Research Article

Machine Learning-Based Intelligent Scoring of College English Teaching in the Field of Natural Language Processing

Wei Wang ⓘ

Changchun Guanghua University, Changchun 130031, Jilin, China

Correspondence should be addressed to Wei Wang; wangweiwei@stu.wzu.edu.cn

The current education evaluation is limited not only to the mode of simplification, indexing, and datafication, but also to the scientific nature of college teaching evaluation. This work firstly conducts a theoretical analysis of natural language processing technology, analyzes the related technologies of intelligent scoring, designs a systematic process for intelligent scoring of college English teaching, and finally conducts theoretical research on the Naive Bayesian algorithm in machine learning. In addition, the error of intelligent scoring of English teaching in colleges and universities and the accuracy of scoring and classification are analyzed and researched. The results show that the error between manual scoring and machine scoring is basically about 2 points and the minimum error of intelligent scoring in college English teaching under machine scoring can reach 0 points. There is a certain bias in manual scoring, and scoring on the machine can reduce the generation of this error. The Naive Bayes algorithm has the highest classification accuracy on the college intelligent scoring dataset, which is 76.43%. The weighted Naive Bayes algorithm has been improved in the classification accuracy of college English teaching intelligent scoring, with an average accuracy rate of 74.87%. To sum up, the weighted Naive Bayes algorithm has better performance in the classification accuracy of college English intelligent scoring. This work has a significant effect on the scoring of the college intelligent teaching scoring system under natural language processing and the classification of college teaching intelligence scoring under the Naive Bayes algorithm, which can improve the efficiency of college teaching scoring.

1. Introduction

As a key link of undergraduate teaching quality evaluation and the implementation of education evaluation reform, college students' evaluation of teaching quality is trend of high-quality development of higher education [1]. However, in recent years, there appear increasingly serious administration and formalization, which deviates from the original intention of mutual teaching and learning [2, 3]. Through policy analysis, teaching practice, and data simulation, it was found that some of the problems that have hampered the evaluation reform are expected to be solved by technological means [4, 5]. Student evaluation of teaching is an important part of the teaching quality guarantee system of colleges and universities. Its original intention is to promote the teaching development of teachers and improve the learning efficiency of students and then fundamentally guarantee the quality of

classroom teaching and the effect of talent training. Meanwhile, effective student evaluation of teaching is also an important reference for the educational reform of colleges and universities [6, 7]. With the opening of the new round of undergraduate teaching evaluation in 2021, the reform of undergraduate teaching and evaluation has once again become an important focus for high-quality talent cultivation system in the 14th five-year plan period [8]. The Implementation Plan for the Examination and Evaluation of Undergraduate Education and Teaching in Ordinary Institutions of Higher Learning (2021–2025) proposes clearly promoting the reform of undergraduate education and teaching to avoid unscientific education evaluation and ensuring the priority of talent training and the core position of undergraduate education and teaching. By making good use of the organic integration of technology and educational governance, the quality of undergraduate education and

teaching can be effectively improved so that students' evaluation of teaching can be credible, feasible, available, and popularized [9].

Some foreign researchers used Hadoop distributed architecture to explore and analyze students' behaviors by collecting data on social platforms [10]. Panessai et al. used data mining to explore the data of teaching evaluation process and established a teaching evaluation prediction model using Iterative Dichotomizer 3(ID3) decision tree algorithm. During this period, association rule algorithm is used to conduct comparative analysis of experimental verification results to explore the reliability of ID3 decision tree model [11]. Pliakos et al. estimated students' learning ability based on the learning information of students through machine learning algorithm and applied item response theory to predict response of the model. By combining the machine learning algorithm and item response theory, it was found that the estimation accuracy of students' learning ability was higher [12]. Kaempf and Finn predicted the feature terms of students' learning ability through diversified questions on the MOOC platform, aiming at solving the problem of slow improvement of students' learning ability [13]. Yuan studied the effectiveness of English composition scoring on the Kouku Correcting website and found that the evaluation results of the machine scoring system were consistent with those of the manual scoring system [14]. Herbart's theory divides interest into four stages: attention, expectation, method, and behavior. According to the psychological changes of students' activities, he puts forward the theory of teaching form stage, which includes teaching methods and teaching forms [15]. Domestic researchers applied machine learning to the teaching evaluation system, modelled the teaching evaluation model of ethnic colleges and universities through the relevant theories of artificial neural network, and conducted quantitative analysis of comprehensive indicators in a quantitative way. It was found that the neural network model based on BP (Back propagation) algorithm can obtain reasonable evaluation and analysis results for teaching evaluation [16, 17].

In this work, natural language processing (NLP) technology is firstly discussed to pave the way for intelligent scoring of college English teaching under machine learning. Then, related technologies of intelligent scoring are briefly introduced. Based on the principles and evaluation standards of intelligent evaluation of college English teaching, the system process of intelligent scoring of college English teaching is designed. Finally, the Naive Bayes (NB) algorithm in machine learning is theoretically studied. To verify the feasibility of this work, the error of intelligent scoring in college English teaching and the accuracy of scoring classification are analyzed and studied. The novelty of this work is that the Bayesian algorithm is applied to the intelligent scoring system of English teaching in colleges and universities, so that the evaluation of efficient English teaching can be truly intelligent. The objective of this work is to apply the machine learning technology to the field of English teaching, to provide a realistic basis for improving the overall teaching level of English teaching.

