Research Article

Construction of Graduate Behavior Dynamic Model Based on Dynamic Decision Tree Algorithm

Fen Yang

Henan University of Animal Husbandry and Economy, Zhengzhou 45004, China

Correspondence should be addressed to Fen Yang; 2016150239@jou.edu.cn

Received 10 November 2021; Accepted 23 December 2021; Published 22 January 2022

Academic Editor: Gengxin Sun

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

With the research of machine learning technology and big data intelligent processing technology in engineering application becoming more and more mature, people gradually combine machine learning technology and big data intelligent processing technology. Aiming at the problem of innovative employment in colleges and universities, this paper proposes a dynamic decision tree algorithm based on these two technologies and constructs a dynamic model of graduates' behavior. Through the analysis of dynamic decision tree algorithm, a big data analysis system is formed. Finally, simulation experiments verify whether the dynamic model can correctly reflect the behavior of college graduates. The results show that the big data integration system based on big data and dynamic decision tree algorithm has high adaptability. Incremental adaptive optimization of the traditional decision tree model can significantly improve the prediction effect and prediction time of dynamic data and provide theoretical support for the industrialization and social significance of big data technology. The dynamic decision tree algorithm of college employment proposed in this paper has good predictability and provides a certain theoretical reference for college graduates' entrepreneurship.

1. Introduction

Big data usually refers to the data with huge volume of information in the three dimensions of "data volume," "data speed," and "data category." People call it big data technology to get some quantitative statistical results through big data analysis, which mainly relies on the internal logic of various specific rules. Among them, machine learning is an effective method to realize this analysis method [1]. Every year, the graduation of college students injects a lot of productivity into the society. Their behavior after graduation is the focus of all sectors of society. At present, most of the graduate behavior analysis still adopts the questionnaire method [2]. Although the statistical results of the questionnaire survey method are easy to quantify, the statistics, processing, and analysis of the survey results can be completed more quickly and economically [3]. However, due to the subjective changes and differences in the intention, thinking, and motivation of the respondents in the questionnaire survey for employment and entrepreneurship, the quality of the survey results is often not guaranteed, and its

recovery rate is difficult to be guaranteed, resulting in a great reduction in the accuracy of the data analysis results [4]. This makes the dynamic model of graduates' whereabouts constructed by questionnaire has some limitations. As the data set of graduates' graduation behavior is changing day by day and its internal inaccessible nodes are increasing, the predictability of the traditional model to each node is not high, which often leads to the quantitative change of the prediction distortion of each node, resulting in the qualitative change of the overall prediction result error. Therefore, it is necessary to study the dynamic model of graduates' employment and entrepreneurship data [5]. Based on this, this study proposes a dynamic model construction method based on big data and machine learning decision tree.

Decision tree is a tree structure similar to flow chart, in which each internal node represents a test on an attribute, each branch represents an attribute output, and each leaf node represents a class or class distribution. The top layer of the tree is the root node. The amount of information is directly related to its uncertainty. To understand a very, very uncertain thing or something we know nothing about, we

need to know a lot of information, and the measurement of the amount of information is equal to the amount of uncertainty. Decision tree machine learning method is an effective method to simulate human learning behavior. Its core problem is how to construct the tree topology based on the known training sample data. Since machine learning is a problem closely related to the problem domain, different application goals will lead to more specific and in-depth topics that cannot be solved or well solved by the existing theoretical framework.

Aiming at the problem that colleges and universities cannot carry out dynamic analysis on graduates' employment and entrepreneurship destination statistics, a big data integrated analysis system based on dynamic decision tree machine learning algorithm is proposed to solve the problem of dynamic analysis and statistics of graduates' destination data. This study is mainly carried out in four chapters. Section 1 gives a framework overview of the research background and research direction of this study; Section 2 introduces the application status of big data analysis technology and its application scope not covered at present. Section 3 introduces the big data technology, machine learning algorithm, the establishment of dynamic decision tree model, and the construction of big data integration system. Section 4 uses the established decision tree model based on machine learning algorithm to practice a large amount of sample data through public data sets and designs confirmatory experiments to verify whether the optimized decision tree model can accurately predict the postgraduation behavior of college graduates.

