Article

Deep Neural Network-Based Prediction and Early Warning of Student Grades and Recommendations for Similar Learning Approaches

Tao Tao [1,2], Chen Sun [1,2,3], Zhaoyang Wu [3], Jian Yang [3] and Jing Wang [3,*]

[1] School of Computer Science and Technology, Anhui University of Technology, Ma'anshan 243032, China; taotao@ahut.edu.cn (T.T.); sunc1006@ahut.edu.cn (C.S.)
[2] Institute of Artificial Intelligence, Hefei Comprehensive National Science Center, Hefei 230088, China
[3] School of Metallurgical Engineering, Anhui University of Technology, Ma'anshan 243032, China; wustwuzhaoyang@163.com (Z.W.); yjxyb@ahut.edu.cn (J.Y.)
* Correspondence: rosanna.happy@163.com; Tel.: +86-182-5695-5954

Abstract: Studies reported that if teachers can accurately predict students' follow-up learning effects via data mining and other means, as per their current performances, and explore the difficulty level of students' mastery of future-related courses in advance, it will help improve students' scores in future exams. Although educational data mining and learning analytics have experienced an increase in exploration and use, they are still difficult to precisely define. The usage of deep learning methods to predict academic performances and recommend optimal learning methods has not received considerable attention from researchers. This study aims to predict unknown course grades based on students' previous learning situations and use clustering algorithms to identify similar learning situations, thereby improving students' academic performance. In this study, the methods of linear regression, random forest, back-propagation neural network, and deep neural network are compared; the prediction and early warning of students' academic performances based on deep neural network are proposed, in addition to the improved K-nearest neighbor clustering based on association rules (Pearson correlation coefficient). The algorithm performs a similar category clustering for early-warning students. Using the mean square error, standard deviation, mean absolute percentage error, and prediction of ups-and-downs accuracy as evaluation indicators, the proposed method achieves a steady improvement of 20% in the prediction of ups-and-downs accuracy, and demonstrates improved prediction results when compared under similar conditions.

Keywords: deep neural network; educational data mining; association rules; clustering algorithm; method recommendation

1. Introduction

Because of the rapid developments in Big Data and internet technology [1], the reformation of school's comprehensive evaluation system has been improved continuously, and students' grades are the most comprehensive reflection of course learning, which directly affects their future development. Teachers, as guides in the process of students' growth and success, should keep pace with times in education. As per the present academic management model, they should consider how to transform its lag into progress, thus helping students make academic plans in advance and comprehensively analysing the academic performance of students, which is an important part of academic work.

Data mining [2] helps researchers discover and understand implicit patterns in the collection using its existing data. Because of the continuous development and evolution of Big Data technology, it has gradually evolved into a data mining technology based on the era of Big Data [3].

Educational data mining [4–6] refers to data mining techniques used to analyse educational data. With the prolonged COVID-19 outbreak [7], as more efficient data are

transferred from offline to online, educational data mining faces bigger challenges, thereby providing researchers with additional raw materials. Feldman-Maggor et al. [8] examined the learning process of undergraduate online general chemistry courses using two logistical regression models and a decision tree algorithm. The submission status of optional assignments and the student's cumulative video open mode have been reported to have strong correlations with the students' grades.

Machine learning (ML) [9–11] is a cutting-edge tool that simulates human behaviour via computer algorithms. Based on the original data, certain potential objective laws are summarised via continuous machine calculations.

Given the rapid development of ML, the aspect of deep learning [12] is particularly prominent, and more deep-level algorithms are being applied in all walks of life. Deep neural networks [13–18] are used for predicting linear regression (LR) problems. The primary advantages are as follows: (1) power consumption is independent of batch size and architecture; (2) accuracy and inference time have a hyperbolic relationship; (3) the energy constraint is the upper limit of the maximum achievable accuracy and model complexity; and (4) the number of operations is a reliable estimate of the inference time.

Although educational data mining and learning analysis have experienced an increase in exploration and use, they are still difficult to accurately define. However, deep learning methods predict college course grades and recommend optimal learning methods for mid- and down-stream students based on focus.

2. Research Actuality

In the 1960s, ML algorithms such as perceptron [19], decision tree [20], and logical regression [21] were first used in teaching management research in major foreign universities, such as predicting the efficiency of online classes, predicting the situation of students staying in school, and assessing the advantages and disadvantages of school teaching.

By searching for ML performance prediction on Google Scholar, 1,670,000 search results were identified, of which 59,700 were reported since 2021. It is obvious that this research area still has global interest. Balqis et al. [22] proposed a custom rule-based model to identify risk students and put forwards reasonable remedial action methods in the research of student achievement analysis and prediction, through which they could recognise the most important and influential attributes to take appropriate corrective measures and help the high-risk students as early as possible. Khakata Esther et al. [23] used the decision tree to predict students' performance in the internet media environment to judge whether they may perform well in using the internet technology in their learning. Alam talha Mahboob et al. [24] proposed a new method to measure the performance of public education institutions using an ML model. They constructed an effective model through five ML algorithms, such as the J48 decision tree, and compared their results. The artificial neural network outperforms other models in some feature selection methods. Hussain et al. [25] used ML (training regression models and DT-classifiers) to predict students' academic performance at secondary and intermediate levels, and then they finally analysed the results obtained by the model. The obtained results demonstrated that ML techniques are effective and relevant for predicting student performance. Berens et al. [26] built an early detection system that uses regression analysis, neural networks, decision trees, and AdaBoost algorithms to identify student characteristics that distinguish potential dropouts and graduates, and its highest accuracy can be improved to 90%. Baashar et al. [27] used artificial neural networks (ANNs) to predict students' academic performance, thus confirming that ANNs are always used along with data analysis and data mining methods, thereby allowing researchers to evaluate the validity of their results in assessing academic achievement.