2. Relevant Technical Theory of Intelligent Grading of English Teaching in Colleges and Universities

2.1. Natural Language Processing. NLP is to develop applications or services that can understand human language. It is practically applied to speech recognition, speech translation, understanding complete sentences, understanding synonyms of matching words, and generating grammatically correct complete sentences and paragraphs. The computer is used as a tool to process various technologies for written or oral forms. Figure 1 shows the task analysis and task generation under NLP.

In Figure 1, word segmentation is the division of a text into the smallest semantic units, i.e., words. For different languages, there are various basic vocabularies and grammars; word disambiguation: it is the process of selecting the correct meaning of a word from multiple meanings in the same context in natural language; named entity recognition: it is the process of extracting entities from a given text. The so-called entities are nouns such as person, place, and organization; lexical tagging: it is to classify a word into one of the different lexical categories such as noun, verb, adjective, and adverb; sentence classification: it is to understand whether a passage expresses a positive or negative meaning, such as teaching evaluation, to distinguish whether it is a good or bad review. It can be seen as a classification task; language generation: based on a text library, it can generate new text. For example, using Jin Yong's martial arts novels for training, the computer automatically generates Jin Yong style martial arts novels; question and answer system: typical application is Apple's Siri (Speech Interpretation and Recognition Interface) voice assistant. It can directly answer and solve the user's problems; machine translation (translation from one language to another): it requires the computer to understand its meaning first, before representing it in another language. Figure 2 shows NLP process under machine learning.

As illustrated in Figure 2, for natural language processing under machine learning, the English corpus is first preprocessed, that is, word classification, and then part-of-speech recovery, named entity recognition and part-of-speech tagging. Secondly, according to feature extraction and feature selection, an appropriate classifier is selected to complete the language classification process. To process natural language with computer, the first thing to do is to represent text with appropriate data. In the early stage, NLP trains models by machine learning, and the text representation is similar to feature engineering in machine learning. Translating an English text into a "word bag" is a strategy to express the text according to the frequency of words. For example, first extract all the words in the text "Written language tends to use nouns instead of verbs" and "Written language tends to use nouns" to form a vocabulary, according to the vocabulary, ("Written," "language," "tends," "to," "use," "nouns," "instead," "of," "verbs"), and then use the frequency of each word to represent the corresponding text. The frequency of each word represents the corresponding text. In addition to the number of times a word appears, there can be other coding methods,
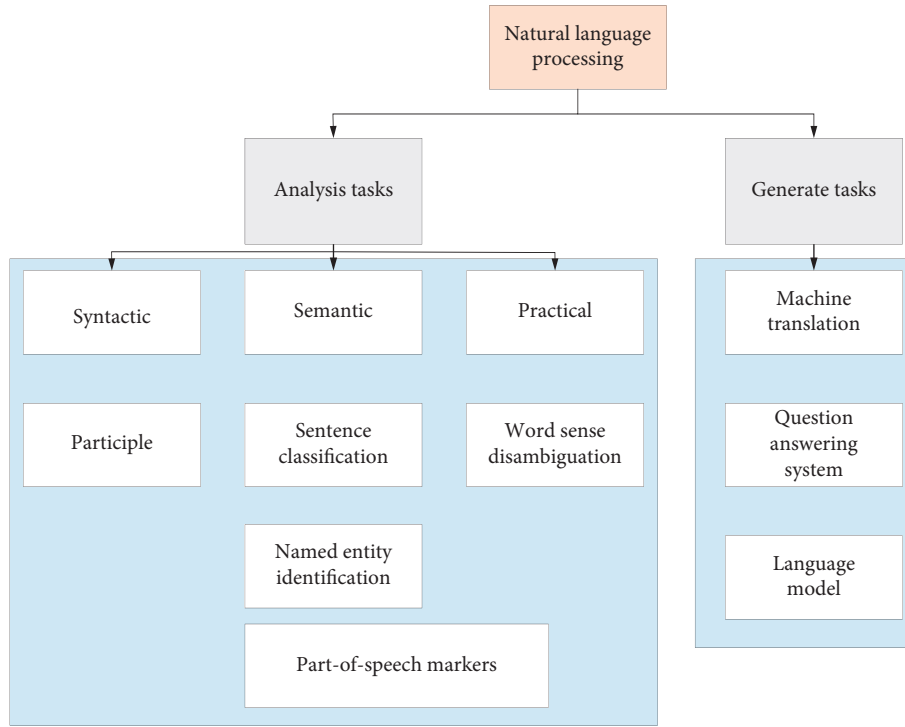
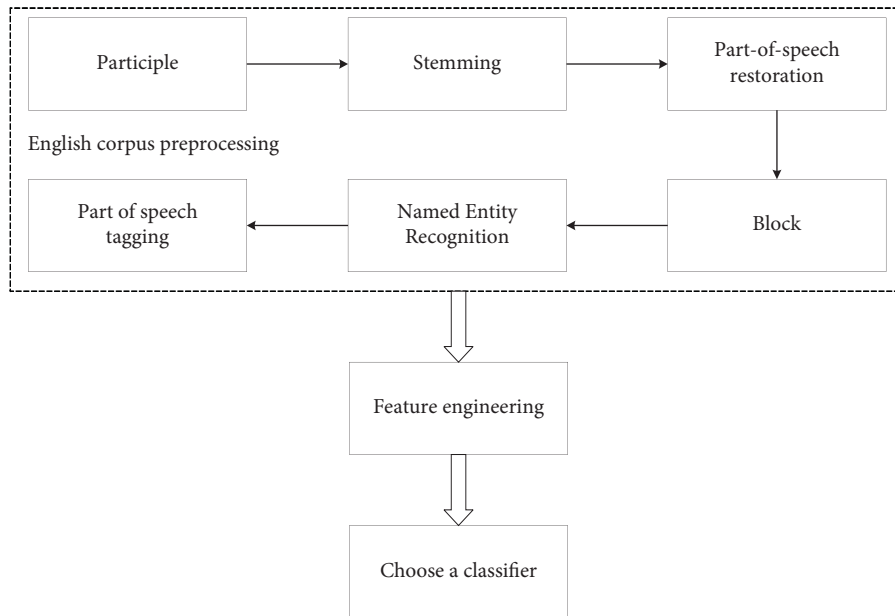FIGURE 1: Task analysis and task generation under NLP.

