Research on quality evaluation of innovation and entrepreneurship education for college students based on random forest algorithm and logistic regression model

Qianqian Lu[1], Yongxiang Chai[2], Lihui Ren[3], Pengyu Ren[3], Junhui Zhou[3] and Chunlei Lin[4]

[1] Administrative Office, Zhejiang Guangsha Vocational and Technical University of Construction, Dongyang, China
[2] School of Information, Zhejiang Guangsha Vocational and Technical University of Construction, Dongyang, China
[3] Department of Public Physical Education, Zhejiang Guangsha Vocational and Technical University of Construction, Dongyang, China
[4] Department of Education and Engineering, Zhejiang Guangsha Vocational and Technical University of Construction, Dongyang, China

ABSTRACT

The quality evaluation of innovation and entrepreneurship (I&E) in the education sector is achieving worldwide attention as empowering nations with high quality talents is quintessential for economic progress. China, a pioneer in the world market in almost all sectors have transformed its educational policies and incorporated entrepreneurial skills as a part of their education models to further catalyst the country's economic progress. This research focuses on building a novel hybrid Machine Learning (ML) model by integrating two powerful algorithms namely Random Forest (RF) and Logistic Regression (LR) to assess the intensity of the I&E in education from the data acquired from 25 leading Higher Educational Institution's (HEI) in different provinces. The major contributions to the work are, (1) construction of quality index for each topic of interest using individual RF, (2) ranking the indicators based on the quality index to assess the strength and weaknesses, (3) and finally use the LR algorithm study the quality of each indicator. The efficacy of the proposed hybrid model is validated using the benchmark classification metrics to assess its learning and prediction performance in evaluating the quality of I&E education. The result of the research portrays that the universities have now started to integrate entrepreneurship skills as a part of the curriculum, which is evident from the better ranking of the topic curriculum development which is followed by the enrichment of skills. This comprehensive research will help the institutions to identify the potential areas of growth to boost the economic development and improve the skill set necessary for I&E education among college students.

INTRODUCTION

Innovation and entrepreneurship (I&E) is an important driver to accelerate the economic growth, population integration, social mobility, cultural formation and financial immunity of any country (*Wang, 2022*). They form the key elements of digital age. More precisely, the culture of I&E supports the country to combat the economic instability and shock caused due to unprecedented conditions like the pandemic by transforming the genre of education through innovations which greatly relieves the employment pressure. China has emerged as a forerunner in cultivating and incubating the entrepreneurship qualities of the students by reforming the country's educational policy. In spite of the efforts took by the state and Higher Educational Institutions (HEIs), the development of I&E is still in its stage of infancy. According to the Chinese College Students Entrepreneurship Report published in 2021, nearly 96.1% of the students from HEIs have good entrepreneurial ideas, out of which only 14% of the students are able to realise it. Globally, China has reached second position in I&E during 2021 from being eleventh in the year 2002 (*Huajie et al., 2022*). The increasing rise of I&E education in China is aligned towards building innovation-oriented country by promoting high-quality education which imbibes holistic knowledge about economics, science and technology, employ ability skills, analysis of job markets, and formalizing the employment structures. The Ministry of Education has issued certain significant key points in April 2012 such as:

- I&E should be implemented in HEIs through personnel training
- Channelizing fundamental requirements on I&E in HEIs
- Developing world-class curriculum with a vision of I&E
- Training teachers
- Extending hands to students with good I&E skills.

The various stakeholders which are a dynamic attribute are listed in Table 1. The list is not exhaustive but throws light on few examples where the interest of stake holders deserves special mention.

Invigorating the I&E education will be the promising direction for educational reforms to envisage a holistic economic growth. Integrating I&E as a part of professional education has emanated as a new challenge to developing the curriculum. New policies, views and insights are quintessential to implement vocational knowledge; improve the quality of technical and digital talents to integrate innovation education and entrepreneurship education into professional education. Almost all the HEIs in China have initiated to promote I&E as a part of the education to eventually improve the quality of entrepreneurial abilities among students. Nevertheless, hurdles such as proper design of curriculum, training the mentors/teachers and provision resources are to be addressed by the HEI's due to the dynamically evolving entrepreneurship nature.

Promoting I&E in HEIs is a culmination of series of micro and macro managed activities that involves right amalgamation of various stakeholders as given in Fig. 1. Proper integration of the factors to develop I&E will eventually help the country to attain an

Table 1 Various stakeholders and their scope for business.

| Stakeholder	Creates values for	Service	Scope
Well established business	Customers, client, employees and investors	Commercial products or services	Financial
Business entrepreneur	Customers, client, employees and investors	Novel Commercial products or services	Financial
Social entrepreneur	Society	Novel social services and products	Financial and social
State	Citizens	Welfare services	Cultural, financial and social
Student	Future employees	Adopting work life	Cultural, financial and social
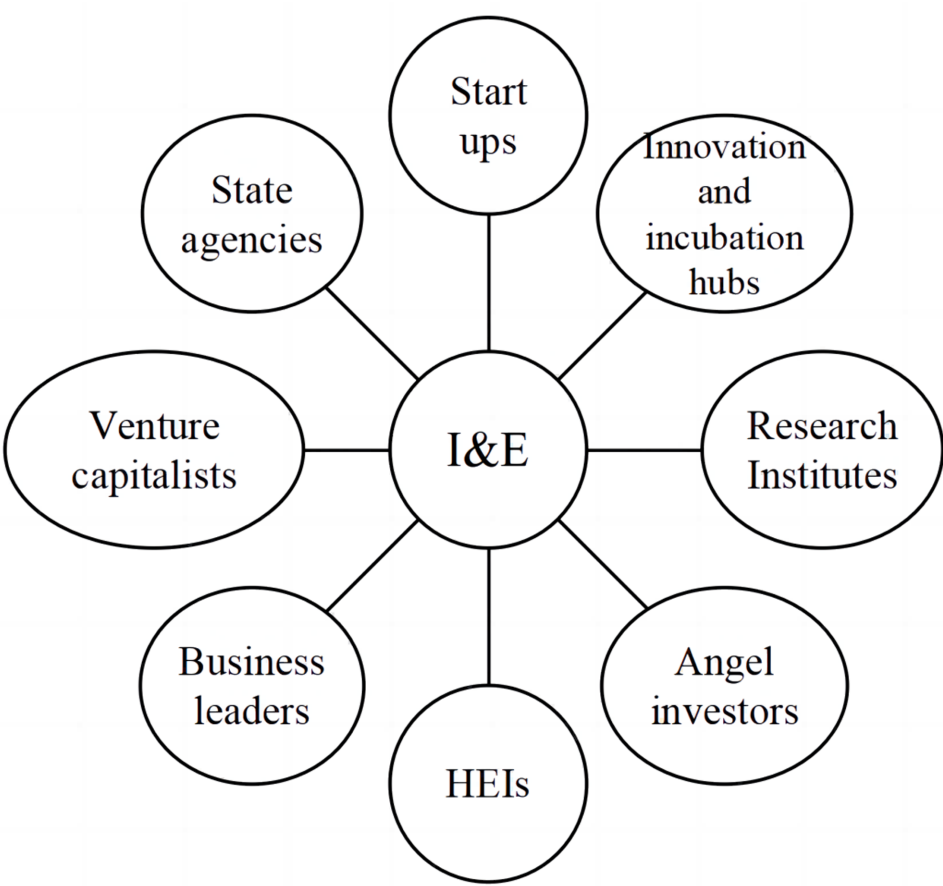
Figure 1 Stakeholders involved in I&E.

Full-size ☒ DOI: 10.7717/peerjcs.1329/fig-1

excellency and become a leading pioneer in global entrepreneurship opportunities that can be realised through the activities mentioned in Fig. 2.

