Research Article

Evaluation and Analysis of the Informatization Degree of College English Education Based on Big Data Technology

Zongying Wang

School of Foreign Languages, East China University of Technology, Nanchang 330013, China

Correspondence should be addressed to Zongying Wang; zywang@ecut.edu.cn

Received 29 June 2022; Revised 19 July 2022; Accepted 29 July 2022; Published 21 August 2022

Academic Editor: Hangjun Che

Copyright © 2022 Zongying Wang. 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.

Since the 21st century, thanks to the continuous achievements in the computer field, the Internet technology has developed rapidly in just a few years until it has swept the world. In the torrent of the information age, big data technology came into being. The application of big data technology is extremely high, and it is widely used in many industries. Based on the excellent performance of big data technology in the education industry, this paper will discuss the application of big data technology in the evaluation and analysis of college English education informatization. Under the condition that the evaluation standard is relatively abstract, the traditional method of evaluating through the subjective perspective of human beings has a large interference surface, and it is difficult to make the evaluation work substantive, data-based, and fair, which directly leads to evaluation and analysis work being difficult to carry out. Based on the abovementioned situation that the evaluation and analysis are relatively limited, this article will introduce the role of the big data technology algorithm model in the evaluation and analysis work. Through big data technology algorithms, qualitative problems that are difficult to evaluate are converted into quantitative problems that are easier to analyze and compare and then get rid of the limitations faced by traditional evaluation algorithms. The interference caused by human subjective factors to evaluation and analysis should be reasonably avoided as much as possible. This makes the evaluation and analysis results more objective and fair. This article is mainly based on the relevant algorithms under the big data technology to mine, classify, and purify the data with strong correlation with the important relevant indicators that affect the degree of college English informatization and carry out the experimental calculation of the evaluation of the relevant indicators. The results are as follows: we provide a reliable basis for the evaluation of the degree of informatization in college English and then conduct more scientific and rigorous evaluation and analysis. In the future development of big data technology, the database will be continuously supplemented with valuable teaching data, so that the results of big data technical analysis are more and more consistent with the real situation.

1. Introduction

The logistic regression model algorithm based on big data technology [1–4] determines the quantitative relationship between two or more variables that depend on each other, so as to find the data with strong correlation with the evaluation index. Complete the preliminary classification of the data and then use data mining. The technology [5–8] mines the data with the same or strong correlation with the data feature information in the same database by comparing the data feature information, so as to achieve the preliminary mining classification of the data, so as to facilitate the subsequent data cleaning, data changes, and data purification. Taking the above as the core idea, try to establish an evaluation system for the degree of informatization in college English education. To a certain extent, put aside the interference of human subjective consciousness and use data and evaluation indicators to establish a more objective evaluation behavior. Based on such prerequisites, it is necessary for the algorithm model and the evaluation system we established to be regarded as systems with two correlations. Using the method of data mining classification, the data related to the indicators affecting the evaluation are classified into the database [9–13], and then to a certain extent, the data related to the evaluation indicators are classified. In order to transform it into a geometric mathematical model with high flexibility to
deal with the rigid conditions of lack of transformation ability, the unnatural conditions of the mathematical discussion form are combined with the mathematical model, and finally, the combination of classical numbers and shapes is used to form images [14, 15]. A more intuitive performance evaluation system is with strong correlation indicators that affect the evaluation results. Standardize the evaluation system and make it fair. Numbers are the most concise and powerful language of reality, and the mathematical expression of everything is true, valid, and concise enough, provided the result is correct. The algorithm optimization and more scientific improvement under the big data technology are also worth looking forward to and paying attention to in the future.

Downtime is a deadlock situation that is very likely to occur in computer computing, and the possibility of server database deadlock is not ruled out. For this, we have considered setting up a framework that includes alerts and monitoring. In the event of downtime, our alert monitoring framework can detect and diagnose problems in a timely manner, reducing the possibility of data loss.

