Migration Letters

Volume: 21, No: 4, pp. 1109-1117

ISSN: 1741-8984 (Print) ISSN: 1741-8992 (Online)
www.migrationletters.com

Student Single Subject Grade Prediction Model Based on Feedforward Neural Networks

Shuang Shuang[1], He Yuchu[2], Tan Wee Hoe[3]

Abstract

Predicting student grades can assist teachers and educational administrators in timely under-standing of students' learning situations, thereby intervening and tutoring students who may fail in academic tests in advance, or providing personalized academic advice and support. This optimizes and enhances teaching quality, promoting the transformation of education towards digitization and intelligence. This paper analyzes and preprocesses academic performance data collected during online and offline teaching, considering various factors affecting college stu-dents' grades, to build a dataset with 10 features and 3102 real data entries of college students' academic performance. Based on this, a multilayer neural network model using feedforward neural network technology was constructed to predict grades in a college computer basics course, assess if students can pass the final exam, and the model was trained and tested with real data. Experimental results show that our proposed model achieved 92.15% and 90.74% accuracy on training and testing datasets, respectively, supporting further development of educational support applications.

Keywords: *student academic performance; neural networks; single subject grade prediction.*

1. Introduction

Student grade prediction, also known as academic performance prediction, is one of the important research issues in the field of educational data mining [1-4]. The goal of student grade prediction is to use academic performance information and basic student information collected during the teaching process, or survey data, to predict students' future academic performance, such as course grades. In recent years, with the popularization and expansion of higher education in China, the number of college students has been increasing steadily, while the growth of the number of teachers is limited by various factors, leading to an imbalance in the teacher-student ratio in many schools [5]. This phenomenon has made classroom teaching increasingly large-scale, making it difficult for teachers to track and understand each student's learning situation, thereby leading to a decline in educational quality. For example, students failing course assessments can lead to consequences like retakes, repetitions, academic probation, or even dropout, severely affecting their academic development. In this context, the study of student grade prediction is of significant importance. Student grades are not only an important indicator for assessing academic ability but also an essential tool for schools, educational institutions, and governments to monitor and improve educational quality. At

[1] Faculty of Social Sciences & Liberal Arts, UCSI University, Malaysia.
School of Elementary Education, Zhengzhou Normal University, China.
[2] School of Information Science and Technology, Zhengzhou Normal University, China.
[3] Faculty of Social Sciences & Liberal Arts, UCSI University, Malaysia.

the same time, accurate prediction of student grades has a profound impact on the fairness and efficiency of the educational system. By understanding students' potential performance in advance, educators can take early measures to provide better academic support, promoting their academic success. On the other hand, the uncertainty and variability of student grades also make it difficult for schools and education policymakers to make effective resource allocation and decisions.

The goals of grade prediction vary, including predictions of individual course grades, overall performance, average grade point average (GPA), and course pass/fail outcomes. Traditional methods of student grade prediction often rely on teachers' subjective judgments, standardized test scores, or students' historical grades. However, these methods have limitations such as subjectivity and inapplicability to personalized predictions. Additionally, the diversity and individual differences among students make personalized grade prediction increasingly important. Different students have unique learning styles, interests, and academic needs, making a one-size-fits-all approach insufficient.

With the advancement of information technology, the large-scale collection of student data, including learning habits, online behaviors, and social media activities, has become easier. This data provides a valuable resource for machine learning-based grade prediction. Machine learning algorithms can analyze and uncover patterns and correlations in this data to more accurately predict student grades. These algorithms offer robust personalized modeling capabilities, providing individualized grade predictions and suggestions based on each student's characteristics and historical performance.

There have been numerous studies in the field of predicting student grades and academic performance using machine learning algorithms. For example, Cherny and others found that students are more fearful of statistics courses compared to other types of psychology lectures. They conducted a survey on 154 undergraduates using a 22-item scale comprising three components: perception of mathematics, perception of statistics, and perception of correlation. The study concluded that over a semester, perception of math anxiety decreased, perception of correlation increased, but the negative perception of math anxiety increased. This study used traditional methods for grade prediction analysis, but its small sample size and lack of advanced automated methods limited its efficiency and generalizability.