Quantifying and measuring the intensity of I&E in education is the greatest challenge faced by almost all HEIs (*Wang & Liu, 2022*). The Entrepreneurial Intention (EI) within the students is one of the best predictors to quantize the entrepreneurial behaviour (*Lihua, 2022*). In addition to this, the degree of correlation between the factors such as
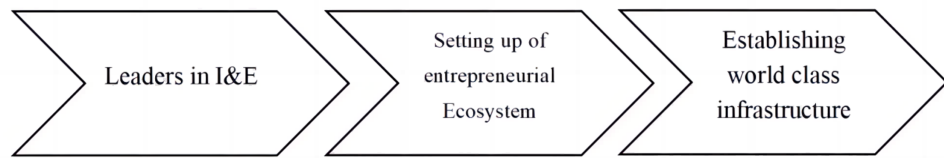
Figure 2 Milestones in achieving entrepreneurial excellence.

Full-size ⬚ DOI: 10.7717/peerjcs.1329/fig-2

EI, feasibility and desirability must be high to incubate the entrepreneurial ecosystem in HEIs. In other words, EI imparts better entrepreneurship thus enabling or enriching the entrepreneurial activities (*Liñán & Fayolle, 2015*). Most importantly, entrepreneurship education (EE) is a tool to nurture entrepreneurship as it demands lifelong learning mandates (*Boldureanu et al., 2020*). China, has well-understood the significance of I&E and hence it integrates the EE into classrooms either as core or elective course for college students, helping college students better plan their time and future career development (*Ratten & Usmanij, 2020*). The ultimate objectives of all these innovative endeavors are to motivate the young students to start their own business ventures. The entrepreneurial teaching must incorporate business plans, develop entrepreneurship courses, and other related entrepreneurial talents training. Figure 3 (*Karen, 2008*) shows the factors that great influence the development of comprehensive entrepreneurial training in HEIs. A major transformation can happen by proper entrepreneurial training by shifting from the teacher central training to student central cultivation.

Challenges confronted by I&E in Chinese HEIs

The evolution of I&E in China shows a very promising development. However, there are few points to be addressed to integrate I&E holistically as a part of the regular curriculum. Few salient points are mentioned here:

- Very limited number of courses and resources: US, a potential rival of China in the global market is offering much more entrepreneurial courses with excellent infrastructure (*Yulian et al., 2014*). This limits the opportunities of native students to develop their entrepreneurial talents.
- Lack of trained skillset: As it is evident that I&E is a comprehensive discipline that demands skilled manpower to train the students with all necessary traits. Unfortunately, the present trend indicates that there is scarcity of well-qualified teachers in this genre (*Zhou, Li & Shahzad, 2021*).
- Limited specializations: Some of the courses that are offered by pioneering universities in China centers on entrepreneurial management, business communication and practices, business finances. team and development training, entrepreneurial market analysis, corporate and contract law etc. These numbers are far lows than other developed countries.
- Single course structures: The conventional education coins around the traditional closed teaching model, which does not provision the inclusion of any new reforms. Integration
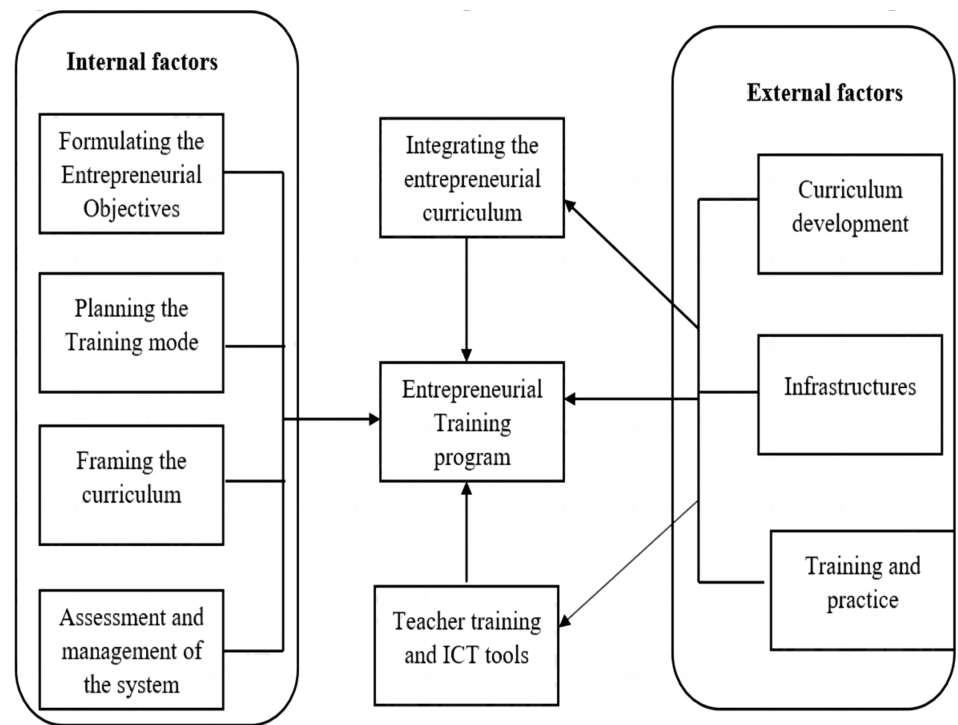
Figure 3 Internal and external factors involved in promotion I&E in education.

of any new measures like embedding I&E through simulation, training or any practical experience is possible only in very few ventures (*Guo, 2022*).

- Lack of formal education model: In reality there is no formalized model that could guide the entrepreneurial teaching in the HEI's. A wide gap can be sensed in integrating all the essential elements of teaching like curriculum development, teaching pedagogy, training and guidance schemes. Also, the implementation of the entrepreneurial education should always align to educational goals, program specific outcomes, and the assessments. EE in most of the Chinese universities are not adhering to these demands (*He & Dong, 2022*).

- Lack of long term goals and support mechanisms: Lack of long run goals in HEI's and rushing behind immediate results ultimately leads to unsuccessful EE with no funding mechanism, hatch and policy safeguard which takes a great toll in the development of EE (*Chen, 2022a*; *Chen, 2022*).

- Insufficient funds: At present, funds in Chinese HEIs are utilized mainly to meet the need of entrepreneurship that are sourced by the government, private entrepreneurs, and very rarely from the funds allotted by schools (*Han et al., 2018*). These funds are highly insufficient.

- Lack of proper incubator space: Incubator space is the ground in which the I&E can be practised by the students. It can include laboratory space or office area where the student can practice business activities, perceive the real market with state of art technological innovation (*Gong et al., 2020*). Insufficient support system in China, has made these unrealistic due to various limitations like conditions, funds and lack of teachers etc.

The above-described issues are in the front line of the I&E in China. These challenges hinder the development of formalized curriculum as each of them are very versatile. Also, the development in I&E is happening in a very slow pace, as the issues can be combated in long run.In spite of all these bottlenecks the Chinese universities are gearing up towards developing I&E. Many researchers, scholars and experts are committed to model the innovative entrepreneurship traits in students with a more formalised research and analysis procedures, to evaluate the quality of I&E in HEI's of China. A well-defined perfect I&E education quality assessment model will eventually help the community to cater the immediate needs of quality I&E.

Educational data mining

Modern developments in education have inspired the growth of Educational Data Mining (EDM). Multitude of research has enforced and identified new ventures and avenues for imparting technologically enhanced learning models and systems based on the societal and students' needs. I&E, which is a new sector of education has leveraged the EDM to model more precise assessment systems. EDM is perceived as a holistic and comprehensive model that integrates valuable insights from multiple domains like Computer Science, Data Mining, E-Learning, data analytics, Machine Learning (ML) and statistics. The relationship among these domains is represented in Fig. 4 (*Gordon & Helen, 2022*). Among these, ML is a field that has witnessed rampant advancements which is further accelerated by learning analytics (*Trakunphutthirak & Lee Vincent, 2022*).

