Journal of Education and Educational Research
ISSN: 2957-9465 | Vol. 5, No. 3, 2023

Study on Predicting University Student Performance Based on Course Correlation

Wanqing Ma, Yuan Yuan *, Jun Feng

Zhejiang University of Science and Technology, Hangzhou, Zhejiang, 310023, China

* **Corresponding author:** Yuan Yuan (Email: yuanyuandr@126.com).

Abstract: Performance prediction has proven to be an effective method for monitoring learning progress, managing student performance, and enhancing teaching quality. In an effort to analyze and predict students' grades in colleges and universities, a comprehensive database of student information is utilized along with big data technology to mine the correlation between courses. To achieve this, a student performance prediction model (SPCA) based on course association is proposed. The model selects 29 course grades from industrial engineering students in a particular school's class of 2018-2020. The courses are then clustered into three categories: mathematical computation, general and professional fundamentals, and practical application. This clustering is accomplished using the Self-Organizing Map (SOM) algorithm. Subsequently, the Apriori algorithm is employed to mine association rules among the courses. Finally, a decision tree algorithm is utilized to predict the grades of previous courses within the same category, based on the association rules discovered. The outcomes of this research can optimize course scheduling, assist students in planning their study plans, and provide practical reference value for improving teaching quality and teaching management.

Keywords: Performance Prediction; Course Correlation; Machine Learning.

1. Introduction

In recent years, the number of students entering colleges and universities has increased dramatically due to the continuous expansion of college enrollment. Students not only have to adapt to the new environment after entering university but also must change their lifestyles and study methods. Some students have difficulties in learning, and every year a certain percentage of students fail exams, repeat grades, or even drop out, which has a great impact on their future growth. From the student's perspective, achieving good academic performance will increase the chances of getting a job [1]. University administrators use many methods to monitor and guide students, but for a variety of reasons, it can be challenging to identify problems early on and provide students with the guidance they need. With the rapid development of information technology and the widespread use of digital education management, massive amounts of student learning and behavior data, as well as faculty education data, are being effectively recorded and stored. How to effectively use these data to analyze and evaluate students' learning and teachers' educational status and outcomes through certain technical means to both improve the relevance and effectiveness of teacher education, school decision-making, and management and provide timely information to different stakeholders to improve the quality of the educational process, it is a more urgent research topic nowadays [2].

Student performance prediction and analysis is an important tool to help educators identify students' weaknesses and improve their performance. Prediction of student achievement is one of the more mature applications of data mining in the field of big data in education and aims to estimate an unknown value of a variable describing the student, usually, the predicted value is performance, knowledge, or score [4]. This value can be numeric/continuous (regression tasks) or categorical/discrete (classification tasks). Today, the most common way of predicting student performance is classification, using algorithms such as decision trees (DT), artificial neural networks (ANN), K-nearest neighbors (KNN), and support vector machines (SVM), and applying techniques and models such as neural networks, bayes networks, regression, and correlation analysis.

However, the majority of earlier research did not take into account the interconnections between various courses and merely used basic information about students, such as their past academic achievement, dietary habits, and demographic characteristics, to predict performance. Previous research has not considered if there are parallels between the knowledge structures of other courses or whether the knowledge of one course aids in the study of another. The main contributions of this paper are as follows:

1. Find out the correlation between the courses in the existing training program, classify them into different categories, and provide a basis for the course setting sequence and the optimization of the training program.

2. Proposed a model called students performance prediction using course association (SPCA). On the basis of previous studies, relevant course grades are used as partial independent variables, and student behavior and demographic factors are combined to predict grades;

3. Verify the accuracy of the SPCA model, and the feasibility and practicability of the model is demonstrated from the perspectives of students, teachers and management departments, which provides a strong basis for effectively improving the teaching quality.

There are five sections to this paper's structure. Section I is an introduction, which mainly describes the research background and research significance of performance prediction based on course association and decision tree algorithm. Section II reviews related research. Section III is the methodology. Firstly, the process of model construction and the related theories and techniques of the study are

introduced. Then the data selection and pre-processing are performed. Finally, data analysis is conducted to mine the association relationship between courses through the graphical description, cluster analysis, and association rule algorithm. Section IV is experiments and analysis, where relevant features are selected and the established model is used to predict student performance and assess accuracy and feasibility. Finally, Section V outlines the conclusions of this paper and the outlook for future work.

2. Related Work

2.1. Achievement Prediction

In terms of predicted characteristic attributes, Alsariera YA et al [3] classified the attributes used to evaluate academic performance into seven categories: demographic factors, academic, internal assessment, communication, behavioral, psychological, and family/personal attributes, as shown in Table 1. Among these attributes, the most commonly used are attendance, grade point average (GPA), and final grades [4]. This is because they are the most significant and direct variables and because GPA have tangible value for future educational and career mobility [5]. The second commonly used attribute is the student's demographic factors (including gender, age, and nationality). Meit et al [6] found that female students tend to have more positive learning styles and behaviors, and are more self-disciplined and self-motivated than male students. The third commonly used attribute is parental education and income, which are family/personal attributes. Several researchers have also used psychological factors to predict student performance [7-9], including students' interests, learning behaviors, engagement time, and family support. However, it focuses more on qualitative data, and is difficult to obtain valid data, so it is not often used as a primary attribute factor. A number of scholars have applied student behavioral factors such as campus one-card data to predict student performance. Zhou Qing, Wang Weifang et al [11] proposed a method to predict students' academic risk based on one-card data and course classification. Song Yuntao [12] designed and implemented an achievement prediction model based on multiple linear regression using one-card data, which provides a reference basis for teachers and parents to analyze students' performance and take corresponding measures. The innovation of this paper is to add the factor of relevant courses to build a predictive model based on previous studies.

In terms of forecasting algorithms, Ma et al [10] used support vector machine algorithm and decision tree algorithm optimized by a grid search to predict student performance and used information gain to identify factors that have a high impact on student performance. The model can be used to identify important characteristics related to student performance and student examination pass rate. Yang et al [11] used the decision tree algorithm to predict student performance by building a student attribute containing performance-related attributes and unrelated attributes matrix, quantified student attributes, applied back propagation neural network algorithm to predict student performance, and identified key factors affecting student performance. Abu-Naser, S.S. et al [12] used ANN model to use factors such as student high school entrance exam scores, college performance points, and gender as indicators of based on multilayer perceptron topology model training data, student grades were predicted with an accuracy of more than 80%. Mueen, A. [13] et al. selected student data from two undergraduate courses and used three data mining methods to make predictions and compare performance and found that the plain bayes classifier had an accuracy of 86%, which was better than the other two classifiers.

Alsariera, Y.A. et al [7] counted the most used achievement prediction machine learning algorithms and algorithm accuracies for the period 2015-2022, as shown in Table 2. the most used algorithm is artificial neural network (ANN) [14-21] with the highest accuracy of 98.3% [14] and the lowest 73.5% [19]. This is followed by decision tree (DT) [22-25] with a maximum accuracy of 98.2% [23] and a minimum of 66% [26]. Then, support vector machines (SVM) [26-29] , linear regression (LinR) [30, 31], naive bayes (NB) [32-34] and K-nearest neighbour (KNN) [35-37]. Alsariera, Y.A. et al [3] compared the average accuracy of various achievement prediction algorithms for the period 2015-2022 and the results are shown in Figure 1, with the highest accuracy being artificial neural network (ANN) and decision tree (DT).