Domestically, by searching Google Scholar for predicting Chinese ML results, >6000 articles written in Chinese can be obtained in China, of which only 500 articles have been published after 2021. Because certain outstanding domestic scholars publish articles in English (published here indicates that there are fewer related articles written in Chi-

nese), Yuling et al. [28] proposed a multi-task MIML learning method for pre-class student achievement prediction, which can predict students who are at risk of failing to a certain extent before the course starts, thus improving the predictability of traditional methods. Hongchen et al. [29] proposed a method for predicting student achievement based on factor analysis and back-propagation (BP) neural network. The experiments on the Q&A event log demonstrate that the current method has certain effects. Junnan et al. [30] predicted students' online grades using data mining and predicted learners' course grades based on their online learning status. Jia et al. [31] analysed and interpreted large educational data using five typical ML algorithms: logistic regression, decision tree, naive Bayes algorithm, back-propagation, and random forest, and finally confirmed that the logistic regression algorithm is the most accurate student achievement prediction algorithm. Guang et al. [32] used the Bayesian network based on evidence-based reasoning to effectively solve low credibility and weak interpretability problems. Compared with the commonly used ML and deep learning models, it demonstrated lower error and stronger interpretability.

Countries attach considerable importance to monitoring the academic performance of students. Students are the future and hope of a country, and their academic status is something we should all focus on. The analysis and research of ML data in education is extensively involved at home and abroad, thus aiding the education of students, teachers, and institutions. This is used to analyse and process the historical data of students' academic performance, classify and manage students, and propose targeted plans suitable for students' academic study to assist teachers in formulating reasonable teaching plans more effectively.

Although several experts have conducted relevant research on students achieving early-warning and similar learning method recommendation, few substantive applications have been performed. In this regard, we can further improve the prediction effect by anticipating the likely pressure and challenges in the courses ahead of time and sending different course analyses to students so that each student receives a clear message about each course based on the powerful fitting and generalisation ability of the deep neural network (DNN). For students who may fail the course exam, we could store their data in the early-warning information table. By integrating the clustering algorithm of association rules, we match a senior student, who has a similar learning situation but has passed or been promoted to the subsequent courses, and then find the student and seek his information such as his daily learning, homework, review, and exam preparation method in the corresponding course. Later, we can provide accurate feedback to the students taking the course, achieving the person-to-person improvement effect. The specific model design idea is shown in Figure 1.
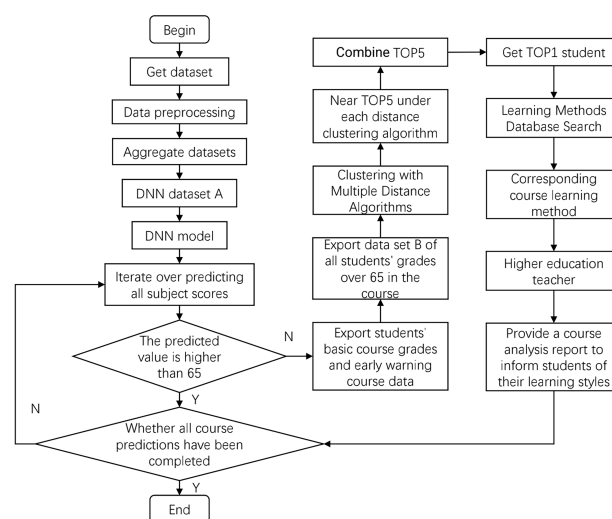

Figure 1. DNN-based student achievement warning and similar learning method recommendation model.

3. Dataset

3.1. Data Sources

The dataset used in this study comes from students' performance data in the College of Metallurgical Engineering at a university in the Anhui Province. After cleaning, the major achievement data of 1683 students from 2017 to 2020 were selected, including elective and compulsory scores. In this study, the compulsory score data were solely used because of the lack of consistency in students' elective courses. Considering the different courses of different majors and the principle of diversity of sample selection, we chose the Metallurgical Engineering major, which has the largest number of students, and obtained 21,350 compulsory score data from 675 students as the total dataset.

For our daily teaching work, we have a certain tendency towards the relevance of the same type of course. Because of the universality of laboratories, we chose professional courses and basic public courses (language courses and logic courses) as the representative of general courses and directly used the specific courses (Metallurgical Physical Chemistry 1, College English and Linear Algebra), using the Pearson correlation coefficient method to obtain the correlation table to confirm the universality of the current data.

Since the system uses two ML models, we must carry out some processing on the total dataset and obtain two types of datasets, namely the conventional dataset A (including all students' grades). Dataset B, where grades are higher than the early-warning line (all grades higher than 65 in the corresponding courses are selected), can be generated according to the grades of each subject requiring an early warning.

3.2. Data Pre-Processing

To provide a dataset suitable for the DNN and clustering algorithm based on association analysis, as much as possible, the pre-processing steps should be standardised. The detailed pre-processing process in this study is as follows:

(1) **Data integration.** The data obtained from the school's educational administration system is a single piece of score data containing several attributes of the course, which is first stored in the MySQL server of the student integrated management system independently developed by the college, and the current 21,350 scores are integrated and exported to the table using database statements.

(2) **Data cleaning.** Data cleaning refers to screening and reviewing the original data, as well as addressing the inconsistency of data, missing data, wrong data, and duplicate data. It should be considered that as the training scheme evolves, the curriculum system in different grades will change, and some courses will be replaced or deleted from the perspective of professional personnel.

(3) **Data conversion and processing.** Using the Python Pandas framework, the String type is converted to the double type, and the excellent (A), good (B), moderate (C), pass (D), and fail (E) types are converted to numerical values of 95, 85, 75, 65, and 55, respectively. Then, missed exams are converted to 0 points, and the missing values are converted into the average of the student's average score and the average score of the course using the horizontal and vertical categories for the student's grade data below 10 points. The mask and number start with student number 0.

(4) **Training data division.** After the cleaned data are randomly disrupted, the original data are randomly divided into a training set and a verification set in a 3:1 ratio. The training set is used to train the DNN model, and the verification set is used to test the generalisation ability of the model after each iteration to prevent overfitting.

4. Grade Prediction Method Based on DNNs

4.1. Introduction of DNN Algorithm

A neural network is the repeated superposition of the perceptron [33] model. The deep neural network (DNN) [13–18] can be understood as a large network structure composed of neurons with nonlinear units. Compared with the NN, the DNN can help obtain the

extremely complex multivariate multi-function model, and solve more complex problems, as shown in Figure 2.