Jia and others introduced the SNMB model to address the issue of poor accuracy in predicting student grades. The SNMB model, a fusion of Sequential Minimal Optimization (SMO) and Naive Bayes, improves prediction accuracy compared to Cherny's method. It predicts students' performance in professional courses based on their performance in compulsory courses. First, the SMO algorithm predicts academic performance and generates initial results; then Naive Bayes resolves any inconsistencies in these results; finally, it generates the final prediction for students' professional course grades. To validate their proposed model, they conducted extensive trials comparing SNMB with four alternative prediction methods. The results showed that the proposed SNMB model outperformed all other examined techniques. However, this method also has drawbacks: it relies on manually extracted student data features and is sensitive to the extracted features. With large data volumes, the workload for feature engineering is substantial.

Wang and others conducted a discourse analysis of a 16-week internet-based chat room conversation in a statistical psychology course. The study showed that various topic categories (e.g., the total number of student comments) correlated with the class's final grade for the semester. Additional examination of the third-week chat room conversation revealed a correlation between two topic categories (i.e., student responses to questions or examples raised in lectures and the total number of student comments) and the class's final grades. They explored the implications of these findings on teachers' ability to detect early warning indicators of student performance in virtual classrooms. This study also relied on traditional analysis methods.

Alvin and others conducted a complete statistical test on students' learning habits and assignment grades on an online learning platform, determining the correlation coefficients between various elements and the final outcomes. Michelle and others used maximum likelihood estimation to evaluate the learning motivation questionnaire data of over 200 students, identifying the correlation between various components of learning motivation and academic performance. Xenos and others adopted a Bayesian model to predict the risk of student dropout based on previous learning behavior and performance on an online learning platform. This study also faced issues of small sample size and difficulty in generalizing and applying the analysis methods.

Existing related research often uses traditional statistical analysis methods for student modeling. These probability and statistical techniques frequently impose restrictions and assumptions on the data distribution. For example, Bayesian techniques presuppose that each element is independent of others. If these conditions or assumptions contradict reality, the prediction results will be impacted. Additionally, these processes are less automated and require significant manual labor. They also demonstrated that by examining estimated models and student-specific combination functions, they could gain insights into the effectiveness of teaching materials provided during courses in various departments.

2. Selection and Analysis

Analyzing the factors affecting student grades and using them as the basis to construct a predictive model is a necessary condition for developing a well-performing model. There are many factors influencing student grades, including social and family environmental factors, educational conditions, and individual student factors (such as intelligence, physical health, psychological aspects, and knowledge background). Our goal is to establish a predictive model to forecast whether students will pass the final exam of 'College Computer Basics'. After comparing and analyzing various potential influencing factors and considering the cost of data collection, we collected information from 3102 students. We selected the following 9 factors that may influence the grades in 'College Computer Basics' course for the grade prediction, as shown in Table 1.

Table 1. Factors affecting performance used in this model

Property name	Meaning	Numeric range	category
questions answered	This is an indicator of student engagement or classroom participation.	positive integer	
online study hours	This metric is used to assess the student's commitment to learning and self-study capabilities.	Positive integer	learning habit
weekly online hours computer basics	This reflects the student's allocation of study time to this specific subject.	Positive integer	learning habit
failed other courses	This is an indicator used to assess a student's academic performance in other subjects.	0 or 1	prior knowledge level
failed course	This directly reflects the student's performance in the course and is the	0 or 1	prior knowledge

	target column for prediction.		level
is class leader	This is an indicator to measure the student's leadership skills or level of involvement in the class.	0 or 1	personality
department	This is used to differentiate students from various aca-demic fields or majors.	string	learning motivation
study pressure	Indicates the level of academic stress perceived by the student, such as "low," "medium," or "high.".	High，medium，low	learning motivation
course satisfaction	This reflects the student's overall satisfaction with aspects such as course content and teaching methods.	0~1	learning motivation
average grade previous courses	This is an indicator used to assess the student's past aca-demic performance.	0~100	emotional factors

Figure 1. Histogram of course satisfaction data

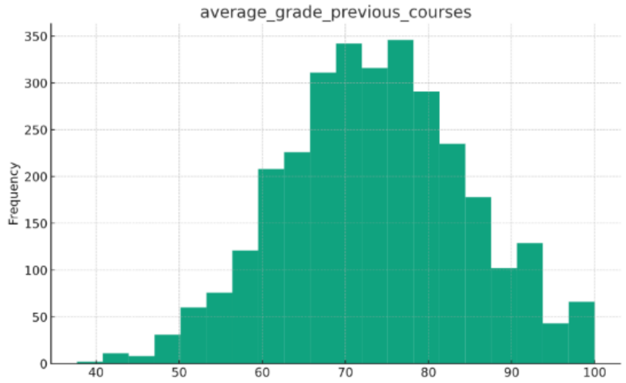

Figure 2. Average grades in previous courses

Figures 1 and 2 respectively show the survey results of students' satisfaction with the course and the average grades of prior courses. These graphs indicate that both sets of data largely follow a normal distribution, reflecting the overall satisfaction of students with aspects like course content and teaching methods, as well as their past academic performance.

