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

https://doi.org/10.70917/ijcisim-2026-0173
Article

Research on Intelligent Development of Talent Cultivation Mode of Higher Vocational Innovation and Entrepreneurship Education Based on Data Mining under the Perspective of Industry-Teaching Integration

Yinxing Zhou *

Teaching Affairs Department, Jiangxi Technical College of Manufacturing, Nanchang 330000, Jiangxi, China; 18296220806@163.com

Abstract: Against the backdrop of the integration of industry and education, vocational college innovation and entrepreneurship education has entered a new phase characterized by diversification, emphasis on intrinsic development, and intelligent development. This article constructs a student learning behavior analysis model, collects student learning behavior data, and proposes the use of the k-means algorithm to analyze student behavior. Additionally, course access, video viewing, assignment performance, daily performance, and interaction are selected as online learning behavior characteristics. The results reveal that vocational college students exhibit a high proportion of passive learners, with relatively low proportions of passive and active learners. Passive learners account for the largest proportion at 87.5%, and learning quality is primarily influenced by factors such as teacher attention, chapter quizzes, assignments, and course credits. The construction of a vocational college innovation and entrepreneurship education talent cultivation model can be approached from three aspects: establishing a comprehensive innovation and entrepreneurship talent cultivation system, implementing an apprenticeship system for innovation and entrepreneurship talent cultivation, and developing an innovation and entrepreneurship course system.

Keywords: k-means; innovation and entrepreneurship education; industry-education integration; learning behavior analysis model

1. Introduction

Innovation and entrepreneurship education is an educational approach aimed at cultivating individuals with innovative and entrepreneurial spirit, awareness, capabilities, and high-quality innovative talent. It integrates elements, content, models, and mechanisms of innovation and entrepreneurship into the teaching and talent cultivation processes of higher vocational education [1-4]. Under the backdrop of industry-education integration, vocational colleges collaborate with enterprises on innovation, engaging in deep cooperation across multiple areas such as talent cultivation, joint research, team building, and platform development to advance the intelligent development of innovation and entrepreneurship education talent cultivation models. This is not only a social responsibility for universities to adapt to the demands of the times but also an intrinsic need to drive their own development [5-8].

Entrepreneurship is challenging, and it is even more so for college students. For vocational college graduates who have recently graduated, to successfully embark on an entrepreneurial journey, they require external support. This includes the implementation of policies and regulations by the national government, local governments, and vocational colleges to vigorously support graduate entrepreneurship and assist them in overcoming various obstacles and challenges encountered during the entrepreneurial process [9-12]. On the other hand, they need to possess basic entrepreneurial knowledge. Educational authorities must strengthen entrepreneurial education for students, enhance their entrepreneurial qualities and capabilities, and ensure that entrepreneurs also possess robust psychological resilience and mental fortitude [13-16]. Especially when facing difficulties and setbacks, a strong will, mature psychological resilience, and solid entrepreneurial skills are essential to mitigate the risks associated with entrepreneurship [17-18]. Therefore, how to strengthen innovation and entrepreneurship education to align with national conditions and societal needs urgently requires a shift in traditional educational philosophies and talent cultivation models [19-20]. The intelligent development of innovation and entrepreneurship education talent cultivation models can play a significant role in addressing the challenge of “entrepreneurial difficulties” [21]. This model emerged alongside the industrial transformation and technological upgrades driven by the development of artificial intelligence. It involves restructuring higher vocational innovation and entrepreneurship education based on intelligent technologies to promote the integration of talent cultivation and industrial demands, which complement each other and hold significant implications for enhancing the innovation and entrepreneurship capabilities of higher vocational students [22-25].

The article first explores the construction of a model for analyzing student learning behavior in vocational college innovation and entrepreneurship education, proposing a method for clustering behavioral data using the K-means algorithm. Next, online learning behavior characteristics are selected, defined, described, and their feature values are chosen. Based on the six elements of online learning behavior, these characteristics are categorized. Using an attribute-oriented inductive method, the collected behavioral characteristics are quantified. Taking the “Innovation and Entrepreneurship Education” course as an example, behavioral data on students' use of digital teaching resources over one semester is collected. Three experiments are conducted: descriptive analysis, correlation analysis, and clustering analysis of online learning behavioral characteristics, to analyze online learning behavioral data and identify key factors influencing students' innovation and entrepreneurship performance. Finally, the article explores talent cultivation models from three aspects: constructing a comprehensive innovation and entrepreneurship talent cultivation system, establishing a high-end innovation and entrepreneurship talent cultivation apprenticeship system based on deep school-enterprise collaboration, and constructing a comprehensive innovation and entrepreneurship course system.

2. Student Behavior Data Analysis Method Based on K-means

2.1. Student Learning Behavior and Data Collection

Student learning behavior analysis refers to researchers using behavioral theory, learning theory, and other theoretical frameworks to analyze the elements of student learning behavior and the relationships between these elements. It also involves analyzing the implementation units of learning behavior and the internal transformative relationships between these units to reveal the connection between student learning behavior and learning outcomes. To analyze student learning behavior, it is necessary to first study student learning behavior.

1) Learning Behavior

With the transformation of teaching models, blended online and offline teaching has become the norm in primary, secondary, and higher education. As a result, students' learning behaviors now include both online and offline learning behaviors. Generally, pre-class and post-class activities are primarily conducted online, which constitutes online learning behavior. During class, activities are mainly offline, though some online activities may also occur, such as accessing online learning resources or watching micro-lecture videos in the classroom, which are examples of online learning behavior.

2) Collection of Learning Behavior Data

This paper sets up measurement data for various learning behaviors from two perspectives: learning behavior analysis applied to course unit learning and whole-course learning. Analysis of learning unit learning enables teachers to promptly grasp students' learning dynamics and intervene and adjust in a timely manner. Analysis of whole-course learning primarily provides analytical data for administrators and teachers, serving as a basis for administrators' decision-making in teaching management and assisting teachers in optimizing the overall teaching design of the course.

(1) Quantification of Basic Learning Information

Student basic information is a critical dataset in learning behavior analysis. It is necessary to quantify student basic information, including gender, age, grade, and student type. The student population in vocational colleges is highly diverse, and different student types exhibit slight variations in learning styles and characteristics. Age itself is a data point and does not require quantification.

(2) Online Learning Behavior Data Collection

Online learning behavior data is relatively easier to collect compared to offline learning behavior data, as it can be directly exported from learning platforms.

(3) Offline learning behavior data collection

Offline learning behavior data collection is more challenging. The data collection method proposed in this paper, based on a collaborative learning model, involves members of a learning community recording the offline learning behavior of other members within the same community. Currently, some systems utilize classroom video surveillance systems to automatically collect student offline learning behavior data.

(a) Attendance in class. This refers to whether students attend class. Middle schools generally do not have absenteeism issues, but absenteeism occurs from time to time in universities and vocational colleges. Therefore, attendance in class is an important data point. This data can be assigned values of 1 or 0 for present or absent, or 1, 0.5, or 0 for on time, late, or absent.

(b) Classroom listening. This data measures how attentively students listen in class. It can be categorized as attentive, somewhat attentive, or inattentive. This is primarily determined based on the number and duration of times students engage in activities unrelated to class, such as fidgeting, looking at their phones, or chatting with classmates.

(c) Answering questions. This data is collected based on the number of times students answer questions in class and the accuracy of their answers. This requires the instructor to provide evaluation records after students answer questions. Students who do not answer questions receive 0 points, while those who answer multiple questions receive the average score of the evaluations, with a maximum score of 5 points.

(d) Classroom discussion. The performance of students during classroom discussions is difficult to quantify, so it can be graded as active, relatively active, inactive, or very inactive, with corresponding values of 3, 2, 1, and 0.

