

Retraction

Retracted: Developing Countermeasures of Integrating Entrepreneurship Education with Professional Education in Colleges and Universities Using Data Mining

Journal of Electrical and Computer Engineering

Received 19 December 2023; Accepted 19 December 2023; Published 20 December 2023

Copyright © 2023 Journal of Electrical and Computer Engineering. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

- (1) Discrepancies in scope
- (2) Discrepancies in the description of the research reported
- (3) Discrepancies between the availability of data and the research described
- (4) Inappropriate citations
- (5) Incoherent, meaningless and/or irrelevant content included in the article
- (6) Manipulated or compromised peer review

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

In addition, our investigation has also shown that one or more of the following human-subject reporting requirements has not been met in this article: ethical approval by an Institutional Review Board (IRB) committee or equivalent, patient/participant consent to participate, and/or agreement to publish patient/participant details (where relevant).

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

- [1] L. Deng, "Developing Countermeasures of Integrating Entrepreneurship Education with Professional Education in Colleges and Universities Using Data Mining," *Journal of Electrical and Computer Engineering*, vol. 2022, Article ID 1723712, 10 pages, 2022.

Research Article

Developing Countermeasures of Integrating Entrepreneurship Education with Professional Education in Colleges and Universities Using Data Mining

LiLi Deng 

School of Economic Management and Law, Shenyang Institute of Engineering, Shenyang 110136, Liaoning, China

Correspondence should be addressed to LiLi Deng; dengll@sie.edu.cn

Received 20 April 2022; Accepted 31 May 2022; Published 17 November 2022

Academic Editor: Wei Liu

Copyright © 2022 LiLi Deng. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The purpose is to promote the integrated development of innovation and entrepreneurship education (IEE) and professional education (PE) in colleges and universities (CAUs) more accurately and efficiently. Data mining (DM) technology is introduced to study IEE + PE in CAUs. At present, although the state strongly supports the IEE + PE, the integration progress is still slow, and IEE and PE are still completely independent of each other. Therefore, it is urgent to combine IEE with PE organically. Accordingly, based on DM technology, this work analyzes the current situation and existing problems of IEE + PE and puts forward suggestions and corresponding countermeasures for the future development of IEE + PE. The results show the following: (1) College students have high expectations of cultivating their innovation and entrepreneurship ability. (2) The internal and external influencing factors of IEE must be merged. (3) The implementation methods of IEE should be diversified. (4) The current resource allocation of IEE is unreasonable. Based on the above analysis, the existing problems are studied in detail and the corresponding countermeasures are put forward. Meanwhile, DM technology provides a theoretical data basis, making the research results more authentic and reliable and providing reference suggestions for IEE + PE.

1. Introduction

These days, the innovation and entrepreneurship education (IEE) programs have become increasingly prevalent in Chinese colleges and universities (CAUs) to catch up with the new requirements of the times. The fast domestic economic growth and solidifying national strength pose ever-more strict demands on higher education [1]. Unlike traditional education, modern higher education pays more attention to students' diversified and all-around development. IEE programs can help students establish entrepreneurship and innovation awareness in the new era and cultivate more comprehensive and application-oriented innovative talents. It further helps China achieve the goal of becoming a new innovative country [2].

Previous professional education (PE) was not at all innovation-oriented. Thus, to meet China's national conditions and development goals, the integration of PE and IEE must

be stepped up. Such fusion serves as an expansion and supplement of PE. It facilitates students to improve their job competitiveness after graduation. In recent years, the countless proposals and suggestions somehow reflect China's determination and practice of combining IEE and PE [3]. IEE + PE affects college education and talent cultivation, as well as China's successful economic transformation, and it has become a matter of urgency. Abundant research can be found on the IEE + PE, mostly focusing on the text and employing subjective judgments. There is a lack of substantive research. In the current big data era, people's interactions or actions with the outside world are all mediated by data information. Data mining (DM) collects data in daily life through various methods. Then, it processes the collected data, classifies, and stores it in the basic database [4]. Analyzing the basic database can discover the laws of daily life and provide corresponding suggestions for problem-solving. The results are usually objective and have strong credibility.

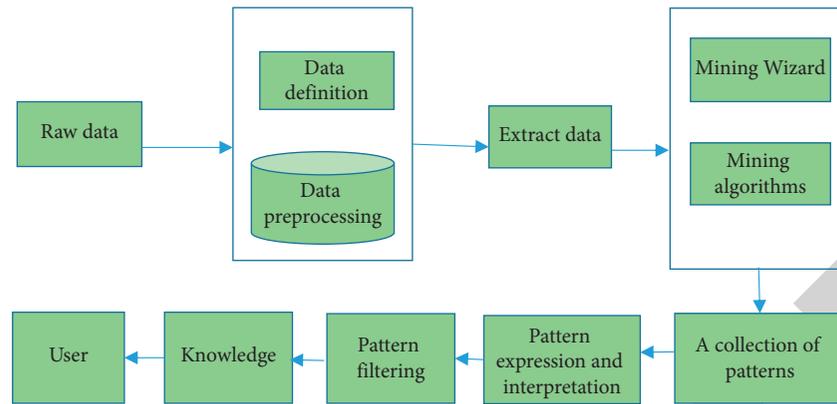


FIGURE 1: Data mining workflow.