2. Evaluation and Analysis of College English Education Informatization Degree under Big Data Technology

This paper mainly evaluates and analyzes the informatization degree of college English education under the big data technology, while the data mining technology based on the big data technology mainly collects and integrates various and complicated data information and then obtains more accurate and representative data information so that we can more objectively evaluate and analyze the degree of university education informatization [16–19]. Firstly, we analyze the ideas according to the data mining technology and construct the idea map. We construct the five stages in the big data mining process. Based on the big data mining technology, we construct the model and its algorithm and then evaluate the informatization degree of college English education, as shown in Figure 1.

At this stage, data mining is done through big data technology, and the generated model can be used later to solve more complicated problems.

In order to avoid security vulnerabilities as much as possible, first of all, we need to understand which security vulnerabilities are most likely to cause security threats to the database. The first and most direct one is that the username and password in the database are too simple, which leads to some malicious hackers. It is easy to steal user information from our database, leading to security breaches, followed by unpatched databases, insufficient authentication, and other related issues. In response to these problems, we have targeted management personnel awareness, systems, and technical means and follow the basic threat prevention guidelines, but if only from a technical perspective, we will consider the use of monitoring (DMI) system, which is so-called database auditing systems, to circumvent security breaches.

2.1. Big Data Technology Mining Learning Model and Its Algorithm

2.1.1. Logistic Regression Model. This paper mainly discusses the evaluation of the informatization degree of college English education based on the big data technology algorithm. Here, we introduce the logistic regression model algorithm to calculate the correlation strength of the correlation data with the relevant indicators and then make a preliminary classification of our data. We use this model to fit the data, and the logistic regression we use is normalized on the basis of linear regression [20–22]. It minimizes the difference between the data value predicted by the model and related algorithms and the real value.

The logistic regression model algorithm mainly multiplies each attribute of the data sample participating in the experiment by the corresponding parameter value and accumulates the results we get. The formula of its model is

\[ h(x) = w_0 x_0 + w_1 x_1 + w_2 x_2 + \ldots + w_m x_m. \]  

(1)

The vectorized formula is expressed as

\[ h(x) = w^T x. \]  

(2)

Calculate the value of the sigmoid function as follows. Substitute the calculated result of the above formula into the sigmoid function. The calculated result obtained by the function calculation will be between (0, 1). The calculated value is compared with the set threshold value, which is greater when the threshold value is positive class; otherwise it is negative class, and its calculation formula is as follows:
2.1.2. Model Calculation. Here, we assume that \( n \) samples are used for calculation training. It is known that the probability of occurrence of each assumed sample conforms to the Bernoulli distribution, and the probability of occurrence is calculated experimentally for each sample.

Calculate the probability of occurrence of positive and negative classes:

\[
p(y_i = 1 \mid x_i),
\]

\[
1 - p(y_i = 1 \mid x_i).
\]

(4)

Calculate the posterior probability of each sample:

\[
p(y \mid x, w) = p(y_i = 1 \mid x_j) y_i^i (1 - p(y_i = 1 \mid x_j))^{1-y_i}.
\]

(5)

The computational model experiment in the article is a special computational experiment, and its computation is nothing more than two results, either success or failure, and each experimental sample exists independently and is not disturbed, and each experiment has a fixed value. The probability of success \( p \) and then the probability of occurrence of the experimental calculation sample are assumed to conform to the Bernoulli distribution, which not only simulates the real situation of the model calculation to a great extent, but also facilitates the calculation of the mathematical expectation and variance of the distribution due to the simple distribution of the results.

2.2. Log-Likelihood Function. Due to the extremely large amount of data, in our calculation process, the model and data overfitting state will inevitably occur. In order to avoid this problem, here we introduce the loss function \( l(w) \), by adding the loss function \( l(w) \) plus a penalty term for \( w \), making the penalty a Regularizer. Its calculation formula is as follows:

\[
l(w) = \prod_{i=1}^{m} p(y_i = 1 \mid x_j)^{y_i} (1 - p(y_i = 1 \mid x_j))^{1-y_i}.
\]

(6)

Expand and solve for it and take the derivative of \( w \)

\[
\frac{\partial l(w)}{\partial w} = \sum_{i=1}^{m} (y_i - g(z))x_i.
\]

(7)

2.3. Naive Bayes Algorithm. Naive Bayesian model originated from classical mathematical theory. Its stable classification efficiency and simultaneous multitask processing, especially when the amount of data information is huge, greatly improve the efficiency of classification and sorting of our data mining information. The function model is as follows:

\[
C_x = \arg \max_{k \in \{1, 2, \ldots, K\}} \left( P(C_K) \prod_{j=1}^{n} p(x_j \mid C_K) \right).
\]

(8)

The above formula is calculated, and the probability is estimated by frequency. The calculation formula is as follows:

\[
p(C_K) = \frac{m_k}{m}.
\]

(9)

Here, we make reasonable assumptions about the distribution of data characteristics of the samples and calculate separately.

