WILEY | Hindawi

Retraction

Retracted: Intelligent Prediction Model of CET-4 Passing Rate Based on Multisource Heterogeneous Data Mining

Wireless Communications and Mobile Computing

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope

(2) Discrepancies in the description of the research reported

(3) Discrepancies between the availability of data and the research described

(4) Inappropriate citations

(5) Incoherent, meaningless and/or irrelevant content included in the article

(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] X. Cheng, "Intelligent Prediction Model of CET-4 Passing Rate Based on Multisource Heterogeneous Data Mining," *Wireless Communications and Mobile Computing*, vol. 2022, Article ID 1189835, 10 pages, 2022.

WILEY | Hindawi

Research Article

Intelligent Prediction Model of CET-4 Passing Rate Based on Multisource Heterogeneous Data Mining

Xiaofang Cheng [ORCID]

Henan University Minsheng College, Henan 471400, China

Correspondence should be addressed to Xiaofang Cheng; 2016123687@jou.edu.cn

With the growing software of English in society, English gaining knowledge is turning into extra and greater important. The software and lookup of fact mining science in English educating and gaining knowledge will turn out to be a new improvement trend. This study is based on the information of non-English majors' learning situation of Public English course and CET-4 scores. It selects the data related to CET-4 scores as the characteristics, takes CET-4 passing rate as the goal, and makes use of selection tree classification mannequin to predict the passing likelihood of CET-4. After comparison and analysis, it is observed that the multisource heterogeneous statistic mining science can precisely predict students' grades, which verify the software of multisource heterogeneous information mining technological know-how in university English studying and have an impact on of the mannequin on the end result prediction. The results of College English final examination and gender are the main factors affecting the passing rate of CET4. The higher the score of College English final examination, the higher the passing rate of CET-4. The results of this study show that girls' passing rate and average score of CET-4 are higher than boys'. Other factors have little influence on CET4 passing rate and can be ignored. There is a significant positive relationship between the scores of CET-4, CET-1, and Putonghua. After analysing the number attributes of the model, blended with the statistical consequences of information analysis, this paper offers some ideas and hints to enhance the instructing administration and the passing charge of CET-4.

1. Introduction

With the advancement of facts science in the subject of education, multisource heterogeneous statistics round the everyday instructing operation of faculties and universities, students' getting to know and different factors are gathered and saved through quite a number facts administration platform. Massive data can help schools understand and master the teaching operation and students' learning situation. However, at present, the use of these data in most ordinary universities is still in the simple stage of data collection and analysis, and there is no in-depth mining [1]. How to make the most of the data stored in the database? For most colleges and universities, it is a huge challenge. In latest years, with the upward jab of the notion of huge data, record mining and information analysis technological know-how have steadily penetrated into the subject of education, which

makes it viable to discover the price of multisource heterogeneous records in college database.

Data mining science is to extract the implicit and tremendous know-how by means of analysing a massive quantity of incomplete statistics in the database. The software of facts mining science to the evaluation of university students' overall performance can correctly analyse and technique the overall performance records at a deeper level. On one hand, it can enhance the work effectivity of training managers and assist them modify training selections in actual time. On the different hand, it can assist instructors understand the studying repute of students, so as to arrange instructing extra flexibly and successfully [2]. At the identical time, it can additionally assist college students enhance their tutorial overall performance scientifically and pertinently. Using fact mining technological know-how to analyse the performance, we can discover the guidelines or patterns hidden in the

overall performance data and mine more than one affiliation relationship hidden in the overall performance data. The two fundamental lookup instructions of multisource heterogeneous statistics primarily based on students' CET-4 rankings in the area of schooling are academic statistics mining and mastering evaluation [3]. The former refers to the complete use of mathematical statistics, desktop mastering, and information mining technological know-how and strategies to method and analyse the large statistics of education. Through information modelling, we can discover the correlation between learners' studying consequences and variables such as studying content, getting to know sources, and educating behaviour, so as to predict learners' future mastering trend; the latter refers to the complete use of the theories and strategies of facts science, sociology, pc science, psychology, and studying science, through the processing and evaluation of generalized schooling large data, the usage of acknowledged fashions and strategies to provide an explanation for main troubles affecting learners' learning, evaluate learners' getting to know behaviour, and supply synthetic adaptive remarks for learners.

The typical overall performance evaluation will cowl up the precise elements that have an effect on students' performance and cannot discover the relationship between the factors. It can solely get an end result of the discussion, however, cannot play a guiding position in the action. With the speedy improvement of the data age, the know-how contained in the big schooling statistics has attracted greater interest [4]. Therefore, the most essential hassle of schooling informatization is to use facts mining science to analyse all types of training and educating information, discover the significant understanding hidden in the data, and make this understanding higher serve the schooling and instructing workers. In this study, record mining is used to analyse the relationship between curriculum and CET, and a fact mining mannequin is installed primarily based on the records of university English curriculum and CET rankings of non-English Majors in undergraduate colleges. By exploring the relationship between the two, it offers a foundation for enhancing educating administration and enhancing the excellent of teaching.

The rest of this paper is organized as follows: the related work is discussed in Section 2. The third section analyses the related theory and technology of data mining. Section 4 constructs the prediction model of CET-4 results. Section 5 analyses the prediction model and expounds the application of the conclusion. Section 6 summarizes the whole paper.

