Hindawi
Complexity
Volume 2021, Article ID 5531595, 10 pages
https://doi.org/10.1155/2021/5531595

Research Article

Distribution of Large-Scale English Test Scores Based on Data Mining

Na Chu[1] **and Wanzhi Ma**[2]

[1]*School of Foreign Languages, Ningxia Normal University, Guyuan, Ningxia Province, 756000, China*
[2]*Department of Educational and Culture Contents Development, Woosuk University, Jeonju 55338, Republic of Korea*

Correspondence should be addressed to Wanzhi Ma; 997443418@stu.woosuk.ac.kr

Received 17 February 2021; Revised 10 March 2021; Accepted 15 March 2021; Published 24 March 2021

Academic Editor: Wei Wang

Data mining technology is an effective knowledge mining and data relationship induction technology based on massive data, which is widely used in data analysis in many fields. In order to improve the utilization effect of students' performance and meet the teaching needs of modern education, data mining technology can be applied to the existing performance database to mine the data information and treatment. Data mining technology is used to analyse and process the data stored in the student achievement management system, which provides the basis for improving the teaching quality and optimizing the teaching resources. Based on the analysis of the relevant data of large-scale English test results, this paper finds out the relevant rules that affect college English test results, forms the corresponding performance prediction rules, uses data mining technology to more comprehensively analyse the factors that affect students' performance, establishes a model, and uses data mining tools to mine and analyse students' English test data. It is of great practical significance to select the model with high accuracy, further optimize the parameters, make good use of the data, and then take targeted measures to guide the teaching reform, help students make more efficient learning plans, and improve and perfect the existing problems in teaching.

1. Introduction

Data mining, also known as knowledge discovery in database, refers to a process of extracting valuable, potential, and available knowledge from many noisy and incomplete databases. It integrates the knowledge of statistics, database, machine learning, pattern recognition, artificial intelligence, information retrieval, and other fields and is a hot issue in the current information technology research. Data mining technology is to solve the contradiction of a large amount of data and lack of knowledge [1]. It can automatically mine people's unknown and valuable knowledge or patterns from the known data set by using a computer and provide scientific guidance for our work and study. Compared with traditional data analysis, data mining has shown unique advantages in many aspects, so it has been successfully applied to the fields of finance, retail, manufacturing, insurance, engineering design, and scientific exploration [2]. In recent years, with the rapid development of big data industry and artificial intelligence, information education has become a development trend of China's education industry. The traditional performance analysis will cover up the specific factors that affect students' performance and cannot find the relationship between the factors. It can only get a result of the discussion but cannot play a guiding role in the action. With the rapid development of the information age, the knowledge contained in the massive education data has attracted more attention. Therefore, the most important problem of education informatization is to use data mining technology to analyse all kinds of education and teaching information, explore the meaningful knowledge hidden in the data, and make this knowledge better serve the education and teaching workers [3].

The continuous promotion and deepening of education informatization not only improve the teaching and management efficiency of schools at all levels but also produce

many education data resources, namely, education big data. Education data cover a wide range, including the whole education chain from enrolment to graduation, such as students' basic information, students' achievement information, and students' evaluation information [4]. Using data mining technology to serve the teaching management and decision-making in colleges and universities is an important basis for the reform of teaching management in colleges and universities. At present, the application of data mining technology in the field of education mainly includes two aspects: the evaluation of education and teaching quality and the application of student management system. But overall, the means of using data mining technology to deal with education big data in colleges and universities are still relatively limited. Accustomed to the traditional mode of education management, the innovative mode seems to be not progressing smoothly. In addition to the database used to store data and various management systems used in management, it is rare to use data mining technology to guide education and teaching [5]. With the continuous development of the information age, data mining in the field of big data has become a trend. Using various data mining software to analyse educational data will make future teaching and management more scientific and efficient [6].

In this paper, a data mining algorithm is used to analyse the results of a large-scale English test. This paper studies the main factors affecting the performance of large-scale English majors and uses these factors to predict the passing situation of the examination, which provides a scientific basis for teachers to reform English teaching methods and guide students to learn English. At the same time, it puts forward constructive suggestions for students with a low passing rate, to improve their employment competitiveness. Through the systematic theoretical research on data mining, this paper deeply analyses the concept, process, function, and common algorithms of data mining. This paper makes a brief statistical analysis and cleaning of the scores of large-scale English tests, explores the specific factors affecting the scores of large-scale English tests, and selects more reliable features to put into the model. Data mining tools are used to mine and analyse students' English test scores. The model with high accuracy is selected and the parameters are further optimized to make the result more accurate and reliable.

2. Related Theories and Technologies of Data Mining

2.1. Overall System Structure of Data Mining. The overall structure of data mining system is mainly composed of the following parts: database and data warehouse. It means that the data mining object is composed of database, data warehouse, data form, or other information databases [7]. Data cleaning and data integration operations are usually used to process these data objects. The database or data warehouse server is responsible for reading the relevant data according to the user's data mining request. The structure of data mining is shown in Figure 1.

Knowledge stock is the area of information that records mining needs, which will be used to inform the search

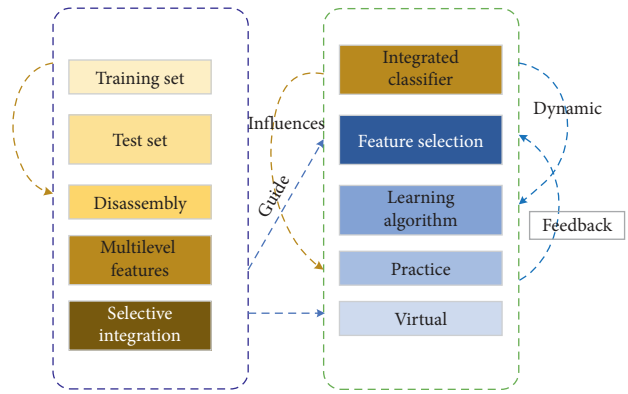

Figure 1: The structure of data mining.