It is a multi-layer neural network that examines the output characteristics of the upper layer as the input characteristics of the next layer. After multi-layer nonlinear transformation training of the original characteristics, the characteristics of the initial sample are gradually converted to another feature space as per requirements to examine the characteristics of the existing input with better effects. Because of the rich dimension of the massive data, the DNN is capable of mining the potential information of the massive data more comprehensively and accurately by establishing multi-level mathematical models and training massive data to improve the information value.
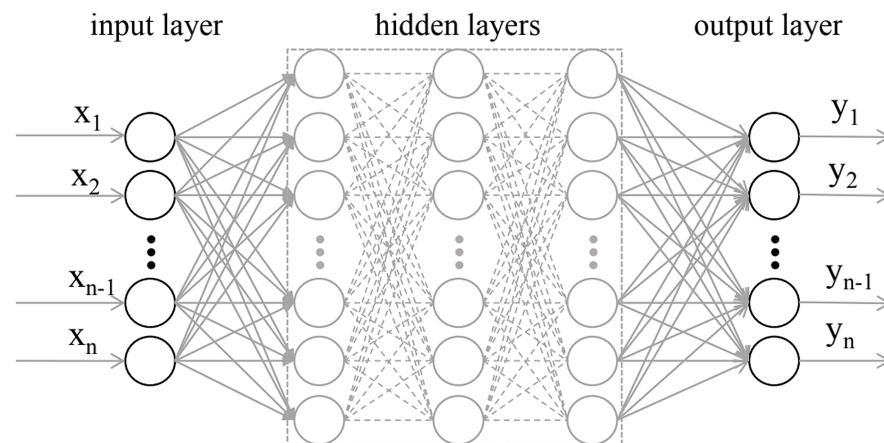

Figure 2. Deep neural network structure diagram.

The DNN is divided according to the positions of the different network layers. The first layer is the input layer, the last layer is the output layer, and the middle layers are concealed layers. As shown in Figure 2, any neuron in each connected layer is fully connected. The weighted coefficient matrices W and bias B are used to perform a series of linear and activation operations of feed-forward propagation on the input vector. Finally, the model output results are obtained. Nevertheless, the training of the DNN model is completed via back-propagation. In the process, the loss function is used to measure the loss between the model output value and the real value (the model loss function in this study uses the mean square error loss (MSEL) function, as shown in Formula (1)). The weight coefficient matrices W and bias B are constantly updated until the training is completed.

$$J(W, b, x, y) = \frac{1}{2} \left\| y^L - y \right\|_2^2 = \frac{1}{2} \left\| \sigma\left(W^L y^{L-1} + b^L\right) - y \right\|_2^2 \tag{1}$$

The pseudocode of the DNN algorithm is shown in Algorithm 1.

4.2. Grade Prediction Based on the DNN

The DNN is built using the Python language and the Pytorch framework [34], with 28-course scores as the featured item in the input and output layers predicting the scores of three courses. The network is also constructed with six hidden layers of 1375 neurons in total. Then, the number of neurons in the eight layers is 28, 256, 128, 512, 256, 128, 64, and 3, respectively. As shown in Algorithm 1, a number of parameters, such as activation, optimiser, loss, batch size, weight decay, epochs, and the early termination of iteration threshold, are set. The previously processed dataset A is used as a sample for training. Each layer's W and B values are solved using iterative training with the BP until the current training iterations reach the specified number or the loss reaches the set threshold. Then, the trained model is solidified to the specified position to complete the training. Many experimental tests are performed following relevant steps. Finally, the MSEL between test and training data is used as an indicator to obtain and set multiple attribute parameters

consistent with the project, thus obtaining the final training and results. Figure 3 shows the decline process of the DNN training the MSEL.

Algorithm 1 The Pseudocode of DNN Algorithm

Input: 'data set': the training set, validation set, test set;
  1: 'epochs': the maximum number of iterations;
  2: 'batch size': the minimum batch size of the data loader;
  3: 'optimizer': the optimisation algorithm (optimiser in torch.optim);
  4: 'optim hparas': the learning rate, momentum, regularisation factor;
  5: 'early stop': the number of epochs since the model was last improved;
  6: 'save path': the model will be saved here;
Output: The trained deep neural network model;
  7: initialize the parameters according to the input;
  8: data normalisation;
  9: create a network;
10: train the network:
11: **repeat**
12:    repeat for training process:
13:    forward-propagation;
14:    back-propagation;
15: **until** for reaches the end condition
16: Using the network;
17: The data are reversely normalized;

When the DNN training is completed, the model parameter is saved to the memory. Subsequently, the offline model is then read, and the trained network model is used to predict the grades. Finally, the predicted value is compared with that of real data, and dark-green and orange-dotted lines are used to mark the initial and predicted data values, respectively.
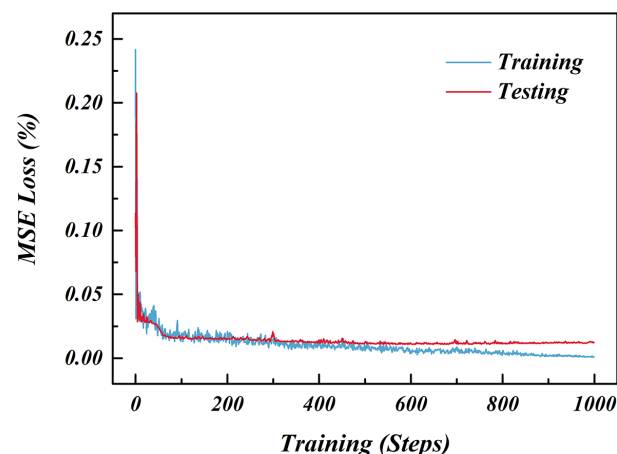

Figure 3. DNN training.

