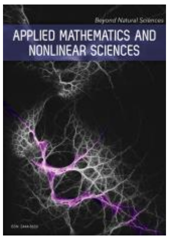
Applied Mathematics and Nonlinear Sciences, 9(1) (2024) 1-17



Applied Mathematics and Nonlinear Sciences

https://www.sciendo.com

Exploring the Potential Application of Big Data in the Digital Transformation Practice of English Education

Lingling Ding[1,†]

1. School of Western Languages and Cultures, Harbin Normal University, Harbin, Heilongjiang, 150025, China.

Submission Info

Communicated by Z. Sabir
Received July 3, 2024
Accepted October 9, 2024
Available online November 27, 2024

Abstract

With the increasing development of the digital economy and Internet technology, in order to adapt to the demand for teaching, many colleges and universities are changing their education mode through digital transformation. Starting from English teaching, this paper seeks to uncover a suitable way for the digital transformation of English education. Based on a collaborative filtering recommendation algorithm, this paper constructs an English teaching model that integrates personality recommendations. The model can use clustering algorithms to classify learners into different groups, which facilitates the subsequent implementation of personalized recommendations. The performance comparison and examples of English teaching in colleges and universities are used in this paper to analyze the efficient performance of the recommendation model and evaluate its usefulness in the smart classroom. The recommendation model designed in this paper divides learners into five clusters, which provides strong support for us to provide personalized recommendations of English learning resources in the future. This paper's system, for example, in basic English, has a recommendation accuracy rate of 10% and 21% higher than reference systems 1 and 2, according to performance comparison. Applying this paper's recommendation model in an efficient and intelligent English teaching classroom, after the experiment, the performance of class T improved by 4.29 points compared to class CK. To conclude, the recommendation system of this paper is very adaptable to the smart classroom and has a greater role in promoting the digital transformation of English teaching mode in colleges and universities.

Keywords: Collaborative Filtering; Cluster Analysis; Personalized Recommendation; English Education.
AMS 2010 codes: 68T05

†Corresponding author.
Email address: 13936000874@163.com

ISSN 2444-8656
https://doi.org/10.2478/amns-2024-3508

1 Introduction

English teaching is an important part of cultivating applied talents and has always been a compulsory course in Chinese schools. English teaching in China has long been characterized by teaching concepts, teaching modes, curricula and teaching methods. Under such an English teaching environment, most students in China do not have a strong ability to apply English comprehensively, and their mastery of English is not comprehensive enough. The single teaching mode causes students' lack of learning motivation and poor performance in English learning, which leads to some students' anorexia towards English [1-4].

English teaching under the background of big data pays more attention to the combination of classroom teaching and extracurricular teaching and satisfies the students' demand for knowledge with a more open classroom teaching mechanism. Under the background of big data, the core of China's education reform is to cultivate students into internationally competitive talents, which requires students to accurately master English and other foreign languages [5-8]. English digital teaching is to strengthen the practicality of English teaching, students through the network platform and other educational media, to improve the comprehensive application of English, especially to strengthen the training and cultivation of English language communication skills [9-11]. Under the leadership of big data, traditional English teaching breaks the old model of teachers, classrooms and blackboards and uses digital facilities to integrate the teaching environment, such as listening and speaking, into the real English environment so that students can face foreign professors intuitively, feel the pure English teaching atmosphere, and truly synchronize their learning and development with the world. This kind of English teaching environment is crucial for Chinese students to learn English and cope with the global competition for talent [12-15].

Literature [16] in order to test the level of preparation of English teachers to introduce technology into their teaching activities, a framework of knowledge of technology content teaching was used to investigate, and after analyzing the results, it was shown that there was no difference between male and female STs in the preparation of integrating technology into the English classroom. Literature [17] proposed a system that enables resource management in distance English education based on big data mining techniques. It was shown through practice that the system helps to improve learning outcomes and also proves the effectiveness of big data technology in managing resources in distance English education. Literature [18] introduces the definition of e-textbooks and analyzes the advantages and disadvantages of e-textbooks by taking high school English textbooks as an example. The impact of digital elements on textbooks and e-learning is discussed, and the results point out that Chinese textbooks have not yet been digitized, but e-learning has become an important part of English teaching. Literature [19] studied the impact of digitization on language education. Applying the systematic codification method to analyze academic data, the results specified that digitization of language education has great differences in terms of types of studies and language skills, while there are no statistically significant differences in terms of types of schools. Literature [20] examines the characteristics of digital education in the framework of foreign language teaching and describes the components of the digital system of teaching foreign languages, as well as the advantages and disadvantages of the psychological, organizational and pedagogical nature of the system. Literature [21] addresses the use of digital solutions in foreign language learning by considering a specific curriculum. It is emphasized that digitalization has clear advantages in the field of education, but it needs to be guided by the classroom in order to achieve student motivation. And digitization is not equal to a completely new way of teaching. Literature [22] proposed an empirical analysis model of university English teaching reform model based on computer big data. Through analysis, it is concluded that the method can be effectively used to evaluate the efficiency of university English reform, and it is highly accurate, which is conducive to promoting the optimization of teaching reform.

The deep integration of technology is crucial for human society's progress, and human civilization is closely tied to the development of educational activities. In recent years, more and more research has focused on the digital transformation of education, hoping to provide scholars with more convenient and efficient learning methods using big data. Based on English teaching, this paper analyzes the learning styles and abilities of English learners in the context of big data. Based on the collaborative filtering personalized recommendation algorithm, a personalized recommendation model for English teaching has been developed. The recommendation model collects learners' learning data and personal information and divides them into multiple clusters by clustering. The objective is to provide English learning resources that are both accurate and effective for learners in different clusters. Additionally, the recommender system is applied to English intelligence teaching in colleges and universities, and comparative experiments are conducted with other reference systems to further validate the excellent performance of the system described in this paper.

2 Methodology

2.1 Intelligent platform empowers the transformation of English teaching and learning

2.1.1 Digitization reshapes English language teaching and learning

Driven by big data and digital technology, digital transformation and innovation in education have become the key to promoting education reform and development. The essence of teaching transformation is digital transformation. It emphasizes taking the data of teaching and learning as the core, using effective means such as measurement and evaluation, data collection, analysis, and application from learning resources, teaching process, teaching evaluation, and other aspects. It promotes the digitalization of the teaching process and the deep integration of technology education into teaching. The essence of teaching innovation is to cultivate innovative talents. To promote innovative character and practical ability, education should alter the concept of teaching, utilize digital technology for teaching, discover students' personalized learning needs and precise training, and recognize the mutual promotion of innovative character and practical ability.