The naive Bayes model that conforms to the multinomial distribution is calculated as follows:

\[
p(x_j \mid C_k) = \frac{m_{kj}}{m_k}.
\]

(10)

Sometimes if the value of a feature in the sample is 0, it will seriously affect the probability distribution of the feature, so we use Laplace smoothing to avoid this situation, namely,

\[
p(x_j \mid C_k) = \frac{(m_{kj} + \lambda)}{(m_k + \lambda)}.
\]

(11)

For the Naive Bayes model conforming to the Bernoulli distribution, the formula is as follows:

\[
p(x_j \mid C_k) = \begin{cases} p(x_j = 0 \mid C_k) = 1 - p(x_j = 1 \mid C_k), \\ p(x_j = 1 \mid C_k). \end{cases}
\]

(12)

The naive Bayes model that conforms to the Gaussian distribution is calculated as follows:

\[
f(x, \mu, \sigma) = \frac{1}{\sigma \sqrt{2\pi}} \exp \left( -\frac{(x - \mu)^2}{2\sigma^2} \right).
\]

(13)

2.4. Decision Tree Model and Its Data Purification. After the data mining is collected, the data will be sorted and summarized to get a database with a huge amount of information. In our database, the invalid information we have collected is often retained. At this time, the introduction of the decision tree model can effectively solve the data (purity issues). Decision tree model is mainly a nonparametric classifier that is simple to use and less difficult to operate. Here, we refer to the ID3 algorithm, as well as the C4.5 algorithm.

The commonly used algorithms in the decision tree model [23] mainly include the ID3 algorithm and the C4.5 algorithm. These two algorithms can be used to divide the data set, and the ultimate goal of the decision tree node splitting is to make the nodes that fall on each branch node. The samples are in the same category to the greatest extent possible, which means that the node purity is higher.

A decision tree is a tree structure consisting of nodes and directed edges. Its essence is a set of causal rules. The decision tree model we introduced in this article is a simple and easy-to-use nonparametric classifier that does not require any assumptions on the data.
2.4.1. ID3 Algorithm. Aiming at the problem of data purity after data mining classification, we introduce the concept of information entropy [24] to measure the data purity after classification. Its calculation formula is as follows:

\[ H = - \sum_{k=1}^{L^y} p_k \log_2 p_k. \]  (14)

\( H \) represents the information entropy \( D \), and the smaller the calculated \( H \) value, the higher the purity of the information entropy.

The impact of data purity on the results is undoubtedly the most direct. It is undeniable that such a problem does exist in this article. However, based on the diversity and quantity of data types, it is difficult to formulate a unified, specific, and standardized measurement system for the measurement of the purity of multiple types of data. Therefore, in this article, by purifying the data multiple times, the information entropy value is reduced as much as possible, and the impact of data purity on the results is minimized.

The ID3 algorithm quoted here conforms to the data gain criterion. The so-called data gain is the positive change of the original data and the classified data after the data is classified, and the gain of an indicator after the data classification is the data set. The difference between the information entropy and the empirical conditional entropy of \( D \) under this condition is

\[ H(D, a) = H - \sum_{v=1}^{v} \frac{|D^v|}{|D|} H(D^v). \]  (15)

3. Algorithm Improvement

In summary, we have made a preliminary framework for the big data technology algorithm, but there may still be deficiencies or loopholes. Next, we will optimize and improve our big data counting algorithm to improve the computing power and accuracy of the algorithm.