ML, a subset of Artificial Intelligence (AI) is the method of imparting power to the model or machine to access the data and learn the inherent patterns and trends by itself (*Edwards Audrene, Kaplan & Jie, 2020*). The field of ML progressed by programming the computers to optimize or fine tune a performance criterion through sample or example data (*Alexandros & Alexandra, 2018*). This model is generally perceived as the black box where the data enters the model in its beginning, and output is obtained out at another end. However, the processes that happen in the black box is quite complex. Few potential business implementations of ML in the field of education are:

- Predict student performance (*Rastrollo-Guerrero, Gómez-Pulido & Durán-Domínguez, 2020*), where a ML model would be deployed in predicting the performance of a student by learning the qualities of individuals.
- Assess students and grade them using ML model which creates automated adaptive assessments (*Liu & Ardakani, 2022*). This model will provide continuous feedback to the teachers about the student's learning ability and their progress towards learning goals.
- ML models are used to predict the retention rates, which helps the institution to adapt strategies to help students to continue their education (*Samer et al., 2022*).
- Automatic evaluation by deploying handwriting recognition systems (*Vij, Tayal & Jain, 2020*).
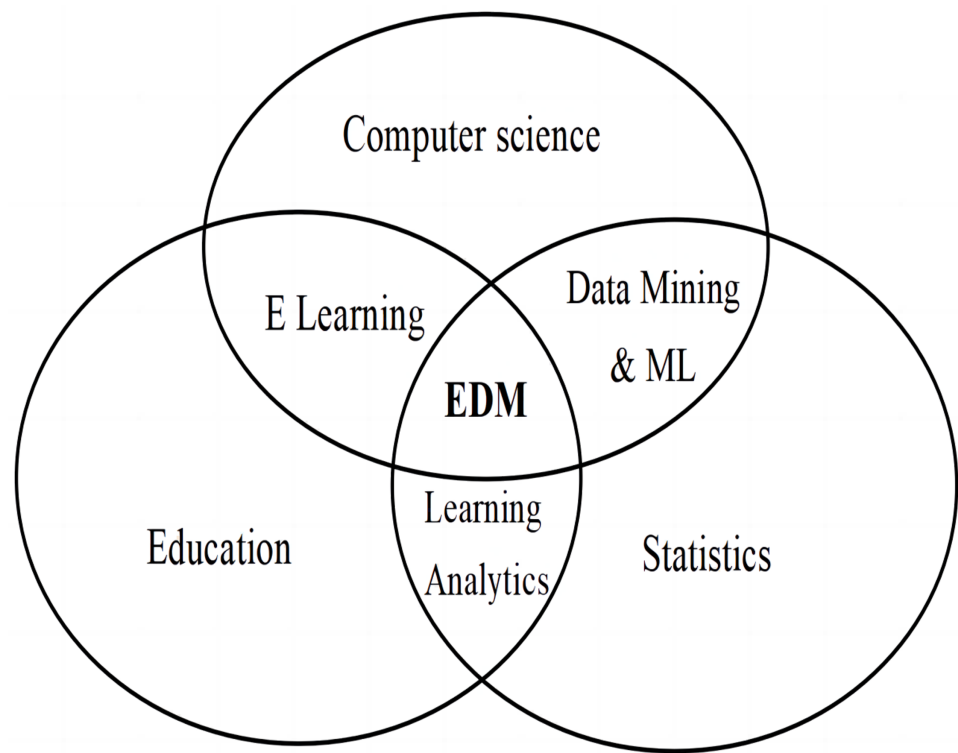- Quality assessment about the course from the formal and informal feedback and student qualities (*Li, 2022*).

Figure 4 Educational data mining.

Literature in the field witnesses the deployment of all types of ML algorithms namely supervised unsupervised and semi supervised algorithms. Mathematically, the supervised algorithms predict the output by learning the history of previous input and output. The model's mapping will be $Y = f(x)$ where Y is the output to be predicted based on the input variable/ variables (x). On the other hand, unsupervised learning algorithms tends to explore the relationships between the inputs and predict the output. The training data will not contain output component. The semi supervised learning is an amalgamation of the former methods, where only few data will contain outputs. A wide variety of ML algorithms like back propagation network, Linear regression, Logistic Regression (LR), Support Vector Machine (SVM), Decision trees, etc.

This work focuses on applying ML algorithms to analyse the quality of I&E education among the College Students. The biggest challenge faced by the educationalists and entrepreneurs is the lack of formal and tractable assessment method that learns the qualitative data to assess the quality of I&E education. The proposed research works attempts to analyse and explore the survey results, which is procured from nearly 25 HEI's in China. The questionnaire majorly focused on assessing the quality on five significant domains namely Curriculum Design, Training and skill set enrichment, Resources, Talent acquisition by teachers and setting Entrepreneurial environment. The responses were scaled and fitted individually into RF algorithm is to construct a novel Quality Index (QI) that characterizes the eminence of the institutions. This QI is learnt by the RF algorithm

by number of individual Classification and Regression based DTs with pruning. Ensemble of the results of the DTs will output the QI, a regressive value. This QI is then fitted into an LR model that finally evaluates the Total Quality Evaluation (TQE) based on the results of applying sigmoid function on the linear regression expression. The major research contributions of the proposed work are:

- A comprehensive analysis of the factors affecting the EE and the hurdles faced by the HEIs in incorporating EE as a part of the curriculum.
- Formulation of the assessment criteria with its relevant indicators
- Development and implementation of Random Forest (RF) and LR model for evaluating the quality of the I&E education
- Validating the models based on standard performance measures

The organisation of the work is as follows: Section 2 gives a brief outline of the important related works. Section 3 presents the information about data sourcing and research methodology. Section 4 portrays the results and the important implications. Conclusions and future research directions are given in Section 5.

RELATED WORKS

Literature witnesses the presence of notable works in the field of assessing the performance of students, curriculum and other entrepreneurial elements offered by the HEIs. Some of the prominent works are briefed in this section.

Huang et al. (2019) developed a BackPropagation (BP) model with optimization to assess quality of I&E using an evaluation index system. The data was collected from major engineering colleges and HEIs to find that the innovative methods must be implemented to foster the I&E in HEI's. Another notable work using back propagation network on quality assessment is done by *Wang et al. (2022a)* and *Wang et al. (2022b)*. This empirical study is primarily focused on qualitative and quantitative assessment made on local University students. A fuzzy based fault tree analysis approach is deployed by *Wang et al. (2022a)* and *Wang et al. (2022b)* for assessing the reliability of I&E in classrooms. The efficacy of the model is validated using a comprehensive questionnaire. A holistic platform and core business processes for assessing the quality of I&E in HEI's is proposed by *Chen (2022a)* and *Chen (2022)*. The proposed system integrated many modules like skill diagnosis, project management skills, online teaching-learning, team building resource availability, testing and assessment evaluation, information and resource sharing etc. A BP model is used to verify the findings of the research.