Table 1. Attributes Previously Used for Student Performance Prediction

Attribute category	Attribute
Demographic	Gender, age, nationality, place of birth, marital status, guardianship, home address, transport
Academic	Grade ID, Class ID, semester, attendance, final grade, GPA
Internal assessment	Assignments, quizzes, lab tests, midterms, daily study time, group presentations, individual reports
Family/personal attributes	Parental status, parent survey, parental satisfaction, family size, family education, parental work, parental income, free time, physical condition
Behavioral	Number of hands raises, retrieval of resources, presentations, discussions
Communication	Message, email, response time, login/logout time, time spent, word count
Psychological	Personality, motivation, environmental influences, learning strategies, socioeconomic status, learning methods

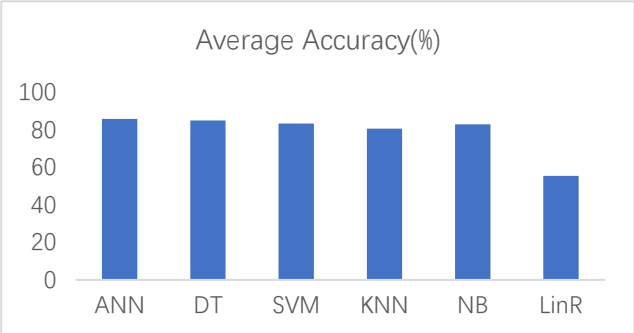


Figure 1. Prediction Accuracy Categorized by Methods from 2015 to 2021

2.2. Course Affiliations

The association is a relationship, connection, or union between two or more objects or things. Association rules were first proposed by Agrawal [38] et al. and are usually used to mine the connections between sets of items in a database and find the interrelationships between items that occur frequently in all subsets of events or attributes. Wang Xinyan et al [39] proposed a parallel split-based FP-Growth algorithm,

implemented on the Hadoop platform, to mine course grades between of association rules, which provides suggestions for students' learning styles and teachers' teaching styles. Gabriela Czibula et al [40] proposed a new classification model named SPRAR (Students Performance prediction using Relational Association Rules) that uses relational association rules (RAR) to predict students' final grades in a subject with better results. Shatnawi et al [41] used association rules to mine associations between courses in which students were enrolled and used them to guide specific students to select courses enrolled by similar students.

Table 2. Performance Prediction Algorithms and Accuracy Used During 2015-2022

Predictive Model	Study	Year	Predictive features	Accuracy (%)
ANN	Zacharis, N.Z.[14]	2016	Internal assessment	98.3
	Mondal, A. and J. Mukherjee[15]	2018	Demographic, academic, family/personal attributes	85.4
	Arunachalam, A. and T. Velmurugan[16]	2018	Demographic, academic, family/personal attributes	68.3
	Lau, E.T., L.J. Sun, and Q. Yang[17]	2019	Demographic, academic	84.8
	Aydoğdu, Ş.[18]	2019	Demographic, academic	80.5
	Li, F.[19]	2019	Demographic, academic	73.5
	Altaf, S., W.J. Soomro, and M.I.M. Rawi[20]	2019	Academic, internal assessment	97.4
	Giannakas, F., et al.[21]	2021	Demographic, academic	86.6
DT	Gupta, J. and K. Garg[22]	2021	Academic, internal assessment	98.2
	Deepika, K. and N.V. Sathyanarayana[23]	2019	Demographic	97.9
	Imran, M., et al.[24]	2019	Demographic	95.8
	Alsalman, Y.S., et al.[25]	2019	Demographic, academic	66
SVM	Agaoglu, M.[26]	2016	Academic, internal assessment	91.3
	Roy, K., et al.[27]	2018	Internal assessment	90.3
	Erickson, V.L.[28]	2019	Demographic, academic	77
	Francis, B.K. and S.B. Suvanam[29]	2019	Demographic, academic, behavioral	66
LinR	Eagle, M., et al.[30]	2018	Academic, internal assessment, behavioral	69
	Nguyen, V.A., Q. Nguyen, and V.T. Nguyen[31]	2018	Demographic, academic, internal assessment	50
NB	Vasić, D., et al.[32]	2015	Demographic, academic, internal assessment	96.9
	Kaur, P., M. Singh, and G.S. Josan[33]	2015	Demographic, internal assessment	65.1
	Hasan, R., et al.[34]	2018	Academic, internal assessment	90
KNN	Vijayalakshmi, V. and K. Venkatachalapathy[35]	2019	Demographic, academic, internal assessment	69
	Turabieh, H.[36]	2019	Demographic, familiy/personal attributes	88
	Kurniadi, D., et al.[37]	2018	Academic, family/persona attributes	95.8

He Chu et al [21] spectral clustering algorithm clusters the courses and divides them into four categories: general knowledge, basic and professional foundation, professional direction, and practical teaching. Chen Bo [22] classified 30 online open courses of Ningbo City Vocational and Technical College into four categories according to their operational effects through clustering analysis, and put forward rationalization suggestions for the operation and management of online courses. Huang Lijun [23] used factor analysis and cluster analysis to classify 35 courses of logistics management majors into seven categories: professional foundation category, job core operation category, international logistics category, professional foreign trade English category, logistics job expansion category, core backbone course category, and mathematics category. Several scholars have used the Kmeans clustering algorithm for educational data mining. Lailiyah et al [42] applied Kmeans and fuzzy C-means clustering algorithm (FCM) to cluster students' learning styles and explored the relationship between learning styles on students' knowledge absorption ability and academic performance. Mondal. B. et al [43] used the Kmeans clustering algorithm to cluster students based on their past performance and built a course recommendation system using a collaborative filtering algorithm. However, Kmeans has a problem in that the number of clusters has to be determined beforehand before clustering. However, in practical applications, especially when the data is large, it is difficult for us to determine the appropriate number of clusters before the experiment. Kohonen [44], a professor at the University of Helsinki, Finland, proposed the self-organizing map (SOM) network based on the phenomenon of biological "lateral inhibition", which is capable of self-organizing the training of input patterns. Delgado, S. et al [45] used the self-organizing map (SOM) artificial neural network model to cluster 1709189 online students enrolled in 2015-2019 at the International University of La Rioja (UNIR) and showed that specific clusters were associated with the average admission to the university and that there was a significant relationship between user interaction and higher performance. relationship.

In summary, most of the clustering algorithms in previous educational data mining studies have been applied to students themselves, and few studies have been conducted for course clustering, and even fewer have combined it with association rules, and the clustering algorithms used need to be optimized. This study attempts to construct a clustering-association-based performance prediction model called Student Performance prediction using Course Association (SPCA), which focuses on considering the influence between different courses, including the sequential row relationship of courses and the influence relationship of course knowledge, to further

explore the association between courses and continuously improve the accuracy of grade prediction.

3. Methodology

3.1. Building a Model for Student Performance Prediction Using Course Association

Most of the forecast accuracies are low when compared to the earlier models. First, the different prediction features taken into account, whose correlation is not immediately apparent. Second, the algorithms selected are different, and different algorithms train different models with different effects, and even if the same features are selected, their accuracy rates will be different. There are certain correlations between courses themselves, between courses and courses, and between course settings and course grades. In this study, the feasibility of the model is explored by taking the "Introduction to Intelligent Manufacturing" course conducted by a university's industrial engineering program as an example.

The high dimensionality and the large number of elements in the dataset and the objectives of this study require the use of data mining tools to explore the potential associations of the course with student performance data. In this study, we attempt to identify hidden relationships in a dataset consisting of independent variables without an objective function. In this case, clustering is the most appropriate technique to0 extract common patterns of behavior in the dataset. Clustering is an unsupervised technique that groups vectors in a dataset based on distance or similarity criteria, such as Euclidean distance. The grouped vectors usually share common properties, and the subsequent analysis of clustering makes it easier to discover the properties that describe them, thus obtaining a description of complex, high-dimensional datasets. Also, association rules in unsupervised techniques are a common method for mining data associations.

In order to exclude the interference of irrelevant factors, the following assumptions are made first:

1) The weight of the usual grades in each semester remains the same, and the difficulty of the test papers is approximately the same;
2) Each student's learning ability and usual performance remain the same;
3) Students devote about the same amount of time to review each course;
4) Each student is in the same examination environment;
5) Students' performance is authentic and free from cheating and other problems;
6) Teachers' marking standards remain essentially the same.