2.1.2 Teaching and learning with human-computer collaboration

Figure 1 shows the collaborative teaching environment between humans and computers. The essence of intelligent platform-enabled teaching transformation and innovation lies in teaching under human-computer collaboration, focusing on building an open, shared and sustainable educational ecosystem, emphasizing the complementary advantages and interactive cooperation between teachers and students led by the intelligent platform, and promoting the diversified development of educational informatization, online learning, and intelligence, paving the way for the cultivation of talents with core literacy in innovative thinking and creativity. It provides comprehensive support for classroom teaching and is effective in resolving issues like the single teaching method in traditional teaching.

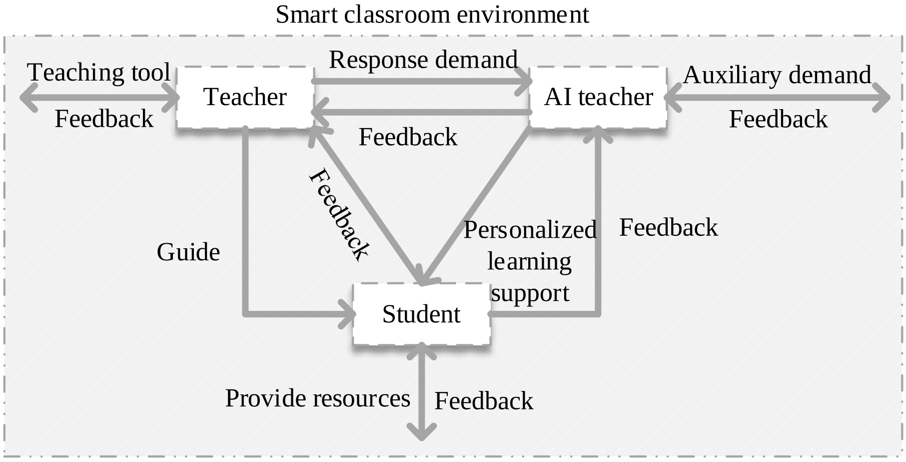


Figure 1. Man-machine collaborative pastoral environment

Taking the KDDI smart classroom as an example, Figure 2 shows the collaborative classroom teaching environment between humans and machines. The high-speed wireless network is configured for the flat-paneled classroom, the teacher is provided with a KDDI teacher machine, the students use the KDDI student machine classroom version, both of them are installed with KDDI's smooth speech intelligent classroom system, and the teacher machine can initiate the classroom, and the student machine can join the classroom and carry out classroom interactions. The intelligent teaching system can realize information interaction in the process of collaborative teaching with teachers, and in the process of collaborative learning with students by means of human-computer interaction, information interaction can be completed through the collection of learning information and the acquisition of knowledge and skill information, and teachers and students can realize interpersonal interaction in the process of classroom activity organization.

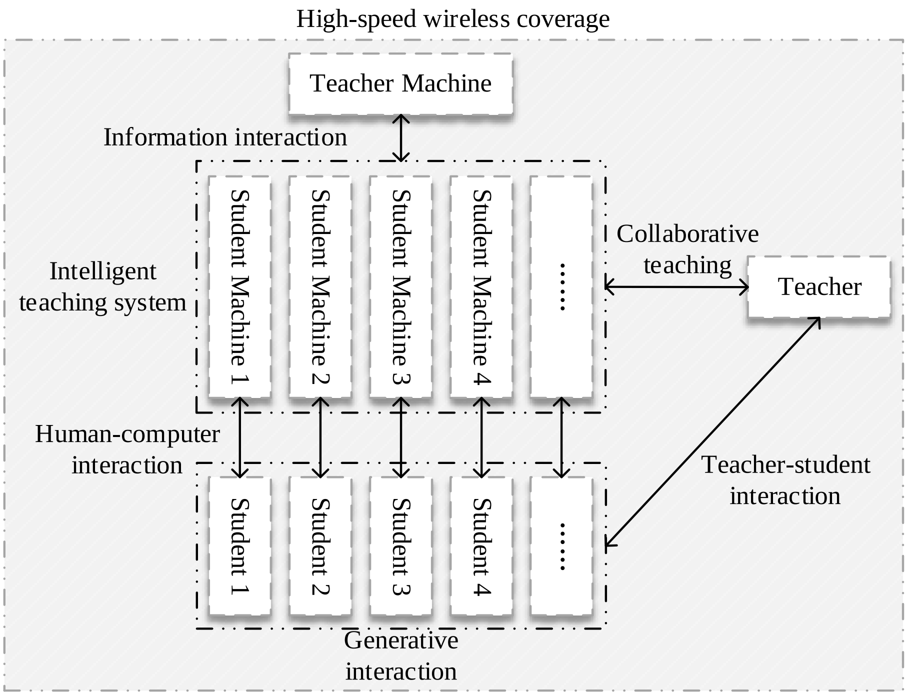


Figure 2. Man-machine collaborative classroom teaching environment

2.1.3 Digital teaching realization path

1) Intelligent learning environment

An intelligent environment is capable of recognizing the situation, recognizing the characteristics of learners, providing them with learning resources and tools, and recording the learning process and results. Relying on the intelligent learning environment, teachers will be relieved from low-value mechanical labor and put more energy into teaching design, teaching organization, and teaching guidance, while students will focus on high-value learning activities such as creative creation, collaborative tasks, knowledge expression, and work presentation.

2) Structuring of learning content

Intelligent platforms play an important role in promoting the structured improvement of learning content, which can be reflected in three aspects. First, the intelligent platform can simplify the learning content, improve students' learning efficiency, and create more learning opportunities and conditions. The second is to enhance the learning content. In order to emphasize the typical characteristics of the learning content and deepen students' memory of learning, the intelligent platform can be applied to strengthen the learning content. The third goal is to reproduce the learning content. With the help of an intelligent platform, learners can acquire abstract knowledge or learned knowledge to help them comprehend.

3) Personalized Learning Path

There are two types of personalized learning paths empowered by the intelligent platform. One thing is to analyze students' learning data, capture their learning status, discover their learning problems, and provide personalized guidance and assistance. The students' self-selected learning path is another way to mobilize their learning enthusiasm and enhance their initiative in learning through the selection and building of learning content according to their strengths.