Therefore, this work will analyze and research the countermeasures of merging IEE with PE in CAUs from the aspect of DM. The problems in the fusion process are analyzed through a model implementation, and corresponding countermeasures are put forward. The DM technology makes the research process more objective and rational and can better discover the shortcomings. The experimental results are also clear and easy to understand, with solid credibility. The term Knowledge Discovery in Database (KDD), which is very similar to DM, first appeared at the Symposium of the 11th International Joint Conference on Artificial Intelligence (AI) held in Detroit in August 1989. Since 1993, the American computer association has held special meetings every year to study and explore DM technology. Compared with foreign countries, China's research on DM and KDD started late and did not form a unified system. In 1993, the National Science Foundation of China (NSFC) first supported the research project of the Hefei Branch of the Chinese Academy of Sciences in the DM field. At present, the researchers engaged in DM are mainly in CAUs, and some are in research institutes or companies. The research field generally focuses on the learning algorithm, the practical application of DM, and DM theory. Most of the current research projects are funded by the government, such as the NSFC, the 863 programs, and the ninth five-year plan [5].

To sum up, DM technology has certain advantages in analyzing and studying the problems in consolidating college IEE with PE. There are few relevant studies. Firstly, this work analyzes DM technology, regression algorithm, and decision tree (DT) regression algorithm and introduces the relevant basic concepts, DT classification algorithm, and several processing processes of DM technology. Some key contents of the DT algorithm have been introduced in detail, focusing on the classification and regression tree (CART) algorithm. Secondly, the sample dataset is collected by questionnaire survey (QS) and segmented and pruned by the CART algorithm to mine the sample dataset deeper. Then, it analyzes and compares the classified data and the experimental results. Finally, the research content is summarized, pointing out some shortcomings and prospecting future research and improvement direction. The findings provide some reference significance for developing IEE programs in the future.

2. Methods

2.1. Analysis of Data Mining Technology. DM technology is a catalyst in the information age. It filters, searches, and converts valuable information and data from large amounts of data through relevant algorithms. The obtained data and information will be reflected in real-life applications. Data extraction and transformation might involve multi-disciplinary techniques, such as intelligent identification, computer technology, database technology, and statistical analysis. Most commonly, DM assists decision-making to help decision-makers adjust methods and strategies and avoid wrong decisions [6]. The data analysis process is shown in Figure 1.

DM technology obtains the potential internal data laws by analyzing the statistical data. The specific process is unfolded into several steps. Step 1: sort out the data, compare the data with the definition database, select the relevant experimental data, and sort them into the basic dataset. Step 2: use the DM algorithm to mine the basic dataset, analyze the evolution of the internal relevance of the data, and find out the potential internal laws of the data. Step 3: express the law simply and understandably [7].

The commonly used methods for data analysis using DM mainly include classification, regression analysis, clustering, association rules, feature, variation and deviation analysis, Web mining, etc. They mine data from different perspectives.

- (1) Classification. It is to find out the common characteristics of a set of data in the database and divide it into different categories according to the classification model. The purpose is to map the data items in the database to a given category through the classification model.
- (2) This method reflects the temporal characteristics of attribute values in transactional databases, generates a function that maps data items to a real-valued predictor variable, and discovers dependencies between variables or attributes. Its main research problems include the trend characteristics of data series, prediction of data series, and correlation between data.

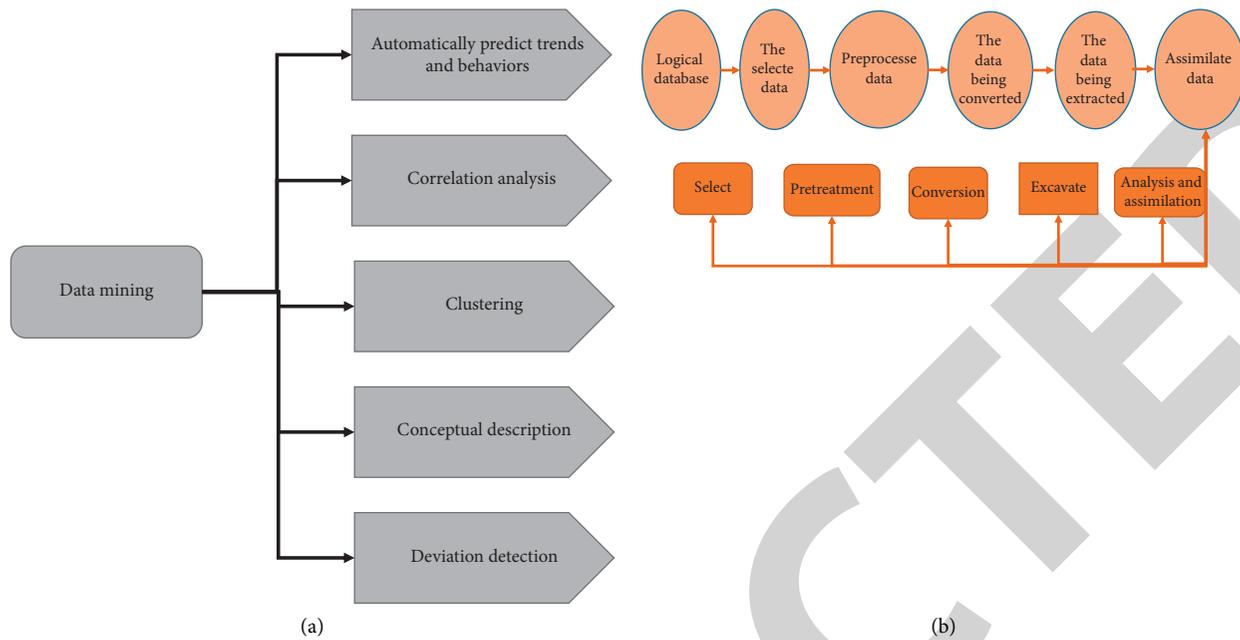


FIGURE 2: DM functions and flowcharts. (a) DM feature diagram. (b) DM flowchart.

