mu_k)^T}{\sum_{i=1}^{N} P(z_i=k \mid x_i)} \tag{5}$$

$$\alpha_k=\frac{\sum_{i=1}^{N} P\left(z_i=k \mid x_i\right)}{N} \tag{6}$$

Here, N denotes the total number of data points.

Repeat steps E and M until convergence. The convergence condition of the algorithm can be that the increment of the log-likelihood function is less than a certain threshold, or that the changes in the mean vector, covariance matrix, and weights of each cluster are less than a certain threshold.

III. A. 2) Algorithm Flow

The core idea of Gaussian mixture clustering is to fit a Gaussian mixture model through an iterative optimization algorithm called EM, and update the cluster center parameters, which include the mean, covariance matrix, and weights. In each iteration, the algorithm calculates the probability of each sample point belonging to each cluster based on the current cluster center parameters, and updates the cluster center parameters based on these probabilities. Through continuous iteration, the algorithm eventually converges to the optimal cluster center parameters. The specific steps are as follows:

(1) Set the number of k, i.e., initialize the number of components in the Gaussian mixture model.

(2) Initialize the Gaussian distribution model parameters for each cluster.

(3) Calculate the probability of each data point belonging to each Gaussian model, i.e., calculate the posterior probability.

(4) Calculate the parameters that maximize the probability of the data points, using the weighted sum of the data point probabilities to compute these new parameters. The weights are the probabilities of the data points belonging to that cluster.

(5) Repeat steps (3) and (4) until convergence.

III. B. Construction of a learning behavior model

The modeling of learning behavior based on activity theory offers two primary benefits for teaching practice: first, it centers on learning activities, providing a systematic framework model for instructional design and resource utilization; second, it promotes the standardization of teaching and enhances the universality of instructional design. The learning behavior-style model constructed in this paper is a structural representation of students' learning activity sequences and cognitive styles. The model can describe general learning activities, providing constructive references for teachers in developing instructional plans guided by constructivist principles, selecting instructional tools, and conducting learning activities. Based on this model, different learning sequences of students can be described, and their distinct learning styles identified, thereby enabling adjustments to instructional methods.

By statistically analyzing students' learning behaviors, such as their enthusiasm and participation frequency in different types of activities, their learning styles can be determined. To better apply activity theory and the characteristics of activity sequences to analyze learning behaviors, this paper establishes a learning behavior model based on activity sequences.

(1) Model Overview

From an information processing perspective, the essence of the learning behavior model is to analyze and process learning activity information, i.e., activity sequences, and output result information. The information flow of the learning behavior model is shown in Figure 1.

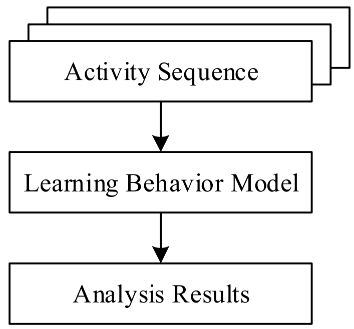

Figure 1: The information flow of the learning behavior model

As shown in Figure 1, the input of the learning behavior model is one or more activity sequences. After analyzing and processing the activity sequences, the output is the direct reference data required to determine students' learning styles.

(2) Design Concept

The order in which students engage in learning activities reflects their learning preferences. This study is based on the following premise: in a series of learning activities performed by students, the first learning activity is generally the one they prefer the most, the next learning activity is generally the one they prefer the second most, and the last learning activity is generally the one they do not prefer. Therefore, by segmenting the series of learning activities performed by students according to their preference levels and analyzing the distribution of different learning activities in each segment, we can obtain the students' learning preferences.

