J. COMBIN. MATH. COMBIN. COMPUT. 127b (2025) 65--82

Journal of Combinatorial Mathematics and Combinatorial Computing
www.combinatorialpress.com/jcmcc

Combinatorial Press

Dynamic Analysis of College Students' Employment and Entrepreneurship Market Trends Based on Time Series Forecasting Models

Shang Sun[1,✉], Di Yang[2], Juan Hu[3]

[1] *School of Economics and Management, Anhui University of Science and Technology, Huainan, Anhui, 232001, China*

[2] *School of Electrical and Information Engineering, Anhui University of Science and Technology, Huainan, Anhui, 232001, China*

[3] *Huainan Vocational and Technical College, Huainan, Anhui, 232001, China*

ABSTRACT

With the deterioration of the global economic situation and the stagnation or regression of the development of enterprises, the problem of college students' employment and entrepreneurship has been particularly prominent in recent years, and it is also one of the key points that can not be ignored in carrying out economic construction. The article realizes the prediction of college students' entrepreneurship and employment market trends based on ARIMA-LSTM by designing the ARIMA algorithm model and combining it with the LSTM model architecture, taking the college students' entrepreneurship and employment data from 2010 to 2022 as the research data, and using two evaluation indexes, namely, the mean absolute percentage error (MAPE) and the root mean square error (RMSE), to predict the results. Evaluation. From the analysis results ARIMA model prediction fit is high. Comparing the prediction results of the combined model with those of the LSTM model and the ARIMA model, the comparison results show that the combined model constructed in this paper can effectively fit the linear and nonlinear intertwined and superimposed trends of the time series compared with a single model, and the relative error of prediction is smaller at 33.78, which makes the results more accurate. The combined model can help the management department related to college students' employment and entrepreneurship make reasonable decisions and improve efficiency.

Keywords: ARIMA algorithm; LSTM model; combined prediction; college employment and entrepreneurship

✉ Corresponding author.
E-mail address: sunshang24@163.com (S. Sun).

Received 07 January 2024; Revised 02 August 2024; Accepted 21 November 2024; Published Online 16 April 2025.
DOI: 10.61091/jcmcc127b-005

1. Introduction

At present, youth employment has become a common problem in China and even in the world, but they are also the main force of innovation and an important force for social development. Therefore, youth groups, especially college students, are a force that cannot be ignored in practicing the concept of “mass entrepreneurship and innovation” [1-2]. Entrepreneurship refers to the process in which a person discovers and captures an opportunity and creates a new product or service from it, and the main symbol and characteristic is the creation of a new enterprise or a new organization. The social concept in the past is that entrepreneurship is something done by a few social elites, and entrepreneurs should have resources and conditions that ordinary people do not have. But in fact, the history of the development of human society is a history of mass entrepreneurship, and many major scientific and technological inventions have been accomplished by ordinary skilled workers [3-5]. “Mass entrepreneurship” is intended to provide entrepreneurial opportunities for all entrepreneurial willingness, whether it is a highly educated scientific and technological workers, college students, or rural migrant workers can get the opportunity to start their own business [6].

Innovation is the recombination of production factors, the core of innovation is technological innovation, but also includes institutional mechanism innovation, management innovation and model innovation. Scientific invention and creation is innovation, the adoption of new products, the use of new production methods, the opening of new markets and so on also belong to the scope of innovation [7-8]. Innovation cannot be separated from talent support, innovation drive is ultimately talent drive, talent is the first resource for realizing innovation and development, and we should let talents with dreams and abilities give full play to their talents and potentials [9-10]. Youth groups are the main force of innovation and entrepreneurship, and college students are the pioneers of innovation and entrepreneurship. College students have a strong ability to accept new knowledge and new things, and a strong desire for knowledge and curiosity. They have not experienced too many bumps in life, often full of confidence in the future, and more daring to think and fight [11-12]. At the same time, they have mastered more systematic professional knowledge in their college studies and have leading scientific and technological advantages and entrepreneurial ideas. Mass entrepreneurship and innovation of all people support each other, entrepreneurship must be built on the basis of innovation, and only entrepreneurship that has experienced innovation is entrepreneurship in the true sense of the word [13-14].

College graduates are a dynamic and creative group, and also a valuable human resource in the society. At present, the sharp growth in the number of college graduates far exceeds the growth in the number of social jobs, and the employment situation of college students is becoming more and more serious, which has become one of the focuses of attention of the whole society [15-16]. In order to solve the problem of difficult employment of college students, the relevant departments have introduced a series of policies and measures in an attempt to create more jobs, while encouraging entrepreneurship to promote employment. Therefore, it is necessary to analyze the characteristics of college students' employment, and then put forward corresponding countermeasures to help college students better employment [17-18].

Literature [19] empirical examination revealed that students' entrepreneurial intention and entrepreneurial education effects showed differences in gender, entrepreneurial experience, and family background, where entrepreneurial education and entrepreneurial self-efficacy promoted entrepreneurial intention. Literature [20], based on related studies, clarified that entrepreneurial self-efficacy mediates the relationship between entrepreneurship education and entrepreneurial intention. Literature [21], based on data from the Entrepreneurial Influence Questionnaire Scale of the University of Kosovo, learned that individuals' attitudes and behavioral content towards entrepreneurship significantly influence entrepreneurial intentions. Literature [22] suggested that students' personality traits, self-efficacy, and entrepreneurial attitudes could be used to predict students' entrepreneurial behaviors and entrepreneurial intentions, whereas nationality and social identity did not have a significant predictive effect through empirical investigations. Literature [23] analyzed the demographics related to students' entrepreneurial intention and pointed out that males and students with entrepreneurial family background showed higher entrepreneurial intention. Literature [24] affirmed the positive role played by the theory of entrepreneurial personality traits in predicting the entrepreneurial behavior of students and corroborated it through an empirical approach. Literature [25] used the social cognitive approach of the Theory of Planned Behavior (TPB) and the motivational organism theory pair of the Self-Determination Theory (SDT) to explain the role of students' autonomy, competence, and kinship psychological needs exerted on entrepreneurial intentions in the form of attitudes, subjective norms, and perceived behavioral control. The above research mainly studied students' entrepreneurial trends at the individual level, and the main influencing factors include family background, gender, and entrepreneurial education, etc., with little mention of the objective market environment and other elements, and the research on the prediction of students' entrepreneurial trends at the macro-market level is rare, so there is a need to improve the relevant aspects of the research.

The article proposes a prediction algorithm based on ARIMA-LSTM. Firstly, an ARIMA model is built and used to fit the linear trend in the series, and the model is validated and fitted in the data by using the data of college students' employment and entrepreneurship from 2010 to 2022 as the research data to observe its forecasting effect. Then the ARIMA-LSTM combined forecasting model was constructed based on the ARIMA model and the LSTM forecasting model, and the forecasting results of the combined model were evaluated using the mean absolute percentage error and the root mean square error, and the forecasting results of the combined model were compared and analyzed with the forecasting results of the LSTM model and the ARIMA model, respectively.