2. Related Work

The two predominant lookup instructions of information in the subject of training are academic information mining and mastering analysis. The former refers to the complete use of mathematical statistics, computer getting to know, and information mining science and techniques to procedure and analyse the huge statistics of education. Through information modelling, we can discover the correlation between learners' gaining knowledge of consequences and variables such as getting to know content, studying sources, and edu-

cating behaviour, so as to predict learners' future mastering trend. The latter refers to the complete use of the theories and strategies of facts science, sociology, laptop science, psychology, and getting to know science, through the processing and evaluation of generalized schooling massive data, the use of recognised fashions and techniques to give an explanation for essential issues affecting learners' learning, consider learners' studying behaviour, and supply synthetic adaptive remarks for inexperienced persons [5]. In current years, multisource heterogeneous information mining has acquired big interest and development. In latest years, there are many advances in multisource heterogeneous information mining and utility research.

Relevant students use record mining science to learn about training data and get extra lookup results. Liang uses stepwise regression and neural community science to analyse university students' mastering fulfilment and its influencing elements [6]. With the assist of statistical evaluation and visualization, affiliation rule algorithm, and clustering algorithm, Holena et al. analysed a massive wide variety of studying records produced in the method of community learning, and in accordance to the evaluation results, they gave some ideas and hints on the supervision and administration of community gaining knowledge of technique [7]. Kornilov and Yakovlev apply the couple of linear regression mannequin to predict the educational overall performance of college students in the hybrid college publications of bodily lecture room and cloud getting to know platform and include out educating intervention in accordance to the prediction to enhance their getting to know effect. The nested ensemble getting to know approach is used to assemble the classification prediction mannequin of online learners' educational performance, which offers a reference for the lookup on the influencing elements and prediction modelling of online learners' tutorial performance and additionally contributes to the exercise of online mastering tutorial early warning, tutorial overall performance prediction and contrast [8]. The clustering algorithm is utilized to the evaluation of English associated direction rankings and gaining knowledge of records of undergraduate college students in community diploma education, and the subdivision prediction of person diploma English take a look at rankings is realized.

By sorting out the lookup effects of home students on the utility of academic facts mining, it is located that many researches are incredibly cantered on record mining in the discipline of online education, whilst there is much less lookup on the statistics generated by way of usual lecture room instructing saved in the database. Network training platform can reveal the total procedure of students' learning and construct an extra complicated and scientific assessment gadget by way of placing up a giant range of facts series factors [9]. Compared with the online training platform, the standard facts have the troubles of much fewer monitoring factors in the mastering method and single rating structure. These issues convey challenges to information mining [10]. Based on the statistics mining lookup on e-learning, this paper constructs a choice tree classification mannequin by using including modelling attributes, decomposing the

composition of grades, reworking the which means of grades, and the usage of the common facts saved in the database, such as the nonpublic records of three grades of non-English majors, college English direction ratings, and grade examination scores, the likelihood of college students passing CET-4 is predicted, and the important attributes of the mannequin are analysed.

3. Related Theories and Technology of Data Mining

3.1. Overall System Structure of Data Mining. We followed the methods of Chu and Ma [11]. The average shape of records mining machine is ordinarily composed of the following parts: database and records warehouse. Its capacity is that the information mining object is composed of database, information warehouse, records shape, or different statistics database. Data cleaning and records integration operations are normally used to manner these information objects. The database or record warehouse server is accountable for studying the applicable statistics in accordance to the user's facts mining request. The data mining system is shown in Figure 1.

Knowledge stock is the area expertise that statistic mining needs, which will be used to information the search method of records mining, or to assist consider the mining results. The user-defined threshold used in mining algorithm is the easiest area know-how [12]. Data mining engine, which is the most primary aspect of information mining system, typically includes a set of mining characteristic modules to entire the mining features such as qualitative induction, affiliation analysis, classification induction, evolutionary computation, and deviation analysis. The sample understanding comparison module can assist the statistics mining module focal point on mining extra significant sample information in accordance to the pastime popular [13]. Of course, whether or not the module can be blended with the statistics mining module is associated to the unique mining algorithm used in the statistics mining module. Obviously, if the fact mining algorithm can be mixed with the know-how comparison method, it will assist to enhance the effectivity of information mining. Visual person interface helps customers talk with information mining gadget itself. On one hand, customers publish their mining necessities or duties to the mining gadget via the module and furnish applicable know-how wished for mining search; on the different hand, the device indicates or explains the outcomes or intermediate consequences of statistics mining to customers thru the module. In addition, the module can additionally assist customers browse the content material of record objects and record definition patterns, consider the mined sample knowledge, and show the mined sample understanding in a number form. From the viewpoint of statistics warehouse, fact mining can be viewed as the superior stage of online evaluation and processing; however, the fact evaluation capacity of information mining based totally on a range of superior facts appreciation applied sciences is a long way greater than the online evaluation and processing
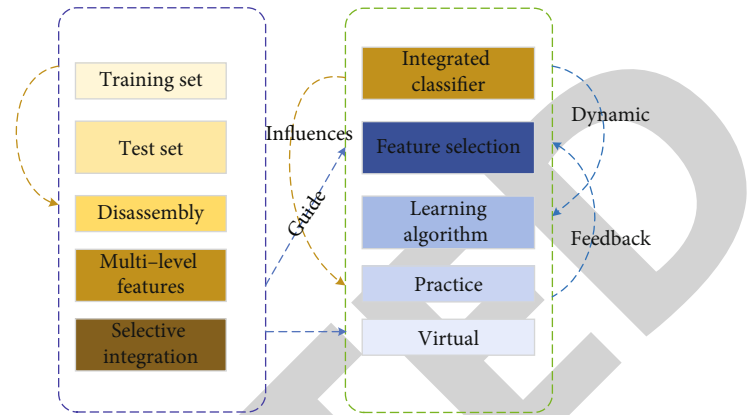

Figure 1: The structure of data mining.

characteristic of record warehouse based totally on information aggregation [14].

