International Journal for Housing Science and Its Applications
Publish August 5, 2025. Volume 46, Issue 3 Pages 4286-4296

https://doi.org/10.70517/ijhsa463366

# Research on the Development Path of Intelligent Student Innovation and Entrepreneurship Education Platform Constructed on the Basis of Big Data in Career Planning

**Gang Wang[1,*], Yongning Qian[2], Jing Zhao[3] and Yifan Xue[4]**

[1] The Office of Student Affairs, College Student Employment Guidance Center, School of Innovation and Entrepreneurship, Shaanxi Technical College of Finance and Economics, Xianyang, Shaanxi, 712000, China
[2] The School of Humanities and Arts, Shaanxi Technical College of Finance and Economics, Xianyang, Shaanxi, 712000, China
[3] The School of Accounting, Shaanxi Technical College of Finance and Economics, Xianyang, Shaanxi, 712000, China
[4] The Academic Affairs Office, Shaanxi Technical College of Finance and Economics, Xianyang, Shaanxi, 712000, China

Corresponding authors: (e-mail: wgangqq@sina.com).

**Abstract** With the rapid development of economy and society, the employment pressure of college students increases year by year. In order to improve the employment competitiveness of college students, innovation and entrepreneurship education has become an important part of higher education. This paper constructs a big data platform for innovation and entrepreneurship education in colleges and universities, which includes data collection sources, data analysis, and data prediction. Apriori algorithm is used to correlate positive-willing students with negative-willing students, and K-means algorithm is used to assess students' entrepreneurial thinking and explore new ways to personalize innovation and entrepreneurship education. The results of the study show that 70.3% of students believe that they choose to start their own business after graduation, and that being able to grasp entrepreneurial policies in a timely manner during college is an important reason for their decision to start their own business, and that the combined effect of motivational and resistance factors affects the willingness of college students to improve their entrepreneurial ability through participation. The results obtained by using the K-means algorithm provide colorful and targeted career planning development paths for students with different characteristics, which provides a good development direction for the personalized teaching method of innovation and entrepreneurship education.

**Index Terms** innovation and entrepreneurship education, decision tree algorithm, K-means algorithm, career planning

## I. Introduction

Innovation and entrepreneurship education, or “double creation”, which is centered on the concept of “mass entrepreneurship and innovation”, aims to cultivate the innovation consciousness and entrepreneurial ability of college and university students, and equip them with professional ability and comprehensive quality to adapt to the society [1], [2]. With the advancement of educational reform in higher education institutions, the importance of innovation and entrepreneurship education has become more and more prominent, which reflects the basic direction and main goal of educational reform in colleges and universities [3]. Career planning, on the other hand, is the basic expectation and scientific planning for future career of college students in view of the elements such as professional characteristics, employment demand and their own ability [4]. For students, the correct career planning will directly affect the quality of their learning and employment quality [5], [6]. From the existing experience, innovation and entrepreneurship education has an important impact on students' career planning, and colleges and universities should pay attention to innovation and entrepreneurship education, exert its positive influence, and guide students to make scientific and reasonable career planning according to the professional needs, real abilities and industry prospects, etc [7]-[9].

However, the current innovation and entrepreneurship projects of college students cultivated by some universities present problems such as low scientific and technological content, insufficient market-oriented analysis, and low success rate, which is due to the lack of a platform for sharing and exchanging information on innovation and entrepreneurship [10], [11]. Some college students are more difficult to obtain the information of innovation and entrepreneurship hotspot information, market statistics, patent technology data and other information, and the cultivated innovation and entrepreneurship project only stays at the level of college entrepreneurship, with serious homogenization of the project and low technological content [12], [13]. Therefore, with the support of big data platform for innovation and entrepreneurship education, the correlation between data can be found from the massive

data, and the connotative value of the data can be deeply excavated, thus analyzing the past, grasping the present and predicting the future, and effectively serving the innovation and entrepreneurship education [14], [15]. In the future, big data will still have a significant impact on innovation and entrepreneurship education, and it is necessary to continue exploring and researching it at the specific application level to play its great role.

The article constructs the framework of big data platform for innovation and entrepreneurship education in colleges and universities from three parts: data collection and storage, data analysis, and prediction. Decision tree algorithm combined with feature selection algorithm is used to realize data collection, Apriori algorithm is used to correlate and analyze students' willingness to improve entrepreneurial ability through participation, and by introducing K-means algorithm as the core method, it is imported into the innovation and entrepreneurship education big data intelligent platform, and targeted entrepreneurship education is carried out in different situations of each student. This paper points out the development path of future career for college students, prompts college students to have clearer and firmer learning goals, and helps college students reduce their psychological burden. At the same time, colleges and universities should also clarify the nurturing purpose of innovation and entrepreneurship education.