system of information mining or to assist in considering the mining results. The user-defined threshold used in the mining algorithm is the easiest area know-how [8]. Data mining engine, which is the most simple issue of the facts mining system, commonly includes a set of mining feature modules to the entire mining features such as qualitative induction, affiliation analysis, classification induction, evolutionary computation, and deviation analysis. The pattern knowledge evaluation module can help the data mining module focus on mining more meaningful pattern knowledge according to the interest standard [9]. Of course, whether the module can be combined with the data mining module is related to the specific mining algorithm used in the data mining module. Obviously, if the data mining algorithm can be combined with the knowledge evaluation method, it will help to improve the efficiency of data mining. The visual user interface helps users communicate with the data mining system itself. On the one hand, users submit their mining requirements or tasks to the mining system through the module and provide the relevant knowledge needed for mining search; on the other hand, the system shows or explains the results or intermediate results of data mining to users through the module; in addition, the module can also help users browse the content of data objects and data definition patterns and evaluate the mined pattern knowledge, as well as a variety of forms to show the pattern of mining knowledge [10]. From the perspective of data warehouse, data mining can be regarded as the advanced stage of online analysis and processing, but the data analysis ability of data mining based on a variety of advanced data understanding technologies is far more than the online analysis and processing function of data warehouse based on data aggregation.

2.2. Algorithm of Data Mining. The data mining algorithm is a group of heuristics and calculations to create a data mining model based on data. It analyses the data provided by users and finds specific types of patterns and trends. The algorithm uses the results of this analysis to define the best parameters for creating mining models. These parameters are applied to the whole data set to extract feasible patterns and detailed statistical information. Most data mining algorithms use one or several objective functions and use several search

methods, such as heuristic algorithm, maximum and minimum method, gradient descent method, and network deduction method to obtain a point or a small area in the data body or in the data space where the distance relationship is established [11, 12]. According to the mining methods, data mining algorithms can be divided into teacher type and nonteacher type, also known as supervised learning and unsupervised learning. In supervised learning, a teacher's signal is given first, which can provide category label and classification cost for each input sample in the training sample set and find the direction to reduce the total cost. There is no explicit teacher in the unsupervised learning algorithm, and the system clusters the input samples automatically [13]. From the perspective of application, data mining algorithms can be divided into the following six categories: classification algorithm, regression algorithm, clustering analysis algorithm, association rules, timing, and deviation checking algorithm [14]. This paper mainly uses a regression algorithm. Linear regression is a kind of regression algorithm. In linear regression, data are modelled by a straight line. Bivariate regression takes a random variable y as a linear function of another random variable x.

$$\gamma = \tau + \ell x + \|\tau - \ell x\|, \tag{1}$$

where the variance of y is a constant and α and β are regression coefficients, which, respectively, represent the intercept and slope of the line on the y-axis. These coefficients can be solved by the least square method, and the error between the actual data and the estimation of the line can be minimized. Given s samples or data points in the form of $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$, the regression coefficient can be calculated by the following formula:

$$\ell = \frac{\sum_1^n (x_i - \overline{x})}{\sum_1^n (x_i - \overline{x})^2} + \frac{y_i - \overline{y}}{\sum_1^n (x_i - \overline{x})}, \tag{2}$$

where $\overline{x}$ is the average value of $x_1, x_2, \ldots, x_n$ and $\overline{y}$ is the average value of $y_1, y_2, \ldots, y_n$. Multiple regression is an extension of linear regression, which designs multiple predictors. The corresponding variable y can be a linear function of a multidimensional eigenvector. Multiple regression based on two predictors x_i and x_{i-1} is as follows:

$$y = (\tau - \ell x_i)\ell x_{i-1}. \tag{3}$$

2.3. Process Analysis of Data Mining. The application of data mining algorithm to the analysis of college students' performance needs to go through three stages: data preparation stage, data mining stage, and data result expression and interpretation stage. The data mining process is shown in Figure 2.

2.3.1. Data Preprocessing. This stage is used to provide data information that can be used for direct processing and analysis, so in this stage, it is necessary to integrate, filter, and process the source data appropriately according to the data information requirements of the algorithm, to obtain the analysis results with high reliability. This part of the work occupies a large proportion in the whole performance analysis work [15]. In the analysis of college students' performance, the information used for data mining may involve multiple databases or disciplines, so it is necessary to collect and sort these data, eliminate the semantic fuzziness between data sources, deal with the existing information defects, and sort them into a unified and standardized data format. There may be many irrelevant data in the data analysis space formed by the collection of source data. These data do not provide support for the development of data mining but will increase the workload [16, 17]. Therefore, the second content of data preparation is the selection of data. The selected data should be the relevant data content that is useful for analysis and can effectively narrow the processing range. There may be noise problems, incomplete problems, or inconsistent problems in the filtered data. At this time, data preprocessing operation is also needed to further improve and enrich the data structure in the data analysis database to ensure the reliability and credibility of the analysis results. In order to facilitate the algorithm analysis, it is also necessary to transform the attribute field information in the database into recognizable and processable coding data.

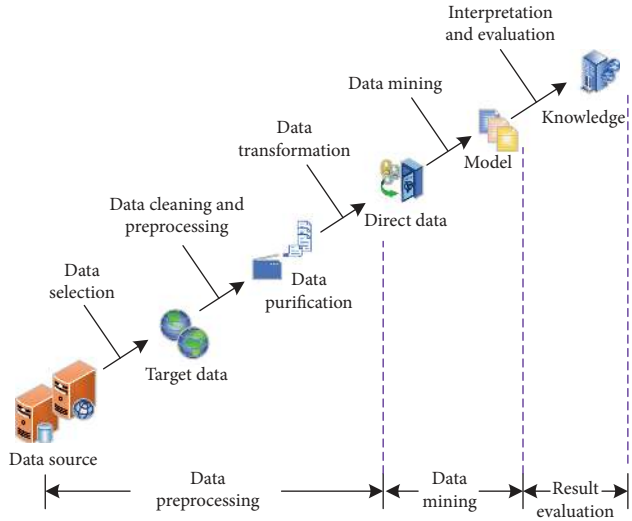

Figure 2: The process of data mining.

