International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18 (2026) pp. 1-15

https://doi.org/10.70917/ijcisim-2026-0056
Article

Research on Analyzing the Competitive Situation of College Students' Innovation and Entrepreneurship Market Based on the Algorithm of Multilayer Perceptual Machine

Juan Chen [1] and Xiaofei Li [2,*]

[1] College of General Education, Guangxi Industrial Vocational and Technical College, Nanning 530000, Guangxi, China
[2] School of Sports Economics and Management, Guangxi University of Finance and Economics, Nanning 530000, Guangxi, China
* Correspondence author: lll1230607@126.com

Abstract: In recent years, along with more and more employment problems, the state has formulated a series of policies to support college students' innovation and entrepreneurship, and all colleges and universities are actively exploring the competitive market dynamics. This paper starts from the current situation of college students' innovation and entrepreneurship to understand its main influencing factors. The traditional TF-IDF algorithm is then used to extract text features, and it is found that there are defects in the algorithm, for which the importance value is introduced on the basis of the original algorithm. The extracted text feature vector of college students' innovation and entrepreneurship market competition situation is used as the input layer of the multilayer perceptron algorithm, and the output layer is the prediction value of the market competition situation, and the BP neural algorithm is used for training, which finally completes the task of constructing the prediction model of the college students' innovation and entrepreneurship market competition situation, and empirically analyzes the model. In terms of AUC area, the performance of BP algorithm (index difference: 0.1052) is better than that of random forest algorithm (index difference: 0.1194), which proves that this paper's algorithm is prioritized in the analysis of competitive situation of college students' innovative entrepreneurship market. In addition, this paper's algorithm detects that the number of employment in the tertiary industry increases by 1,479,600 people every year, and the corresponding number of employment in the primary industry and the secondary industry shows a decreasing trend, which quantitatively reflects the current competitive situation in the market of college students' innovation and entrepreneurship.

Keywords: BP neural algorithm; multilayer perceptron algorithm; TF-IDF algorithm; market competition situation

1. Introduction

At present, cloud computing, big data, artificial intelligence and other advanced information technology is gradually realized with all walks of life mutual integration and interoperability, the development mode of Internet technology to help various industries to realize the development of new modes and new business models. For the innovation and entrepreneurship of college students, it is necessary to break through the traditional entrepreneurial thinking mode, and fully recognize the great impact of the combination of the Internet and traditional industries on industrial upgrading [1-2]. “Internet +” provides opportunities for college students' innovation and entrepreneurship, but also brings challenges. On the one hand, the industry fields and jobs for innovation and entrepreneurship in the Internet era have greatly increased, and the entrepreneurial environment of the whole society is getting better and better [3-4]. However, because the old business model and management concepts of traditional

industries are no longer suitable for the development requirements of the “Internet +” era, and their entrepreneurial space is squeezed by the Internet model, this actually puts forward higher requirements for college students to search for and determine the innovative entrepreneurial industries and projects [5-8].

On the other hand, the development of Internet information technology makes the market competition more intense, some traditional industries have declined or even disappeared in the fierce market competition, and a large number of newly founded enterprises are also struggling in the fierce market competition [9-11]. The entrepreneurship of college students in the school stage is constrained by factors such as capital, technology and management experience, and most of them choose service-oriented projects or small-cost projects, and most of the entrepreneurial projects take school students as the target customers [12-13]. Once the entrepreneurial leaders graduate from university and leave the school, an environment with relatively weak market competition, it is still questionable whether their entrepreneurial projects can gain a foothold in the competitive society [14-15]. Based on this, it is necessary to establish an analytical model of the competitive market situation to create favorable conditions for entrepreneurship of college students during school or after graduation.

In this paper, through the theoretical analysis of the status quo of college students' innovation and entrepreneurship, the main influencing factors of college students' innovation and entrepreneurship market competition dynamics are drawn out. In order to obtain the text feature data, the traditional TF-IDF algorithm is used to extract the text features, and it is found that there are defects in the algorithm, and in view of this problem, the traditional TF-IDF algorithm is optimized by adding the importance value. The acquired text features are used as the input of the multilayer perceptual machine algorithm, while the predicted value is the output, and the BP neural network algorithm is used to iteratively train the input, and finally design a prediction model for the competitive situation of college students' innovation and entrepreneurship market. With the support of research data, the model of this paper is validated and analyzed, and the analysis of the application effect of the prediction model of this paper is also added to make the research structure of this paper more rigorous.

2. An Exploration of the Dynamics of Competition in the Innovation and Entrepreneurship Market for College Students

2.1. Challenges of Innovation and Entrepreneurship for University Students

As we all know, the Internet is the foundation of the current information society and has become an indispensable infrastructure for social life. “Internet+” refers to the new development mode arising from the fusion of the Internet and traditional industries by using information and communication technologies and network platforms, which is characterized by cross-border integration, innovation-driven, reshaping structure, respecting human nature, open ecology and connecting everything, and also brings a lot of challenges for college students' innovation and entrepreneurship.

2.1.1. Inherent Weaknesses in Experience and Capacity

From the viewpoint of our current education path, most of the college students receive knowledge in school from elementary school to university and lack the experience of business activities. In society, business activities are extremely complex, especially in the real business environment in the age of the Internet, students with limited knowledge and experience in entrepreneurial activities often seem to be overwhelmed, and need to catch up on all aspects of knowledge.

2.1.2. Unclear Entrepreneurial Goals of University Students

In the face of innovation and entrepreneurship activities, some college students, driven by the consciousness of following the trend, see that their classmates are starting their own business, so they also join the entrepreneurial army, but without a clear entrepreneurial goal. Moreover, although some college students have mastered rich professional knowledge in their long-term school life, they have insufficient business management concepts and no scientific awareness of entrepreneurial risk. Thus, in the innovative entrepreneurial activities, there is no actual investigation of the market, the development of the market dynamics, product marketing methods, reasonable product prices and other aspects are not understood, resulting in the process of entrepreneurship has taken a lot of detours, and even in the face of difficulties to give up, entrepreneurial failure abounds.