2. Time-series analysis of market trends in entrepreneurship among university students

2.1. *Trend Forecasts for the Entrepreneurship Market Based on ARIMA Modeling*

The following modeling steps are often followed when performing ARIMA modeling on time series, as shown in Figure 1.

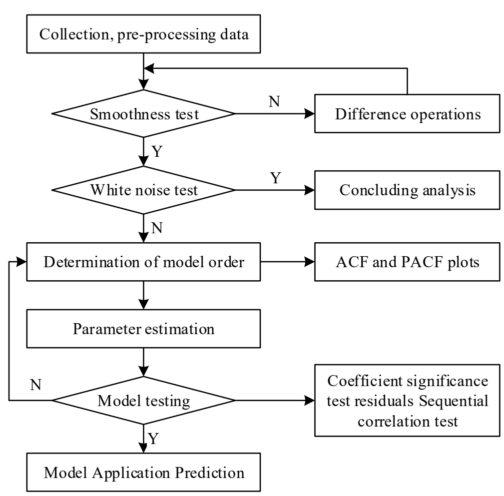

Fig. 1. Flowchart of arima model

2.2. Trend Forecasting for the Entrepreneurship Market Based on LSTM Modeling

2.2.1. LSTM model. LSTM models have good long-term dependencies in time-series based tasks, and the main advantage of LSTM over RNNs is its ability to memorize patterns over long periods of time because its high-level structure relies on feedback connections [26]. LSTM uses multiple gates as storage units to manage inputs and outputs within the network structure, instead of the traditional hidden layer with nonlinear activation architecture, LSTM has a complex hidden layer containing many parameters compared to a traditional RANN.

At time t, the current input element x, and the input of the hidden state produced in the previous step h_{t-1} 3, the network can map a sequence of inputs x_t into a sequence of outputs y_t, each y_t depending on all previous x_{t^1}, with the following recursive relationship:

$$h_t = \eta\left(w_h h_{t-1} + u_i x_t + b_h\right) \tag{1}$$

$$y_t = \eta\left(u_0 h_t + b_0\right) \tag{2}$$

In Eqs. (1) and (2) η is the activation function, b_h and b_0 are the applied deviations, and u_i, w_h and u_0 are the layer weight values.

The main difference between LSTM networks and RNN networks is that LSTM networks use gates that are not recursive hidden neurons. These gates have a self-connection at the next time step and have a decision unit that controls memory clearing and the gates control the flow of information within the network. There are various types of gates; forgetting, input and output. The forgetting gate obtains the usefulness of the information and decides to remove the useless information from the cell state, this decision has two main forms: one for clearing the previous cell state and one for keeping the previous cell state. Its expression formula is as follows:

$$F_t = \eta\left(U_f x_t + w_f h_{t-1} + b_f\right) \tag{3}$$

After that, the decision is made to store the information in the cell state. This decision is executed through two gates: the first is an input gate that determines useful information for updating the cell state; while the second is an activation gate that applies an tanh activation function on the information to update state c_t.

The LSTM model differs from the traditional neural network model in that the basic unit of the hidden layer is called the Mem-ory Block, and the structure of the memory block is shown in Figure 2:

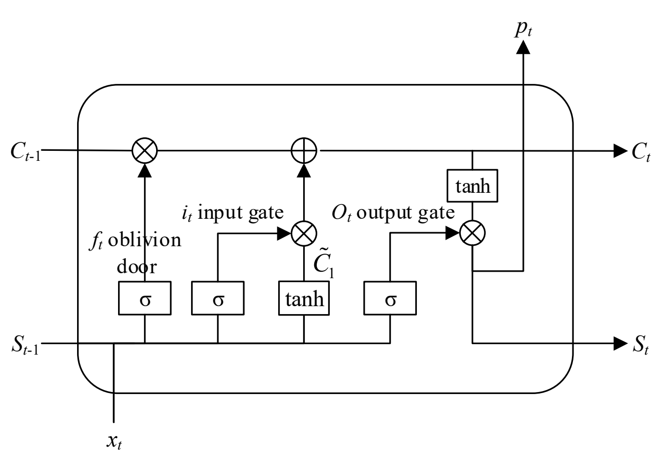

Fig. 2. LSTM Model memory block structure

In the figure, the following three gates are included: input gate (i), output gate (o) and forgetting gate (f), memory cell (c), $\oplus$ symbols represent the addition of two vectors, $\otimes$ symbols represent the dot product of two vectors, σ is the sigmoid activation function, and $\tanh$ is the hyperbolic tangent activation function, and the computational formulas are as follows, respectively:

$$\sigma(x) = \frac{1}{1+e^{-x}} \tag{4}$$

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \tag{5}$$

The LSTM network model with its input to output specific formula is as follows:

$$f_t = \sigma\left(w_f \circ [S_{t-1}, x_t] + b_f\right) \tag{6}$$

$$i_t = \sigma\left(w_i \circ [S_{t-1}, x_t] + b_i\right) \tag{7}$$

$$o_t = \sigma\left(w_o \circ [S_{t-1}, x_t] + b_o\right) \tag{8}$$

$$\tilde{c}_t = \tanh\left(w_c \circ [S_{t-1}, x_t] + b_c\right) \tag{9}$$

$$c_t = f_t \circ c_{t-1} + i_t \circ \tilde{c}_t \tag{10}$$

$$s_t = o_t \circ \tanh\left(c_t\right) \tag{11}$$

In equations (6) to (11) above, the "$\circ$" symbol is the dot product of two or two vectors; w_f is the forgetting gate, w_i is the input gate, w_o is the output gate, and w_c is the weight value vector of the memory unit. b_f, b_i, b_o and b_c are the offset vectors of the forgetting gate, the input gate, the output gate and the memory unit, respectively; x represents the input value of the network value at time t.

2.2.2. LSTM modeling steps. The use of LSTM model for univariate time series prediction, first need to set the neuron structure of the LSTM model, and then use the hyper-parameter tuning method to optimize the hyper-parameters of the LSTM, and then choose the Bayesian optimization algorithm to optimize the hyper-parameters, this is because like genetic algorithms and particle swarm algorithms they need a sufficiently large number of initial sample points and the optimization efficiency can not be guaranteed. When the optimal hyperparameters are obtained, the LSTM model is constructed according to the optimal parameters, and then the optimized LSTM model is run in the test set to make predictions and get the predicted market trend of college students' employment and entrepreneurship. The modeling ideas are as follows:

1) Data preparation, data preprocessing;
2) Divide the training set and test set;
3) Build the .LSTM model on the training set, initially design the number of network layers of LSTM, loss function, choose Adam as the optimizer, set the initial parameter of the time window width to 12, and use the college employment and entrepreneurship market trend of the previous 12 months to predict the college employment and entrepreneurship market trend of the next year, for example, by the college employment and entrepreneurship market trend from June to December 2015, the predict the July 2016 college student employment and entrepreneurship market trend.
4) Optimize the supercalculus in the LSTM model based on Bayesian optimization algorithm to get the best hyperparameters.
5) Train the BO-LSTM model in the training set to get the final model.
6) Put the final model in the test set for prediction and judge it with the three metrics of RMSE, MAE, and MAPE.