Because multiple courses require prediction, we selected three representative subjects to demonstrate in the experiment project, namely Physical Chemistry of Metallurgy 1, University English 4, and Linear Algebra B. As shown in Figure 4, a broken line comparison diagram of predicted and real grades in the corresponding courses is generated using the abovementioned approach, with 65 and 60 as the early warning and passing grades, respectively.
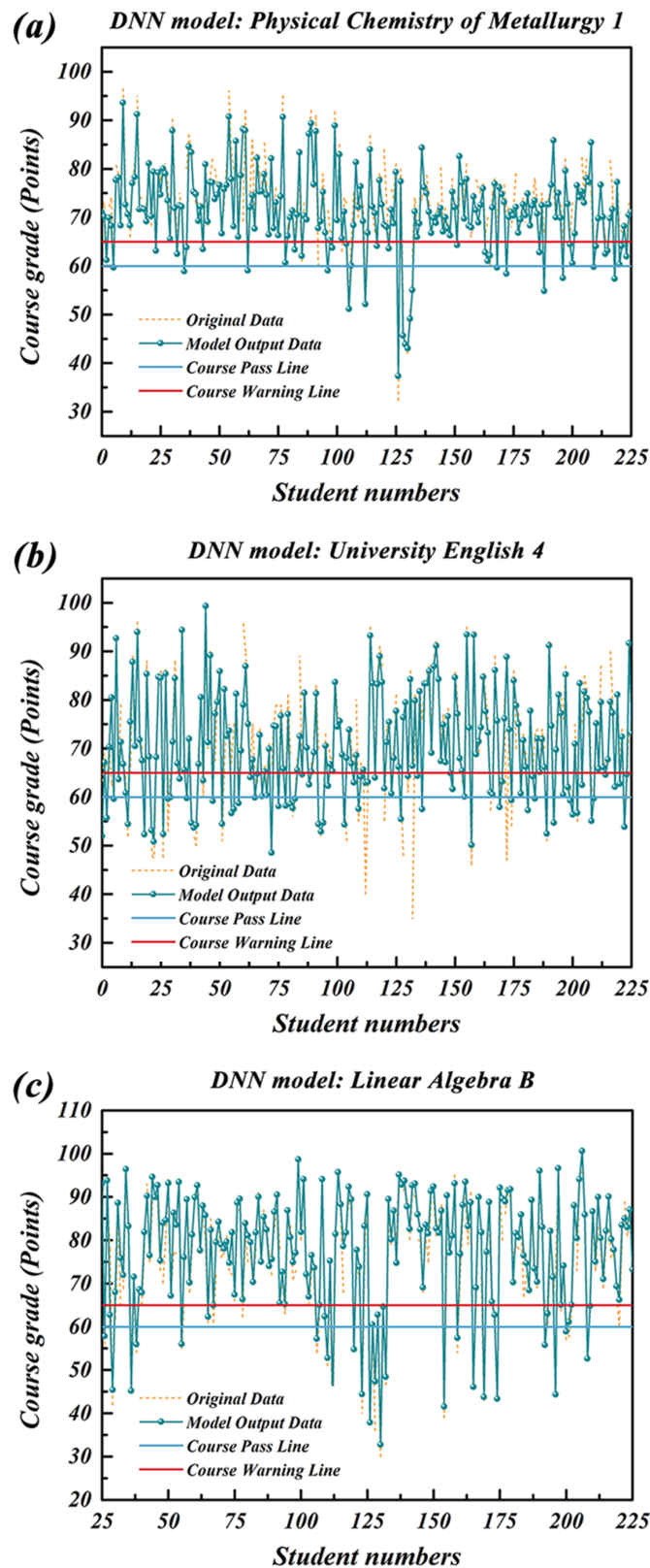
Figure 4. DNN training.

The overlapped brownish-red line shows that the predicted data agree with the real data when the two lines are overlayed with a degree of transparency. However, the non-overlapping parts indicate certain errors in the model. Students' grades in different semesters may fluctuate with attitude changes. However, they are not distinct in most

data, indirectly demonstrating that the volatility is in the acceptable range. Accordingly, the current model and parameters can be used for stable test set fitting.

An automated batch programme is set up as per the process to predict the grades of all courses in the current semester. The prediction results are shown in Table 1.

Table 1. DNN predictions.

Number	Features					Model Predictions		
	1	2	...	27	28	1	2	3
1	56	82	...	80	61	28.00	29.00	30.00
2	80	73	...	81	60	66.93	73.42	52.83
3	67	84	...	74	64	77.96	72.83	69.05
4	86	82	...	70	72	64.09	63.56	64.24
5	89	96	...	77	64	86.95	72.19	71.22
...	...	...	...	...	...	...	...	...
189	48	69	...	71	50	66.59	65.13	65.81
190	62	89	...	92	70	85.84	75.39	74.43
191	62	67	...	61	61	74.51	52.94	62.44
192	77	76	...	78	40	68.56	71.96	50.28
193	90	91	...	85	71	93.22	74.32	92.03

We automatically acquired the predicted grades of <65 and requested the student's e-mail IDs from the database. E-mails were generated in fixed forms using the predicted grades via an automatic batch mail system and then sent to students from the college teacher's official email to remind them to master the important learning points.

Students whose predicted grades were less than or equal to 65 are selected, and all students' course information were obtained. In the total dataset, the pre-warning course of the student was higher than 65 points for screening, and the selected data were used as the dataset B above the pre-warning line. The dataset of the association analysis and clustering algorithm was used to perform the next test.

4.3. Comparative Experiment Based on LR

LR [35] is a popular statistical method used to describe a continuous output variable related to multiple independent variables X_i, $i = 1, 2, \ldots, n$. Random forest (RF) [36] is an ensemble of tree-structured classifiers that have effectively been used for classification and regression problems. The BP neural network (BPNN) [37] has many applications. It is a type of network with one-way propagation and multi-layer forward characteristics. It belongs to the feed-forward neural network and adopts the error BP algorithm and has been effectively used for classification and regression problems. By reading the relevant ML materials, the above three ML algorithms are reproduced, various model parameters are set according to the instructions, and the processed data set A is used for model training.

The mean absolute error (MAE), mean square error (MSE), mean absolute percentage error (MAPE), and prediction of ups-and-downs accuracy (PUDA) are used as indicators of the merits of the model, and their mathematical representations are shown in Table 2.

y_i is the ith true value and $\hat{y}_i$ is the ith-predicted value. The performance of various algorithms in the current dataset is shown in Table 3. The larger the corresponding values of MAE, MSE, and MAPE in the table, the smaller the model effect and fitting performance of the model. The larger the value of the PUDA index, the better the fitting performance of the model. For the training (train) and test sets (test), if the difference between the train and test indicators is greater, the generalisation performance of the model will be worse and the model becomes over-fitted, and vice versa. The current model has a good generalisation performance.

Table 2. Mathematical meaning of each model indicator.