2.1.3. Relative Shortage of Resources for University Student Entrepreneurship

Financial support is the key to entrepreneurship, without financial support everything is on paper. However, for college entrepreneurs, the funds that can be used to start a business are very small, and

college students' entrepreneurial funds mainly come from the following ways: self-financing from relatives and friends. Self-financing from relatives and friends is the main choice for college students to start their own business. According to the data, 75.3% of undergraduate entrepreneurial funds come from personal and family funds, which is almost impossible for students with poor families or in general conditions to start their own business. Borrow from the bank, but considering the security, the bank is not willing to lend money to these fledgling college students. Financial support from corporate investors. For enterprise investors, they pay more attention to the benefits, college students have not yet been to the social experience, the lack of a comprehensive understanding of the market economy, there will be a greater risk of investment.

2.2. Textual Characterization of Competitive Market Dynamics

2.2.1. Traditional TF-IDF Feature Extraction Method

Feature extraction is one of the most common and effective methods to reduce the dimensionality of the feature space, the feature extraction methods are divided into many kinds according to the different feature scoring functions, TF-IDF is one of the most common feature extraction methods and the extraction effect is better compared with the other methods [16-17].TF-IDF is usually used to measure the importance of a word or phrase in a text set for a text that contains the word or phrase. TF-IDF is actually the product of TF and IDF, and the feature extraction function of TF-IDF is:

$$f(w) = TF(w) \cdot IDF(w) = TF(w) \cdot \log \frac{N}{n(w)+1} \tag{1}$$

Feature term frequency TF is the quotient of the number of occurrences of a feature term in a text to the total number of occurrences of all feature terms in the text. A feature term can be a word or a phrase.The main idea of TF is that if a feature term occurs more times in a text, it indicates that the feature term may describe the main information of the text better and is suitable for classification.

If the number of texts in a text set containing feature word w is less, it means that the inter-feature w category differentiation is better.IDF can weaken the importance of feature words that are contained in a large number of texts, or strengthen the importance of feature words that are contained in only a small number of texts.The commonly used formula for IDF is as in Equation (2). For:

$$IDF(w) = \log \frac{N}{n(w)+1} \tag{2}$$

where N is the total number of texts and $n(w)$ is the number of texts containing w. The traditional TF-IDF feature extraction method uses formula (1) to calculate the TF-IDF weight value of each feature word in the text and descending order, and then screens the first n feature words that satisfy the requirements according to the pre-set filtering conditions, thus realizing the dimensionality reduction of the original feature space.

2.2.2. Improved TF-IDF Feature Extraction Method

It is found that the traditional TF-IDF feature extraction method still has the following two shortcomings: (1) TF only considers the information contained in the text in terms of word frequency, lacks the consideration of the contextual environment of the feature words, and ignores the text structure information. (2) The main idea of IDF is that in a text set with a small number of texts containing the feature word ω, the larger its IDF value is, the stronger the category differentiation of the feature word ω is. This paper makes corresponding improvements for these shortcomings. Finally, this paper obtains the improved TF-IDF feature extraction function as in Equation (3). For:

$$f(w) = TF(w)o(1+PR(w)) \cdot IDF_{\text{change}}(w) \tag{3}$$

where $TF(w)$ is the feature term frequency value of feature word w, $PR(v)$ is the PR value of feature word w in its text network, also known as the importance value, and $IDF_{\text{change}}(w)$ is the inverse document frequency value of feature word w computed by utilizing the improved IDF computation method.

$$PR(p_i)=\frac{1-d}{N}+d\sum_{p_j=M(p_i)}\frac{w_{ji}\cdot PR(p_j)}{\sum_{i=Q(p_i)}w_{ji}} \tag{4}$$

where N is the total number of representative nodes, w_{ji} is the weight of the directed edge between node p_j to node p_i, $M(p_i)$ is the set of nodes containing directed edges pointing to node p_i, $Q(p_j)$ is the set of adjacent nodes to node p_j the set of neighboring nodes, and w_{jk} is the weight of the directed edges between node p_j to node p_i.

After we calculate the improved TF-IDF weight value of each feature word in the original feature space using equation (3), the steps of feature extraction are roughly the same as those based on the traditional TF-IDF feature extraction method, by calculating and descending the weight value of the feature words, and then filtering out the first n feature words that satisfy the requirements of their workflows according to the pre-set filtering conditions, so as to downsize the original feature space. The original feature space is downsized.

2.3. Predictive Modeling of Competitive Market Dynamics

2.3.1. Perceptual Machine (MLP) Neural Networks

The significance represented by the single-layer perceptron as the first artificial neural network is significant. However, its disadvantages are also very obvious, the network structure is too simple to solve nonlinear problems. For this reason, multilayer perceptron was proposed. A multilayer perceptron is a neural network that introduces one or more hidden layers on top of a single-layer neural network, so that the neural network has more than one network layer, and is therefore called a multilayer perceptron [18-19]. In this case, the hidden layer is located between the input layer and the output layer.

Given a small batch of samples $X\in R^{n\times d}$ with batch size n and number of inputs d. Let the multilayer perceptron machine have only one hidden layer, where the number of hidden units is h. Remembering that the output of the hidden layer is H, we have $H\in R^{n\times h}$. Because both the hidden layer and the output layer are fully connected layers, the weight parameter and the deviation parameter of the hidden layer can be set to be respectively:

$$W_h\in R^{d\times h},\quad b_h\in R^{1\times h} \tag{5}$$

The weight parameter and bias parameter of the output layer are respectively:

$$W_0\in R^{d\times q},\quad b_0\in R^{1\times q} \tag{6}$$

A design of a multilayer perceptual machine containing a single hidden layer. Its output $0\in R^{n\times q}$ is computed as:

$$\begin{aligned} H&=XW_h+b_h\\ O&=HW_0+b_0 \end{aligned} \tag{7}$$

The output of the hidden layer is directly used as the input of the output layer. If Eq. (6) and Eq. (7) are associated, it can be obtained:

$$O=(XW_h+b_h)W_0=XW_hW_0+b_hW_0+b_0 \tag{8}$$

It can be seen from Eq. (8) after association that although a hidden layer is introduced into the neural network, it is still equivalent to a single-layer neural network: where the weight parameter of the output layer is W_hW_0 and the bias parameter is $b_hW_0+b_0$. Therefore, the above design, even if more hidden layers are added, can only be equated to a single-layer neural network containing only output layers.