## II. Construction of a big data platform for innovation and entrepreneurship education in higher education institutions

### *II. A. Framework of Big Data Platform for Innovation and Entrepreneurship Education in Colleges and Universities*

The big data platform framework is mainly composed of three parts: data collection and storage, data analysis, and prediction as shown in Figure 1. The collection and storage of massive data provides the data source for the big data platform for innovation and entrepreneurship education in colleges and universities, which is the foundation and source power of the whole platform; data analysis processes the data and digs deep into the connotative value of the data. Big data can analyze the past, grasp the present and predict the future, effectively serving innovation and entrepreneurship education.

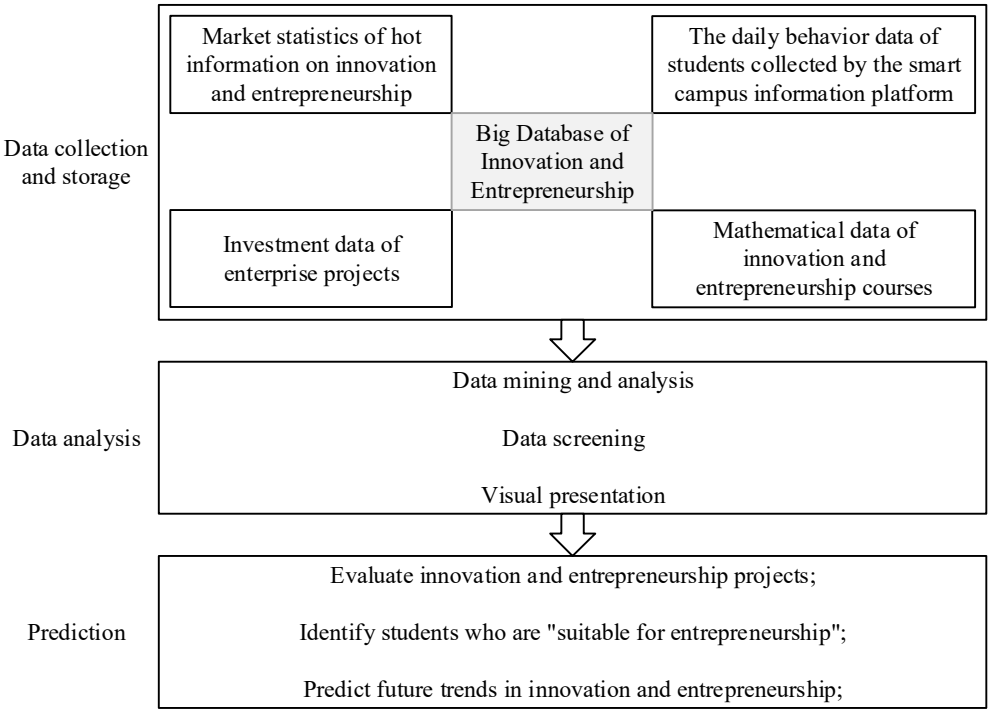


Figure 1: University innovation entrepreneurship education is based on the platform

### *II. B. Data collection sources*

#### II. B. 1) Hot information and market statistics on innovation and entrepreneurship

Collect multi-channel information collection on innovation and entrepreneurship categories, such as national policies, scientific and technological innovation achievements, etc. Market statistics are sourced from the National Bureau

of Statistics, publicly available data sources from local government portals, and databases of economic and management disciplines.

#### II. B. 2) Data on Students' Daily Behavior in Innovation and Entrepreneurship Education

The data collection of the Smart Campus Information Platform is mainly through two channels: first, the databases of each administrative department of the school, such as the database of students' course learning and performance, the database of students' daily management work, and the database of the transcripts of the “second course”, etc.; and second, the data collected through the trajectory of the students' school activities and behaviors, such as the following The second is data collected through students' trajectories and behavioral habits in school, such as data on library books and study hours, innovation and entrepreneurship activities, network usage, and consumption behavior in school. Colleges and universities can establish an integrated and unified student data information platform to provide data support for student education, management and service.

#### II. B. 3) Data on teaching and investment in innovative entrepreneurship programs

(1) Teaching data of innovation and entrepreneurship courses

The main sources of data are: first, the teaching data of innovation and entrepreneurship general course, and second, the teaching data of innovation and entrepreneurship virtual simulation experiment.

(2) Investment data of investment enterprise projects

It mainly comes from the data of investment in entrepreneurial projects by national science and technology parks, financial institutions, intermediary organizations and so on.