4) Diversified learning evaluation

The intelligent platform empowers teaching evaluation, diversifying evaluation content and methods, automating and collaborating in evaluation implementation, and applying evaluation results in a precise and scientific manner. The intelligent platform can collect learners' data in various aspects, find relevant evidence for each type of evaluation content, reduce teachers' burden of collecting and analyzing data, and provide technical support for teachers' process evaluation, comprehensive evaluation and characteristic evaluation.

2.2 User collaborative filtering recommendation algorithm based on location scenarios

2.2.1 Clustering algorithm based on k-means++

In this paper, the Calinski-Harabasz [23-24] metric is used as a measure of the algorithm's clustering effectiveness to determine the optimal number of clusters. Assuming $X=\{x_i\}, i=1,\dots,n$ it represents a collection of samples containing n D-dimensional vectors, the samples are divided into k clustering clusters, denoted $C=\{C_i\}, j=1,....k$, while μ_i is the clustering center of cluster c_i. If the data sample is considered as an overall clustering cluster, its clustering center is:

$$\mu_c = \frac{1}{n}\sum_{i=1}^{n} X_i \tag{1}$$

The inter-cluster dispersion of the clusters is denoted by BGSD, which is calculated by the formula:

$$BGSD = \sum_{i=1} n_j d(\mu_j, \mu_c)^2 \tag{2}$$

where n_j denotes the total number of samples within the j nd cluster. Also the intra-cluster tightness is denoted by WGSD, which is calculated by the formula:

$$WGSD = \sum_{j=1}^{k}\sum_{x_i=i} d(x_i, \mu_j)^2 \tag{3}$$

Calinski-Harabasz, on the other hand, is the ratio of the intercluster dispersion, BGSD, to the intracluster tightness, WGSD, and is calculated as:

$$Calinski - Harabasz = \frac{BGSD}{WGSD} \times \frac{n-k}{k-1} \tag{4}$$

2.2.2 Near-neighbor user selection

The specific ideas for dividing and finding the clusters to which the target user belongs and the near-neighbor users of the target user are:

1) According to the existing clustering results, obtain the location coordinates of each cluster center;

2) Calculate the Euclidean distance value between the target user and each cluster center, and sort them from small to large;

3) Divide the target user to the clustering cluster with the smallest distance value;

4) Calculate the Euclidean distance value between the target user and all users in the cluster cluster where it is located, and sort them;

5) Take out the first 50 users with the smallest distance value from the target user, and this set of users is used as the nearest neighbor users of the target user.

2.2.3 Invisible weight calculation

In this paper, we normalize and assign weights to the behavioral data of “learning time” and “number of visits”, and use the results as the user’s preference value for resources, as discussed below.

1) Learning time t represents the duration of users' browsing and learning for a certain resource, and the calculation method of converting users' learning time into specific values is as follows:

$$T(u_{si}) = \begin{cases} 0 & t < 30 \\ t_{(u_s)} & 30 < t < 300 \\ 1 & t > 300 \end{cases} \tag{5}$$

Where t represents the length of time the user stays on a resource. When the user stays less than 30s, it indicates that the user is not interested in the resource and its length of interest $T(u_{ii}) = 0$; when the user stays longer than 30s and less than 300s, the user's length of interest $T(u_{si}) = \frac{t}{300}$; and the rest indicates that the user has a high degree of interest in the resource, and the user's length of interest $T(u_{si}) = 1$.

2) The larger the ratio of the number of user visits to a resource to the number of user visits to all resources in the context of the location to which the resource belongs, the higher the user's interest in the resource, and vice versa, the smaller. Its specific calculation is as follows:

$$R(u_{si}) = \frac{u_{si}}{\sum_{i=1}^{N} u_{si}} \tag{6}$$

$R(u_{si})$ represents the ratio of the number of accesses to resource i by user u in location context s to the number of accesses to all resources by user u in location context s. The numerator represents the number of accesses to resource i by user u in location context s, the denominator represents the number of accesses to all resources by user u in location context s, and N represents the number of resources.

$$P(u_{si}) = 0.5 \times T(u_{si}) + 0.5 \times R(u_{si}) \tag{7}$$

$P(u_{si})$ denotes the overall preference value of the user u for the resource i in the location context s, $P(u_{si})$ has a value in the range of [0, 1], $T(u_{ij})$ denotes the duration interest level of the user u for the resource i in the location context s, and $R(u_{si})$ denotes the access interest level of the user u for the resource i in the location context s.

2.2.4 Algorithm implementation steps

The flow chart of the user-based collaborative filtering recommendation algorithm can be seen in Figure 3. In this paper, the user-based collaborative filtering recommendation algorithm [25-26] firstly realizes user location clustering as well as the acquisition of the target user's nearest neighbor users by the above method; then it needs to calculate the Pearson's similarity between the target user and the nearest neighbor users; then it selects the top five users who have the highest similarity with the target user and takes them as the target user's similar users; finally, it utilizes the similar user's ratings of the resources to predict the target user ratings, and sort the rating values and recommend the top N resources with high rating values to the target user.