In deep learning, according to previous studies, it is found that different data sizes in the dataset will have some effects on the prediction effect of the data. Therefore, in order to eliminate the influence of the data magnitude, the data are usually standardized before using them, i.e., the data are scaled to fall into a small specific interval, so that the data are in the same magnitude. The next three commonly used standardization methods are introduced.

1) Min-Max standardization. Min-Max normalization is a linear variation for the original data that maps the values in $[-1,1]$. Noting the original sequence $A=(a_1,a_2,\ldots,a_T)^T$, find the maximum value $a_{\max}$ and the minimum value $a_{\min}$ of it. For a_i satisfying $1 \le i \le T$, the following formula is used:

$$a_i = \frac{2 * a_i - (a_{\max} + a_{\min})}{a_{\max} - a_{\min}} \tag{12}$$

Denote the normalized factor as $A' = (a'_1, a'_2, \ldots, a'_T)$. For ease of presentation, define the above normalization process as a function $Norm(\ )$:

$$A' = Norm(A) \tag{13}$$

2) Z-score normalization. Z-score normalization is based on the mean and standard deviation of the data to be processed, applicable to the original sequence of the most value of the unknown and there are outliers beyond the range of values. Remember the original sequence $X=\{x_1,x_2,\ldots,x_n\}$, whose mean is μ and standard deviation is σ. For x_i which satisfies $1 \le i \le n$, the formula is as follows:

$$x'_i = \frac{x_i - \mu}{\sigma} \tag{14}$$

In machine learning, the premise is that the data need to meet the independent distribution, so generally through the Z-score standardization is the data can obey the normal distribution. Neural network is different from machine learning, in its network structure, each layer has its own inputs and outputs, Z-score standardization is to standardize the output values in each layer of the network, so that it meets $N(0,1)$.

3) $\log$ function normalization. Remembering the original sequence $X=\{x_1,x_2,\ldots,x_n\}$, in which the largest element is $x_{\max}$, then for each element in this sequence is normalized, the result is shown below:

$$X_{norm} = \frac{\log x}{\log x_{\max}} \tag{15}$$

Also considering that Z-score normalization is not uniform in the scaling of the data, it is also not used and Min-Max is finally chosen for data normalization.

2.3. *Trend Forecasting Based on Combined ARIMA-LSTM Models*

2.3.1. ARIMA-LSTM model overall framework design. The data of college students' employment and entrepreneurship market includes different kinds of data and the data scale is large, so the ARIMA-LSTM model is designed. Firstly, the ARIMA model is used to obtain the characteristics of the factors affecting the smoothness of the data of college students' employment and entrepreneurship market, and the residual data is calculated according to the characteristic parameters, and the nonlinear characteristic data in the original data are hidden and processed, so as to avoid the interference of the nonlinear data to the overall data; secondly, the LSTM model is constructed. The LSTM model is constructed secondly, the hidden residual data is input into the model, and the residual data is corrected by the LSTM model, and finally the long-term trend prediction of the college employment and entrepreneurship market is completed in the form of the fusion of the ARIMA model and the LSTM model and the corresponding prediction results are obtained. The framework of the ARIMA-LSTM model is shown in Fig. 3.

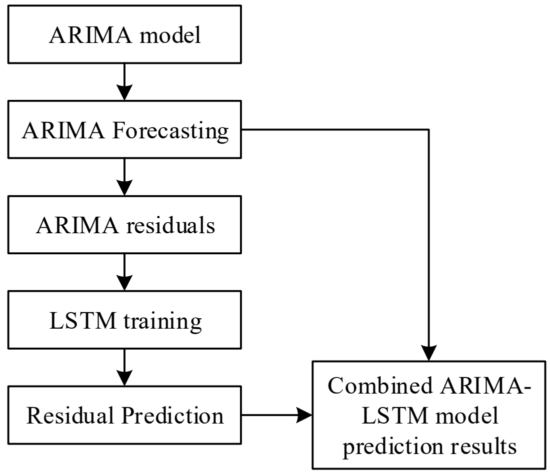

Fig. 3. The trend prediction of college students' employment and entrepreneurship market trends

2.3.2. ARIMA model based hidden data processing. The ARIMA model includes an autoregressive (AR) process and a moving average (MA) process to accomplish data preprocessing in a time series manner [27]. Constructing the ARIMA algorithm model applied to college students' employment and entrepreneurship market trend prediction requires college students' employment and entrepreneurship data to have smoothness. However, college students' employment and entrepreneurship data are easily affected by factors such as debt repayment ability, profitability, operation ability, development ability and other factors in the actual collection and processing process, which in turn leads to the acquisition of the employment rate, salary level, entrepreneurship, etc., there is an unsteady situation. By adding the differential transformation method, the random perturbation is solved and the series smoothing processing is completed, which in turn improves the smoothness of college students' employment and entrepreneurship market data.

2.3.3. LSTM model based residual data correction. The LSTM neural network optimizes the feedback error function in the gradient propagation process through the forgetting gate, corrects the residual data with multiple iterations of the input gate, updates the computational state using the update gate, and controls the output of valid information by the output gate, thus improving the convergence of the neural network. According to the LSTM structure, the unit state vector is obtained by the logic function (S-type function) to complete the neuron state activation, and the specific expression is:

$$u_t = f_t c_t + i_t \tanh\left(W_{xc} x_t + W_{kc} h_t + b_t c_v\right) \tag{16}$$

In Eq. (16), f_t is the result of forgetting gate calculation, i_t is the result of input gate data collection, W_{xc} is the neural network weight coefficient, W_{kc} is the memory module weight coefficient; b_t is the neuron bias. The LSTM neural network input gate calculates the different types of financial data and determines whether it is activated or not, and the specific expression is:

$$i_t = \frac{s\left(W_{xc} x_t + W_{kc} h_t + b_t c_v\right)}{\sigma} \tag{17}$$

In Eq. (17), s is the residual data at the previous moment of input gate input, and σ is the bias gradient. At when $i_i \leq 0$, the data is rejected in the LSTM neural network. At when $i_1 > 0$, the data will enter the update gate for memory unit calculation, and combined with dynamic control in the output gate to complete the result output. The update gate calculation formula is:

$$o_t = \left(1 - z_i\right) \times r_i \tag{18}$$

In Eq. (18), z_i is the sum of the loss function on all moments and r_i is the hidden vector. The results of the update gate calculation are passed to the output gate, and the residual data correction output of the LSTM neural network is completed according to the weighted influence of the output gate.