The innovation of this study is to solve the problem of statistics and analysis of employment and entrepreneurship dynamics of college graduates through big data analysis technology. This study combines the big data of graduates' behavior after graduation with machine learning algorithm. Aiming at the problem that graduates' behavior data changes year by year, a dynamic decision tree model is proposed based on machine learning algorithm. Even if the data sample changes, it can still get a high prediction accuracy without rebuilding the new model, which solves the problem of low energy efficiency when the traditional model analyzes dynamic data. The effective application of the algorithm can greatly improve the model structure of each branch of the decision tree and then improve the efficiency of retrieving dynamic data.

2. Related Work

Although researchers have studied the complex system strategy based on big data for many years, there are still some deficiencies in the construction scenario of big data dynamic model [6]. Elsayad et al. proposed to use whale optimization algorithm under the background of big data of college graduates and carry out algorithm loading through machine learning H_2O framework to find the best feature subset, so as to maximize the prediction accuracy of employment of college graduates [7]. Through experiments, Dommaraju et al. have proved that using the multilayer perceptron deep neural network learning technology based on expectation

condition maximization clustering and Ruzicka regression to analyze the big data of cellular network can provide theoretical support for the intelligent suggestions on employment and entrepreneurship of college graduates [8]. Yahia et al. proposed to transform the common feature data in big data technology into the deep feature data filtered by the feature selector and, combined with the hybrid machine learning algorithm, analyzed the incentive elements of the whereabouts of college graduates. This method shows that business trip practice is the main incentive to determine the behavior of graduates and achieves an accuracy rate of more than 95% [9]. With the help of random forest algorithm, Gui et al. overcome the shortcomings of overfitting algorithm by using the weather conditions, flight time, airport location, and other employment registration information of college graduates, so as to improve the accuracy of prediction and analysis of college graduates in employment to more than 90% [10]. Latif et al. analyzed the big data according to 70 attributes that may affect the delayed graduation of college graduates and obtained the prediction accuracy of delayed graduation events greater than 98% based on convolutional neural network (CNN) algorithm. Based on this result, it has good social significance for college graduates to change their strategies according to the current learning situation and improve their school efficiency [11]. Dijk et al. proposed to establish the largest information database for college graduates, which aims to use big data technology to predict the whereabouts of college graduates, so as to improve the statistical efficiency of the whereabouts of graduates [12]. Xiao et al. analyzed the data of the ship automatic identification system by using adaptive learning, motion modeling, and particle filter technology through big data acquisition and aiming at the direction of intelligent maritime traffic and carried out the collision risk assessment of ship navigation dynamic trajectory. They found that big data analysis technology has a good effect in processing dynamic data analysis [13]. Liu et al. verified the effectiveness of the cooperative spectrum sensing network based on the two terminal machine learning models for the intelligent spectrum sensing method by detecting the probability simulation method. It realized the diversified statistics of employment and entrepreneurship of college graduates, but could not achieve dynamic real-time analysis [14]. Through the particle swarm optimization algorithm, Nogueira et al. generate a large data set of dynamic feasible solutions and obtain the optimal solution based on big data, so as to establish a probability confidence region, find the optimal conditions for analyzing graduates' employment and entrepreneurship choices, and realize the dynamic analysis of equipment status based on big data technology [15]. Fnyes et al. put forward the theory of using machine learning linear parameter change control to analyze big data, obtained a highly robust vehicle selection function for college graduates to check in and take a ride, and verified its effectiveness through simulation. This method can intelligently analyze the choice of graduates' whereabouts, but can not carry out high accuracy analysis and statistics on dynamic data [16].

To sum up, it can be seen that, in the process of dynamic statistics and analysis of employment and entrepreneurship

of college graduates, there are some problems, such as inability to effectively obtain real-time data of graduates, large error of analysis model, small database of graduates' statistical information, and poor coupling of analysis model [17, 18]. In terms of the research on the dynamic model of employment and entrepreneurship of college graduates, the current research results are mostly implemented based on statistics, but rarely combined with big data technology and artificial intelligence algorithms for diversified and convenient analysis [19–21]. On the other hand, big data analysis technology is mostly coupled with traditional engineering and less combined with science education. In terms of algorithm structure, most machine learning algorithms only perform final calculation for static data, and few dynamic algorithm models are designed for complex dynamic discrete data [22]. Therefore, under the background of big data in today's era, it is of great significance to carry out the research on the dynamic model of employment and entrepreneurship of college graduates.