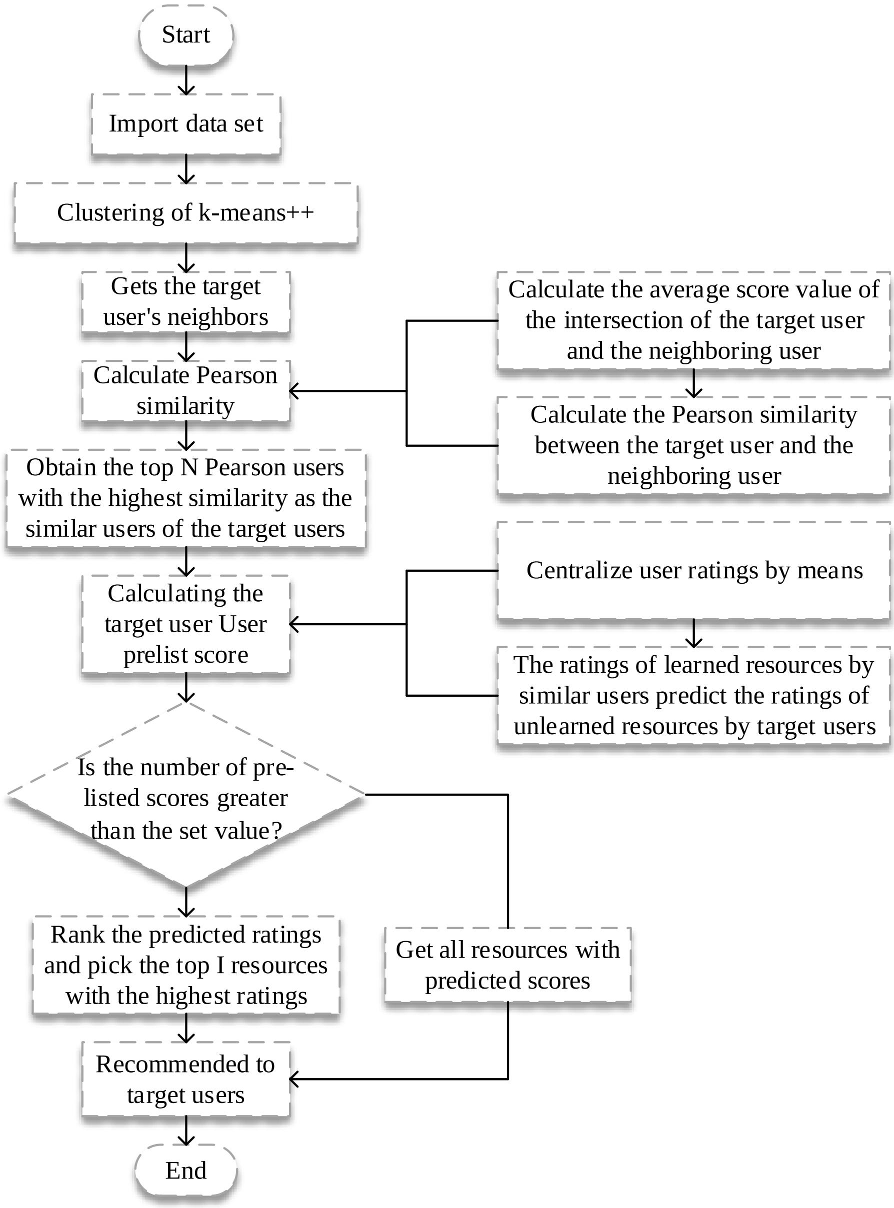


Figure 3. Collaborative filter recommendation algorithm flowchart

2.3 Recommendation system architecture design

The system is built to help designers develop more accurate recommendation strategies through effective data. The English learning recommendation system's development architecture is depicted in Fig. 4. The overall framework of the system is divided into three main layers, specifically including the user layer, business layer, and data layer.

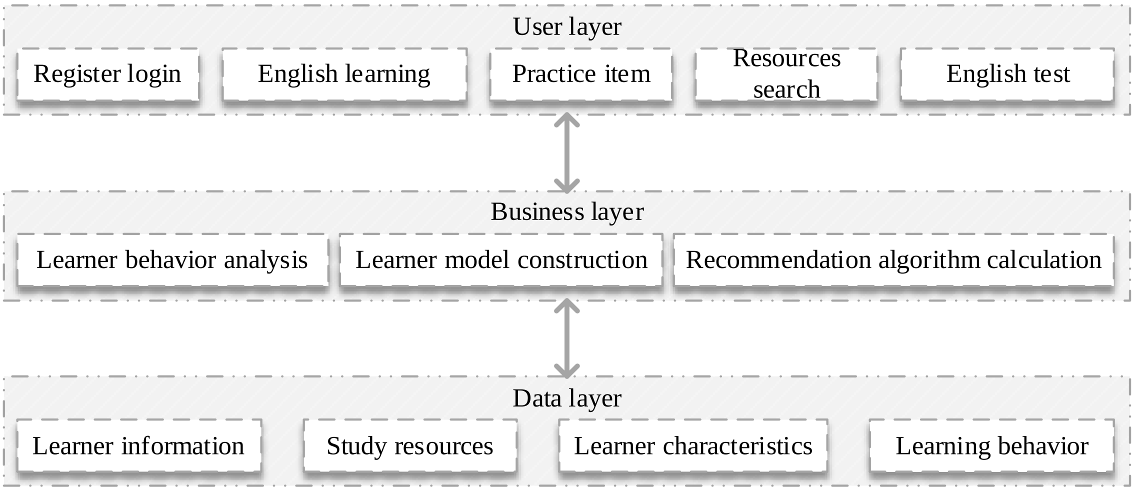


Figure 4. Platform architecture diagram

1) Data Layer

The data layer is the first layer of the English learning recommendation system, which mainly operates and manages system data, including English resource data, learner information data, learner behavior data, and learner characteristics. Among them, resource data include vocabulary resources, test question resources, etc. Learner information data includes learners' basic characteristics, which are generally stable and unchanged. Basic learner information includes grade and major, which are basic conditions for personalized learning. The learner's behavioral data includes both explicit and implicit data, with explicit data mainly consisting of the behaviors that learners actively participate in while learning in the system. Implicit behavioral data mainly refers to the learner's experience in the system, such as the number of times of browsing the knowledge of each module, the residence time, and the search records and the number of searches, which can indirectly show the learner's preference. The combination of explicit and implicit learning behavior data makes the construction of the learner model more complete, which provides a guarantee for the subsequent resource recommendation of the system. The data layer collects, organizes, and stores the above data.

2) Business Layer

The data layer and user layer are connected by the business layer, which is the middle layer of the system architecture. The business layer interacts more with the data layer, and the learner data generated by the business layer will be recorded to the data layer, and the data from the data layer will also update the business layer. The learner's vocabulary level model and dynamic interest model are updated, and the clustering algorithm based on the learner's English level and the collaborative filtering recommendation algorithm based on the learner's dynamic interest model is constructed.

3) User Layer

The user layer includes functions such as English learning, English testing, registration and login, and searching for learning resources. This layer records learner behavior data in the system database. The user layer presents the recommendation services provided by the English resource recommendation system for learners, which are mainly reflected in the system's functions of login and registration, recommendation of learning resources, and practice test questions. Among them, registration and login are used for learners' identity verification, in which the registered professional information can also be used for the determination of learners' initial learning interests; English resource recommendation is mainly used for the system to recommend personalized learning resources for learners and the preference of learners for resources at the current stage is mainly determined according to learners' English proficiency and learning interests. Based on the learner's behavioral data, the system updates the learner's level and learning interest to provide more accurate learning resources for the learner; the testing function is mainly used to test the learner's English level, which helps the system recommend learning resources that meet the learner's level. These data provide the basis for the system to recommend personalized learning resources to learners.

3 Results and analysis

3.1 Cluster analysis of learners' English level

English proficiency test data from 100 learners randomly selected from the recommender system of this paper were collected and analyzed by clustering. The level test is divided into four segments: listening, speaking, reading, and writing. In this paper, the age, gender, occupation type, and learning needs of 100 learners are used as the basis for analyzing the cluster mining results. In addition, the data description of the basic information of the sampled research subjects is in accordance with the classification criteria of the recommendation system described in this paper. The basic information obtained from the learners is shown in Table 1.

