

The Role of AI In Predicting Students' Academic Grades

Zihan Yuan *

School of Internet finance and information engineering, Guangdong University of Finance,
Guangzhou, Guangdong, China

* Corresponding Author Email: 221543417@m.gdof.edu.cn

Abstract. Students' academic performance is crucial to educators, parents, and researchers. It is a key basis for evaluating education quality and making decisions. It is necessary to understand and predict performance accurately. This study focuses on the prediction of students' academic performance and comprehensively explores the factors affecting performance and prediction methods. The factors influencing academic performance were analyzed from multiple perspectives, including personal perspective, learning motivation and time management, and family and teacher factors. In terms of prediction methods, an overview of machine learning and deep learning prediction models is given, such as Extreme Gradient Boosting (XGBoost), Logistic Regression (LR), Back Propagation Neural Networks (BPNN), etc. These methods all show good prediction results. The study found that multiple factors are intertwined to form a complex performance-influencing mechanism. Although existing prediction methods are diverse and effective, they still face problems such as difficulty in obtaining data sets, high costs, incomplete data, and insufficient model adaptability, which can easily lead to prediction bias. This study aims to improve prediction accuracy, serve educational and teaching practices, and promote the improvement of education quality by optimizing data set utilization, developing adaptive models, and interdisciplinary collaboration.

Keywords: Student learning performance prediction; influencing factors; machine learning; deep learning.

1. Introduction

In today's education field, academic performance always attracts much attention. As the scale of education expands and educational concepts are updated, accurately understanding and predicting academic performance is crucial for educators, parents, and researchers. Grades are not only a reflection of the level of knowledge mastered, but also the key to education quality assessment and educational decision-making. Chen et al. analyzed the factors affecting grades through big data and used Back Propagation Neural Networks (BPNN) to predict grades. Experiments have shown that the model can accurately predict students' academic performance and provide an effective neural network method for grade prediction [1]. According to Li's research, using student behavior data and integrating five machine learning algorithms through Stacking for grade prediction improved the accuracy of predictions. The study emphasized the importance of behavioral data in grade prediction [2]. At the macro level, it reflects the effectiveness of the education system; at the micro level, it is related to the individual development of students.

When discussing student academic performance, it is necessary to be aware of its complexity and multifaceted nature. Performance is influenced by a complex mix of factors, including the individual student, motivation and time management, family, and teachers. For example, students with psychological problems such as depression or excessive anxiety may have their grades affected. Similarly, time management tendencies, parenting styles, and teacher gender may all be key variables influencing achievement.

To gain a deeper understanding of learning performance and its prediction methods, this paper conducts a systematic review. In terms of factors affecting academic performance, first of all, from the perspective of individual students, mental health, gender, and emotions have a significant impact on academic performance. Mental health is the cornerstone of learning, and psychological problems may lead to distraction and thus affect academic performance. Gender differences affect learning styles, which in turn affects academic performance. Emotions are also crucial as non-intellectual

factors. Moderate anxiety may promote performance, but excessive or insufficient anxiety may have the opposite effect. The second is learning motivation and time management factors. Learning motivation is divided into surface motivation, deep motivation, and achievement motivation, which are significantly correlated with academic performance. Deep and achievement motivation helps students set long-term goals and overcome difficulties. Reasonable time management can help high school students study and review in an organized manner and prepare for exams with ease. Furthermore, family factors also affect students' academic performance. Parents' occupations and educational levels indirectly affect academic performance through the family learning atmosphere and the degree of care and management. Family education style plays an important role in academic performance, whether directly or through mediating factors. In addition, the impact of teacher factors, especially teacher gender, on student performance is particularly significant among female students. In terms of prediction methods, machine learning, and deep learning models are widely used. Extreme Gradient Boosting (XGBoost) and Logistic Regression (LR) performed well in analyzing student performance data. BPNN and Gradient Boosting Regression Tree (GBRT) have obvious advantages in dealing with complex nonlinear problems. These models provide powerful tools for predicting student performance. However, the current research has limitations in dataset acquisition and model adaptability. Prospects are given for future research through interdisciplinary collaboration, improved dataset utilization methods, and the development of adaptive models to improve the accuracy of academic performance prediction and provide better guidance for educational practice.

2. Factors Affecting Student Performance

2.1. Student's Perspective

Students' mental health has a crucial impact on academic performance. Research by Lang et al. shows that there is a significant negative correlation between students' depression and their academic performance. This study used the CESD scale to investigate and found that a 20-point cutoff was used to distinguish depressed people from non-depressed people. The higher the score, the higher the degree of depression [3]. This is because psychological problems may lead to students' inattention and lack of motivation to learn. For example, when students are depressed for a long time, they may lose interest in learning and be unable to devote themselves to learning.