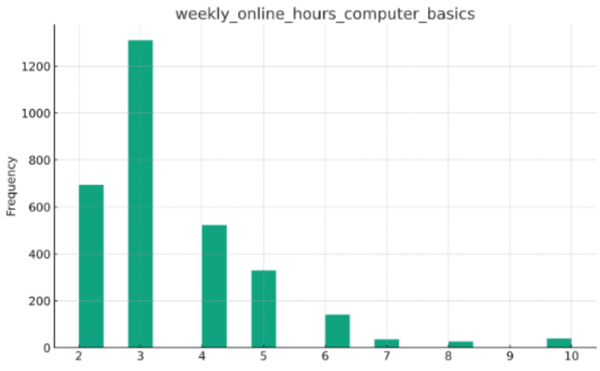

Figure 3. Time spent by students in the "College Computer Fundamentals" course per week

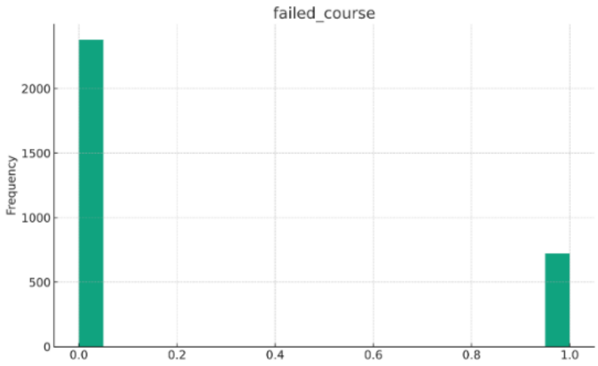

Figure 4. Whether you passed the exam of "College Computer Fundamentals" course

Figure 3 shows the amount of time students spend weekly on the 'College Computer Basics' course, indicating their study habits. Figure 4 shows whether students ultimately passed the exam for the 'College Computer Basics' course. This indicator is the label for the data and the target of the neural network model's prediction. From Figure 4, it can be seen that the pass rate of students is around 70%. This means that our neural network model's accuracy must be significantly higher than 70% to be meaningful.

Among the 10 types of data shown in Table 1, the students' department and academic pressure data are non-numerical and cannot be directly input into the neural network model. We use one-hot encoding to convert these into numerical data. After processing, the data comprises 21 columns, with the 'failed_course' column as the target column, indicating whether the 'College Computer Basics' course was passed, serving as the label for the data. The remaining 20 columns are feature columns, serving as inputs for the neural network model.

3. New Prediction Model

Based on the size of the data and empirical rules, we designed a neural network model with a three-layer perceptron structure. The first layer is a fully connected layer with 20-dimensional input, including all numerical data after preprocessing, as described in Section 2. The output of the first layer is set to 64 dimensions, activated using the ReLU function. The width of the middle layer in the neural network is a hyperparameter set according to empirical rules; generally, a wider middle layer can handle a larger amount of data. Both the second and third layers use 64-dimensional data for input and output, where the 64-dimensional input corresponds to the 64-dimensional output of the previous layer, and the 64-dimensional output is also a hyperparameter. Similarly, we use the ReLU function for activation in the second layer.

The third layer of the neural network, which is the output layer, uses a 1-dimensional output. This is because our prediction target is whether a student can pass the 'College Computer Basics' course, where the outcome of not passing is represented as 1 and passing as 0. In other words, the output is a single data point indicating the probability of passing. Therefore, the dimension of the output layer is 1, and it is activated using the Sigmoid function. The structure of the feedforward neural network used is illustrated in Figure 5., as illustrated in Figure 5.