Table 1. Learners basic information

Attribute	Data type	Range of values
Age	Number(10)	≥18
Gender	Char(15)	Male or Female
Occupational type	Char(100)	Students, accounting, agricultural technology, foreign trade personnel, tourism, medicine, other
Learning demand	Char(100)	Academic, professional, travel, and other

The clustering algorithm obtains the final clustering center results, and the clustering center results based on different learners are shown in Figure 5. Five colors are depicted in the figure, which represent the classification levels of the recommendation system in this paper. Table 2 demonstrates the clustering statistics results based on English proficiency levels.

Based on the clustering results, we categorize learners' English proficiency into five clustering groups, i.e., Yellow (Primary level), Red (Entry level), Purple (Qualified level), Blue (Mastery level), and Green (Proficiency level). The result is the clustering of learners' scores corresponding to the five aspects of English listening, speaking, reading, writing and overall proficiency by the recommended system in this paper. For example, learners at the beginner's level of English account for 25% of the statistical population. Their intervals for listening, speaking, reading, writing, and overall scores are [0,50], [0,53], [0,47], [0,54], and [0,50], respectively.

This result provides strong support for us to provide personalized recommendations of English learning resources in the following. Take learner number 1 as an example. In Fig. 5, this learner belongs to the Yellow cluster, which shows that this person's English level is Primary level. The recommendation system in this paper requires the background to prioritize the recommendation of primary English learning materials for him. According to the level of strength of different modules, the learning contents of the weaker modules are recommended as a priority. In addition, the system described in this paper can also provide recommendations for specific learning needs. For example, the learner numbered 12 in the figure is a college student, and his learning purpose is to pass English 6. So our system can arrange for him the content related to Grade 6 vocabulary, listening, translation and writing. Accurate and effective recommendations can often result in learners achieving twice the result with half the effort.

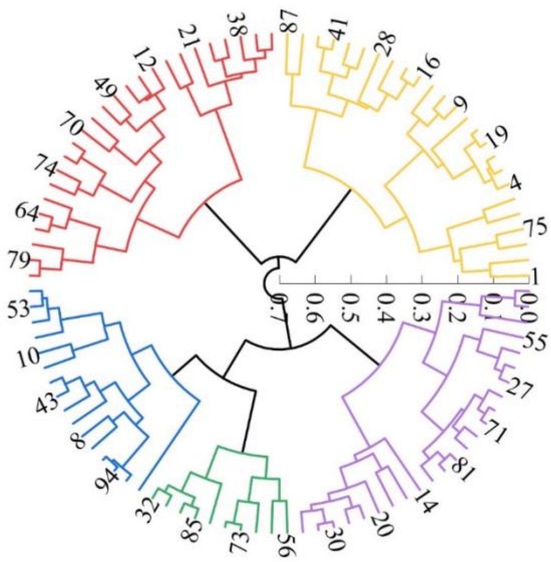


Figure 5. Cluster center diagram based on different learners

Table 2. Statistical results based on the English level

Clustering (Grade)	Proportion	Listening Score	Speaking Score	Reading Score	Writing Score	Overall Score
Yellow (Primary level)	25%	[0,50]	[0,53]	[0,47]	[0,54]	[0,50]
Red (Entry level)	25%	(50,68]	(53,73]	(47,71]	(54,75]	(50,70]
Purple (Qualified level)	24%	(68,78]	(73,81]	(71,82]	(75,83]	(70,80]
Blue (Mastery level)	16%	(78,87]	(81,90]	(82,91]	(83,92]	(80,90]
Green (Proficiency level)	10%	(87,100]	(90,100]	(91,100]	(92,100]	(90,100]

3.2 Comparison of Personalized Recommender Systems for Students

Take a certain two classes of students majoring in a certain university as an example. To carry out the research on the English online learning system based on students' individualized needs, in order to verify the effectiveness and superiority of the system in this paper, it is compared with the intelligent teaching system based on deep learning (notation system 1) and the online open course system based on B/S architecture (notation system 2), respectively. It is assumed that all students have the highest demand for the six major courses of Basic English (notation number 1), Advanced English (notation number 2), English Vocabulary (notation number 3), English Writing (notation number 4), English Translation (notation number 5), and English Audio-Visual (notation number 6), and according to the one-to-one collection, the actual needs of the students are shown in Table 3.

The comparison results of the recommendation accuracy of the three groups of recommender systems for all students in two classes are shown in Figure 6. The figure shows that the accuracy rates of the six English courses recommended by the recommender systems in this paper are higher than those of the reference systems. The system recommendation accuracy of this paper is above 90%, and the accuracy rate of Basic English is the highest, at 95%, which is 10% and 21% higher than System 1 and System 2, respectively.

Table 3. The requirements of students' English courses are investigated

Course name	Basic English	Advanced English	English vocabulary	English writing	English translation	English audio-visual
Demand degree	23	10	17	15	10	5

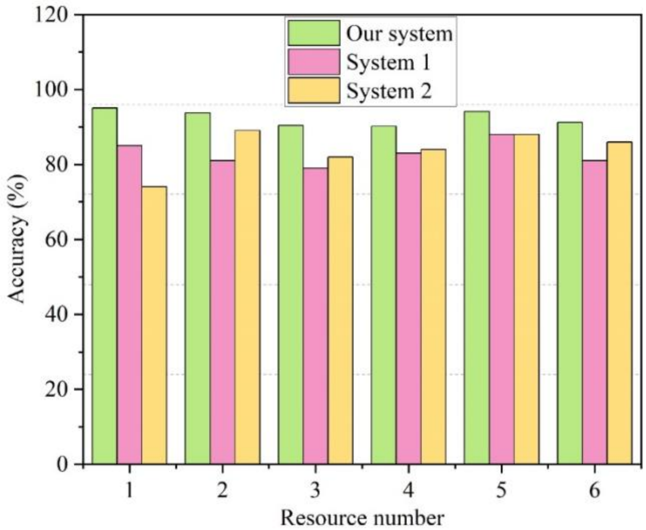


Figure 6. Comparison diagram of recommendation accuracy of system

In order to further verify the applicability of this paper's system, Fig. 7 shows the results of comparative experiments on the recommendation time of students' course requirements. As can be seen from Fig. 7, this paper's system has a clear advantage in the recommendation time. The recommendation time remains within 4s even with data size of 10MB, which is approximately 4s and 3s less than the two reference systems, respectively. The personalized demand recommendation using the system in this paper has the advantage of being faster, as demonstrated.