2.3.4. Forecasts of Long-Term Changes in Market Trends Based on Model Fusion. By constructing a combination model of ARIMA model and LSTM model, obtaining the response time employment and entrepreneurship forecasting index data, filtering out the nonlinear regular data, completing the college students' employment and entrepreneurship time series feature extraction in ARIMA model, and completing the residual data processing in LSTM model. Finally, the results are superimposed to obtain the prediction results of the long-term change trend of college students' employment and entrepreneurship market, in order to solve the prediction of large-capacity and multi-dimensional data. The nonlinear component residual mathematical expression is:

$$e_t = l_t - U_t \tag{19}$$

In equation (19), l_t is the correlation index of each element of corporate finance; U_t is the predicted value of ARIMA model. After obtaining e_t, the LSTM model is used to model the long-term trend of changes in corporate finance and obtain the predicted value of financial changes, which is expressed as:

$$F_t = f\left(e_{t-1}, e_{t-2}, \ldots, e_{t-m}\right) \times v \tag{20}$$

In Eq. (20), f () is the relationship function established by the LSTM model; v is the limiting variable. Finally, U, and F, are summed to obtain the results of predicting the long-term trend of corporate financial changes. As in equation (21):

$$D_t = U_t + F_t \tag{21}$$

In order to ensure the convergence of the combined model, it is necessary to combine the method of iterative forecasting with the results of the forecast of the long-term trend of financial changes in the enterprise for the next moment of the financial time series value forecast, so as to complete the forecast of the long-term trend of financial changes in each future moment.

3. Empirical analysis and analysis

3.1. Analysis of ARIMA model prediction results

3.1.1. Census X12 analysis. X12 allows a suitable ARMA (p, d, q) model to be built for the adjusted series before seasonal adjustment, and the monthly trend from January 2010 to August 2022 is shown in Figure 4, and it can be seen that the college entrepreneurship and employment market trend has a more obvious non-stationary characteristics, using the EViews software system, and through the ADF and Correlogram test, the choice of the ARMA (1, 2, 2) model is more effective, after processing the data, split to get TC, S, I factor data as shown in Figure 5 - Figure 7.

From the long-term trend, the bottom climbed slowly before 2012, and the peak value (the difference between the peak and the trough) was small, and after 2012, the vibration amplitude increased significantly and the volatility was stronger. From the point of view of seasonal factors, the months that cause the seasonal rise in the demand for college entrepreneurship and employment are mainly in September and May (graduation season), while the months that cause the seasonal decline in college entrepreneurship and employment are mainly in January month (during the New Year). In terms of the irregular factor, the irregular factor has a very large impact on the demand in some months, such as the 2016 4 trillion yuan investment.