2.3.2. Data Mining. This work is the executive part of the whole student achievement analysis. It needs to apply a variety of data mining algorithms to process and analyse the data information in the database and explore the available internal relations or knowledge map [18]. First, we need to determine the mining target or task, then select the appropriate mining algorithm according to the mining target to construct the data model and the specific parameters that need to be analysed, and use the model to mine and analyse the relevant parameters in the database, find out the association rules and data regression structure that meet the requirements, and give the pattern expression that can be used for evaluation and analysis [19, 20]. In practical application, after algorithm selection, data mining can be directly selected to complete data mining automatically.

2.3.3. Results Evaluation and Interpretation. After the completion of data mining, users need to evaluate and judge the obtained pattern analysis results or pattern expressions to see whether they are effective and can meet the needs of performance analysis. If users are not satisfied with the mining results, they can change the algorithm or reexecute the data mining process.

3. Construction of Data Mining Model for Large-Scale English Test Results

3.1. Demand Analysis of e-Commerce Logistics Cloud Service. According to the principle of data warehouse and data mining, this paper adopts the related technology of data mining, such as association rules and decision tree, combined with the actual situation of the school and many performance data, using advanced data warehouse and data mining technology, and the school needs to form the data mining rules of students' performance [21]. Based on these rules, we implement multilevel analysis and classification of the performance data and finally mine the data.

Data preprocessing is the preparation work before data mining, which aims to provide the standard format and targeted data for data mining, reduce the amount of data processing of data mining algorithm, improve mining efficiency, and ultimately improve the accuracy of the model. Data preprocessing methods include data cleaning, data integration, data conversion, and data specification [22]. The main task of data preparation is to preprocess the original data according to the goal of data mining and audit and judge the data source before data mining. High-quality data is the premise of data analysis and the guarantee of the reliability of analysis conclusions. These data include student number, name, gender, age, and major, scores of each course in the entrance test, and scores of each course. Based on the analysis of the specific situation of the students' relevant data, referring to the effective variables predicted by the unified English test, considering the differences between the written test and the computer-based test, we only retain the "student number," "entrance age," "online learning situation," "entrance test English," "entrance test computer," "college English 2," "college English 3," "average course," and "unified college English." There are 10 variables of "degree English" and "degree English." After removing the records with invalid exception and variable value of "0," the retained records are saved in the form of an Excel data table as the training and test data source of decision tree construction. In addition to retaining the variables of "student number" and "degree English," all other variables are deleted and saved as the data source of prediction target.

This research will use the data classification technology of data mining technology to realize the prediction of large-scale English achievement. It will go through several steps, such as data extraction, data preprocessing, decision tree construction, decision tree optimization, and prediction rule implementation. The specific implementation process is shown in Figure 3.

4. Analysis of the Results of Large-Scale English Test

In order to facilitate the analysis of the results of large-scale English tests, this paper takes the relationship between whether college students can pass CET-4 and various factors as an example. According to the results of CET-4, more than 425 points mean that they have passed the test, while less than 425 points mean that they have not passed the test. In this paper, the CET-4 score of 425 or above is recorded as 1, while the score of 425 or below is recorded as 0.

4.1. The Relationship between Examination Results and Students' Gender and Major Categories. In the 2821 person-times of CET-4, there are 1518 male students, 566 of whom have passed the test, accounting for 37.29% of the total number of male students, and 1303 female students, of whom 839 have passed the test, accounting for 64.39% of the total number of female students. As shown in Figure 4, whether CET-4 is passed or not has a lot to do with gender; girls are easier to pass CET-4 than boys.

Figure 4 shows the proportion of people taking CET-4 and the passing rate of the three colleges. The number of participants in the three colleges is basically the same. The passing rates of "college 1" and "college 2" are 41.6% and 43.7%, respectively, but the passing rate of "college 3" is 62.5%, which is 20% higher than that of the former two colleges, which is obviously higher than that of the former two colleges. The passing rate of the two colleges is basically the same, but there is a significant difference with the other college. In other words, college is not the most fundamental factor affecting the passing rate of CET-4. The following will be from a professional point of view to see the impact of CET-4 through the degree.

As shown in Figure 5, the passing rate of some majors is very high, while the passing rate of some majors is very low. The passing rate of two majors in CET-4 is more than 70%. They are "major 12" and "major 13." The passing rate of three majors is more than 60%. The passing rate of one major is just more than 50%, while the passing rate of "major 12" is more than 50%. The passing rate of CET-4 of 8 majors is between 40% and 50% and that of the other two majors is less than 40%. Whether a student can pass CET-4 is closely related to his major.

4.2. The Relationship between Examination Results and Semester. In this paper, "semester proportion" refers to the proportion of the total number of participants in each semester, and "pass rate" refers to the ratio of passing CET-4 in each semester. Some students take part in CET-4 earlier than others, which is distributed in three semesters. The number of participants in different semesters is different, and the passing rate of each college is very different. As shown in Figure 6 below, the number of students taking CET-4 in the second semester and the third semester is significantly less than that in the fourth semester, but the passing rate of CET-4 in the fourth semester is the lowest, and the passing rate of CET-4 in the third semester shows a downward trend. In