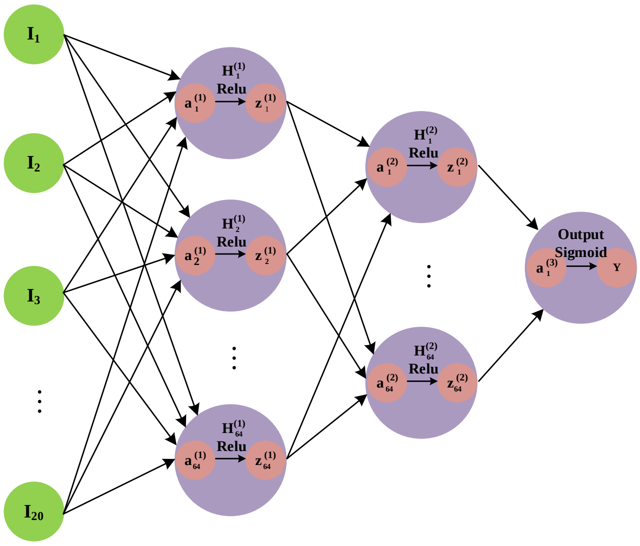

Figure 5. Feedforward neural network structure

4. Experimental Results and Analysis

For the experiment, we used 2326 data entries as training data and 776 as test data. The setup of other experimental parameters is detailed in Table 2.

Table 2. Experimental parameter settings

parameter name	value	illustrate
Training data size	2326	The size of the training dataset
Testing data size	7 76	Test dataset size
batch	6 4	Randomly selected training dataset size for each batch
e poch	500	training times
Learning rate	0.0001 _	Learning rate used by gradient descent algorithm
optimization	Adam	Optimizer used
Loss function	BCELoss	Use cross entropy loss function

Figures 6-8 display the training results of the neural network model. Figure 6 shows the decline in the average value of the loss function during the training process. Figures 7 and 8 respectively depict the changes in the model's accuracy on the training and test datasets as the training progresses. From these figures, it's observed that the final accuracy of the model reached 92.15% on the training dataset and 90.74% on the test dataset. This indicates that the final model can predict with 90.74% accuracy whether a student will pass the 'College Computer Basics' course exam in the final test.

The similar accuracy rates on both training and test datasets suggest that the model did not experience overfitting. Additionally, the accuracy did not significantly improve with more training iterations, indicating no underfitting occurred.4.2. Factor Selection and Data Augmentation.

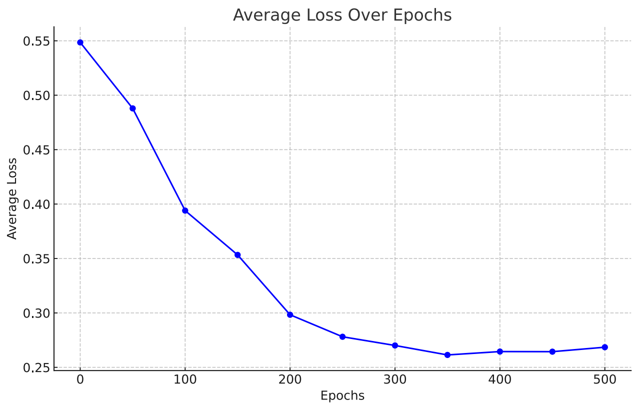

Figure 6. As the number of training times increases, the changes in the average value of the loss function

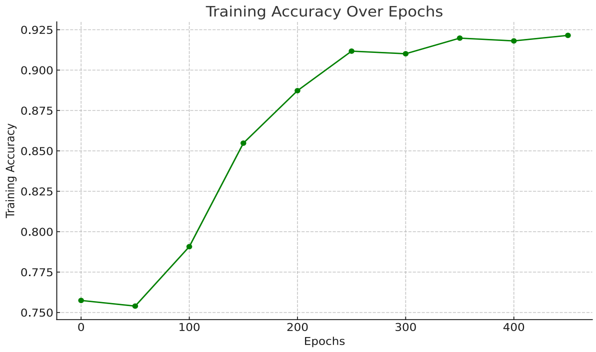

Figure 7. As the number of training times increases, the accuracy of the model on the training dataset changes.

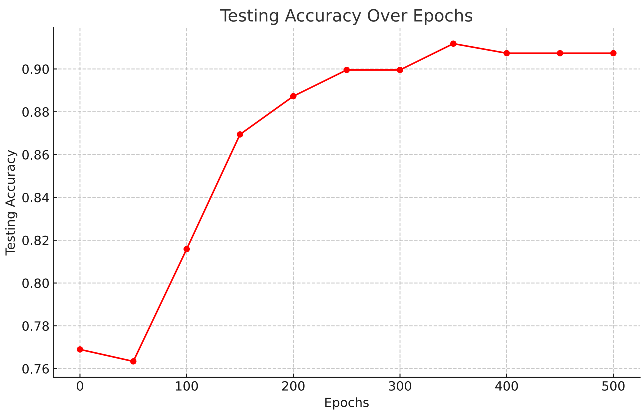

Figure 8. As the number of training times increases, the accuracy of the model on the test dataset changes.