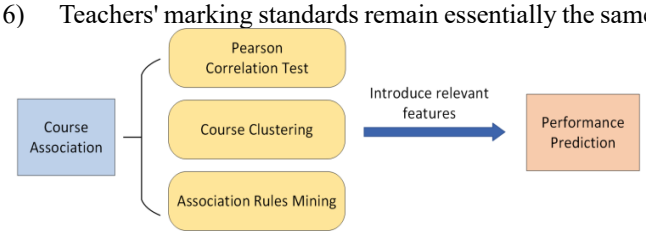


Figure 2. Model Building Process

As shown in Figure 2, the model construction process can be divided into the following steps:

1) First, we conduct a preliminary exploration of the performance data and observe the distribution characteristics and trends. Several courses with similarities in traditional concepts, such as Introduction to Marxism and Mao Zedong Thought, Probability and Mathematical Statistics, and Linear Algebra, were selected to compare the trends of grades and observe their similarity. Subsequently, the correlation coefficients of 29 courses were calculated to observe the correlation between courses quantitatively.

2) Then, course clustering was performed to cluster the 29 courses into three different categories, and the categories were named according to the characteristics of each category.

3) Third, association rule mining is performed. The clustering is only a preliminary way to derive the general categories of courses, and the specific association relationships need to be obtained by using association rule algorithms, and the relationships between course grades are inferred from the derived association rules, and then the intrinsic connections between courses are derived. For example, the association between Principles of Management-B and Economic Law-B has a lift of 1.29 at the support of 0.20 and confidence of 0.70, in this one rule, Principles of Management is the antecedent course and Economic Law is the posterior course. From this, we can conclude that there may be some intrinsic connection between the two knowledge systems, and learning management principles can improve the learning of economic law.

4) Finally, the prediction model was constructed under the condition of known course association. The model construction process is shown in Figure 3. The figure indicates the association network among some courses, and the three colors represent three-course categories. Suppose we want to predict the grade of course A, screen the courses whose antecedents are opened before it and belong to the same clustering category as some influencing factors, i.e., the grade of course D and course E, and combine the usual grade, consumption data, library reading data and college entrance examination grade data for grade prediction.

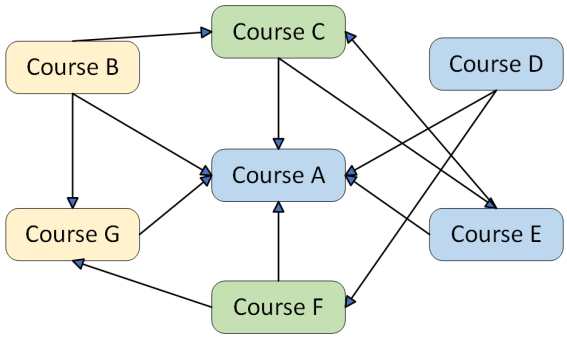


Figure 3. A Predictive Model for Student Performance Based on Course Association

3.2. Introduction of Related Algorithms

3.2.1. SOM Neural Network

The inability to guarantee independence between subject scores and linear separability between categories makes it inappropriate to classify courses using cluster analysis in statistics. Jiménez, R. et al [46] showed that educational and psychometric data can be processed efficiently using artificial neural networks. The self-organizing mapping (SOM) algorithm is an unsupervised learning neural network algorithm, a dimensionality reduction algorithm that allows the projection of high-dimensional data onto the node plane for self-organization and thus visualization and clustering of high-dimensional data. Compared with other clustering algorithms, SOM does not need to set the number of categories in advance, and it can automatically find the intrinsic laws and essential properties in the data with strong

self-adaptability. As shown in Figure 4, the SOM algorithm consists of an input layer and a competition layer (output layer). The neurons in the competition layer compete for the input vector, the more similar to the input vector, the more competitive it is, and finally only one neuron wins. Each winning neuron weight vector maintains a tendency to approximate the input vector so that the input vectors with similar features are clustered together. This is repeated so that each winning neuron becomes matched to only one specific class, thus achieving the clustering of data points.

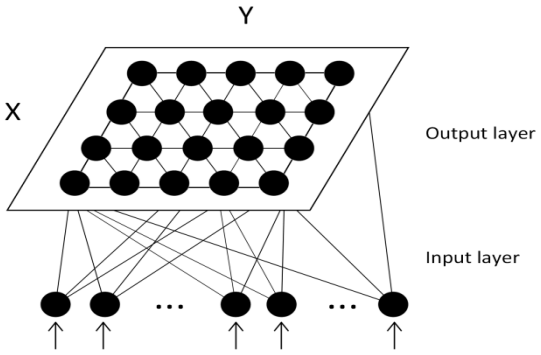


Figure 4. Topology of SOM

3.2.2. Apriori Association Rules Algorithm

Among the many algorithms for association rules, the Apriori algorithm is the most classical one, and many later algorithms are extensions and improvements of the Apriori algorithm. The Apriori algorithm is an efficient algorithm for mining frequent itemsets of Boolean association rules designed and proposed by Agrawal [47] et al. in 1994. It is based on the a priori principle that if an itemset is frequent, then its subsets must also be frequent. Conversely, all supersets of non-frequent itemsets are non-frequent itemsets. The exploration process is as follows: first, the database is scanned and the counts of each item set are accumulated. Then, the support of each item set is compared with the minimum support, and the set of items greater than or equal to the minimum support is obtained, i.e., the set of frequent 1-itemsets is generated, which is denoted as L_1. L_1 is used to generate the set of frequent 2-itemsets L_2, and then L3 is generated by L_2, and this step is iterated until the set of maximum frequent k-itemsets L_k is generated. The relevant definitions are as follows:

Definition 1. Association rules: Let $L = l_1, l_2, \ldots, l_m$ be the set of items, the transaction database *D* consists of a series of transactions with a unique flag *TID*, each corresponding to a subset on *L, i.e.,* $T \subseteq L$. The association rule of course grades can be expressed as the logical implication of $l_iA \rightarrow l_jB$, where $l_iA \subseteq I, l_jB \subseteq I$, and $l_iA \cap l_jB = \emptyset, l_iA, l_jB$ denote the grades of two different courses, respectively.

Definition 2. Support: Support indicates the probability that l_iA and l_jB occur simultaneously.

$$Support(l_iA, l_jB) = P(l_iA, l_jB) = \frac{num(l_iAl_jB)}{num(AllSamples)} \quad (1)$$

Definition 3. Confidence: The confidence level represents the probability that l_jB also occurs in the presence of l_iA.

$$Confidence(l_iA \rightarrow l_jB) = P(l_jB|l_iA) = \frac{P(l_iAl_jB)}{P(l_iA)} \quad (2)$$

However, support and confidence alone often cannot provide the value of evaluation rules, so sometimes strong association rules meet the minimum support and minimum confidence without practical significance. Therefore, in this experiment, we introduce the degree of lift to mine strong association rules satisfying the minimum degree of lift.

Definition 4. Lift: The lift of $l_iA \rightarrow l_jB$ is the ratio of the confidence of l_iA on l_jB to the overall probability of occurrence of l_jB.

$$Lift(l_iA \rightarrow l_jB) = \frac{P(l_jB|l_iA)}{P(l_jB)} = Confidence(l_iA \rightarrow l_jB)/P(l_jB) \quad (3)$$

As shown in Figure 5, to find association rules involves three steps, i.e., extracting frequent item sets from the dataset, then generating rules from these frequent items, and finally filtering the valuable strong association rules based on the boosting degree.

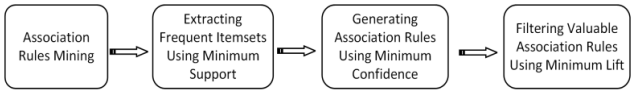


Figure 5. Association Rules Mining Process

3.2.3. Decision Tree

The decision tree is one of the most commonly used data mining classification algorithms, widely used in retail target marketing, customer retention, fraud detection, and medical diagnosis. A decision tree has a tree-like structure containing three elements: root node, internal node, and leaf node. Different values in the dataset are tested according to the tree-like structure of the decision tree, and a path from the root node to the leaf nodes represents a decision, hence the name of the decision tree. Compared to other classification techniques, decision tree classifiers are more understandable and reasonably quick. Even for small datasets, techniques like neural networks can require very long training durations, whereas decision trees can be reduced to straightforward classification rules. And compared to other classification techniques, tree classification occasionally achieves higher accuracy [48]. The decision tree classification algorithm is generally divided into two steps: first, a decision tree is formed using the training set starting from the root node and judging from the top down in this way to build a model; then the model built is used to classify the samples to be classified. The decision rules that form the decision tree are information gain, Gini coefficient, information gain ratio, etc. In this study, decision tree algorithm will be used to build the grade prediction model. GPA, grades, and demographic factors are the factors that have the highest accuracy in predicting student grades in most of the algorithms using DT. DT can handle both forms of data and performs well in massive data sets and the relationship between variables is easy to understand [49].