(e) Classroom assignments. Classroom assignments or tests are not conducted in every class, so these two data points can be combined. This data is directly collected based on the grades given by the instructor for students' completion of classroom assignments or tests.

(f) Practical training and presentations. In vocational colleges, practical training is an important teaching component. Therefore, students' performance during practical training and their presentation of training results are crucial learning behavior data. This data can be collected by combining peer evaluations of performance during practical training with instructors' evaluations of group training results presentations.

2.2. Student Learning Behavior Analysis Model

Based on students' learning behaviors and their corresponding datasets, a learning behavior analysis model has been developed. This model consists of several modules, including a learning behavior data collection module, a learning outcomes module, a learning behavior analysis module, and a feedback module for analysis results. The model not only analyzes students' learning behaviors after completing a single unit but can also analyze their behaviors after completing an entire course.

(1) Learning Behavior Collection Module

The Learning Behavior Collection Module digitizes students' online and offline learning behavior data and collects it. After collection, the data requires necessary cleaning and processing.

(2) Learning Effectiveness Acquisition Module

The Learning Effectiveness Acquisition Module obtains students' learning effectiveness, which is reflected in their comprehensive grades. This requires determination based on the needs of learning behavior analysis. The learning outcomes of a single class can be reflected through evaluations of online tests and online assignments. The learning outcomes of a task can be reflected based on the instructor's grading after the task is submitted. The learning outcomes of a course can be reflected based on students' final exam scores. Learning outcomes should reflect students' mastery of knowledge, problem-solving abilities, and the development of their overall qualities. Therefore, instructors should scientifically design online test questions, online assignments, tasks, and final exam questions.

(3) Learning Behavior Analysis Module

The learning behavior analysis module utilizes clustering algorithms, Pearson correlation analysis, and other analytical tools to conduct comprehensive, multi-angle data mining on students' learning behavior data, generating analytical results. This is the core module of the learning behavior analysis model.

2.3. K-means algorithm

The K-means algorithm is an unsupervised learning algorithm that is simple to implement and highly effective. In data exploration and cluster discovery, samples within a cluster are similar to each other but dissimilar to samples in other clusters. It is characterized by its ease of implementation and efficiency.

Let the sample set X be a set of behavioral data containing n student samples, i.e., $X = \{x_1, x_2 \ldots x_n\}$. Each sample consists of q categories of resource data [26]. The i th sample is represented as $x_i = (x_{i1}, x_{i2} \ldots x_{iq})$, and the Euclidean distance between any two samples is expressed as:

$$d(x_i, x_j) = \sqrt{\sum_{r=1}^{q} (x_{ir} - x_{jr})^2} \quad (1)$$

The K-means algorithm implementation process can be described as follows:

Step 1: Input a set of behavior resource data for n samples and the number of clusters k.

Step 2: Initialize the cluster centers.

Step 3: Calculate the Euclidean distance between the samples and each center point using the Euclidean distance formula (1), and assign the sample to the cluster with the closest Euclidean distance to the center point.

Step 4: Calculate and update the center point of each cluster.

Step 5: Check whether the cluster center points have changed (no objects have been reassigned to different clusters). If no change, proceed to Step 6; otherwise, proceed to Step 3.

Step 6: Output the clustering results.

To determine the optimal number of clusters k, the sum of squared errors of clustering can be defined as an evaluation metric for clustering performance as follows:

$$SSE = \sum_{i=1}^{k} \sum_{x \in c_i} \| x - u_i \|^2 \quad (2)$$

Among them, SSE represents the sum of squared errors of clustering, which is the sum of the squares of the distances between sample points and cluster centers. The better the clustering effect, the smaller its value. c_i represents the i th cluster, and u_i represents the center point of the i th cluster.

2.4. Clustering of behavioral data based on the k-means algorithm

In order to obtain more authentic, effective, and comprehensive behavioral data on students' use of digital teaching resources, we extracted behavioral data on all students' use of digital teaching resources on the platform during a one-semester period for a certain course offered by the college on the online teaching platform.

2.4.1. Acquisition of behavioral data

This paper uses the “Innovation and Entrepreneurship Education” course at a higher vocational college as an example to conduct a cluster analysis of student behavioral data. The course organizes blended learning through a “three-in-one, two-platform, three-stage” approach combining online learning, classroom instruction, and social practice, and features abundant digital teaching resources. The online teaching platform has offered multiple sessions, with each session attracting 200 to 1,500 students enrolling on campus. This paper analyzes five categories of resource behavior data collected from students in the most recent semester of the course, including video resource completion rates, participation in discussions, chapter learning frequency, chapter quiz scores, and course assignment scores. The video resources consist of 30 chapter-based video learning resources. The video resources consist of 32 chapter-based video learning resources, with student behavior data reflecting the completion rate of resource viewing. The number of discussions refers to the number of discussion topics students participated in, serving as a form of course interaction. The number of chapter learning instances indicates the number of times students accessed chapter learning materials, reflecting their learning interest and the course's difficulty level. Chapter quizzes and assignments represent the scores students achieved through participating in chapter quizzes and completing course assignments, serving as indicators of learning outcomes.

2.4.2. Data normalization

(1) Video resources

The numerical value of video resources represents the completion rate of the resources, i.e., the completion status of video learning. A value exceeding 100% (inclusive) can be considered as completion of resource learning. Normalization is performed as shown in Equation (3).

$$x_{ij} = \begin{cases} 100 & y_{ij} \geq 100\% \\ y_{ij} * 100 & else \end{cases} \tag{3}$$

(2) Other resources

The discussion count, chapter study count, chapter quiz count, and assignment count are normalized as shown in Equation (4).

$$x_{ij} = \frac{y_{ij}}{\max(y_j)} * 100 \tag{4}$$

Among them, $\max(y_j)$ represents the maximum value of the j th type of resource.

2.4.3. Clustering using the k-means algorithm

The collected data is normalized and clustered using the K-means algorithm. The specific process is as follows:

Algorithm input, i.e., the resource value set y .

Algorithm output, i.e., the clustering results and the sum of squared errors SSE .

(1) After normalization, the resource value set y is represented as a behavior data set X containing n student samples.

(2) Set the number of clusters k , input the behavioral data set X , and perform clustering according to the algorithm process.

(3) Output the clustering results and the sum of squared errors SSE .

(4) Increase the number of clusters k and perform clustering again.

(5) Output all clustering results that meet the conditions.

2.4.4. Determination of K value and best clustering results

Clustering of behavioral data based on the K-means algorithm yields clustering results.

3. Characteristics of online learning behavior

3.1. Extraction of learning behavior characteristics

According to behavioral science theory, learners use online learning platforms for learning based on their own needs. Behavioral science focuses on observable and measurable overt behavioral activities. During the online learning process, learners' operations can be observed and quantified. Therefore, online learning behavior analysis takes online operations as a breakthrough point, observing, describing, and refining learners' learning operations for research. Learners engage in online learning on the platform, generating a series of different behaviors that lead to varying learning outcomes. While the behavioral subject, object, environment, and outcomes are universal, the behavioral tools and methods employed, as well as the behavioral processes, are unique to each learner [27]. Therefore, based on the behavioral tools and methods, as well as the behavioral processes within the six elements, online learning behavior is categorized into five types: course access, video viewing, assignment performance, regular performance, and interaction.

