Advances in Computer and Communication, 2025, 6(1), 27-34
https://www.hillpublisher.com/journals/acc/
ISSN Online: 2767-2875

Prediction of Student Performance Based on MPL

Peng Zhao[1,2,*], Junye Yang[1,2], Rongmu Cai[1,2], Jia Gao[1,2]

[1]School of Computer Science and Technology, Taiyuan Normal University, Jinzhong 030619, Shanxi, China.
[2]Shanxi Key Laboratory of Intelligent Optimization Computing and Blockchain Technology, Taiyuan Normal University, Jinzhong 030619, Shanxi, China.

How to cite this paper: Peng Zhao, Junye Yang, Rongmu Cai, Jia Gao. (2025) Prediction of Student Performance Based on MPL. *Advances in Computer and Communication*, 6(1), 27-34.
DOI: 10.26855/acc.2025.01.005

Received: December 29, 2024
Accepted: January 27, 2025
Published: February 25, 2025

***Corresponding author:** Peng Zhao, School of Computer Science and Technology, Taiyuan Normal University, Jinzhong 030619, Shanxi, China; Shanxi Key Laboratory of Intelligent Optimization Computing and Blockchain Technology, Taiyuan Normal University, Jinzhong 030619, Shanxi, China.

Abstract

As the rapid development of digitalization has led to profound changes in the field of education and the integration of artificial intelligence technology, early education is facing problems such as a lack of funds and uneven teacher resources. Although the country has promoted artificial intelligence in education to improve hardware, teachers, and students lack application capabilities and lack guidance programs. This study predicts student scores based on a multi-layer perceptron model, aiming to explore the application of artificial intelligence in the field of educational evaluation. It can provide teachers with forward-looking teaching references to adjust strategies and implement teaching in accordance with students' aptitude, help students discover weak links and plan their learning, provide new ideas for solving educational practice problems, promote the development of educational artificial intelligence, and provide empirical support for achieving high-quality and equitable education.

Keywords

Artificial intelligence; Student performance prediction; Educational evaluation; Multi-layer perceptron model; Educational equity

1. Introduction

In today's era of rapid digital development, the field of education is undergoing profound changes. The integration of artificial intelligence technology has undoubtedly opened up new paths for the development of education, but it has also brought many challenges. In the early days, balanced education and teaching students in accordance with their aptitude faced many difficulties, such as lack of funds and uneven distribution of teaching staff. These problems seriously hindered the overall improvement of education quality and the realization of educational equity. To this end, the country has invested a lot of money to promote the process of artificial intelligence in education, which has enabled many regions to make significant progress in the introduction of hardware facilities and technologies, equipped with advanced artificial intelligence education equipment and resources, and to some extent alleviated the initial resource shortage. dilemma.

However, as the advancement of AI in education has progressed, new problems have gradually emerged. Although hardware conditions have improved, teachers and students still have obvious shortcomings in the practical application of AI. Most people lack AI thinking and find it difficult to effectively integrate advanced technologies into the teaching and learning process. In addition, there is currently a lack of a systematic and specific plan to guide how to use artificial intelligence for teaching and learning, which has become a key factor restricting the further development of educational artificial intelligence.

In this context, it is particularly important to predict student scores. Predicting student scores by using artificial intelligence technologies such as multi-layer perceptron models can play a key role in the field of education evaluation. On the one hand, it can provide teachers with more forward-looking teaching references, allowing teachers to understand students' learning conditions and possible problems in advance, so as to adjust teaching strategies in a targeted manner, implement teaching in accordance with students' aptitude, and improve teaching quality. On the other hand, for students, grade predictions can help them discover their weak links in learning in a timely manner, plan their study plans reasonably, and improve their learning efficiency. From a macro perspective, this initiative provides new ideas and methods for solving practical problems in educational practice, can effectively promote the development of educational artificial intelligence, and provides solid empirical support for achieving better and fairer educational goals.