Number	Indicator Name	Mathematical Meaning
1	MAE	$\frac{1}{m}\sum_{i=1}^{m}\lvert(y_i - \hat{y}_i)\rvert$
2	MSE	$\frac{1}{m}\sum_{i=1}^{m}(y_i - \hat{y}_i)^2$
3	MAPE	$\frac{1}{m}\sum_{i=1}^{m}\lvert(y_i - \hat{y}_i)/y_i\rvert \times 100\%$
4	PUDA	$\frac{1}{m-1}\sum_{i=1}^{m-1}\{(\hat{y}_{i+1} - \hat{y}_i) \times (y_{i+1} - y_i) \geq 0?1:0\} \times 100\%$

Table 3. The results of each model indicator.

Number	Indicator Name	LR	RF	BPNN	DNN
1	Train MAE	0.1364	0.0601	0.0738	0.0743
2	Train MSE	0.0289	0.0059	0.0099	0.0092
3	Train MAPE	21.55	31.70	353.1	35.68
4	Train PUDA	58.52	93.56	76.24	85.57
5	Test MAE	0.1499	0.1217	0.2036	0.0865
6	Test MSE	0.0319	0.0229	0.0966	0.0115
7	Test MAPE	22.60	19.45	33.31	34.59
8	Test PUDA	52.38	50.00	59.09	80.95

By comparing the specific performance of multiple algorithms in the training and test sets divided by the current data set through the four train parameters, it becomes obvious that although RF better fits the training set, its performance in the test set is less effective. Furthermore, by comparing the two indicators of train MAE and MSE, the BPNN and the DNN yielded similar results; however, by using test MAE and MAE, it is clear that the BPNN is overfitted. Furthermore, in this study, the DNN used has a better generalisation performance than the above-mentioned three algorithms.

5. Most Similar Sample Recommendation Model Based on the Pearson Correlation Coefficient and Distance Proximity Algorithm

5.1. Pearson Correlation Coefficient, Definitions of Distance and Thought on the KNN Algorithm

5.1.1. Pearson Correlation Coefficients

The Pearson correlation coefficient [38] is used to eliminate the covariance of dimensional influence, thus reflecting whether there is a linear relationship between two variables with a value range of $[-1, 1]$. The current two variables have a strong linear relationship when the coefficient value is in the range of $(0.8, 1)$. Relativity weakens as the absolute value of the coefficient decreases, and the negative coefficient indicates that the two variables present certain negative relativity, expressed in Formula (2).

$$\rho_{XY} = \frac{\text{Cov}(X, Y)}{\sqrt{D(X)}\sqrt{D(Y)}} = \frac{E((X - EX)(Y - EY))}{\sqrt{D(X)}\sqrt{D(Y)}} \tag{2}$$

5.1.2. Euclidean Distance

Euclidean distance [39] is the most common and intuitive distance measurement method, usually used to describe the distance between two points in a 2D space. The Euclidean distance between the two points $a(x_{11}, x_{12}, \cdots x_{1n})$ and $b(x_{21}, x_{22}, \cdots x_{2n})$ in an n-dimensional space is expressed, as shown in Formula (3).

$$d_{ab} = \sqrt{\sum_{k=1}^{n}(x_{1k} - x_{2k})^2} \tag{3}$$

The Pearson correlation coefficient is used to propose an improved Euclidean distance measurement method. Assume that the n + 1 dimension is the dimension that must be calculated, it is weighted as per the correlation coefficient. Its expression is shown in Formula (4).

$$d_{ab} = \sqrt{\sum_{k=1}^{n} \rho_{X_k X_{n+1}} (x_{1k} - x_{2k})^2} \tag{4}$$

5.1.3. Manhattan Distance

The Manhattan distance [40] is used to calculate the addition of the absolute distances between two points in space in the coordinate system direction, as proposed by Hermann Minkowski. The Manhattan distance between two points $a(x_{11}, x_{12}, \cdots x_{1n})$ and $b(x_{21}, x_{22}, \cdots x_{2n})$ in an n-dimensional space is expressed as Formula (5).

$$d_{ab} = \sum_{k=1}^{n} \sqrt{(x_{1k} - x_{2k})^2} \tag{5}$$

An improved Manhattan distance measurement method based on the Pearson correlation coefficient is proposed. Assuming that the n + 1 dimension is the dimension that must be calculated, it is weighted as per the correlation coefficient. Its expression is shown in Formula (6).

$$d_{ab} = \sum_{k=1}^{n} \rho_{X_k X_{n+1}} \sqrt{(x_{1k} - x_{2k})^2} \tag{6}$$

5.1.4. Chebyshev Distance

The Chebyshev distance [41] is a measure in vector space derived from the upper limit norm, indicating that the distance between these two points is the maximum distance in each dimension. The Chebyshev distance between these two points and in the n-dimensional space is expressed, as shown in Formula (7).

$$d_{ab} = \max\left(\sqrt{(x_{1k} - x_{2k})^2} \right) \tag{7}$$

The Pearson correlation coefficient is used to propose an improved Chebyshev distance measurement method. The n + 1 dimension is assumedly the dimension that must be calculated, and it is weighted as per the correlation coefficient. Its expression is shown in Formula (8).

$$d_{ab} = \sum_{k=1}^{n} \rho_{X_k X_{n+1}} \sqrt{(x_{1k} - x_{2k})^2} \tag{8}$$

5.1.5. KNN Algorithm

The k-nearest neighbour (KNN) [42] algorithm is one of the popular algorithms in ML to perform supervised learning classification and regression. If K samples closest to the centre of a sample are reported by a certain distance (Euclidean distance) measurement, the K samples belong to this category.

5.2. Most Similar Sample Recommendation Model Based on the Distance Proximity Algorithm Improved with the Pearson Correlation Coefficient

As shown in Table 4, a Pearson correlation table is established using the basic dataset and grade data to be predicted. College English 2 demonstrated a strong correlation with College English 4. Although English courses correlate with Physics and Mathematics courses, they are less relevant than similar English courses. Metallurgical Physical Chemistry 1 has a close relationship with Physics and Chemistry and correlates with Mathematics courses whose correlation coefficient is higher than that of English courses. Another example is the Employment Guidance course for college students, which does not show a great correlation with the Physical Chemistry, Mathematics, and English courses. Their scores only show a weak correlation.