### *II. C.Analysis and projections*

#### II. C. 1) Data analysis

Data analysis is the processing of collected data information. Relying on the powerful computing power of cloud computing, it provides users of big data platform with a number of data analysis functions. In the Data Analysis Center, users of the big data platform can design different mathematical models to explore and analyze the potential value of the data, such as discovering new market opportunities, consumer consumption behaviors and preferences, competitor analysis, and government policy regulation trends. Data analysis can better restore and simulate the real entrepreneurial environment, so that college students can cultivate innovative entrepreneurial projects no longer on paper, but based on the results of data analysis. Finally, the data analysis center can use visualization technology to present the results of data analysis, and users can intuitively analyze, utilize and modify the data.

#### II. C. 2) Projections

The predictive function of innovation and entrepreneurship education big data platform is mainly reflected in three aspects: First, the big data platform can evaluate the innovation and entrepreneurship projects, and realize the functions of entrepreneurial program guidance, project risk assessment, project feasibility assessment and so on. Secondly, through the collection of data generated spontaneously by students in the process of innovation and entrepreneurship activities, the innovation and entrepreneurship character portrait model is established, and students with “innovation and entrepreneurship potential” are identified by cloud computing, providing data support for educators. Thirdly, educators can integrate national and corporate information resources on innovation and entrepreneurship, analyze future innovation and entrepreneurship projects that are in line with national economic development by using the prediction model of big data, and provide targeted counseling to college students to cultivate innovation and entrepreneurship projects.

## III. Algorithm for data collection of innovation and entrepreneurship case data in career planning

### *III. A. Data Collection and Classification of Innovation and Entrepreneurship Case Data*

#### III. A. 1) Apriori algorithm

The association rule is defined as follows: suppose $I=\{i_1,i_2,i_3,...,i_m\}$, where $I$ is the set of items, given a database of transactions; $i_m$ is the $m$th item of data. $D=\{t_1,t_2,t_3,\ldots,t_m\}$, where each transaction $t$ is a non-empty subset of $I$, i.e., $t\in I$, and each transaction corresponds to a unique identifier TID.

The association rule is an implication of $X\Rightarrow Y$, where $X,Y\in I$ and $X\cap Y=\phi$, $\phi$ is the empty set, and $X$ and $Y$ are referred to as the prior (LHS) and the successor (HS) of the association rule, respectively.

Probability $P(X\cap Y)$: the support of the association rule $X\Rightarrow Y$ in $D$ is the percentage of $D$ where the transaction includes $X\cap Y$; the confidence level is the percentage that includes $X\cap Y$.

$P(Y|X)$ is the conditional probability, which indicates the probability of $Y$ under the condition of $X$, and the confidence, i.e., the reliability, which can be judged by its association characteristics.

Association rules can be used to mine the data: 1) Frequent itemset query. 2) Strong association rules are generated from frequent itemsets.

Apriori is an iterative method of layer-by-layer search, and $k$ itemsets are used to explore $(k+1)$ itemsets. Combined with data processing, the set is reached by itemset query, counted as $L_1, L_2$, which can be generalized as a collection of 2-item sets, and $L_2$ is used to search for $L_3$ until $k$-itemsets are reached.

The Apriori algorithm improves the efficiency of frequent itemsets by analyzing them based on frequent occurrences, all of whose non-empty subsets must be frequent, and determining whether the threshold min_sup is satisfied with the itemset $I$. If $I$ satisfies the threshold, then $I$ is frequent, i.e., $P(I) < \min\_\sup$. The set of terms (i.e., $I'A$) cannot be more frequent than $I$ [16]. Neither is frequent, i.e., $P(I'A) < \min\_\sup$.

#### III. A. 2) Decision tree algorithm

Decision tree for analyzing the prediction model. Combined with the decision tree needs, can be based on the algorithm using top-down recursion, and combined with the attributes of the judgment, through the decision-making to understand the characteristics of its attributes, and then data mining, the formation of data branching, so that it can solve the classification problem, to reach a prediction, and to achieve the classification.

#### III. A. 3) Feature selection algorithm

Feature selection is a description of an application domain that combines features that are common or relevant to the thing itself. Feature selection selects a minimal subset of $N$ original features, including $M(M \le N)$, such that the probability distribution values of different categories in the subset of $M$ features are close to those of the $N$ original features. If $F_N$ is the original set of features and $F_M$ is the selected subset of features, the conditional probability of the possible category $C$, $P(C|F_M = f_m)$, should be close to $P(C|F_N = f_N)$, with $f_M$ and $f_N$ are the value vectors of the corresponding feature vectors $F_M$ and $F_N$.

Overall, feature selection algorithms can simplify the data description and make the data collection task simpler, which in turn can solve problems and improve the quality of the dataset. When the number of features in a dataset is large, it needs to be cooled down.CFS (Correlation based Feature Subset) algorithm, which is a correlation based feature selection algorithm. For continuous variables, the CFS method measures the correlation in terms of feature subset scores as shown in equation (1).

$$Merit_s = \frac{k\overline{r_{cf}}}{\sqrt{k + k(k-1)\overline{r_{ff}}}} \tag{1}$$

where: $k$ is the number of variables in the subset; $\overline{r_{cf}}$ is the mean value of the correlation between all independent variables and the target variable in the feature subset; $\overline{r_{ff}}$ is the mean value of the correlation between two independent variables in the feature subset.