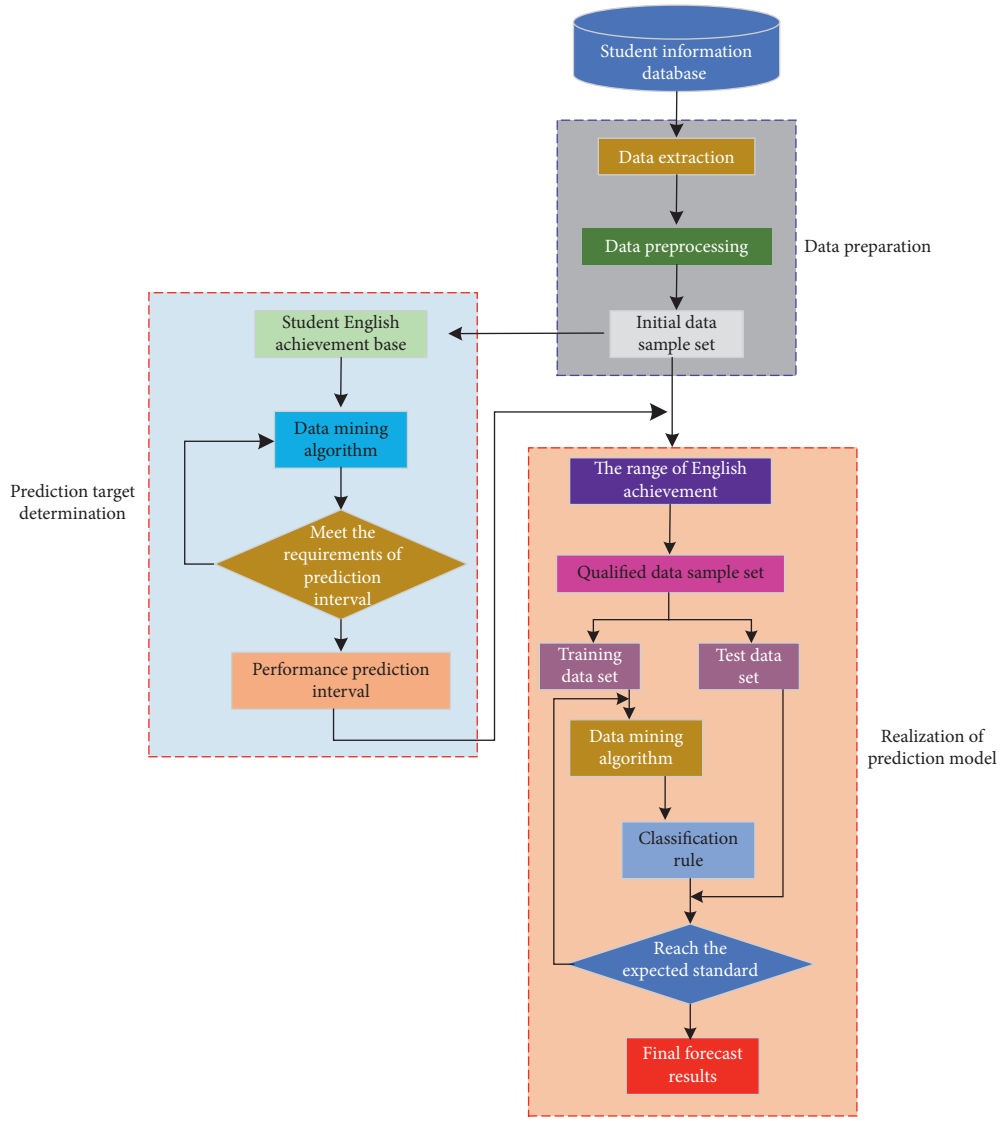

Figure 3: Process of English achievement prediction.

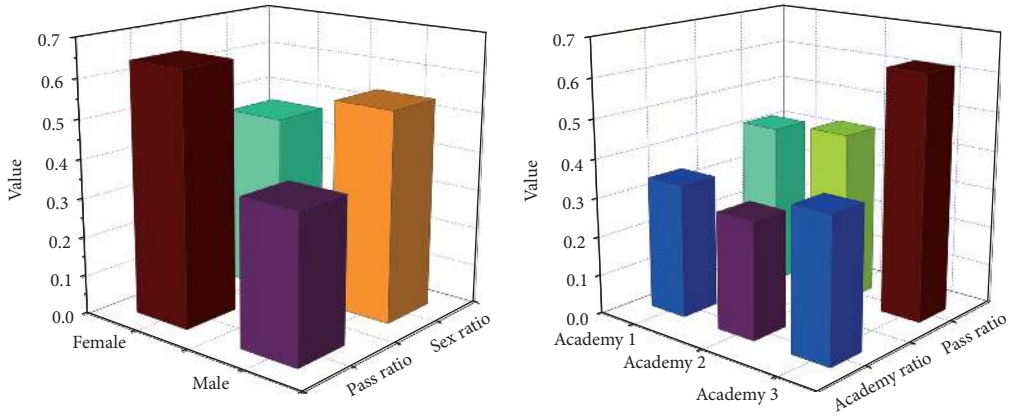

Figure 4: Relationship between examination results and students' gender and major categories. (a) Relationship between examination results and gender. (b) Relationship between examination results and colleges.

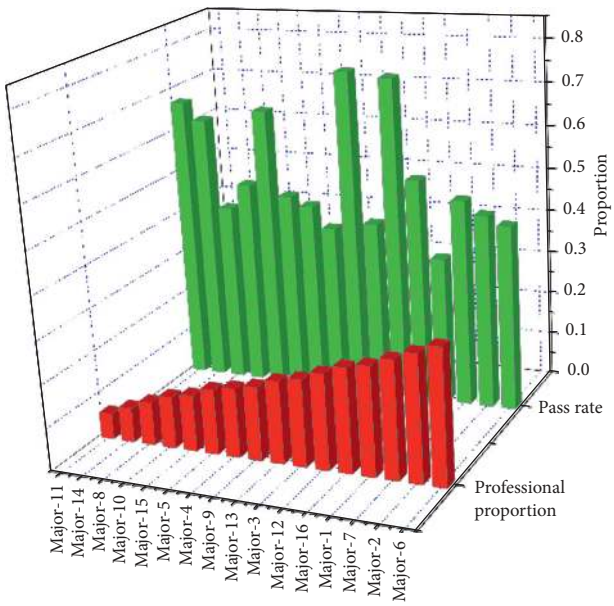

Figure 5: The relationship between passing rate and specialty.

other words, the earlier the test is, the more likely it is to pass CET-4, and the later the test is, the less likely it is to pass CET-4. Under normal circumstances, the longer the preparation time is, that is, the longer the learning time is, the more conducive it is to pass the exam. But as shown in Figure 6, the result is just the opposite. Therefore, this paper infers that the students who take the examination in "semester 2" and "semester 3" are selected from the scores of previous semesters. In order to verify this inference, this paper analyses the students' college English scores in the three semesters and finds that the lowest scores of "English score 1" in the first semester before CET-4 are 74, 60, and 14, respectively. The scores of "English score 1" in the first two semesters before CET-4 are much higher than those in the last semester. Therefore, students who can take CET-4 in the first two semesters must have a high "English score 1"; otherwise, they will not have the chance to take CET-4.