FIGURE 2: NLP process under machine learning.

such as whether a word appears or not, with 0 indicating no appearance and 1 indicating appearance. This method is called One-Hot, that is, one-hot coding [18]. You can also adopt the strategy of TF-IDF (term frequency, inverse document frequency) [19, 20], in which TF (Term frequency) is called word frequency, and its importance is shown in

$$tf_{i,j} = \frac{n_{i,j}}{\sum_k n_{k,j}}.$$ (1)

In (1), $n_{i,j}$ is the frequency of d_j that is the word appearing in the document d_j, $\sum_k n_{k,j}$ is d_j that is the sum of the times of all words appearing in the document d_j, and $tf_{i,j}$ is the importance of words t_i in a specific document. IDF (inverse document frequency) is a method of weighting word frequency, expressed as

$$idf_i = \log \frac{|D|}{\left|\{j: t_i \in d_j\}\right|},$$ (2)

where $|D|$ is the total number of $|\{j: t_i \in d_j\}|$ files in the corpus, $|\{j: t_i \in d_j\}|$ is the number of files containing words t_i, and idf_i is the frequency of reverse files. If the word is not in the corpus, the dividend will be 0, so $1 + |\{j: t_i \in d_j\}|$ is generally used. Calculation of the high word frequency of a particular file and the low file frequency of the word in the whole file set can produce $tfidf_i$ with a high weight, and the $tfidf_i$ calculation expression is

$$tfidf_i = tf_i * idf_i. \tag{3}$$

2.2. Related Technologies for Intelligent Scoring.
More and more colleges and universities regard student evaluation as an important mechanism of teacher evaluation. This mechanism is aimed at arousing the enthusiasm of students to participate in classroom teaching, encouraging teachers to make full preparations and give lectures seriously, and urging teachers to pay attention to the teaching effect. The original intention of this mechanism is worthy of affirmation and has also received good results, but there are still some areas to be improved in the teaching practice, for example, one-size-fits-all, unification, and formalization. These deficiencies restrict the improvement of teaching quality and scientific research level to some extent. In terms of its negative impact on the field of economics, some colleges and universities pay too much attention to those scholars who teach modern mainstream economics and not enough attention to other types of scholars when introducing and assessing talents. Some "oratory" teachers are popular, while "thinking" scholars who are good at quiet thinking tend to get lower marks. Figure 3 shows the intelligent evaluation principle of college English teaching.

In Figure 3, the intelligent evaluation principle of English teaching in colleges and universities includes three parts, namely, the principle of diversity, the principle of practicality, and the principle of coherence. The plurality principle means that the intelligent English teaching not only should be limited to the evaluation of students, but also needs to introduce the participation of professional English teaching researchers to diversify teaching effect evaluation. The principle of practicability means that, according to the current evaluation system of English teaching in colleges and universities, there are many shortcomings. Practicability requires not only the improvement of students' and teachers' English ability, but also the comprehensive consideration of students' professional English ability and oral English level. Consistency principle suggests that evaluation of English teaching in colleges and universities should not have time limitations. It must follow the teachers' teaching theories and methods and make the students have a basic understanding of the English learning ability. At the same time, the teaching evaluation not only exists in classroom teaching, but also is performed through practice and activities. Figure 4 shows the evaluation criteria of college English teaching.

College English teaching is evaluated in terms of teaching purpose, teaching material processing, teaching process, and teaching evaluation based on the principle of
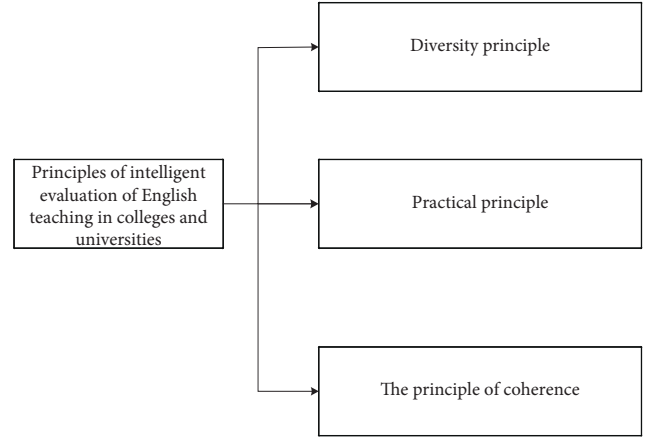

Figure 3: Intelligent evaluation principles of college English teaching.

comprehensive consideration and scientific practice. The evaluation results are divided into excellent (above 85 points), good (84–75 points), qualified (74–60 points), and unqualified (below 60 points). Figure 5 is the system process of intelligent scoring in college English teaching.