2.3.2. Activation Functions

The problem of introducing a neural network with hidden layers can be equated to that of a single-layer neural network containing only output layers. A fully connected layer only affine transforms

the data, and the superposition of multiple affine transformations is still an affine transformation. One way to solve the problem is to introduce an affine transformation, where the hidden variables are transformed using an element-wise affine function, which is then used as the input to the next fully connected layer. The nonlinear function is called an activation function. The most commonly used activation functions are the linear rectifier unit (ReLU), the Tanh (hyperbolic tangent) function, etc. In this paper, the ReLU function is used. The most commonly used activation function is the ReLU function because of the simplicity of its implementation as well as its good performance in various prediction tasks, ReLU provides a very simple nonlinear transformation. Given an element x, the ReLU function is defined as the maximum value of that element with respect to zero:

$$\mathrm{Re}\,LU\left(x\right)=\max\left(x,0\right) \tag{9}$$

2.3.3. Error Back Propagation Algorithm

In a multilayer perceptron, the input data is fed from the input layer, passed through the hidden layer and finally output from the output layer. The most commonly used training method for the connection weights between multilayer networks is the error back propagation algorithm (BP algorithm). w_{1ij} denotes the connection weights between the input layer and the middle layer, and w_{2ik} denotes the connection weights between the middle layer and the input layer. i denotes the input layer units, j denotes the units in the middle layer, and k denotes the units in the output layer. The error function E of the multilayer perceptron is equal to the sum of the errors of multiple output units:

$$E=\frac{1}{2}\sum_{j=1}^{q}(r_j-y_j)^2 \tag{10}$$

Consider first the adjustment of the connection weight w_{zjk} between the output layer and the intermediate layer. Take the activation function as a ReLU function, for example, and set $y=f(u)$, then $\{j=y(1-y)$. Derive the error function E for the weights w_{2jk}:

$$\frac{\partial E}{\partial w_{2jk}}=\frac{\partial E}{\partial y_k}\frac{\partial y_k}{\partial u_{2k}}\frac{\partial u_{2k}}{\partial w_{2jk}} \tag{11}$$

Among them:

$$\frac{\partial E}{\partial y_k}=-(r_k-y_k) \tag{12}$$

$$\frac{\partial y_k}{\partial u_{2k}}=y_k(1-y_k) \tag{13}$$

$$\frac{\partial u_{2k}}{\partial w_{2jk}}=z_j \tag{14}$$

After the derivation of the output y_k via the error function E in Eq. (10), the output y_k in Eq. (11) on the activation value u_{2k}, and the activation value u_{zk} on the connection weights w_{zjk} in Eq. (12), substituting into Eq. (9) yields the following equations:

$$\frac{\partial E}{\partial w_{2jk}}=-(r_k-y_k)y_k(1-y_k)z_j \tag{15}$$

The partial derivatives of the connection weights w_{1ij} between the input and intermediate layers:

$$\frac{\partial E}{w_{1ij}} = \sum_{k=1}^{q}\left[\frac{\partial E}{\partial y_k}\frac{\partial y_k}{\partial u_{2k}}\frac{\partial u_{2k}}{\partial w_{1ij}}\right] \tag{16}$$

The unit j in the middle layer is connected to all the units in the output layer, so as shown in equation (16), when the error function E is calculated by taking the partial derivative of the connection weight w_{1ij}, the result is equivalent to weighting and summing the derivatives of all the output units, and in fact the synthesis of the connection weights of all the output units is used. Substituting the derivatives of the sigmoid function and the error function into the above equation yields:

$$\frac{\partial E}{w_{1ij}} = -\sum_{k=1}^{q}\left[(r_k - y_k)y_k(1-y_k)\frac{\partial u_{2k}}{\partial w_{1ij}}\right] \tag{17}$$

Since the role played by the connection weights w_{1ij} only affects the state of the middle layer z_j, the remaining part of Eq. (15) is calculated as follows after derivation:

$$\frac{\partial u_{2k}}{\partial w_{1ij}} = \frac{\partial u_{2k}}{\partial z_j}\frac{\partial z_j}{\partial w_{1ij}} \tag{18}$$

The activation value u_{2k} is derived from z_j to obtain the connection weight w_{2ijk}, and by combining it with the following equation, the adjusted value of the connection weight w_{1ij} between the intermediate layer and the input layer can be derived:

$$\frac{\partial z_j}{\partial w_{1ij}} = \frac{\partial z_j}{\partial u_{1j}}\frac{\partial u_{1j}}{\partial w_{1ij}} = z_j(1-z_j)x_i \tag{19}$$

$$\Delta w_{1ij} = \eta\sum_{k=1}^{q}\left[(r_k - y_k)y_k(1-y_k)w_{2jk}\right]z_j(1-z_j)x_i \tag{20}$$

where η is the learning rate, taking the value range of (0, 1]. As can be seen from the above equation, the learning rate η is the coefficient used to determine the degree of adjustment of the weight connection. In the training process of the neural network, if the learning rate is chosen too large, it may lead to overcorrection, the error can not converge, and the neural network is not well trained. On the contrary, if the learning rate is chosen too small, it will lead to slower convergence and longer training time. In most cases, we will first determine a larger value for training based on experience, and then slowly reduce this value according to the training effect.

2.3.4. Model Construction

The first layer is the input layer, which is the textual feature vector of the competitive situation of university innovation and entrepreneurship market extracted from the improved TF-IDF features, and the second and the third layers are the hidden layers, which are used to extract higher-level feature information from the basic features of the input layer. The fourth layer is the output layer, which is used to output the predicted value of the competitive situation of the university innovation and entrepreneurship market. In this case, the output of the neuron nodes in each layer is a function of the neuron nodes in the previous layer. In order to be able to predict the competitive situation of college students' innovative entrepreneurship market based on the constructed multilayer perceptual machine model, we first need to learn the multilayer perceptual machine model. The learning of the multilayer perceptron model is mainly to learn the connection weights between the nodes of each layer, and its learning process is usually realized by using the error back-propagation (BP) algorithm. That is, first set a small random number to the initial weights of the multilayer perceptron, and then input the training samples into the multilayer perceptron network, use the error back propagation (BP) algorithm based on the principle of stochastic gradient descent to train the network, and then adjust the network parameters to make the actual output value as close as possible to the expected output value of the multilayer perceptron model operation. In this paper, we have tried to make the calculated predicted value of the

competitive situation of the university innovation and entrepreneurship market as close as possible to the actual value of the competitive situation of the university innovation and entrepreneurship market. We use the mean square error as the loss function in the training process, and the loss function J(w) can be expressed as:

$$J(w)=\frac{1}{m}\sum_{i=1}^{m}(h_w(x^{(i)})-y_i)^2 \tag{21}$$

where: $h_w(x^{(i)})$ is the predicted value of the i th university innovation and entrepreneurship market competitive situation calculated using the multilayer perceptron model in the paper. y_i is the label of the i th actual text feature pre-labeled to indicate whether or not the text belongs to the competitive landscape of the university innovation and entrepreneurship market (if $y_i=1$, it means it belongs, and if $y_i=0$, it means it does not belong). m is the total number of text features.

