Research Article

Optimization of Entrepreneurship Education for College Students Based on Improved Random Forest Algorithm

Dongfeng Jia\(^1,2\) and Hui Zhao\(^1,2\)

\(^1\)Hunan Institute of Technology, Hunan 421002, Hengyang, China
\(^2\)Adamson University, Manila 900, Philippines

Correspondence should be addressed to Hui Zhao; 2020001067@hnit.edu.cn

Received 12 May 2022; Revised 17 June 2022; Accepted 27 June 2022; Published 31 July 2022

Academic Editor: Shadi Aljawarneh

Copyright © 2022 Dongfeng Jia and Hui Zhao. 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.

As we all know, the random forest algorithm has the advantages of high classification intensity and wide application range. Nevertheless, it still has a lot of room for improvement. This paper introduces the basic idea and working principle of classification algorithm and random forest algorithm, so this paper proposes some improved algorithms on the basis of all current research studies. College students, as the most innovative and adaptable group, have gradually become a development direction of China’s education reform to cultivate their innovative employment ability and develop innovative employment education. By constructing a model of the relationship between innovation and entrepreneurship education and college students’ employability, this paper examines the impact of innovation and entrepreneurship education mode on college students’ employment and entrepreneurship. At the same time, the model is tested by improving the random forest algorithm. The process of talent quality evaluation is to select talents with the best comprehensive quality based on various indicators of students’ performance in school, which can be regarded as a classification problem of unbalanced data sets. The improved random forest algorithm proposed in this paper has little difference in precision and recall rate when it is used for talent training evaluation but has a certain degree of improvement in accuracy rate, which meets the design requirements.

1. Introduction

Because the random forest algorithm has many excellent features, when the random forest is used in biological prediction, fault detection, network attacks, and other fields, good results are achieved, but it is worth mentioning that there are relatively few improvement studies on the random forest algorithm itself [1]. It can be seen that the random forest has many advantages, but the random forest algorithm has some disadvantages that cannot be changed [2]. For example, the random forest cannot balance the data well, control or detect the specific conditions in the model, but simply try through parameter adjustment and random data [3]. Therefore, this article needs to optimize and improve the random forest algorithm and explain some situations about random forest. If we want to improve the random forest, we generally need to consider two aspects. The first aspect is to improve the classification strength of each classifier and each tree, which is determined by the overall characteristics of the entire random forest decision, which can show the strength of the classification of the random forest [4]. The other is to reduce the correlation between the various departments of the classifiers, so that each classifier becomes an “expert” in its own field [5]. Because trees exist independently of each other and determine the final result together, they become experts in their respective fields. The stronger the correlation between the trees, the worse the classification performance of RF [6]. The essence of innovation and entrepreneurship education is to improve the comprehensive education and employability of college students and enhance the employment and entrepreneurial capital of college students [7]. This article constructs a model of the relationship between innovation and entrepreneurship education and college students’ employability and conducts related research on
college students’ innovation and entrepreneurship against the background of the "Internet+.” After several years of development in China, it can be seen that its innovation and entrepreneurship education has made great progress, and many experts and scholars have conducted research on the training effects of innovation and entrepreneurship education in colleges and universities [8]. Through combining the existing literature, it is found that the current research focus is mainly on the impact of innovation and entrepreneurship education on innovation and entrepreneurship ability and entrepreneurial willingness. From these studies, it can be seen that such educational activities as innovation and entrepreneurship courses, entrepreneurship competitions, innovation and entrepreneurship lectures, and entrepreneurial clubs can significantly improve the innovation and entrepreneurship ability and willingness of college students [9]. Entrepreneurship education and ideological education for college students are based on the theory of the all-round development of people, so that the combination of entrepreneurship education and ideological education for college students can play the guiding role of ideological education in entrepreneurship education and make entrepreneurship education the ideological education [10].