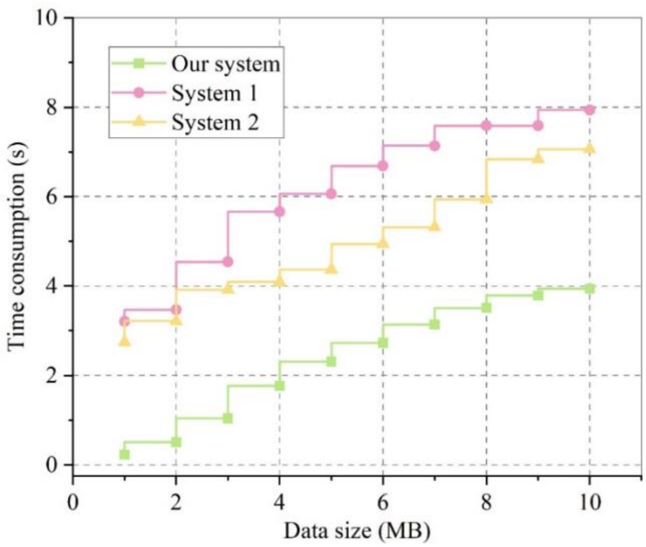


Figure 7. Recommended consumption time for course requirements

3.3 The Effect of Personalized Recommendations on Students' English Achievement

In this paper, pre-test and post-test English scores were collected from 50 students in each of the control classes (CK) and the experimental class (T), respectively. In the data analysis stage, the scores of the pre-test and post-test of the experimental and control classes were first compared horizontally to explore the differences in scores between the two classes after the experiment, and then the pre-test and post-test scores of the two classes were analyzed vertically to explore the trend of the changes in the English scores of the two classes.

3.3.1 Comparison of English proficiency before experimentation

Before the experiment, this paper tested the normality of the English scores of the two classes, and the normal Q-Q plots of the English scores of the two classes are shown in Figures 8 and 9. As can be seen from the graphs, the expected values of the scores of the two classes are distributed around the red reference line. It indicates that the English scores of the two classes are consistent with the normal distribution. Before the experiment, the average English score for the CK class was 69.89, and the average English score for the T class was 69.86, with no significant differences in mean values between the two classes. In addition, Figure 10 demonstrates the normal distribution graph of the English scores of the two classes. The graph indicates that Class T's English scores are more discrete than those of Class CK. The results of the statistical analysis of the pre-test English achievement data are shown in Table 4. Among them, the independent samples t-test obtained a p-value of 0.846, which is greater than 0.05, and there is no statistically significant difference in the data, which is in line with the requirements of the parallel class and allows for the English teaching experiment.

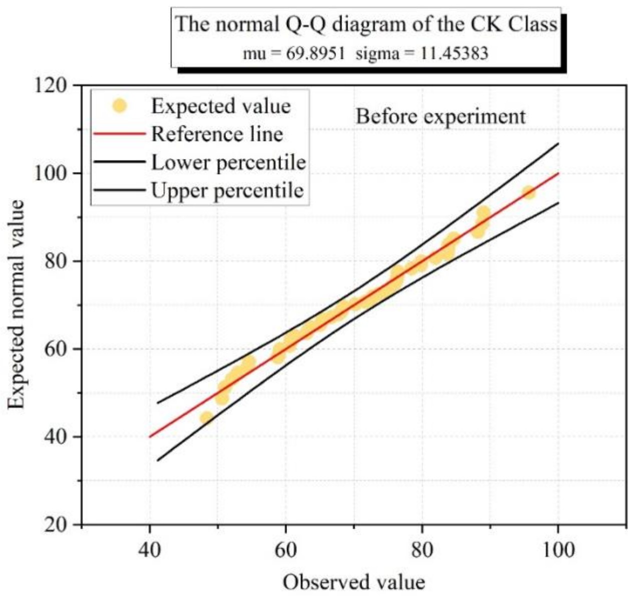


Figure 8. The normal Q-Q diagram of the CK Class

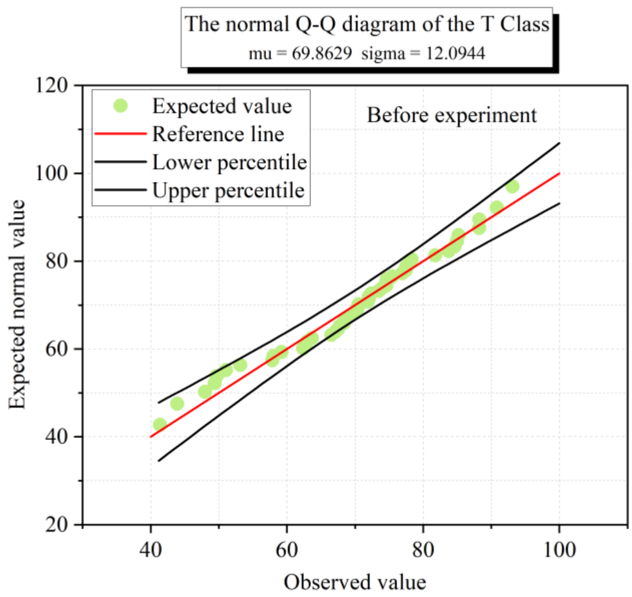


Figure 9. The normal Q-Q diagram of the T Class

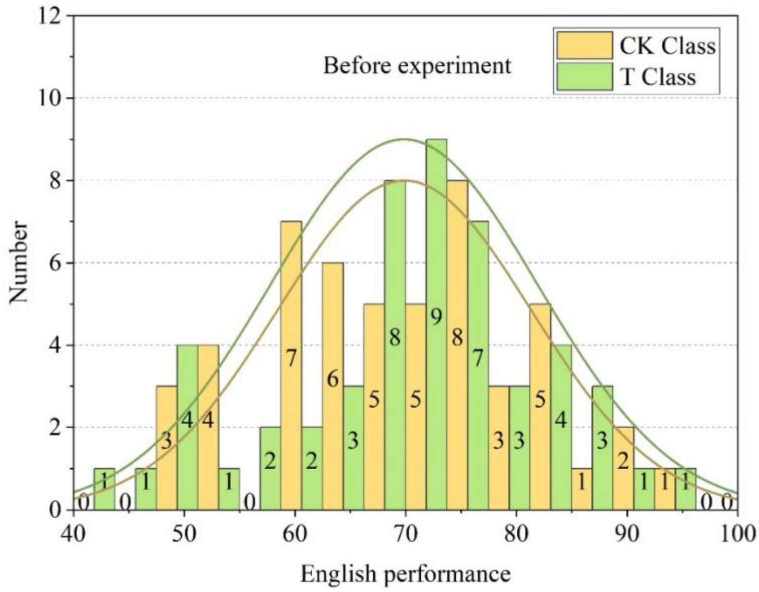


Figure 10. The positive state distribution of the previous English grade

Table 4. Before experiment English performance statistics

	Class	Number	Average value	Standard deviation	Standard error mean	P value
Pretest	CK Class	50	69.89	8.169	1.2659	0.846
	T Class	50	69.86	10.261	1.7162	