4.3. The Relationship between the Passing of CET-4 and the Scores of Entrance English. Every year when freshmen enter the school, they must organize a freshmen entrance examination, mainly English examination and mathematics examination. The purpose is to divide the students into different classes and levels of teaching. In order to better understand the relationship between entrance English score and CET-4 pass, this paper gives a bar chart between them.

It is found from Figure 7 that the higher the students' scores in the entrance English test are, the more likely they are to pass the CET-4. However, there are also some cases in which students fail to pass CET-4 because of their high English scores. Many students fail to pass CET-4 because their scores exceed 90. If the entrance English score (high school English level) is regarded as the students' English foundation, the better the students' English foundation is, the more conducive it is to pass CET-4. However, a good foundation in English does not guarantee that we can pass CET-4. To pass CET-4, we need to continue to study English hard in the university. In addition, some students' entrance scores are very low or even zero, but they have also passed CET-4. Through the communication and investigation with students, it is learned that these students do not pay attention to the entrance examination organized by the school and choose to abandon the examination or deal with the examination passively, so that the entrance score is low, which does not reflect the real level of students' English. Therefore, it will not affect them to pass CET-4 smoothly.

There are four college English scores, namely, college English (1), college English (2), college English (3), and college English (4). However, the two college English scores closest to CET-4 are more closely related to CET-4 scores, which has been verified in the previous correlation coefficient diagram. Therefore, this paper only discusses the relationship between the latest two college English scores and the results of CET-4 and presents them by using Figure 6 and 7. In the two pictures, "English score 2" is the college English course score that students are studying when they take CET-4, while "English score 1" is the college English score of the previous semester relative to "English score 2."

It can be seen from Figure 8 that the higher the "English score 1" is, the more likely it is to pass CET-4. If the "English score 1" is more than 75, the more likely it is to pass CET-4. If the "English score 1" is less than 50, the less likely it is to pass CET-4. The overall change trend of CET-4 passing rate is basically consistent with that in Figure 8, and the possibility of CET-4 passing increases with the increase of "English score 2." In addition, a few students with very low "English score 2" also passed CET-4. It is verified that these students all took CET-4 in the fourth semester, but they did not take CET-3 in the third semester. They only had their usual scores, but no paper scores. Therefore, the situation of "English score 2" was very low, but they could also pass CET-4. If this situation is removed, it can be said that the change trend is consistent. This content can be verified by the correlation coefficients of "English score 1," "English score 2," and the results of CET-4, and their correlation coefficients are 0.58 and 0.62, respectively.

4.4. The Relationship between Test Scores and Reading, Listening, and Writing. The score of CET-4 is composed of listening, reading, and writing. They are used to test students' abilities in listening, reading, writing, and translating. Through the analysis of the results of the three parts and the influence on the passing of CET-4, we can better understand the ability and level of students in listening, reading, writing, and translation and provide clear aspects and ideas for the future English teaching reform. Therefore, this paper calculates the average scores of reading, listening, and writing of students in CET-4, as shown in Figure 9.

It can be seen from Figure 9 that the average scores of listening, reading, and writing are all higher than 0.6 among the subscores of CET-4, among which reading is the highest (0.678), followed by writing (0.668), and listening is the lowest (0.633), while the average scores of listening, reading, and writing are lower than 0.6 among the scores of CET-4.

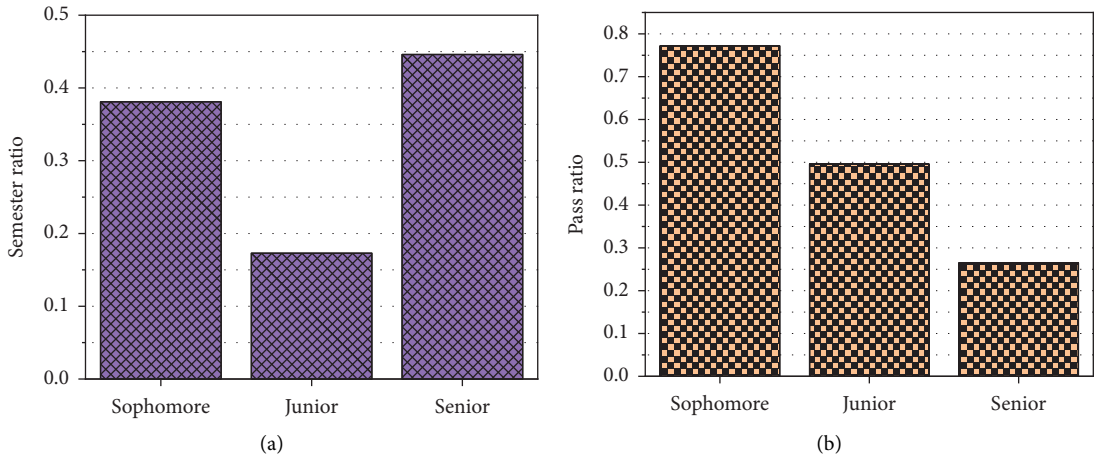

Figure 6: The relationship between grade passing and semester. (a) Percentage of semesters passing the examination. (b) Percentage of semesters passing the examination.