As for continuous-discrete variables, continuous variables need to be discretized. If the variables after discretization are $X$ and $Y$ respectively, the calculation formula is shown in Equation (2) and Equation (3).

$$H(Y) = -\Sigma p(y)\log_2 p(y) \tag{2}$$

$$H(Y|X) = -\sum_{x\in X} p(x)\sum_{y\in Y} p(y|x)\log_2 p(y|x) \tag{3}$$

where: $p(y)$ is the probability of $y$; $p(y|x)$ is the probability of $Y$ under the condition of $X$; $H(Y)$ is the information entropy of $Y$; $H(Y|X)$ is the information entropy of $Y$ under the condition of $X$.

Then calculate the information gain, i.e., the difference between the a priori information entropy and the a posteriori information entropy, as shown in Equation (4).

$$gain = H(Y) - H(Y|X) = H(Y) + H(X) - H(X,Y) \tag{4}$$

$$symmetrical\ uncertat\ int\ y = 2.0 \times \frac{gain}{H(Y) + H(X} ) \tag{5}$$

The uniform uncertainty between the variables is then calculated as shown in equation (5). If the uncertainty is larger, the correlation is smaller. Evaluating the correlation between the variables in each feature subset enables data categorization based on the features and is more conducive to data processing.

### *III. B. K-MEANS-based clustering model of innovation and entrepreneurship capabilities*

#### III. B. 1) Data preparation and pre-processing

(1) Diversified data were collected from the school, including test scores, classroom interactions, homework completion, and participation in extracurricular activities. These data provide a solid foundation for the comprehensive assessment of students' abilities.

(2) Data cleaning

For the problem of missing data, different strategies are adopted according to the proportion of missing data and the characteristics of data distribution. Small-scale missing is filled in using mean, median, or plural interpolation to maintain data consistency, while large-scale missing considers the importance of the data points and eliminates them if necessary.

For outliers, statistical tests such as the Z-score test or the IQR (interquartile range) method were applied to identify and deal with them. Selection is made to exclude or replace them with values within a reasonable range to avoid their distortion of the clustering results.

Duplicate records were deleted, and the uniqueness of each observation was ensured by comparison with unique identifiers (e.g., student IDs).

(3) Data transformation

Text-based assessment data were converted to numerical scores through sentiment analysis or thematic modeling. Specific methods include using sentiment analysis tools to score the text for affective tendencies, or using thematic modeling to extract key themes in the text and assign corresponding weights.

Time series data, such as daily learning time changes, are converted into fixed-length numerical vectors by extracting key features through time series analysis.

(4) Data standardization

Z-score standardization: make the data conform to a standard normal distribution by subtracting the mean and dividing by the standard deviation. This is calculated as $Z = \frac{(X - \mu)}{\sigma}$, where $X$ is the original data, $\mu$ is the mean, and $\sigma$ is the standard deviation. This method is suitable for continuous variables without clear boundaries.

Min-Max normalization: scaling the data to the interval [0, 1]. The specific calculation method is:

$$X_{\text{norm}} = \frac{(X - X_{\min})}{(X_{\max} - X_{\min})} \tag{6}$$

where: x is the raw data; $X_{\min}$ and $X_{\max}$ are the minimum and maximum values of the data, respectively. This method is applicable to variables with natural upper and lower bounds, such as classroom participation scores.

#### III. B. 2) Evaluation model based on K-means algorithm

When constructing the assessment model of students' innovation and entrepreneurship ability based on K-means algorithm [17], feature selection and extraction as the initial key step, the core of which is to accurately select the feature set that can comprehensively reflect the level of students' innovation and entrepreneurship ability, and the selection of features needs to comprehensively consider its relevance, stability and differentiation, in order to ensure that the accuracy and objectivity of the assessment results.

The core iterative process of K-means algorithm can be expressed as follows:

$$C_i^{(t+1)} = \arg\min_m \sum_{x_j \in S_i^{(t)}} \| x_i - m \|^2 \tag{7}$$

where $C_i^{(t+1)}$ denotes the centroid (center of mass) of the $i$ th cluster after $t+1$ iterations; $m$ denotes the candidate centroid, which traverses all points in the cluster $S_i^{(t)}$ to find the point that minimizes the sum of the squares of distances of all the points to that point; $x_i$ denotes the data points in the cluster $S_i^{(t)}$; $S_i^{(t)}$ denotes the set of all data points belonging to the $i$ th cluster after $t$ iterations; $\|\cdot\|$ denotes the common Euclidean distance.