In Figure 5, the intelligent scoring module of college English teaching is analyzed from the three modules of data collection, evaluation process, and data storage.

2.3. NB Classification.
Machine learning is essentially an approximation of a problem by a real model. Among them, supervised classification algorithm has been widely used in many business scenarios. Classification problem belongs to the prediction task, which is to obtain an objective function f through the learning of existing data sets and to map each attribute set x to the target attribute y. The y must be discrete. There are many methods to solve the classification problem. The basic classification methods mainly include decision tree, NB, artificial neural network, K-nearest neighbor, and support vector machine. In addition, there are ensemble learning algorithms for combining basic classifiers. Representative ensemble learning algorithms include random forest [21, 22]. In this work, the basic principles and advantages and disadvantages of each basic classifier are summarized. Figure 6 shows the NB model.

The NB model in Figure 6 is a simple directed graph model that learns the probability of $y(x)$ belonging to each classification [23–25]. y is the class variable, and $x_1, x_i, \ldots,$ and x_d are the attribute variables. First, the training sample set is defined in

$$\left(x_1^{(1)}, x_2^{(1)}, \cdots, x_n^{(1)}, y_1\right)\left(x_1^{(2)}, x_2^{(2)}, \ldots, x_n^{(2)}, y_2\right), \ldots, \\ \left(x_1^{(m)}, x_2^{(m)}, \ldots, x_n^{(m)}, y_m\right), \tag{4}$$

where m represents the number of samples, x_i is the feature vector of sample X, n is the sample feature, y_1, y_2, and y_m are the output variables, and the feature output has k categories, which are defined as $C_1, C_2, \ldots, C_k$. The sample number of each feature output category is $m_1, m_2, \ldots, m_k$. In the kth category, if the feature is discrete, the value of x_i category is
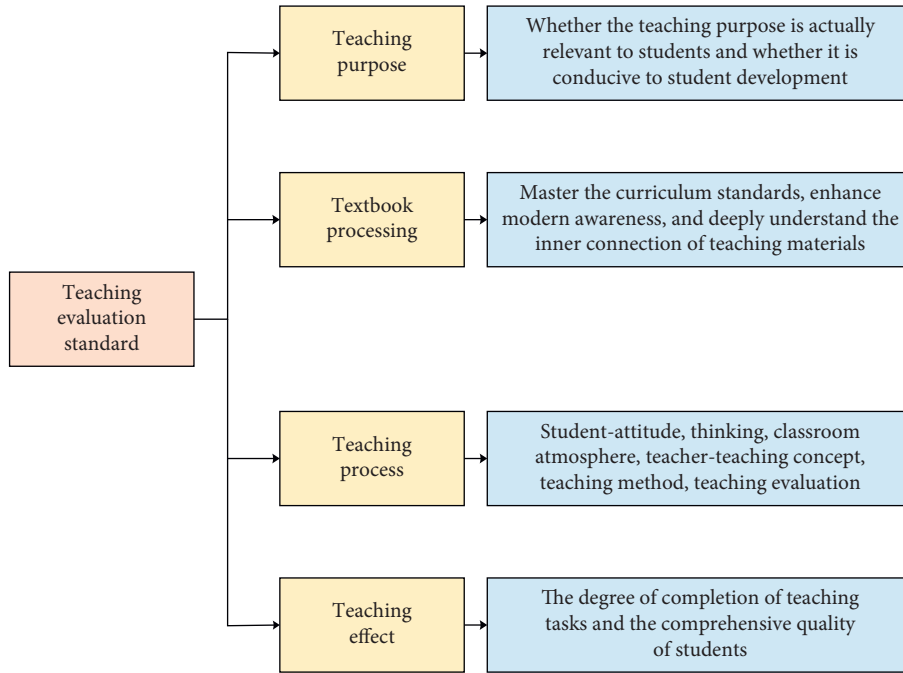
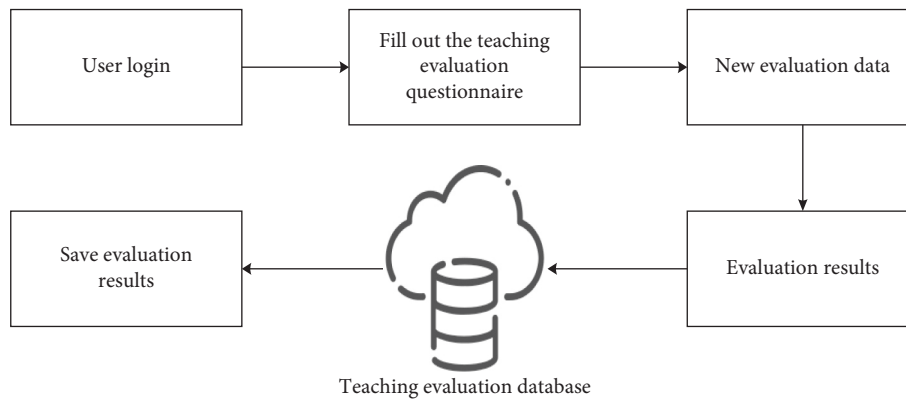FIGURE 4: Evaluation criteria for college English teaching.