2. Related Work

The literature random forest is mainly applied to two types of related problems, classification and regression, and at the same time it also has the ability to reduce data and is suitable for dealing with data problems such as fault values, missing values, noise, and outliers. Random forest algorithm can also solve some problems that traditional algorithms cannot solve [11, 12]. For example, random forest algorithm can avoid problems caused by data imbalance. Due to the combination of models, the random forest algorithm reduces some of the bad effects caused by overfitting, finally having a better effect than that of the decision tree, and it has improved a lot [13]. According to the literature, under a large data set, oversampling is not as good as undersampling. The literature proposes an essential feature of the random forest algorithm-stratified sampling, and the use of a vector machine as a base classifier at the node splitting makes it a huge advantage and makes it better than traditional classification on text imbalanced data [14]. The literature shows that the radial basis function neural network of the random forest algorithm is based on the basis classifier, and it is suitable for the classification of work balance and unbalanced data. The definition of the essential neural network of the radial basis function neural network as the basis classifier of the random forest algorithm is also proposed; it can be applied to the classification problems of medium and high imbalance and balanced data that appear in the work [15]. In the literature, since the weighted weights of decision trees in the random forest algorithm have the same value, the idea of scientists wanting to modify the weighted weights has become popular and widespread [16]. This led to the emergence of weighted random forest algorithm. The weight setting is based on the similarity size and proportion between each tree. The weighted random forest algorithm has some better characteristics, which makes the effect and display better than the traditional random forest algorithm [17]. The literature combines the essential idea algorithm of k-nearest neighbor with the random forest algorithm, makes improvements, and then proposes the algorithm of k-PNNs, which has a huge impact and enlightenment on the algorithm circle [18].

3. Research on Improved Random Forest Algorithm

3.1. Overview of Random Forest Algorithms. For supervised learning, the steps are as follows: obtain several training samples, input X into each sample obtained, get one and only one output result Y, and then train a specific model (mapping f : X → Y); at the same time, this unknown sample X' can make predictions about the result Y'. If Y' is a discrete variable, such as whether smoking is harmful to your health and whether a robot can replace all human work, such a problem is a classification problem; but if it is a continuous variable, such a problem is a regression problem.

The combination of the two algorithms of bagging and random subspace is called the random forest algorithm. Decision trees are the most basic neural network unit and thus play an important role. By this way, the important role played by the combination of multiple decision trees can be improved. Through the classification accuracy such as h1(x), h2(x), . . . , hKTr(x), we can get the random forest classifier we want (Figure 1). If we want to solve the problem caused by regression disorder, the result of the weighted average of the common results of all decision trees is required.

Randomly generating training set: Use the famous bootstrapping method to randomly sample the data of each tree. As shown in Figure 1, the data set of each tree is randomly composed of the original data set, and may contain some data repeatedly or may not contain some data.

Random selection of feature subsets: When the tree nodes of each tree start to split, we need to randomly select a subset of the overall data attributes without replacement, and the size of the subset K is required to satisfy the number of features less than the total attribute M. Then, we can randomly extract the selected data K from M, then calculate the various indicators of the continuous data changes under K splits, select the best split feature index, continue to divide, and at this time make the selection field MK.

According to the definition of entropy value we know, the entropy value of sample data T can be obtained:

\[
\text{Entropy}(T) = -\sum_{i=1}^{c} P_i \log_2 P_i. \tag{1}
\]

Among them, \(P_i\) indicates how much of the sample of category \(i\) accounts for the total number of samples. When the obtained data feature A comes into play, the sample T will be divided into k parts. At this time, we can know that the obtained information entropy and information gain are as follows:
Entropy(T, A) = \sum_{j=1}^{k} \frac{|T_j|}{|T|} \cdot \text{Entropy}(T_j), \quad (2)

Gain(T, A) = \text{Entropy}(T) - \text{Entropy}(T, A).

And according to the definition of entropy in information theory, the gain efficiency of information is proportional to the effect of classification:

\[ \text{GainRation}(T, A) = \frac{\text{Gain}(T, A)}{\text{SplitEntropy}(T, A)}. \quad (3) \]

SplitEntropy(T, A) = -\sum_{i=1}^{c} \frac{|T_i|}{|T|} \log_{2} \frac{|T_i|}{|T|}.

And according to the definition of entropy in information theory, the negative ratio between Gini coefficient and classification effect is

\[ \text{Gini}(T) = 1 - \sum_{i} p_{i}^{2}, \quad (4) \]

\[ \text{Gini}(T, A) = \sum_{j=1}^{k} \frac{|T_j|}{|T|} \text{Gini}(T_j). \]

Through the comparison of the above three attribute selection rules, it is found that the Gini coefficient measurement has less impact, so this article studies the Gini coefficient measurement.