3.1. Logistic Regression Model Algorithm Optimization. In the logistic regression algorithm, it is not excluded that the last derivative is 0. In this case, we cannot solve \( w \), and we need to use the gradient iterative optimization algorithm to optimize the algorithm. The combination of stochastic gradient descent and batch gradient descent and derivation of the objective function can help us solve the above problems. Here, we use the batch gradient descent method to find the partial derivative of \( w \) and get the gradient corresponding to each \( w \). The calculation formula is as follows:

\[ \frac{\partial l(w)}{\partial w_j} = \frac{1}{m} \sum_{i=1}^{m} (y_i - h_w(x_i)) x_{ij}. \]  (16)

Since we need to minimize the risk function, we need to update \( w \) in the negative direction of \( w \). The calculation formula is as follows:

\[ w^*_j = w_j + \frac{1}{m} \left( y^j - h_w(x^j) \right) x^j. \]  (17)

By calculating the formula, we will finally get a comprehensive optimal solution, but every change of \( w \) requires all the training data. If the data is too large, it will greatly affect the speed of \( w \) change, so we use randomly the gradient descent method which is combined, and the calculation formula is as follows:

\[ l(w) = \frac{1}{2m} \sum_{i=1}^{m} \left( y_i - h_w(x_i) \right)^2 = \frac{1}{m} \sum_{i=1}^{m} \text{cost}(w_i, (x^i, y^i)), \]  (18)

\[ \text{cost}(w_i, (x^i, y^i)) = \frac{1}{2} \left( y^i - h_w(x^i) \right)^2. \]

Obtain the corresponding gradient by taking the partial derivative through the loss function of each sample, and then update \( w \):

\[ w^*_j = w_j + \left( y^j - h_w(x^j) \right) x^j. \]  (19)

3.2. Improvement of Naive Bayes Algorithm. Among the three classification algorithms based on the naive Bayes algorithm, the best classification effect is the multinomial naive Bayes algorithm classification model, but the disadvantage is that the algorithm automatically defaults to the same weight of all features and ignores the features of data. To a certain extent, it will reduce the accuracy of our data classification. Therefore, we need to combine other algorithms for related optimization. Here, we introduce the TDF-IDF algorithm and improve and optimize the original algorithm before applying it to the data processing module.

Combine the improved TDF-IDF-LD into a multinomial naive Bayesian algorithm to get the final formula:

\[ C_x = \arg \max_{k \in \{1,2,3\}} \left( \frac{m_k}{m} \prod_{j=1}^{n} \frac{m_{kj} + \lambda}{m_{kj} + s \lambda} \right) \ast \text{TF} \ast \text{IDF} \ast \text{LF} \ast \text{DTF}. \]  (20)

3.3. Improvement of Decision Tree Model Algorithm. The main idea of the ID3 algorithm is a top-down greedy strategy from the root node to the leaf node. First, the information gain of each feature is calculated according to the above formula, and finally the feature with the largest information gain is selected as the node of the decision tree. Splitting further improves the purity of the child nodes of the decision tree, and the ability to divide samples into corresponding categories is stronger, and the representativeness of such features is stronger. However, the shortcomings of ID3 are also obvious. The algorithm has a preference for attributes with a large number of values. The decision tree created only by the ID3 algorithm obviously cannot achieve the expected effect for unknown data. At this time, we use the C4.5
algorithm for collaborative calculation, so that the decision tree we create is sufficiently convincing [25].

3.3.1. C4.5 Algorithm. In view of the limitations of the ID3 algorithm, to minimize the adverse effects of the ID3 algorithm, we use the C4.5 algorithm for collaborative calculation. For attribute preference problem, the calculation formula is

\[ H(D, a)_{\text{ratio}} = \frac{H(D, a)}{H(D)}. \]  

(21)

3.3.2. Classification and Regression Tree Algorithm. Classification and regression trees are a type of decision tree and are very important. The generation of classification tree and regression tree can be realized at the same time. The CART algorithm we introduce here is a binary recursive segmentation technique. The internal node of the generated decision tree has only two branches and only two categories of yes or no. Even if a feature or attribute has multiple values, it is divided into two parts.

Create a classification tree: in the recursive process of creating the split tree, the CART algorithm selects the feature with the smallest Gini index in the current data set as the node to divide the decision tree. The Gini index is similar to the information entropy and is usually used to measure the purity of the data set \( D \). The calculation formula is as follows:

\[ \text{Gini}(D) = 1 - \sum_{k=1}^{[y]} p_k^2. \]  

(22)

The Gini index is obtained by calculation, and the purity is estimated by observing the value of the final calculation result of the index. The smaller the value, the higher the purity.