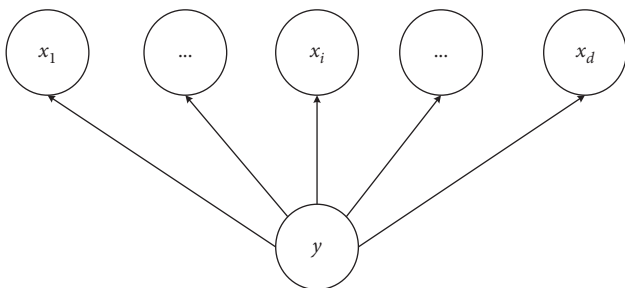FIGURE 5: The systematic flow of intelligent grading of English teaching in colleges and universities.

FIGURE 6: NB model.

m_{kjl}, where l is $1, 2, \ldots, S_j$. S_j is the different values of feature j. The prior probability of y is calculated as

$$P(y = C_k) = \frac{m_k + \lambda}{m + k\lambda}, \tag{5}$$

where λ is an arbitrary constant and the conditional probability distribution is expressed in

$$P(X = x | y = C_k) = P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n | y = C_k). \tag{6}$$

The probability of determining the data set sample belonging to a certain category is expressed as

$$P(C_j | X) = \text{argmax} \frac{P(X | C_j) P(C_j)}{P(X)}. \tag{7}$$

In (7), sample X of data set can be represented as $\{x_1, x_2, \ldots, x_n\}$, and C_j is the class label of the corresponding training data set. Assuming that the samples are independent from each other, the calculation expression of the category to which the data set belongs is shown in

$$P(X|C_j) = \prod_{i=1}^{n} P(x_i|C_j). \tag{8}$$

From (8), $P(x_1|C_j)$, $P(x_2|C_j)$, and $P(x_n|C_j)$ can be obtained, and then the category of test data set is deduced. Novendri et al. first adopted the idea of attribute feature weighting in NB classifier, improving the classification accuracy of NB method [26]. To take the correlation degree and accuracy coefficient of attributes into account at the same time, a new attribute calculation method is defined in

$$w_i = \frac{(w_k + w_s)}{2}, \tag{9}$$

where w_k is the weighted coefficient of accuracy, w_s is the weighted coefficient of correlation degree, and w_i is the weight of total associated attribute. Then, NB assigned with a weight w is expressed as

$$P(X|C_k) = w * \prod_{i=1}^{n} P(x_i|C_k). \tag{10}$$

The calculation process of NB classification is shown in Figure 7.

In Figure 7, the NB is divided into three stages. For the preparation stage, its task is to make necessary preparations for the NB classification. Its main work is to determine the attributes and characteristics of each attribute according to the specific circumstances. Then, each characteristic attribute is properly divided, and a part of the items to be classified are manually classified to form a training sample set. In this stage, all the data to be classified are input, and the output is characteristic attributes and training samples. The training stage of classifier is the generation of classifier. The main work is to calculate the occurrence frequency of each category in the training sample and the conditional probability estimation of each category based on the classification of each characteristic attribute and to record the results. Its input is characteristic attribute and training sample, and its output is classifier. This stage is mechanical and can be calculated automatically by the program according to the formula discussed earlier. In the application phase, its task is to classify items using classifiers. Its input is classifiers and items to be classified, and its output is the mapping relationship between items to be classified and categories. This stage is also mechanical and is done by a computer.

The data set used for the above model is the UCI data set IRIS. According to the needs of English teaching evaluation work, the above classification algorithm can be applied to the English teaching evaluation process. Taking a series of evaluation attribute values as input data and the comprehensive evaluation level as class label, learning a classifier through a certain classification algorithm can give the most likely class label for the new English teaching evaluation attribute value, that is, the evaluation result. In order to ensure the credibility of the evaluation results, it is necessary to select a suitable algorithm to construct the classifier. Accuracy is an indicator for evaluating the performance of a classifier, which is specifically defined as the ratio of the

number of samples that the classifier can correctly classify to the total number of samples for a given test data set. The calculation equation is given as follows:

$$p = \frac{N'}{N}. \tag{11}$$

In (11), P represents the accuracy rate, N' represents the number of correctly classified samples, and N represents the total number of samples.

3. Results and Discussion

Based on the Nation Survey of Student Engagement, undergraduates in colleges and universities in Changchun were taken as the research subjects and graded according to the evaluation results of excellent, good, qualified, and unqualified. The evaluation content of college English teaching is the behaviors of teachers and students in the teaching process, as shown in Table 1.

3.1. An Analysis of Intelligent Scoring Errors in College English Teaching. Through NLP, the key words in the intelligent scoring process of college English teaching are extracted. For example, the scoring key words are good, not bad, understandable, abstract, and difficult. According to the score level, above 85 is excellent, between 84 and 75 is good, between 74 and 60 is qualified, and below 60 is not qualified. The error in the intelligent scoring process of Changchun colleges and universities is analyzed, and 30 score data sets are randomly selected on the Internet for testing. Figure 8 shows the error of intelligent scoring in college English teaching.

Figure 8 shows the manual scoring and machine scoring under NLP of college English teaching. It is noted that the error curves of both manual scoring and machine scoring show irregular fluctuations, the error between manual scoring and machine scoring is basically less than 2 points, and the error of intelligent scoring of college English teaching under machine scoring can reach 0 points at the minimum. The mean test error is 1.41 points under manual scoring and 1.02 points under machine learning. According to the theory of human emotion, there is a certain deviation in the manual scoring, but the machine scoring can reduce this kind of error.