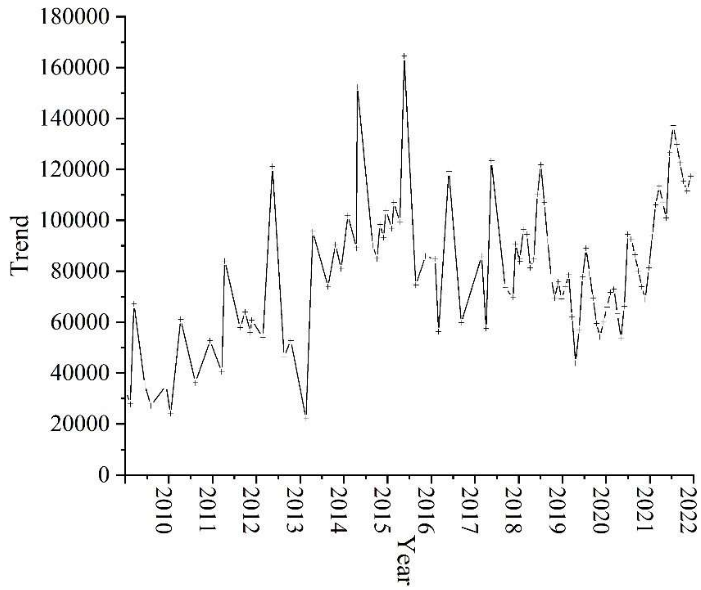

Fig. 4. January 2010 to 2022

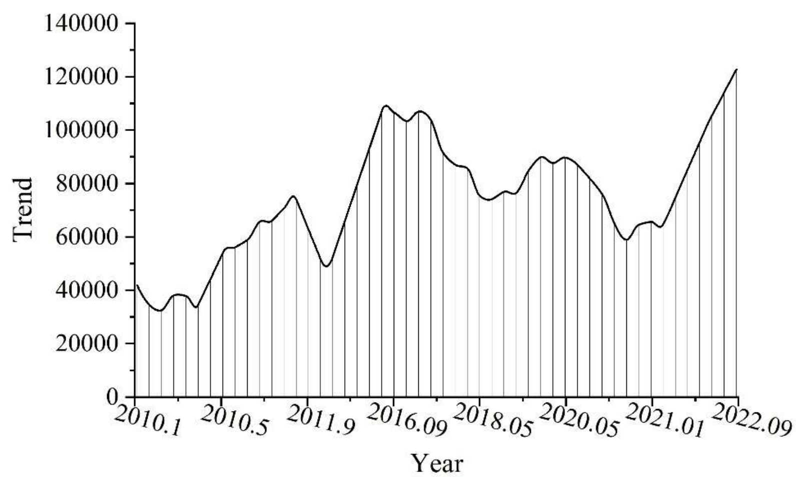

Fig. 5. TC-trend cycle factor

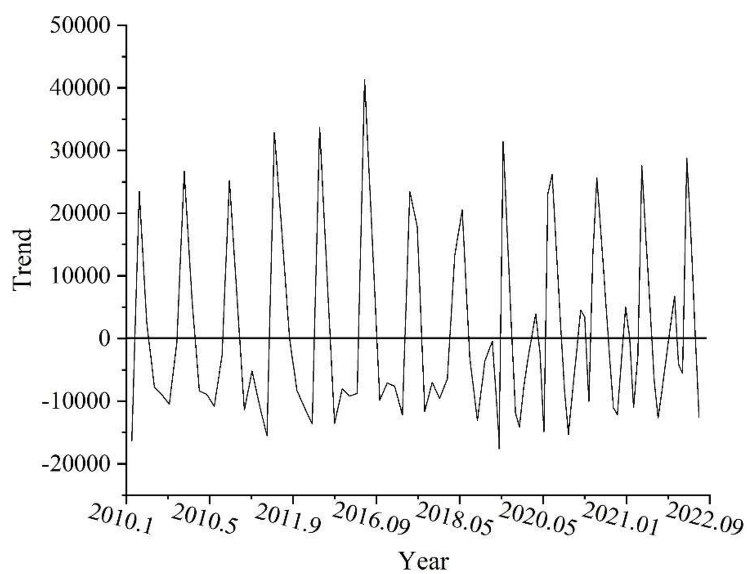

Fig. 6. S-seasonal factors

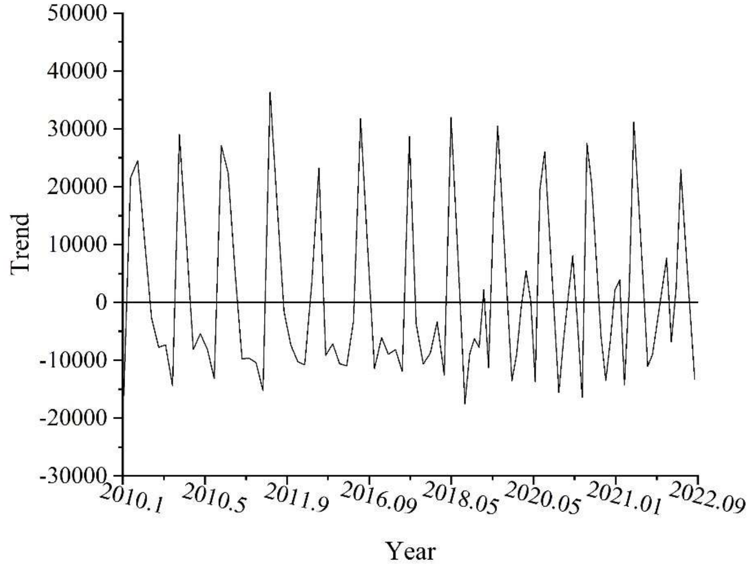

Fig. 7. I-irregular factor

3.1.2. TC sequence modeling. In order to forecast the trend, the TC series needs to be modeled. Before modeling, the smoothness of the data is first checked and the results of Table 1 are obtained using the ADF test.

Table 1. Tc sequence ADF test results

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-1.569	0.293
Test critical values:	-3.352	
	5%level	-2.743
	10% level	-2.451

From the results, it can be seen that the TC sequence is not smooth $(P = 0.29 > 0.05)$ at the 95% level of significance, and therefore the first-order differencing of the TC sequence is required:

$$K_t = TC_t - TC_{t-1} \tag{22}$$

The ADF test is then performed on the K-series and the results are shown in Table 2.

Table 2. K sequence adf test results

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-3.012	0.0275
Test critical values:	-3.3672	
	5% level	-2.71
	10% level	-2.43

From the results, it can be seen that the K-series rejects the single-root hypothesis of having 1 at the 95% significant level (P = 0.0275 < 0.05), which can be utilized to build the ARMA model. Through several model simulations, the regression of ARMA(3, 1) was fitted best, and its results are shown in Table 3.

Table 3. The results of the arma(3, 1) model

Variable	Coefficient	Std.Error	t-Statistic	Prob.
K(-1)	2.356	0.057	37.677	0.000

K(-2)	-2.11	0.101	-19.15	0.000
K(-3)	0.723	0.051	11.56	0.000
MA(1)	0.396	0.077	4.533	0.000
R-squared		0.989	Mean dependent var	611.56
Adjusted R-squared		0.985	S.D.dependent var	2543.477
S.E.of regression		241.18	Akaike info criterion	11.741
Sum squared resid		8321477.	Schwarz criterion	11.926
Log likelihood		-1019.233	Durbin-Watson stat	1.903

Based on the significance, the model can be shown to be:

$$K_t = 2.356K_{t-1} - 2.11K_{t-2} + 0.723K_{t-3} + \xi_t + 0.396\xi_{t-1} \tag{23}$$

Combining Eq. (22), Eq. (23) shows that:

$$\begin{aligned} TC_1 &= 3.412TC_{t-1} - 4.622TC_{t-2} + 2.977TC_{t-3} \\ &\quad -0.767TC_{t-4} + \xi_t + 0.41\xi_{t-1} \end{aligned} \tag{24}$$

Then combined with formula (22), formula (24) can be calculated model predicted value, will be compared with the actual value of demand, the comparison results shown in Figure 8, the two fit is very high, most of the trend is basically the same.

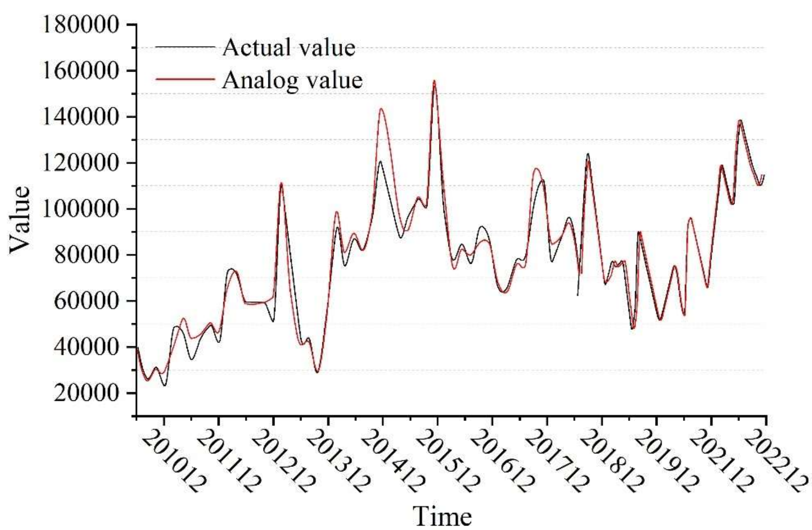

Fig. 8. The actual value of the demand and the prediction of the model

3.1.3. Example of a trend forecasting algorithm. Using the EViews software forecast tool, combined with the formula (23), it can be deduced that the forecast value of the TC in September 2021 is 127,451, the quarterly factors, the impact of irregular factors with reference to the trend in September in the last three years (take the average), the final market demand forecast value is 132,261, according to the data of college students' entrepreneurship and employment shows that the actual market demand is 124,562, the deviation is only 3.1%.

3.2. Analysis of LSTM model prediction results

3.2.1. Data construction. In order to give full play to the role of college student entrepreneurship and employment data, and at the same time to improve the prediction accuracy of the model, this paper adopts three consecutive years of college student entrepreneurship and employment data to recursively predict the next year's data for the construction of the LSTM model sample data, that is, the use of the college student entrepreneurship and employment data from 2019 to 2022 to predict the college student entrepreneurship and employment data for the year of 2026, and in total, we can build a total of 117 sets of sample data, and use the the first 93 sets of data as the training dataset and use the next 24 sets of data as the test dataset.