3.3. Data Processing

3.3.1. Data Preparation

The data for this experiment were obtained from the course grades of 134 students in industrial engineering in the 2018-2020 classes at a university in the university's academic affairs management system, including 41 in the 2018 class, 54 in the 2019 class, and 39 in the 2020 class. The school public elective courses were excluded because the data were incomplete due to the large variation among different students. In addition, some elective courses with a low number of students taking them were also excluded, and 29 courses were finally selected for this study.

3.3.2. Data Cleaning

The first step is to remove sensitive information, such as student names and ID numbers, to ensure the privacy of the information. The next step is to remove irrelevant data. Since the exported raw data included information such as the

college offering the course, examination method, course nature, instructor, major name, and class, which were not considered in this experiment, they were removed. Then similar courses were combined, and then the average method was used to take the values (rounded to one decimal place). After merging, 29 courses remained. At the same time, there are cases of missing data, mainly because: students change their majors in the middle of the course, different students take different courses, and students change their courses in the middle of the course for their reasons. For these reasons, if a student has more missing grades, the record will be deleted, and if only one or two courses are missing, the missing grades will be filled in as the average score of the course.

3.3.3. Data Conversion

Table 3. Part of the Itemsets

Student ID	Itemset
118*	C Programming-D; College English-C; ERP Principles and Applications-A; Introduction to KAB and Entrepreneurship-A; Management Statistics-B
118*	C Programming-D; College English-C; ERP Principles and Applications-C; Introduction to KAB and Entrepreneurship-C; Management Statistics-C
118*	C Programming-E; College English-D; ERP Principles and Applications-C; Introduction to KAB and Entrepreneurship-C; Management Statistics-D
118*	C Programming-E; College English-C; ERP Principles and Applications-C; Introduction to KAB and Entrepreneurship-C; Management Statistics-C
118*	C Programming-E; College English-D; ERP Principles and Applications-C; Introduction to KAB and Entrepreneurship-B; Management Statistics-D

In the classical Apriori algorithm, the transaction database mostly adopts a horizontal structure, however, the current grade database in the academic system adopts a vertical structure, where each student has multiple records and each record contains relatively more information.

The Apriori algorithm is used to mine association rules for the above-structured data, the data structure should be varied and converted into data suitable for mining, and because each course grade is evaluated differently, some course grades have specific values, while others are graded. In the course clustering, the grades must be converted to specific values to facilitate the experiment. The specific rules are: excellent: 95; good: 85; moderate: 75; passing: 65; failing: 0. And because the Apriori algorithm only supports correlation analysis of sub-typed data, student grades need to be discretized and divided into four grades of A, B, C and D. The specific rules: 90-100: A; 80-90: B; 70-80: C; 60-70: D; <60: E.

To perform association rule mining, we also need to connect these attribute-value pairs to form an item, and the item attribute is "course name - grade level" transformed from "attribute-value" pairs, and the result of the transformation is shown in Table 3. The table contains many "attribute-value" pairs, such as "C Programming-D," and each row in the table represents a transaction. In this research, we investigate the relationship between various "attribute-value" pairs that frequently occur together.

3.4. Data Analysis

3.4.1. Characteristics of the Distribution of Achievement Data

First, descriptive statistical analysis was conducted on the 30 courses' performance data. Five courses, C Programming, College English, ERP Principles and Applications, KAB and Introduction to Entrepreneurship, and Management Statistics, were taken as examples, and their distribution characteristics were listed as shown in Table 7.

The means and medians in the above table show that the average grades in each course were generally moderate, with most grades not varying much. The large variance and standard deviation indicate a large dispersion of grades across courses. Skewness is small, mostly close to 0, indicating a more symmetrical distribution of grades in most courses. The absolute value of kurtosis is small, indicating that the overall distribution of course grades is close to a normal distribution.

Then the course grades were tested for normality. Taking Outline of Modern Chinese History and College English as examples, histograms and Q-Q plots of grade distributions were drawn, as shown in Figures 6 and 7. From the histograms, we can see that there are more students in the middle score range and fewer students in the high and low score ranges for these two courses. From the Q-Q plots, it can be seen that the points on the plots drawn with the grade data of the two courses all fall approximately around the y=x line, indicating that the grades of these two courses approximately obey a normal distribution. From the results of descriptive statistics and the normal distribution test, we know that the course distribution is relatively continuous and follows a normal distribution, with a trend of "small at the end and large in the middle"

Table 4. Statistics of Grade Distribution Characteristics of Some Courses

	C Programming	College English	ERP Principles and Applications	KAB and Introduction to Entrepreneurship	Management Statistics
Max	30	47.5	48	62	54
Min	98	85.0	95	96	94
Average	67.59	70.726	80.44	83.22	75.81
Standard deviation	15.281	7.0668	9.434	6.938	8.714
Variance	233.5	49.940	88.994	48.131	75.942
Skewness	-0.233	-0.615	-1.085	-0.685	-0.163
Kurtosis	-0.664	-0.434	1.036	0.777	-0.502

For empirical reasons, if two courses are similar in nature and course content, then the course grades can verify the similarity. Take linear algebra and probability and mathematical statistics, for example, both courses are mathematics courses that involve calculation, understanding, logic, and other related knowledge. If the same student has the relevant ability, the scores obtained in these two courses should not be very different, as shown in Figure 8, we can see that the trend of the scores is very similar, reflecting that these two courses are indeed intuitively related. The same applies to the two courses of Basic Principles of Marxism, Introduction to Mao Zedong Thought and Socialist Theoretical System with Chinese Characteristics, where the scores of the two courses are not very different and most of them are concentrated between 70 and 90.

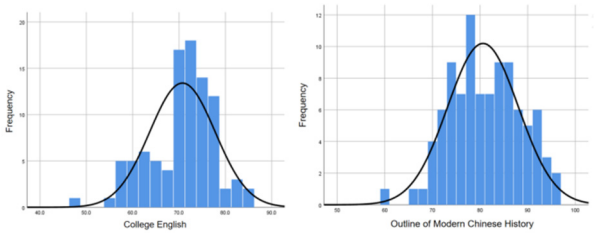


Figure 6. Histogram of Course Grade Distribution

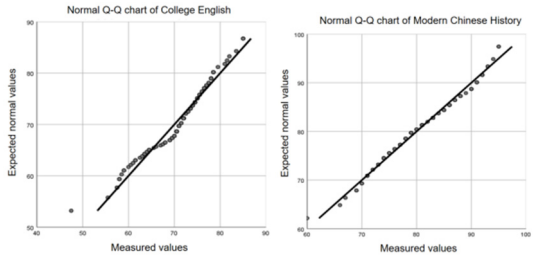


Figure 7. Q-Q Chart of Course Grade Distribution

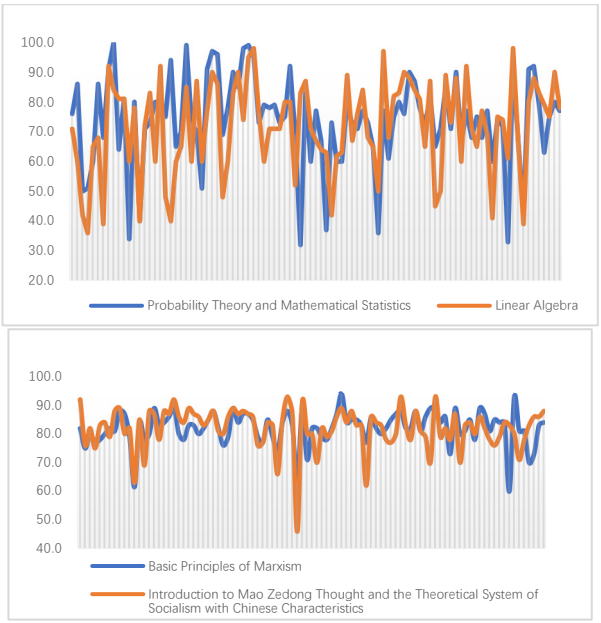


Figure 8. Performance Chart

3.4.2. Correlation Test

Measuring the closeness of correlation between two things can be achieved using correlation analysis, i.e., calculating the correlation coefficient. The Pearson correlation coefficient method was used to derive the correlation coefficient matrix for 29 courses, each represented by l_1-l_{29} respectively, and the correlation heat map was drawn as shown in Figure 9.