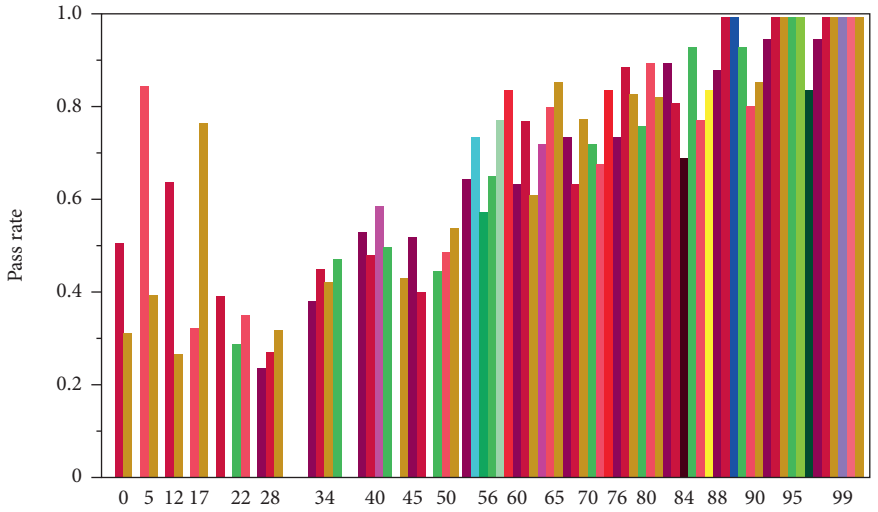

Figure 7: The relationship between CET-4 passing and entrance examination results.

The lowest score was listening (0.519), followed by reading (0.526), and the highest score was writing (0.575).

By comparing the average scores of passing CET-4 and failing CET-4, it is found that the average scores of reading, listening, and writing are 0.152, 0.114, and 0.093, respectively. Combined with the correlation between reading, listening, and writing and the results of CET-4, we can draw the following conclusions: in CET-4, reading is the biggest factor that affects students' passing of CET-4; writing and translation have the least influence on the passing of CET-4, while listening has more influence than writing and translation, but less than reading comprehension; no matter whether it is CET-4 pass or not, listening is the short board for most students to pass CET-4.

There are obvious differences between CET-4 reading comprehension and high school reading comprehension, mainly involving two aspects. On the one hand, the types of reading comprehension are different. CET-4 requires students to have a large vocabulary. In addition, we need to master the characteristics of question types and reading skills and also pay attention to the input and output of language. The low scores of CET-4 are mainly manifested in the poor foundation of listening, the weak mastery of listening vocabulary, and the lack of listening training skills. Students should listen more, practice more, combine extensive listening with intensive listening, and have a deep understanding of different listening subjects. It can be seen from Figure 9 that, among the students who have passed CET-4, the students with high scores in writing are more than those with high scores in reading comprehension and listening, but there are not a few students with low scores. The students who want to pass CET-4 should also pay attention to the accumulation of writing.

5. Prediction and Analysis of the Results of Large-Scale English Test

According to the correlation between students' entrance English score, English score 1, English score 2, and CET-4, they are relatively important characteristics and should be input variables for prediction and classification. For

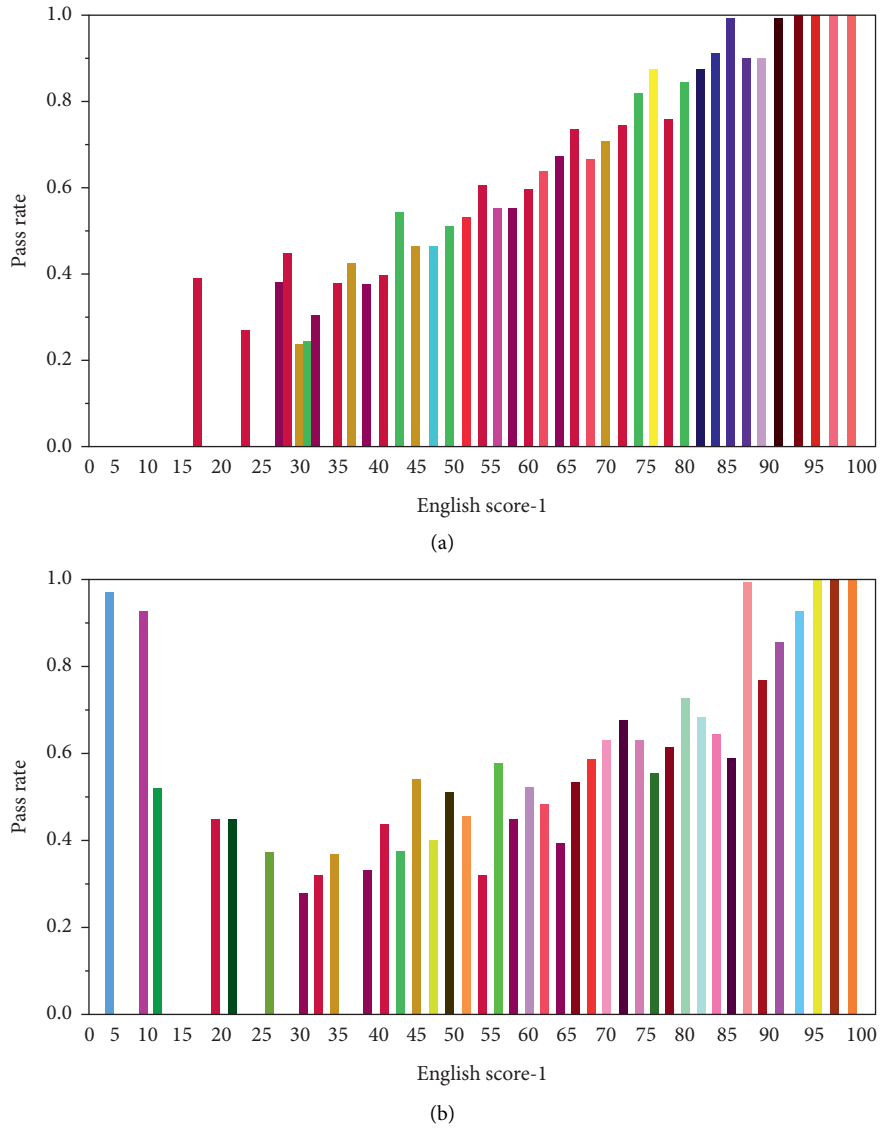

Figure 8: The relationship between CET-4 passing and English examination results. (a) The relationship between CET-4 passing and English-1. (b) The relationship between CET-4 passing and English-2.