With the BP algorithm, we can learn the parameters in the multilayer perceptron efficiently. Thus, in the testing stage, we can use the trained multilayer perceptron model to calculate the predicted values of the text features of the competitive situation of the university innovation and entrepreneurship market, and then realize the purpose of predicting and analyzing the competitive situation of the university innovation and entrepreneurship market.

3. Empirical Validation Analysis

3.1. Market Competitive Landscape Text Feature Extraction Analysis

3.1.1. Data Sources

The experimental data comes from the University Innovation and Entrepreneurship Platform - a textual dataset of the competitive market situation of university students' innovation and entrepreneurship provided by the platform, which contains 10 categories (the number of projects X1, the diversity of fields X2, the fierce competition in the industry X3, the serious homogenization X4, the uneven distribution of resources X5, the personal experience X5, the market awareness X6, related policies X7, innovation awareness X8, technology transformation X9, key role of universities X10), each article is saved in plain text format, including journal literature, magazine bibliography and other categories, in order to validate the effectiveness of the improved TF-IDF feature extraction method, the above 10 categories are used as the test text set.

3.1.2. Assessment of Indicators

Since the performance of general keyword extraction algorithms is evaluated by comparing them with a specific evaluation function, two metrics of text keyword extraction effectiveness-accuracy, recall, and F1 value-are used to measure the algorithm performance. Accuracy and recall are defined as:

$$P=a/b \tag{22}$$

$$R=a/c \tag{23}$$

$$F1=\frac{2*P*R}{(P+R)} \tag{24}$$

where P is the accuracy rate, R is the recall rate, a is the number of correctly extracted keywords, b is the number of extracted keywords, and c is the number of keywords of the text of the competitive market situation in the college students' innovation and entrepreneurship platform.

3.1.3. Analysis of Results

Comparison experiments were conducted using TF, TF-IDF, and the improved TF-IDF algorithm. The results of the experiments are shown in Figures 1 to 3. All the data in Figures 1 to 3 are the mean values of the calculations made for the text on the competitive landscape of the university innovation and entrepreneurship market. The results show that the accuracy of the TF algorithm is lower than that of the TF-IDF algorithm. It was analyzed that it was due to the fact that when calculating the accuracy rate, the TF algorithm misused the number of feature words in the text as the calculation standard, and the number

of feature words in the individual text was lower than 10, which led to the local TF accuracy rate being too low, and the error was corrected in the later experiments, and a more stable result was obtained. In the calculation of the TF-IDF algorithm recall, due to the calculation of the recall of 1 text less added, so the average recall of the TF-IDF algorithm is lower, the subsequent calculation corrected the error, the recall obtained by the calculation of 0.7232, basically stable, the same phenomenon exists in terms of the comparative analysis of the results of the F1 value. This paper's algorithm extracted keywords recognition accuracy and recall rate is significantly better than the traditional algorithm, resulting in more precise text keywords. Since the traditional TF and TF-IDF algorithms do not take into account the word length of the feature words, the error in the text recognition accuracy and recall rate is larger. Due to the frequent occurrence of abstract words such as “risk” and “enterprise” in the text of the whole category, they are ranked very low in the TF algorithm, resulting in low local TF values, while the improved TF-IDF algorithm filters these texts according to the proportion of word length. , which well excludes such words from irrelevant text features and produces relatively balanced accuracy, recall, and F1 values, further validating the desirability of the algorithm in this paper.

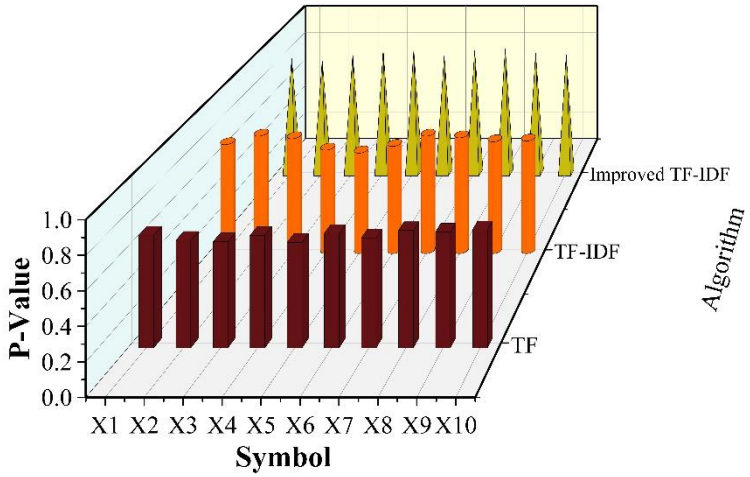

Figure 1. Comparative analysis of Precision rate.

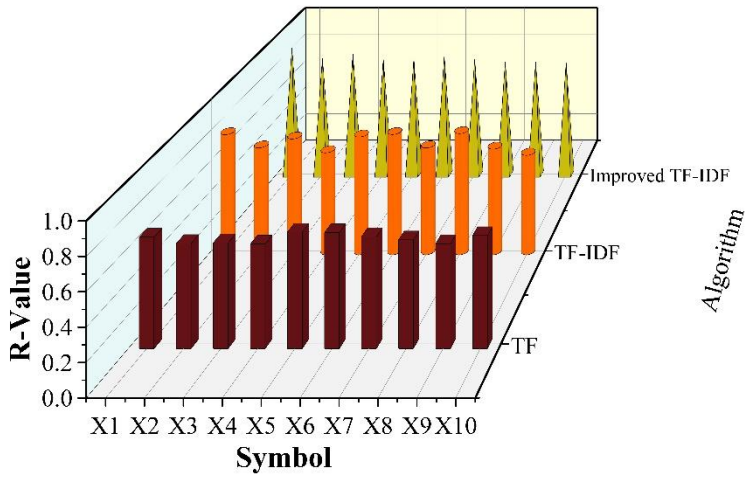

Figure 2. Comparative analysis of Recall rate.