In this figure, the positive and negative correlation coefficients are indicated by different colors, green indicates a positive correlation and red indicates a negative correlation. The magnitude of the correlation coefficient can be indicated by the color shade, the darker the color, the stronger the correlation. As can be seen from the figure, most of the course grades are positively correlated with each other, and only a few course grades are negatively correlated, and the absolute value of the correlation coefficient of most of the course grades is concentrated between 0.3 and 0.8, indicating that the course grades have a certain degree of correlation with each other.

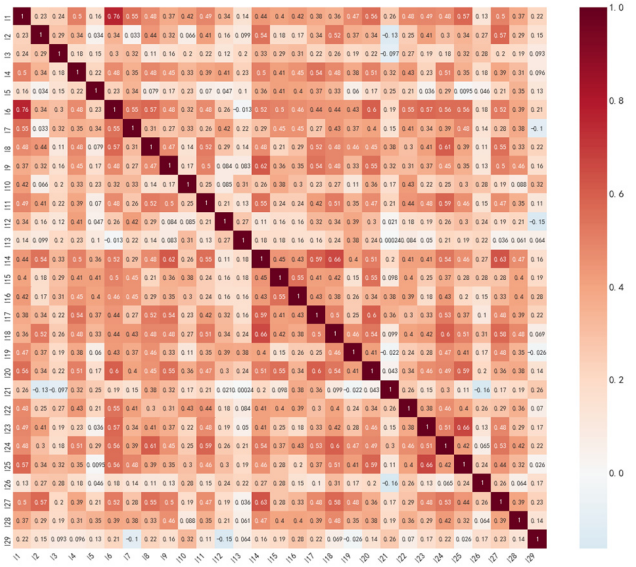


Figure 9. Correlation Heat Map

3.5. Analysis of Course Association

3.5.1. Course Clustering

To exclude the influence of other factors and to ensure the reliability of the results, we standardized the student performance. Due to the large number of courses in this study and the high dimensionality of clustering, SOM neural network was selected for course clustering. The SOM neural network was constructed using python, and the standardized 29 course grades were used as the input variables of this network. In order to take the minimum value of quantification error, the maximum number of iterations finally chosen for this experiment was 160.

Then the number of clusters is determined, and the evaluation index used in this experiment is the contour coefficient. The contour coefficient is an evaluation index of the degree of density and dispersion of classes in the following form:

$$SC(i) = \frac{b(i)-a(i)}{max\{a(i),b(i)\}} \tag{4}$$

$a(i)$ represents the average distance of sample i to other samples in the same cluster, $b(i)$ represents the average distance of all samples from sample i to some other cluster.

By clustering the original data into two, three, four, five, six, and seven categories and observing the magnitude of the contour coefficients in each case, the results obtained from the clustering are shown in Table 5 and Figure 10.

Table 5. Contour Coefficients for Different Numbers of Clusters

Number of clusters	Profile coefficients
N=2	0.15938
N=3	0.12998
N=4	0.12294
N=5	0.09716
N=6	0.09570
N=7	0.09647

Considering the total number of course samples, either too coarse or too fine classification is inappropriate, so dividing into two categories is not considered. And it can be seen from Table 5 and Figure 10 that when the number of clusters is 3, the contour coefficient is the largest and the best clustering map is drawn. Therefore, the number of output layer neurons is 3, and the 3×1 two-dimensional SOM competitive layer

neurons are used as the container for clustering. The results after course clustering are shown in Table 6.

As can be seen from Table 6, the clustering results are more consistent with the actual situation. Referring to He Chu [21] for the naming of categories, the four courses in the first category require certain logical thinking ability and computational ability, such as "Advanced Mathematics" and "Linear Algebra", so they are categorized as mathematical computation. Most of the courses in the second category are specialized basic courses, such as "KAB and Entrepreneurship Fundamentals", "Logistics Engineering", "Introduction to Intelligent Manufacturing", as well as some general education courses such as "Outline of Modern Chinese History", "Basic Principles of Marxism", etc., so they are categorized as general knowledge and professional foundation. Most of the courses in the third category require practical ability and hands-on skills, such as "Engineering Graphics", "Basic Mechanical Design", etc., focusing on the application of the knowledge gained to the real world, and are therefore categorized as practical applications.

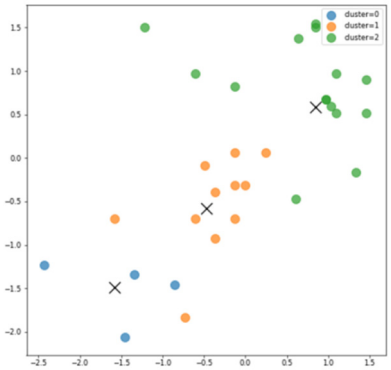


Figure 10. The Result Graph

Table 6. Course Clustering Results

Mathematical Calculation	General Studies and Professional Foundations	Practical Applications
C programming, Probability and Mathematical Statistics, Calculus, Linear Algebra	ERP Principles and Applications, KAB and Introduction to Entrepreneurship, College English, College Language, Economic Law, Lean Production, Basic Principles of Marxism, Introduction to Mao Zedong Thought and Socialist Theory System with Chinese Characteristics, Ergonomics, Ethics and Legal Foundations, Physical Education, Logistics Engineering, Situation and Policy, Introduction to Intelligent Manufacturing, Outline of Modern Chinese History	Engineering Graphics, Management Statistics, Management Information Systems, Principles of Management, Fundamentals of Mechanical Design, Computer Applications, Production Planning and Control, Data Warehousing and Big Data Mining, Database Principles and Applications, Microeconomics, Quality Engineering

3.5.2. Course Association Rules Mining

Based on the courses already clustered, mining the specific association relationships between courses requires mining association rules using the Apriori algorithm. Set the minimum support to 0.2, and dig out all frequent item sets that meet the conditions. Then use the frequent item set to generate association rules, set the minimum confidence level to 0.6, and filter according to the condition that the lifting degree is greater than 1. A total of 1079 association rules are derived, and part of strong association rules as shown in Table 7.

Table 7. Part of Strong Association Rules

Previous item	Latter-item	Support	Confidence	Lift
Introduction to Mao Zedong Thought and the Socialist Theory System with Chinese Characteristics-B	Basic Principles of Marxism-B	0.52	0.78	1.06
Basic Principles of Marxism-B	Introduction to Mao Zedong Thought and the Socialist Theory System with Chinese Characteristics-B	0.52	0.70	1.06
Management Information System-B	Quality Engineering-B	0.22	0.64	2.02
Management Information System-C	Quality Engineering-C	0.24	0.64	1.35
Quality Engineering-B	Management Information System-B	0.22	0.70	2.02
Quality Engineering-B	Management Statistics - B	0.20	0.63	1.88
Management Statistics -C	Quality Engineering-C	0.22	0.60	1.27
Introduction to Mao Zedong Thought and the Socialist Theory System with Chinese Characteristics-B	Basic Principles of Marxism-B, Introduction to Smart Manufacturing-B	0.40	0.60	1.19
KAB and Introduction to Entrepreneurship-B	Introduction to Mao Zedong Thought and the Socialist Theory System with Chinese Characteristics-B, Basic Principles of Marxism-B	0.38	0.60	1.16

Further analysis of the association rules in Table 7 provides a reasonable interpretation of them:

Rule 1: Introduction to Mao Zedong Thought and the Socialist Theory System with Chinese Characteristics-B→Basic Principles of Marxism-B (Support=0.52, Confidence=0.78, Lift=1.06);

Rule 2: Basic Principles of Marxism-B→Introduction to Mao Zedong Thought and the Socialist Theory System with Chinese Characteristics-B (Support = 0.52, Confidence = 0.70, Lift= 1.06).