3.2. Quantification of behavioral characteristics

3.2.1. Course Access

In course access statistics, the attendance rate reflects the most basic learning situation of learners, as shown in formula (5):

$$A_i = \frac{\sum_{j=1}^{k} m_{ij}}{\sum_{j=1}^{k} n_{ij}} \left(i = 1, 2, \ldots, s; j = 1, 2, \ldots, k\right) \tag{5}$$

In the formula, A_i represents the attendance rate of the i th student, m_{ij} represents the number of times the i th student attended the j th class, n_{ij} represents the number of times the i th student did not miss the j th class, k represents the total number of attendance checks conducted by the teacher, and s represents the total number of students. Considering factors such as random events and emergencies, and in conjunction with expert recommendations, learners with an attendance rate of 80% or higher are recorded as 1, while those below 80% are recorded as 0.

The number of active days reflects whether learners are logging into the course to study. Generally, the number of days the course lasts is the number of days learners should be active online, denoted as a. The mode a_0 is taken as the threshold. Compared to the median and mean, the mode represents the central tendency of a set of data and is not affected by extreme values, reflecting the normal active days of most learners. The median represents the median level of a set of data. and the mean represents the average status of a set of data, but it is easily influenced by extreme values. Therefore, the mode is selected as the threshold. If there are multiple modes, one option is to use the median as an alternative threshold, and the other is to perform threshold selection and classification discussion on the multiple modes again, as described below. Let the learner's actual active days be denoted as A. If $a_0 \geq a$, the expression is:

$$\text{Active days} = \begin{cases} \text{High activity} & A > a_0 \\ \text{Medium activity} & a < A \leq a_0 \\ \text{Low activity} & A < a \end{cases} \tag{6}$$

If $a_0 < a$, then the expression is:

$$\text{Active days} = \begin{cases} \text{High activity} & A \geq a \\ \text{Medium activity} & a_0 \leq A < a \\ \text{Low activity} & A < a_0 \end{cases} \tag{7}$$

3.2.2. Video viewing status

In terms of video viewing, the viewing frequency reflects the learner's commitment to learning. Generally, the total number of videos is taken as the minimum number of times a learner should watch, denoted as t. The mode of the viewing frequency of all learners is taken as t_0. If there are multiple modes, the median is used as the backup threshold, denoted as t_0'. The learner's actual viewing frequency is denoted as T. If $t_0 \geq t$, the expression is:

$$\text{Viewing frequency} = \begin{cases} \text{High} & T > t_0 \\ \text{Normal} & t < T \leq t_0 \\ \text{Low} & T < t \end{cases} \tag{8}$$

If $t_0 < t$, the expression is:

$$\text{Viewing frequency} = \begin{cases} \text{High} & T \geq t \\ \text{Normal} & t_0 \leq T < t \\ \text{Low} & T < t_0 \end{cases} \tag{9}$$

Viewing duration can reflect learners' mastery of the course. Let the total video duration be l, i.e., the duration that learners should watch, and let the mode be l_0. If the actual viewing duration of learners is L and $l_0 \geq l$ (there are repeated viewings), the expression is:

$$\text{Viewing duration} = \begin{cases} \text{Long} & L > l_0 \\ \text{Normal} & l \le L \le l_0 \\ \text{Short} & L < l \end{cases} \tag{10}$$

If $l_0 < l$, the expression is:

$$\text{Viewing duration} = \begin{cases} \text{Long} & L \ge l_0 \\ \text{Normal} & l_0 < L \le l \\ \text{Short} & L < l_0 \end{cases} \tag{11}$$

3.2.3. Work performance

In terms of assignment performance, the threshold for assignment submission is denoted as m, where m represents the number of assignments assigned by the instructor, i.e., the number of assignments learners are expected to complete. If the number of assignments submitted by the learner reaches the threshold, i.e., every assignment is completed and submitted, it is recorded as excellent. Based on actual teaching situations and expert advice, reaching 70% of the threshold is recorded as average, and the rest are recorded as unsatisfactory. The threshold for whether the assignment is repeated is set to 1. If the learner submits the assignment repeatedly, it is recorded as 1; otherwise, it is recorded as 0.

The time of assignment submission reflects the learner's self-efficacy. Let the assignment submission deadline be d days, take the median value of d days, d', sort the assignment submission times in ascending order, and record the learner's assignment submission time as Job. Since learners' study times vary on online learning platforms, take the median value Job_{sub}. If $Job_{sub} < d'$, the expression is:

$$\text{Assignment submission time} = \begin{cases} \text{Positive} & Job \le Job_{sub} \\ \text{Average} & Job_{sub} < Job \le d' \\ \text{Negative} & Job > d' \end{cases} \tag{12}$$

If $Job_{sub} > d'$, then the expression is:

$$\text{Assignment submission time} = \begin{cases} \text{Positive} & Job \le d' \\ \text{Average} & d' < Job \le Job_{sub} \\ \text{Negative} & Job > Job_{sub} \end{cases} \tag{13}$$

3.2.4. Regular performance and interaction

In terms of performance, chapter test scores reflect learners' mastery of the content in that chapter. The maximum score for a chapter test is 100 points, with scores of 90 or above considered excellent, scores in the range of $[80, 89]$ considered average, scores in the range of $[60, 79]$ considered passing, and scores below 60 considered failing [28]. Assuming there are a total of x chapter tests, the mode of the grades from all chapter tests is taken to determine the overall chapter test score. The mode is chosen because the grade that appears most frequently across all chapter tests is more likely to accurately reflect the learner's overall chapter test performance. If the grades are evenly distributed, it is recorded as passing. Assignment grades can indicate a learner's attitude toward the course. Let the total number of assignments assigned be y. teachers evaluate the assignments submitted by learners, with grades divided into "excellent, good, and poor." If using a percentage system, scores of 80 or above are considered excellent, the range $[60, 79]$ is good, and scores below 60 are poor. The mode of all assignment grades is taken as the learner's overall assignment score. If the grades are evenly distributed or multiple modes appear, the assignment score is recorded as good.

In terms of interaction, the total length threshold for post content is the median of the post content contribution rate, which is calculated as follows (14):

$$C_{ij} = \frac{f_{ij}}{\sum_{j=1}^{n} f_{ij}} \left(i = 1, 2, \ldots, m; j = 1, 2, \ldots, n \right) \tag{14}$$

In the formula, C_{ij} is the contribution rate of the number of words in the j th post of the i th student to the total number of posts of the individual, and f_{ij} is the number of words in the j th post of the i th student.

The post depth threshold is the median of the post quantity contribution rate, and the formula for calculating the post quantity contribution rate is as follows (15):

$$C_i = \frac{\sum_{j=1}^{n} f_{ij}}{\sum_{i=1}^{m} \left(\sum_{j=1}^{n} f_{ij} \right)} \left(i = 1, 2, \ldots, m; j = 1, 2, \ldots, n \right) \tag{15}$$

Among these, C_i is the contribution rate of the number of posts or replies made by the i th student to the total number of posts or replies made by all students, f_{ij} is the number of posts made by the i th student in the j th post, m is the number of students, and n is the number of posts.

The quantification of other behavioral features is similar to the process described above and will not be repeated here.

4. Visualization and analysis of student behavior data

4.1. Descriptive analysis of online learning behavior characteristics

4.1.1. Analysis of Learners' Course Learning Frequency

The online learning platform records the log data of the learner's online learning behavior, and in the process of the learner's online learning, each operation will be recorded, and the operation frequency refers to the number of times the learner operates on the learned course resources when the learner is learning the course, and the number of each operation record is increased by 1. Statistics can be obtained on the number of course learning records of each learner, and the learner's use of learning resources can be analyzed, and then the access to the learning platform can be further understood. According to the teaching plan of online learning, the log data generated by the online learning platform in 2023 was selected for analysis, and the data of 5629 students were selected for analysis by sampling method. Figure 1 shows the statistics of course page views. As can be seen from the figure, there is a clear difference in the frequency of learners' learning of the course. Among them, in a semester of study, the minimum number of course views is 0, and the maximum number of views reaches 13,566.