A novel ML based system is proposed by *Jiang & Li (2021)* that considers the curriculum, teaching and training modes, reforms in practice and teaching in I&E. An intelligent strategy to assess the performance is designed using SVM, Extreme Learning Machines (ELM) and Radia Basis Function (RBF) that uses the facial expressions of the students. An improved collaborative filtering model that uses behavior graph and route of students to assess the performance of EE (*Geng, 2022*). This path is and the multidimensional behavior path is analysed for extracting the similarities between each dimension. A deep learning model that relies on human comprehensive development to assess the innovative quality in EE is

Lu et al. (2023), *PeerJ Comput. Sci.*, DOI 10.7717/peerj-cs.1329

8/22

designed by *Zhang (2020)*. This model understands the learning status of students to assess the quality of I&E. A deep learning approach that is rooted on the understanding, creation and application of knowledge in I&E was proposed by *Jing (2022)*. This model analyses the performance by monitoring the training, critical thinking, knowledge transfer, resource utilization, availability, flexible learning methods etc as the performance indicators. A deep learning model based on the empirical and theoretical research foe assessing the quality of I&E was proposed by *Li & Zhao (2021)*. The learning algorithm models the experimental data to predict the quality.

Principle Component Analysis (PCA) and Linear Discriminant Analysis (LDA) was employed by *Chen (2022a)* and *Chen (2022)* to check the effectiveness of the I&E in modern education system. Entrepreneurship orientation among the individuals is studies and analysed using the predictive analytics approach using ridge, lasso, SVM, BP, and RF (*Sabahi & Parast, 2020*).

Graham & Bonner (2022) designed a decision tree approach on a large-scale dataset to monitor the I&E. The result indicates that entrepreneurial self-efficacy, networks and cultural perceptions are more dominant factors that determine the quality of education. The work by *Qing & Zejun (2021)* built ML based entrepreneurial policy analysis model to study the effect of entrepreneurial policies. The projection pursuit is optimized by applying genetic algorithm and results are validated using Entrepreneurial Activity index. RF is deployed to assess the level of entrepreneurial knowledge possessed by various sectors of students (*Setiawan, 2021*). Integrating the autonomy, critical thinking and automation in I&E using intelligent machines is described by (*Knox, 2019*).

The brief literature in the field of deploying I&E education among students in HEIs using ML, Deep Learning and other Artificial Intelligent techniques shows that there is still room for improvement by including more features that necessarily depicts the quality of the education mode.

DATA AND METHODS

Data collection

The data was collected with the informed consent of the participants. Participants choose "Yes" or "No" in the questionnaire to agree or reject the questionnaire. Procedures associated with data collection were approved by the Zhejiang Guangsha Vocational and Technical University of Construction of Technology Committee.

The data for this study is procured from 25 colleges in China through questionnaires that will best describe the opinion of students about the I&E culture promoted in their institutions. Data collection was done randomly irrespective of gender in all fields of study like literature, language, Science and humanities, economics, law, management, arts, education, agriculture, Engineering, medicine etc. The students were asked to express their views in five scales as given in Table 2. The comprehensive questionnaire contains questions that assess the quality of I&E in various dimensions such as curriculum design, building entrepreneurial skills, resource availability, talent acquisition by teachers and entrepreneurial environment as given in Table 3.

Table 2 Five-point scale of the survey questionnaire.

S.No	Explanation	Scale
1	Strongly Disagree	1
2	Disagree	2
3	Neutral	3
4	Agree	4
5	Strongly Agree	5

Table 3 Cronbach reliability analysis and explanatory indicators for the study.

Topic	Subtopic	CITC	α coefficient of deleted items	Cronbach α coefficient	Variable
Curriculum Design	The resources, contents and materials of I&E are complete	0.707	0.904		X1
	The teaching method is unanimous	0.577	0.908		X2
	The training and practices of the I&E method is good.	0.518	0.910		X3
	The contents of the curriculum have good relevance to professional courses	0.585	0.908		X4
	The curriculum is globally competitive	0.575	0.908		X5
Training and skill set enrichment	Proper and relevant training is incorporated as a professional activity	0.623	0.907		X6
	Training is given to students by well qualified trainers	0.555	0.909		X7
	Sufficient resources like smart rooms, work spaces, project development platforms, activity rooms etc are provided	0.628	0.906		X8
Resources	Usage of Information and Communication Technology (ICT) tools	0.474	0.911	0.913	X9
	Financial policy and support to I&E activities	0.592	0.908		X10
	Well qualified teachers	0.684	0.905		X11
Talent acquisition by teachers	Development of I&E culture in teachers	0.599	0.907		X12
	Enough number of teachers	0.659	0.906		X13
	Availability of teachers in all disciplines	0.738	0.903		X14
	Experience and research capability of teachers	0.625	0.907		X15
Entrepreneurial environment	Constitution of entrepreneurial clubs	0.473	0.912		X16
	Conduction of seminars, guest lectures, symposiums, contests, workshops, conferences and outreach activities.	0.421	0.913		X17

As can be seen from the Table 3, the reliability coefficient value is 0.913, greater than 0.9, indicating that the reliability quality of the research data is high. For the "α coefficient of deleted items", the reliability coefficient does not increase significantly after any item is deleted, indicating that the item should not be deleted. For "CITC value", the CITC value of analysis items is all greater than 0.4, indicating that there is a good correlation between analysis items and a good level of reliability. In summary, the reliability coefficient of the research data is higher than 0.9, which indicates that the data reliability is of high quality and can be used for further analysis (*Eisinga, Grotenhuis & Pelzer, 2013*).

Methods

The research deploys two important ML algorithms namely Random Forest (RF) and Logistic Regression (LR) for assessing the effectiveness of I&E in colleges using the data collected from the questionnaires under the five topics of interest.

Overview of the random forest algorithm

RF is a supervised algorithm that absorbs training samples and then input (X) into each sample to get the one output result (Y), through a mapping (f: X $\longrightarrow$ Y). Once the mapping is established by the model, predictions (Y') for any unknown sample (X') can be obtained. RF can be used for both classification and regression problems. The fundamental unit of RF is the Decision Tree (DT) and it works by integrating bagging with a random subspace. This combination helps the RF to improve its prediction accuracy.

The RF classifier integrates the decision of individual DTs $d_1(x)$, $d_2(x)$, …,$d_n(x)$ till sufficient accuracy is reached. This process is shown in Fig. 5 (*Onesmus, 2020*). The results of the individual trees are combined through majority voting technique to give the result (*Fawagreh, Gaber & Elyan, 2014*). Two important phases in modelling RF are:

- Randomized training data: Bootstrapping method is employed in RF where the data samples are randomly picked by each DT.
- Randomized features: At each partitioning point of the individual DTs, the RF algorithm selects subset of features in random fashion without any replacement. K number of features will be selected, and then various indicators of the data will change under the K splits so as to choose the best split each time. This process will be continued till desired accuracy is reached.

The partitioning at each tree depends on two main metrics namely the entropy and information gain. The entropy of the sampled data T is found using Eq. (1).

$$\text{Entropy}(T) = -\sum_{i=1}^{n} P_i \log 2\, P_i. \tag{1}$$

The term P_i signifies the quantity of data samples belonging to a particular category i that contributes to the total samples. For every feature (A) the data will be sampled into k splits. The information entropy and gain will proceed according to Eqs. (2) and (3).

$$\text{Information Entropy} = -\sum_{j=1}^{k} \frac{|T_j|}{|T|} \text{Entropy}(T) \tag{2}$$

$$\text{Information gain} = \text{Entropy}(T) - \text{Information Entropy}. \tag{3}$$

The variance or Mean Square Error is a measure that quantifies the best fit of the RF tree. It estimates the deviation of the estimated value from its actual value (*Meyners & Hasted, 2022*) and its expression is given in Eq. (4).