2. Related Work

Zongwen Fan et al. [1] used the residual-based complementary CatBoost method (C-CatBoost) for student performance prediction, which improved the effectiveness of student performance prediction. Mahdi-Reza Borna et al. [2] focused on inclusive and adaptive educational assessment based on the effectiveness of various AI models, including random forests, XGBoost, and recurrent neural networks (RNNs), in identifying at-risk students and differentiating between academic achievement levels. Zheng Luo et al. [3] used a method to construct a multidimensional spatio-temporal feature dataset, taking into account factors such as students' basic information, performance at each stage of the semester, and education indicators of their place of origin, to effectively predict students' academic performance. Adnan Zeb al. [4] used a deep learning (DL)-based model named Student Academic Performance Prediction Network (SAPPNet) to predict student grades. They considered a questionnaire-based Jordanian University dataset that contained demographic information, the use of digital tools before and after COVID-19, and other information to effectively predict student grades. Kaitong Wang et al. [5] combined DistilBERT with a hybrid LSTM (DBTM) method and Spotted Hyena Optimizer (SHO) to change parameters, taking into account the existing needs of predicting student performance in education and the shortcomings of traditional models when faced with increasing amounts of data, to effectively evaluate student performance. Dien Tran Thanh et al. [6] used various deep-learning techniques to predict student performance. In addition, we also analyzed and introduced several data preprocessing techniques (e.g., quantile transformation and MinMax Scaler), extracted them into well-known deep learning models (such as long short-term memory (LSTM) and convolutional neural network (CNN)) to perform prediction tasks, and the proposed method provided good prediction results.

3. Data Description

3.1 Data Profiling

Some of the data sets used in the experiment in this paper are shown in Table 1. In order to achieve the goal of predicting student scores, this paper conducts an in-depth analysis of these indicators.

Table 1. Table showing some data sets

id	Gender	Eth-nicGroup	ParentEduc	LunchType	TestPrep	ParentMari-tal-Status	PracticeS-port	IsFirstChild	NrSiblings	Transport Means	Wkly Stud-y-Hours	MathScore	ReadingScore	WritingScore
1	1	2	4	1	0	3	3	1	4	0	2	87	93	91
2	0	3	5	1	0	2	3	1	0	0	1	76	78	75
3	1	2	1	1	0	2	2	1	1	0	1	73	84	79
4	1	2	5	1	1	4	1	0	1	1	1	85	93	89
5	0	2	5	0	0	2	3	1	1	1	3	41	43	39
6	0	4	3	0	1	3	3	0	3	1	3	65	64	68
7	0	4	1	1	0	1	3	1	1	0	1	40	52	43
8	1	2	3	1	0	2	2	0	1	1	1	66	82	74

Table 2. Data field description

Fields	Field Description
id	Student Number
Gender	Student Gender
EthnicGroup	The ethnic groups of students are group A, group B, group C, group D, and group E, which correspond to 1, 2, 3, 4, and 5 respectively.
ParentEduc	The educational levels of the students' parents are college, undergraduate, high school, master's, college, and high school, corresponding to 1, 2, 3, 4, 5, and 6 respectively.
LunchType	The student lunch types are standard, free/reduced, and correspond to 1 and 0 respectively.
TestPrep	Students prepare for the exam. Completed and None correspond to 1 and 0 respectively.
ParentMaritalStatus	The marital status of the student's parents is divorced, married, single, widowed, corresponding to 1, 2, 3, 4 respectively.
PracticeSport	The frequency of students' participation in sports is never, often and sometimes, corresponding to 1, 2 and 3 respectively.
IsFirstChild	Is the student the first child in the family? If yes, it corresponds to 1; if no, it corresponds to 0.
NrSiblings	Number of siblings of the student
TransportMeans	The student's way of going home is private and school_bus, which represent 1 and 0 respectively.
WklyStudyHours	If a student studies 5 to 10 hours, less than 5 hours, and more than 10 hours per week, the corresponding values are 1, 2, and 3 respectively.
MathScore	Student math scores
ReadingScore	Student reading scores
WritingScore	Student Writing Scores

Figure 1. Visualization of some data of math scores, reading scores, and writing scores.