Table 4. Intercourse relevance.

Number	Course Name	Metallurgical Physical Chemistry 1	College English 4	Linear Algebra B
1	Advanced Mathematics A1	0.5121	0.3701	0.4000
2	Basic Computer Science	0.1870	0.3915	0.1566
3	Physical Chemistry D1	0.5599	0.3543	0.4446
4	University Physics B2	0.5956	0.3521	0.5446
5	Career Development and Employment Guidance for College Students 1	0.2625	0.1713	0.1713
6	Outline of Modern Chinese History	0.1841	0.0846	0.1286
…	…	…	…	…
23	Freshman Seminar[Y]	0.2135	0.2049	0.0631
24	Advanced Mathematics A2	0.5256	0.4032	0.4926
25	Introduction to Business Management	0.3046	0.4612	0.1180
26	Engineering Graphics B	0.3519	0.3216	0.3370
27	Probability Theory and Mathematical Statistics C	0.6296	0.2384	0.5973
28	College English 2	0.3426	0.6213	0.2227

The correlation coefficient of the course 'Physical Chemistry of Metallurgy' can be taken as an example. Based on the correlation coefficient, when taking the Euclidean distance, Manhattan distance, Chebyshev distance, and their improved algorithms as the distance criteria, and the test sample as the centre for clustering with the KNN algorithm, the top five similar student data can be obtained. Considering the minimum value of each group and comparing the correlation coefficients in Table 5, the specific situation can be observed.

Table 5. Sorted-neighbourhood top 5 based on various distance algorithms.

Distance Type	Proximity Sort Top 5				
	1	2	3	4	5
Euclidean distance	31	207	11	163	127
Manhattan distance	31	207	11	127	125
Chebyshev distance	31	207	11	112	192
Euclidean distance based on the Pearson coefficient	217	2	50	207	11
Manhattan distance based on the Pearson coefficient	217	11	207	2	125
Chebyshev distance based on the Pearson coefficient	2	31	217	50	164

Based on multiple common distance standards, the data most similar to the learning situation of the target student is Student No. 31. Based on multiple distance standards of the Pearson correlation coefficient, the data most similar to the learning situation of the target student are Student No. 217. The grade data of Student Nos. 31 and 217 are obtained from the original dataset. As per expert experience, the connection between each course and target course is obtained. The courses that are highly and less correlated with the target course are divided into groups. Then, the grade data between the target students and most similar students reported by the two algorithms are compared; moreover, the comparison line graph is drawn and shown in Figures 5 and 6.

As shown in Figure 5, the trend of the blue line is closer to that of the red line and more dispersed than that of the orange line. The difference between Student Nos. 0 and 217 in corresponding courses is smaller than that between Student Nos. 0 and 31. As shown in Figure 6, the shape of the blue and red lines is similar, indicating that the difference between Student Nos. 31 and 0 is small; they are the closest in the case of clustering when the correlation is not considered. The red and blue lines considerably differ in direction, indicating that Student No. 217 in the normal range benchmark clustering has not entered the top five sorted neighbourhood. Nevertheless, Student Nos. 217 and 0 are the closest in consideration of correlation. With the enlargement of the sample data, the student most similar to the target student can be determined using a correlation coefficient from whom the corresponding course learning methods can be obtained. The methods will be

recommended to the students that risk failing the course. Thus, recording and rewarding students who offer help will strengthen academic supervision in advance.
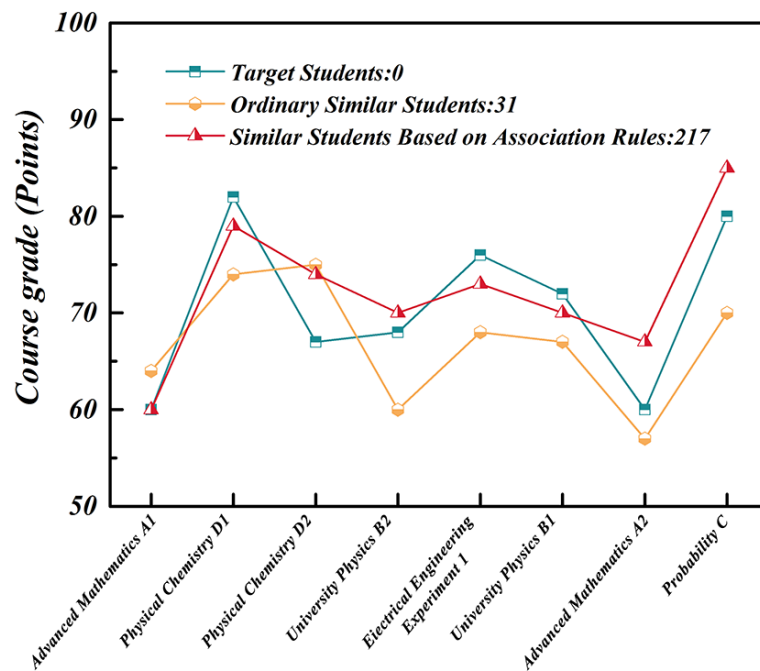

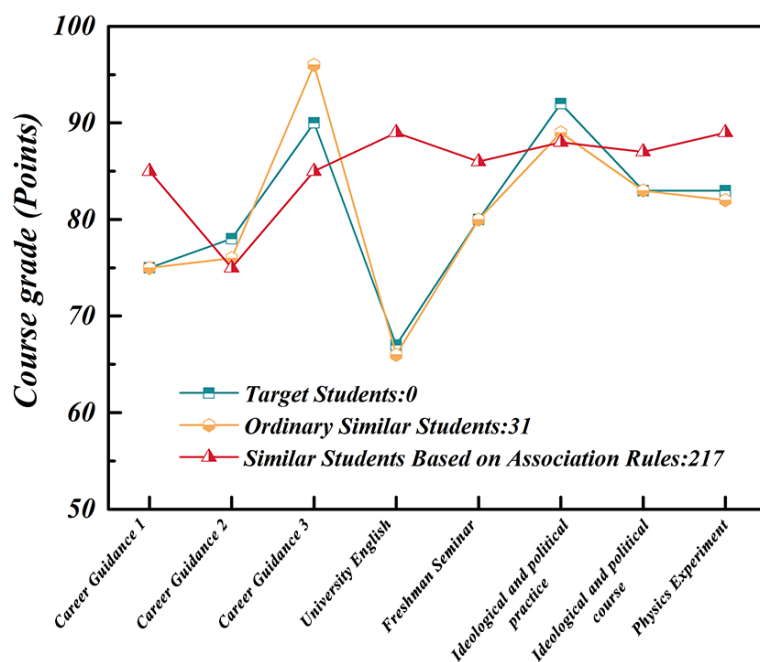**Figure 5.** Comparison chart of courses that are highly correlated with the target course.