(3) Parameter Definitions

First, we define the parameters and variables used in this section.

a) Activity: A single learning action, denoted by A.

b) Activity sequence: A set of sequential activities within a certain period of time, denoted by S. An activity sequence containing n activities can be represented as $S:\{A_1, A_2, \cdots, A_n\}$.

c) Activity category: A set of activities with common characteristics, denoted by T. An activity category containing k activities can be represented as $T:\{A_1, A_2, \cdots, A_k\}$ Different activity categories do not intersect, i.e., $\forall A \in T_1$, $T_1 \neq T_2 \Rightarrow A \notin T_2$.

d) Activity Priority P: This is a division of an activity sequence. An activity sequence is divided into several segments, each with a different activity priority. Segments closer to the start of the activity sequence have higher priority. That is, activities in segments with higher activity priority are performed before activities in segments with lower activity priority.

e) Activity Priority Weight W: This is the numerical representation of activity priority. Activities with higher activity priorities have greater weights than those with lower activity priorities, and all activities within the same activity priority share the weight of that activity priority equally.

f) Category Score R: Represents the total sum of weights obtained by all activities within an activity category in an activity sequence.

g) Category Score Index I: Represents the proportion of a category score relative to the total sum of all category scores. This serves as a direct basis for assessing a student's learning style.

(4) Activity Sequence Processing Flow

a) Activity Sequence Segmentation

For an input activity sequence containing n activities, the learning behavior model first segments the sequence based on activity priority and assigns each segment a corresponding activity priority weight. Assuming there are m predefined priorities, the segmentation of the activity sequence is shown in Figure 2, where the sequence is divided into m segments; the number of activities contained in each segment are respectively $L_1, L_2, \cdots, L_m$; the activity priorities of each segment are respectively $P_1, P_2, \cdots, P_m$; the corresponding priority weights are $W_1, W_2, \cdots, W_m$, and $P_1 > P_2 > \cdots > P_3$ and $W_1 > W_2 > \cdots > W_3$.

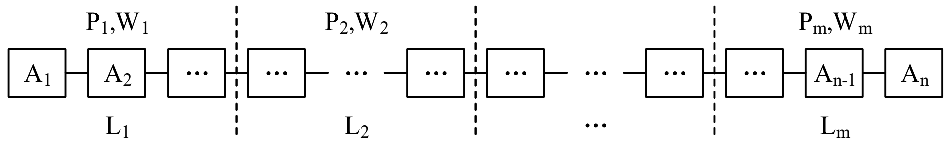

Figure 2: Segmentation of the activity sequence

b) Calculate the category scores R for each activity category

Assume that the activity sequence contains k activity categories, denoted as $T_1, T_2, \cdots, T_k$. There are q_1 activities belonging to T_j in P_1, q_2 activities belonging to T_j in P_2, …, q_m activities in P_m belong to T_j. Then, the category score R_j for the j th activity category in this activity sequence is given by equation (7):

$$R_j = \sum_{u=1}^{m} \frac{q_u}{L_u} \cdot W_u \tag{7}$$

c) Calculate the category score index I for each activity category.

Based on the definition of the category score index I, the category score index for each activity category can be calculated. Assuming there are a total of k activity categories, with their respective category scores being $R_1, R_2, \cdots, R_k$, then the category score index I_j for the j th activity category is given by equation (8):

$$I_j = \frac{R_j}{\sum_{w=1}^{k} R_w} \tag{8}$$

d) Output the category score index for each activity category

Output the category score index for each category as the basis for determining the learning style of the activity sequence. When conducting a comprehensive learning style analysis of multiple activity sequences, the category score for each activity category is the sum of its category scores across all activity sequences.

(5) Determination of learning style

This learning behavior model analyzes the category score indices for each activity category and determines the student's learning style based on the predefined category score index range.

IV. Application and verification of learning behavior models

Second-year students majoring in a certain subject at University E were selected as the research subjects. Six sets of data were collected from the subjects during the semester: (A1) overall course evaluation, (A2) final exam scores, (A3) exam time, (A4) midterm exam scores, (A5) attendance records, and (A6) course performance. After data consolidation and missing value handling, the dataset used in this study (E-ipl) was obtained.

IV. A. Gaussian Mixture Clustering Analysis

After data transformation of the six data performance metrics in the E-ipl dataset, three indicators—(B1) normalized course performance, (B2) normalized daily average performance, and (B3) normalized course assessment scores—were obtained for cluster analysis.