In the process of classifying data, the Gini index is calculated for the indicator \( a \), and the formula is as follows:

\[ \text{Gini}_{\text{index}}(D, a) = \sum_{v=1}^{D} \frac{|D_v|}{D} \text{Gini}(D^v). \]  

(23)

Create a regression tree: the regression tree created by CART uses the principle of least mean square variance to determine the optimal division of the regression tree, so that our final data result prediction is closest to the true value. Assuming that the mean squared error is calculated for a feature, and the feature with the smallest error is found, it is theoretically the optimal splitting point. The squared error formula is

\[ \sum_{x \in D} (y_i - f(x_i))^2. \]  

(24)

In this article, we only quoted a part of the cart algorithm, and we used the Gini index algorithm to improve the other part. Compared with the original algorithm, the GIN index algorithm is simpler and faster. The original algorithm is added as follows:

\[ R_1(j, s) = \{ x \mid x^j \leq s \}, \]

\[ R_2(j, s) = \{ x \mid x^j > s \}, \]

\[ c_m = \frac{1}{N_m} \sum_{x \in R_m(s)} y^j, x \in R_m, m = \{1, 2\}. \]  

(25)

4. Experiment and Test Based on Big Data Technology Algorithm Evaluation

This article mainly evaluates and analyzes the informatization degree of college English education based on big data technology. Here, we will consider the selection of student learning environment, information equipment construction, teaching equipment, voice environment, and other indicators to use our proposed big data technology algorithm for experimental testing.

4.1. Feasibility Experiment Test of Big Data Technology Algorithm. Since the degree of feasibility of the algorithm proposed in our article is still unknown, in order to ensure that the test results obtained by our subsequent experimental tests based on this algorithm are more credible, we first test the feasibility of the algorithm. Taking a university as an example, we will conduct experiments on the hardware facilities of a university based on the traditional evaluation method of the results calculated by the big data counting algorithm and compare and analyze the obtained results. The comparison results are shown in Table 1, and the visualization is shown in Figure 2.

Through our big data counting algorithm and artificial traditional algorithm to calculate some indicators of a university’s hardware facilities construction, it can be clearly observed that the results obtained by our big data algorithm are highly consistent with the results obtained by traditional artificial algorithms, and there are slight differences. We can think that our proposed big data technology algorithm is still feasible in the experiment, and the high accuracy is in line with the real situation. You can continue to participate in follow-up testing.

4.2. Big Data Technology Algorithm Optimization Test. This paper mainly evaluates and analyzes the degree of college English informatization, which requires us to take into account the important indicators that affect the degree of college English informatization, such as college English education funding, English e-book resources, and student learning environment. Use our big data technology algorithm to quantitatively analyze the important influencing factor indicators, and compare the experimental results obtained with the real situation to reflect the advantages of the optimization algorithm.

4.2.1. C4.5 Algorithm and ID3 Algorithm Test. In the previous article, based on the ID3 algorithm introduced by data processing and purification, we further optimized it and
introduced the C4.5 algorithm. We compared the advantages brought by the improvement of this algorithm through a more direct experimental test. Taking a university as an example, we use two algorithms to calculate the university’s investment in English education and compare it with the actual situation, and then we compare the advantages and disadvantages of the two algorithms. The calculation results are shown in Table 2, and the visualization is shown in Figure 3.

According to the visualization in the above figure, the data results obtained by our optimized C4.5 algorithm compared to the ID3 algorithm after data purification are compared with the actual investment of the university, and the optimized C4.5 algorithm is closer to the real situation of the experiment.

4.2.2. Classification Tree and Regression Tree Algorithm Test. In this article, we have created the classification tree and regression tree model algorithms, respectively. Here, we will use the collection of English e-books in the university library as an indicator to use the two algorithms for experimental comparison. The experimental results are shown in Table 3. The visualization is shown in Figure 4.

According to the visual display of the experimental results in the figure, there are differences within a certain range between the two algorithms we use, respectively, but the degree of error is extremely small and can be basically ignored. Both algorithms have excellent data processing and analysis capabilities.