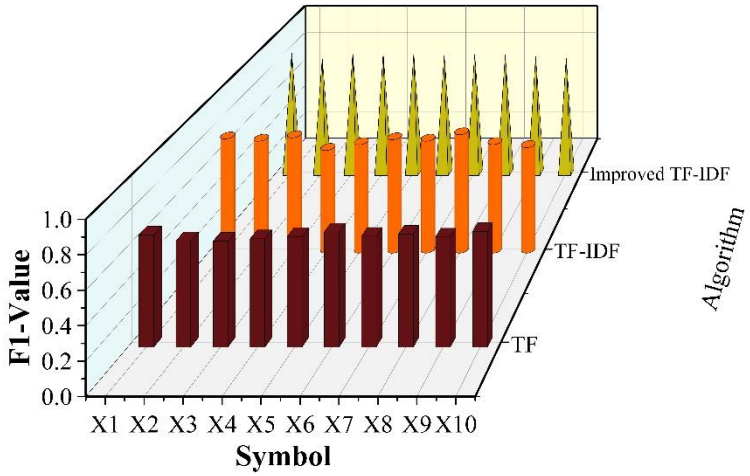

Figure 3. Comparative analysis of F1-Value rate.

3.2. Predictive Model Validation Analysis

3.2.1. Parameter Selection

There are important parameters such as excitation function, learning rate, initial weights, number of nodes in the hidden layer, error function, etc. in multilevel perceptron networks. In this paper, the ReLU function with unidirectional incremental and inverse function monomial incremental properties is used as the excitation function of the BP neural algorithm. To ensure stability, the learning efficiency is set between [0.005-0.1]. For the initial weights, they are set based on the product of the initial value and the largest of the absolute values of x being approximately 1. The number of nodes in the hidden layer is often set between 1.0 and 2.0 times the number of nodes in the input layer. The expected error is often set as a mean square error function (MSE).

3.2.2. Confusion Matrix

In this paper, the text feature dataset of 0.6, 0.5, and 0.4 is selected as the training set, and the rest of the samples are quizzed as the test set, and the computed sensitivity, specificity, false-positive rate, and false-negative rate are shown in Fig. 4, and we choose the better model among them for the visualization and analysis. The data realization in the figure reveals that the sensitivity of the 50% training set is improved by 0.00324 and 0.00474 compared to the 60% and 40% training sets, respectively. The specificity was improved by 0.00935 and 0.00374 over the 60% and 50% training sets, respectively.

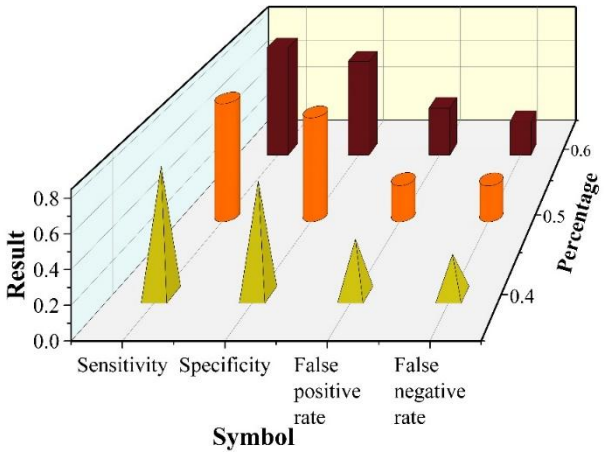

Figure 4. Model classification results of different training sets and sample sets.

Thus 50% of the training set is best in terms of model performance and the visualized confusion matrix is shown in Figure 5 below. The horizontal coordinates in the figure represent the labels of the neural network model's prediction of success or failure, True represents the prediction of success and False represents the prediction of failure. The vertical coordinate represents the label value of the

competitive situation of the university innovation and entrepreneurship market, True represents the actual success and False represents the actual failure. From the prediction results, it is known that out of 150 samples, the total number of samples predicted as success for positive examples is 98. The number of samples predicted to be successful positive examples of failure totaled 6. The total number of samples predicted to be successful for the negative case is 11. A total of 35 samples were predicted to be successful counter-examples. The overall prediction accuracy was 0.887 and the overall prediction error rate was 0.113.

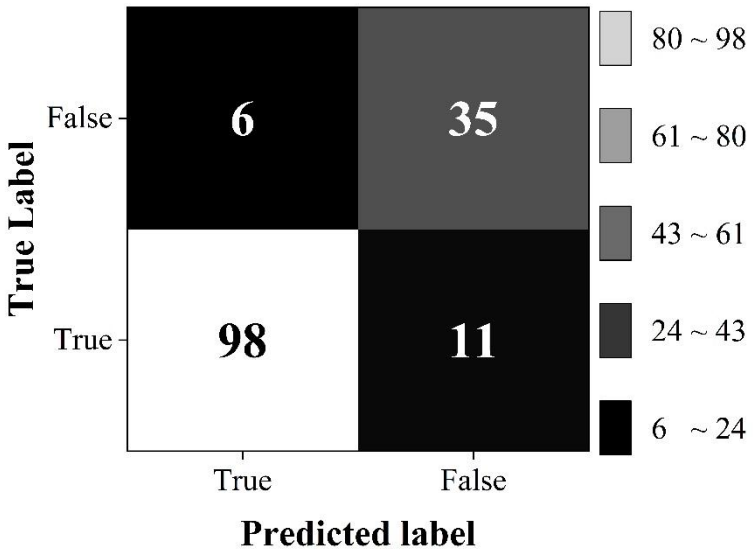

Figure 5. Confusion matrix.

3.2.3. ROC Curves

The ROC curves of the better models in different sample sets are selected for visualization and analysis as shown in Figure 6 below. The area under the curve is 0.887. Indicating that the prediction accuracy of this model is moderately high.