After determining the optimal number of clusters K=3 using the silhouette coefficient method, the six proposed data metrics were imported into SPSS software for Gaussian mixture clustering analysis, with the number of cluster centers set to 3. After 20 iterations, the final cluster centers for the normalized values are shown in Table 1.

Table 1: Cluster centers (normalized values)

Cluster	Cluster1	Cluster2	Cluster3
B1	0.5489	0.0281	0.0267
B2	0.0077	0.0423	0.0479
B3	0.95	0.72	0.97

Convert the normalized values back to their original form to obtain the behavioral feature data of each cluster learner, as shown in Table 2, i.e., the final cluster centers of the converted values.

Table 2: Cluster centers (conversion values)

Cluster	Cluster1	Cluster2	Cluster3
A1	81.64	70.23	62.58
A2	79.26	69.24	60.47
A3	75.71	88.71	92.63
A4	83.92	75.43	68.95
A5	9.47	7.89	6.21
A6	9.03	7.56	6.54

The visualization of the Gaussian mixture clustering results is shown in Figure 3, where the X-axis represents course performance, the Y-axis represents daily performance, and the Z-axis represents course assessment scores. The three cluster centers in the figure do not overlap significantly, indicating that the Gaussian mixture clustering model can effectively classify student behavior in ideological and political education classes.

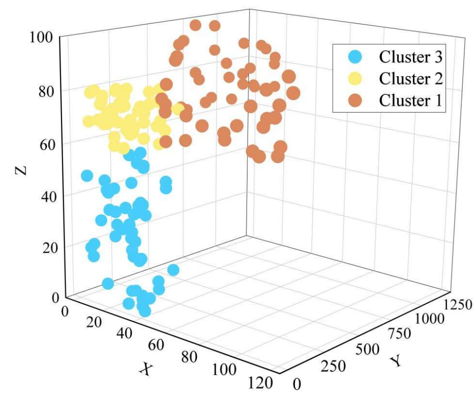

Figure 3: Scatter plot of learning behavior

IV. B. Performance Evaluation of Learning Behavior Models

IV. B. 1) Parameter optimization analysis of the model

The probability $p(\mathrm{T})$ is used to describe the learner's knowledge level, ranging from not yet mastered to already mastered. To optimize the parameters of the proposed model, a learning curve approach is adopted to find the optimal $p(\mathrm{T})$. Three data points (D1-D3) were randomly selected from the experimental dataset, and their change curves are shown in Figure 4. It can be observed that the prediction accuracy of the D1-D2 data points gradually decreases to around 0.45 as the $p(\mathrm{T})$ parameter increases. That is, if the $p(\mathrm{T})$ parameter is set improperly, it will significantly impact the model's prediction accuracy.

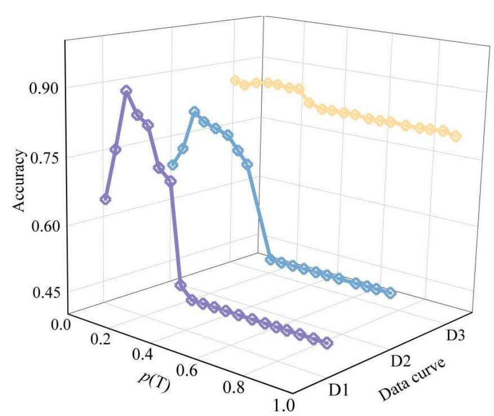

Figure 4: The influence of $p(\mathrm{T})$ parameters on model performance

For all datasets of the experimental subjects, the distribution of prediction accuracy for the (C1) unoptimized model and the (C2) optimized model is shown in Figure 5. The prediction accuracy of the (C1) unoptimized model primarily ranges between 0.0 and 0.7, while the (C2) optimized model is concentrated between 0.7 and 1.0. Over 40.00% of the prediction accuracy falls within 0.9, indicating that the optimized model achieves higher prediction accuracy, thereby significantly improving overall prediction performance.