The accuracy of classification can be defined as the accuracy of classification representing the total sample size:

\[ \text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} = \frac{TP + TN}{N}. \quad (5) \]

The current F value, is an important indicator for evaluating unbalanced data sets. The size of \( \beta \) can indicate that the recall rate and the precision rate are more important and more prominent. When \( \beta > 1 \), the recall rate is more important, and when \( \beta < 1 \), the recall rate is less important. Take the value according to the actual situation; generally, \( \beta = 1 \) means that the recall rate and the precision rate are of the same importance. Only when both are high, the F value is larger, which means a better classifier effect.

\[ F = \frac{1 + \beta^2}{\beta} \frac{\text{Recall} \cdot \text{Precision}}{\text{Recall} + \text{Precision}} + (1/\text{Recall}) \quad \beta \in (0, 1], \quad (6) \]

\[ F = \frac{2}{(1/\text{Precision}) + (1/\text{Recall})}. \]

G-mean, because in order to judge the fairness of the standard, we need to have the F value and the geometric mean, so we need the following formula:

\[ G - \text{mean} = \sqrt{\text{Sensitivity} \cdot \text{Specificity}}. \quad (7) \]

The contrast classifier is the purpose and theme of the simulation experiment in this section. First, we need to test some data to select the UCI data set; then, we divide these data into two categories, need to make their imbalance ratio gradually lower, and also make the agreed subcategory data not a negative category (Table 1).

Normalization processing: through Table 1, it can be found that due to the large difference in the attribute characteristics of the above data, they are not of the same order of magnitude, so before we build the data model, first, we need to normalize the input attribute characteristic matrix and carry out chemical treatment at the same time, and place it in \([-1,1]\).

**3.2. Optimization of Random Forest Algorithm.** When the data obtained in the training set is unbalanced, the random forest algorithm will not balance the training data of each tree randomly selected in its self “random” process.

The small-class sample synthesis resampling technology is a relatively mature-SMOTE algorithm. Its specific thought
characteristics are as follows: firstly give a small-class sample $X_i$, obtain $k$ adjacent samples for it, then randomly select $n$ among these $K$ samples, denote it as $X_{\text{new}}$, and add a new sample $X_{\text{new}}$ by interpolation:

$$X_{\text{new}} = X_{\text{origin}} + \text{rand} \times (X_i - X_{\text{origin}}), \quad i = 1, 2, \ldots, n. \quad (8)$$

Although the SMOTE algorithm has some advantages, it also has many problems. For example, when selecting neighbors, you need to manually set the value of $k$ by yourself, which is false and self-deceptive. Therefore, great scholars have made improvements and optimizations day and night and finally proposed the Borderline-SMOTE algorithm. The definition of the algorithm is shown in Table 2.

The traditional classification algorithm aims to maximize the classification accuracy of all data when the hypothesis is classified into the wrong class. Cost-sensitive learning is also based on traditional algorithms; however, it does not change the algorithm but assigns weights based on the cost caused by the wrong classification. Finally, the cost-sensitive classification algorithm is designed, but this design will classify small-class samples. The cost is not low.

Due to the imbalance of sample size data, the distribution of features also appears imbalanced. For example, some situations that appear in this article may only appear in some major categories. Therefore, we need to select the classification with the most discriminative ability based on the problems caused by the imbalance of the data classification, which will help improve the discrimination rate of the subcategories.

The CURE algorithm is a clustering algorithm for large databases developed by scientists:

**Step 1.** We need to normalize the data set, which has been mentioned above; then extract the data subsample $X_i$ and then calculate the mutual distance dist of the data subsamples. For a cluster class $U$ that we obtain, $U_r$ and $U_c$ represent the collection point and center point of the data class $U$. Assuming that $p$ and $q$ are data items, we can get the formula for the distance between two clusters $U$ and $V$:

$$\text{dist} (U, V) = \min_{p \in U_r, q \in V_r} \text{dist} (p, q). \quad (9)$$

**Step 2.** Set the number of clusters $c$, then cluster them, then merge the two clusters with the closest interval, and finally update the center point and the representative point.