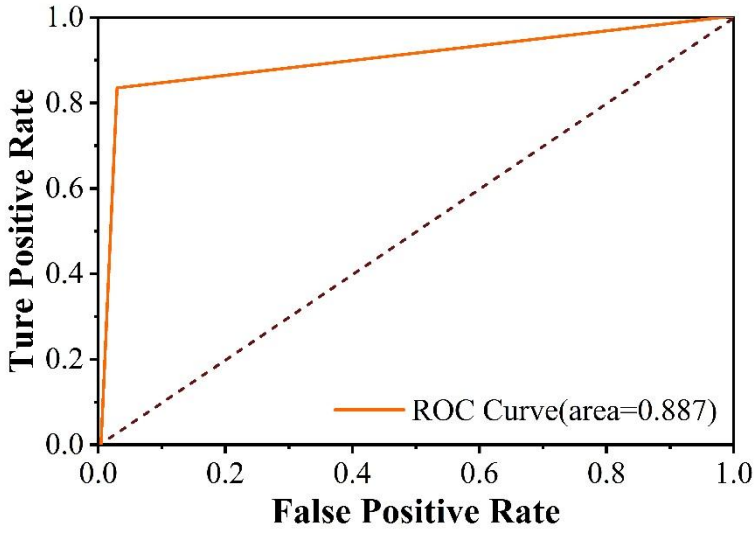

Figure 6. ROC curve.

3.2.4. Analysis of the Model's Discriminatory Power

The analysis of the discriminative ability of the models is shown in Figures 7 to 8, and the comparison reveals that the difference between the specificity and false positive rate of the two models on different

scaled training sets is small, but the difference between the sensitivity and false negative rate is large. The specificity of the BP neural algorithm on 60%, 50%, and 40% training sets decreased by 0.0334, 0.0316, and 0.0407 compared to the random forest algorithm, and the sensitivity of the BP neural algorithm on 60%, 50%, and 40% training sets increased by 0.0579, 0.0717, and 0.0687 compared to the random forest algorithm, respectively. It shows that the P-neural algorithm as a whole is more capable than the random forest algorithm in predicting the competitive dynamics of the university innovation and entrepreneurship market.

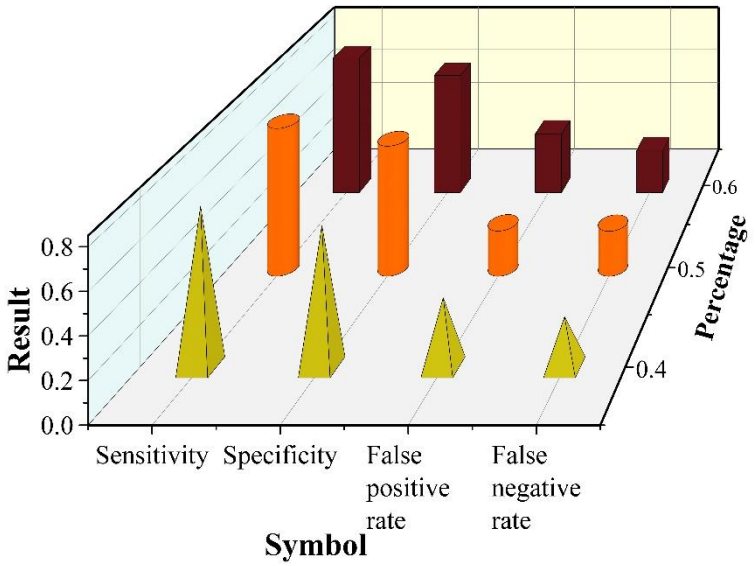

Figure 7. Analysis of the discriminative ability of the model(BP).

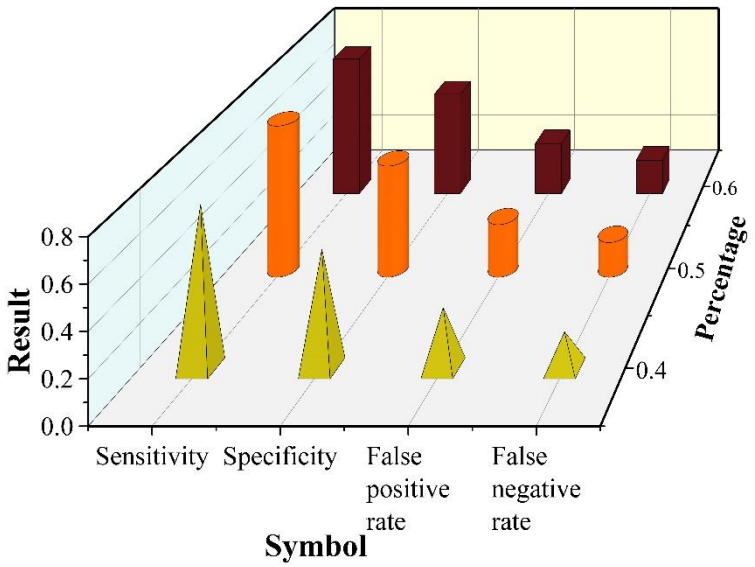

Figure 8. Analysis of the discriminative ability of the model(RF).

3.2.5. Analysis of the Model's Generalization Capacity

Generalization ability is also commonly used to assess the performance of classification prediction models. The size of the difference in test results between the model training set and the test set determines the strength of the generalization ability. If the gap between the misclassification rate, accuracy and AUC between the training set and the test set of the prediction model is larger, the weaker the generalization ability of the model is, and the less suitable the prediction model is to be applied and promoted on other sample datasets. On the contrary, if the gap is smaller, the stronger the model's generalization ability is, and the more suitable the prediction model is to be applied on other datasets. In

this paper, in order to determine the strength of the generalization ability, the predictive index results of the constructed prediction model are compared between the training set and the test set, in order to adjudicate the expected applicability of the two models to the competitive situation of the college students' innovation and entrepreneurship market, and the results of the model's analysis of the generalization ability are shown in Fig. 9~Fig. 10. As can be seen from the data in the figure, the BP algorithm prediction model is found to have strong generalization ability. In the comparison of overall misclassification rate and Accuracy, the difference between the training set and test set of BP algorithm is 0.0216, while that of Random Forest algorithm is 0.0547.In the comparison of AUC area, the difference between the training set and test set of BP algorithm is 0.1052, while that of Random Forest algorithm is 0.1194.This indicates that the difference between the results of the training set and the test set of the Random Forest algorithm is larger than that of BP algorithm, and therefore the generalization ability is weaker. In addition, the BP algorithm can ensure the accuracy of prediction on the basis of reducing the running time of the construction program, the need to adjust the parameters and the complexity of the operation, so as to streamline the amount of computing. In conclusion, for the sample data environment of this paper, the BP algorithm is more suitable for predicting and analyzing the competitive situation of college students' innovation and entrepreneurship market.