3. Dynamic Model Construction Method Based on Big Data and Dynamic Decision Tree Algorithm

3.1. Application of Machine Learning Algorithm in Digital Model Construction. Machine learning algorithm is a subject that can optimize computer algorithm by improving the previous data. The three main elements of machine learning are "algorithm," "strategy," and "model," in which "algorithm" fits the training data set through "strategy," so as to create a new "model," which makes the algorithm very important in machine learning [23]. Common machine learning algorithms are usually trained only for fixed data sets. Once the data sets change, the algorithm will rescan all the data sets, which makes the algorithm inefficient in solving dynamic big data [24]. In fact, in solving practical engineering problems, big data with dynamic attributes are often used as training data sets [25]. In recent years, with the deepening of the special work of Higher Vocational enrollment expansion by the Ministry of Education, the Ministry of Human Resources and Social Security, and other departments, the number of college graduates in China has increased from 2.8 million per year in 2004 to an estimated 9 million in 2021. According to the regularity and uncertainty of social development every year, the data of employment and entrepreneurship of college graduates also show dynamic changes.

Based on this, in order to solve the lag analysis problem in graduates' employment and entrepreneurship data, this study improves the traditional decision tree algorithm for static data and proposes a big data integration system based on dynamic decision tree algorithm, so as to build a new dynamic model in the complex discrete data of graduates' employment and entrepreneurship changing year by year.

3.2. Data Analysis Process of Machine Learning Dynamic Decision Tree Algorithm in Dynamic Model. The traditional decision tree algorithm selects the optimal subnode by

calculating the information gain of different attributes. For the uncertainty of college graduates' employment and entrepreneurship, this study extends a decision tree algorithm with dynamic incremental solution based on the traditional decision tree algorithm, reprocesses the discrete data at all levels in the dynamic model structure, and extracts the rules of each node. The process of dynamic decision tree data analysis is shown in Figure 1.

As can be seen from Figure 1, the relationship between the decision tree analysis data is based on information entropy. Under the influence of dynamic data, when the change of a child node of the decision tree reaches the threshold, the child node becomes invalid, so it is automatically replaced with another child node. When the dynamic data does not touch the threshold, it means that the subnodes of the decision tree are probabilistic classified based on the node attributes.

Make a subnode of the decision tree algorithm model have two attributes to be selected: A and B . If $\text{gain}(A) > \text{gain}(B)$ is found after calculating the information gain of a and B , that is, in terms of subnodes of the decision tree, the decision of attribute a is better than attribute B . When the data is dynamically incremented, if the current decision tree branch node subpoint has and only has two attributes a and B , then Ea and Eb are used to represent the entropy of attributes a and B , respectively, and then there will be

$$\begin{aligned} \text{gain}(A) > \text{gain}(B) &\Leftrightarrow I(p, n) - E(A) > I(p, n) - E(B) \\ &\Leftrightarrow E(A) > E(B) \Leftrightarrow Ea > Eb, \end{aligned} \quad (1)$$

where p represents the number of positive examples with a certain attribute value record set, n represents the number of counterexamples with the same attribute value record set, and their number count is expressed in i , $I(p, n)$ represents the amount of information estimated based on the overall data source, and $p/(p+n)$ and $n/(p+n)$ are the possibility of positive examples and the possibility of counterexamples, respectively. At this time, the method of calculating information entropy E is as follows:

$$\begin{aligned} E &= \sum_{i=1}^m \frac{p_i + n_i}{p + n} \left(-\frac{p_i}{p_i + n_i} \log \left(\frac{p_i}{p_i + n_i} \right) - \frac{n_i}{p_i + n_i} \log \left(\frac{n_i}{p_i + n_i} \right) \right) \\ &= \frac{\theta}{p + n}, \end{aligned} \quad (2)$$

where

$$\theta = - \sum_{i=1}^m \left(p_i \log \left(\frac{p_i}{p_i + n_i} \right) + n_i \log \left(\frac{n_i}{p_i + n_i} \right) \right). \quad (3)$$

In the formula, $\theta \geq 0$ and m represent the values of the branch subnodes of a decision tree. The value includes the number of values of the node attribute, and the bottom of $\log$ is 2. When the dynamic incremental data is introduced into the decision tree, the information gain of attributes A and B will change accordingly. This change may lead to the final change of the node classification attribute under some