3.2.2. Model parameterization. In this paper, Adam's algorithm is used as the optimizer, tanh is used as the trigger function, and the number of iterations is set to 100. In order to achieve the best prediction performance of the LSTM model, the parameters such as the number of hidden layers, the number of neurons, and the learning rate of the network model need to be set, and the model is predicted in several attempts using historical data, and the analysis of the results based on the prediction results reveals that the model fits best when the number of hidden layers is 2, number of neurons is 4, and the learning rate is 0.001, the model fits best.

3.2.3. Forecasting results and analysis. By using 96 sets of training data from the college entrepreneurship employment sample data to train the LSTM model and 24 sets of validation data to verify the model, the trend of the loss function is shown in Figure 9.

From the figure, it can be seen that due to the increase in the number of iterations, the value of the loss function continues to decrease, and when the model is trained to roughly the 5th time, the value of the loss function decreases significantly and remains stable, at which time the model training is the most effective.

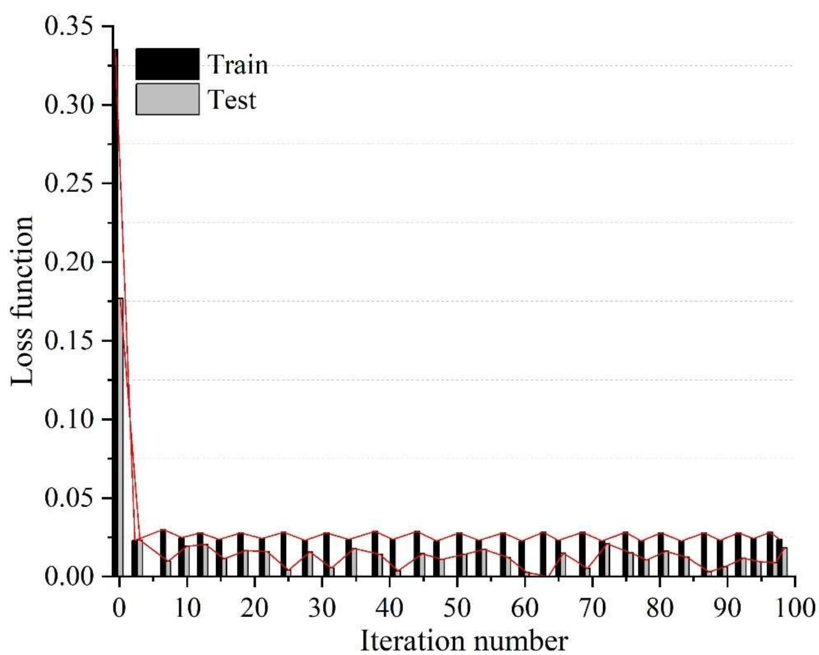

Fig. 9. Loss function diagram

The trained LSTM model is used to predict the test set and the obtained prediction results are shown in Figure 10. From the prediction results of the test set, it can be seen that the LSTM model predicted value and the actual value of the volatility of the trend of the fit is better, but with the prolongation of time, by the influence of external reasons, its prediction error is gradually becoming larger but the overall prediction effect is better than the ARIMA model, which indicates that because of the comprehensive consideration of a variety of factors interacting with each other at the same time, combined with the history of throughput data, it can be very well fitted to the time series of the nonlinear characteristics. This is related to the nonlinear ability of the neural network. The above prediction results are evaluated using evaluation indexes, and the results are shown in Table 4.

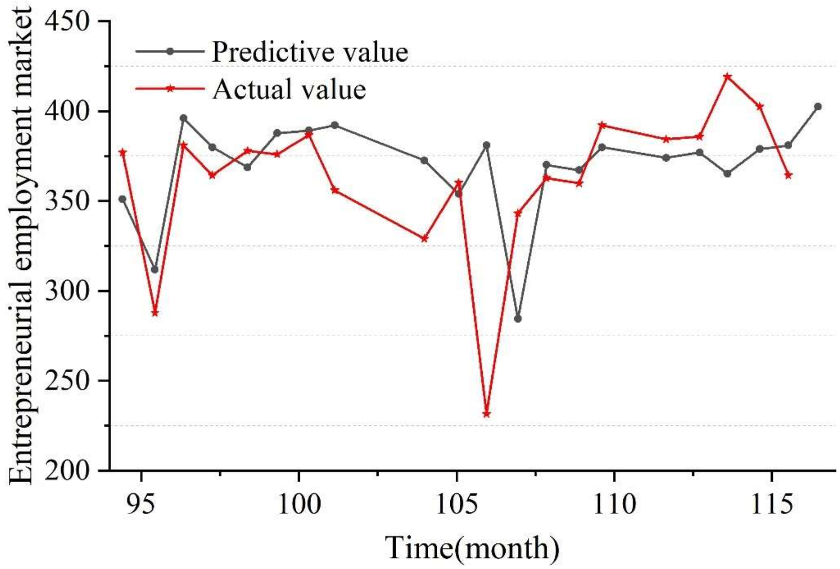

Fig. 10. LSTM model prediction results

Table 4. The LSTM model predicts the value of the result

Evaluation index	LSTM
MAPE	5.152
RMSE	31.245

3.3. Analysis of Combined Model Prediction Results

From the analysis of the prediction results of the above two models, it can be seen that the ARIMA model fits the linear trend of the time series better, but with the extension of time, the error gradually becomes larger, and the LSTM model fits the nonlinear ability of the time series better. The combined ARIMA-LSTM model developed in this paper can combine the advantages of both, accurately predicting the trend as well as fitting the volatility in the series to improve the prediction accuracy.

In this section, the combined model constructed in section 2.3 is used to forecast the trend of the college entrepreneurship employment market, firstly, the test data in the college entrepreneurship employment market is input into the ARIMA (2,1,1) model for forecasting, the linear part of the final forecast result is obtained, and the forecast result is subtracted from the actual value to obtain the forecast residual, and secondly, the residual is combined with the trend of the college entrepreneurship employment market determined above Secondly, the residuals are combined with the data of the influencing factors identified above and fed into the LSTM model to get the nonlinear part of the final prediction results, and the results of the prediction residuals are shown in Figure 11.