<table>
<thead>
<tr>
<th>Serial number</th>
<th>Data set</th>
<th>$N$</th>
<th>$M$</th>
<th>Positive</th>
<th>Negative class</th>
<th>IR</th>
<th>Category</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Iris</td>
<td>100</td>
<td>4</td>
<td>50</td>
<td>50</td>
<td>1:1</td>
<td>Versicolor: virginica</td>
</tr>
<tr>
<td>2</td>
<td>Pokertrain</td>
<td>23092</td>
<td>10</td>
<td>10559</td>
<td>12493</td>
<td>0.8484:1</td>
<td>1:0</td>
</tr>
<tr>
<td>3</td>
<td>Wine</td>
<td>130</td>
<td>13</td>
<td>59</td>
<td>71</td>
<td>0.8310:1</td>
<td>1:2</td>
</tr>
<tr>
<td>4</td>
<td>Magic</td>
<td>19020</td>
<td>10</td>
<td>6688</td>
<td>12332</td>
<td>0.5423:1</td>
<td>g:h</td>
</tr>
<tr>
<td>5</td>
<td>Breast Cancer Wisconsin</td>
<td>702</td>
<td>10</td>
<td>243</td>
<td>459</td>
<td>0.5249:1</td>
<td>1:0</td>
</tr>
<tr>
<td>6</td>
<td>Haberman’s survival</td>
<td>306</td>
<td>3</td>
<td>81</td>
<td>225</td>
<td>0.0360:1</td>
<td>2:1</td>
</tr>
<tr>
<td>7</td>
<td>Blood transfusion</td>
<td>748</td>
<td>4</td>
<td>178</td>
<td>570</td>
<td>0.3123:1</td>
<td>4:2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 2: Definitions in Borderline-SMOTE.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Point</strong></td>
</tr>
<tr>
<td>Noise point</td>
</tr>
<tr>
<td>Boundary points/dangerous points, easy to make mistakes</td>
</tr>
<tr>
<td>Safe point</td>
</tr>
</tbody>
</table>

$$U_c = \frac{|U| \cdot U_c + |V| \cdot V_c}{|U| + |V|},$$

$$U_r = \{ p + \alpha \cdot (U_c - p) | p \in U_r \}. \quad (10)$$

**Step 3.** According to the newly revised SMOTE interpolation algorithm formula by the scientists, new data samples are generated. $XX$ represents the data sample points obtained after the CURE algorithm clustering. It is worth mentioning that the results we obtained will move randomly to representative points.

$$X_{\text{new}}^n = \overline{X} + \text{rand} (0, 1) \times (U_r - \overline{X}), \quad (11)$$

**Step 4.** We need to calculate the imbalance IR obtained by the data set. If the data set we obtain does not reach the lowest IR0, then return to Step 3.

**Step 5.** We need to use the convenience of the random forest algorithm to point out the new processed data set $X_{\text{new}} = \overline{X} \cup \{ X^p_{\text{new}} \}$ and large class samples, and then start classification and a series of other operations.

So we can get the algorithm design flow chart as shown in Figure 2.

When performing data set selection and artificially generating data circle and unbalanced data set class in UCI, it can be seen that the training set and test set are divided into 3:1 stratified random sampling, and SPECT uses training and test files.

In this section, the random oversampling algorithm is used as an example to study the rules of each evaluation index of F, D-mean, AIC, and OOB error.

**Objective function** is as follows: $f (n \text{Tree}^*, K^*, \{ \text{Attribute}|i = 1, 2, \ldots, M \}) = \arg \min \text{avg OOB error}$.

Optimization variables are as follows: $n \text{Tree}$, $K$, $\{ \text{Attribute}|i = 1, 2, \ldots, M \}$.

**Binary encoding:** We can see that Figure 3 contains two tangent points and three fragments. $n \text{Tree}$ and $K$ are the embodiment of binary data, 0 of $\{ \text{Attribute}|i = 1, 2, \ldots, M \}$.
means this part is not selected, and 1 means this part is selected. And the constraint condition is \( K \leq \sum \text{Attribute}_i = 1 \).

Step 1. Combine the RF classification calculator to calculate \( F = \max(1/f) \), gen = 1.