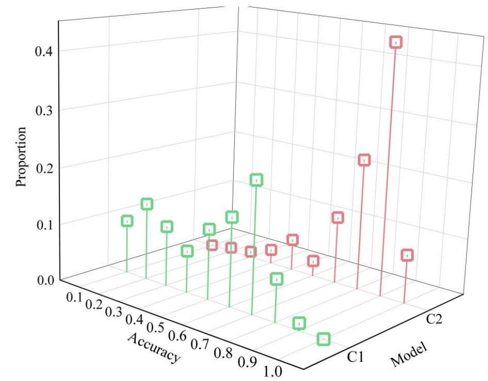

Figure 5: Comparison of prediction accuracy distribution

IV. B. 2) Analysis of Prediction Results

This section compares the predictive performance of the proposed learning behavior models under different clustering methods. The experiment still uses the E-ipl dataset of the ideological and political classroom performance of second-year students in a certain major at E University, with the (F1) learning behavior model, which does not use any clustering methods to assist in predicting student behavior, as the reference. The proposed learning behavior models are compared with the (F2) K-Means clustering method, (F3) hierarchical clustering method, and (F5) Gaussian mixture clustering method, as shown in Table 3. The evaluation metrics used are: prediction accuracy (ACC), root mean square error (RMSE), and AUC.

Table 3: Comparison of model prediction results

Evaluation index	ACC	RMSE	AUC
F1	0.7875	0.5681	0.7078
F2	0.8035	0.4723	0.7188
F3	0.8398	0.4265	0.7709
F4	0.8839	0.3275	0.8758

Overall, the learning behavior model that did not utilize any clustering methods for prediction had the lowest prediction accuracy (0.7875). This is because clustering methods can fully account for the individual differences in knowledge levels among students, thereby effectively improving the prediction accuracy of the learning behavior model for students' learning behavior and performance in ideological and political courses. Among the three clustering algorithms, the Gaussian mixture clustering method selected in this study not only achieved a prediction accuracy of 0.8839 but also had a root mean square error of only 0.3275. This validates that the Gaussian mixture clustering method has the ability to handle complex distribution data, enhancing the model's smoothness and consistency. Therefore, the learning behavior model under the Gaussian mixture clustering method is more suitable for tracking and evaluating the learning behavior and performance of students in ideological and political education courses.

V. Conclusion

This paper, based on the characteristics of ideological and political education, employs the Gaussian mixture model clustering algorithm as a tool for clustering analysis of student learning behavior data. It establishes a learning behavior model to conduct learning analysis and behavior tracking for different students, thereby providing insights into the effectiveness of ideological and political education. For the ideological and political education dataset after Gaussian mixture clustering, the proposed learning behavior model achieves a prediction accuracy of 0.8839 for student behavior, with an error as low as 0.3275. This method, based on a large amount of teaching fact data, enables precise classification and even prediction of student learning and behavioral performance, providing reliable data support and references for the optimization and innovation of ideological and political education teaching, thereby effectively promoting the high-quality development of teaching.

Funding

General Project of Philosophy and Social Sciences Research in Jiangsu Universities in 2025 (Ideological and Political Special Project) (Project Number: 2025SJSZ0442); Title: Exploration of Promoting the Collaborative Education Model of "Great Ideology and Politics" in Higher Vocational Colleges under the Concept of Moral Education.

References

[1] Hu, Z., & Li, J. (2018). Innovative methods for ideological and political education of college students. Educational Sciences: Theory & Practice, 18(5).

[2] Dong, W., Gao, W., Zhang, S., & Ma, X. (2017). Effectiveness of ideological and political education in colleges and universities under the outbreak of internet media. Eurasia Journal of Mathematics, Science and Technology Education, 13(11), 7605-7610.

[3] Gao, H. W. (2023). Innovation and development of ideological and political education in colleges and universities in the network era. International Journal of Electrical Engineering & Education, 60(2_suppl), 489-499.

[4] Wang, P. (2020). Realization of the effectiveness of discourse in ideological and political education. World Scientific Research Journal, 6(4), 308-311.