3.2. Process Analysis of Data Mining. The utility of information mining algorithm to the evaluation of university students' overall performance wishes to go via three stages: data preprocessing stage, data mining stage, and result evaluation stage. The data mining process is shown in Figure 2.

(1) *Data Preprocessing.* This stage is used to supply information data that can be used for direct processing and analysis, so in this stage, it is fundamental to integrate, filter, and method the supply statistics accurately in accordance to the records statistic's necessities of the algorithm, so as to reap the evaluation effects with excessive reliability [15]. This phase of the work occupies a massive percentage in the entire overall performance evaluation work. In the evaluation of university students' performance, the facts used for statistics mining may additionally contain more than one database or disciplines, so it is critical to gather and type these data, get rid of the semantic fuzziness between statistics sources, deal with the current statistics defects, and type them into a unified and standardized information layout [16]. There might also be a giant quantity of inappropriate statistics in the information evaluation house shaped by way of the series of supply data. These statistics do now not grant help for the improvement of statistics mining, however, will extend the workload [17, 18]. Therefore, the 2 d content material of records practise is the resolution of data. The chosen records must be the applicable records content material that is beneficial for evaluation and can efficaciously slim the processing range. There may additionally be noise problems, incomplete issues, or inconsistent troubles in the filtered information [19]. At this time, information preprocessing operation is additionally wanted to similarly enhance and enrich the facts shape in the records evaluation database to make sure the reliability and credibility of the evaluation results. In order to facilitate the algorithm analysis,
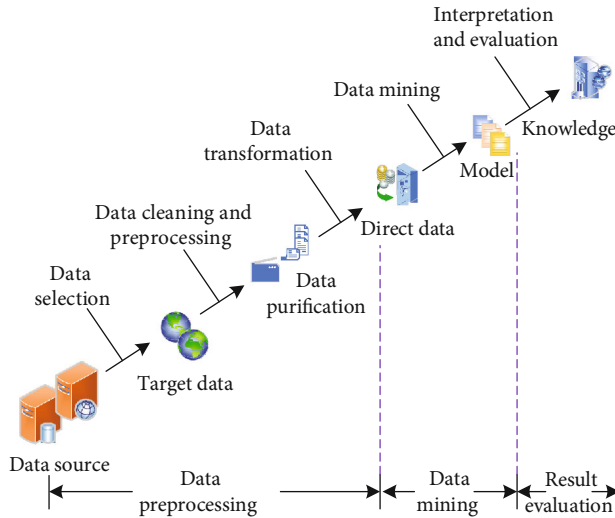
Figure 2: Process of English ache process of data mining.

it is additionally vital to radically change the attribute area data in the database into recognizable and processable coding facts [20]

(2) *Data Mining*. This work is the government phase of the total scholar fulfilment analysis. It desires to observe a range of statistics mining algorithms to procedure and analyse the facts statistics in the database and explore the handy interior members of the family or information map [21]. First of all, we want to decide the mining goal or task and then pick out the gorgeous mining algorithm in accordance to the mining goal to assemble the statistics mannequin and the precise parameters that want to be analysed and use the mannequin to mine and analyse the applicable parameters in the database, discover out the affiliation regulations and information regression shape that meet the requirements, and supply the sample expression that can be used for comparison and analysis. In sensible application, after algorithm selection, information mining can be immediately chosen to entire records mining robotically [22, 23]

(3) *Result Evaluation*. After the completion of records mining, customers want to consider and decide the bought sample evaluation outcomes or sample expressions to see whether or not they are tremendous and can meet the desires of overall performance analysis. If customers are no longer cosy with the mining results, they can alternate the algorithm or re execute the facts mining manner [24]

4. Construction of Prediction Model for CET-4 Results

The fact classification technological know-how of information mining science is used to recognize the prediction of CET-4 results, which goes via the steps of statistic extraction,

statistic processing, choice tree construction, end result prediction, and choice tree optimization, as shown in Figure 3.

4.1. Model Building. Data preprocessing is the preparation work before data mining, which aims to provide 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 [25, 26]. Data preprocessing methods include data cleaning, data integration, data conversion, and data specification. According to the research objectives of this paper, the basic information of students of three grades, college English course scores, and three types of grade examination scores are extracted from the database of the university, and the data are processed according to the following rules: (1) the basic information of students only retains five attributes: student number, name, gender, college, and major. (2) The results of college English course include usual scores, final scores, and final scores; and the number of semesters of college English course in three grades is 2 semesters, 3 semesters, and 4 semesters, respectively. The author divides all the college English course scores into three categories: usual, midterm, and final. After summation, the average scores are obtained by taking the number of semesters as the denominator. This paper classifies the number of semesters of college English courses to form the learning duration attribute, which is used to measure students' usual classroom performance, students' English knowledge level and examination ability, whether to complete college English courses and the learning duration of college English courses. (3) Students can participate in three types of grade examinations organized by the school, and the number of times is unlimited. The three tests are CET-1, CET-4, and Putonghua proficiency test. In the process of data preprocessing, this paper only takes the highest score of each type of grade examination, excluding the rest. As the problem studied in this paper belongs to the classification problem, the CET-4 score is divided into "qualified" and "unqualified" according to 425, if the score is greater than or equal to 425, it will be deemed as passing; if the score is lower than 425, it will be deemed as failing, forming the attribute of "whether to pass the CET-4." After excluding the invalid records and preprocessing the data according to the above three rules, a data set with 7867 samples is formed, including 2611 samples passing the fourth level and 5256 samples not passing the fourth level. The sample number of 2015, 2016, and 2017 students in the data set is 2806, 2614, and 2447, respectively.