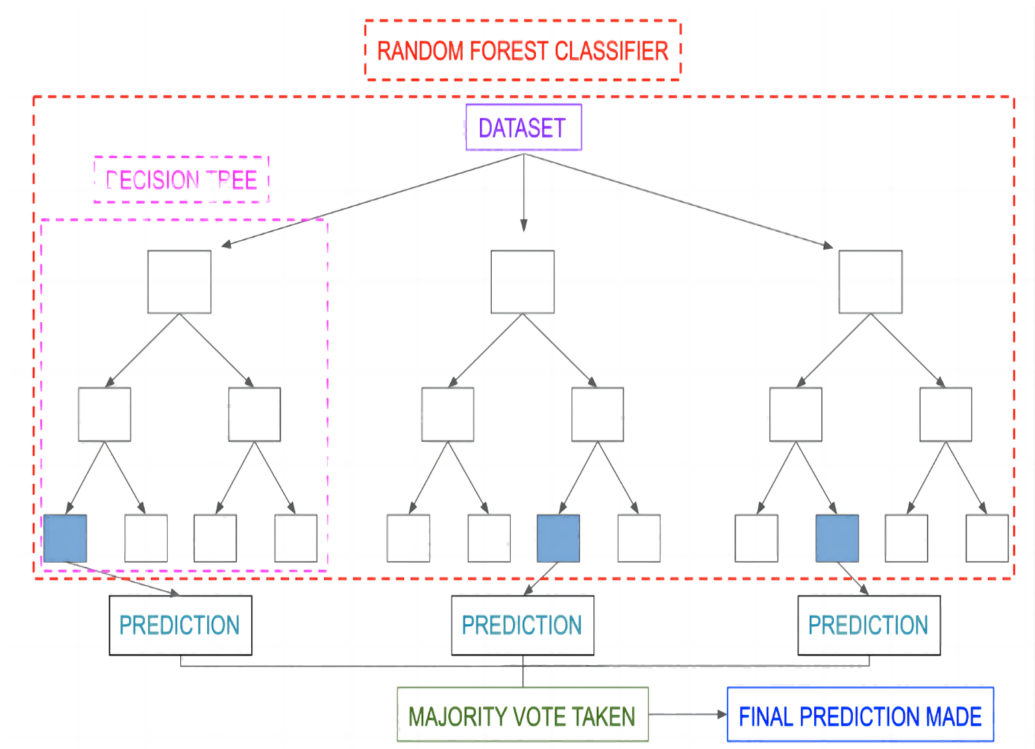$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n} (y_{i-}\mu)^2. \tag{4}$$

Figure 5 Outline of RF classifier.

Full-size ▣ DOI: 10.7717/peerjcs.1329/fig-5

The term y_i is the label of the n number of instances used for the construction of trees. The value μ represents the mean of the y_i for all the n instances considered in the study.

Overview of logistic regression

The logit or LR is a statistical model used for classification problems which estimates the likelihood of an event based on the independent variables (*Niu, 2020*). The result of LR is probability bounded within [0,1] and the transformation is obtained by the odds: ratio of success probability to the failure probability. Equation (5) depicts the Logit function.

$$\mathrm{Log}\left[\frac{Y}{(1-Y)}\right] = b_0 + b_1 X_1 + \ldots + b_n X_n. \tag{5}$$

The response variable (Y) is measured as the sigmoid function that estimates the values between [0-1]. The beta parameters are coefficients learned by the model using maximum likelihood estimation and are multiplied with the indicator or explanatory variables (X). The algorithm is run iteratively to find the best estimate. After obtaining the optimal coefficient the conditional probabilities for each record is determined and logged. The aggregation of this logged value will be the prediction or the response variable using the method of Goodness of fit.

Lu et al. (2023), *PeerJ Comput. Sci.*, DOI 10.7717/peerj-cs.1329

12/22

Proposed model to assess the quality of Innovation and Entrepreneurship in college students

The research on evaluating the quality of I&E education based on the questionnaire collected from college students is done based on the model given in Fig. 6. This model integrated two ML algorithms namely RF and LR as discussed earlier. This hybrid model analysis the explanatory variables under five different genres namely Curriculum Design, Training and skill set enrichment, Resources, Talent acquisition by teachers and Entrepreneurial environment. Each of the topics is individually fit using RF algorithm to predict the Quality Index (QI). Each of the QI from QI_1 to QI_5 represents the quantitative value of the topic. The RF algorithm is an ensemble model that integrates individual decision trees till sufficient GI, Information gain and entropy is reached. This index is the quantitative metric that gauges the performance of the I&E of the colleges. To complete the research the QIs are fit using the Logit model to give the classification of the that is the final outcome. Table 3 lists the explanatory variables considered for the research.

The estimation of individual QIs are given in Eqs. (6) to (10).

$$QI\_1 = \text{Mean}(f\_IDT(X1, X2, X3, X4, X5)) \tag{6}$$

$$QI\_2 = \text{Mean}(f\_IDT(X6, X7)) \tag{7}$$

$$QI\_3 = \text{Mean}(f\_IDT(X8, X9, X10)) \tag{8}$$

$$QI\_4 = \text{Mean}(f\_IDT(X11, X12, X13, X14, X15)) \tag{9}$$

$$QI\_5 = \text{Mean}(f\_IST(X16, X17)). \tag{10}$$

The proposed method uses predictive RF technique that averages the estimations of Individual Decision Trees (IDT) that are constructed in the implementation. Thus, it is evident from the Eqs. (6)–(10) that each RF explores the data space of the questionnaire independent of each other, to estimate the QI. Internally, the individual RFs namely RF_1, RF_2, RF_3, RF_4 and RF_5 were constructed using the information and entropy gain as discussed in the previous section. Constructing the trees based on these metrics will help the trees to assign higher adaptive weights to the instances or variables that holds greater significance. The models will be trained iteratively till better MSE is achieving, which is a direct implication of the error component between actual and predicted value.

The proposed novel hybrid model finally deploys LR or logit model that combines the statistically powerful linear regression and squeezes the estimated Y value into two classes. This result will be the comprehensive and holistic TQE that is the quality assessment outcome using the explanatory variables considered in this study. The estimation of TQE is given in Eqs. (11) and (12).