4.2.3. Simulation Test of Unit Computation Time Efficiency of the Stochastic Gradient Descent Method. Since the amount of data is too large, we need to test the time-consuming comparison of different algorithms when processing the same unit capacity of data. Here, we introduce the control variable method to fix the difference of irrelevant variables brought by the experimental equipment and only change the type of the algorithm and carry out the experiment with a single variable. As a simulation test, taking the time efficiency as the evaluation standard for the calculation results reflects the simplicity of the algorithm and the simplification of complex problems. The test result data is shown in Table 4, and the visual chart is shown in Figure 5.

Considering the efficiency of the three algorithms used in this paper for data processing, we will transform qualitative problems into quantitative problems for experimental testing. The evaluation results of qualitative understanding are presented in the form of relatively intuitive data. The time consumption remains high in an advantageous state. In similar algorithms, the measurement to be measured is simplified, and the simplified data to be measured is calculated, thereby reducing the time consumption of the algorithm. Minimize the time consumption of the algorithm due to the huge amount of data, and reduce the algorithm speed.

In the article, only some indicators that affect the degree of informatization in college English education are tested and evaluation standards are formulated, but in fact, there are many factors that affect the degree of informatization in college English education. There are many indicators for
Table 2: Comparison of the results of ID3 algorithm and C4.5 calculation of investment capital (unit: ten thousand yuan).

|                        | Hire an English teacher | English teaching equipment | English professional equipment | E-books in English | Other |
|------------------------|-------------------------|----------------------------|--------------------------------|-------------------|-------|
| ID3 algorithm calculation input | 22.3                    | 72.2                       | 88.7                           | 11.4              | 7.8   |
| C4.5 algorithmic computing input | 21.9                    | 74.8                       | 90.6                           | 10.1              | 7.5   |
| Actual input           | 21.5                    | 74.3                       | 90.2                           | 10.5              | 7.2   |

Figure 3: Comparison and visualization of the results of the two algorithms.

Table 3: Library’s English e-book collection data.

|                      | English reading and writing | English listening and speaking | English extracurricular readings | Other literature |
|----------------------|-----------------------------|--------------------------------|---------------------------------|------------------|
| Classification tree algorithm | 1648                        | 2071                           | 3011                            | 1024             |
| Regression tree algorithm | 1692                        | 2104                           | 3048                            | 1133             |

Figure 4: Visualization of e-book collections.
reference. In this article, we only select a few of the more important reference indicators for testing. But that does not mean we are turning a deaf ear to other influencing factors. The final evaluation and analysis results must be discussed under the influence of various factors. Due to the similarity of the calculation methods, they are not tested and displayed in this paper.

4.3. Evaluation Experiment of College English Education Informatization Degree Based on Big Data Technology Algorithm. In summary, after the feasibility and optimization tests of our algorithm have been carried out, we use the algorithm to conduct experiments on the evaluation and analysis of college English education informatization. Here, we select three indicators, learning environment, English e-books, and language environment, to evaluate the degree of English informatization in a university.

4.3.1. Learning Environment Assessment Experiment. Aiming at the pros and cons of students’ learning environment, it can reflect the degree of college English teaching informatization to a certain extent. Taking a university as an example, we use the big data technology algorithm proposed by us to make the evaluation criteria for students’ different learning environments evaluation and compared it with human subjective evaluation. The experimental results are shown in Table 5, and the visualization is shown in Figure 6.

To a certain extent, the evaluation of students’ learning environment in school reflects the degree of informatization of the university’s learning environment construction. There is a certain difference between the experimental results obtained using our big data technology and the human subjective evaluation. However, considering that human subjective evaluation is influenced by human subjective psychological differences, the error brought about by the evaluation results is relatively large. We can think that the evaluation results made by big data algorithms are fairer and more accurate.

4.4. Derivative Simulation of Similar Algorithms. The big data technology algorithm needs to perform weighted calculation on the evaluation index factor set when mining the correlation data of participation evaluation indicators, which brings certain challenges to the derivative performance of the algorithm. The ability to derive a single impact factor into multiple impact factors is a necessary ability of the algorithm when the conditions for participating in the evaluation in the actual calculation process are insufficient or the conditions for participating in the evaluation are single. The starting point of this simulation is to derive the number of participating factors after the basic evaluation factor is tested by the algorithm. The evaluation factor data of similar algorithm tests are shown in Table 6, and the data visualization chart is shown in Figure 7.