3.2. Classification Accuracy of Intelligent Scoring in College English Teaching Using Different Models. In this work, NB is compared with DT (decision tree), BP algorithm, KNN (K-nearest neighbor), and SVM. After 10 times of testing, the intelligent scoring data are classified. Figure 9 shows the classification accuracy of intelligence scoring under different models.

NB algorithm has the highest classification accuracy of 76.43% compared with SVM algorithm of 69.56%, KNN algorithm of 70.00%, DT algorithm of 67.37%, and BP algorithm of 72.27%. It can be concluded that it is reasonable to choose this algorithm in machine learning. By introducing weights into the NB algorithm, the classification
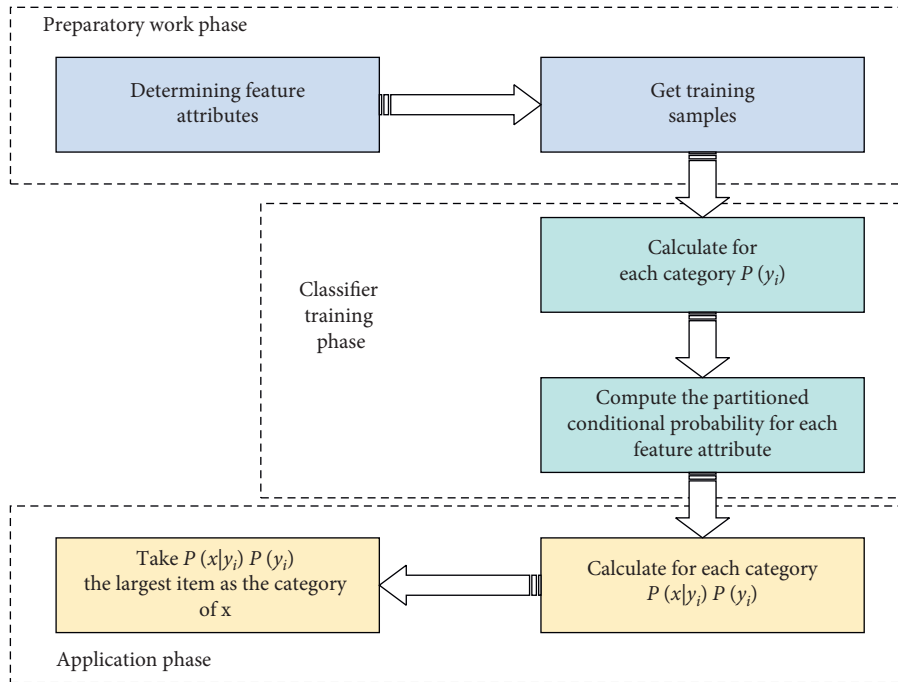
Figure 7: The classification process of NB.

Table 1: The content of college English teaching evaluation.

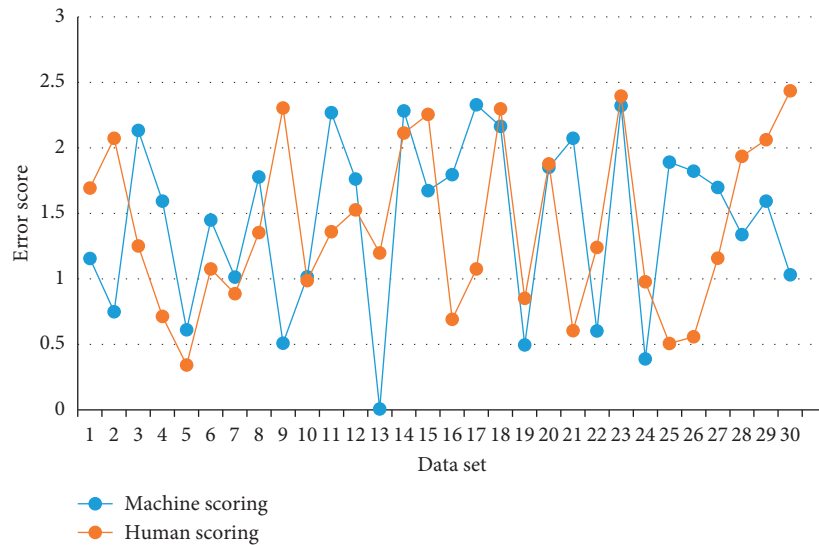| The subjects	Content
Teacher	Classroom atmosphere, homework completion, teaching content extension, etc.
Student	Learning initiative, awareness of cooperation and communication, learning attitude, etc.

Figure 8: Intelligent scoring error in college English teaching.

accuracy before and after weighting is compared. The classification accuracy under weighted Naive Bayes (WNB) is shown in Figure 10.