3.3.2 Comparison of English proficiency after the experiment

Maintaining one month of English smart teaching for an experimental class, the personalized recommendation system constructed in this paper is utilized in a smart classroom environment to make full use of the learning resources of the smart teaching system in order to judge whether the smart classroom integrating personalized recommendation in this paper has practical feasibility.

The normal Q-Q diagrams of the English scores of the two classes after the test are shown in Figures 11 and 12, and the expected values of the scores of the two classes are distributed around the red reference line. It indicates that the English scores of the two classes are in line with the normal distribution. As shown in Figure 13 of the normal distribution, Class T has more students with high scores than Class CK after the experiment. The results of the data analysis are shown in Table 5. The post-test mean of English scores of Class T is 77.38, the mean of Class CK is 73.09, and the mean of Class T is 4.29 points more than that of Class CK. In terms of standard deviation, it is 5.5128 for class T and 7.165 for the control class, and there is not much difference between the standard deviation of the students in the two classes. The p-value of independent samples T-test was 0.026, which is less than 0.05. Therefore, the difference in vocabulary scores between the experimental and control classes after the experiment was large. There is a statistically significant difference in the data.

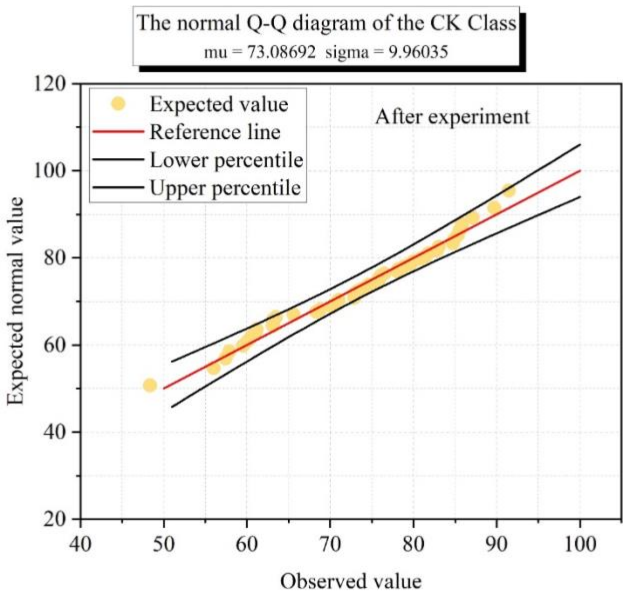


Figure 11. The normal Q-Q diagram of the CK Class

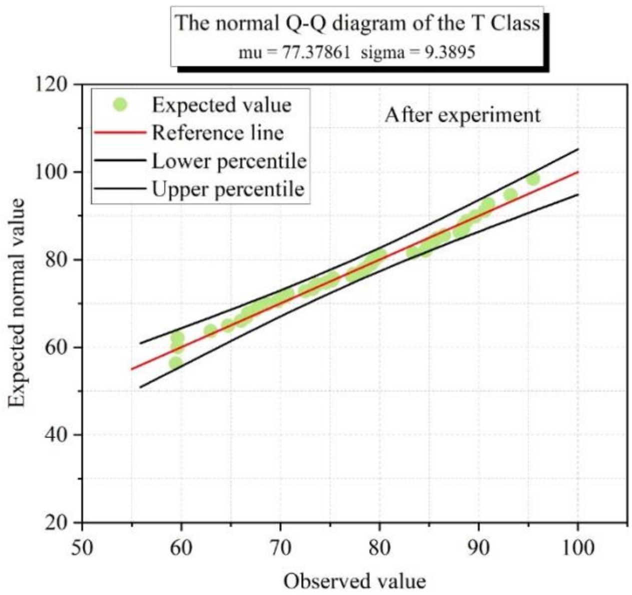


Figure 12. The normal Q-Q diagram of the T Class

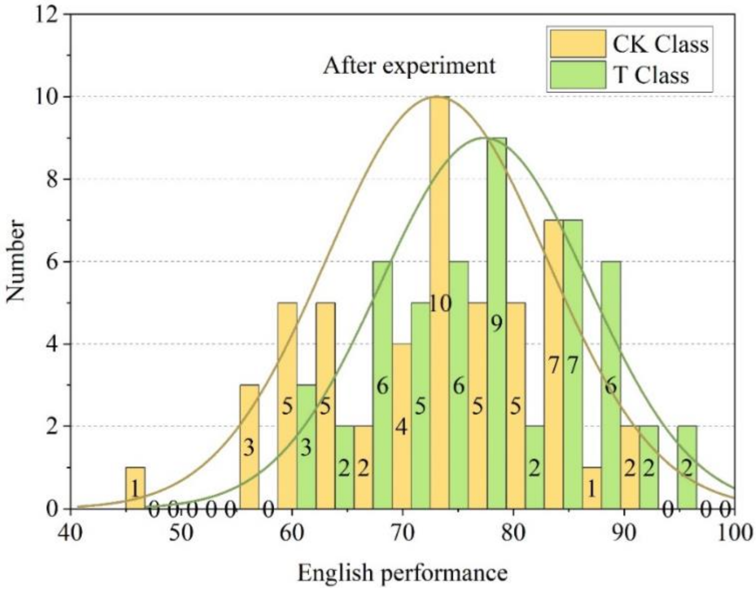


Figure 13. The positive state distribution of the previous English grade

Table 5. After experiment English performance statistics

	Class	Number	Average value	Standard deviation	Standard error mean	P value
Aftertest	CK Class	50	73.09	7.165	1.4626	0.026
	T Class	50	77.38	5.5128	1.5315	

The results of the data analysis show that English teaching in a smart classroom incorporating the personalized recommendation system of this paper is more effective than traditional teaching.