Step 2. Selection operation: decide whether to select the individual according to the proportion of fitness value.

\[
P_i = \frac{F_i}{\sum_{i=1}^{\text{popsize}} F_i},
\]

(12)

Step 3. Intersection operation: according to the single-point crossover method, we randomly select two individuals, make a random change of 0-1, and then judge whether the solution is correct.

Step 4. Change operation: we randomly select an individual, randomly select the changed position \( j \), make the 0-1 switch, then judge whether the solution is correct, calculate the fitness value, and update the optimal solution.

Step 5. When the result of \( \text{gen} > \text{maxgen} \) appears, it means the end; otherwise, it returns to the third step.

Hybrid PSO-RF process design is as follows:

Step 1. Population initialization: the number of population \( \text{popsize} \), the number of changes maxgen; the binary data position of the particle \( X_k = \{Z_{k,1}, Z_{k,2}, \ldots\} \), \( K = 1, 2, \ldots \ \text{popsize} \); the required data speed \( V \); the learned data factors \( c_1, c_2 \); and the false weight number \( w \).

Step 2. We combine the RF classifier to start calculating the fitness \( F = \max(1/f) \), and input the parameter combination into the RF algorithm, gen = 1.

Step 3. The velocity \( V^{(K+1)} \) and position \( X^{(K+1)} \) of the particles to be used, and \( kP \ PK \) is expressed as part of the best position, and \( P_{jk} \) is the best overall, where rand represents a random number.
Hybrid AFSA-RF process design is as follows:

**Step 1.** Population initialization: the number of population is \( n_p \), the number of changes is \( m \), \( K = 1, 2, \ldots, n_p \); place where the fish school is \( X(\mathbf{z}) = \{Z_{k,1}, Z_{k,2}, \ldots\} \), \( k = 1, 2, \ldots, n_p \); the range visible that the artificial fish can see is \( \mathbf{p} \), the crowding factor is \( \mathbf{f} \); the maximum number of attempts of the fish is \( \mathbf{a} \); the maximum number of generations is \( \mathbf{g} \). When \( \mathbf{g} > \mathbf{a} \), return to Step 2; otherwise, return to Step 3.

**Step 2.** We need to combine the RF classifier to calculate the concentration of fish food \( \mathbf{m} \).

**Step 3.** At the same time, we need to carry out experiments on the grouping behavior and rear-end collision behavior of fish:

(a) Group behavior of fish: the current state of artificial fish is \( X_1 \), the number of fish partners in the viewing range is \( \mathbf{f} \), and the center position of the lake is \( X_2 \). When \( F_{c,f} > \mathbf{f} \), change the position of the fish according to the following formula, and we start foraging behavior if it does not match.

\[
Z_{k+1,j} = \begin{cases} 
Z_{k,j}, & Z_{k,j} = Z_{c,j}, \\
0, & Z_{k,j} \neq Z_{c,j}, \text{rand} > 0.5, \\
1, & Z_{k,j} \neq Z_{c,j}, \text{rand} \leq 0.5
\end{cases}
\]

(b) Tail-catch behavior of fish: we find the fish \( X_{\mathbf{m}} \) with the maximum food concentration \( F_{\mathbf{m}} \) in the fish school. If \( F_{\mathbf{m},\mathbf{f}} > \mathbf{f} \), change the place of the fish with the following formula, calculate the concentration of fish food, and update the food concentration; if not, perform other foraging behaviors.

\[
Z_{k+1,j} = \begin{cases} 
Z_{k,j}, & Z_{k,j} = Z_{\mathbf{m},j}, \\
0, & Z_{k,j} \neq Z_{\mathbf{m},j}, \text{rand} > 0.5, \\
1, & Z_{k,j} \neq Z_{\mathbf{m},j}, \text{rand} \leq 0.5
\end{cases}
\]

(c) Foraging behavior of fish: the existing fish state is \( X = \{Z_k\} \), and the random behavior performed by \( d_{ij} \) visual within the visible range starts to select the state of the data \( X_j = \{Z_j\} \).

If \( F > F_{\mathbf{p}} \), you need to randomly regenerate their state \( X_{(k+1)} \) for the fish and then finally perform the experiment of calculating the food concentration.