Figure 6. Comparison chart of courses that are less correlated with the target course.

6. Conclusions

The grade prediction and learning recommendation model proposed in this study is combined with the DNN algorithm. Compared with LR, random forest, and BP neural network, the current model improves the prediction accuracy to some extent and can fit the data better. Using the improved clustering algorithm, to further migrate the day-to-day teaching of student grades ahead of the final exam. The early-grade warning work helps to improve students' learning situations and helps them plan their studies better. Based on the model examples in this study, the proposed model can be extensively used in daily life,

industrial production, and intelligent manufacturing. Replacing the existing complex tasks with relevant ML algorithms will ease labour for more creative work. With the development of deep learning, certain seemingly unrealistic ideas are becoming a reality.

Finally, the findings from this research are especially relevant in the context of the COVID-19 pandemic and will continue to be important in the post-COVID-19 world. Predicting student achievement has become extremely important with the rapid transition from offline learning to online learning in both schools and universities following the coronavirus outbreak in recent years. This dramatic change continues to attract educators, researchers, policymakers, and the media to focus their attention to a variety of learning theories. More factors, such as students' learning, daily performance, and family influence, can be utilised as basic features for deep learning prediction with increasingly convenient internet tools and continuous optimisations and improvements in Big Data platforms. An overall evaluation of students' learning will help locate students' performance information more accurately, thus making it more convenient for teachers to have a comprehensive understanding of students and complete student work in advance.

References

1. Gao, H. Big data development of tourism resources based on 5G network and internet of things system. *Microprocess. Microsyst.* **2021**, *80*, 103567. [CrossRef]
2. Witten, I.H.; Frank, E. Data mining: Practical machine learning tools and techniques with Java implementations. *ACM Sigmod Rec.* **2002**, *31*, 76–77. [CrossRef]
3. Kumar, S.; Mohbey, K.K. A review on big data based parallel and distributed approaches of pattern mining. *J. King Saud-Univ.-Comput. Inf. Sci.* **2019**, *34*, 1639–1662. [CrossRef]
4. Hicham, A.; Jeghal, A.; Sabri, A.; Tairi, H. A survey on educational data mining [2014–2019]. In Proceedings of the 2020 International Conference on Intelligent Systems and Computer Vision (ISCV), Fez, Morocco, 9–11 June 2020; pp. 1–6. [CrossRef]
5. Crivei, L.M.; Czibula, G.; Ciubotariu, G.; Dindelegan, M. Unsupervised learning based mining of academic data sets for students' performance analysis. In Proceedings of the 2020 IEEE 14th International Symposium on Applied Computational Intelligence and Informatics (SACI), Timisoara, Romania, 21–23 May 2020; pp. 000011–000016. [CrossRef]
6. Figueroa-Cañas, J.; Sancho-Vinuesa, T. Early prediction of dropout and final exam performance in an online statistics course. *IEEE Rev. Iberoam. Tecnol. Del Aprendiz.* **2020**, *15*, 86–94. [CrossRef]
7. Akour, I.; Alshurideh, M.; Al Kurdi, B.; Al Ali, A.; Salloum, S. Using machine learning algorithms to predict people's intention to use mobile learning platforms during the COVID-19 pandemic: Machine learning approach. *JMIR Med. Educ.* **2021**, *7*, e24032. [CrossRef] [PubMed]
8. Feldman-Maggor, Y.; Blonder, R.; Tuvi-Arad, I. Let them choose: Optional assignments and online learning patterns as predictors of success in online general chemistry courses. *Internet High. Educ.* **2022**, *55*, 100867. [CrossRef]
9. Janiesch, C.; Zschech, P.; Heinrich, K. Machine learning and deep learning. *Electron. Mark.* **2021**, *31*, 685–695. [CrossRef]
10. Kwekha-Rashid, A.S.; Abduljabbar, H.N.; Alhayani, B. Coronavirus disease (COVID-19) cases analysis using machine-learning applications. *Appl. Nanosci.* **2021**, 1–13. [CrossRef]
11. Bell, J. What Is Machine Learning? In *Machine Learning and the City: Applications in Architecture and Urban Design*; John Wiley & Sons: Hoboken, NJ, USA, 2022; pp. 207–216. [CrossRef]
12. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [CrossRef]

13. Canziani, A.; Paszke, A.; Culurciello, E. An analysis of deep neural network models for practical applications. *arXiv* **2016**, arXiv:1605.07678.

14. Kwon, H.; Kim, Y. BlindNet backdoor: Attack on deep neural network using blind watermark. *Multimed. Tools Appl.* **2022**, *81*, 6217–6234. [CrossRef]

15. Lieu, Q.X.; Nguyen, K.T.; Dang, K.D.; Lee, S.; Kang, J.; Lee, J. An adaptive surrogate model to structural reliability analysis using deep neural network. *Expert Syst. Appl.* **2022**, *189*, 116104. [CrossRef]

16. Xia, L. Research on Optical Performance Monitoring Based on Deep Learning. Master's Thesis, University of Electronic Science and Technology of China, Chengdu, China, 2020. Available online: https://cdmd.cnki.com.cn/Article/CDMD-10614-1020736441.htm (accessed on 10 June 2022).

17. Jie, H. Large-scale task processing method based on big data deep neural network and agent. *Comput. Technol. Autom.* **2021** . Available online: https://www.cnki.com.cn/Article/CJFDTotal-JSJH202104023.htm (accessed on 10 June 2022).

18. Xu, W. Design and Implementation of Long Text Classification Algorithm Based on Deep Neural Network. Master's Thesis, Nanjing University of Posts and Telecommunications, Nanjing, China, 2020. Available online: https://cdmd.cnki.com.cn/Article/CDMD-10293-1020427857.htm (accessed on 10 June 2022).