After starting SPSS modeler and creating a new stream file, select variable file in the source submenu at the bottom of the interface and drag it into the panel. Double click the "variable file" icon in the panel and select "import file" in the pop-up editing interface. Select the C 5.0 node in the modelling tab, drag it into the panel, and connect it with the filter node generated by executing the feature selection node. Set "output type" of "C 5.0" node to "decision tree," select "cross validation," and set "fold times" to 10. Select "expert" mode, the initial value of "build severity" is set to 85, the initial value of "minimum records per branch" is set to 5, and check "global pruning." After executing the
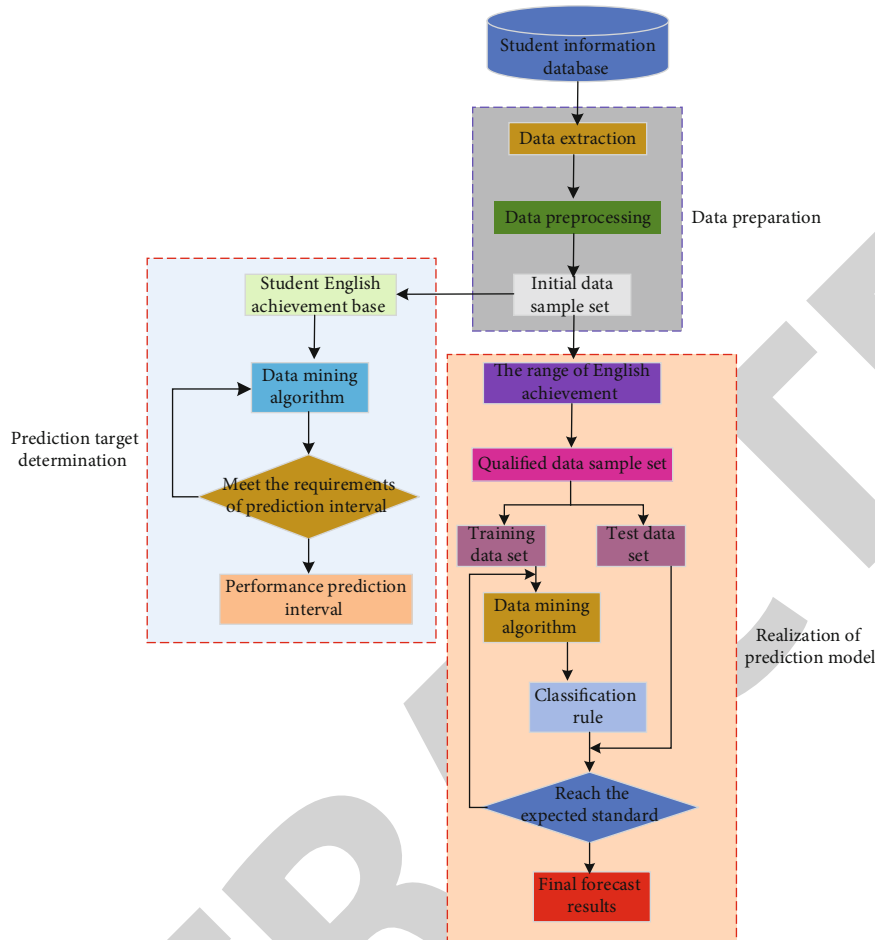
FIGURE 3: The flowchart of constructing the prediction model of CET-4 passing rate.

"C 5.0" node, the "pass level 4" icon appears in the management area and data flow design area, and the modelling is completed. Double click the icon to view the modelling results. Select the analysis node in the output tab to connect it to the go through level 4 icon. Run the analysis node to view the prediction accuracy of the model, as shown in Figure 4.

4.2. Model Optimization. Through the above steps, the selection tree mannequin includes 9 attributes, the cost of "construction severity" is 85, the cost of "minimum wide variety of documents per sub branch" is 5, and the prediction accuracy is 79.62%. The prediction accuracy of the selection tree mannequin is high; however, it is nonetheless unsure whether or not the prediction accuracy of the mannequin is optimal. In view of the reality that "average rating of university English commonplace assessment" and "average rating of university English customary assessment" are no longer used in modelling; however, these two attributes are associated to English learning. In addition, it is feasible to alter the values of "construction severity" and "minimum wide variety of information per branch." In this paper, the mannequin optimization steps are as follows:

(1) The common rating of university English well-known contrast and the common rating of university English traditional overall performance are introduced to the selection tree mannequin to shape the prediction end result two and the prediction end result 3

(2) The "construction severity" parameter determines the pruning degree of the decision tree. Increasing the value can get a more concise small tree, and decreasing the value can get a more accurate tree. In this paper, we try to set the values of "construction severity" to 75, 65, and 55 in order to form prediction result 4 to prediction result 6