**Step 4.** When the result of \( \mathbf{g} > \mathbf{a} \), the optimal artificial fish appears, it means the end.

3.3. Application of Random Forest Algorithm in the Evaluation of College Students’ Talent. The process of talent quality evaluation is to select talents with the best comprehensive quality based on the various indicators of student performance in school, which can be regarded as a classification problem of unbalanced data sets. If you do not consider the balance of the indicators to directly model the original data, it is difficult to get an ideal model. You may need to use training data to improve the imbalance rate. The main way to achieve this is to measure the importance of data indicators through feature importance metrics. Using this as the index weighting standard, this paper chooses the Ranking method to measure the importance of the index points; on the other hand, because the random forest algorithm uses an average voting mechanism for the classifier, this voting mechanism makes the weak classifier to the final results of the evaluation have an impact. In this paper, the F-measure algorithm is used to weight the classifiers to reduce the impact of weak classifiers on the results.

In order to solve the above problems, this paper uses the importance of each sample feature as the basis and selects the sample features with higher importance to reduce the possibility of weak classifier generation. There are many ways to measure the importance of features. This article uses a random forest sorting algorithm to calculate the importance of features. The following are the main steps of the method:

**Step 1.** Select a sample feature \( X \), randomly introduce noise data, and calculate OOB again. The result is recorded as \( \text{errOOB}_2 \), and the initial OOB calculation result is recorded as \( \text{errOOB}_1 \). When there are \( N \) decision trees in the random forest, the formula for feature \( X \) is as follows:

\[
I_x = \frac{\sum_{i=1}^{N} (\text{errOOB}_{2,i} - \text{errOOB}_{1,i})}{N}
\]
Step 2. Follow the ordered features obtained in Step 1, select 75% of the features, and remove the next 20% of the features from the feature set.

Step 3. Repeat the above two steps until the number of features drops to $M$, a value set in advance. Finally, $m$ final feature sets are obtained.

The traditional random forest method uses the average majority voting method when making classification decisions. Each decision tree outputs its own classification label, and the final result is the class with the most output. But in the classification process, the classification effect of the decision tree is different, and the poorly effective classifier has a negative impact on the result.

Recall $\text{Recall} = \frac{TP}{TP + FN}$

$\text{ACC} = \frac{(TP + TN)}{TP + FP + FN + TN}$

Among them, TP represents the actual number of high-quality graduates predicted to be high-quality graduates, and TN represents the actual number of low-quality graduates predicted to be low-quality graduates. FP represents the actual low-quality graduates predicted to be high-quality graduates, and FN represents the actual high-quality graduates predicted to be low-quality graduates.

According to the F-measure calculation formula, the F-measure value of each decision tree that forms the random forest classifier is calculated.

$F\text{-measure} = \frac{2 \times \text{recall} \times \text{precision}}{\text{recall} + \text{precision}}$ (19)

In the above formula, recall represents the recall rate, and precision represents the accuracy rate. First, input the data of the verification set into each decision tree; then, each decision tree will have a category prediction for each record in the verification set; and compare the prediction results of the decision tree with the real results.

The improved random forest algorithm reduces the impact of the average voting mechanism, reduces the impact of weak classifiers on the results, and improves the overall performance of the algorithm. It can be applied in talent quality evaluation or in other applications.

It can be seen that although the convergence speed, accuracy, and local optimization ability of the ASSABC algorithm on some test functions have been improved, the results on complex multimodal functions are not very satisfactory, and this is likely to be due to insufficient mining capacity and step length variables. The sensitivity is not enough to fall into the local optimum. The test results are shown in Figures 4 and 5.

The data source of this article is mainly the student data collected by a university’s information college during the national engineering professional certification process. The data is composed of college archives, questionnaire surveys, comprehensive evaluation results, etc., including detailed data of more than 2000 graduates from 2011 to 2020, each of which contains about 35 fields, a total of 80,000 data.

According to the needs of talent training quality evaluation, we only need to select data closely related to the evaluation content to establish a database. In the end, only the fields shown in Table 3 are retained.

Among them, the sample data for a total of 8 years from 2011 to 2018 is used as the original training set, accounting for 80% of the total number of samples, and the sample data of two years from 2019 to 2020 are used as the test set.