3.2 Pearson Correlation Coefficient Method

A frequently utilized correlation is known as the Pearson correlation, which is named in honor of Karl Pearson, who built upon a concept that was proposed by Francis Galton during the 1880s. This correlation coefficient is often referred to as the "Pearson Product-Moment Correlation." Typically, the Pearson correlation coefficient is denoted by the letter r, serving as a measure of the linear relationship (or association) that exists between two random variables. For a population, the Pearson correlation coefficient between two variables is defined as the ratio of the product of covariance to the standard deviation of those variables (or alternatively, the normalized covariance). In general, the simple correlation coefficient for a sample is also represented by r, with its calculation formula as follows:

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

Let n represent the sample size, while the terms observed values and means refer to the two variables in question. The term r quantifies the strength of the linear relationship between these variables. The range of r spans from -1 to

+1. When r is greater than 0, it suggests a positive correlation between the two variables; in this case, as the value of one variable increases, so does the other. Conversely, if r is less than 0, this indicates a negative correlation, meaning that as one variable's value rises, the other variable's value falls. A larger absolute value of r signifies a stronger correlation. It is important to note that this does not imply any causative link. When r equals 0, it signifies that the two variables do not exhibit a linear correlation, although they may be related in other, possibly non-linear ways (such as through curves). The sample correlation coefficient serves as a basis for deducing whether a correlation exists between the two variables within the larger population. To test the null hypothesis that the population correlation coefficient is 0, the t statistic can be applied. A significant t-test results in the rejection of the null hypothesis, which indicates a linear correlation between the two variables; should the t-test lack significance, the null hypothesis remains unrefuted, suggesting no linear correlation.

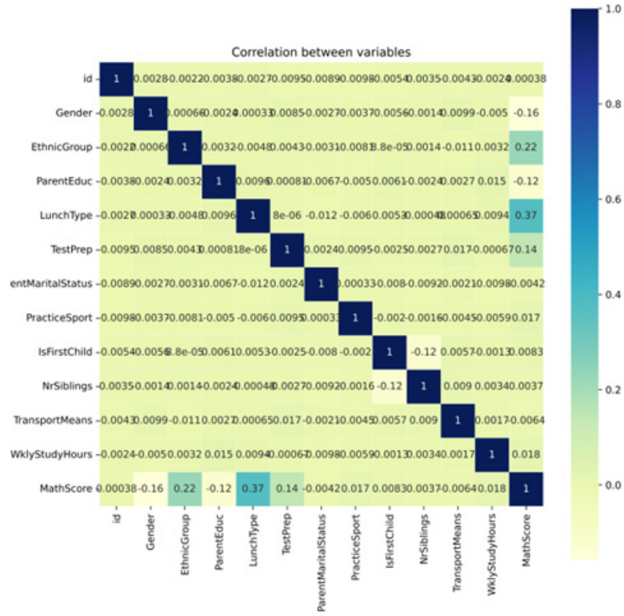

Figure 2. Correlation diagram of various parameters of mathematics scores.

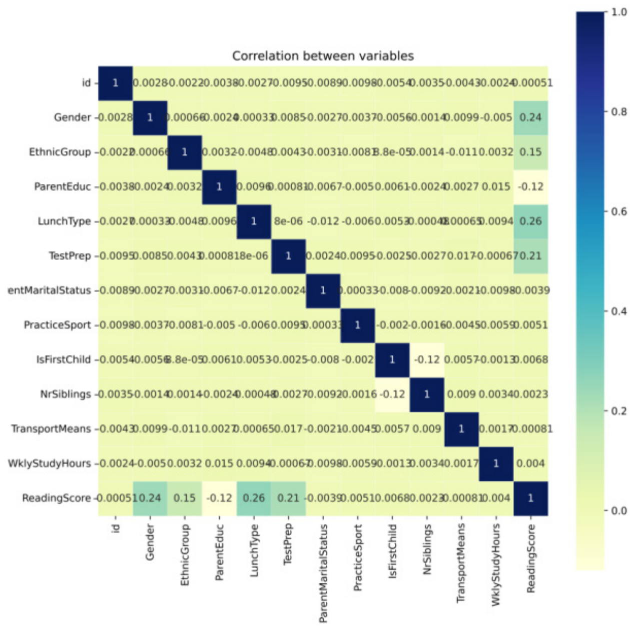

Figure 3. Correlation diagram of various parameters of reading scores.