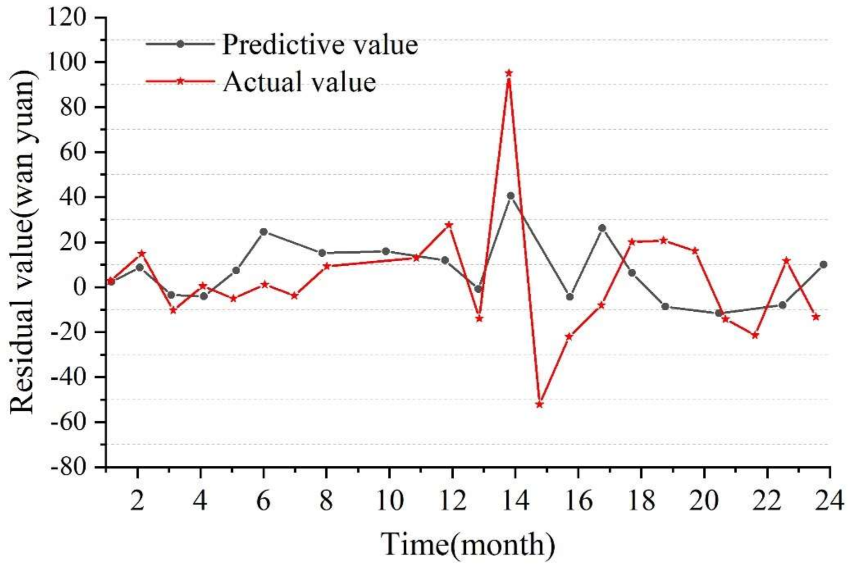

Fig. 11. The ARIMA-LSTM composite model predicts the residual results

The linear and nonlinear prediction results of the ARIMA (2,1,1) model and the two parts of the LSTM model are superimposed to derive the final predicted value of the college entrepreneurship and employment market trend, and the combined model prediction results are shown in Figure 12. From the figure, it can be found that the predicted value of the combined model is closer to the direction of the actual value, and its long-term trend and volatility are better fitted.

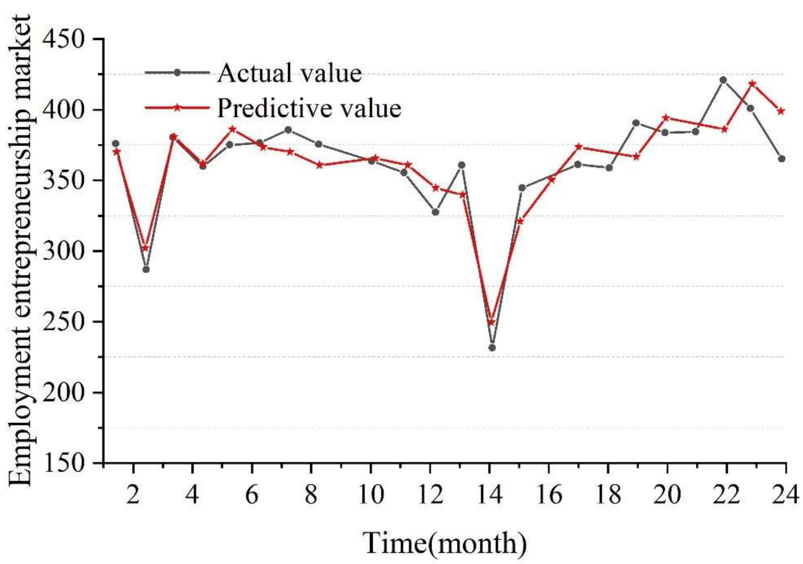

Fig. 12. The ARIMA-LSTM composite model predicts thultse res

In order to more accurately compare the prediction effect of the three models, this paper calculates the MAPE and RMSE of the combined model, and compares the values of the evaluation indexes and the error values of the prediction results of the ARIMA (2,1,1) model and the LSTM model calculated as a result, and the results are shown in Table 5 and Table 6.

Table 5. comparison

Evaluation index	ARIMA (2,1,1)	LSTM	Combination model
MAPE	9.411	5.293	2.145
RMSE	47.268	32.244	11.677

Table 6. Model predictive value and error

Serial number	Actual value /wan TEU	Predictive value / wan TEU			Error value / wan TEU		
		ARIMA	LSTM	Combination model	ARIMA	LSTM	Combination model
97	372.90	374.48	348.19	368.11	1.43	-22.51	-4.57
98	283.20	297.69	308.09	298.92	13.41	23.72	14.28
99	378.20	367.69	391.32	378.89	-9.48	12.87	0.65
100	358.90	358.39	373.58	358.69	-0.45	13.79	-0.17
101	373.30	368.03	367.83	384.37	-4.99	-5.32	10.12
102	373.30	373.90	383.46	371.40	0.57	9.91	-1.41
103	383.20	378.96	384.14	367.52	-3.79	0.91	-14.52
104	373.70	382.26	387.76	358.45	7.88	13.96	-14.09
105	368.20	377.76	377.89	361.73	8.73	8.83	-6.37
106	360.60	371.36	373.00	362.59	9.59	11.37	2.00
107	352.90	364.20	368.59	358.05	10.24	14.58	4.09
108	324.30	351.44	348.54	343.02	26.84	23.17	17.87
109	358.30	342.91	378.56	337.51	-14.29	19.83	-19.69
110	227.50	323.17	278.28	246.49	94.58	49.86	18.00
111	340.90	286.83	365.03	283.36	-53.11	23.27	-21.48
112	348.40	325.57	361.04	347.98	-21.74	11.78	-0.37
113	359.40	351.45	376.07	370.59	-6.95	15.75	11.09
114	357.90	377.69	372.48	367.37	18.67	13.93	8.57
115	388.00	407.69	371.10	363.30	18.55	-15.78	-23.58
116	381.70	397.69	373.01	392.51	15.00	-7.87	9.49
117	383.00	367.69	361.97	385.96	-14.71	-20.11	2.54
118	418.00	395.96	376.07	382.62	-21.69	-40.64	-34.64
119	398.20	410.14	378.91	416.16	10.72	-18.27	16.84
120	361.30	347.69	399.22	396.17	-12.71	36.93	33.58

Comparing the prediction results of ARIMA model and LSTM model with the prediction results of the combination model respectively, the MAPE of the prediction value of the combination model is reduced to 0.978, the RMSE is reduced to 9.452, and the maximum prediction error is 33.78, all of which are smaller than that of the prediction results of the single model in terms of the evaluated value and prediction error, which means that the combination model constructed in this paper is able to effectively fit the time series compared to the single model with a linear and nonlinear intertwined superposition trend, the relative error is smaller, which can improve the scientificity and accuracy of the prediction of the college entrepreneurship employment market trend. Applied to the prediction of the trend of college student entrepreneurship and employment market, it can help the relevant departments of the college student entrepreneurship and employment market to make reasonable decisions and plans, and improve the operational efficiency of the college student entrepreneurship and employment market.

4. Conclusion

The study designed an ARIMA-LSTM-based algorithm for predicting the trend of college entrepreneurship employment market and analyzed the prediction results. The characteristics of the college entrepreneurship employment market time series linear and nonlinear intertwined and overlapped, and based on the different performance of the model, the ARIMA linear model and LSTM nonlinear model were combined, and the residuals of the ARIMA model prediction combined with the college entrepreneurship employment market influencing factors were jointly inputted into the LSTM model. The experiment proves that ARIMA-LSTM is better than the single model compared with the prediction results of ARIMA model and LSTM model, with the model prediction value MAPE reduced to 0.978 and RMSE reduced to 9.452. As a result, the model is better adapted to the data of college students' entrepreneurship and employment market, and the accuracy of the prediction results is improved to provide a reliable reference for the college students' entrepreneurship and employment market.