This article has made two improvements to the RF algorithm. In order to verify that both improvements can have a positive impact on the evaluation results, this article verifies the two improvements separately to prove their effectiveness. In order to verify the enhancement of algorithm performance by feature selection, this paper compares the RF algorithm without feature importance weighting to the feature importance weighting algorithm. The results are shown in Table 4.

In order to verify the influence of the F-measure weighting algorithm on the performance of the algorithm, the RF algorithm of the ordinary voting mechanism is compared to the RF algorithm with the F-measure weighted voting mechanism. The results are shown in Table 5.

It can be seen from Table 6 that the random forest algorithm model based on the F-measure weighted voting mechanism has higher performance than the traditional one.
In order to prove the role of FWRF algorithm in the evaluation of college students’ graduation quality, this paper makes a horizontal comparison between several classical improved RF algorithms and FWRF algorithm. After the final screening, this paper selects the hybrid particle swarm random forest algorithm, hybrid genetic random forest algorithm, hybrid fish school random forest algorithm, and original random forest algorithm as the comparison algorithms. The performance evaluation index is mainly accuracy. The Python language is used in PyCharm. The above 4 algorithms are implemented using the scikit-learn library on the development platform. The experiment uses a tenfold cross-validation method to analyze the sample set and evaluates the classification results based on accuracy, recall, and F1 value. Table 7 shows the experimental results of the five methods.

It can be seen from Table 7 that compared with several classical modified random forest algorithms, the improved random forest algorithm proposed in this paper has little difference in accuracy and recall rate when used in the evaluation of talent training, but it is accurate. The rate has been improved to a certain extent, which meets the design requirements.

4. Optimization of College Students’ Entrepreneurship Education in the Era of Artificial Intelligence

4.1. Research Content of the Employment Ability Training Model of College Students in the Era of Artificial Intelligence

The education of traditional colleges and universities is mostly limited to the theoretical level, which leads to the fact that college graduates cannot keep up with the practical needs of the workplace after graduation. Many companies may need to spend a long time to cultivate talents that meet their own business practice requirements. Therefore, theoretical teaching in Colleges and universities needs to be combined with practical teaching. For example, innovative colleges and universities such as cash apprenticeship and...
order classes are combined with enterprise teaching to officially try to solve the above problems and innovative attempts. When students are in school, they can be given company-based teaching or go to a company for practical learning, which can not only meet the requirements of improving the ability of college students in the artificial intelligence era, but also improve the learning efficiency of students.

Generally, college students have limited awareness and have less exposure to new artificial intelligence technologies. In particular, many vocational and technical college students are limited to employment in traditional industries. Therefore, schools should stimulate students' innovative potential. Every year, campus culture and art festivals and science and technology festivals are held to encourage students to participate in professional skill competitions, to study and ponder on their counterparts, and to cultivate the spirit and perseverance of learning, loving, and working hard. Innovation leads ideas; technology revolutionizes the world.

4.2. Innovativeness of the Employment Ability Training Model for College Students in the Era of Artificial Intelligence. After college students acquire entrepreneurial knowledge, companies can also sign some contracts with the colleges and universities the students belong to, to provide more outstanding graduates with more job opportunities. It is possible to further consolidate and apply knowledge in the practice base of undergraduate entrepreneurship education. And college students can also make full use of the school's equipment and facilities to make a judgment on their own abilities and then gradually enhance their own comprehensive quality.

Pay attention to the combination of theory and practice. At present, many colleges and universities still use the past teaching materials, so they will face the problem of gaps between the teaching content and materials and the cutting-edge technology of the society. Many college teachers have a single teaching method. Therefore, colleges and universities should increase the ability of masters, promote the reform of teaching courses, establish professional counterpart training rooms, and focus on the combination of professional practical operation and extracurricular research.

The current national employment and entrepreneurship policies run through it. By publicizing the deeds of party members in the new era and the deeds of outstanding students close to students' lives in the school, postgraduate entrance examination policies, college students' entrepreneurship, etc., college students can understand the current policies related to graduates and make their own college career and career planning.