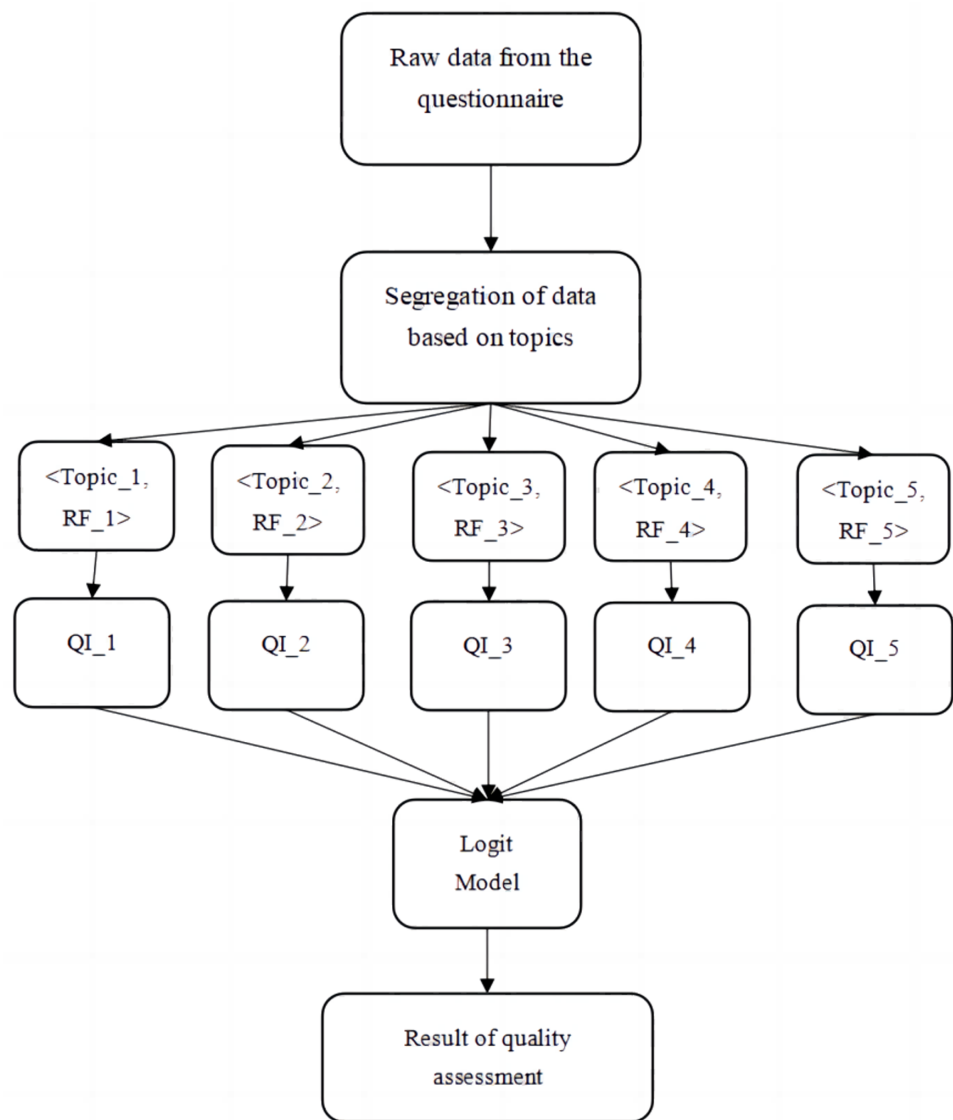$$Y = b_0 + b_1 QI_1 + b_2 QI_2 + b_3 QI_3 + b_4 QI_4 + b_5 QI_5 \tag{11}$$

Figure 6 **Proposed model to evaluate the quality of Innovation and Entrepreneurship using Random Forest and Logit model.**

Full-size ⬛ DOI: 10.7717/peerjcs.1329/fig-6

$$TQE = \frac{1}{1+e^{-Y}}.$$ (12)

The coefficients b_0, b_1, b_2, b_3, b_4 and b_5 are computed adaptively by the login algorithm by training over the presented instances of the study. The proposed novel hybrid model is is motivated to achieve the following three major goals:

1. Prediction of the quality of I&E in education using the hybrid model constructed using RF and LR.
2. Assessment of individual topic of interest by constructing the QI using the RF algorithm.
3. Ranking the HEI's performance in various topics as mentioned in Table 3.

Table 4 Ranking based on RF.

Rank	Topic	QI
1	Training and skill set enrichment	0.65
2	Resources	0.59
3	Talent acquisition by teachers	0.54
4	Curriculum Design	0.43
5	Entrepreneurial environment	0.23

Table 5 Summary of classification metrics.

S.No	Metric	Value
1.	Classification Accuracy	98.53%
2.	Precision	94.25%
3.	Recall	93.64%
4.	F1-score	93.95%

RESULTS

The training and test data is taken in the ratio of 80:20. In the pre-processing step the data is segregated into five topics and each of them is fiting individually using five RFs. Table 4 presents the ranking of the topics based on the predicted indicator (QI_1 to QI_5). The individual estimator in each RF algorithm are taken as 10, which is a hyper parameter that is learnt by the proposed ML model after series of iterations. The label for the training data is computed based on the weighted scales as mentioned in Table 2 scaled between 0-1.

The results indicate that the HEIs are more focussed on talent and skill enrichment as most of the students who answered the questionnaire gave better rating to the training programs which is due to the enforcement of vigorous reforms in educational policies in China (*Xiong, Yang & Shen, 2022*). However, the country still has room for developing the curriculum design and constructing a good entrepreneurial ecosystem for the young students. Also, the performance of the hybrid model in assessing the quality of I&E in education is validated using the following classification metrics: accuracy, precision, recall, F1-score and support (*Mehbodniya et al., 2022*).

The complete metrics of the proposed hybrid model are presented in Table 5. The hybrid model have predicted the quality with accuracy of 98.53% which is a promising value. The model also has very low misclassification rate that are witnessed through the high precision, recall and F1-Score values. The graphical representation of the efficacy of the model in evaluating the quality is given as Fig. 7.

Table 6 shows the comparative performance analysis of the proposed work with other state of art techniques like ANN, SVM and DTs.It can be seen that SVM, ANN and individual DTs gives much lower accuracy when compared to the proposed work. This work trains individual DT on each topic and later combines the results of the trees to train the logit model. In addition to classification accuracy, the proposed work also assess the
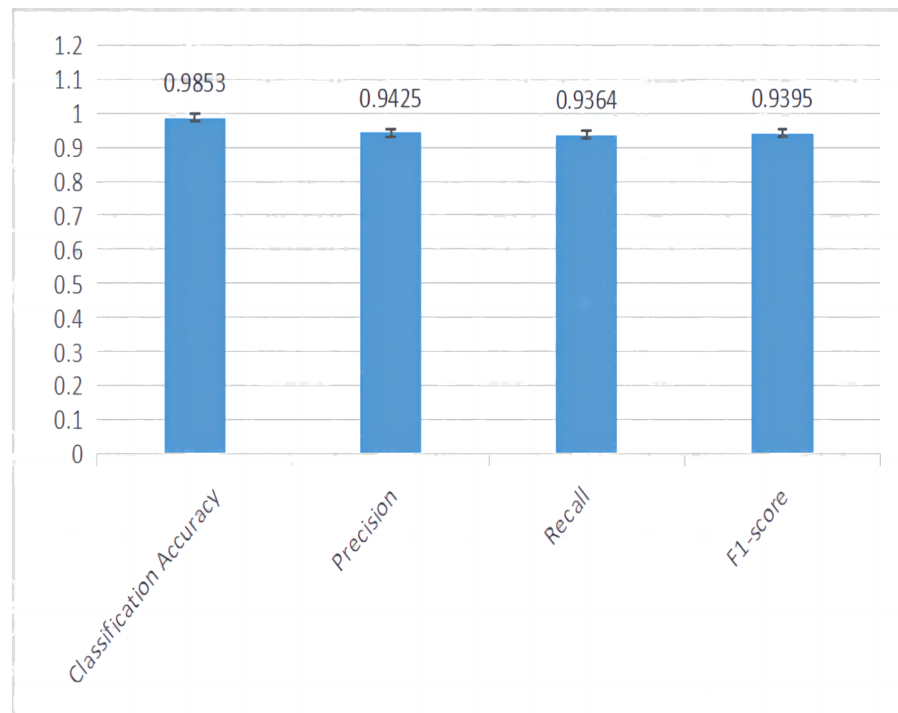
Figure 7 Graphical representation of the performance of proposed model.

Table 6 Performance comparison with other state of art techniques.

S.No	Method	Classification accuracy	F1-Score
1.	SVM	90.81%	88.46%
2.	ANN	93.86%	90.68%
3.	DT	91.86%	84.13%
4.	Proposed work	98.53%	93.95%

model's performance on the misclassification rate that is ,assured through F1-score. The model has better F1-score of 93.95% followed by ANN model.

The above research infers some important research implications. Table 7 gives the assessment results of the 25 HEI's that are considered in this study.Few of the significant findings are explained below:

• The results reveal that the overall impact on entrepreneurial education in terms of I&E in the colleges can be improved by integrating the elements of I&E as a part of the curriculum design based on the learning abilities and teaching efficiency of the students and teachers respectively. Also, it is important to include I&E in almost all disciplines of education based on the absorption ability and the skill set of the students. Special emphasis must be given to students in higher years of education to decide on their future career inclination either to start own enterprise or getting employed in firms.