Based on Figure 2, we can see:

1) LunchType (student lunch type): This indicator has a significant positive correlation with students' math scores (the correlation coefficient is about 0.37). This suggests that math scores may involve grade movement differently than the type of lunch a student receives, such as standard or free/reduced.
2) Ethnic Group (student ethnic group): has a significant positive correlation with students' math scores (correlation coefficient is about 0.22). This indicates that different ethnic groups may be involved in different changes in students' math scores.
3) TestPrep (students' preparation before the test): has a significant positive correlation with students' math scores (the correlation coefficient is about 0.22). This suggests that mathematics learning scores may be related to students' preparation before the exam, such as whether they are well prepared or not, which will more easily affect the level of mathematics scores.

Based on the Figure 3, we can see:

1) LunchType (student lunch type): This indicator has a significant positive correlation with students' reading scores (the correlation coefficient is about 0.26). This suggests that reading scores may be associated with grade movement differently than the type of lunch a student receives, such as standard or free/reduced.
2) Gender: It has a significant positive correlation with students' reading scores (the correlation coefficient is about 0.24). This indicates that students' gender may be involved in changes that are different from students' reading scores.
3) TestPrep (Student Exam Preparation): The student's test result is a positive correlation (correspondence ratio is approximately 0.21). This is how you can learn fractions, and how many students have before taking the exam.
4) EthnicGroup (student ethnic group): Positive correlation between students and students (correspondence number approximately 0.15). Therefore, it is possible to express different ethnic groups and students' participation in different fractions.

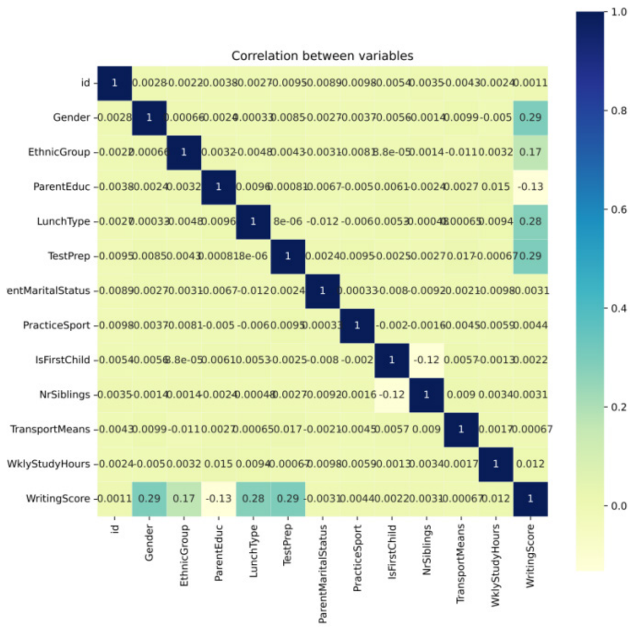

Figure 4. Copying fractions of each number index correlation diagram.

Based on Figure 4, we can see:

1) TestPrep (preparation for student examination): The positive correlation of the student's copy fraction with the correct correlation (correspondence ratio of approximately 0.29). It is possible to make a copy of a fraction in this case, and the situation before the student examination is as low as possible.
2) Gender (student sex distinction): The positive similarity of student copying fractions by Yukio (correspondence number approximately 0.29). Therefore, the student's character can be changed and the student's copying fraction is different.

3) LunchType (Student Lunch Type): This is the first indicator and the student copy fraction has a positive correlation (correspondence number is approximately 0.28). This expression can be used to copy fractions, and students can change the number of different fractions in the lunch category.