It can be clearly seen from Figure 10 that the WNB algorithm improves the classification accuracy of intelligent scoring in college English teaching, with an average accuracy of 74.87%, while the NB algorithm has an average classification accuracy of 71.29%. Above, the WNB algorithm has a good performance in the classification accuracy of intelligent scoring of college English teaching.
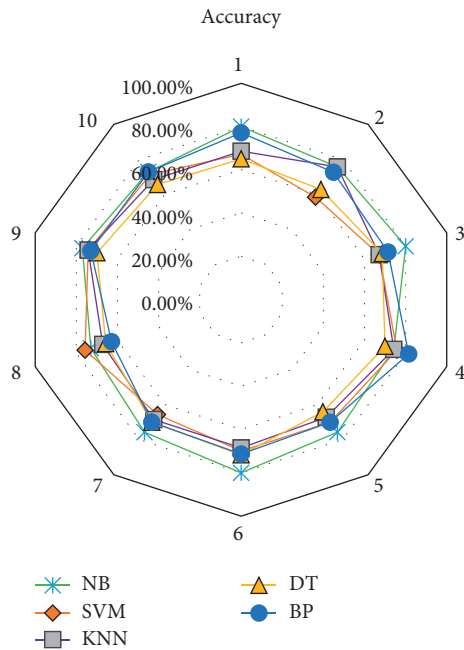
FIGURE 9: Classification accuracy of intelligent scoring in college English teaching using different models.
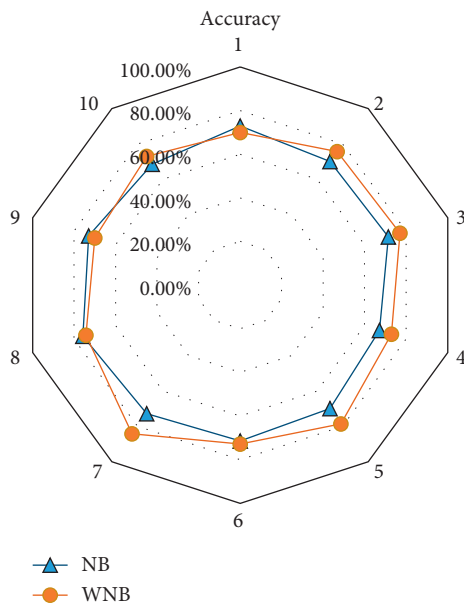

FIGURE 10: An analysis on the classification accuracy of college English teaching intelligence scoring based on WNB.

Lin applied the classification algorithm in machine learning to the construction of the evaluation model, which further improved the scientificity and feasibility of teaching evaluation. In addition, the empirical algorithm was used as the basic algorithm to evaluate the teaching quality, and the subject word distribution obtained by the joint model training was used as the original knowledge. The scoring of the research model was close to the standard manual scoring, which can provide a theoretical reference for subsequent related research [27]. This is similar to the research in this

work, which can provide a basis for college teaching. Sun et al. established an English teaching evaluation implementation model based on machine learning decision tree technology. It provided valuable data from a wide range of information, summarizing rules and data to help teachers improve education and students' English achievement. This system embodied the idea of artificial intelligence expert system. Test applications showed that the system can help students improve their learning efficiency and make learning content more relevant [28]. This is similar to the results of this work, which can also apply machine learning to English teaching. This shows that the integration of cutting-edge technologies (such as machine learning technology under artificial intelligence) in college teaching, especially English teaching, is crucial to the development of English education.

4. Conclusions

In the work, the NB algorithm is used to study the classification accuracy of intelligent scoring in college English teaching, and the error of intelligent scoring in C=college English teaching under NLP is analyzed. According to the analysis of the research results, the average test error under manual scoring is 1.41 points, while the average test error under machine learning is 1.02 points. According to the theory of human emotion, there is a certain deviation in the manual scoring, but the machine scoring can reduce this kind of error. The average classification accuracy of NB algorithm is 76.43%, higher than SVM algorithm of 69.56%, KNN algorithm of 70.00%, DT algorithm of 67.37%, and BP algorithm of 72.27%. The WNB algorithm has an average accuracy of 74.87% in intelligent scoring classification of college English teaching, and that of NB algorithm is 71.29%. In conclusion, the WNB algorithm has a good performance in the classification accuracy of intelligent scoring. This study plays a significant role in the scoring of intelligent teaching scoring system in colleges and universities under NLP and the classification of intelligent teaching scoring, which can improve the work efficiency of teaching scoring in colleges and universities and evaluate the teaching process scientifically to a certain extent. However, the number of data sets selected is limited, and the researchers could increase the amount of data, so as to strengthen the findings of the study.

Data Availability

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

Conflicts of Interest

The author declares no potential conflicts of interest.

Acknowledgments

References

[1] M. Antonopoulou, M. Mantzorou, A. Serdari, and C. Bonotis VasiosPavlidouTrifonosVadikoliasPetridisGiaginis, "Evaluating Mediterranean diet adherence in university student populations: does this dietary pattern affect students' academic performance and mental health?" *The International Journal of Health Planning and Management*, vol. 35, no. 1, pp. 5–21, 2020.

[2] A. A. Lapada, F. F. Miguel, D. A. R. Robledo, and Z. F. Alam, "Teachers' Covid-19 Awareness, distance learning education experiences and perceptions towards institutional readiness and challenges," *International Journal of Learning, Teaching and Educational Research*, vol. 19, no. 6, pp. 127–144, 2020.

[3] S. Pokhrel and R. Chhetri, "A literature review on impact of COVID-19 pandemic on teaching and learning," *Higher Education for the Future*, vol. 8, no. 1, pp. 133–141, 2021.

[4] M. Assunção Flores and M. Gago, "Teacher education in times of COVID-19 pandemic in Portugal: national, institutional and pedagogical responses," *Journal of Education for Teaching*, vol. 46, no. 4, pp. 507–516, 2020.

[5] S. Nuere and L. de Miguel, "The digital/technological connection with Covid-19: an unprecedented challenge in university teaching," *Technology, Knowledge and Learning*, vol. 26, no. 4, pp. 931–943, 2021.

[6] J. Ives and M. Castillo-Montoya, "First-generation college students as academic learners: a systematic review," *Review of Educational Research*, vol. 90, no. 2, pp. 139–178, 2020.