FIGURE 1: The process of analyzing data by dynamic decision tree.

columns of complex discrete calculations. When the dynamic incremental data with attribute p appears in a value segment k and the value segment is between A and B , the method for calculating information entropy E_1 is as follows:

$$\begin{aligned}
 E_1 &= \frac{-1}{p+n+1} \sum_{i=1}^m \left(p_i \log \left(\frac{p_i}{p_i + n_i} \right) + n_i \log \left(\frac{n_i}{p_i + n_i} \right) \right) \\
 &+ \frac{-1}{p+n+1} \left(-(p_i+1) \log \left(\frac{p_j+1}{p_j+n_j+1} \right) \right. \\
 &\quad \left. - n_j \log \left(\frac{n_j}{p_j+n_j+1} \right) \right) \\
 &= \frac{\theta + J}{p+n+1},
 \end{aligned} \tag{4}$$

where

$$\begin{aligned}
 J_1 &= p_j \log \left(\frac{p_j}{p_j + n_j} \right) + n_j \log \left(\frac{n_j}{p_j + n_j} \right), \\
 J_2 &= (p_j + 1) \log \left(\frac{n_j + 1}{p_j + n_j + 1} \right) - n_j \log \left(\frac{n_j}{p_j + n_j + 1} \right).
 \end{aligned} \tag{5}$$

After transforming the deformation formula,

$$\begin{aligned}
 J_1 &= p_j \log p_j + (p_j + n_j + 1) \log(p_j + n_j + 1), \\
 J_2 &= (p_j + n_j) \log(p_j + n_j) - (p_j + 1) \log(p_j + 1).
 \end{aligned} \tag{6}$$

After calculation, when the dynamic incremental information is added to the training data set, the change of information entropy before and after is

$$E_1 - E = \frac{\theta + J}{p+n+1} - \frac{\theta}{p+n}. \tag{7}$$

In order to evaluate the impact of dynamic incremental information on attributes A and B , it is necessary to calculate the maximum and minimum values of information entropy before and after the emergence of dynamic incremental information. According to the characteristics of functions J and $\log$, when $J = 0$, $\theta = p + n$, $E_1 - E$ takes the minimum value:

$$E_{\min} = \frac{-1}{p+n+1}. \tag{8}$$

When $J = (p_j + n_j + 1) + \log e$, $E_1 - E$ takes the maximum value:

$$E_{\max} = \frac{\log(p_j + n_j + 1) + \log e}{p+n+1}. \tag{9}$$

3.3. Analysis Process of Machine Learning Dynamic Decision Tree Algorithm in Dynamic Model. When the new dynamic data increment is added to the data set, instead of rescanning all the data and building a new model, adaptive correction is carried out on the basis of the model obtained from the previous training data by the decision tree algorithm, so as to improve the classification accuracy of the new decision tree. If x samples are dynamically added to a branch subnode of a decision tree of the previously constructed data model, the newly added x samples will meet the following inequalities according to the previous classification attributes:

$$E_{\max} - E_{\min} \leq \left(\sum_{i=1}^x \frac{\log(p_j + n_j + i) + \log e}{p+n+1} \right) + \frac{x}{p+n+1}. \tag{10}$$

If the number of newly added samples x is less than the number of samples $p_j + n_j$ data on the j attribute value segment, (10) can be scaled as

$$E_{\max} - E_{\min} \leq \frac{x \log(p_j + n_j) + 2 + \log e}{p+n+1}. \tag{11}$$

To sum up, the value of $E_1 - E$ determines the number of new samples x :

$$E_1 - E \leq \frac{x \log(p_j + n_j) + 2 + \log e}{p+n+1}. \tag{12}$$

That is,

$$x_{\max} = \frac{(p+n+1)(E_1 - E)}{\log(p_j + n_j) + 2 + \log e}. \tag{13}$$

Taking (13) as a guide, the refitting method of dynamic decision tree can be obtained. If the dynamically added data samples meet $x \leq x_{\max}$, after the data is dynamically added and updated, the entropy of the classification attribute is still less than that of the substitute attribute. In other words, the information gain calculated based on the classification attribute will still be greater than that calculated based on the substitute attribute. Therefore, the subnode classification attributes of the decision tree model remain unchanged compared with the previous classification attributes. If the number of samples x meets $x \leq x_{\max}$, the corresponding entropy of the classification attribute will become greater than that of the substitute attribute after the dynamic new data appears. At this time, the dynamic decision tree algorithm will replace the original classification attribute of the branch subnode of the decision tree with the substitute attribute, so as to update the dynamic decision tree.