(3) The "minimum number of records per sub branch" parameter can be used to limit the number of partitions in any branch of the tree. Increasing this value helps to prevent over training with noisy data. Since the data type used in this paper is relatively simple, and the invalid data has been deleted in the preprocessing process, in the process of model optimization, we try to reduce the value of "minimum number of records per branch" in turn, and set it
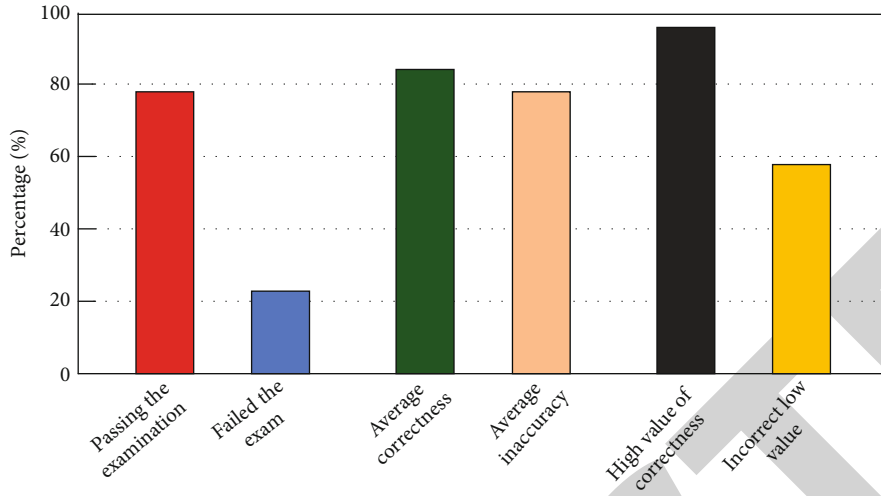
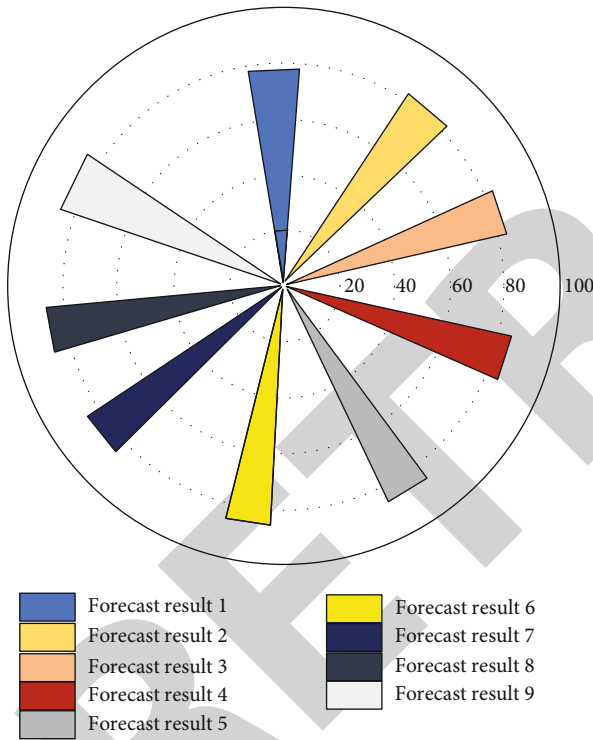FIGURE 4: Accuracy analysis of model prediction.

FIGURE 5: Optimization of decision tree.

to 4, 3, and 2, to form prediction result 7 to prediction result 9. The optimization results of decision tree formed according to the above steps are shown in Figure 5

As can be seen from Figure 5, in the first step of the optimization process, the accuracy of the choice tree mannequin can be extended via including attributes associated to English semester whilst maintaining the values of "pruning severity" and "minimum variety of information per sub branch" unchanged. In the second step, when the range of attributes stays at 9 and the minimal quantity of documents

per subdepartment stays unchanged, the prediction accuracy of the mannequin will be multiplied with the aid of decreasing the price of "construction severity." It should be noted that the values of prediction result 5 and prediction result 6 are equal, which indicates that reducing the value of "construction severity" to 65 and 55 does not help to improve the prediction accuracy of the model. Finally, 55 is chosen as the final parameter of "construction severity." In the third step, when the number of attributes is kept at 9 and the value of "construction severity" is set to 55, the parameter value of "minimum number of records per sub branch" is decreased. By decreasing the parameter value, it can be found that when the parameter value is reduced to 3, the prediction accuracy reaches the highest, and when the parameter value is reduced to 2, the prediction accuracy decreases. To sum up, this paper selects prediction result 8 as the final result. In this case, the value of "construction severity" is 55, the value of "minimum number of records per sub branch" is 3, the depth of decision tree is 12, and the number of included nodes is 435.

5. Model Analysis and Conclusion Application

5.1. Model Analysis. In this study, the selection tree classification technique based totally on C 5.0 algorithm in records mining is adopted. Based on the simple records of students, university English direction ratings, and a number grades of examination scores, in the records mining surroundings of SPSS modeler, with the aid of including modelling attributes, adjusting mannequin parameters and different steps, the prediction mannequin of CET-4 passing likelihood is subsequently constructed. When building a selection tree in SPSS modeler environment, the significance of every attribute can be seen via double clicking the "prediction results" node. The importance value of each attribute is shown in Figure 6.