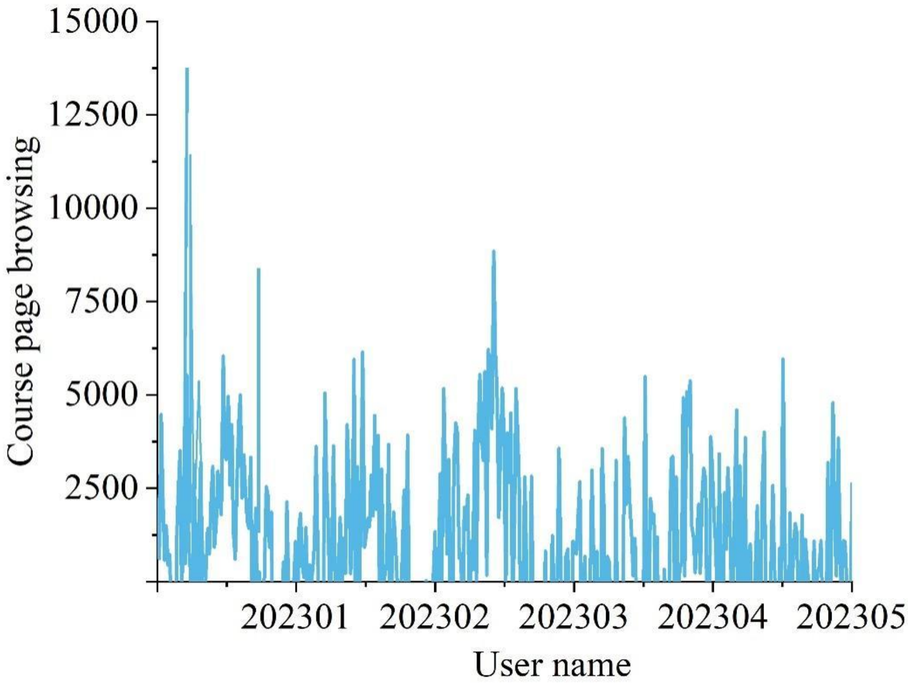

Figure 1. Course page browsing statistics.

The number of times learners have viewed course pages is now sorted in descending order. To analyze learners' course viewing behavior more clearly, a cutoff of 1,000 course page views is used to divide the data into two groups: learners who viewed course pages fewer than 1,000 times during the semester and those who viewed course pages more than 1,000 times. The statistics for learners who viewed course pages fewer than 1,000 times are shown in Table 1. It can be seen that the group with fewer than 1,000 course page views accounts for 97.9% of the total number of learners. Among this group, approximately 86.28% had fewer than 200 course page views per semester.

Table 1. The number of pages of the course page is less than 1000.

Interval	0~200	200~400	400~600	600~800	800~1000
Number	4755	638	54	36	28

The number of people who viewed the course page more than 1,000 times is shown in Table 2. It can be seen that the number of people who viewed the course page more than 1,000 times only accounts for 2.1% of the total number of people. Among them, about 67.8% of this group viewed the course page less than 4,000 times during the semester.

Table 2. The number of pages of the course page is greater than 1000.

Interval	1000~4000	4000~7000	7000~10000	10000~14000
Number	80	25	10	3

Overall, in terms of course learning frequency, the majority of learners viewed the course pages fewer than 1,000 times per semester. Given that course learning encompasses multiple aspects such as video learning, course assignments, and online exercises, the relatively low number of course page views despite the availability of diverse learning resources suggests that, following the shift from traditional classroom instruction to online learning, learners face conflicts between learning, work, and family responsibilities. Additionally, the absence of teacher supervision and the interactive learning environment associated with traditional classroom settings may lead to reduced learner motivation and lower learning frequency. This could result in learners failing to complete the course, thereby impacting their learning outcomes.

4.1.2. Analysis of learners' learning time preferences

Continue to select 2024 learners' online learning behavior data for sampling analysis. Based on the behavioral data generated through the online learning platform, a total of 615,295 log entries were statistically derived regarding learners' interactions with the course during online learning. The analysis provides an overview of the overall time distribution of students' participation in online learning, with the overall time distribution of course learning shown in Figure 2. As shown in the figure, learners' learning time exhibits a certain degree of regularity. Specifically, learners access the online learning platform more frequently at the beginning and end of the semester, with a notable increase in access volume toward the end of the semester, reaching a peak to some extent. The following explanations can be provided for the aforementioned learning behaviors: first, at the beginning of the semester, learners need to log in to the online learning platform for course selection, payment, and other tasks, and subsequently require some understanding and learning of the courses they are taking. Therefore, during this phase, learners have a higher volume of access to the courses they are studying. Second, toward the end of the semester, as final exams approach, learners need to complete required tasks such as course assignments and self-assessment exercises, and prepare adequately for the final exams through online Q&A sessions and discussions to achieve satisfactory academic results. Therefore, toward the end of the semester, learners' access to the learning platform increases significantly and becomes more concentrated. Finally, during other periods, learners may have limited time for online learning and lower engagement due to conflicts between work, family, and study. Therefore, teachers should provide reasonable learning activities and teaching plans to enhance learners' enthusiasm for online learning. Regularly monitor students' online learning progress and increase the proportion of daily learning activities in exam evaluations. During periods when learners show lower interest in learning, teachers should intervene appropriately to enhance learners' initiative and enthusiasm for learning.

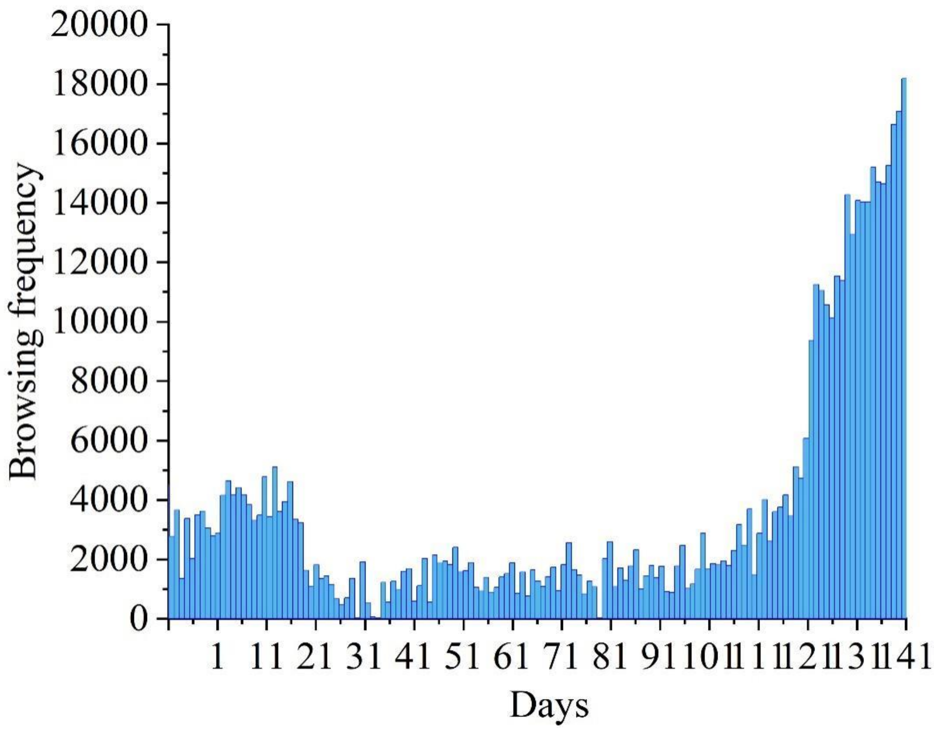

Figure 2. The course studies the overall time distribution.