- (3) Clustering. It is to divide a set of data into several categories according to their similarity and difference. Its purpose is to make the similarity between data belonging to the same category as large as possible and the similarity between data in different categories as small as possible. It can be applied to the classification of customer groups, analysis of customer background, trend prediction of customer purchase, market segmentation, etc.
- (4) Association rules. It is a rule that describes the relationship between data items in the database. That is, according to the appearance of some items in a transaction, it can be concluded that other items will also appear in the same thing, namely, the association or interrelationship hidden in the data.
- (5) Features. It is to extract the feature expressions about the data from a set of data in the database, and these feature expressions express the overall characteristics of the dataset.
- (6) Variation and deviation analysis. Deviation includes a large class of potentially interesting knowledge, such as anomalous instances in classification, exceptions to patterns, and deviations of observations from expectations. The purpose is to find meaningful differences between observations and reference quantities. In enterprise crisis management and its early warning, managers are more interested in those unexpected rules.
- (7) Web mining. With the rapid development of the Internet and the global popularity of the Web, the amount of information on the Web is extremely rich. Through the mining of the Web, the massive data of the Web can be used to analyze and collect research-

related information, to further concentrate on analyzing and processing external environmental information and internal information that have a significant or potentially significant impact on research. According to the analysis results, various problems and factors that may cause the crisis in the research process are identified.

The goal of DM is to find implicit and meaningful knowledge from the database. The main functions and processes of DM are manifested in Figure 2.

As in Figure 2, DM has five functions: (1) automatically predict trends and behaviors. DM automatically looks for predictive information in large databases and can quickly and directly draw conclusions from the data itself. (2) Correlation analysis. Data association is an important knowledge found in databases. There is a correlation if there is some regularity between two or multiple variables. Correlation can be divided into simple, temporal, and causal correlations. Association analysis can find out the hidden association network in the database. The rules generated by association analysis have credibility. (3) Clustering. The database records can be divided into meaningful subsets, namely, clustering. In turn, clustering enhances people's understanding of objective reality and is a prerequisite for concept description and deviation analysis. Clustering technology mainly includes traditional pattern recognition methods and mathematical taxonomy. (4) Concept description. Concept description describes the connotation of a certain kind of object and summarizes its relevant characteristics. (5) Deviation detection. The basic deviation detection method finds the meaningful difference between the observation results and the reference values.

The process of DM includes the following: (1) determining business objects and clearly defining business

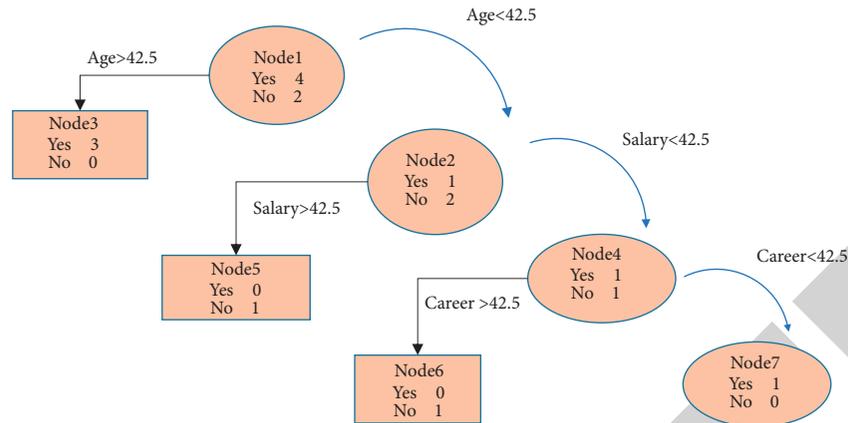


FIGURE 3: DT workflow.

problems. (2) Data preparation. It searches all internal and external data information related to business objects and selects the data suitable for data mining applications. (3) DM. It selects the appropriate mining algorithm to mine the converted data. (4) Result analysis. The analysis methods used to interpret and evaluate the results are generally based on DM operations, and visualization technology is usually used. (5) Assimilation of knowledge. It integrates the knowledge obtained from the analysis into the organizational structure of the business information system.

2.2. Classification and Regression Analysis. Classification and regression technology is most commonly used in DM. The classification technique uses a group of determined variables to form a classification function model. Then, it processes the classified data, denotes each data with a determined variable value, and then forms a new function dataset [8]. Essentially, classification analyzes the data to introduce the prediction classification model. Then, the classification prediction model is used to segment the unknown data into deterministic categories. Regression analysis reflects the characteristics of attribute values of data in the database and finds the dependencies between attribute values by expressing the relationship of data mapping through functions. It can be applied to the study of data series prediction and correlation. Alternatively, regression analysis uses the data changing trend of variables to predict the possible changes of data in the future and judge the correlation between data [9]. The difference between the two is that the distribution state of the predicted value is different. The classification analysis data are discrete distribution, while the regression analysis is a continuous distribution. Classified regression technology includes decision tree (DT), Bayesian network (BN), logistic regression model (LRM), random forest (RF), and genetic algorithm (GA) [10].

2.3. Decision Tree. DT mainly carries out regression prediction according to the basic database by using the inductive algorithm to construct a regression tree. The generated regression tree can analyze the data in other similar cases and

predict the future trend [11]. The workflow of DT is unfolded in Figure 3.

The workflow of DT is mainly divided into two steps. Step 1 establishes the basic training dataset through the classified analysis data and uses the regression algorithm to establish the corresponding DT according to the training dataset. The second step is to test the error of the DT step by step.