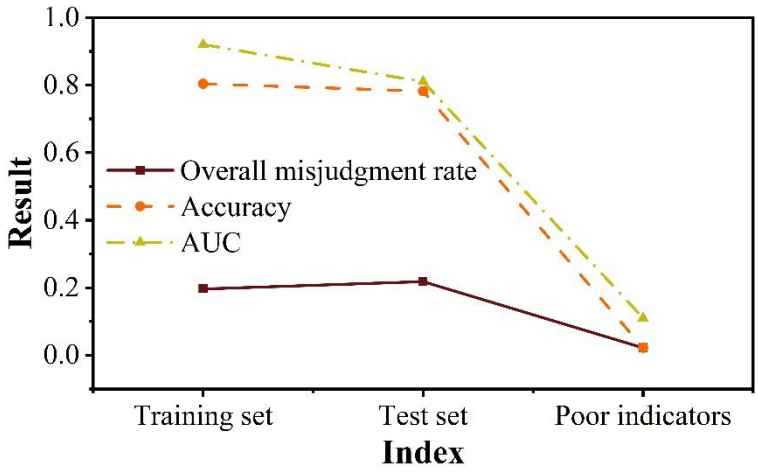

Figure 9. Analysis results of the generalization ability of the model(BP).

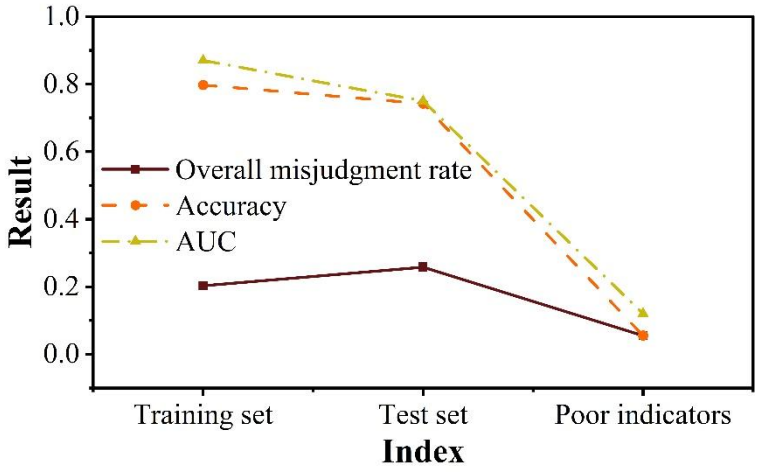

Figure 10. Analysis results of the generalization ability of the model(RF).

3.3. Forecast Analysis of the Competitive Landscape of the Innovation and Entrepreneurship Market

Based on the data of "employment population growth rate and employment population" in 2023, the model of this paper is used to predict the future competition of college students' innovation and

entrepreneurship market in Province A. The results of the employment population growth rate prediction are shown in Table 1, and the number of employment prediction is shown in Table 2. Comprehensive Table 1~Table 2 shows that there is an obvious downward trend in the number of employed people in the primary and secondary industries in Province A during 2023-2028, with an average decline of 2.247% in the employed population in the primary industry, and a slightly more obvious decline in the number of employed people in the secondary industry, down by 1.66%. There is a clear upward trend in the number of employed persons in the tertiary industry, with an average annual increase of 6.105%. This is mainly due to the flow of employed people within the industry caused by industrial upgrading, and the tertiary industry absorbs a large number of employed people from the primary and secondary industries, indicating that the flow of people within the industry leads to the formation of the competitive situation in the market of college students' innovation and entrepreneurship.There is a steady growth in the overall number of employed people in Province A from 2023 to 2028, with an average annual growth rate of 1.328%, and the forecast is 6586.5 percent in 2028. In 2028, it will be 65.865 million people, with an average annual rise of about 597,000 people in total employment. There is an overall downward trend in the number of people employed in the primary sector, with an average annual decline of 1.66%. It declines slightly in 2023, but rises in 2024 and 2025 and declines after 2026. The statistics show that the number of people employed in the primary industry was about 13,262,000 in 2023, and is forecast to be 11,941,000 in 2028, with an average annual decline of about 264,200 people. There is an overall declining trend in secondary industry employment, with an average annual decline of 2.48%. The number of employed people in the secondary industry will have a rapid downward trend after 2023, and the data statistics show that the number of employed people in the secondary industry will be about 25,144,000 in 2023, and the forecast will be about 22,052,000 in 2028, with an average annual decline of about 618,400 people. There exists a rapid upward trend in the number of employed in the tertiary industry in general, with an average annual increase of 5.038%. There exists a clear upward trend in the tertiary industry after 2023, and the data statistics show that the number of employed in the tertiary industry in 2023 is about 24,474,000 people, and it is predicted that it will be about 31,872,000 people in 2028, with an average annual rise of 1,479,600 people. In general, from 2023 to 2028, there is a clear downward trend in the number of employed people in primary and secondary industries in Province A, and the decline in the number of employed people in secondary industry is more obvious. There is a clear upward trend in the number of people employed in the tertiary industry. This is mainly due to the flow of people within industries caused by industrial transformation and upgrading, and the tertiary industry absorbs a large number of employed people from the primary and secondary industries.

Table 1. Prediction of the growth rate of the innovative and entrepreneurial population.

Year	Total employment growth rate/%	Employment growth rate in the primary industry/%	Employment growth rate in the secondary industry/%	Employment growth rate in the tertiary industry/%
2023	4.18	-3.341	-2.057	9.578
2024	3.006	1.116	-0.547	2.437
2025	1.455	0.128	-5.337	6.664
2026	1.068	-3.443	-2.207	6.718
2027	-0.626	-2.327	-3.886	5.587
2028	-0.814	-5.616	-0.845	5.647
Mean	1.378	-2.247	-2.480	6.105

Table 2. Forecast of employment numbers (Unit: Ten thousand people).