When the big data technology algorithm explored in this article has a single or insufficient condition among similar algorithms, the ability to derive the known condition is twice that of the original condition. Compared with similar algorithms under the same conditions, the derivation ability of the participating factors is only about 1.5 times of the
original factor. When faced with the actual situation, the evaluation factors often appear in the order of ten. The stronger the derivative ability of the factors participating in the evaluation, the more comprehensive the expanded evaluation factor set, which provides the accuracy of the big data technology algorithm in this article (data help).

Because the existing big data technology has a relatively complete system, and combined with the experimental tests made in this article, the big data algorithm has been able to process data more flexibly and build a more perfect evaluation system accordingly. Then, in the future development of big data technology, it is worth paying attention to the

| Basic evaluation factor | Big data evaluation method | Factor analysis evaluation method | Subjective assessment |
|-------------------------|---------------------------|----------------------------------|-----------------------|
| 10                      | 21                        | 18                               | 16                    |
| 15                      | 37                        | 36                               | 34                    |
| 25                      | 58                        | 54                               | 51                    |

Table 6: Algorithm condition derived performance data table.
improvement of data accuracy and data processing speed. It will bring more convenient and substantial help to the education industry.

5. Conclusion

With the rapid development of big data technology, there are many examples of big data technology being applied to the education industry. It has gradually evolved into a major trend. After years of technical improvement and optimization, a large amount of valuable teaching data in the database has provided theoretical basis and practical cases for big data mining technology. The algorithm deduced based on big data technology is introduced into the evaluation and analysis of the informatization degree of college English education. Mining and technical algorithms combine qualitative subjective events into the scope of mathematics to carry out rigorous quantitative analysis. In our simulation experiments, we mainly study the rational application of big data technology algorithms to the evaluation and analysis of the degree of informatization in college English education performance. If and only when we transform qualitative problems into quantitative problems under big data technology and quantify useless parameters for evaluation, we cleverly use the C4.5 algorithm to eliminate useless parameters and ensure the parameters are involved in evaluation and analysis. All of them have a strong correlation with the informatization of college English education, so that the evaluation results are more scientific and rigorous, close to the real situation. To sum up, the big data technology is enough to provide substantial help to the evaluation and analysis personnel in the evaluation of the informatization degree of college English education, and it also meets the requirements of the evaluation system for quantitative algorithms.

First of all, considering the technical difficulty, most data mining tools are user-friendly, easy to understand, and easy to use, which greatly reduces the difficulty for analysts or industry evaluators to mine value from massive data. Secondly, data mining technology is the product of countless experiments and is widely recognized and accepted by everyone. It can clean, calculate, and visualize data through various built-in programs and realize automatic management and control of multitasking, which can significantly reduce the user’s time and cost and reduce the workload, and provide substantial help for analysts and evaluators.

Data Availability

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

Conflicts of Interest

The author declares no conflicts of interest regarding this work.

Acknowledgments

This work was sponsored in part by 14th Five-Year Plan of Jiangxi Education Science (21YB082) and Fuzhou Social Science (20SK20).