The choice of K value needs to be based on the intrinsic structure of the dataset, which can be scientifically evaluated by methods such as ElbowMethod or contour coefficient in order to find a balance point that avoids the clustering from being too coarse and also prevents it from being too fine-grained. After completing the K-value determination, we enter the construction and training phase of the model.

# IV. Innovation and entrepreneurship assessment and career planning pathways

## *IV. A. Survey and analysis of innovation and entrepreneurship*

The questionnaire has a total of 10 questions, and the students were selected for the survey. The questions with distinctive features are: timely grasp of entrepreneurship policies, ability to take risks, development of life views and values, systematic grasp of entrepreneurship theory, and strong interpersonal skills. The results of the survey are shown in Table 1, Table 2 and Table 3. Under the topic of "timely mastery of entrepreneurship policy", 70.3%, 56.8% and 43.2% of the entrepreneurial students, employed students and students pursuing further studies chose it respectively. Among the students who chose to start their own business, 70.3% of them chose to agree to grasp entrepreneurship policies in time, which is more than the proportion of the other two categories of students, which also indicates that the college students who chose to start their own business think that they choose to start their own business after graduation, and that the ability to grasp the entrepreneurship policies in time during the college period is one of the most important reasons for their entrepreneurial decision.

Table 1: Select the results of the study of the students

| Topic | Proportion (Unit: %) | | |
|---|---|---|---|
| | Consent | Different meaning | Neutrality |
| Master entrepreneurship policy in time | 70.3 | 16.8 | 17.5 |
| Ability to take risks | 89.8 | 4.7 | 9.8 |
| Culture of conception and values | 66.4 | 21.5 | 15.2 |
| The system holds the theory of entrepreneurship | 81.5 | 9.1 | 12 |
| Strong interpersonal skills | 91.3 | 1.5 | 9.5 |

Table 2: Select the results of the employment student survey

| Topic | Proportion (Unit: %) | | |
|---|---|---|---|
| | Consent | Different meaning | Neutrality |
| Master entrepreneurship policy in time | 56.8 | 31.4 | 14.2 |
| Ability to take risks | 68.1 | 13.9 | 21.4 |
| Culture of conception and values | 60.3 | 20.7 | 23.5 |
| The system holds the theory of entrepreneurship | 61.7 | 11.6 | 30 |
| Strong interpersonal skills | 90.6 | 5.5 | 6.4 |

Table 3: Choose the results of the students' survey

| Topic | Proportion (Unit: %) | | |
|---|---|---|---|
| | Consent | Different meaning | Neutrality |
| Master entrepreneurship policy in time | 43.2 | 45.2 | 17.7 |
| Ability to take risks | 51.6 | 25.3 | 29.2 |
| Culture of conception and values | 60.2 | 23.4 | 23.9 |
| The system holds the theory of entrepreneurship | 37.2 | 42.3 | 25.6 |
| Strong interpersonal skills | 75.7 | 9.4 | 21.2 |

## *IV. B. Data Mining Results and Analysis*

### IV. B. 1) Decision tree feature selection

(1) For the identification of college students' innovative and entrepreneurial quality, three factors, namely, completion degree, own shortcomings and innovative and entrepreneurial thinking, play a decisive role, while the influence of gender and the number of types of projects participated in is not particularly large, and the two factors, grade and major, have almost no influence.

(2) Utilizing the classification rules of decision tree, top-down judgment.

**IV. B. 2) Apriori Algorithm Mining Results and Analysis**

A willingness model is a relational network model that includes both willingness types and motivational factors, and the Apriori algorithm is used to construct a positive willingness model and a negative willingness model based on the two types of willingness, namely, positive willingness type and negative willingness type.

First of all, the Apriori algorithm can be used to uncover the motivational (resistance) factors of the motivation and resistance to participation of positively willing students and negatively willing students (i.e., willingness to improve entrepreneurial ability through participation). The association rule is constructed by linking the employment "willingness" (e.g., positive willingness, negative willingness) and motivation (resistance) factors. The information content at both ends of the association rules was controlled: only motivational (resistance) factors (but not willingness to participate) appeared at the front end, and only willingness to participate but not motivational (resistance) factors appeared at the back end, and an appropriate rule evaluation index was set up in the model, which was used to perform an initial screening of the association rules. Then, the association rules after the initial screening that rank in the top 10 in terms of enhancement are selected from the positive willingness model and the negative willingness model, respectively, to obtain the strong association rules. Finally, the strong association rules network graph relationships in the positive willingness mode and negative willingness mode are laid out by automatic intelligent layout algorithm.

The strong association rule network diagram of positive willingness pattern and the strong association rule network diagram of negative willingness pattern are shown in Fig. 2 and Fig. 3, and the variable codes and attribute names of strong association rules in each willingness pattern are shown in Table 4.