A training dataset is a special dataset selected from the basic dataset generated after the operation of the DM algorithm. It is mainly used to help induce, analyze, construct, and generate the corresponding DT. The training dataset is the basis of building DT, so it has special requirements for data selection. That is, no "ambiguity" data are allowed. The data attributes of the training dataset must be highly consistent with the decision attributes. Similarly, the testing dataset is also a special dataset that meets the requirements selected from the basic dataset. Nevertheless, the testing dataset cannot overlap with the training dataset. After the DT is constructed, the testing dataset is imported into the DT system for operation. The operation results are compared with the estimated results, the different results are further analyzed, and the erroneous data and variables are removed to refine the DT [12].

The DT is a tree structure comprising data nodes and corresponding split edges. The specific structure of the DT is detailed in Figure 4.

Furthermore, Table 1 expounds on the basic structure of the DT [13].

(1) Different stages of DT have different operation standards. To illustrate, the data preprocessing stage ensures data quality. Low-quality data will affect the DT operation. The low-quality testing dataset will affect the judgment quality of the DT model, thus affecting the final result. Therefore, ensuring the quality of the dataset is the basic link to building a perfect DT. The following methods are usually used to ensure dataset quality [14]. (2) Data cleaning fills out lost key data during transmission. Manual data completion has very low efficiency. Automatic data completion uses the system default value and the mean value of the data. (3) Data conversion transforms the basic dataset into easily analyzable data to reduce data complexity and speed up algorithm

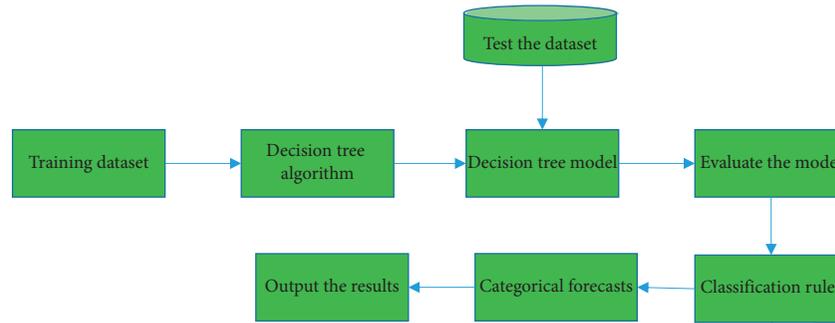


FIGURE 4: Basic structure diagram of DT.

TABLE 1: Basic structure of DT.

Component name	Component function
Node 1	Input point of data, the “root” of DT
“Yes” or “no”	Details of the amount of data inside the node and the number of “yes” + “no” equals the total amount of data inside the node.
Node 3, node 5, and node 6	The leaf node of the DT can predict the future trend of the same situation.
“Age <42.5.”	The split condition can judge and segment the data when the data decision attribute values are different to construct the DT further.

TABLE 2: DT prepruning process.

Method	Concrete content
Tree height inhibiting standard	When the DT reaches a certain height, it inhibits the tree growth. It classifies the DT nodes according to the decision data.
Data threshold inhibiting standard	When the node data amount is less than the threshold, it inhibits the DT growth. It takes the data value with the largest number of decision attributes as the node classification standard.
Decision attribute inhibiting standard	When the proportion of a “decision attribute” is greater than a threshold, it inhibits DT growth. It takes the “decision attribute” as the classification standard of the node.

operation. (4) Dataset. Multiple scattered data with potential correlation are sorted and collected into one category, convenient for comprehensive reference during subsequent data analysis. Meanwhile, several data processing methods are employed to screen and eliminate differential data to ensure dataset quality.

The splitting standard also affects the DT data splitting process. The splitting standard is a key index to distinguish different DT internal regression algorithms. Therefore, selecting the splitting condition standard is essential for DT research. The selection criteria mainly include (1) information gain rate. Information gain refers to the information difference of variable before and after a feature is removed. The information gain rate is based on information gain. It takes the proportion of information gained as the selection standard of splitting standard. (2) Gini split index (Gini index). It is the complexity of the category to which the analysis data belongs. The higher the complexity is, the higher the Gini index is [15].

A newly generated DT must be pruned to adapt to the actual experimental test. At the same time, DT pruning is also a critical factor in determining DT quality. Pruning mainly includes two links. (1) Prepruning inhibits DT generation in advance and prunes the DT. It has a simple

operation, can reduce the DT generation time, and is especially apt for large-scale data. The prepruning algorithm flow is unfolded in Table 2 [16].

Nevertheless, prepruning also has obvious disadvantages. Interrupting the DT growth will cause an imperfect DT, thus losing some attributes and reducing the advantages of DT.

(2) Postpruning technology refers to pruning the data nodes after the DT is completely grown to ensure the DT performance and the integrity of internal data attributes. However, the postpruning method is cumbersome. The commonly used postpruning algorithms include (1) cost complexity pruning (CCP), (2) reduce error pruning (REP), (3) minimum error pruning (MEP), and (4) pessimistic error pruning (PEP) [17].