By analysing Figure 6 and combining with the model, the following conclusions can be drawn:

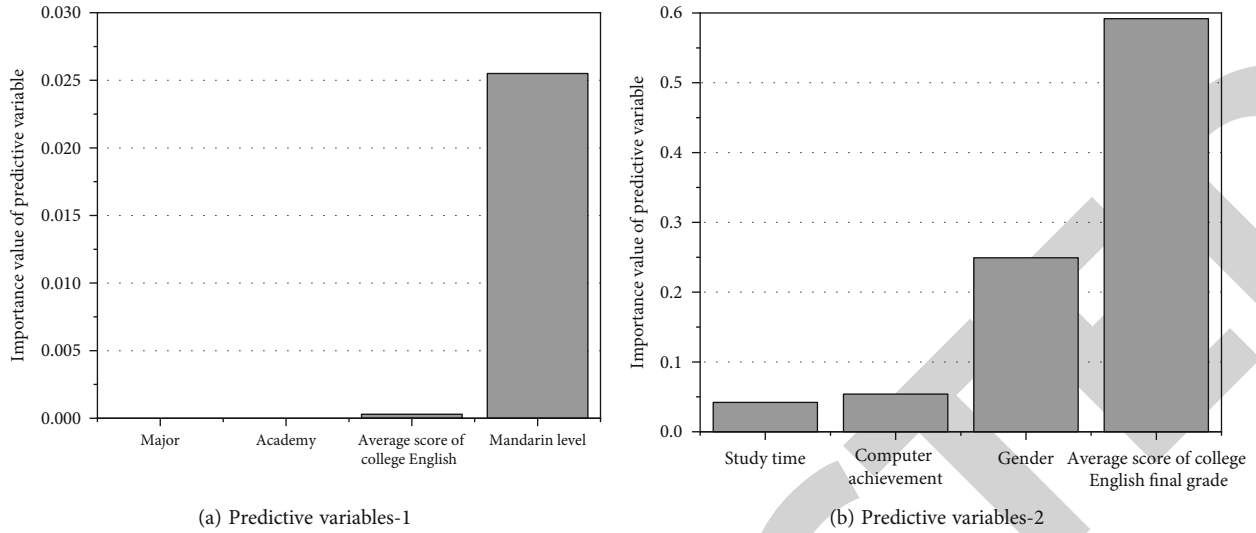(1) The average score of college English final examination is the most important to pass the prediction of

(a) Predictive variables-1

(b) Predictive variables-2

Figure 6: Importance value of each predictive variable.

(a) Comparison of the number of people
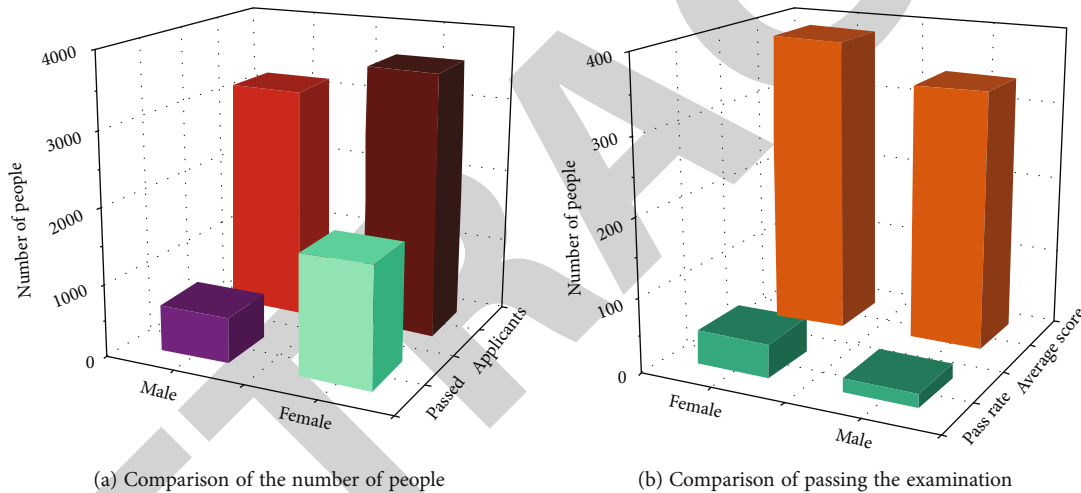
(b) Comparison of passing the examination

Figure 7: Average score and passing rate of CET-4 for male and female students.

CET-4. The main reason is that the requirements of the final examination of college English course are close to those of CET-4, and the types and quantity of questions are basically the same. From the examination form, each are written exams and use reply playing cards to do questions. Both are the consideration of students' understanding degree and examination mentality. Therefore, the probability of passing CET-4 will naturally increase if we get a good result in college English examination

(2) "Gender" is the second most important feature in the model, which shows that there is a huge difference in the passing rate of CET-4 between boys and girls

From the analysis of data samples in Figure 7, it can be seen that the passing rate and average score of girls are higher than boys.