Year	Total number of employed people	The number of employed people in the primary industry	The number of employed people in the secondary industry	The number of employed people in the tertiary industry
2023	6288	1326.2	2514.4	2447.4
2024	6333	1338.2	2487.4	2507.4
2025	6369.7	1339.2	2354.3	2676.2
2026	6456.8	1296.1	2304.3	2856.4
2027	6491.3	1265.4	2209.4	3016.5
2028	6586.5	1194.1	2205.2	3187.2
Mean	6420.9	1293.2	2345.8	2781.9

4. Conclusion

Based on the theoretical analysis of college students' innovation and entrepreneurship dilemma, this paper adopts the traditional TF-IDF feature extraction algorithm to obtain the textual features of college students' innovation and entrepreneurship market competition situation, and finds that there are certain limitations, for this reason, the introduction of the value of the degree of importance to complete the design of the improved TF-IDF feature extraction algorithm. After that, based on the perspective of multilayer perceptual machine algorithm, the BP neural network algorithm is proposed to construct a prediction model of college students' innovation and entrepreneurship market competition situation, and the model is verified and analyzed. In terms of Accuracy, the BP algorithm (index difference: 0.0216) is better than the Random Forest algorithm (index difference: 0.0547), which indicates that the BP algorithm is more suitable for analyzing the competitive situation of college students' innovation and entrepreneurship market. In addition, the model also predicts that the number of employment in the tertiary industry will rise by 1,479,600 people every year during the period from 2023 to 2028, indicating that the mobility of industrial personnel and the transformation of the industry have led to the competitive posture of the college students' innovation and entrepreneurship market.

Acknowledgements

The special project of the "14th Five Year Plan" of Guangxi Education Science in 2022, "Research on the Influencing Factors and Countermeasures of the COVID-19 on the Entrepreneurship Behavior of Undergraduate Graduates in Guangxi" (project number: 2022ZJY2908); Special Project for Innovation and Entrepreneurship Education in Higher Education Institutions in Guangxi's 14th Five Year Plan for 2023, titled "Research on Cultivating Sports Innovation and Entrepreneurship Talents in Guangxi Universities from the Perspective of Strengthening the Awareness of the Chinese National Community" (Project No.: 2023ZJY1760).

References

1. Zhiwen, W., & Chuanchao, C. (2021, June). Innovation and Entrepreneurship Education Reform and Innovation Based on “Internet plus” Thinking. In 2nd International Conference on Language, Art and Cultural Exchange (ICLACE 2021) (pp. 231-236). Atlantis Press.
2. Zhang, W. Innovation and Entrepreneurship of College Students in The Era of Internet Plus Under the New Normal. International Journal of Education and Teaching Research, 10.
3. Qiao, W., Khatibi, A., & Tham, J. (2023). New Changes in the Development Opportunities of College Students Innovation and Entrepreneurship under the background of Internet+. Revista Ibérica de Sistemas e Tecnologias de Informação, (E62), 638-649.
4. Hu, B., & Luo, Q. (2018, July). College Students’ Opportunities and Challenges in Innovation and Entrepreneurship Based on “Internet Plus”. In IOP Conference Series: Materials Science and Engineering (Vol. 394, No. 3, p. 032069). IOP Publishing.
5. Butenko, N., Mykhaylovych, O., Bincheva, P., Lyndyuk, A., & Luchnikova, T. (2023). The role of internet marketing in the strategy of forming entrepreneurial activity. Economic affairs, 68(1s), 73-82.
6. Rybakova, E. V., & Nazarov, M. A. (2021). Entrepreneurship in digital era: prospects and features of development. Current achievements, challenges and digital chances of knowledge based economy, 105-112.
7. Hsieh, Y. J., & Wu, Y. J. (2019). Entrepreneurship through the platform strategy in the digital era: Insights and research opportunities. Computers in Human Behavior, 95, 315-323.
8. Tan, Y., & Li, X. (2022). The impact of internet on entrepreneurship. International Review of Economics & Finance, 77, 135-142.
9. Matusik, S. F. (2016). Entrepreneurship, competition, and economic development. The Antitrust Bulletin, 61(4), 561-563.
10. Block, J. H., Kohn, K., Miller, D., & Ullrich, K. (2015). Necessity entrepreneurship and competitive strategy. Small Business Economics, 44, 37-54.
11. Whalen, P., Uslay, C., Pascal, V. J., Omura, G., McAuley, A., Kasouf, C. J., ... & Deacon, J. (2016). Anatomy of competitive advantage: towards a contingency theory of entrepreneurial marketing. Journal of Strategic Marketing, 24(1), 5-19.
12. Guerrero, M., Urbano, D., & Fayolle, A. (2016). Entrepreneurial activity and regional competitiveness: evidence from European entrepreneurial universities. The Journal of Technology Transfer, 41(1), 105-131.
13. Jansen, S., Van De Zande, T., Brinkkemper, S., Stam, E., & Varma, V. (2015). How education, stimulation, and incubation encourage student entrepreneurship: Observations from MIT, IIIT, and Utrecht University. The International Journal of Management Education, 13(2), 170-181.
14. Bauman, A., & Lucy, C. (2021). Enhancing entrepreneurial education: Developing competencies for success. The International Journal of Management Education, 19(1), 100293.
15. Din, B. H., Anuar, A. R., & Usman, M. (2016). The effectiveness of the entrepreneurship education program in upgrading entrepreneurial skills among public university students. Procedia-Social and Behavioral Sciences, 224, 117-123.

16. Wei Yang,Weicong Tan,Zhijian Zeng,Ren Li,Jie Qin,Yuting Xie... & Dongliang Xiao. (2025). A Two-Stage Feature Extraction Approach for Green Energy Consumers in Retail Electricity Markets Using Clustering and TF–IDF Algorithms. Energy Engineering,122(5),1697-1713.
17. Vaishali Ingle & Sachin Deshmukh. (2017). Predictive mining for stock market based on live news TF-IDF features. Int. J. of Autonomic Computing,2(4),341-365.
18. Jayaraman Kumarappan,Elakkiya Rajasekar,Subramaniyaswamy Vairavasundaram,Ketan Kotecha & Ambarish Kulkarni. (2024). Federated Learning Enhanced MLP–LSTM Modeling in an Integrated Deep Learning Pipeline for Stock Market Prediction. International Journal of Computational Intelligence Systems,17(1),267-267.
19. Sonal Gupta,Deepankar Chakrabarty & Rupesh Kumar. (2023). Predicting Indian electricity exchange-traded market prices: SARIMA and MLP approach. OPEC Energy Review,47(4),271-286.