2.4. CART Algorithm. The classification and regression tree (CART) algorithm, proposed by Breiman et al. in 1984, has composite functions. It can classify and regress data. CART algorithm can generate binary DT through binary division [18]. Moreover, CART uses the Gini index as the attribute selection standard to quickly find the segmentation edge points. The functional expression of the Gini index reads [19]

$$\text{Gini}(S) = 1 - \sum_{i=0}^n p_i^2. \quad (1)$$

In equation (1), S , n , and p_i are the sample, the value of decision attribute, and the relative probability of decision attribute i in sample S . The data sample is segmented and judged by Gini index. The expression of the Gini index reads

$$\text{Gini}_{\text{split}}(S) = \frac{m_1}{m} * \text{Gini}(S_1) + \frac{m_2}{m} * \text{Gini}(S_2). \quad (2)$$

In equation (2), m , m_1 , and m_2 are the data entry in sample S . The smaller the Gini index is, the better the segmentation effect is.

CART algorithm operates differently on different data types. For discrete data, the algorithm will analyze all possible attribute values. It combines the possible values of the data one by one and segments the data in the sample according to the combination results. Then, it calculates the Gini index of the segmented data to verify the segmentation result. In comparison, continuous data are arranged in positive order according to the decision attribute value. The average decision attribute value of the two adjacent data is taken as the segmentation standard σ [20].

CCP algorithm is usually used to prune CART-generated DT. The algorithm will prune step by step from the root of the decision tree. First, the original data T_0 of the DT is split step by step to generate subtrees: T_1, T_2, \dots, T_n . Then, the testing dataset is used to test the generated subtrees, and the optimal DT is selected as the classification standard according to the test results. The splitting node of the sub-DT is the branch with the lowest error gain rate of the previous subdecision tree. The calculation of gain error reads

$$\alpha = \frac{R(t) - R(T_t)}{|N_{T_t}| - 1}. \quad (3)$$

In equation (3), $R(T_t)$ represents the sum of error costs of all leaf nodes of subtree T_t except node t [21]. $|N_{T_t}|$ is the number of leaf nodes in subtree t . $R(t)$ denotes the error cost of no-leaf node t , calculated as follows:

$$R(t) = r(t) * p(t). \quad (4)$$

In equation (4), $r(t)$ is the error rate of node t . $p(t)$ signifies the proportion of subtree T_t data in the total data.

2.5. Data Model Construction. The research data are obtained through questionnaire survey (QS) by randomly investigating students and teachers in an IEE program. Overall, 500 QSs are distributed and 436 recovered, with 418 valid. The distribution details of the initial QS sample dataset are explained in Figure 5.

As in Figure 5, 147 freshmen, 92 sophomores, 154 junior students, and 25 IEE teachers are in this survey. The QS is designed based on the current situation of IEE and PE and their differences. The survey mainly includes three parts. First is the investigators' IEE cognition, including the concept and purpose of IEE. The second part analyzes the influencing factors of IEE from three aspects: individual, school, and

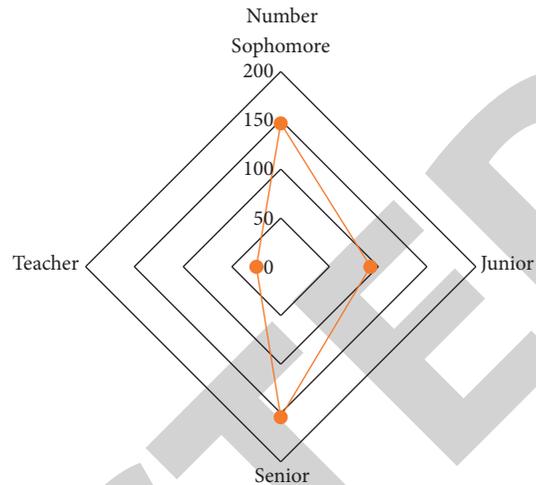


FIGURE 5: Basic information of respondents.

society, Third, it explores the implementation mode, the curriculum teaching mode, and the practice form of IEE [22]. The investigation process of this work is described in Figure 6.

According to Figure 6, the specific process of QS data includes selecting the internal data from the database to form the target dataset, filtering the target dataset to form the test dataset, data conversion and mining of the test dataset, and data coding after outputting the results. Finally, the assimilated knowledge data are outputted:

- (1) Data preparation imports the QS data into the original database as the source of DM.
- (2) Data preprocessing filters the data, deletes or reduces the noise, and supplements the missing decision attributes.
- (3) Data modeling utilizes the CART algorithm. DT splitting parameters are set according to the Gini index of DT segmentation and growth. At the same time, the maximum growth level of DT is set as five layers [23–25].

3. Results

3.1. CART Classification Results. Here, the collected data are inputted into the CART algorithm. The DT splits the sample dataset to generate the corresponding attribute list. The results obtained are listed in Table 3 (due to too much data, only some of them are listed).

The attribute list obtained by differentiating the initial sample data is presented in Tables 4–6.

Next, the classified data are sorted forward according to their respective attributes to generate a new sorted attribute list. Then, according to the attribute list, the data are segmented by the median value of attribute A of every two consecutive sample instances. The Gini segmentation index is calculated. In the instance data, after arranging the age attribute list in positive order, the median is selected so that the Gini segmentation index under the five segmentation values of 19.5, 20, 20.5, 21, and 28 needs to be solved. The calculation results are plotted in Figure 7.

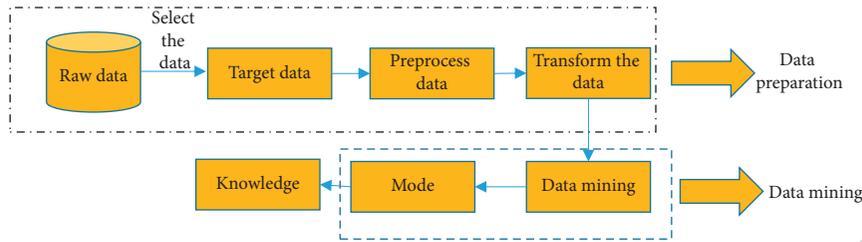


FIGURE 6: Detailed diagram of QS.