students' gender, college, and major, based on the classification effect, this paper chooses to investigate the importance of features from the perspective of classification effect (accuracy rate, accuracy rate, and recall rate). The idea of selection is to remove features one by one. If a feature is removed and the effect of classification and prediction is significantly reduced, it means that the feature is relatively important and should be retained. If the feature is removed and the effect of classification is not significantly reduced or the effect of classification is better, it means that the feature is not important or even counterproductive and should be removed.

After data cleaning and processing, there are 2674 pieces of data, in which the results of CET-4 have been marked. If we use all the data to train the model and use it to predict the unlabelled data, we will not be able to evaluate the prediction effect of the algorithm at all. In order to solve this problem, this paper divides the data into training set and test set according to 7 : 3: one part is used to train the number and characteristics of the nearest neighbours, that is, to train the k-nearest neighbour model, and the other part is used to evaluate the prediction and classification effect of the k-nearest neighbour model. In order to avoid a tie, let K take an odd number, from 1 to 35, a total of 18. It lists eight cases of input features. The first case is the case with the most input features, and the other cases are the cases without some features.

From Figure 10 that in the accuracy chart, when k is small, the accuracy of eight cases increases with the increase of k. when $k \geq 11$, the accuracy of various cases basically tends to be stable. Although it fluctuates with the increase of k, the amplitude of fluctuation is relatively small, and the amplitude of most fluctuations is less than 2%. "Case 1" is one of the most input features, but from the accuracy chart, its prediction accuracy is not the highest. "Case 6" has only three features,

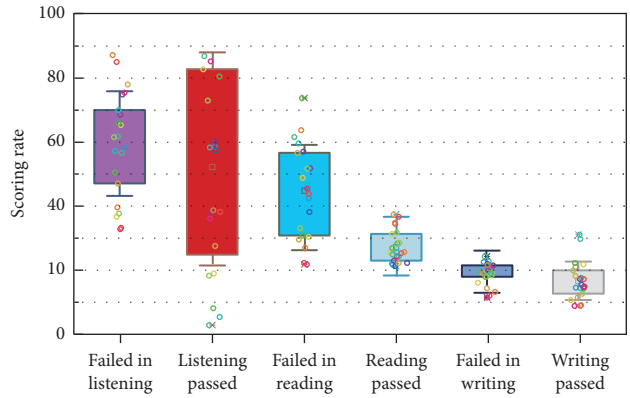

Figure 9: Relationship between test scores and reading, listening, and writing.

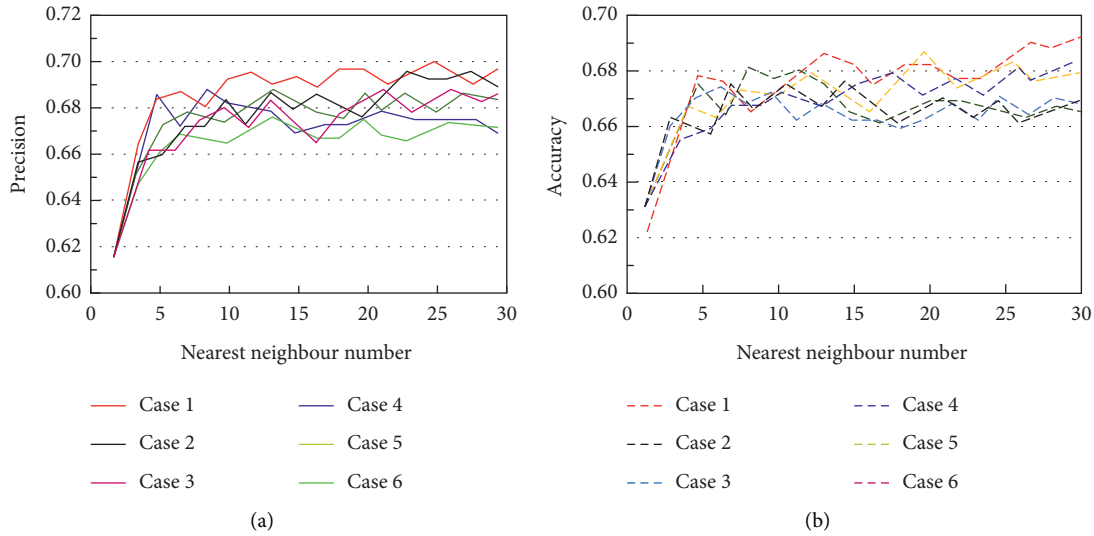

Figure 10: Graph of prediction test. (a) Chart of precision. (b) Chart of accuracy.

namely, "entrance English score," "English score 1," and "English score 2." Although the input features are the least, in this period, the average prediction accuracy is the highest, and the fluctuation range is relatively small. In the accuracy chart and recall chart, the fluctuation range of each curve is larger than that in the accuracy chart and f-value chart, which indicates that the accuracy and recall of prediction are greatly affected by the nearest neighbour number k and characteristics. In these eight cases, the accuracy rate, recall rate, and f-value of "Case 8" are not the highest, while the accuracy rate, recall rate, and value of "Case 6" are relatively high. Considering the purpose of this paper, we hope to predict whether the students can pass the CET-4 with high accuracy. Combined with the meaning of each index, we only choose to predict the results of the CET-4 with the three characteristics of "entrance English score," "English score 1," and "English score 2."