In addition to tracking course browsing activity, we also analyzed discussion activity within the online learning platform, identifying patterns similar to those observed in course browsing. Learners' discussion and evaluation activity was concentrated in the first and last months of the semester, with nearly zero discussions occurring during other periods. The overall discussion activity statistics are shown in Figure 3. At the beginning of the semester, learners need to understand relevant course information, and interactions occur between teachers and students, as well as among students, enabling them to gain a comprehensive understanding of the course, thus participating in discussions and evaluations. By the end of the semester, discussions increase in number to complete discussion tasks required during the learning process. However, unlike course browsing activity, online learning lacks the emotional exchange effects of face-to-face instruction, resulting in a loss of the immediacy of teaching. Additionally, interactions between teachers and students are often not subject to assessment requirements, leading to a relatively low overall number of posts by learners. In particular, the level of learning interaction participation during the mid-semester period requires improvement, and the enthusiasm of most learners to participate in discussions needs to be enhanced. Therefore, teachers should regularly remind students to actively participate in course discussions to foster a positive interactive learning environment.

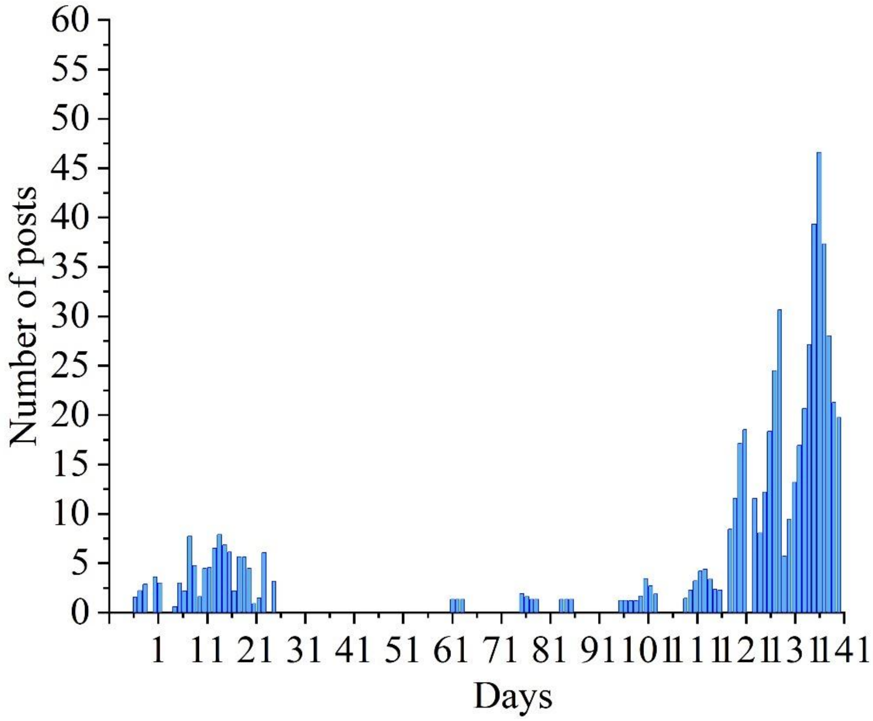

Figure 3. Overall discussion statistics.

Overall, in terms of study time, learners' frequency of online learning initially rises rapidly, then declines and stabilizes, before surging again toward the end of the semester. This indicates that learners' time allocation during online learning is unreasonable, specifically manifested in lower levels of participation and enthusiasm during the mid-semester period, with learners only dedicating significant time and effort to complete academic exams at the end of the semester. Such a learning approach is inconsistent with the principles of effective learning. Therefore, teachers should incorporate diverse learning activities into their instruction, enhance synchronous interaction between teachers and students, and implement reasonable measures such as monitoring learners' progress, understanding the challenges and needs they encounter during learning, to stimulate learners' enthusiasm for learning. This should encourage balanced participation in learning, provide personalized learning services, and intervene in the learning process to help learners reasonably allocate their time, actively and proactively engage in learning, and develop good learning habits.

4.2. Algorithm Practice

The results of the K-means clustering analysis of entrepreneurial ability are shown in Figure 4. When we quantify the data of some of the students who participated in the survey, write a program, and import it into the algorithm mentioned in this paper, we can see that after using the K-means clustering algorithm, we can well describe the three important post-graduation paths of choosing entrepreneurship, choosing employment, and choosing further education among graduates. This also indirectly verifies that different post-graduation paths will inevitably differ in certain factors. Entrepreneurship education teachers can utilize the results obtained through algorithms based on students' big data to effectively stimulate their entrepreneurial potential.

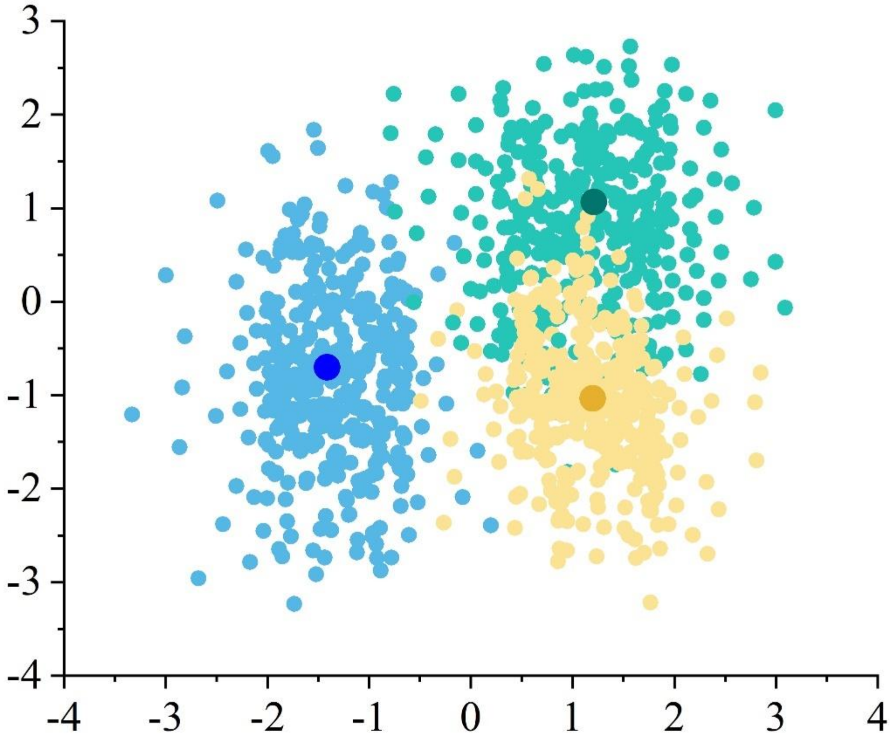

Figure 4. K-means clustering analysis of entrepreneurial ability clustering results.

To verify the universality of the algorithm, the study selected 200 sophomore students and conducted a one-month innovation and entrepreneurship course training program. At the beginning of the course, a questionnaire survey was conducted to collect relevant information from the 200 students. After processing the designated information, the initial clustering distribution of the students is shown in Figure 5. As shown in Figure 5, before the course, only 8 students demonstrated a certain level of entrepreneurial thinking, while the remaining students lacked various aspects required for entrepreneurial thinking to varying degrees.