TABLE 3: QS initial sample dataset.

ID	Age	Gender	Grade	Do you like IEE?
1	20	Female	Sophomore	No
2	21	Male	Junior	Yes
3	35	Female	Freshman	Yes
4	20	Male	Sophomore	Yes
5	19	Female	Freshman	No
6	21	Male	Junior	Yes

TABLE 4: Age attribute list.

ID	Age	Do you like IEE?
1	20	No
2	21	Yes
3	35	Yes
4	20	Yes
5	19	No
6	21	Yes

TABLE 5: Gender attribute list.

ID	Gender	Do you like IEE?
1	Female	No
2	Male	Yes
3	Female	Yes
4	Male	Yes
5	Female	No
6	Female	Yes

TABLE 6: Grade attribute list.

ID	Grade	Do you like IEE?
1	Sophomore	No
2	Junior	Yes
3	Freshman	Yes
4	Sophomore	Yes
5	Freshman	No
6	Junior	Yes

As in Figure 7, the Gini segmentation index of the five age segmentation values of 19.5, 20, 20.5, 21, and 28 is 0.222, 0.435, 0.138, 0.324, and 0.457, respectively. That is, when the age segmentation value is 20.5, the Gini segmentation index is the smallest. Hence, the best segmentation value of the age attribute for the initial sample is 20.5. Similarly, the QS sample dataset can be calculated and classified to obtain the best segmentation index value to classify further and analyze the data.

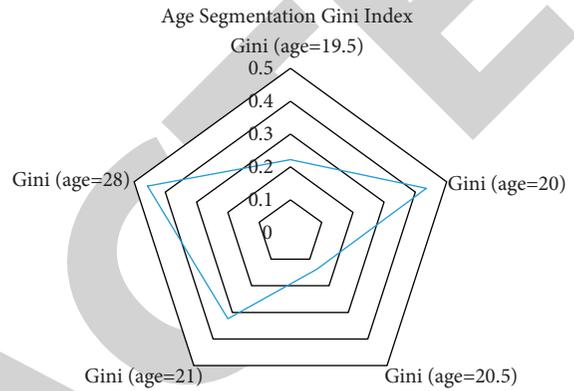


FIGURE 7: Gini segmentation index of age attributes.

3.2. QS Results of Talent Ability Training and Entrepreneurial Motivation. The experimental results analyzed by DT are as follows: college students are the most direct beneficiaries of higher education reform. Therefore, investigating their views on PE reform has important reference significance. This section makes a random survey of college students using the QS method to get the opinions and views of college students on IEE and PE reform. This section designs two questions to investigate college students' cognition of IEE. First question is as follows: what are the five main abilities you think need to be improved in talent cultivation? The QS results are demonstrated in Figure 8.

As in Figure 8, 47.3%, 38.6%, 32.4%, 30.2%, 26.2%, and 25.7% of the respondents believe they need to cultivate innovation, practical, cooperation, career planning, communication, and work ability, respectively. Figure 8 shows that college students pay more attention to innovation and entrepreneurship ability cultivation. In the era of the knowledge economy, college students urgently need to cultivate their own innovation ability in daily teaching. Only by continuous innovation can they excel in the fierce competition.

Second question is as follows: what is your entrepreneurial motivation if you start a business (multiple choices are allowed)? The QS results are revealed in Figure 9.

As in Figure 9, respondents' entrepreneurial motivation to create wealth accounts for the highest: 68.2%. The second most entrepreneurial motivation is to realize self-worth, accounting for 55.4%. The third most is personal interest, accounting for 30.3%. The entrepreneurial motivation for career planning accounts for 30.2%, and the entrepreneurial motivation forced by reality accounts for 20.61%. Lastly,

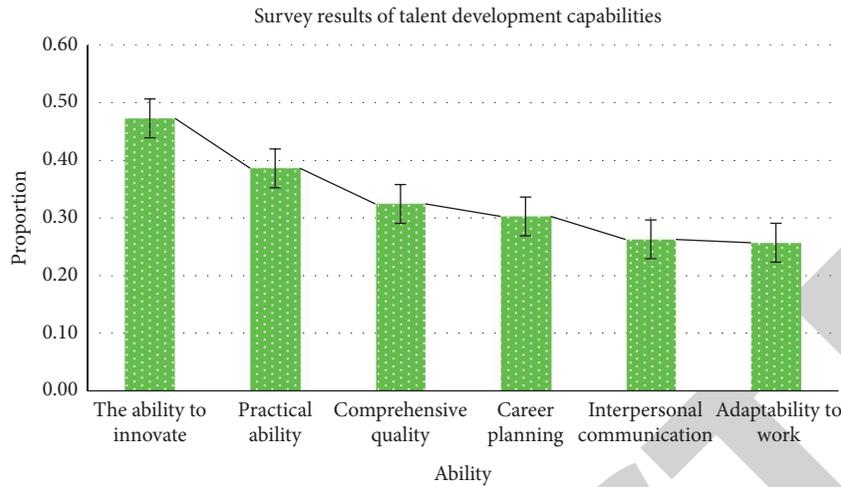


FIGURE 8: QS results of talent cultivation ability.