3.4. Prototype Construction of Machine Learning Dynamic Decision Tree Algorithm Fusion Big Data Integration System. The machine learning dynamic decision tree algorithm fusion big data integration system needs to run in real time to complete the analysis task. Users can quantify the performance of big data according to the output behavior of discrete data dynamic fusion in the system process. The big data system designed in this research mainly includes three modules: import module, management module, and analysis module. The analysis process of graduate employment data by big data system is shown in Figure 2.

After data loading, the big data system imports the extracted data into the database of the management module to meet the data analysis of graduates' employment and entrepreneurship. The data is listed according to the data name, data type, and data length compiled by the database type, and the management module has the functions of adding, deleting, and modifying the imported data. The analysis module mainly integrates the algorithm of dynamic decision tree, which belongs to an enhanced decision tree algorithm. The underlying logic of the algorithm has been described above. In this stage, the employment simulation analysis results of senior, junior, and senior graduates are shown in Figure 3.

As can be seen from Figure 3, under the dynamic decision tree algorithm, with the increase of analysis times and data samples, the proportion of dynamic employment choice of senior, junior, and senior graduates is also increasing. The growth rate of senior is the slowest, and the growth rate of junior and senior graduates is similar, which is also in line with the actual law of employment choice. And in the data operation process of dynamic decision tree algorithm, this research uses Python software which is brilliant in the field of data processing as the development language of this part, develops the IDE, and uses the Tensorflow framework which is more open for machine learning to complete the real-time simulation analysis of the dynamic discrete data of graduates' behavior, so the simulation results are closer to the real situation. Tensorflow is an open source software library using data flow graphs for numerical calculation. Nodes represent mathematical operations in the graph, and edges in the graph represent multidimensional data arrays, namely, tensors, which are interrelated between nodes.

According to the characteristics of big data system integration, this research designs the SSM big data system framework based on MVC traditional design pattern. Its structure mainly includes DAO layer, service layer, Controller layer, and View layer. The big data system framework is used to simulate and analyze the entrepreneurship of senior, junior, and senior graduates, and the results are shown in Figure 4.

As can be seen from Figure 4, under the dynamic decision tree algorithm, with the increase of analysis times and data samples, the proportion of dynamic entrepreneurial choice of senior, junior, and senior graduates is also increasing. The growth rate of senior graduates is the slowest, and the growth rate of junior and senior graduates is similar. This is because the dynamic decision tree model uses real graduate data for in-depth learning, and senior graduates

have less entrepreneurial skills (such as business skills, financial skills, business thinking skills, and management skills), while doctoral and master's students have relatively certain technical ability and business thinking.

For complex discrete dynamic data, a big data processing system based on dynamic decision tree is designed. The designed big data system structure is simulated through trial data. The simulation uses the employment and entrepreneurship of graduates in 2021 disclosed by a municipal government in China as training data and test data. The prediction analysis results are shown in Figure 5.

It can be seen from Figure 5, under the dynamic decision tree algorithm, with the increase of analysis times, the change rules of the training data group and the test data group are different. The result jump percentage of the training data group is almost unchanged and changes slightly around 20%, but the result jump percentage of the test data group decreases gradually with the increase of analysis times (the maximum is 88.1%). It is close to 90%, the minimum is 2.1% (close to 0%), because with the increase of analysis times, the stability of test data in dynamic analysis stage becomes stronger due to random variation factors, while the training sample is because the data will not change randomly with the increase of analysis times. Therefore, the big data model proposed in this study has good accuracy and stability in the process of dynamic analysis under static data.

4. Result Analysis and Discussion

4.1. Confirmatory Experiment and Data Analysis. In the specific experimental process, the dynamic model algorithm and big data system proposed in Section 3 of this paper are used to start the experimental process. Before the experiment, it is necessary to explain various characteristic parameters in the algorithm in advance. For convenience, this study analyzes the dynamic data of graduates' employment and entrepreneurship from 2012 to 2019 published by a municipal government in China.