Funding

1) General Project of Humanities and Social Sciences Research of Ministry of Education in 2023 (23JDSZ3081).
2) Key research project of Anhui University of Science and Technology's school-level teaching reform in 2023 (JG202344).

References

[1] Hilorme, T., Shurpenkova, R., Kundrya-Vysotska, O., Sarakhman, O., & Lyzunova, O. (2019). Model of energy saving forecasting in entrepreneurship. Journal of Entrepreneurship Education, 22, 1-8.

[2] Kwilinski, A., Litvin, V., Kamchatova, E., Polusmiak, J., & Mironova, D. (2021). Information support of the entrepreneurship model complex with the application of cloud technologies. International Journal of Entrepreneurship, 25(1), 1-8.

[3] Singh, S. H., Bhowmick, B., Eesley, D., & Sindhav, B. (2021). Grassroots innovation and entrepreneurial success: Is entrepreneurial orientation a missing link?. Technological Forecasting and Social Change, 164, 119582.

[4] Park, C. (2017). A study on effect of entrepreneurship on entrepreneurial intention: Focusing on ICT majors. Asia Pacific Journal of Innovation and Entrepreneurship, 11(2), 159-170.

[5] Miller, K., Alexander, A., Cunningham, J. A., & Albats, E. (2018). Entrepreneurial academics and academic entrepreneurs: A systematic literature review. International Journal of Technology Management, 77(1-3), 9-37.

[6] Farinha, L., Ferreira, J. J., & Nunes, S. (2018). Linking innovation and entrepreneurship to economic growth. Competitiveness Review: An International Business Journal, 28(4), 451-475.

[7] Buckley, P., & Doyle, E. (2017). Individualising gamification: An investigation of the impact of learning styles and personality traits on the efficacy of gamification using a prediction market. Computers & Education, 106, 43-55.

[8] Zheng, L. J., Xiong, C., Chen, X., & Li, C. S. (2021). Product innovation in entrepreneurial firms: How business model design influences disruptive and adoptive innovation. Technological Forecasting and Social Change, 170, 120894.

[9] Terwiesch, C. (2023). Would Chat GPT3 get a Wharton MBA? A prediction based on its performance in the operations management course. Mack Institute for Innovation Management at the Wharton School, University of Pennsylvania, 45.

[10] Guerrero, M., & Urbano, D. (2017). The impact of Triple Helix agents on entrepreneurial innovations' performance: An inside look at enterprises located in an emerging economy. Technological forecasting and social change, 119, 294-309.

[11] Kerrin, M., Mamabolo, M. A., & Kele, T. (2017). Entrepreneurship management skills requirements in an emerging economy: A South African outlook. The Southern African Journal of Entrepreneurship and Small

Business Management, 9(1), 1-10.

[12] Si, S., Hall, J., Suddaby, R., Ahlstrom, D., & Wei, J. (2023). Technology, entrepreneurship, innovation and social change in digital economics. Technovation, 119, 102484.

[13] Al-Mamary, Y. H. S., Abdulrab, M., Alwaheeb, M. A., & Alshammari, N. G. M. (2020). Factors impacting entrepreneurial intentions among university students in Saudi Arabia: testing an integrated model of TPB and EO. Education+ Training, 62(7/8), 779-803.

[14] Rippa, P., & Secundo, G. (2019). Digital academic entrepreneurship: The potential of digital technologies on academic entrepreneurship. Technological Forecasting and Social Change, 146, 900-911.

[15] Núñez-Canal, M., de Obesso, M. D. L. M., & Pérez-Rivero, C. A. (2022). New challenges in higher education: A study of the digital competence of educators in Covid times. Technological Forecasting and Social Change, 174, 121270.

[16] Shah, N., & Soomro, B. A. (2017). Investigating entrepreneurial intention among public sector university students of Pakistan. Education+ Training, 59(7/8), 841-855.

[17] Klofsten, M., Fayolle, A., Guerrero, M., Mian, S., Urbano, D., & Wright, M. (2019). The entrepreneurial university as driver for economic growth and social change-Key strategic challenges. Technological Forecasting and Social Change, 141, 149-158.

[18] Williamson, B. (2021). Making markets through digital platforms: Pearson, edu-business, and the (e) valuation of higher education. Critical Studies in Education, 62(1), 50-66.

[19] Wu, L., Jiang, S., Wang, X., Yu, L., Wang, Y., & Pan, H. (2022). Entrepreneurship education and entrepreneurial intentions of college students: The mediating role of entrepreneurial self-efficacy and the moderating role of entrepreneurial competition experience. Frontiers in psychology, 12, 727826.

[20] Liu, X., Lin, C., Zhao, G., & Zhao, D. (2019). Research on the effects of entrepreneurial education and entrepreneurial self-efficacy on college students' entrepreneurial intention. Frontiers in psychology, 10, 869.

[21] Youssef, A. B., Boubaker, S., Dedaj, B., & Carabregu-Vokshi, M. (2021). Digitalization of the economy and entrepreneurship intention. Technological Forecasting and Social Change, 164, 120043.

[22] Fragoso, R., Rocha-Junior, W., & Xavier, A. (2020). Determinant factors of entrepreneurial intention among university students in Brazil and Portugal. Journal of Small Business & Entrepreneurship, 32(1), 33-57.

[23] Nguyen, C. (2018). Demographic factors, family background and prior self-employment on entrepreneurial intention-Vietnamese business students are different: why?. Journal of Global Entrepreneurship Research, 8(1), 1-17.

[24] Obschonka, M., Moeller, J., & Goethner, M. (2019). Entrepreneurial passion and personality: the case of academic entrepreneurship. Frontiers in psychology, 9, 2697.

[25] Al-Jubari, I., Hassan, A., & Liñán, F. (2019). Entrepreneurial intention among University students in Malaysia: integrating self-determination theory and the theory of planned behavior. International entrepreneurship and management journal, 15(4), 1323-1342.

[26] Eko Putra Wahyuddin,Rezzy Eko Caraka,Robert Kurniawan,Wahyu Caesarendra,Prana Ugiana Gio & Bens Pardamean. (2025). Improved LSTM hyperparameters alongside sentiment walk-forward validation for time series prediction. Journal of Open Innovation: Technology, Market, and Complexity(1),100458-100458.

[27] Ahmad Dar,Akshat Jain,Mehak Malhotra,Ataur Rahman Farooqi,Olayan Albalawi,Mohammad Shahfaraz Khan & Hiba. (2024). Time Series analysis with ARIMA for historical stock data and future projections. Soft Computing(21),1-12.