4. MLP Model Introduction

In recent years, the prominence of deep learning has transformed multi-layer perceptrons into a focal point of research. Particularly within the realms of image, speech, and natural language processing, deep neural networks built upon multi-layer perceptrons have achieved significant advancements, including target detection, image classification, speech recognition, semantic segmentation, machine translation, and various other tasks.

The multilayer perceptron (MLP) architecture represents a crucial form of artificial neural network. It operates as a feedforward neural network that translates a collection of input vectors into corresponding output vectors. This structure can typically be depicted as a directed graph, comprising several layers of interconnected nodes, with each layer fully linked to its successor. Excluding the input nodes, every node functions as a neuron equipped with a nonlinear activation function, trained via the back-propagation algorithm. To establish a multi-layer perceptron model, the initial step involves weight initialization, followed by the feeding of input samples into the network for forward propagation. For each input sample, the weighted sum from the input layer to the hidden layer is computed and then fed into the activation function to derive the hidden layer's output. This output is subsequently relayed to the next layer until the output layer generates the classification result. The network's output is compared against the sample's label, and the loss function is defined. Frequently employed loss functions include Mean Squared Error and Cross Entropy. The back-propagation algorithm serves as a prevalent optimization technique, which reduces the loss function by calculating gradients and adjusting weights and biases, with the gradient of each weight concerning the loss being computed through the back-propagation process.

This study utilizes a multi-layer perceptron (MLP) architecture with eight layers, comprising one input layer, six hidden layers, and a single output layer. This configuration allows the MLP to effectively manage intricate data patterns.

Input Layer: The initial layer of the network, known as the input layer, is responsible for accepting raw data. Each node within this layer corresponds to a distinct feature of the data.

Hidden layer: Following the input layer are the hidden layers, which perform most of the computational work. Each hidden layer is fully connected to all nodes in the previous layer. There are six hidden layers in the eight-layer MLP model, and the number of neurons and activation functions of the hidden layers can be adjusted according to the specific task. Commonly used activation functions include ReLU, sigmoid, and tanh. This article uses the Relu function.

Output layer: The final layer of the MLP represents the concluding layer of production. It encompasses the export quantity level, the type of activation function, and particular assignments. For instance, in a binary classification scenario, the output layer could consist of a single neuron equipped with a sigmoid activation function. Conversely, in multi-classification cases, the output layer's number of neurons corresponds to the number of categories, employing a softmax activation function.

MLP has excellent global optimization and generalization capabilities, and its generalization ability can be enhanced through appropriate regularization and early stopping techniques. Its simple and efficient implementation enables it to show superior prediction accuracy on small-scale datasets. These advantages make MLPs ideal for fitting and predicting in datasets with fewer features.

5. Experimental Analysis

This paper uses the MLP model and LSSVM model to predict the results of the above problem. First, the target variables 'MathScore, ReadingScore, and WritingScore are binarized. Select feature columns (id, Gender, EthnicGroup, ParentEduc, LunchType, TestPrep, ParentMaritalStatus, PracticeSport, IsFirstChild, NrSiblings, TransportMeans, WklyStudyHours), The dataset is then divided into a training set and a test set, with the test set accounting for 20%. One-hot encodes categorical data. And set the random state to 42 to ensure the repeatability of the division. The maximum number of iterations is 1000, which is used to control the training time and avoid overfitting or underfitting problems. Then use the model's training function to input the features of the training set and the target data for training, so that the model can learn the mapping relationship from features to target variables.

After training is completed, the prediction function of the model is used to make predictions and obtain the prediction results. Then the mean absolute error (MAE), mean square error (MSE), and coefficient of determination (R^2) are calculated.

Table 3. Experimental values of MAE, MSE, and R^2 for MLP and LSSVM

	MLP			LSSVM		
	MAE	MSE	R^2	MAE	MSE	R^2
MathScore	4.290421	28.791305	0.882189	4.584754	33.562849	0.862665
ReadingScore	3.308117	16.905099	0.925017	3.545130	19.956587	0.911482
WritingScore	2.927973	13.401510	0.945806	3.147241	16.376517	0.933776

This problem uses a dataset of nearly 200,000 records. Feature columns are selected based on the correlation between various indicators and student scores, including id, Gender, EthnicGroup, ParentEduc, LunchType, TestPrep, ParentMaritalStatus, PracticeSport, IsFirstChild, NrSiblings, TransportMeans, and WklyStudyHours to predict students' math scores, reading scores, and writing scores.