As seen from the above rules, the results of the association rules obtained by the Apriori algorithm are roughly similar to the previous hypothesis that there is indeed a certain connection between “Mao Zedong Thought and Introduction to the Socialist Theory System with Chinese Characteristics” and “Basic Principles of Marxism”, and the course grades of

the two affect each other, and the results obtained from the previous descriptive statistics show that the correlation coefficient is 0.76, and the two rules above The confidence level of both rules is 70%.

Rule 3: Management Information Systems-B→Quality Engineering-B (Support = 0.22, Confidence = 0.64, Lift= 2.02);

Rule 4: Management Information Systems-C→Quality Engineering-C (Support = 0.24, Confidence = 0.64, Lift = 1.35);

Rule 5: Quality Engineering-B→Management Information Systems-B (Support=0.22, Confidence=0.70, Lift=2.02).

As shown in rules 3, 4, and 5, quality engineering and MIS grades are closely linked, and a higher or lower grade in either course affects the other course, and the degree of improvement is similar for both courses regardless of which course is used as the antecedent.

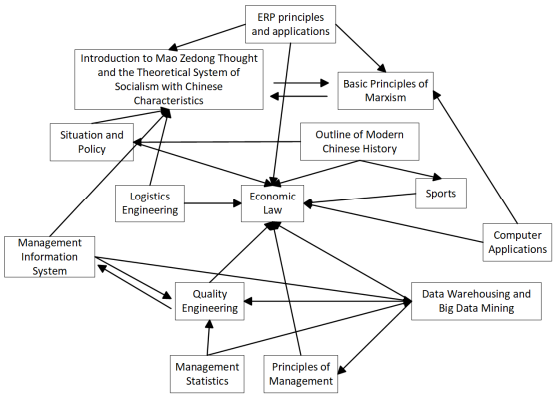


Figure 11. Part of Course Association

The final map of partial course association is shown in Figure 11. Course correlation quantifies to a certain extent the degree of influence of one course on another course, explores the intrinsic correlation between two courses, and by the relationship between the front and back terms of the rules, the sequence of course settings can be derived, which is convenient for students to lay a good foundation in the learning process, and is also conducive to the management to optimize and update the training program and set a more reasonable curriculum system.

4. Experimental Results and Discussion

4.1. Feature Extraction

Some of the strong association rules obtained with the "Introduction to Smart Manufacturing" as the latter item course are shown in Table 8. The strong association rule obtained by the Apriori algorithm is considered valid as long as the semester of the previous course is before the semester of the latter course. After checking the training plan, all the courses in the former part of the above association rules are offered before "Introduction to Smart Manufacturing", so they all satisfy the validity condition. From the clustering results, we know that the above courses belong to the same category as "Introduction to Intelligent Manufacturing", including "KAB and Entrepreneurship Foundation", "College English", "College Language", "Economic Law", "Moral Cultivation and Legal Foundation", "Sports", "Situation and Policy", and "Outline of Modern Chinese History". Therefore, the grades of these eight courses are used as some independent variables, and the model is trained to predict the grades of "Introduction to Smart Manufacturing" by combining library access and borrowing data, canteen consumption data, and enrollment slotting data

Table 8. Part of Strong Association Rules

Previous item	Latter item	Support	Confidence	Lift
College Language - B	Introduction to Smart Manufacturing-B	0.38	0.82	1.18
College English - B	Introduction to Smart Manufacturing-B	0.35	0.75	1.18
KAB and Introduction to Entrepreneurship-B	Introduction to Smart Manufacturing-B	0.47	0.75	1.08
Outline of Modern Chinese History-C	Introduction to Smart Manufacturing-B	0.32	0.75	1.08
Linear Algebra-D	Introduction to Smart Manufacturing-B	0.20	0.73	1.05
Economic Law-B	Introduction to Smart Manufacturing-B	0.40	0.73	1.05
Microeconomics-C	Introduction to Smart Manufacturing-B	0.34	0.73	1.05
Situation and Policy-A	Introduction to Smart Manufacturing-B	0.33	0.72	1.04
Computer Applications-B	Introduction to Smart Manufacturing-B	0.29	0.72	1.03
Engineering Graphics-B	Introduction to Smart Manufacturing-B	0.28	0.71	1.02
Sports-B	Introduction to Smart Manufacturing-B	0.41	0.71	1.02
Moral and Legal Foundations-B	Introduction to Smart Manufacturing-B	0.38	0.71	1.02
Computer Applications-C	Introduction to Smart Manufacturing-B	0.25	0.71	1.02
Microeconomics-B	Introduction to Smart Manufacturing-B	0.24	0.70	1.00

Previous studies have shown that students' grades are also closely related to their performance in the classroom, and their daily habits in ordinary life, such as cafeteria consumption, library access, and book borrowing, and are also related to the students' source profile. We collected data related to all industrial engineering students of a school from 2018-2020 to build a student performance prediction model. To avoid the influence of different magnitudes on the accuracy of the model, course grades, regular grades, and total cafeteria consumption were standardized and their values were transformed to be between 0 and 1. The relevant attributes and attribute descriptions are shown in Table 9.

Table 9. Related Attributes and Descriptions

Attribute category	Attributes	Attribute description
Related course grades	Final grades for related courses	0~1
Class performance	Usual grades	0~1
Cafeteria consumption data	Total cafeteria spending during the first to the second year of college	0~1
Library management data	Number of library checkouts	≥0
	Number of library visits	≥0
Enrollment slotting data	Province of origin	No.0~13

4.2. Training Models

After the data preparation was completed, we obtained the dataset used to perform student performance prediction. It was partitioned into a training set and a test set in the ratio of 8:2, and the features were input according to Table 9, and the model was trained using the decision tree algorithm.

Using the GridSearchCV function for multi-parameter tuning, the accuracy under different parameters was calculated using the ten-fold cross-validation method, i.e., the data set was divided into an equal number of ten copies, and each time the classification was performed, one copy was selected as the test set and the remaining nine copies were used as the training set, which was repeated ten times, and each time the training set data and the test set data were independent of each other and completely covered the whole data set. The final model parameters are determined as shown in Table 10, and the accuracy curves of the model under different maximum depths of the decision tree are plotted using the matplotlib package in Python as shown in Figure 12. After parameter adjustment, the generated decision tree model is shown in Figure 13, and the feature importance of each attribute is shown in Figure 14. The accuracy of the model on the test set is 90.2%, and the accuracy rate is 88.9%, which verifies the effectiveness of the model.

Table 10. Parameters of the Models

Parameter name	Parameter value
max_depth	2
criterion	entropy
random_state	20
splitter	random
min_samples_leaf	4
min_samples_split	2

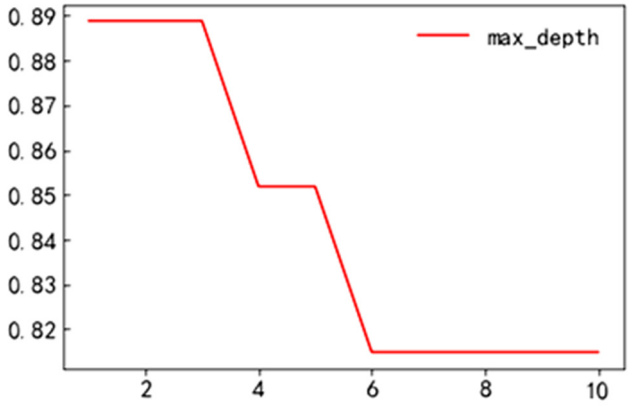


Figure 12. Accuracy Curves of the Model at Different Maximum Depths of the Tree

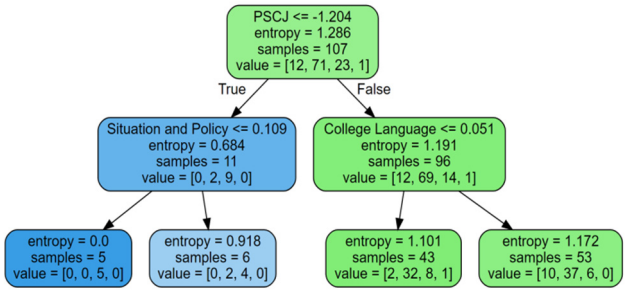


Figure 13. The Decision Tree Model for Predicting Student Performance Based on Course Association