4.3. Research on Innovation and Entrepreneurship Education and University Students' Employability. Employability is the embodiment of college students' comprehensive ability, and its level directly determines the employment quality and work performance of college students. Therefore, this article constructs a model to examine the influence of innovation and entrepreneurship education model on college students' employability and entrepreneurship. Through analysis, it is found that the current level of employability of college students is average. The level of teamwork, communication, and coordination skills is relatively high, while the level of professional innovation ability is relatively low. On the other hand, it reflects innovation and entrepreneurship courses, innovation and entrepreneurship simulation training. The education model is closely related to the employability and various dimensions of college students. Therefore, in order for innovation and entrepreneurship education to better play its role in promoting the employability of college students, the following suggestions are put forward:

1. In order to improve the educational effect of innovation and entrepreneurship courses, colleges and universities can design innovation and entrepreneurship curriculum systems hierarchically and categorically according to the different stages of student learning and individual differences. Sophomore and junior students mainly study professional courses. At this stage, it is necessary to solve the problem of the integration of innovation and entrepreneurship courses and professional courses. Students should be instructed in how to apply the entrepreneurial awareness, skills, and spirit acquired in entrepreneurship education courses to their professional studies so as to obtain good professional knowledge understanding and absorption ability. The senior year is often a period when students are preparing to start a business or get a job. Therefore, special entrepreneurial intensification courses can be provided for students with entrepreneurial inclination and a clear career direction.

2. Give full play to the leading role of entrepreneurship lectures. Colleges and universities can also invite entrepreneurs, especially successful alumni, to come to the school to conduct innovation, entrepreneurship, and employment seminars; combine their own professional experience and professional vision; and share innovation and entrepreneurship experience and employment status in the professional field, so that students can understand the development of society and enterprises. The requirements for the quality and ability of college students provide a direction for the development of high-quality employment for students.

3. Colleges and universities should establish complete competition guarantees and incentive measures to organically integrate the participation in entrepreneurial competitions, evaluation, employment recommendation, and other matters closely related to college students' academic and employment, so as to increase students' enthusiasm for participation and improve project quality. For outstanding entrepreneurial competition project results, school incubators or cooperation with off-campus enterprises and
investment institutions can be used to promote the market transfer of projects.

(4) For students who intend to start a business or entrepreneurial projects with development potential, the school can assist in the formation of a cross-disciplinary and cross-professional entrepreneurial team and the establishment of the innovation and entrepreneurship practice base.

(5) Universities should strengthen laboratory construction, such as business simulation laboratory, consumer cognition laboratory, and business negotiation laboratory, and ensure the class hours of experimental courses. Through playing a role in the simulation experiment operation, students can experience the business process of the enterprise, feel the nature of different jobs, and perceive the correct management mode. At the same time, the school also needs to strengthen school-enterprise cooperation. Students can regularly go to the off-campus training base to receive entrepreneurship training and directly participate in the production and operation activities of the enterprise.

5. Conclusion

Nowadays, data mining algorithms are widely used. After studying several classification algorithms (LR, NB, SVM, GRNN, and RF), this paper focuses on the RF algorithm. For unbalanced data processing, we propose the CURE-SMOTE algorithm, which selects representative data points to prevent noise data, and artificially generated samples follow the original distribution. Under this influence and under the leadership of innovation driven in the era of artificial intelligence, college students have been integrated with career planning through education and services, and each region combines local economic characteristics, industrial upgrading, and integration, changing the traditional cheap labor model and optimizing the local economy. Against the background of the new era, college students need to give full play to their value as knowledge-based talents to give their country their strength and contribute their value. In the new era, college students also need to innovate themselves, strengthen their comprehensive quality capabilities, adapt to the changing times, and realize the reform and innovation of political, economic, and social talents in the era of artificial intelligence. At the same time, colleges and universities need to optimize education in terms of undergraduate entrepreneurship. Undergraduate entrepreneurship education and ideological and political education are based on the nature of different jobs, and perceive the correct management mode. At the same time, the school also needs to strengthen school-enterprise cooperation. Students can regularly go to the off-campus training base to receive entrepreneurship training and directly participate in the production and operation activities of the enterprise.

Data Availability

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

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the Teaching Reform fund XJT [2019] No. 291: Research Projects of Teaching Reform in Colleges and Universities of Hunan Province in 2019, “Design and Practice of the Flipped Classroom for Guidance of Career Development, Innovation and Entrepreneurship for College Students’ Based on MOOC” (Fund Project No. 1024).

References