6. Conclusion

This paper studies the application of data mining technology in the analysis and prediction of English test scores in our school, which provides a decision-making basis for the scientific management of improving students' scores. This paper first analyses the history and value of the topic and then describes the applicable expertise of facts mining technology. Then, referring to the current papers, this paper analyses the unique elements that may additionally have an effect on university English take a look at scores and contains out records preprocessing for the records set that can be obtained. Then, it selects the first-rate features, establishes a model, and makes use of information mining equipment to mine and analyse the students' English take a look at scores. It is of great practical significance to select the model with high accuracy, further optimize the parameters, make good use of the data, and then take targeted measures to guide the teaching reform. There are still some problems that need further research and improvement. Because of the confidentiality of students' information, much information cannot be obtained, such as students' college entrance examination results, registered residence information, and other pieces of information that may be related to student achievement. In the future, we can get more comprehensive information through searching and other ways and continue to carry out data mining analysis.

Data Availability

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

Conflicts of Interest

The authors declare that they have no known conflicts of interest or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This research was supported by the Ningxia Normal University School-Level Scientific Research Project: Research on the Path to Promote Rural Education Revitalization in Southern Ningxia, Approval no. NXSFYB2112.

References

[1] F. Hu, Z. Li, R. Hu, and Y. Zhou, "Research on the deformation characteristics of shear band of soil-rock mixture based on large scale direct shear test," *Chinese Journal of Rock Mechanics and Engineering*, vol. 37, no. 3, pp. 766–778, 2018.

[2] M. Allam, J. Chao, and N. Xiaohong, "The application of power network data mining and optimization processing based on distribution network," *IOP Conference Series: Earth and Environmental Science*, vol. 242, no. 2, p. 22045, 2019.

[3] R. N. Boubela, K. Klaudius, H. Wolfgang, N. Christian, and M. Ewald, "Big data approaches for the analysis of large-scale fMRI data using Apache spark and GPU processing: a demonstration on resting-state fMRI data from the human connectome project," *Frontiers in Neuroscience*, vol. 9, no. 62, pp. 492–495, 2016.

[4] W. Yaqin, "Research on the data mining analysis system implementation based on network news," *Technical Bulletin*, vol. 55, no. 19, pp. 677–683, 2017.

[5] C. B. Li, W. Dong, and S. S. Lin, "Research on analytic model of project chain risk elements transmission based on risk ranking method and data mining," *International Journal of Multimedia and Ubiquitous Engineering*, vol. 11, no. 11, pp. 353–364, 2016.

[6] S. Liang, "Research on the method and application of Map-Reduce in mobile track big data mining," *Recent Advances in Electrical & Electronic Engineering (Formerly Recent Patents on Electrical & Electronic Engineering)*, vol. 14, no. 1, pp. 20–28, 2021.

[7] M. Holena, L. Bajer, and M. Scavnicky, "Using copulas in data mining based on the observational calculus," *IEEE Transactions on Knowledge and Data Engineering*, vol. 27, no. 10, pp. 2851–2864, 2015.

[8] S. V. Kornilkov and V. L. Yakovlev, "Methodology-based approach to the research in the area of mineral exploration and mining based on systematic, integrated, inter-disciplinary and innovation strategy," *Journal of Materials Science Materials in Medicine*, vol. 21, no. 1, pp. 440–445, 2015.

[9] L. J. Guo, "Practical exploration of college English teaching reform for arts and sports majors," *Contemporary Educational Practice and Teaching Research*, vol. 7, no. 12, pp. 192-193, 2020.

[10] J. Wang, "Can the development of non-cognitive ability explain the gender differences in the distribution of academic achievement?—empirical evidence from the urban functional development area of Beijing," *World Economic Review*, vol. 9, no. 6, pp. 49–69, 2018.

[11] Y. L. Wu, "A study on the relationship between English learners' personality types and English reading ability," *Journal of Minnan Normal University*, vol. 32, no. 2, pp. 74–82, 2018.

[12] K. R. Zhao, "Exploring the current situation of FLTRP's high school English textbooks," *Intelligence*, vol. 6, no. 3, pp. 117–119, 2016.

[13] C. Q. Zhu, "Analysis of CET-4 score based on SPSS—taking Wanxi university as an example," *Journal of Suzhou Institute of Education*, vol. 17, no. 1, pp. 135–137, 2014.

[14] C. J. Xu and G. B. Zhu, "Research and application of data mining in NCRE score analysis," *Computer Applications and Software*, vol. 37, no. 8, pp. 64–67, 2020.

[15] W. C. Guo, "Analysis of the application of data mining technology in the analysis of self-taught examination results," *Journal of Jilin Radio and TV University*, vol. 12, no. 8, pp. 37-38, 2019.

[16] D. D. Cheng, "Application of clementine data mining tool in computer rank examination results," *Journal of Qilu University of Technology*, vol. 31, no. 6, pp. 52–56, 2017.

[17] J. Dai and J. Li, "Research on the relationship between students' individual attributes and test scores based on data mining," *China Education Informatization*, vol. 14, no. 3, pp. 49–51, 2017.

[18] W. J. Yin, "Analysis and research of data mining in higher vocational computer level one examination results," *Fujian Computer*, vol. 33, no. 01, pp. 50-51, 2017.

[19] L. Sun, K. Zhang, and B. Ding, "Research and implementation of online education performance segmentation prediction based on data mining: a case study of undergraduate adult degree English test," *China Distance Education*, vol. 10, no. 2, pp. 22–29, 2016.

[20] H. Lin, "Application of data mining technology in the analysis of computer grade examination results," *Information and Computer*, vol. 12, no. 7, pp. 131-132, 2016.

[21] Z. Q. Liu, "Research on the application of data mining in the analysis of computer grade examination results in sports colleges," *Journal of Jianghan University*, vol. 44, no. 04, pp. 377–381, 2016.

[22] S. Zeng, "Analysis and research of data mining technology in computer rank examination results," *Computer Knowledge and Technology*, vol. 11, no. 13, pp. 14-15, 2015.