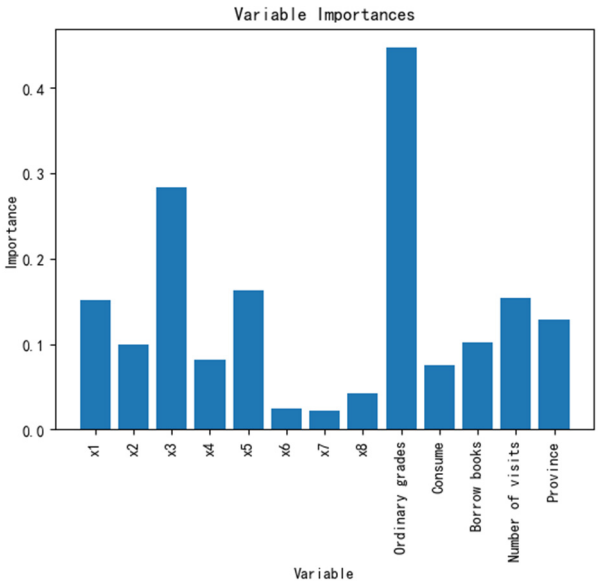


Figure 14. Variable Importances

4.3. Results Analysis

There are many metrics to evaluate the merits of decision tree algorithms, and in binary classification, the commonly used metrics are confusion matrix, precision, recall, accuracy, F1-score, etc. The formulas are as follows:

$$precision = TP/(TP + FP) \quad (5)$$

$$recall = TP/(TP + FN) \quad (6)$$

$$accuracy = (TP + TN)/(TP + FP + TN + FN) \quad (7)$$

The meanings represented by the parameters in the above equation are as follows:

TP: true positive, positive samples classified as positive samples;

FP: false positive, negative samples classified as positive samples;

TN: true negative, negative samples classified as negative samples;

FN: false negative, positive samples classified as negative samples.

In multi-categorization, these metrics are often aggregated in some way, and the three ways of aggregation are Micro, Macro, and Weighted. The following is an example of how these three approaches differ, using recall as an example:

Micro: Summarize all classes together with the following formula:

$$Recall = \frac{TP_0+TP_1+TP_2}{TP_0+FN_0+FN_1+TP_2+FN_2} \quad (8)$$

Macro: The recall rate of each category is calculated separately and then averaged, with the following formula:

$$Recall = \left(\frac{TP_0}{TP_0+FN_0} + \frac{TP_1}{TP_1+FN_1} + \frac{TP_2}{TP_2+FN_2}\right) * \frac{1}{3} \quad (9)$$

Weighted: the recall of each category is calculated separately, multiplied by the respective weights, and then summed, with the following formula:

$$Recall = \frac{TP_0}{TP_0+FN_0} * W_0 + \frac{TP_1}{TP_1+FN_1} * W_1 + \frac{TP_2}{TP_2+FN_2} * W_2 \quad (10)$$

In this experiment, the weighted parameter was chosen to calculate the indicators, and the training effect indicators of the model were obtained as shown in Table 11. The comparison graph of prediction results is shown in Figure 15 and Figure 16. It can be seen that the model has high accuracy, good training effect and high feasibility.

Table 11. Model Training Effectiveness Indicators

Evaluation indicators	Precision	Recall	Accuracy
Indicator value	90.2%	88.9%	88.9%

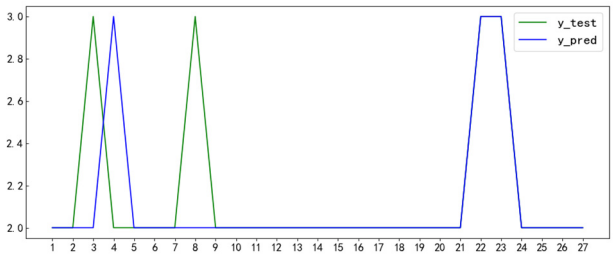


Figure 15. Comparison Line Chart of Prediction Results

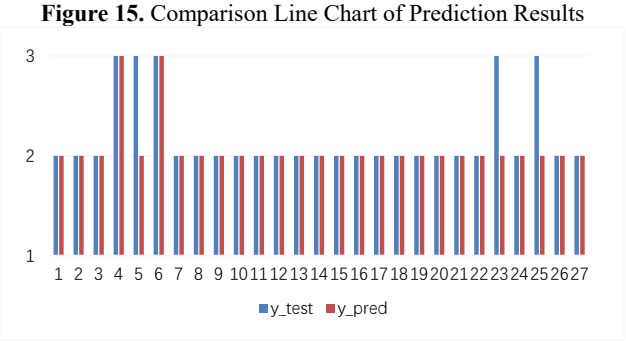


Figure 16. Comparison Bar Chart of Prediction Results

5. Conclusion

In this study, course correlations were introduced into the grade prediction model, and the grade prediction model was built using real data as the research object. With a prediction accuracy of 90.2% and an accuracy of 88.9%, the results indicate that the model and its construction process are effective. Compared with traditional grade prediction studies, the model proposed in this study incorporates multiple course correlations and provides more perspectives for the subsequent interpretation of prediction results. From the student's perspective, students can identify potential academic risks based on the prediction results; for example, if they did not achieve a satisfactory grade or even failed a course in the previous semester, they can predict that they may also fail a related course carried out in the next semester and should strengthen their study of that course to avoid the risk of failing it. From the teacher's point of view, based on the results of the grade prediction, the teacher can reasonably adjust the progress and content of the lecture, and teach the students according to their abilities to promote their understanding and absorption of knowledge. If a student fails in a subject, the teacher can use the model results to focus on the learning of subjects associated with that subject in subsequent courses. From the administration's perspective, academic administrators identify struggling students on time based on the grade prediction results and provide grade alerts.

Traditional grade warning methods only provide a rough summary of student performance data by counting students' current GPAs and credits and including students who are seriously lagging in the warning range, which is often lagging and not targeted to detect and intervene at an early stage and is ineffective. The method proposed in this study can identify struggling students at an early stage and is prospective and relevant. The establishment of an inter-course correlation model and correlation analysis, it provides new ideas and a strong basis for developing a reasonable training program and optimizing the curriculum system.

Although the current study has achieved preliminary results, there are still some elements that need to be improved. Firstly, the currently selected sample size is small, containing only 29 courses related to one major. In the future, similar courses of different majors can be selected for comparison experiments to observe what are the differences and connections between the same course and different majors. Secondly, the subjectivity of teachers' grading may cause the grades of different courses to not have uniform standards and therefore produce errors. In the future, teachers' grading standards should be standardized and a more scientific grading system should be established. In addition, the currently constructed index system is not perfect, only including objective factors, and students' subjective factors such as cognitive ability, thinking ability, and motivation are not included, and subjective factors can be introduced into the model in the future to establish a richer and more perfect index system.

Acknowledgments

This research was funded by the Liaoning Provincial Social Science Planning Fund Project (Grant No. L21CGL005) and Major Project of Teaching Research and Reform of Zhejiang Institute of Science and Technology (2022-jg06).

References

[1] Baashar, Y., et al. Predicting student's performance using machine learning methods: A systematic literature review. in 2021 International Conference on Computer & Information Sciences (ICCOINS). 2021. IEEE.

[2] Asif, R., et al., Analyzing undergraduate students' performance using educational data mining. Computers & Education, 2017. 113: p. 177-194.

[3] Alsariera, Y.A., et al., Assessment and Evaluation of Different Machine Learning Algorithms for Predicting Student Performance. Computational Intelligence and Neuroscience, 2022. 2022.

[4] Shahiri, A.M. and W. Husain, A review on predicting student's performance using data mining techniques. Procedia Computer Science, 2015. 72: p. 414-422.

[5] Bin Mat, U., et al. An overview of using academic analytics to predict and improve students' achievement: A proposed proactive intelligent intervention. in 2013 IEEE 5th conference on engineering education (ICEED). 2013. IEEE.