[7] V. R. Downing, K. M. Cooper, J. M. Cala, and S. E. GinBrownell, "Fear of negative evaluation and student anxiety in community college active-learning science courses," *CBE-life Sciences Education*, vol. 19, no. 2, p. ar20, 2020.

[8] K. Zakić, "New development paradigm within the Chinese 14th five-year plan-Chinese vision of modern China," *Review of International Affairs*, vol. 72, no. 1183, pp. 67–87, 2021.

[9] G. Sen, H. W. Chau, M. A. U. R. Tariq, A. W. M. Muttil, and A. W. M. Ng, "Achieving sustainability and carbon neutrality in higher education Institutions: a review," *Sustainability*, vol. 14, no. 1, p. 222, 2021.

[10] S. Asadianfam, M. Shamsi, and A. R. Kenari, "TVD-MRDL: traffic violation detection system using MapReduce-based deep learning for large-scale data," *Multimedia Tools and Applications*, vol. 80, no. 2, pp. 2489–2516, 2021.

[11] I. Y. Ismail Yusuf Panessai, M. M. Lakulu, M. H. Abdul Rahman, and N. S. Mohd NoorMat Salleh, "PSAP: improving accuracy of students' final grade prediction using ID3 and C4.5," *International Journal of Artificial Intelligence*, vol. 6, no. 2, pp. 125–133, 2019.

[12] K. Pliakos, C. Vens, and G. Tsoumakas, "Predicting drug-target interactions with multi-label classification and label partitioning," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 18, no. 4, pp. 1596–1607, 2021.

[13] S. Kaempf and C. Finn, "Teaching international relations through the format of a massive open online Course (MOOC)," *International Studies Perspectives*, vol. 22, no. 1, pp. 1–24, 2021.

[14] Z. Yuan, "Interactive intelligent teaching and automatic composition scoring system based on linear regression machine learning algorithm," *Journal of Intelligent and Fuzzy Systems*, vol. 40, no. 2, pp. 2069–2081, 2021.

[15] N. Nurlaila and M. Muassomah, "The collaboration of Herbart method with cooperative learning: optimizing writing skills at IAIN Madura," *ALSINATUNA*, vol. 5, no. 2, pp. 147–164, 2020.

[16] X. Zhai, "Practices and theories: how can machine learning assist in innovative assessment practices in science education," *Journal of Science Education and Technology*, vol. 30, no. 2, pp. 139–149, 2021.

[17] H. S. Alenezi and M. H. Faisal, "Utilizing crowdsourcing and machine learning in education: literature review," *Education and Information Technologies*, vol. 25, no. 4, pp. 2971–2986, 2020.

[18] G. Wei and Z. Wang, "Adoption and realization of deep learning in network traffic anomaly detection device design," *Soft Computing*, vol. 25, no. 2, pp. 1147–1158, 2021.

[19] A. P. Pimpalkar and R. J. Retna Raj, "Influence of pre-processing strategies on the performance of ML classifiers exploiting TF-IDF and BOW features," *ADCAIJ: Advances in Distributed Computing and Artificial Intelligence Journal*, vol. 9, no. 2, pp. 49–68, 2020.

[20] Y. D. Kirana and S. Al Faraby, "Sentiment analysis of beauty product reviews using the K-nearest neighbor (KNN) and TF-IDF methods with chi-square feature selection," *Journal of Data Science and Its Applications*, vol. 4, no. 1, pp. 31–42, 2021.

[21] S. Talukdar, P. Singha, S. Mahato, A. PalLiou, and A. Rahman, "Land-use land-cover classification by machine learning classifiers for satellite observations-A review," *Remote Sensing*, vol. 12, no. 7, p. 1135, 2020.

[22] X. F. Wang, P. Gao, Y. F. Liu, and F. LiLu, "Predicting thermophilic proteins by machine learning," *Current Bioinformatics*, vol. 15, no. 5, pp. 493–502, 2020.

[23] S. Bakheet and A. Al-Hamadi, "A framework for instantaneous driver drowsiness detection based on improved HOG features and naïve bayesian classification," *Brain Sciences*, vol. 11, no. 2, p. 240, 2021.

[24] I. Wickramasinghe and H. Kalutarage, "Naive Bayes: applications, variations and vulnerabilities: a review of literature with code snippets for implementation," *Soft Computing*, vol. 25, no. 3, pp. 2277–2293, 2021.

[25] S. Ruan, B. Chen, K. Song, and Li, "Weighted naïve Bayes text classification algorithm based on improved distance correlation coefficient," *Neural Computing & Applications*, vol. 34, no. 4, pp. 2729–2738, 2022.

[26] R. Novendri, A. S. Callista, D. N. Pratama, and C. E. Puspita, "Sentiment analysis of YouTube movie trailer comments using naïve bayes," *Bulletin of Computer Science and Electrical Engineering*, vol. 1, no. 1, pp. 26–32, 2020.

[27] L. Lin, "Smart teaching evaluation model using weighted naive bayes algorithm," *Journal of Intelligent and Fuzzy Systems*, vol. 40, no. 2, pp. 2791–2801, 2021.

[28] Z. Sun, M. Anbarasan, and D. Praveen Kumar, "Design of online intelligent English teaching platform based on artificial intelligence techniques," *Computational Intelligence*, vol. 37, no. 3, pp. 1166–1180, 2021.