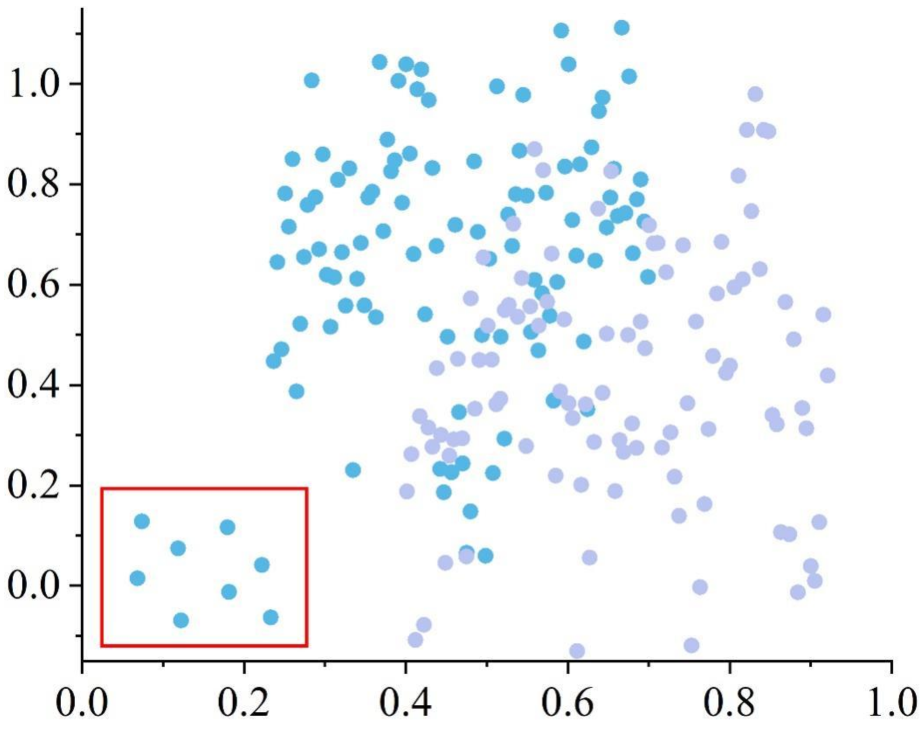

Figure 5. Student clustering in start stage.

Based on the initial results obtained, targeted entrepreneurship education was provided to each student. After the course concluded, the 200 participants were again evaluated using the entrepreneurial thinking algorithm. The final student cluster distribution is shown in Figure 6. It can be observed that after one month of targeted entrepreneurship education, 16 students demonstrated a certain inclination toward entrepreneurship and developed preliminary entrepreneurial thinking. The experiment demonstrated that using the K-means clustering algorithm to conduct quantitative analysis of vocational college students and implement targeted entrepreneurship education yields excellent educational outcomes.

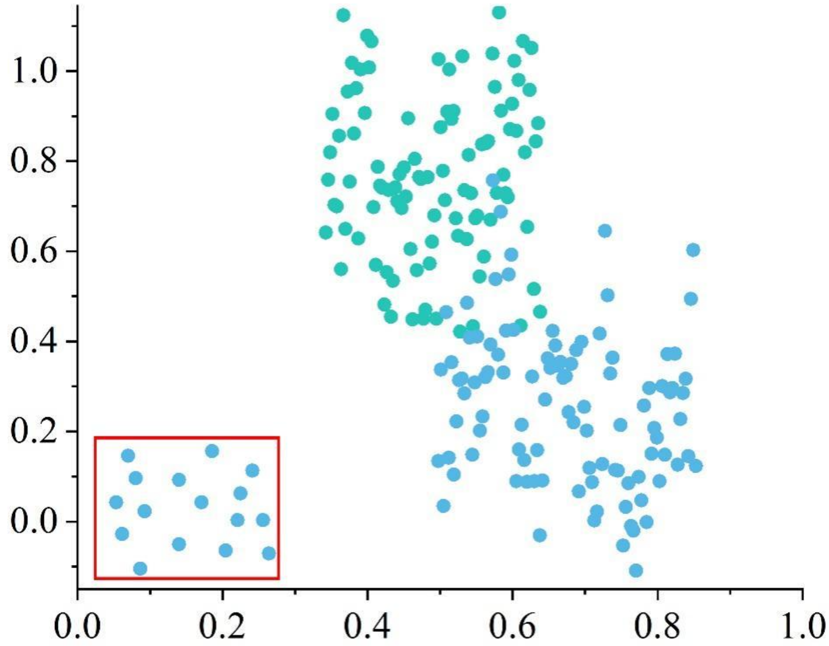

Figure 6. The end stage student cluster distribution.

4.3. Analysis of Student Learning Behavior Characteristics

The experiment continued to select learning behavior data from the online platform courses of the 2023 and 2024 classes of vocational college students in innovation and entrepreneurship courses for analysis.

4.3.1. Descriptive Analysis of Learning Behavior

Passive learning behavior data were included in the calculation of regular grades in this study, specifically video viewing, course access, assignment performance, interaction, and regular performance. The results of the descriptive analysis of learning behavior for the 2023 and 2024 cohorts are shown in Table 3. Video viewing data remained largely consistent, with the 2024 cohort showing slightly higher scores. The 2023 cohort achieved higher scores in course interaction, as most students were experiencing online instruction for the first time, leading to a sense of novelty and higher engagement. The 2024 cohort scored lower, as most courses were conducted offline, resulting in lower student activity levels and greater difficulty in course interaction, leading to a lower average score. Standard deviation section: The 2024 cohort showed more consistent overall performance, particularly with a standard deviation of only 1.05 for course video viewing. This aligns with the requirement for students to complete videos before or during class breaks as per the teacher's offline instruction guidelines.

Table 3. Descriptive analysis of behavioral behavior.

		Video viewing	Course access	Job performance	Interaction situation	Normal performance
Mean	2023	95.54	70.55	95.65	38.65	77.54
	2024	96.33	75.6	88.66	14.65	84.69
Standard	2023	13.66	15.96	11.25	22.65	26.76

deviation	2024	1.05	8.89	22.59	8.63	15.55

4.3.2. Analysis of Learning Behavior Correlations

This study conducted a correlation analysis of the learning behaviors of students in two cohorts. The results of the Spearman correlation analysis of learning behaviors are shown in Figure 7. Course interaction involved only a small number of students each time, and only students who had a good grasp of the course content were able to earn points. This had the highest correlation with final grades, at 0.455. Course interaction and video viewing had a relatively high correlation of 0.388. During the course, it was found that students with lively personalities were more willing to participate in course interaction.

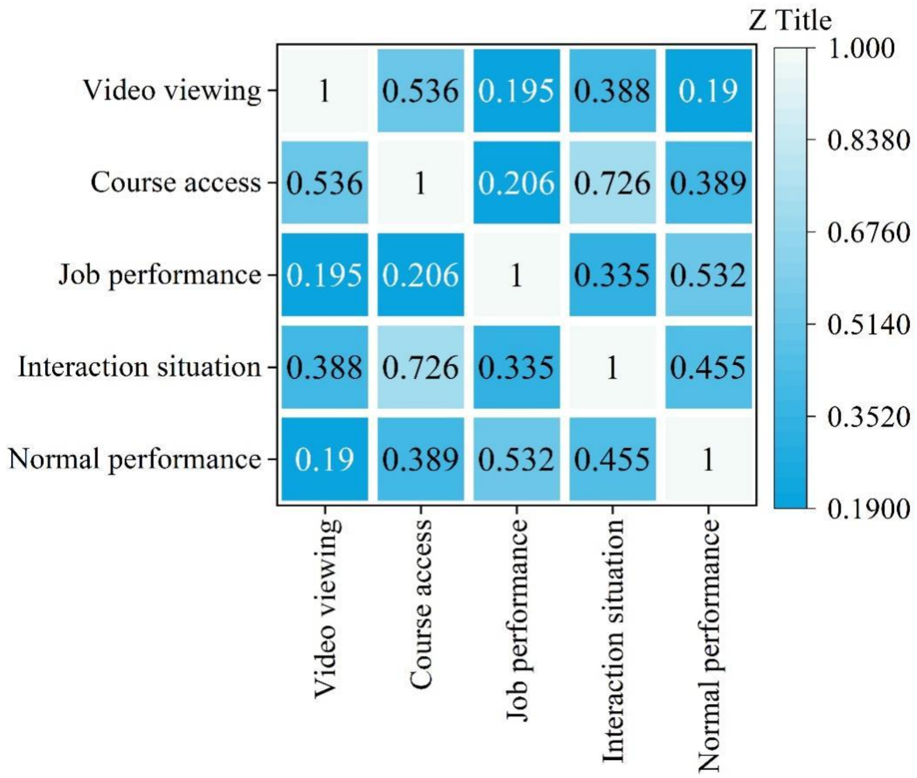

Figure 7. Spearman correlation analysis results.