Table 7 Quality evaluation results of the research using the proposed hybrid ML model.

HEI	Ranking of the topics from (1-5)					TQE
	QI_1	QI_2	QI_3	QI_4	QI_5	
1.	1	4	2	3	5	Yes
2.	2	1	4	5	3	Yes
3.	3	1	2	4	5	No
4.	2	1	5	3	4	No
5.	4	3	5	1	2	No
6.	5	2	3	4	1	No
7.	1	2	5	3	4	Yes
8.	4	5	1	2	4	Yes
9.	4	2	3	1	5	Yes
10.	3	2	1	4	5	No
11.	4	1	2	3	5	No
12.	1	4	5	3	2	Yes
13.	5	2	3	1	4	Yes
14.	3	1	2	5	4	No
15.	4	2	1	3	5	Yes
16.	3	2	5	1	4	Yes
17.	2	3	1	3	5	No
18.	4	1	5	3	2	Yes
19.	2	4	1	5	3	No
20.	1	2	2	3	5	Yes
21.	2	3	4	1	5	Yes
22.	5	1	3	4	2	Yes
23.	2	3	4	5	1	No
24.	4	3	1	5	2	Yes
25.	3	1	2	1	5	No

- Colleges and HEIs should invite notable alumni, employment workshops, entrepreneurs, innovation camps, idea camps and experts with professional experience to foster the entrepreneurial ecosystem. This will motivate the students to combine their experiences and skills to understand and contribute to the development of enterprises.
- Special incentives and awards can be felicitated by the HEI's to students who excel in I&E skills. This will aspire the student's community to participate in more competitions and bring out innovative technology transfer ideas.
- Proper training should be given to teachers in transforming the teaching pedagogy from teacher centric to student centric notion. This will increase the student involvement.
- Multi-disciplinary and cross platform projects should be fostered and encouraged.
- HEIs must pay more attention to build world class laboratories, workspaces, project platforms and business simulation centres so that the student community can experience the business feel.

CONCLUSIONS

The research on quality evaluation of I&E elements in education among the college students is empirically analysed using the hybrid RF- LR model. The educational reforms in China have opened doors for many regional and international students for pursuing entrepreneurial education. The data for the study is collected through a questionnaire and the responses are rated. The proposed model, (1) constructs individual QIs using RFs for five topics considered in the study, (2) ranks the QIs to understand the strength and weaknesses of HEI's, (3) and fits a Logit model that classifies the quality of the I&E. The results indicate that training and skillset enrichment has been developed better than other classes. However, significant focus must be given by the government and HEIs in designing the curriculum and fostering entrepreneurial ecosystem. Also, the effectiveness of the proposed hybrid model is validated using classification metrics namely classification accuracy, F1-score, precision and recall. The results showed that the model showed promising results in assessing the quality of I&E in educational field among the college students. This research is focusses on questionnaires developed based only on six topics, which can be further extended to understand the I&E capability of students of various disciplines and the factors considered can include the cognitive and meta cognitive traits. And the combination of fuzzy algorithm (*Dirik, 2022*; *Karahan et al., 2022*) may make the data analysis more accurate, which is also where this paper will continue to study in the future.

ADDITIONAL INFORMATION AND DECLARATIONS

Competing Interests
The authors declare there are no competing interests.

Author Contributions

- Qianqian Lu conceived and designed the experiments, analyzed the data, authored or reviewed drafts of the article, and approved the final draft.
- Yongxiang Chai conceived and designed the experiments, performed the experiments, authored or reviewed drafts of the article, and approved the final draft.
- Lihui Ren conceived and designed the experiments, performed the experiments, authored or reviewed drafts of the article, and approved the final draft.
- Pengyu Ren analyzed the data, prepared figures and/or tables, and approved the final draft.
- Junhui Zhou analyzed the data, prepared figures and/or tables, and approved the final draft.
- Chunlei Lin analyzed the data, prepared figures and/or tables, and approved the final draft.

Data Availability

The following information was supplied regarding data availability:

The raw data is available in the Supplemental Files.

Supplemental Information

Supplemental information for this article can be found online at http://dx.doi.org/10.7717/peerj-cs.1329#supplemental-information.

REFERENCES

Alexandros G, Alexandra IC. 2018. Cognitive agents and machine learning by example: representation with conceptual graphs. *Computational Intelligence* **34(2)**:603–634 DOI 10.1111/coin.12167.

Boldureanu G, Ionescu AM, Bercu AM, Bedrule-Grigoruţă MV, Boldureanu D. 2020. Entrepreneurship education through successful entrepreneurial models in higher education institutions. *Sustainability* **12(3)**:1267 DOI 10.3390/su12031267.

Chen X. 2022a. Internet plus innovation and entrepreneurship education model based on machine learning algorithms. *Mobile Information Systems* **2022**:6176675 DOI 10.1155/2022/6176675.

Chen Y. 2022b. The effectiveness of innovation and entrepreneurship education in China. *Education, Language and Sociology Research* **3(2)**:1–14 DOI 10.22158/ELSR.V3N2P1.

Dirik M. 2022. Detection of counterfeit banknotes using genetic fuzzy system. *Journal of Fuzzy Extension and Applications* **3(4)**:302–312 DOI 10.22105/jfea.2022.345344.1223.

Edwards Audrene S, Kaplan B, Jie T. 2020. A primer on machine learning. *Transplantation* **105(4)**:699–703 DOI 10.1097/TP.0000000000003316.

Eisinga R, Grotenhuis MT, Pelzer B. 2013. The reliability of a two-item scale: Pearson, Cronbach, or Spearman-Brown? *International Journal of Public Health* **58**:637–642 DOI 10.1007/s00038-012-0416-3.

Fawagreh K, Gaber MM, Elyan E. 2014. Random forests: from early developments to recent advancements. *Systems Science & Control Engineering* **2(1)**:602–609 DOI 10.1080/21642583.2014.956265.

Geng L. 2022. The recommendation system of innovation and entrepreneurship education resources in universities based on improved collaborative filtering model. *Computational Intelligence and Neuroscience* **2022**:7228833 DOI 10.1155/2022/7228833.

Gong YJ, Ma LR, Jin JG, Gu WH. 2020. Research on current construction situation and promotion measures of business incubation platform of colleges and universities in Hebei. *Education Modernization* **7(12)**:19–21 DOI 10.16541/j.cnki.2095-8420.2020.12.007.

Gordon A, Helen B. 2022. Exploration of educational data mining in improving teaching quality. *Journal of Educational Research and Policies* **4(6)**:42–48 DOI 10.53469/JERP.2022.04(06).09.

Graham B, Bonner K. 2022. One size fits all? Using machine learning to study heterogeneity and dominance in the determinants of early-stage entrepreneurship. *Journal of Business Research* **152**:42–59 DOI 10.1016/J.JBUSRES.2022.07.043.

Guo K. 2022. Research on the development of innovation and entrepreneurship education in colleges and universities from the perspective of regional development. *Jiangsu Commercial Forum* **458(12)**:108–110 DOI 10.13395/j.cnki.issn.1009-0061.2022.12.008.

Han G, Huang Y, Yang C, Meng X, Yan M. 2018. Study on construction and bottleneck analysis of college students'venture fund. *Journal of Shenyang Institute of Engineering* **14(01)**:134–138 DOI 10.13888/j.cnki.jsie(ss).2018.01.022.

He W, Dong X. 2022. Innovation and entrepreneurship education development: value concept, ecosystem, practice optimization. *Contemporary Education and Culture* **14(5)**:103–107 DOI 10.13749/j.cnki.cn62-1202/g4.2022.05.001.