Table 3 presents the mean absolute error (MAE), mean square error (MSE), and the coefficient of determination (R^2) for the test set of both the MLP and LSSVM models. An analysis of the error metrics, including MAE, MSE, and R^2, indicates that the MLP model demonstrates a notable level of reliability and fitting accuracy when compared to the LSSVM model in relation to the predicted and actual values of the experimental outcomes.

6. Conclusion

This paper focuses on academic performance. By deeply mining the multi-dimensional indicator data in the data, it analyzes the key driving factors of academic performance and constructs MLP and LSSVM prediction models. This article analyzes the correlation between all indicators in the data and academic performance. Through statistical methods and correlation coefficient calculations, we found that MathScore is positively correlated with LunchType, EthnicGroup, and TestPrep, ReadingScore is positively correlated with LunchTypeGenderTestPrep EthnicGroup, and WritingScore is positively correlated with TestPrep, Gender, LunchType, and EthnicGroup. Based on the correlation of these parameters, experiments were conducted using a multi-layer perceptron model (MLP) and a least squares support vector machine (LSSVM). The results showed that the MLP model performed significantly better than the LSSVM.

References

[1] Fan Z, Gou J, Weng S. Complementary CatBoost based on residual error for student performance prediction. Pattern Recognition. 2025;161:111265.

[2] Borna MR, Saadat H, Hojjati AT, et al. Analyzing click data with AI: implications for student performance prediction and learning assessment. In: Frontiers in Education. Frontiers Media SA; 2024. p. 1421479.

[3] Luo Z, Mai J, Feng C, et al. A method for prediction and analysis of student performance that combines multi-dimensional features of time and space. Mathematics. 2024;12(22):3597.

[4] Junejo NUR, Huang Q, Dong X, et al. SAPPNet: students' academic performance prediction during COVID-19 using neural network. Scientific Reports. 2024;14(1):24605.

[5] Wang K. Optimized ensemble deep learning for predictive analysis of student achievement. Plos One. 2024;19(8):e0309141.

[6] Dien TT, Luu SH, Thanh-Hai N, et al. Deep learning with data transformation and factor analysis for student performance prediction. International Journal of Advanced Computer Science and Applications. 2020;11(8).

[7] Meng F. Research on text recognition methods based on artificial intelligence and machine learning. Advances in Computer and Communication. 2023;4(5):340-344.

[8] Khin Shin Thant, Ei Theint Theint Thu, Myat Mon Khaing, Khin Lay Myint, Hlaing Htake Khaung Tin. Evaluation of student academic performance using Naïve Bayes classifier. Advances in Computer and Communication. 2020;1(1):46-52.

[9] Xu H, Deng Y. Dependent evidence combination based on Shearman coefficient and Pearson coefficient. IEEE Access.

2017;6:11634-11640.

[10] Armstrong RA. Should Pearson's correlation coefficient be avoided? Ophthalmic and Physiological Optics. 2019;39(5):316-327.

[11] Adler J, Parmryd I. Quantifying colocalization by correlation: the Pearson correlation coefficient is superior to the Mander's overlap coefficient. Cytometry Part A. 2010;77(8):733-742.

[12] Gardner MW, Dorling SR. Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences. Atmospheric Environment. 1998;32(14-15):2627-2636.

[13] Adeli H, Yeh C. Perceptron learning in engineering design. Computer-Aided Civil and Infrastructure Engineering. 1989;4(4):247-256.

[14] Rosenblatt F. The perceptron: a probabilistic model for information storage and organization in the brain. Psychological Review. 1958;65(6):386.

[15] Spring J. Research on globalization and education. Review of Educational Research. 2008;78(2):330-363.