The strong association rules of positive willingness pattern and the strong association rules of negative willingness pattern are analyzed as follows:

(1) The positive willingness pattern strong association rule reflects the motivational characteristics of the positive willingness type students in their willingness to participate in entrepreneurship competitions. In addition to improving their own strengths through entrepreneurship competitions (e.g., "practicing their own abilities"), these students often have a deeper understanding of entrepreneurship, and they consider participating in entrepreneurship competitions a valuable opportunity to test the feasibility of their entrepreneurial ideas. In terms of participation in entrepreneurship competitions, this group of students has high intrinsic motivation and strong self-driven force; they are willing to participate in entrepreneurship competitions even when they encounter difficulties such as lack of experience and immaturity of their projects, and they show determination and confidence not to be discouraged in the face of difficulties.

(2) The negative willingness model has strong correlation rules, which reflects the motivation characteristics of negative willingness students in entrepreneurship competition. Through the automatic intelligent layout algorithm based on graph relations, it is found that the two types of factors, "existing entrepreneurial achievements" and "lack of competition experience", are the characteristic motivation factors and resistance factors of negative willing students, respectively, and "existing entrepreneurial achievements" and "lack of competition experience" are located in the central part of the network relationship graph composed of 10 strong association rules in this model.

From the active-willing students, it can be seen that this type of students have similar problem-solving and stress-resistant abilities (e.g., being aggressive, facing up to difficulties, and being brave enough to face risks) as real entrepreneurs.

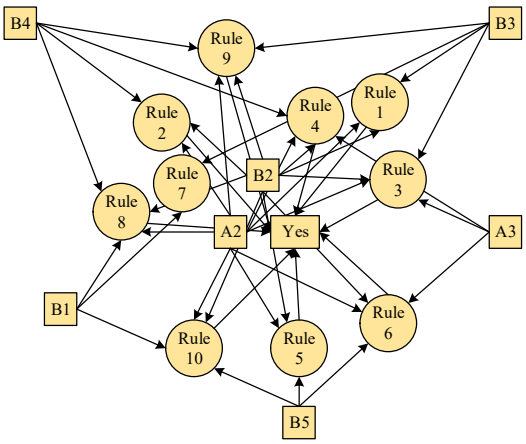


Figure 2: The strong association rules network diagram of the active willingness pattern

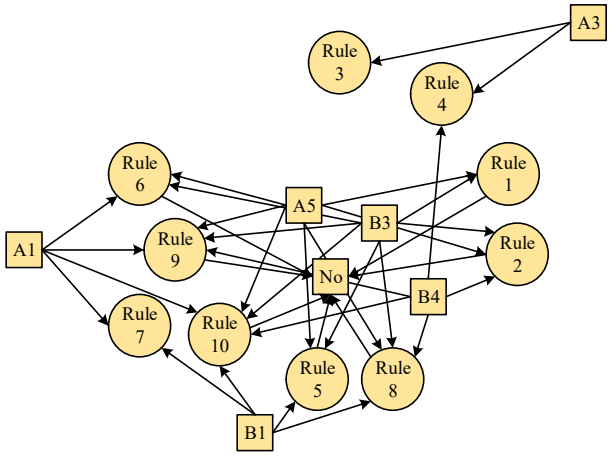


Figure 3: The strong association rules network diagram of the negative willingness pattern

Table 4: Variable code and attribute names for strongly associated rules in each will pattern

| Variable code | Attribute name |
|---|---|
| A1 | award |
| A2 | Testing the idea of entrepreneurship |
| A3 | Exercise oneself |
| A5 | We have started our business and we just played |
| B1 | No suitable field research |
| B2 | Lack of funds |
| B3 | Lack of experience |
| B4 | Without appropriate projects and ideas |
| B5 | No match |
| YES | Willing to compete |
| NO | Unwilling to compete |
| Rule-Rule10 | (Positive/Negative mode) 10 strong association rules |

### *IV. C. Assessment and Analysis of Innovation and Entrepreneurship*

K-means clustering analysis of entrepreneurial ability clustering results are shown in Figure 4. In the figure, some of the students who participated in the survey information for data quantification, and write a program, imported into the algorithm mentioned above, can be found when the use of K-means clustering algorithm, can be very good description of graduates in the choice of entrepreneurship, the choice of employment and the choice of further education, these three important graduation destination. It also indirectly verifies that different graduation destinations are bound to differ in certain factors. Teachers of entrepreneurship education can target the entrepreneurial potential of students based on their big data and the results obtained by the algorithm.