(3) The two attributes of "CET-1 score" and "Putonghua score" can affect the prediction results of the model, which indicates that there is something in common between CET-4 and cet-1. From the prediction results of the model, we can find that there is a significant positive relationship between the scores of CET-4, CET-1, and Putonghua. That is to say, students with higher scores of CET-4 also have higher scores of cet-1 and Putonghua

(4) The attribute of "length of study" indicates that the length of study period has a certain influence on whether students pass CET-4. According to the training plan of 2014, 2015, 2016, and 2017 students, the number of Public English teaching semesters is, respectively, 4 semesters, 3 semesters, and 2 semesters, and the school forbids students to apply for CET-4 in the first semester. Combined with the data
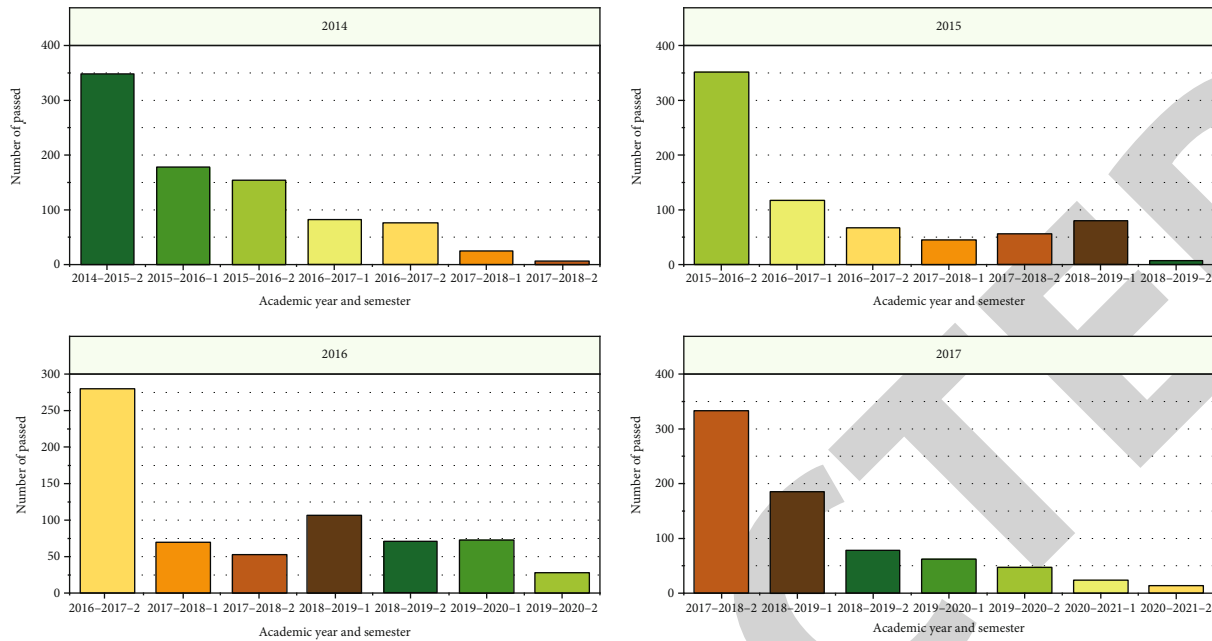
Figure 8: Statistics of the number of students who passed CET-4 in each academic year and semester.

in Figure 8, it can be found that the number of CET-4 students who pass CET-4 in the first semester of college English class can keep in three digits, and the number of students who pass CET-4 will drop sharply after the end of the course

(5) In view of the fact that the usual performance is the evaluation given by the teacher after the comprehensive investigation of the students' attendance, the quality of homework completion, classroom participation, and other indicators, these indicators are closely related to the students' learning attitude towards the course. Therefore, the average score of college English can be used as an attribute to measure students' attitude towards curriculum learning. This attribute still plays a certain role in the modelling, which indicates that students' learning attitude can have a certain impact on whether they can pass CET-4

5.2. Result Application. First, teaching administrators should pay attention to the construction of school style of study, guide students to change from passive learning to active learning in senior high school, help students establish correct learning attitude, and shape a positive and upward learning atmosphere in school. Learning attitude is closely related to learning outcomes. There is no similarity between the contents and forms of CET-4, Putonghua Proficiency test, and CET-4, but there is a positive correlation among the scores of the three tests, which indicates that students' attitude towards CET will have an impact on the test results.

Second, reform college English teaching and create high-quality English courses. College English courses are usually offered from the first semester to the fourth semester. During this period, students are still in the process of changing

from passive learning to active learning. College English classroom teaching plays an important role in this period. High quality courses are conducive to arouse students' interest in learning and help to improve the passing rate of CET-4.

Third, the school administrators should pay attention to the huge gap between boys and girls in passing rate of CET-4. From the data in Figure 8, it can be calculated that the gap between boys and girls in the passing rate of level 4 is as high as 26.1%. Although there are various reasons for this phenomenon, previous studies have also shown that women do have language advantages over men. Female college students are more inclined to make good grades in various examinations, especially in the large-scale CET-4, their goal, However, in addition to gender differences, there are other reasons for the huge gap between boys and girls in the passing rate of CET-4, which is worthy of further study.

6. Conclusions

(1) The results of College English final examination and gender are the main factors affecting the passing rate of CET4. The higher the score of College English final examination, the higher the passing rate of CET-4. The results of this study show that girls' passing rate and average score of CET-4 are higher than boys', the gap between boys and girls in the passing rate of level 4 is as high as 26.1%. Other factors have little influence on CET4 passing rate and can be ignored. There is a significant positive relationship between the scores of CET-4, CET-1, and Putonghua. The number of students who pass CET-4 will drop sharply after the end of the course

(2) With the speedy improvement of a range of online training platforms, a massive wide variety of new

facts are continuously built-in into the college database, however, face-to-face study room educating is nonetheless the most important shape of college education. This potential that whilst the new facts are growing, there are nonetheless a massive wide variety of usual statistics being accumulated and stored. With the nonstop improvement of academic statistics mining research, academic fact mining and evaluation strategies are greater abundant, and the fee of regular instructional statistics will be rediscovered. Compared with the new data, the largest benefit of ordinary schooling statistics lies in the big quantity of information and the longer collected years of data. Therefore, the mining of large normal statistics is extra probably to locate the deep-seated laws in the improvement manner of faculties and universities, teachers' instructing, and students' learning, and supply help for enhancing the instructing management, enhancing the instructing first-class and advertising students' learning

(3) There are other reasons for the huge gap between boys and girls in the passing rate of CET-4, which is worthy of further study

Data Availability

The figures used to support the findings of this study are included in the article.