4 Conclusion

The purpose of this paper is to explore and test the effectiveness of a personalized English teaching recommendation system based on big data, applied to the smart classroom, on the digital transformation of English teaching mode in colleges and universities.

1) The personalized recommendation system designed in this paper divides 100 learners into five clusters by clustering method, combining their English level and basic information. It provides strong support for us to provide personalized recommendations of English learning resources in the follow-up.

2) By comparing with the referenced personalized recommendation systems, the accuracy of the system recommendations in this paper is all higher than 90%, and the accuracy of basic English is the highest at 95%, which is 10% and 21% higher than System 1 and System 2, respectively. The system in this paper has the shortest consumption time recommended. Even when the data size is 10MB, the recommendation time is not more than 4s, which is about 4s and 3s shorter than the two reference systems, respectively. The personalized demand recommendation using the system in this paper has the advantage of being faster, as demonstrated.

3) In this paper, pre-test and post-test English scores were collected from 50 students in each of the control classes (CK) and the experimental class (T), respectively. The English scores of both classes were in line with the normal distribution before the experiment, and the average scores of the CK and T classes were respectively 69.89 and 69.86. After the experiment, the scores of the T class improved by 4.29 points compared to the CK class. The distribution of high scores in class T is greater than in class CK. The gap in vocabulary scores between the experimental and control classes after the experiment is significant. The data shows a statistically significant difference.

Funding:

Practical Research on Foreign Language Awareness from the Perspective of Cultural Consciousness (Project Number: YQJH2023165).

References

[1] Yu, X., & Liu, C. (2022). Teaching English as a lingua franca in China: Hindrances and prospects. English Today, 38(3), 185-193.

[2] He, D. (2017). Perceptions of Chinese English and pedagogic implications for teaching English in China. Researching Chinese English: The state of the art, 127-140.

[3] Liu, L., & Tsai, S. B. (2021). Intelligent recognition and teaching of English fuzzy texts based on fuzzy computing and big data. Wireless Communications and Mobile Computing, 2021(1), 1170622.

[4] Wang, L., & Kokotsaki, D. (2018). Primary school teachers' conceptions of creativity in teaching English as a foreign language (EFL) in China. Thinking skills and creativity, 29, 115-130.

[5] Jia, X. (2021). Research on the role of big data technology in the reform of English teaching in universities. Wireless Communications and Mobile Computing, 2021(1), 9510216.

[6] Zhu, J., Zhu, C., & Tsai, S. B. (2021). Construction and analysis of intelligent english teaching model assisted by personalized virtual corpus by big data analysis. Mathematical Problems in Engineering, 2021(1), 5374832.

[7] Sun, M., & Li, Y. (2020). Eco-environment construction of English teaching using artificial intelligence under big data environment. IEEE Access, 8, 193955-193965.

[8] Chang, H. (2021). College English flipped classroom teaching model based on big data and deep neural networks. Scientific Programming, 2021(1), 9918433.

[9] Shi, L. (2022). Application of big data language recognition technology and GPU parallel computing in English teaching visualization system. International journal of speech technology, 25(3), 667-677.

[10] Maimaiti, K. (2019). Modelling and analysis of innovative path of English teaching mode under the background of big data. International Journal of Continuing Engineering Education and Life Long Learning, 29(4), 306-320.

[11] Du, Z., & Su, J. (2021). Analysis of the practice path of the flipped classroom model assisted by big data in english teaching. Scientific programming, 2021(1), 1831892.

[12] Zeng, Y. (2021). Application of flipped classroom model driven by big data and neural network in oral English teaching. Wireless Communications and Mobile Computing, 2021(1), 5828129.

[13] Mao, W. (2022). Analysis of the psychological barriers to spoken English from big data and cross-cultural perspectives. Frontiers in Psychology, 13, 899101.

[14] Qian, R., Sengan, S., & Juneja, S. (2022). English language teaching based on big data analytics in augmentative and alternative communication system. International Journal of Speech Technology, 25(2), 409-420.

[15] Zang, R., & Wang, L. (2021, April). Personalized teaching model of college English based on big data. In Journal of Physics: Conference Series (Vol. 1852, No. 2, p. 022013). IOP Publishing.

[16] Ghazali, M. A. I. M. (2020). Student teachers' readiness towards the digitalization of the 21st century English language classroom in education 4.0. In ICTE'20: International Conference on Teacher Education (Vol. 2, pp. 63-75).

[17] Huang, H. (2024, July). Research on English distance education network resource management system based on big data mining technology. In Third International Conference on Electronic Information Engineering and Data Processing (EIEDP 2024) (Vol. 13184, pp. 886-893). SPIE.

[18] Li, X. (2021). Textbook Digitization: A Case Study of English Textbooks in China. English Language Teaching, 14(4), 34-42.

[19] Sur, E. (2022). A systematic compilation study on digitalization in language education. Gazi Üniversitesi Gazi Eğitim Fakültesi Dergisi, 42(1), 541-584.

[20] Shcherbakova, I., Kovalchuk, N., Timashova, M., Konkin, B., & Soprantsova, J. (2021). Digitalization of education and its impact on the teaching of foreign languages to students of technical universities. In E3S Web of Conferences (Vol. 273, p. 12019). EDP Sciences.

[21] Zakharova, M. (2020). Digitalization of english language training at tertiary level. In EDULEARN20 Proceedings (pp. 5525-5528). IATED.

[22] Wang, Q., & Jiang, X. Q. (2018, February). Empirical study on reform model of college English teaching model based on computer and big data. In 2018 10th International Conference on Measuring Technology and Mechatronics Automation (ICMTMA) (pp. 412-415). IEEE.

[23] Aik Lim Eng,Choon Tan Wee & Abu Mohd Syafarudy. (2023). K-means Algorithm Based on Flower Pollination Algorithm and Calinski-Harabasz Index. Journal of Physics: Conference Series(1),

[24] Zhang Wenbo,Yue Zixian,Ye Jinmei,Xu Hengying,Wang Yuxiang,Zhang Xiaoguang & Xi Lixia. (2022). Modulation format identification using the Calinski-Harabasz index. Applied optics(3),851-857.

[25] Ni Li & Yinshui Xia. (2024). Movie recommendation based on ALS collaborative filtering recommendation algorithm with deep learning model. Entertainment Computing100715-.

[26] Ke Yan. (2024). Optimizing an English text reading recommendation model by integrating collaborative filtering algorithm and FastText classification method. Heliyon(9),e30413-.