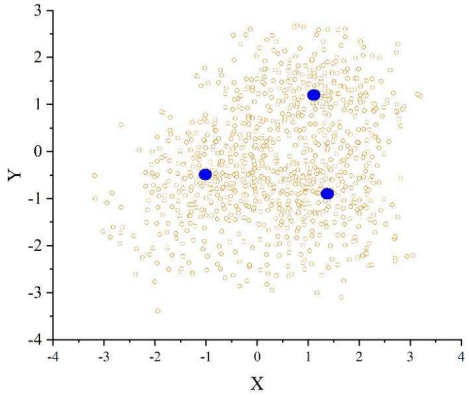


Figure 4: K-means clustering analysis of entrepreneurial ability clustering results

The distribution of students clustered at the beginning stage is shown in Figure 5, from which it can be seen that only six have some entrepreneurial thinking before the instruction, while the rest of the students, to a greater or lesser extent, lack the aspects of entrepreneurial thinking that should be fulfilled. Based on the results obtained at the initial stage, each student was given targeted entrepreneurial education. At the end of the course, the students were again evaluated on the algorithm of entrepreneurial thinking and the results are shown in Figure 6. It can be seen that after one month of targeted entrepreneurship education, a total of 16 people finally have some choice tendency towards entrepreneurship and have initial entrepreneurial thinking. The experiment proves that the use of K-means clustering algorithm, quantitative analysis of college students, and targeted entrepreneurship education has a good teaching effect.

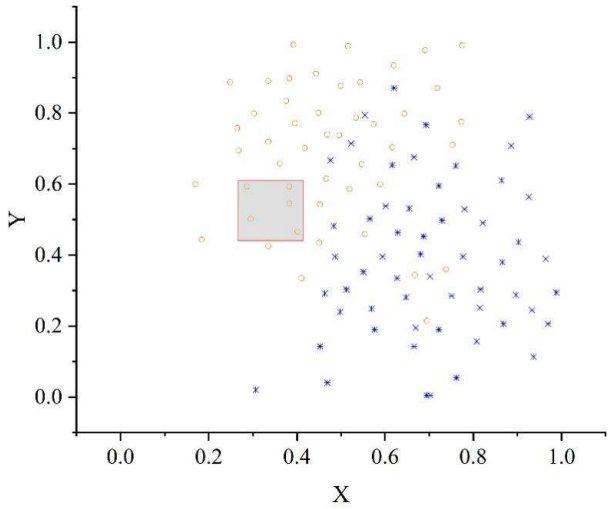


Figure 5: Student clustering distribution at the beginning stage

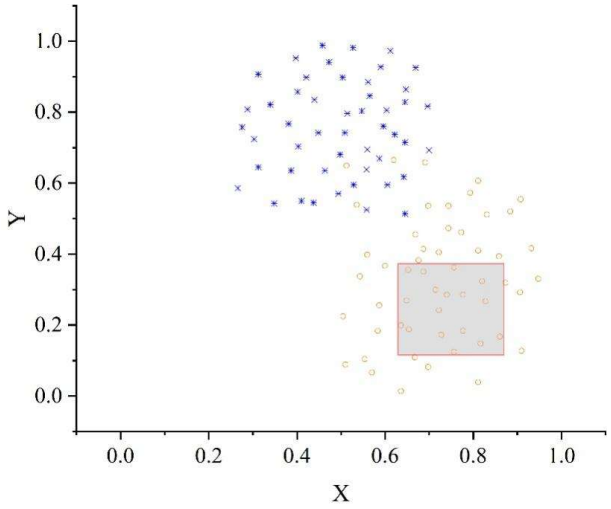


Figure 6: The end stage student cluster distribution

### *IV. D. Career planning paths*

The formulation of career development planning for students is mainly a series of measures taken to ensure the realization of career development goals. The formulation of career development planning for students is mainly a process of continuous communication between the upper and lower levels, the human resources management department should do a good job in guiding career development planning, and students should do a good job in cooperating with career development planning. The first step is to clarify the students' own positioning, that is, the jobs that the students are matched with. The second is to formulate the career development path of young students, according to the students' jobs and competencies, set the appropriate career development goals and use them to carry out targeted training and education. Based on the competency theory of career development planning,

enterprises should set appropriate development goals according to the competency characteristics of students and carry out a series of career development planning. Career planning is divided into competency assessment, job matching and specialized training and education based on the results of competency assessment.

First, competency testing and job matching. The competency test is used to determine whether young students are suitable for a certain position, and then the students' competency coordinates and the enterprise's job competency coordinates are used to determine whether the students are suitable for the position. This requires a good competency model for each position within the enterprise, through which problems in the work of young students can be analyzed and corrected in time. When conducting competency testing enterprises should unify the assessment tools, most of the competency testing tools can test the professional quality, work ability and professionalism of young students. The professional quality test is mainly conducted using questionnaires, and can also be evaluated by the tested person's superiors, subordinates and colleagues of the same level to evaluate their work performance. The work ability test is mainly to test all aspects of the work ability of the students in the enterprise. Occupational quality test includes the test content of both students' work attitude and status.