According to research by Jiang et al., student gender is one of the important factors affecting college students' academic performance. This study used Shapley Additive Explanation (SHAP) based on the Gradient Boosting Decision Tree (GBDT) model to analyze about 5,000 information records of 145 college students and found that boys generally perform better in exams, while girls are relatively weak. This conclusion has been further verified in the prediction of test failure rate by gender indicators, showing that boys have an advantage over girls in predicting academic performance. In addition, the study also pointed out that gender, as one of the personal factors, together with other indicators such as age, whether there is additional work, whether there is a partner, etc., affects the academic performance of college students, but it has a positive effect on predicting academic performance, that is, gender factor Can predict students' academic performance to a certain extent [4].

As an important non-intellectual factor, emotion's impact on students' performance cannot be ignored. According to research by Liu et al., students' academic emotions are not only closely related to test scores but also throughout the entire learning process, including classroom learning, after-school homework, etc. Positive high arousal and positive low arousal emotions can positively predict academic performance, while negative low arousal emotions hurt academic performance. This conclusion is consistent with the research of Pekrun et al., that is, academic emotions are composed of two dimensions: arousal and pleasure (valence). Positive emotions tend to have positive valence, and negative emotions tend to have negative valence. In addition, multiple studies have also confirmed that moderate learning anxiety is beneficial to academic performance, but too high or too low anxiety will reduce academic performance, showing an inverted "U" shaped relationship [5].

Therefore, emotional management is crucial to improving students' academic performance. Teachers should actively pay attention to students' emotional changes, guide students to cultivate positive emotions, and reduce negative emotions, thereby improving academic performance.

2.2. Study Motivation and Time Management

According to the research of Wang et al., there is a significant correlation between learning motivation and students' academic performance. In particular, surface motivation has a direct and significant positive impact on academic performance. However, over-reliance on surface motivation may lead to insufficient learning motivation, which will instead Hurt learning. At the same time, deep motivation and achievement motivation are equally important. They are closely related to academic performance and can motivate students to establish long-term goals and firm beliefs, overcome difficulties in learning, and ultimately achieve excellent results. Therefore, when predicting students' academic performance, learning motivation is an important factor that cannot be ignored, especially the stimulation of deep motivation and achievement motivation, which plays a vital role in improving students' performance and all-round development [6].

According to research by Wang et al., time management tendencies have a significant impact on the academic performance of high school students, and are positively correlated, and can effectively and positively predict academic performance. This study conducted a questionnaire survey on 809 high school students and found that time management tendency and its three dimensions are significantly positively correlated with learning investment and academic performance [7]. This shows that reasonable time management can make high school students calmer when learning new knowledge and reviewing old knowledge, and can also prepare for emergency exams, which is conducive to smooth performance in exams and achieving excellent results. Therefore, to improve the academic performance of high school students, teachers can invite students who are good at using time management strategies to share their experiences, use courses or lectures to teach time management methods and recognize students' progress in time management. Parents should also cooperate, remind, and supervise their children's implementation of plans during holidays.

2.3. Family Factors

According to Zhao's research, parents' occupations and academic qualifications are closely related to the academic performance of high school students. Specifically, the mother's career type directly affects the time and energy she spends caring for and managing her children, while the father's academic level is related to the creation of a family learning atmosphere and the degree of intellectual development of the child [8]. For example, when the father is an intellectual or a cadre and has a higher education level, he is more likely to create a good learning environment for students and promote their academic progress. On the contrary, if the mother can devote more time to the family and provide better life care and study management for the children, the student's academic performance can also be improved. Therefore, parents' occupation and academic background are family factors that cannot be ignored in influencing students' academic performance.

According to Li's research, family parenting style is one of the important factors affecting students' academic performance. Positive family upbringing styles can have a positive predictive effect on students' academic performance, while negative family upbringing styles may have a negative predictive effect. Family parenting style not only directly affects students' academic performance, but also indirectly affects students' academic performance through intermediary factors such as subjective well-being and achievement motivation. Specifically, family upbringing can explain 15.2% of the variation in academic performance variables, and under a positive family upbringing, students tend to have better academic performance [9]. Therefore, when predicting students' academic performance, the impact of family parenting styles should be fully considered to help parents and teachers correctly understand and adopt appropriate parenting styles to promote students' learning progress and all-round development.