Huajie X, Yanping Z, Huiying C, Jing L, Yuanyuan K. 2022. The impact of international technical cooperation in new energy industry on carbon emissions: evidence from the top 30 countries in the global innovation index. *Environmental Science and Pollution Research International* **30(2023)**:21708–21722 DOI 10.1007/S11356-022-23770-7.

Huang W, Liu B, He X, Zhou Y, Ning B, Wang L. 2019. December. BP neur al network study on the quality evaluation system of innovation and entrepreneurship education in local universities. *Journal of Physics:Conference Series* **1423**:012058 DOI 10.1088/1742-6596/1423/1/012058.

Jiang Y, Li B. 2021. Exploration on the teaching reform measure for machine learning course system of artificial intelligence specialty. *Scientific Programming* **2021**:8971588 DOI 10.1155/2021/8971588.

Jing G. 2022. Research on evaluation algorithm of innovation and entrepreneurship effect of college graduates based on DL. *Mathematical Problems in Engineering* **2022**:1802156 DOI 10.1155/2022/1802156.

Karahan M, Lacinkaya F, Erdonmez K, Eminagaoglu ED, Kasnakoglu C. 2022. Age and gender classification from facial features and object detection with machine learning. *Journal of Fuzzy Extension and Applications* **3(3)**:219–230 DOI 10.22105/jfea.2022.328472.1201.

Karen W. 2008. *Entrepreneurship education in Europe. Entrepreneurship and higher education.* OECD Publishing, France.

Knox J. 2019. The 'Creative, problem-solving entrepreneur': alternative futures for education in the age of machine learning? In: Peters M, Jandrić P, Means A, eds. *Education and technological unemployment.* Singapore: Springer, 297–312 DOI 10.1007/978-981-13-6225-5_19.

Li J, Zhao Y. 2021. Construction of innovation and entrepreneurship platform based on deep learning algorithm. *Scientific Programming* **2021**:1833979 DOI 10.1155/2021/1833979.

Li Y. 2022. Similar classification algorithm for educational and teaching knowledge based on machine learning. *Wireless Communications and Mobile Computing* **2022**:7222236 DOI 10.1155/2022/7222236.

Liñán F, Fayolle A. 2015. A systematic literature review on entrepreneurial intentions: citation, thematic analyses, and research agenda. *International Entrepreneurship and Management Journal* **11(4)**:907–933 DOI 10.1007/s11365-015-0356-5.

Lihua D. 2022. An extended model of the theory of planned behavior: an empirical study of entrepreneurial intention and entrepreneurial behavior in college students. *Frontiers in Psychology* **12**:627818 DOI 10.3389/FPSYG.2022.627818.

Liu X, Ardakani SP. 2022. A machine learning enabled affective E-learning system model. *Education and Information Technologies* **27(7)**:21–22 DOI 10.1007/S10639-022-11010-X.

Mehbodniya A, Lazar AJP, Webber J, Sharma DK, Jayagopalan S, K K, Singh P, Rajan R, Pandya S, Sengan S. 2022. Fetal health classification from cardiotocographic data using machine learning. *Expert Systems* **39(6)**:e12899 DOI 10.1111/exsy.12899.

Meyners M, Hasted A. 2022. Reply to Bi and Kuesten: ANOVA outperforms logistic regression for the analysis of CATA data. *Food Quality and Preference* **95**:104339 DOI 10.1016/j.foodqual.2021.104339.

Niu . 2020. A review of the application of logistic regression in educational research: common issues, implications, and suggestions. *Educational Review* **72(1)**:41–67 DOI 10.1080/00131911.2018.1483892.

Onesmus M. 2020. Introduction to random forest in machine learning. *Available at https://www.section.io/engineering-education/introduction-to-random-forest-in-machine-learning/.*

Qing Y, Zejun W. 2021. Research on the impact of entrepreneurship policy on employment based on improved machine learning algorithms. *Journal of Intelligent & Fuzzy Systems* **40(4)**:6517–6528 DOI 10.3233/JIFS-189490.

Rastrollo-Guerrero JL, Gómez-Pulido JA, Durán-Domínguez A. 2020. Analyzing and predicting students' performance by means of machine learning: a review. *Applied Sciences* **10(3)**:1042 DOI 10.3390/app10031042.

Ratten V, Usmanij P. 2020. Entrepreneurship education: time for a change in research direction? *The International Journal of Management Education* **19(1)**:100367 DOI 10.1016/j.ijme.2020.100367.

Sabahi S, Parast MM. 2020. The impact of entrepreneurship orientation on project performance: a machine learning approach. *International Journal of Production Economics* **226**:107621 DOI 10.1016/j.ijpe.2020.107621.

Samer MA, Eman AZ, Anees HR, Basem SA, Samy SAN. 2022. Predicting university student retention using artificial intelligence. *International Journal of Advanced Computer Science and Applications* **13(9)**:315–324 DOI 10.14569/IJACSA.2022.0130937.

Setiawan A. 2021. Analysis of the effect of knowledge on entrepreneurship readiness using random forest classification machine learning. *Technium Social Sciences Journal* **23(1)**:134–149 DOI 10.47577/tssj.v23i1.4500.

Trakunphutthirak R, Lee Vincent CS. 2022. Application of educational data mining approach for student academic performance prediction using progressive temporal data. *Journal of Educational Computing Research* **60(3)**:742–776 DOI 10.1177/07356331211048777.

Vij S, Tayal D, Jain A. 2020. A machine learning approach for automated evaluation of short answers using text similarity based on WordNet graphs. *Wireless Personal Communications* **111(2)**:1271–1282 DOI 10.1007/s11277-019-06913-x.

Wang C, Zheng P, Zhang F, Qian Y, Zhang Y, Zou Y. 2022b. Exploring quality evaluation of innovation and entrepreneurship education in higher institutions using deep learning approach and fuzzy fault tree analysis. *Frontiers in Psychology* **12**:767310 DOI 10.3389/FPSYG.2021.767310.

Wang H. 2022. Ecological system analysis of U.S. innovation and entrepreneurship education and its inspiration for Chinese Universities. *Education Journal* **11(5)**:249–253.

Wang J, Liu H. 2022. Exploration of quantitative evaluation of innovation and entrepreneurship education in colleges and universities: India's Atal ranking and its enlightenment. *Journal of National Academy of Education Administration* **298(10)**:50–61.

Wang W, Qiu D, Chen X, Yu Z. 2022a. An empirical study on the evaluation system of innovation and entrepreneurship education in applied universities. *Computer Applications in Engineering Education* **31(1)**:100–116 DOI 10.1002/cae.22573.

Xiong W, Yang J, Shen W. 2022. Higher education reform in China: a comprehensive review of policymaking, implementation, and outcomes since 1978. *China Economic Review* **2022**:101752 DOI 10.1016/J.CHIECO.2022.101752.

Yulian F, Liping L, Weixia H, Suying Z. 2014. Collegian cultural form of the American entrepreneurial universities and its enlightenment. In: *Proceedings of 2014 2nd international conference on social science and health (ICSSH 2014 V57), Taipei, Taiwan*. 101–104.

Zhang JL. 2020. The application of human comprehensive development theory and deep learning in innovation education in higher education. *Frontiers in Psychology* **11**:1605 DOI 10.3389/fpsyg.2020.01605.

Zhou Y, Li H, Shahzad F. 2021. Does college education promote entrepreneurship education in China? *SAGE Open* **11(3)**:1–10 DOI 10.1177/21582440211031616.

Lu et al. (2023), *PeerJ Comput. Sci.*, DOI 10.7717/peerj-cs.1329

22/22