Conflicts of Interest

The author declares no conflicts of interest.

Acknowledgments

References

[1] F. Hu, Z. Li, R. Hu, Y. Zhou, and R. Yue, "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] A. Maalla, C. Jia, and X. Ning, "The application of power network data mining and optimization processing based on distribution network," *IOP Conference Series: Earth and Environmental Science*, vol. 242, article 022045, 2019.

[3] R. N. Boubela, K. Kalcher, W. Huf, C. Našel, and E. Moser, "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. Yuqin, "Research on the records mining evaluation gadget implementation based totally on community 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 Multime-*

dia and Ubiquitous Engineering*, vol. 11, no. 11, pp. 353–364, 2016.

[6] S. Liang, "Research on the method and application of mapreduce 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," *Knowledge & Data Engineering IEEE Transactions*, vol. 27, no. 10, pp. 2851–2864, 2015.

[8] S. V. Kornilov and V. L. Yakovlev, "Methodology-based method to the lookup in the location of mineral exploration and mining primarily based on systematic, integrated, interdisciplinary 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 university English educating reform for arts and sports activities majors," *Contemporary Instructional Exercise and Educating Research*, vol. 7, no. 12, pp. 192-193, 2020.

[10] J. Wang, "Can the improvement of non-cognitive capability give an explanation for the gender variations in the distribution of tutorial achievement? -empirical evidence from the city practical improvement location of Beijing," *World Economic Review*, vol. 9, no. 6, pp. 49–69, 2018.

[11] N. Chu and W. Ma, "Distribution of large-scale English test scores based on data mining," *Complexity*, vol. 2021, Article ID 5531595, 10 pages, 2021.

[12] Y. L.,. Y. H. J. Wu, "A learn about on the relationship between English learners' character kinds and English studying capability," *Journal of Mannan Normal University*, vol. 32, no. 2, pp. 74–82, 2018.

[13] K. R. Zhao, "Exploring the present day state of affairs of FLTRP's excessive college English textbooks," *Intelligence*, vol. 6, no. 3, pp. 117–119, 2016.

[14] C. Q. Zhu, "Analysis of CET-4 rating based totally on SPSS – taking Wani college as an instance," *Journal of Suzhou Institute of Education*, vol. 17, no. 1, pp. 135–137, 2014.

[15] C. J. Xu and G. B. Zhu, "Research and software of statistics mining in NCRE rating evaluation," *Computer Applications and Software*, vol. 37, no. 8, pp. 64–67, 2020.

[16] W. C. Guo, "Analysis of the software of records mining science in the evaluation of self-taught examination effects," *Journal of Jilin Radio and TV University*, vol. 12, no. 8, pp. 37-38, 2019.

[17] D. D. Cheng, "Application of clementine records mining device in pc rank examination outcomes," *Journal of Qila University of technology*, vol. 31, no. 6, pp. 52–56, 2017.

[18] J. Dai and J. Li, "Research on the relationship between students' man or woman attributes and check rankings based totally on statistics mining," *China Education Informatization*, vol. 14, no. 3, pp. 49–51, 2017.

[19] W. J. Yin, "Analysis and lookup of information mining in greater vocational laptop stage one examination effects," *Fujian Computer*, vol. 33, no. 1, pp. 50-51, 2017.

[20] L. Sun, K. Zhang, and B. Ding, "Research and implementation of on line training overall performance segmentation prediction based totally on statistics mining: a case learn about of undergraduate person diploma English check," *China Distance Education*, vol. 10, no. 2, pp. 22–29, 2016.

[21] H. Lin, "Application of information mining technological know-how in the evaluation of laptop grade examination

consequences," *Information and Computer*, vol. 12, no. 7, pp. 131-132, 2016.

[22] Z. Q. Liu, "Research on the utility of facts mining in the evaluation of laptop grade examination outcomes in sports activities schools," *Journal of Jiangshan University*, vol. 44, no. 4, pp. 377–381, 2016.

[23] S. Zeng, "Analysis and lookup of facts mining technological know-how in laptop rank examination consequences," *Computer Knowledge and Technology*, vol. eleven, no. 13, pp. 14-15, 2015.

[24] C. Chen, T. Feng, M. Shao, and B. Yao, "Understanding the determinants of spatial-temporal mobility patterns based on multi-source heterogeneous data," *Transportation Research Procedia*, vol. 52, no. 2, pp. 477–484, 2021.

[25] Y. Zhu, Y. Zuo, and T. Li, "Modeling of ship fuel consumption based on multisource and heterogeneous data: case study of passenger ship," *Journal of Marine Science and Engineering*, vol. 9, no. 3, pp. 273–275, 2021.

[26] J. Yuan, X. Luo, Z. Li et al., "Sustainable development evaluation on wind power compressed air energy storage projects based on multi-source heterogeneous data," *Renewable Energy*, vol. 169, no. 10, pp. 178–182, 2021.