Secondly, according to the results of competency evaluation to choose the appropriate training and education strategy for young students. Firstly, we judge the work ability of young students according to the completion of their previous work content or tasks, and find the shortcomings of young students. Then use competency theory to find the training methods and paths for young students' deficiencies, and then formulate clear training and education programs and course contents, so as to achieve the effect of improving the comprehensive quality and ability of young students in enterprises.

## V. Conclusion

This paper proposes a big data platform framework for innovation and entrepreneurship education in colleges and universities, combining Apriori algorithm, decision tree algorithm and K-means algorithm to data mine and analyze students' innovation and entrepreneurship data. The results show:

(1) Active-willing students have a more profound and rational understanding of entrepreneurship competitions, their motivation to participate is more active and positive, and they tend to agree with the value recognition of entrepreneurship competitions.

(2) After targeted entrepreneurship education, a total of 16 people finally have a certain selection tendency towards entrepreneurship and have preliminary entrepreneurial thinking. The use of K-means algorithm can truly reflect the situation of each student, grasp the learning needs of different students, provide colorful and targeted educational resources for each student, and provide a good direction for the development of personalized teaching methods for entrepreneurship education.

## References

[1] Ding, Y. Y. (2017). The constraints of innovation and entrepreneurship education for university students. Journal of Interdisciplinary Mathematics, 20(6-7), 1431-1434.

[2] Mallett, O. (2019). Collaboration in entrepreneurship education: challenges, opportunities and innovations. Journal of Small Business & Entrepreneurship, 31(3), 177-182.

[3] Wang, Y., & Ma, Y. (2022). Innovation and entrepreneurship education in Chinese universities: Developments and challenges. Chinese Education & Society, 55(4-5), 225-232.

[4] Luo, Y. (2016, August). Research on career planning of college students. In 2016 International Conference on Economics, Social Science, Arts, Education and Management Engineering (pp. 856-861). Atlantis Press.

[5] Belser, C. T., Prescod, D. J., Daire, A. P., Dagley, M. A., & Young, C. Y. (2018). The influence of career planning on career thoughts in STEM‐interested undergraduates. The Career Development Quarterly, 66(2), 176-181.

[6] Zhang, Z., Yu, X., & Liu, X. (2022). Do I decide my career? Linking career stress, career exploration, and future work self to career planning or indecision. Frontiers in Psychology, 13, 997984.

[7] Peng, X. (2020). Study on the Innovation of College Career Planning Education From the Perspective of “Integration of Innovation & Entrepreneurship and Specialty Education”. Advances in Educational Technology and Psychology, 4(1), 141-148.

[8] Hou, W. (2024). Practical Exploration of Integrating Innovation and Entrepreneurship Education with Employment Guidance in Vocational Colleges. The Educational Review, USA, 8(5).

[9] Wolf, B., Maisch, B., & Sales, A. (2024). Entrepreneurial Life Design: Bridging Innovation and Personal Growth in Entrepreneurship Education. CERN IdeaSquare Journal of Experimental Innovation, 8(3), 14-21.

[10] Jia, Y. (2023, August). University Innovation and Entrepreneurship Education Resource Sharing System Based on Cloud Service Platform. In International Conference on E-Learning, E-Education, and Online Training (pp. 139-155). Cham: Springer Nature Switzerland.

[11] Wei, X. (2023). Research on the Implementation of Innovation and Entrepreneurship Education and Existing Problems. International Journal of New Developments in Education, 5(20).

[12] Jiang, X., & Sun, Y. (2015). Research article study on constructing an education platform for innovation and entrepreneurship of university student. Research Journal of Applied Sciences, Engineering and Technology, 9(10), 824-829.

[13] Yang, Y. (2024). Research on curriculum education of innovation and entrepreneurship education in colleges and universities under the guidance of application-oriented talent training. International Journal of New Developments in Education, 6(1).

[14] Zheng, X. (2024). Construction of an innovative entrepreneurship project learning platform introducing a group recommendation algorithm for college students. Entertainment Computing, 51, 100666.
[15] Wang, J., & Wang, X. (2023, August). A University Innovation and Entrepreneurship Information Sharing Platform Based on Data Mining and Classification Algorithms. In 2023 IEEE 4th Annual Flagship India Council International Subsections Conference (INDISCON) (pp. 1-6). IEEE.
[16] Mahsa Hajihosseinlou,Abbas Maghsoudi & Reza Ghezelbash. (2024). Geochemical Anomaly Detection and Pattern Recognition: A Combined Study of the Apriori Algorithm, Principal Component Analysis, and Spectral Clustering. Minerals,14(12),1202-1202.
[17] Liu Meidan. (2025). A Digital Classification Method of Rural Public Affairs Governance Based on K-means Clustering Algorithm. International Journal of High Speed Electronics and Systems,(prepublish),