[5] Wu, X. (2020, October). Research on the innovation of ideological and political education in universities in the era of big data. In Proceedings of the 2020 International Conference on Computers, Information Processing and Advanced Education (pp. 336-340).

[6] Zhang, Z., & Zhang, B. (2023, June). Research on Ideological and political precision Teaching Model in Colleges and Universities driven by Big Data. In 2023 8th International Conference on Information Systems Engineering (ICISE) (pp. 325-328). IEEE.

[7] Li, N., & Yi, Z. (2020, December). Research on the Effectiveness of Students' Ideological and Political Education Based on Internet Information Technology. In 2020 5th International Conference on Mechanical, Control and Computer Engineering (ICMCCE) (pp. 1899-1903). IEEE.

[8] Hou, H. (2023, July). Big Data Analysis of College Students' Daily Behavior and Ideological and Political Behavior. In Proceedings of the 2023 2nd International Conference on Educational Innovation and Multimedia Technology (EIMT 2023) (Vol. 8, p. 327). Springer Nature.

[9] Zheng, J., Shi, B., & Liu, L. (2025, April). Innovative Application of Big Data Technology in Ideological and Political Education in University: A Text Mining Approach through Online Forum Analysis. In Proceedings of the 2025 International Conference on Artificial Intelligence and Educational Systems (pp. 318-325).

[10] Yu, L., & Qi, W. (2024, December). Assessing the Effectiveness of Ideological and Political Education Based on Big Data. In 2024 7th International Conference on Humanities Education and Social Sciences (ICHESS 2024) (pp. 93-101). Atlantis Press.

[11] Hou, L. L., & Lin, J. (2018). Application of big data in ideological and political education. DEStech Transactions on Social Science Education and Human Science.

[12] Liu, S. (2020, October). Research on university ideological and political education based on big data information teaching mode. In 2020 international conference on computers, information processing and advanced education (CIPAE) (pp. 280-285). IEEE.

[13] Qingyan, M., & Zhenli, C. (2025). Precise Push and Personalized Guidance of Ideological and Political Education in Colleges and Universities Based on Big Data Technology. International Journal of High Speed Electronics and Systems, 2540608.

[14] Feng, X., & Yuchen, Y. (2024). The challenges and countermeasures of college students' ideological and political course education under the background of big data. Adult High. Educ, 6(3), 14-20.

[15] Ning, S. (2021, October). Analysis of college students' behavior based on machine learning and big data technology. In International Conference on Machine Learning and Big Data Analytics for IoT Security and Privacy (pp. 145-151). Cham: Springer International Publishing.

[16] Jiang, W., & Yuankun, R. (2023, August). Dynamic Early Warning System of Ideology Education Based on Psychological Big Data Analysis. In EAI International Conference, BigIoT-EDU (pp. 326-333). Cham: Springer Nature Switzerland.

[17] Xian, L. I., & Rang-xin, W. A. N. G. (2015). Behavioral Logic's Analysis on Promoting Network Ideological and Political Education. Journal of University of Electronic Science and Technology of China (SOCIAL SCIENCES EDITION), 17(4), 108-112.

[18] Schildkamp, K. (2019). Data-based decision-making for school improvement: Research insights and gaps. Educational research, 61(3), 257-273.

[19] Agasisti, T., & Bowers, A. J. (2017). Data analytics and decision making in education: towards the educational data scientist as a key actor in schools and higher education institutions. In Handbook of contemporary education economics (pp. 184-210). Edward Elgar Publishing.

[20] Yang, F., Wang, F., & Wei, Z. (2024). Research on Curriculum Ideological and Political Reform of Data Visualization under Background of New Engineering. Frontiers in Educational Research, 7(4), 75-80.

[21] Bach, B., Keck, M., Rajabiyazdi, F., Losev, T., Meirelles, I., Dykes, J., ... & Carpendale, S. (2023). Challenges and opportunities in data visualization education: A call to action. IEEE Transactions on visualization and computer graphics, 30(1), 649-660.