4.3.3. Factor Analysis of Learning Behavior

A factor analysis was conducted on the statistical data provided by the online platform for the two cohorts of students (five variables: video viewing, course access, assignment performance, interaction, and daily performance). The KMO value was 0.705, and the significance of Bartlett's sphericity test was <0.001, which is less than the significance level of 0.05. The common factor variance and component matrix are shown in Table 4. As shown in the table, the extracted common factors exhibit high commonality with assignment performance, interaction, and regular performance, indicating that the common factors focus on extracting data related to students' learning abilities during the teaching process, providing a basis for classification in cluster analysis.

Table 4. Common factor variance and composition matrix.

Variable	Common factor variance	Constituent
Video viewing	0.366	0.576
Course access	0.241	0.514
Job performance	0.601	0.765
Interaction situation	0.6	0.772
Normal performance	0.592	0.772

4.3.4. Cluster Analysis

Based on the results of factor analysis, students' comprehension abilities and learning attitudes were selected as the classification criteria. Using a systematic clustering method, combined with students' actual performance in class, the final number of categories was determined to be 3. These three categories primarily reflect differences in students' learning attitudes, and are therefore classified as: proactive, passive, and negative. The results of the cluster analysis and their relationship with final grades are shown in Table 5. In the table, the distribution of students across the three types follows a typical educational pattern, with the majority in the middle and smaller numbers at the extremes. The average final grades for the three types of students increase in a stepwise manner from high to low, consistent with the principle that effort in learning is proportional to outcomes. The standard deviation for the negative and passive types is slightly higher, indicating more scattered grades. The positive type has the highest average final grade, with minimal variation within the category.

Through cluster analysis of two cohorts of students, it was found that proactive students had final exam scores above 80. In the teaching process, it was observed that proactive students had stronger self-control, were willing to invest effort into learning, engaged in more post-class communication with instructors, actively addressed in-class difficulties, and could exceed expectations and diligently complete both online and offline teaching tasks assigned by instructors. Passive students constituted the largest proportion at 87.5%, while negative and proactive students were relatively few in number. Among them, the standard deviation of final grades for passive and negative students was slightly higher. Based on the actual circumstances of the students, it was found that a portion of these two types of students had strong comprehension abilities but weak self-control, and their learning attitudes were not proactive or sustainable, leading to more scattered grades. Some students in the passive and negative groups scored above 80 points, able to learn independently through other methods and paying less attention to the teacher's online platform requirements. Compared to the proactive group, the performance gap is significant, indicating that teachers play a significant guiding role in the teaching process. In addition to enhancing students' participation in online platform courses, teachers also guide students to focus on classroom instruction.

Table 5. Cluster analysis results and final results.

Student category	number	Proportion (%)	Final score mean (score)	Standard deviation
Negative type	17	8.5	55.59	16.6
Passive type	175	87.5	71.37	15.75
Positive type	8	4	91.97	6.51

5. Intelligent Development Model for Cultivating Innovative and Entrepreneurial Talent in Higher Vocational Education

5.1. Building a comprehensive system for cultivating innovative and entrepreneurial talent

In the cultivation of innovative and entrepreneurial talent, to achieve this goal, it is essential to establish a comprehensive system for cultivating such talent. This system must integrate the educational objectives for cultivating innovative and entrepreneurial talent across all educational levels and consolidate the superior resources available for cultivating such talent at each stage, thereby maximizing the overall effectiveness of talent cultivation. Education authorities should develop a technical preparation plan for innovation and entrepreneurship talent cultivation, systematically integrating course resources for innovation and entrepreneurship talent cultivation across different educational stages based on the needs of high-tech talent cultivation, to ensure the quality of talent cultivation.

5.2. Apprenticeship system for cultivating innovative and entrepreneurial talent

In the innovation of talent cultivation models for innovation and entrepreneurship, it is essential to strengthen the application of modern apprenticeship-based talent cultivation methods. By adopting modern apprenticeship systems in the innovation of talent cultivation models for innovation and entrepreneurship, we can effectively address students' future employment challenges, foster the development of their specialized skills, and cultivate their technological innovation capabilities through the transmission of skills between mentors and apprentices, as well as the dissemination of long-term experience.

When establishing modern apprenticeship systems, schools should collaborate closely with enterprises to clarify the objectives of innovation and entrepreneurship talent cultivation. They should start from the actual needs of students in innovation and entrepreneurship programs, with the goal of cultivating students' comprehensive professional capabilities. This involves organically integrating theoretical education models with professional practical education models to ensure that theoretical knowledge and skill learning are aligned with students' job-related capability requirements, and that practical activities are aligned with students' job capability needs. The professional competencies of innovation and entrepreneurship students should be analyzed and designed based on the needs of the intelligent era. Through school-enterprise collaboration, the specific professional qualities and competency requirements for cultivating innovation and entrepreneurship talent should be clearly defined. Through cooperation, schools and enterprises should reasonably set up on-the-job internship activities, develop comprehensive job practice content, and establish a complete talent cultivation management plan. They should design job management based on industry needs and adjust talent cultivation plans based on the needs of the era to ensure that the cultivated innovation and entrepreneurship talents can truly meet the needs of specific job positions and the demands of the intelligent era.

5.3. Building a comprehensive innovation and entrepreneurship curriculum system

In response to the new requirements for cultivating innovative and entrepreneurial talent, it is essential to establish a comprehensive innovation and entrepreneurship course system. This system should actively explore methods for developing the professional competencies of innovation and entrepreneurship students and establishing the organizational framework for innovation and entrepreneurship courses. Such innovations must be tailored to the needs of the times to ensure the cultivation of high-quality, application-oriented innovative and entrepreneurial talent.

In the development of innovation and entrepreneurship courses, it is crucial to thoroughly research the demands of intelligent production systems on the professional competencies of innovative and entrepreneurial skilled talent. Based on this research, a comprehensive educational curriculum for innovation and entrepreneurship courses should be constructed. It is necessary to conduct a systematic analysis of work systems within the field of innovation and entrepreneurship. Based on the results of this analysis, a comprehensive set of course content for work system units should be developed to ensure the cultivation of students' systematic capabilities. Efforts should be made to strengthen the construction of practical education content in innovation and entrepreneurship. Guided by the requirements for work system capabilities, comprehensive practical education content should be developed to ensure the operability of each comprehensive practical education module. The effective implementation of this content will ensure the cultivation of students' comprehensive skill competencies. In the development of innovation and entrepreneurship courses, it is important to emphasize the integration of relevant content, based on the needs of cultivating students' professional capabilities, to improve and enrich related educational content. Based on in-depth research into innovation and entrepreneurship work models, it is necessary to conduct in-depth exploration and enrichment of related professional education content to ensure the comprehensive development of relevant content.

6. Conclusion

In recent years, the development of entrepreneurship education in China has entered a phase of deepening and enrichment. With the support of national policies and the active cooperation of local enterprises, entrepreneurship education in higher vocational education has received sufficient attention. This paper attempts to use the K-means algorithm to analyze student learning behavior data and proposes corresponding talent cultivation strategies based on the experimental results.