[6] Meit, S.S., et al., Personality Differences in Incoming Male and Female Medical Students. Online Submission, 2004.

[7] Sembiring, S., et al. Prediction of student academic performance by an application of data mining techniques. in

International Conference on Management and Artificial Intelligence IPEDR. 2011.

[8] Gray, G., C. McGuinness, and P. Owende. An application of classification models to predict learner progression in tertiary education. in 2014 IEEE International Advance Computing Conference (IACC). 2014. IEEE.

[9] Hidayah, I., A.E. Permanasari, and N. Ratwastuti. Student classification for academic performance prediction using neuro fuzzy in a conventional classroom. in 2013 international conference on information technology and electrical engineering (ICITEE). 2013. IEEE.

[10] Ma, X. and Z. Zhou. Student pass rates prediction using optimized support vector machine and decision tree. in 2018 IEEE 8th annual computing and communication workshop and conference (CCWC). 2018. IEEE.

[11] Yang, F. and F.W. Li, Study on student performance estimation, student progress analysis, and student potential prediction based on data mining. Computers & Education, 2018. 123: p. 97-108.

[12] Abu-Naser, S.S., et al., Predicting student performance using artificial neural network: In the faculty of engineering and information technology. 2015.

[13] Mueen, A., B. Zafar, and U. Manzoor, Modeling and Predicting Students' Academic Performance Using Data Mining Techniques. International journal of modern education & computer science, 2016. 8(11).

[14] Zacharis, N.Z., Predicting Student Academic Performance in Blended Learning Using Artificial Neural Networks. International Journal of Artificial Intelligence & Applications, 2016. 7: p. 17-29.

[15] Mondal, A. and J. Mukherjee, An Approach to Predict a Student's Academic Performance using Recurrent Neural Network (RNN). International Journal of Computer Applications, 2018.

[16] Arunachalam, A. and T. Velmurugan, Analyzing student performance using evolutionary artificial neural network algorithm. International journal of engineering and technology, 2018. 7: p. 67.

[17] Lau, E.T., L.J. Sun, and Q. Yang, Modelling, prediction and classification of student academic performance using artificial neural networks. SN Applied Sciences, 2019. 1.

[18] Aydoğdu, Ş., Predicting student final performance using artificial neural networks in online learning environments. Education and Information Technologies, 2019. 25: p. 1913-1927.

[19] Li, F., et al., Which Factors Have the Greatest Impact on Student's Performance. Journal of Physics: Conference Series, 2019. 1288.

[20] Altaf, S., W.J. Soomro, and M.I.M. Rawi, Student Performance Prediction using Multi-Layers Artificial Neural Networks: A Case Study on Educational Data Mining. Proceedings of the 2019 3rd International Conference on Information System and Data Mining, 2019.

[21] Giannakas, F., et al., A deep learning classification framework for early prediction of team-based academic performance. Appl. Soft Comput., 2021. 106: p. 107355.

[22] Gupta, J. and K. Garg, Reflections on Blended Learning in Management Education: A Qualitative Study with a Push-pull Migration Perspective. FIIB Business Review, 2021.

[23] Deepika, K. and N.V. Sathyanarayana, Relief-F and Budget Tree Random Forest Based Feature Selection for Student Academic Performance Prediction. International Journal of Intelligent Engineering and Systems, 2019.

[24] Imran, M., et al., Student Academic Performance Prediction using Supervised Learning Techniques. Int. J. Emerg. Technol. Learn., 2019. 14: p. 92-104.

[25] Alsalman, Y.S., et al., Using Decision Tree and Artificial Neural Network to Predict Students Academic Performance. 2019 10th International Conference on Information and Communication Systems (ICICS), 2019: p. 104-109.

[26] Agaoglu, M., Predicting Instructor Performance Using Data Mining Techniques in Higher Education. IEEE Access, 2016. 4: p. 2379-2387.

[27] Roy, K., et al., Student Career Prediction Using Advanced Machine Learning Techniques. International journal of engineering and technology, 2018. 7: p. 26.

[28] Erickson, V.L. Data-Driven Models to Predict Student Performance and Improve Advising in Computer Science. 2019.

[29] Francis, B.K. and S.B. Suvanam, Predicting Academic Performance of Students Using a Hybrid Data Mining Approach. Journal of Medical Systems, 2019. 43: p. 1-15.

[30] Eagle, M., et al. Predictive Student Modeling for Interventions in Online Classes. in Educational Data Mining. 2018.

[31] Nguyen, V.A., Q. Nguyen, and V.T. Nguyen, A Model to Forecast Learning Outcomes for Students in Blended Learning Courses Based On Learning Analytics. Proceedings of the 2nd International Conference on E-Society, E-Education and E-Technology, 2018.

[32] Vasić, D., et al., Predicting student's learning outcome from Learning management system logs. 2015 23rd International Conference on Software, Telecommunications and Computer Networks (SoftCOM), 2015: p. 210-214.

[33] Kaur, P., M. Singh, and G.S. Josan, Classification and Prediction Based Data Mining Algorithms to Predict Slow Learners in Education Sector. Procedia Computer Science, 2015. 57: p. 500-508.

[34] Hasan, R., et al., Student Academic Performance Prediction by using Decision Tree Algorithm. 2018 4th International Conference on Computer and Information Sciences (ICCOINS), 2018: p. 1-5.

[35] Vijayalakshmi, V. and K. Venkatachalapathy, Comparison of Predicting Student's Performance using Machine Learning Algorithms. International Journal of Intelligent Systems and Applications, 2019. 11: p. 34-45.

[36] Turabieh, H., Hybrid Machine Learning Classifiers to Predict Student Performance. 2019 2nd International Conference on new Trends in Computing Sciences (ICTCS), 2019: p. 1-6.

[37] 3Kurniadi, D., et al., The prediction of scholarship recipients in higher education using k-Nearest neighbor algorithm. IOP Conference Series: Materials Science and Engineering, 2018. 434.

[38] Agrawal, R., T. Imieliński, and A. Swami. Mining association rules between sets of items in large databases. in Proceedings of the 1993 ACM SIGMOD international conference on Management of data. 1993.

[39] Wang, X. and G. Jiao, Research on association rules of course grades based on parallel fp-growth algorithm. Journal of Computational Methods in Sciences and Engineering, 2020. 20(3): p. 759-769.

[40] Czibula, G., A. Mihai, and L.M. Crivei, S PRAR: A novel relational association rule mining classification model applied for academic performance prediction. Procedia Computer Science, 2019. 159: p. 20-29.

[41] Shatnawi, R., et al., Building a smart academic advising system using association rule mining. arXiv preprint arXiv:1407.1807, 2014.

[42] Lailiyah, S., E. Yulsilviana, and R. Andrea, Clustering analysis of learning style on anggana high school student. TELKOMNIKA (Telecommunication Computing Electronics and Control), 2019. 17(3): p. 1409-1416.

[43] Mondal, B., et al., A course recommendation system based on grades. 2020 International Conference on Computer Science, Engineering and Applications (ICCSEA), 2020: p. 1-5.

[44] Kohonen, T., The self-organizing map. Proceedings of the IEEE, 1990. 78(9): p. 1464-1480.

[45] Delgado, S., et al., Analysis of Students' Behavior Through User Clustering in Online Learning Settings, Based on Self Organizing Maps Neural Networks. IEEE Access, 2021. 9: p. 132592-132608.

[46] Jiménez, R., et al., Dimensionality reduction in data mining using artificial neural networks. Methodology, 2009. 5(1): p. 26-34.

[47] Agarwal, R. and R. Srikant. Fast algorithms for mining association rules. in Proc. of the 20th VLDB Conference. 1994.

[48] Mehta, M., R. Agrawal, and J. Rissanen. SLIQ: A fast scalable classifier for data mining. in Advances in Database Technology—EDBT'96: 5th International Conference on Extending Database Technology Avignon, France, March 25–29, 1996 Proceedings 5. 1996. Springer.

[49] Mayilvaganan, M. and D. Kalpanadevi. Comparison of classification techniques for predicting the cognitive skill of students in education environment. in 2014 IEEE International Conference on Computational Intelligence and Computing Research. 2014. IEEE.