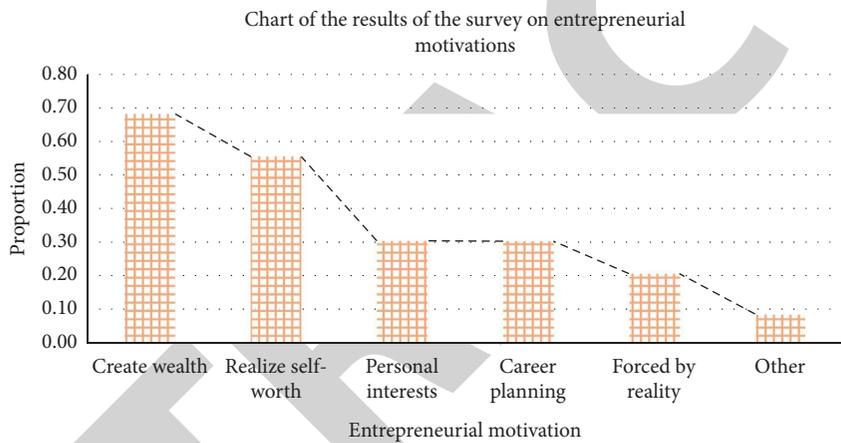


FIGURE 9: QS results of entrepreneurial motivation.

other entrepreneurial motives account for 8.5%. From the QS results, college students' entrepreneurial motivation is diverse. About two-thirds of the respondents choose to create more wealth, showing that in the entrepreneurial motivation of college students, the pursuit of higher economic income is a more critical appeal. More than half of the total people choose to realize their self-worth, which shows that college students choose entrepreneurship with solid subjective initiative. Entrepreneurship is to realize college students' pursuit of life. Additionally, 40% of respondents choose to start a business because of employment pressure, showing that CAUs must carry out IEE to help college students win a better career.

3.3. QS Results of Influencing Factors of Entrepreneurship. Entrepreneurship factors are diverse, and this section selects some indicators for investigation and analysis. The QS results are illustrated in Figure 10.

According to Figure 10, during the growth period of college students, they have received school education and participated less in social practice activities. Their network resources, industry experience, and financing ability are

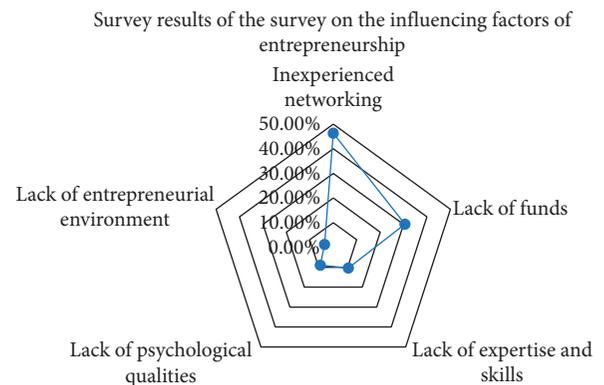


FIGURE 10: QS results of influencing factors of entrepreneurship.

relatively weak. These were the main obstacles in starting a business, where CAUs should strengthen guidance and support. At the same time, creating a good environment for innovation and entrepreneurship is the responsibility of CAU. It requires the joint efforts of the government, social organizations, and college students.

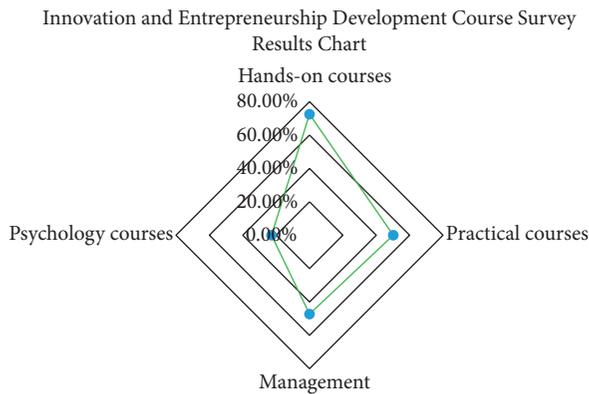


FIGURE 11: Survey results of IEE programs.

3.4. QS Results of Implementation Methods of IEE. So far, the forms of college students' entrepreneurship have also become diverse, among which Internet-based entrepreneurship projects emerge and prevail. The survey results in "what kind of IEE programs do you want the school to offer?" The QS results are illuminated in Figure 11.

Apparently, the demand of college students for IEE shows a diversified trend. When setting up IEE programs, CAUs need to focus on cultivating students' practical ability and innovation and entrepreneurship abilities.

4. Conclusions

In the context of the new era, integrating IEE and PE has attracted considerable attention in China's higher education. The integration process needs the joint efforts of CAUs, the government, and society. Therefore, CAUs should focus on the long-term development of IEE + PE. This work analyzes IEE + PE in CAUs from the DM perspective. Specifically, DM provides a solid and reliable data theoretical basis for this work and improves the objectivity and credibility of the research results. DM technology can classify the initial sample set and calculate the Gini index of each sample attribute classification set to obtain the optimal segmentation point. Thus, DM helps to analyze better and sort out the sample set. At the same time, investigating the evaluation of college students on IEE can judge the results of the current implementation of IEE in CAUs and provides a certain reference basis for the IEE + PE. The innovation of this work is that based on the literature review of IEE + PE, DM technology is added to classify the survey data better and increase the objectivity of the experimental result. The deficiency of this work is that it only investigates the situation of students in one university, the range of data selection is small, and the generalization ability is poor. Whether it can be applied to other schools needs further verification.

Data Availability

The data used to support the findings of this study are available from the author upon request.

Conflicts of Interest

The author declares no conflicts of interest.