According to the algorithm ideas described in Section 3 of this study, this study selected more than 50 attributes of college graduates, such as education, gender, major, age, household registration category, and failure rate during school and graduation years, as the characteristic categories of the decision tree and predicted the whereabouts of college graduates in 2013 based on the employment and entrepreneurship of graduates in 2012. The study also compared with the actual situation of the whereabouts of college graduates in 2013 and the teacher data to verify the accuracy of the algorithm and then fused the dynamic data in 2013 with the base data to predict the whereabouts of college graduates in 2014 and calculate its accuracy and so on. In this study, the data generated are analyzed many times, and the results are recorded. In order to verify that the dynamic decision tree algorithm proposed in this study effectively generates new characteristic attribute branches among the subnodes of the data structure, based on the above experimental methods, this study establishes a retrieval line model system based on Tensorflow framework through the analysis module in Section 4 of this study. For the dynamic decision

FIGURE 2: The big data system's analysis process of graduate employment data.

FIGURE 3: Simulation analysis results of employment of graduates.

tree algorithm, ID3 decision tree, and C4.5 decision tree (ID3 decision tree and C4.5 decision tree are two mainstream decision algorithm models in the latest research results), three different decision tree machine learning algorithms are experimentally studied. The preliminary experimental analysis results are shown in Figure 6.

As can be seen from Figure 6, the accuracy of three different machine learning decision tree algorithms in predicting the dynamic employment and entrepreneurship of college graduates from 2013 to 2019 is different in terms of the reliability of the experimental data results. When the

number of people is less than 1000, the prediction result accuracy of ID3 decision tree algorithm is the highest. When the number of people is greater than 1000, the data result reliability of dynamic decision tree algorithm is the highest and has been maintained at a high level. The data result reliability of ID3 decision tree algorithm is the lowest, and the prediction accuracy tends to decline gradually. When the number of people is greater than 2000, the prediction accuracy of the three groups of results has decreased to varying degrees, because the dynamic decision tree algorithm has continuously generated new subbranches during the

FIGURE 4: Dynamic simulation analysis of graduates' choice of entrepreneurship based on dynamic decision tree algorithm.

FIGURE 5: The simulation analysis result of the dynamic decision tree under the training data and the test data.

experiment. The maximum bearing capacity of the experimental data has been attenuated since 1800 (that is, the data has been cleared and dynamically analyzed again), and the data collection samples have good coupling with the base data samples. It can be understood that the dynamic decision tree algorithm will test, judge, and classify the new dynamic samples under the “guidance” of the category training samples. The experimental time for all three decision trees is shown in Figure 7.

Figure 7 shows the time required to process different amounts of data during the experiment of three machine learning decision tree algorithms in the big data system. It can be seen from Figure 7, the time required by the dynamic decision tree algorithm is significantly lower than that of the other two mainstream big data analysis decision tree

algorithms. This is because the training data of the dynamic decision tree is constructed based on the addition of base data. After each data dynamic increases, the algorithm does not need to rescan all the data, but modifies the previously obtained tree branch structure, which greatly saves the high cost and low efficiency of the decision tree that the algorithm needs to rescan all the data every time the data is changed, and provides efficiency support for the large data system to analyze the characteristics of graduates' employment and entrepreneurship.

4.2. Result Analysis. The experimental results show that, on the premise of meeting the hierarchical optimization model of conceptual dynamic model, the existing decision tree

FIGURE 6: Preliminary analysis results of the experiment.

FIGURE 7: Experimental time for all three types of decision trees.

algorithm is improved by incremental optimization, and the structural analysis of graduates' employment destination is completed combined with big data system. In order to explore the employment and entrepreneurship of college graduates, this study visualized the details of the employment and entrepreneurship of graduates in a city in China from 2018 to 2020, and the calculation results are shown in Figure 8. Based on the employment and entrepreneurship data of college students from 2018 to 2020, this study calculates the whereabouts of graduates in a city in China in 2021 through the dynamic decision tree algorithm of big data system. The comparison between the predicted results and the real results is shown in Figure 9.