2.4. Teacher Factors

According to relevant research by Liu et al., teachers' gender has a significant impact on students' academic performance, especially among girls. Specifically, female teachers have a significant positive impact on girls' academic performance, especially in English subjects, and this impact is significantly greater than the impact on boys. However, for boys, teacher gender has no significant impact on their academic performance [10]. This finding emphasizes the positive role of teacher-student gender matching in girls' learning process and prompts educators to consider the matching of teacher gender and student gender when allocating teaching resources to better promote students' learning progress.

3. Student Performance Prediction Methods

3.1. Prediction Methods Based On Machine Learning

Machine learning is a multi-field interdisciplinary subject that involves probability theory, statistics, approximation theory, convex analysis, algorithm complexity theory, and other disciplines. Machine learning uses algorithms to parse data, learn from it, and make predictions. It is widely used in the field of student performance prediction. Zhang et al. proposed a student performance prediction model based on the XGBoost algorithm and applied it to the open-source student performance dataset for data mining. This method has better prediction results than traditional machine learning models, with a classification accuracy of 71.5% and an AUC value of 0.782 [11]. Li et al. proposed a student performance prediction model based on LR. This method first analyzes students' historical data through visualization, then divides the data set into a training set and test set, establishes the LR model for prediction, and scores the accuracy of the prediction results. The results show that the prediction accuracy of the optimized model is increased to 0.802 after removing irrelevant features, which can effectively predict student scores [12].

3.2. Prediction Methods Based On Deep Learning

Deep learning is a machine learning method based on characterizing data. It automatically learns complex patterns and features from large amounts of data by constructing neural network models with multiple levels. It has significant advantages in dealing with complex nonlinear problems. Chen et al. proposed a student performance prediction method based on BPNN. This method uses big data analysis to identify the key factors that affect grades and uses these factors to build a neural network model for prediction. The experimental results show that this method can accurately predict students' academic performance, effectively making up for the shortcomings of traditional qualitative predictions, and has high prediction accuracy and practical application value [1]. Liu et al. proposed an online learning performance prediction and evaluation model based on deep learning and data mining. The model relies on the online course platform to collect learner behavior data, uses the K-means algorithm to cluster analyze and screen key factors, uses distortion function to judge convergence, introduces GBRT to build the model, and uses root mean square error to describe accuracy. The result is that the root mean square error can be reduced by about 30%, and the accuracy is good [13].

4. Limitations and Future Prospects

4.1. Limitations

The prediction of students' academic performance relies on a large amount of historical data and real-life data. However, it faces many obstacles in the process of data acquisition. On the one hand, there are difficulties in effectively obtaining data due to various factors. For example, some schools or educational institutions are unwilling to provide data related to student performance due to privacy protection and other reasons, or some data records are incomplete and the format is not uniform,

making it difficult to integrate utilization. On the other hand, obtaining these data often requires high economic and time costs. For example, collecting long-term and comprehensive performance data from multiple schools and multiple grades may involve the cost of communication and coordination with numerous education departments and schools, as well as manual labor. These factors, such as the large amount of time it takes to organize data, limit the size and quality of the data set, which in turn affects the accuracy and reliability of grade prediction.

Most of the currently used student performance prediction methods are quantitative prediction methods, which are mainly based on past conditions to predict the future. However, student performance does not only depend on historical performance, but is also affected by many external factors, such as sudden changes in students' family environment, the impact of major social events on students' psychological states, and adjustments to school teaching methods. However, existing prediction models cannot often adapt to external factors and cannot flexibly adjust prediction logic and parameters promptly according to changes in these external factors. As a result, when the model faces complex and changeable real-life situations, the prediction results may be inconsistent. Large deviation makes it difficult to accurately reflect students' true academic performance development trends.

4.2. Prospects

In the future, various advanced learning methods can be used to alleviate the problem of insufficient data in the data set. For example, reinforcement learning, by setting up a reasonable reward mechanism, allows the model to continuously explore and optimize based on limited data, and mine more valuable information for performance prediction. Pre-training methods can use large-scale general data for preliminary training, so that the model can learn some common patterns and features, and then fine-tune it on a small-scale data set related to student performance, thereby improving the performance of the model under limited data. In addition, metric learning can help measure the similarities and other relationships between different data samples and better organize and utilize student performance-related data. Through the comprehensive application of these methods, it is expected that the limitations caused by insufficient data volume can be overcome to a certain extent and the quality of performance prediction can be improved.