After conducting targeted entrepreneurship education for 200 students, it was found that 16 students exhibited a certain degree of entrepreneurial inclination and developed preliminary entrepreneurial thinking, an increase of 8 students compared to before the targeted innovation and entrepreneurship education. The experiment demonstrates that using the K-means clustering algorithm to conduct quantitative analysis of vocational college students and implementing targeted entrepreneurship education yields good educational outcomes.

Through cluster analysis of two cohorts of students, their understanding abilities and learning attitudes can be categorized into proactive, passive, and negative types. Among these, passive students account for the largest proportion at 87.5%, while negative and proactive students are relatively few in number.

In summary, under the backdrop of industry-education integration, innovation and entrepreneurship talent cultivation also requires innovation in talent cultivation models based on the needs of intelligent production, actively transforming traditional talent cultivation models, and exploring innovative talent cultivation measures. It is necessary to construct a comprehensive talent cultivation system based on the needs of the times, actively implement the modern apprenticeship talent cultivation model, and establish a comprehensive curriculum system. Through multifaceted efforts, the effective cultivation of innovation and entrepreneurship talent must be ensured.

Funding
This research was supported by the Jiangxi Higher Education Association project，titled”Research on the Innovative Talent Cultivation Model Driven by the Dual Wheels of Science and Education Integration and Industry-Education Integration”。Number:PX-2251441. This research was supported by the Jiangxi province education science planning general project and Jiangxi Higher Education Association project， titled”Research on the Path of Integrated Education of Enrollment, Training and Employment in Higher Vocational Colleges from the Perspective of Industry-Education Integration”。Number:645253.

References

1. Guo, J., Khatibi, A., & Tham, J. (2024). The Role of Innovation And Entrepreneurship Education In Promoting The Growth Of Students In Higher Vocational Colleges. Library of Progress-Library Science, Information Technology & Computer, 44(3).
2. Liu, C. (2025). Evaluation of vocational education innovation and entrepreneurship education based on fuzzy information. Australian Journal of Electrical and Electronics Engineering, 1-16.
3. Li, Y., Wei, H., Wei, S., & Zhang, L. (2021, April). Research on innovation and entrepreneurship of higher vocational education based on citespace. In Journal of Physics: Conference Series (Vol. 1883, No. 1, p. 012154). IOP Publishing.
4. Yang, Y. (2018). Teaching research on higher vocational pre-school education of professional art course based on innovation and entrepreneurship education. Creative Education, 9(05), 713.
5. Chu, Y. (2022). Research On The Improvement Of College Students’Anxiety By The Integration Mode Of Industry And Education Of Vocational Education In The New Era. Psychiatria Danubina, 34(suppl 1), 879-880.
6. Hu, K., & Li, X. (2024). Research on Industry-education Integration and Talents Cultivation of Vocational Education Empowering New Quality Productive Forces. Journal of Higher Vocational Education (ISSN: 3005-5784), 1(2), 211.
7. Choy, S., Wärvik, G. B., & Lindberg, V. (2018). Integration of vocational education and training experiences. Springer Singapore.
8. Wang, J. (2024). Developing Vocational Education and Deepening the Integration of Industry and Education Providing Human Resources Support for Building a Strong Provincial Centre-Taking the Development of Vocational Education in Sanmenxia as an Example. Journal of Social Science Humanities and Literature, 7(5), 49-54.
9. Liu, X., Liu, Y., & Wan, Z. (2015). Difficulties of college students’ business startups under economic new normal and their countermeasures. Cross-cultural communication, 2015, 107-112.
10. Ooi, Y. K., & Ahmad, S. (2012). A study among university students in business start-ups in Malaysia: Motivations and obstacles to become entrepreneurs. International Journal of Business and Social Science (IJBSS), 3(19), 181-192.
11. Kvedaraitė, N. (2014). Reasons and obstacles to starting a business: Experience of students of Lithuanian higher education institutions. Management-Journal of Contemporary Management Issues, 19(1), 1-16.
12. An, C., & Ma, Y. (2024). Research on the Strategy of Promoting College Students’ Employability in Student Work. Journal of Humanities, Arts and Social Science, 8(6).
13. Chu, J., Liu, J., & Liu, S. (2019). Research and countermeasure analysis of college students’ employment issues. Open Journal of Social Sciences, 8(1), 129-138.
14. Liu, Y., Gao, J., & Liang, Z. (2024). Research on the Working Mechanism of Counselors in Vocational Undergraduate Colleges to Promote High-Quality Employment of College Students. Journal of Modern Education and Culture, 1(3).
15. Yan, L. (2025). Research on the Core Employment Ability of University Students in Applied Universities and the Promotion Path of Employment Guidance Work in the New Era—Taking Chengdu Technological University as an Example. Journal of Advanced Research in Education, 4(2), 56-63.
16. Alferaih, A. (2022). Starting a new business? Assessing university students’ intentions towards digital entrepreneurship in Saudi Arabia. International Journal of Information Management Data Insights, 2(2), 100087.
17. Sulistyowati, R. (2021). The effectiveness of the Business Incubator and Entrepreneurial Education in Interest to Start a Business in Vocational School students majoring in Marketing. Studies in Learning and Teaching, 2(1), 29-40.
18. Bergmann, H., Hundt, C., & Sternberg, R. (2016). What makes student entrepreneurs? On the relevance (and irrelevance) of the university and the regional context for student start-ups. Small business economics, 47, 53-76.
19. Liu, H., Gao, F., & Wang, L. (2022). The Practice and Reflection on the Cultivation of New Technology Applied Entrepreneurial Talents in Higher Vocational Colleges. International Journal of Social Science and Education Research, 5(9), 307-314.

20. Xiong, T., Zhang, J., & Huang, H. (2023). Entrepreneurship education for training the talent in China: exploring the influencing factors and their effects. Sustainability, 15(15), 11664.
21. Liu, Y. (2022). Construction of talent training mechanism for innovation and entrepreneurship education in colleges and universities based on data fusion algorithm. Frontiers in Psychology, 13, 968023.
22. Zhao, X., & Wang, X. (2022). The influence of college entrepreneurship education system on the cultivation of applied innovative talents. Frontiers in Psychology, 13, 844234.
23. Xiao, X., Zhu, Y., Guo, Z., Ma, Y., Zhang, Z., Cao, L., ... & Wang, W. (2025). An Innovation Talent Cultivation Mechanism for Robotics in the Digital-Intelligent Era: Exploration and Practice at Wuhan University. Frontiers of Digital Education, 2(1), 1-7.
24. Cai, X. (2024). Digital Intelligent Transformation and Promotion Path of New Business Talent Training Model under the Background of Artificial Intelligence. Journal of Industry and Engineering Management (ISSN: 2959-0612), 2(1), 51.
25. Han, K., & Zhang, L. (2021, December). Exploration on the path of cultivating innovative talents under the background of intelligent era. In 2021 International Conference on Forthcoming Networks and Sustainability in AIoT Era (FoNeS-AIoT) (pp. 270-274). IEEE.
26. Wang Shaohua & Xu Xiaoxiong. (2023). A retrieval method of learners' behavior features based on K-means clustering algorithm. Cluster Computing,27(2),2049-2058.
27. Xiangyuan Liu. (2024). Push method of online learning resources based on user behaviour characteristics. International Journal of Business Intelligence and Data Mining,24(3-4),324-339.
28. Christy Jacqueline & Ranjith K. (2024). A Framework for Performing an Analysis on Behavioral traits using Machine Learning.(eds.)Research Scholar, Department of Computer Science, Karpagam Academy of Higher Education, Coimbatore; Assistant Professor Research Guide, Department of Computer Science, Karpagam Academy of Higher Education, Coimbatore.