References

- [1] A. Colombelli, "Entrepreneurship education: the effects of challenge-based learning on the entrepreneurial mindset of university students," *Administrative Sciences*, vol. 10, no. 2, 2022.
- [2] Q. Wang, Z. Zhe, and Y. Xing, "Application and research of VR technology in art design teaching," *Journal of Physics: Conference Series*, vol. 1345, no. 4, Article ID 042026, 2019.
- [3] L. Yi, H. C. Wang, and W. Hong, "Sustainable development evaluation of innovation and entrepreneurship education of clean energy major in colleges and universities based on SPA-VFS and GRNN optimized by chaos bat algorithm," *Sustainability*, vol. 1744, no. 13, 5960 pages, 2021.
- [4] D. Yan, "Construction of innovative talent ecosystem under the background of enterprise digital transformation," *Lecture Notes in Electrical Engineering*, vol. 791, no. 1, pp. 803–809, 2022.
- [5] D. Alican and D. Birant, "Machine learning and data mining in manufacturing," *Expert Systems with Applications*, vol. 166, no. 2, Article ID 114060, 2021.
- [6] T. Harshil, V. Shah, H. Yagnik, and M. Shah, "Comparative anatomization of data mining and fuzzy logic techniques used in diabetes prognosis," *Clinical eHealth*, vol. 4, pp. 12–23, 2021.
- [7] K. Kannan and A. Menaga, "Risk factor prediction by naive bayes classifier, logistic regression models, various classification and regression machine learning techniques," *Proceedings of the National Academy of Sciences, India - Section B: Biological Sciences*, vol. 166, no. 14, pp. 1–17, 2021.
- [8] E. Kannatey-Asibu and F. R. Tsai, "Modeling of conduction mode laser welding process for feedback control," *Journal of Manufacturing Science and Engineering*, vol. 122, no. 3, pp. 420–428, 2016.
- [9] S. Belagoune, N. Bali, A. Bakdi, B. Baadji, and K. Atif, "Deep learning through LSTM classification and regression for transmission line fault detection, diagnosis and location in large-scale multi-machine power systems," *Measurement*, vol. 177, no. 6, Article ID 109330, 2021.
- [10] W. Chmiel, J. Dańda, A. Dziech et al., "INSIGMA: an intelligent transportation system for urban mobility enhancement," *Multimedia Tools and Applications*, vol. 75, no. 17, pp. 10529–10560, 2021.
- [11] C. Bahzad and A. Abdulazeez, "Classification based on decision tree algorithm for machine learning," *Journal of Applied Science and Technology Trends*, vol. 41, no. 1, p. 2, 2019.
- [12] L. Durjoy, "Determinants of aphasia recovery: exploratory decision tree analysis," *Language, Cognition and Neuroscience*, vol. 9, pp. 62043–62054, 2021.
- [13] M. Wei and X. Fu, "A method of enterprise financial risk analysis and early warning based on decision tree model," *Security and Communication Networks*, vol. 79, pp. 1–8, 2019.
- [14] Y. H. Miao, J. Wang, B. Zhang, and H. Li, "Practical framework of Gini index in the application of machinery fault feature extraction," *Mechanical Systems and Signal Processing*, vol. 165, no. 1-2, Article ID 108333, 2022.
- [15] F. M. Javed, S. Chakraborty, M. M. Billah, P. Das, J. N. Muna, and R. Ranjan, "A comprehensive study on pre-pruning and post-pruning methods of decision tree classification algorithm," in *Proceedings of the 2021 5th International Conference*

- on *Trends in Electronics and Informatics (ICOEI)*, vol. 168, no. 8, Article ID 10432, Tirunelveli, India, June 2021.
- [16] G. Amr and S. Mohamed, "Construction facilities location selection for urban linear infrastructure maintenance projects using uniform cost search method," *Soft Computing*, vol. 26, no. 3, pp. 1403–1415, 2022.
- [17] N. Dong, "Cervical cell classification based on the CART feature selection algorithm," *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, no. 2, pp. 837–1849, 2021.
- [18] J.-y. Zhang, X.-f. Liu, R.-q. Zhang, and H. Wang, "Gini correlation for feature screening," *Acta Mathematicae Applicatae Sinica, English Series*, vol. 37, no. 3, pp. 590–601, 2021.
- [19] S. Singh and P. Gupta, "Comparative study ID3, cart and C4. 5 decision tree algorithm: a survey," *International Journal of Advanced Information Science and Technology*, vol. 1, no. 3, pp. 271–289, 2019.
- [20] L. F. Zhou, "Automatic segmentation of liver from CT scans with CCP-TSPM algorithm," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 1, no. 3, pp. 4–7, 2021.
- [21] J. M. Saskia, "Incorporating student-centred learning in innovation and entrepreneurship education," *European Journal of Education*, vol. 43, no. 4, pp. 513–526, 2018.
- [22] H. B. Zhu, K. Zhang, and U. S. Ogbodo, "Review on innovation and entrepreneurship education in Chinese universities during 2010-2015," *Eurasia Journal of Mathematics, Science and Technology Education*, vol. 13, no. 8, pp. 4325–4335, 2021.
- [23] A. Debra, "Innovation in entrepreneurship education," *Journal of Entrepreneurship Education*, vol. 9, no. 1, 29 pages, 2016.
- [24] L. Pittaway, "Entrepreneurship education: a systematic review of the evidence," *International Small Business Journal*, vol. 25, no. 5, pp. 479–510, 2020.
- [25] A. Hassan, "Relationship between attitudes and intentions for business start-up: a case of Omani university & college students," *Academy of Entrepreneurship Journal*, vol. 24, no. 2, pp. 1–10, 2018.