According to the data analysis of the experimental results in Figures 8 and 9, with the continuous growth of China's economic strength, the continuous opening of China's College Students' Entrepreneurship Policy, and the

FIGURE 8: Dynamic analysis results of employment and entrepreneurship of graduates in a certain city from 2018 to 2020.

continuous improvement of China's social system, the proportion of graduates in a city in China who choose to start their own business and enter a higher school after graduation has been increasing since 2018~2020, and the proportion of unemployed graduates after graduation has shown an overall downward trend. The overall employment choice of graduates tends to be stable and upward, which is also consistent with the actual situation of the city in 2018~2020. On the other hand, the prediction errors of graduates' employment rate, entrepreneurship rate, and other choice rate in a city in 2021 predicted by the algorithm are 2.8%, 2.6%, and 2.3%, respectively, which is very similar to the whereabouts of graduates in the first half of 2021.

FIGURE 9: Dynamic prediction and error rate analysis of the selection of college graduates in a certain city in 2021.

Therefore, the dynamic analysis model of college graduates' employment and entrepreneurship proposed in this study has high accuracy and practical application value.

5. Conclusion

How to industrialize big data technology with low cost and high efficiency has always been the industry hotspot of big data technology research. Based on this, aiming at the employment and entrepreneurship of college graduates, this study establishes a multilayer dynamic model system based on big data and machine learning dynamic decision tree. Firstly, the development status of various types of big data application technologies and the shortcomings of research progress are introduced, and a dynamic model establishment method based on big data and machine learning dynamic decision tree algorithm is proposed. Secondly, the establishment of big data technology, machine learning technology, and dynamic decision tree algorithm model is introduced, respectively. With the help of these technologies, complex discrete dynamic data can be analyzed and predicted quickly and accurately. Finally, combined with the public data set, the proposed dynamic model is analyzed. Experiments show that the big data integration system based on big data and dynamic decision tree algorithm has high result adaptability. If incremental adaptation optimization is carried out on the traditional decision tree model, the prediction effect and prediction time of dynamic data can be significantly improved, which puts forward theoretical support for the industrialization and social significance of big data technology. This method expands the attributes in the form of dichotomy and only pays attention to the construction of logical attributes in each subbranch of the dynamic model. Therefore, the dynamic model established by this method may lead to too deep decision tree, which needs to be further studied.

Data Availability

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

Conflicts of Interest

The author declares no conflicts of interest regarding this work.

Acknowledgments

The study was funded by the study on the Influence of the Employment and Entrepreneurship of Graduates under the Epidemic Situation of Pneumonia in COVID-19 (JYB2020302).

References

- [1] K. A. Brown, S. Brittman, N. Maccaferri, D. Jariwala, and U. Celano, "Machine learning in n : big data at small scales," *Nano Letters*, vol. 20, no. 1, pp. 2–10, 2020.
- [2] J. Wang, S.-C. Li, and P.-C. Lin, "A psychophysical and questionnaire investigation on the spatial disorientation triggered by cockpit layout and design," *International Journal of Industrial Ergonomics*, vol. 72, pp. 347–353, 2019.
- [3] E. Jouanjus, C. Sans-Pola, B. Mainoli et al., "Establishing and evaluating a study questionnaire on knowledge and attitudes of healthcare professionals towards recreational and medical cannabis across europe," *Clinical Drug Investigation*, vol. 41, no. 8, pp. 701–710, 2021.
- [4] A. Kamei, Y. Nakajima, T. Nakano, and K. Takeyama, "Raising spaces for animals in the current elementary school system," *Journal of Environmental Engineering*, vol. 84, no. 757, pp. 357–366, 2019.
- [5] A. Prochazka, O. Vysata, and V. Marik, "Integrating the role of computational intelligence and digital signal processing in education: emerging technologies and mathematical tools," *IEEE Signal Processing Magazine*, vol. 38, no. 3, pp. 154–162, 2021.
- [6] Y.-T. Chen, E. W. Sun, and Y.-B. Lin, "Coherent quality management for big data systems: a dynamic approach for stochastic time consistency," *Annals of Operations Research*, vol. 277, no. 1, pp. 3–32, 2019.
- [7] A. S. Elsayad, A. I. E. Desouky, M. M. Salem, and M. Badawy, "A deep learning H₂O framework for emergency prediction in biomedical big data," *IEEE Access*, vol. 8, pp. 97231–97242, 2020.
- [8] V. S. Dommaraju, K. Nathani, U. Tariq et al., "ECMCRR-MPDNL for cellular network traffic prediction with big data," *IEEE Access*, vol. 8, pp. 113419–113428, 2020.
- [9] N. B. Yahia, J. Hlel, and R. Colomo-Palacios, "From big data to deep data to support people analytics for employee attrition prediction," *IEEE Access*, vol. 9, no. 99, p. 1, 2021.
- [10] G. Gui, F. Liu, J. Sun, J. Yang, Z. Zhou, and D. Zhao, "Flight delay prediction based on aviation big data and machine learning," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 1, pp. 140–150, 2020.
- [11] R. M. A. Latif, S. B. Brahim, S. Saeed, L. B. Imran, M. Sadiq, and M. Farhan, "Integration of Google play content and frost prediction using CNN: scalable IoT framework for big data," *IEEE Access*, vol. 8, pp. 6890–6900, 2020.