19. Gamarnik, D.; Kızıldağ, E.C.; Perkins, W.; Xu, C. Algorithms and Barriers in the Symmetric Binary Perceptron Model. *arXiv* **2022**, arXiv:2203.15667.

20. Tanha, J.; Van Someren, M.; Afsarmanesh, H. Semi-supervised self-training for decision tree classifiers. *Int. J. Mach. Learn. Cybern.* **2017**, *8*, 355–370. [CrossRef]

21. Wang, J.J.; Liang, Y.; Su, J.T.; Zhu, J.M. An Analysis of the Economic Impact of US Presidential Elections Based on Principal Component and Logical Regression. *Complexity* **2021**, *2021*, 5593967. [CrossRef]

22. Hussain, S.; Khan, M.Q. Student-performulator: Predicting students' academic performance at secondary and intermediate level using machine learning. *Ann. Data Sci.* **2021**, 1–19. [CrossRef]

23. Albreiki, B.; Habuza, T.; Shuqfa, Z.; Serhani, M.A.; Zaki, N.; Harous, S. Customized Rule-Based Model to Identify At-Risk Students and Propose Rational Remedial Actions. *Big Data Cogn. Comput.* **2021**, *5*, 71. [CrossRef]

24. Alam, T.M.; Mushtaq, M.; Shaukat, K.; Hameed, I.A.; Umer Sarwar, M.; Luo, S. A novel method for performance measurement of public educational institutions using machine learning models. *Appl. Sci.* **2021**, *11*, 9296. [CrossRef]

25. Khakata, E.; Omwenga, V.; Msanjila, S. Student performance prediction on internet mediated environments using decision trees. *Int. J. Comput. Appl.* **2019**, *975*, 8887. [CrossRef]

26. Berens, J.; Schneider, K.; Görtz, S.; Oster, S.; Burghoff, J. Early Detection of Students at Risk–Predicting Student Dropouts Using Administrative Student Data and Machine Learning Methods. Available at SSRN 3275433. 2018. Available online: https://ssrn.com/abstract=3275433 (accessed on 10 June 2022). [CrossRef]

27. Baashar, Y.; Alkawsi, G.; Mustafa, A.; Alkahtani, A.A.; Alsariera, Y.A.; Ali, A.Q.; Tiong, S.K. Toward predicting student's academic performance using artificial neural networks (ANNs). *Appl. Sci.* **2022**, *12*, 1289. [CrossRef]

28. Ma, Y.; Cui, C.; Yu, J.; Guo, J.; Yang, G.; Yin, Y. Multi-task MIML learning for pre-course student performance prediction. *Front. Comput. Sci.* **2020**, *14*, 1–10. [CrossRef]

29. Bao, Y.; Lu, F.; Wang, Y.; Zeng, Q.; Liu, C. Student performance prediction based on behavior process similarity. *Chin. J. Electron.* **2020**, *29*, 1110–1118. [CrossRef]

30. Wu, J. Prediction of Students' Online Grades Based on Data Mining Technology. Master's Thesis, Chang'an University, Xi'an, China, 2021. Available online: https://cdmd.cnki.com.cn/Article/CDMD-10710-1021890378.htm (accessed on 10 June 2022).

31. Yu, J.; Bai, S.; Wu, D. Research on student achievement prediction based on machine learning in online teaching. *Comput. Program. Ski. Maint.* **2021**. Available online: https://www.cnki.com.cn/Article/CJFDTotal-DNBC202108047.htm (accessed on 10 June 2022).

32. Feng, G.; Pan, T.; Wu, W. Analysis of Online Learning Behavior Based on Bayesian Network Model. *J. Guangdong Univ. Technol.* **2022**, *39*, 41–48. [CrossRef]

33. Luo, W. Research on Human Physiological Signal Classification Based on Genetic Algorithm and Multilayer Perceptron. Master's Thesis, Xiamen University, Xiamen, China, 2018. Available online: https://cdmd.cnki.com.cn/Article/CDMD-10384-1018194858.htm (accessed on 10 June 2022).

34. Talloen, J.; Dambre, J.V.; Esompele, A. PyTorch-Hebbian: Facilitating local learning in a deep learning framework. *arXiv* **2021**, arXiv:2102.00428.

35. James, G.; Witten, D.; Hastie, T.; Tibshirani, R. Linear regression. In *An Introduction to Statistical Learning*; Springer: New York, NY, USA, 2021; pp. 59–128. [CrossRef]

36. Nachouki, M.; Abou Naaj, M. Predicting Student Performance to Improve Academic Advising Using the Random Forest Algorithm. *Int. J. Distance Educ. Technol. (IJDET)* **2022**, *20*, 1–17. [CrossRef]

37. Liu, D.; Li, S.; You, K. Training Load Prediction in Physical Education Teaching Based on BP Neural Network Model. *Mob. Inf. Syst.* **2022**, *2022*, 4821208. [CrossRef]

38. Deng, J.; Deng, Y.; Cheong, K.H. Combining conflicting evidence based on Pearson correlation coefficient and weighted graph. *Int. J. Intell. Syst.* **2021**, *36*, 7443–7460. [CrossRef]

39. Maxim, L.G.; Rodriguez, J.I.; Wang, B. Euclidean distance degree of the multiview variety. *SIAM J. Appl. Algebra Geom.* **2020**, *4*, 28–48. [CrossRef]

40. Wang, Z.; Xu, K.; Hou, Y. Classification of iris by KNN algorithm based on different distance formulas. *Wirel. Internet Technol.* **2021**, *18*, 105–106. Available online: https://www.cnki.com.cn/Article/CJFDTotal-WXHK202113051.htm (accessed on 10 June 2022).
41. Sun, Y.; Li, S.; Wang, X. Bearing fault diagnosis based on EMD and improved Chebyshev distance in SDP image. *Measurement* **2021**, *176*, 109100. [CrossRef]
42. Zheng, T.; Yu, Y.; Lei, H.; Li, F.; Zhang, S.; Zhu, J.; Wu, J. Compositionally Graded KNN-Based Multilayer Composite with Excellent Piezoelectric Temperature Stability. *Adv. Mater.* **2022**, *34*, 2109175. [CrossRef] [PubMed]