Given the current lack of adaptive capabilities in models, future research should focus on developing predictive models with strong adaptive properties. This type of model should be able to perceive changes in various external influencing factors in real time and sensitively, and automatically and flexibly adjust the various parameters involved in the prediction accordingly, to ensure that the prediction results can be more in line with the actual trend of changes in student performance. For example, when it is monitored that there are major adjustments in the allocation of educational resources in the region or that individual students encounter special life events, the model can quickly capture these changes, optimize and adjust the internal parameters accordingly, and output more accurate and reliable performance prediction results. By building such an adaptive model, the accuracy and timeliness of grade predictions will be greatly improved, making them more in line with the needs of complex and changing real-world educational scenarios.

Actively carrying out interdisciplinary cooperation is of great significance to improving the student academic performance prediction model. Multidisciplinary knowledge such as education, psychology, and sociology can be integrated into model construction and optimization. Educational knowledge helps analyze the impact of educational factors on academic performance and provides practical variables and rule bases for the model. Psychological knowledge helps to understand the fluctuation patterns of students' learning behaviors and grades at different stages and psychological states and consider the role of psychological factors in grades. Sociological knowledge can analyze the impact of the social environment on learning motivation and its effect on academic performance from a macro perspective. Through interdisciplinary collaboration, the people can fully grasp the various factors that affect academic performance, thoroughly understand the relevant influencing factors,

improve the prediction effect of the model, and enable it to better serve educational and teaching practice.

5. Conclusion

This article focuses on the prediction of student academic performance, analyzes the influencing factors and methods, and explores the limitations and prospects, providing a comprehensive perspective for improving the effectiveness of performance prediction. In summary, student performance is affected by many factors, including mental health, gender differences, learning motivation, time management tendencies, family education methods, and teacher gender. These factors work together to form a complex mechanism that influences performance. In terms of prediction methods, machine learning, and deep learning methods are diverse and effective. Related models such as XGBoost, LR, BPNN, etc. have good prediction effects and provide a scientific basis for predicting student grades. However, there are currently problems such as difficulty and high cost in obtaining data sets, incomplete data, and the lack of adaptive ability of prediction models to external factors, which can easily lead to prediction bias. Looking into the future, improving data set utilization, developing adaptive models, and conducting interdisciplinary collaboration through reinforcement learning will help improve the accuracy of grade predictions, better serve educational and teaching practices, and promote improvements in education quality.

References

- [1] Cheng Q., Peng W. Short-term load forecasting model and method for power system based on complementation of neural network and fuzzy logic. *Automation Applications*, 2024, 17: 24-26+31.
- [2] Li K. Research on power load forecasting based on improved BP neural network. *Computer Age*, 2023, 12: 220-223. DOI: 10.16644/j.cnki.cn33-1094/tp.2023.12.049.
- [3] Lang X., Luo J., Shi X., Yan Y. The impact of students' mental health on their academic performance. *Science & Education Digest*, 2023, 1: 179-184. DOI: 10.16871/j.cnki.kjwh.2023.01.044.2.
- [4] Jiang S., Jiang S., Zhang J. The exploration of multidimensional factors influencing university students' academic performance—Based on interpretable machine learning models. *Journal of Liaoning Normal University (Social Sciences Edition)*, 2024, 4: 97-105. DOI: 10.16216/j.cnki.lsxwbk.202404097.
- [5] Liu Y., Cai M. The impact of university students' academic emotions on academic performance. *Intelligence*, 2024, 13: 120-123.
- [6] Wang H., Xu M., Qi X. The correlation between university students' academic performance and learning motivation. *China Educational Technology Equipment*, 2022, 14: 46-48.
- [7] Wang S., Zhong S., Li J., Wang Z. The impact of high school students' time management tendencies on academic performance: The mediating role of learning engagement. *Psychological Health Education for Primary and Secondary School Students*, 2023, 5: 14-20.
- [8] Zhao Y. The influence of family factors on high school students' academic performance and intervention strategies. *Contemporary Family Education*, 2020, 8: 4.
- [9] Li X. The impact of family upbringing on academic performance: The mediating role of subjective well-being and achievement motivation. Master's Thesis, Henan University, 2022.
- [10] Liu Z., Guo R., Tian M. The impact of teacher gender on students' academic performance—A study from the perspective of teacher-student gender matching. *Education and Economy*, 2023, 3: 64-74.
- [11] Zhang Z., Chu Z. Research on student performance prediction based on XGBoost and SHAP feature analysis. *Information Research*, 2024, 3: 34-40.
- [12] Li G., Zhang M., Kang R. Research on student performance prediction model based on logistic regression. *China Information Technology Education*, 2023, 15: 77-80.
- [13] Liu J., You Z. Design of an online learning prediction and evaluation model based on deep learning and data mining. *Electronic Design Engineering*, 2023, 15: 131-134+139. DOI: 10.14022/j.issn1674-6236.2023.15.027.