5. Conclusions

Long Short-Term Memory (LSTM) networks can effectively process time-series data. The feedforward neural network model demonstrated effective data fitting and suitability for handling student grade prediction issues. Our three-layer neural network model, designed to predict whether students can pass the 'College Computer Basics' exam, achieved accuracies of 92.15% and 90.74% on the training and test sets, respectively. This model allows teachers to anticipate student performance, take early measures for students at risk of failing, and provide better academic support to promote their academic success. A limitation of this model is its inability to predict specific scores, only pass/fail outcomes. Future work involves using more complex neural network models for regression predictions, offering more effective academic support for teachers, students, and educational administrators.

Acknowledgments

This work was supported by the Henan Province Young Backbone Teachers Project, under Grant No. 2023GGJS169.

References

[1] Romero, Cristóbal, and Sebastián Ventura. "Educational data mining: a review of the state of the art." IEEE Transactions on Systems, Man, and Cybernetics, Part C (applications and reviews) 40.6 (2010): 601-618.

[2] Xiao, Wen , P. Ji , and J. Hu . "A survey on educational data mining methods used for predicting stu-dents' performance." Engineering Reports (2021).

[3] Abu, Marjan Md , U. M. Palash , and I. A. Masud . "An Educational Data Mining System For Predicting And Enhancing Tertiary Students' Programming Skill." The Computer Journal 5(2022):5.

[4] Manga, I , M. B. El-Yakub , and D. T. Nzadon . "AN Intelligent System for Predicting Students Perfor-mance." IOSR journal of computer engineering 1(2022):24.

[5] Ministry of Education, China. "2022 National Education Development Statistical Bulletin." Ministry of Education, July 5, 2023, http://www.moe.gov.cn/jyb_sjzl/sjzl_fztjgb/202307/t20230705_1067278.html.

[6] Macfadyen, Leah P. , and S. Dawson . "Mining LMS data to develop an "early warning system" for educators: A proof of concept." Computers & Education 54.2(2010):588-599.

[7] Christos, et al. "Evaluating and improving educational Material and tutoring aspects of distance learn-ing systems." Studies in Educational Evaluation (2003).

[8] Meehl, Paul E. . "Review of Prediction of the adjustment and academic performance of college students by a modification of the Rorschach method." Journal of Applied Psychology 30.6(1946):660-662.

[9] Tang, Changjie , et al. "personalized courseware construction based on web data mining*." IEEE Com-puter Society (2000).

[10] Baffes, Paul Thomas , and R. J. Mooney . "A novel application of theory refinement to student model-ing." (1996).

[11] Bunkar, Kamal , et al. "Data Mining: A Prediction for Performance Improvement of Engineering Stu-dents using Classification." World of Computer Science and Information Technology Journal, 10.1109/WOCN.2012.6335530. 2012.

[12] Lykourentzou, Ioanna , et al. "Early and dynamic student achievement prediction in e-learning cours-es using neural networks." Journal of the American Society for Information Science & Technology 60.2(2014):372-380.

[13] Zhou Qing et al. "Predicting student failure based on campus card consumption." China Educational Technology and Equipment 24(2017):5.Wassan, and J. Talreja . "Discovering Big

Data Modelling for Educational World." Procedia - Social and Behavioral Sciences 176(2015):642-649.

[14] Wassan, and J. Talreja . "Discovering Big Data Modelling for Educational World." Procedia - Social and Behavioral Sciences 176(2015):642-649.

[15] Xi-Hu, Lin , and F. J. University . "Research on the Application of Big Data of Education in the Educa-tion and Management of University Students." The Journal of Shandong Agriculture and Engineering University (2019).

[16] Liu, Yuping , et al. "Collaborative Learning Team Formation: ACognitive Modeling Perspective." Inter-national Conference on Database Systems for Advanced Applications Springer International Publish-ing, 2016.

[17] Wu, Runze , et al. "Knowledge or Gaming?: Cognitive Modelling Based on Multiple-Attempt Re-sponse." International Conference on World Wide Web Companion International World Wide Web Conferences Steering Committee, 2017.

[18] Mingyu Feng, Joseph Beck . "Educational Data Mining 2009 Back to the future: a non-automated meth-od of constructing transfer models." (2010).

[19] Joshi, Mahesh , et al. "Analyzing Students performance Using Academic Analytics." International Conference on Ict in Business Industry & Government IEEE, 2017.

[20] Bakhshinategh, Behdad , et al. "Educational data mining applications and tasks: A survey of the last 10 years." Education and information technologies (2018).