- [12] L. V. van Dijk, K. A. Wahid, S. Ahmed et al., "Big data statistical learning improves survival prediction for head and neck cancer patients," *International Journal of Radiation Oncology, Biology, Physics*, vol. 108, no. 3, pp. e796–e797, 2020.
- [13] Z. Xiao, X. Fu, L. Zhang et al., "Big data driven vessel trajectory and navigating state prediction with adaptive learning, motion modeling and particle filtering techniques," *IEEE Transactions on Intelligent Transportation Systems*, vol. 2020, pp. 1–14, 2020.
- [14] X. Liu, Q. Sun, W. Lu, C. Wu, and H. Ding, "Big-data-based intelligent spectrum sensing for heterogeneous spectrum communications in 5G," *IEEE Wireless Communications*, vol. 27, no. 5, pp. 67–73, 2020.
- [15] I. B. R. Nogueira, M. A. F. Martins, M. J. Regufe, A. E. Rodrigues, J. M. Loureiro, and A. M. Ribeiro, "Big data-based optimization of a pressure swing adsorption unit for syngas purification: on mapping uncertainties from a meta-heuristic technique," *Industrial and Engineering Chemistry Research*, vol. 59, no. 31, pp. 14037–14047, 2020.
- [16] D. Fnyes, B. Nmeh, and P. Gspr, "Design of LPV control for autonomous vehicles using the contributions of big data analysis," *International Journal of Control*, vol. 2021, pp. 1–23, 2021.
- [17] R. Skuratovskii, A. Bazarna, and E. Osadhy, "Analysis of speech MEL scale and its classification as big data by parameterized KNN," *Artificial Intelligence*, vol. 26, no. 1, pp. 42–57, 2021.
- [18] K. Lu, J. Liu, X. Zhou, and B. Han, "A review of big data applications in urban transit systems," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 5, pp. 1–18, 2020.
- [19] M. Bracher-Smith, K. Crawford, and V. Escott-Price, "Machine learning for genetic prediction of psychiatric disorders: a systematic review," *Molecular Psychiatry*, vol. 26, no. 1, pp. 70–79, 2020.
- [20] C. Salehnasab, A. Hajifathali, F. Asadi, E. Roshandel, A. Kazemi, and A. Roshanpoor, "Machine learning classification algorithms to predict aGvHD following allo-HSCT: a systematic review," *Methods of Information in Medicine*, vol. 58, no. 6, pp. 205–212, 2019.
- [21] K. Ahn, K. Lee, H. Kim et al., "VP40.02: machine learning approaches to the prediction of newborn's body mass index using nationwide multicentre data," *Ultrasound in Obstetrics and Gynecology*, vol. 56, no. S1, p. 229, 2020.
- [22] L. Heikkinen, M. Äijälä, K. R. Daellenbach et al., "Eight years of sub-micrometre organic aerosol composition data from the boreal forest characterized using a machine-learning approach," *Atmospheric Chemistry and Physics*, vol. 21, no. 13, pp. 10081–10109, 2021.
- [23] M. Raissi and G. E. Karniadakis, "Hidden physics models: machine learning of nonlinear partial differential equations," *Journal of Computational Physics*, vol. 357, pp. 125–141, 2018.
- [24] K. T. Butler, D. W. Davies, H. Cartwright, O. Isayev, and A. Walsh, "Machine learning for molecular and materials science," *Nature*, vol. 559, no. 7715, pp. 547–555, 2018.
- [25] J. Meyers, B. Fabian, and N. Brown, "De novo molecular design and generative models," *Drug Discovery Today*, vol. 26, no. 11, pp. 2707–2715, 2021.