References

[1] A. Mcafee and E. Brynjolfsson, “Big data: the management revolution,” Harvard Business Review, vol. 90, no. 10, pp. 60–128, 2012.
[2] H. R. Varian, “Big data: new tricks for econometrics,” The Journal of Economic Perspectives, vol. 28, no. 2, pp. 3–28, 2014.
[3] D. Howe, M. Costanzo, P. Fey et al., “Big data: the future of biocuration,” Nature, vol. 455, pp. 7209, pp. 47–50, 2008.
[4] V. Marx, “Biolog: the big challenges of big data,” Nature, vol. 498, no. 7453, pp. 255–260, 2013.
[5] Z. Zhu, “Application of data mining technology in the information technology of college English teaching,” Advances in Journal of Food Science and Technology, vol. 5, no. 7, pp. 969–975, 2013.
[6] Y. R. Shiau, C. H. Tsai, and Y. H. Hung, “The application of data mining technology to build a forecasting model for classification of road traffic accidents,” Mathematical Problems in Engineering, vol. 2015, pp. 170635.1–170635.8, 2015.
[7] H. Tang, J. Zhou, D. Liao, and R. Liu, “Research on data mining and visualization technology for numerical simulation data based on the PDM platform,” Journal of Computational and Theoretical Nanoscience, vol. 9, no. 9, pp. 1193–1199, 2012.
[8] G. Ji, “The research of decision tree learning algorithm in technology of data mining classification,” Journal of Convergence Information Technology, vol. 7, no. 10, pp. 216–223, 2012.
[9] S. F. Altschul, T. L. Madden, and A. A. Schaffer, “Gapped BLAST and PSI-BLAST: a new generation of protein database search programs,” Nucleic Acids Research, vol. 25, no. 17, pp. 3389–3402, 1997.
[10] S. Altschul, “Gapped BLAST and PSI-BLAST: a new generation of protein database search programs,” Nucleic Acids Research, vol. 25, no. 17, pp. 3389–3402, 1997.
[11] S. T. Sherry and M. Kholodov, “dbSNP: the NCBI database of genetic variation,” Nucleic Acids Research, vol. 29, no. 1, pp. 308–311, 2001.
[12] K. Hofmann, P. Bucher, L. Falquet, and A. Bairoch, “The PROSITE database, its status in 1999,” Nucleic Acids Research, vol. 27, no. 1, pp. 215–219, 1999.
[13] K. Higo, Y. Ugawa, M. Iwamoto, and T. Korenaga, “Plant cis-acting regulatory DNA elements (PLACE) database: 1999,” Nucleic Acids Research, vol. 27, no. 1, pp. 297–300, 1999.
[14] M. Marcotte, C. J. Troupin, and M. Le Maguer, “Mass transfer in cellular tissues. Part I: the mathematical model,” Journal of Food Engineering, vol. 13, no. 3, pp. 199–220, 1991.
[15] A. Wilson, “The mathematical model,” Canadian Medical Association Journal: Canadian Medical Association journal = journal de l’Association medicale canadienne, vol. 139, no. 6, p. 479, 1971.
[16] W. Zheng, “Comparative analysis of the informatization degree of regional economy,” Research on Quantitative Economy and Technology Economy, vol. 1, pp. 27–29, 2001.
[17] F. Chen, “Discussion on gray evaluation of enterprise management informatization,” China Management Information: Comprehensive Edition, vol. 9, no. 2, pp. 2, 2006.
[18] J. Zheng, L. Qi, and Y. Qi, “A correlation analysis between the degree of informatization and economic development in my country,” Journal of Harbin University of Science and Technology, vol. 7, no. 4, p. 4, 2002.
[19] W. Yao, "Measurement of social informatization degree in Jilin province and its influence on economic development [J]," Information Science, vol. 19, no. 012, pp. 1335-1336, 2001.
[20] H. Wang and J. Meng, "Predictive modeling method of multiple linear regression [J]," Journal of Beijing University of Aeronautics and Astronautics, vol. 33, no. 4, p. 5, 2007.
[21] S. Lei, C. Sun, and T. Zhou, "Research on the multivariate time series linear regression prediction method of electric power short-term load [J]," Chinese Journal of Electrical Engineering, vol. 26, no. 2, p. 5, 2006.
[22] L. Chen, Z. Chen, and L. I. Guangxin, "Discussion on linear regression method of shear strength index in triaxial test [J]," Geotechnical Mechanics, vol. 26, no. 11, p. 5, 2005.
[23] S. Li and E. Zhang, "Research on remote sensing image classification method based on decision tree [J]," Regional Research and Development, vol. 22, no. 1, pp. 17-21, 2003.
[24] L. Xu, J. He, and H. Meng, "Application of decision tree based on information entropy in TCM syndrome differentiation of chronic gastritis [J]," Journal of Second Military Medical University, vol. 25, no. 9, p. 4, 2004.
[25] X. Liu and L. Sheng, "Optimization algorithm for decision tree [J]," Journal of Software, vol. 9, no. 10, p. 4, 1998.