Construction and Practical Exploration of a Precise Management Model for Student Aid: A Multi-Dimensional Data-Driven Methodology

Chao Wang*, Xiao Lv, Yang Sun, Zhe Liu

Zhuhai College of Science and Technology, Guangdong Province, China

*Email: 844766468@qq.com

Abstract. This article aims to explore how to construct a precise management model for student funding through a multi-dimensional data-driven methodology, in order to improve funding efficiency and accuracy. This article summarizes the shortcomings of previous research and introduces a new method based on Gradient Boosting Decision Tree (GBDT). Based on the GBDT algorithm, the model achieved an accuracy of 92% and an F1 score of 90.5%. From the data conclusion, it can be seen that the GBDT model is effective in predicting student aid, and it also emphasizes the need to pay special attention to specific features and data quality before optimizing and deploying the model in the future.

Keywords: Student Aid Management; Multidimensional Data Analysis; Gradient Boosting Decision Tree; Model Robustness

1 Introduction

With the popularization of higher education and the widening economic gap, student funding has become an important means to ensure educational equity. However, due to the lack of comprehensive consideration of multidimensional information about students, traditional funding management systems often cannot accurately provide funding, resulting in low efficiency in resource allocation. This article adopts a multidimensional data-driven approach and proposes a predictive model that can accurately evaluate student funding needs.

This article first introduces the background and importance of the research, and then introduces the data integration, feature extraction and selection, and model development in the methodology stage. In the experimental stage, the effectiveness and robustness of the GBDT model were verified through three experiments, and a comprehensive evaluation of the GBDT model was conducted. The final conclusion section discusses the limitations of t.

D. Rad et al. (eds.), *Proceedings of the 2024 5th International Conference on Mental Health, Education and Human Development (MHEHD 2024)*, Advances in Social Science, Education and Humanities Research 857,
https://doi.org/10.2991/978-2-38476-271-2_48

2 Related Works

In past research, many scholars have attempted to address these issues through different methods. For example, Chen Yiting conducted a survey using 1811 eighth grade students and their parents from the "Chinese School Curriculum Teaching Survey Project" as samples. She explored the relationship between family socioeconomic status and math grades, and examined the mechanisms of parent-child communication and academic self-efficacy in it [1]. Wang Min et al. investigated the physical activity levels of high school students in Beijing and explored the relationship between family socioeconomic status and their academic performance [2]. Hascot M aims to test the theoretical expectation of linking family socioeconomic status and parental expectations in the family socialization model [3]. Zhao J's research shows that there is a significant gap in academic performance between urban and rural students in China [4]. Marley S C's research investigated how social support from family and peers for academic performance affects students' learning motivation [5]. Nishina et al.'s study investigated the mediating role of different types of parental involvement in adolescent academic performance [6]. However, these studies often overlook the importance of integrating multidimensional data, leading to incomplete and inaccurate optimization of management models.

The comprehensive application of multidimensional data is particularly crucial in addressing the shortcomings of existing research. For example, Liu J utilized data mining techniques to achieve dynamic management of psychological warning data [7]. The above research did not fully cover all the key factors that affect the effectiveness of student funding. This article will use the GBDT algorithm to evaluate the need for student funding.

3 Methods

3.1 Data Integration

In this study, we integrated information from different data sources. Specifically, we obtained detailed student information from a school management system and mental health support institution for training and testing models such as GBDT. To ensure data consistency and comparability, we have cleaned and preprocessed the data. Meanwhile, we have meticulously integrated data formats from multiple platforms[8-9].

In data processing, we use advanced techniques to transform raw data to fit the GBDT model. By normalizing and adjusting the data range, ensure that the data is suitable for model use. In the feature selection stage, we select the features from the processed data that can best predict student funding needs[10]. The importance of features can be represented by formula (1):

$$\mathrm{Importance}(f_i) = \sum_{t \in T} I(f_i, t) \cdot p(t) \tag{1}$$

In formula (1), f_i represents a specific feature, T is the set of nodes in the decision tree, $I(f_i, t)$ represents the information gain at node t, and $p(t)$ represents the proportion of samples at node t. Through feature extraction and selection operations, we have constructed a precise and efficient GBDT model, which enables the student aid management model to more accurately target and assist students in need of support.

3.2 Model Development

In this study, we decided to use the GBDT model as the primary tool for prediction. The reason for choosing GBDT is its excellent processing ability for complex and multi-dimensional data. Another advantage is that the GBDT model can effectively handle the nonlinear relationships between data features. The core algorithm of GBDT can be represented by formula (2):

$$L(y, F(x)) = \sum_{i=1}^{n} l(y_i, F(x_i)) \tag{2}$$

In formula (2), L is the total loss function, y is the true value, $F(x)$ is the predicted output of the model, l is the loss of a single data point, and n is the number of samples. By adjusting the learning rate, we can find the optimal GBDT model configuration to achieve the highest prediction accuracy. The adjusted learning rate can be expressed by formula (3):

$$F_m(x) = F_{m-1}(x) + \eta \cdot h_m(x) \tag{3}$$

In formula (3), $F_m(x)$ is the model in step m, $h_m(x)$ is the newly added tree in step m, and η is the learning rate. The final step in GBDT model development is model validation. By using an independent test set to evaluate the model, it can be ensured that the GBDT model achieves the expected performance in practical applications.

4 Results and Discussion

4.1 Model Prediction Performance Experiment

In the model prediction performance experiment, the performance of GBDT and other models in predicting student funding needs was evaluated. In the experiment, focusing on the performance of the four models in three key performance indicators: accuracy, recall, and F1 score, and draw these performance indicators into graphs to visually demonstrate the performance of the model.

From Figure 1, it can be seen that the GBDT model demonstrated the best performance on the dataset, with an accuracy of 92%, a recall rate of 89%, and an F1 score of 90.5%. The accuracy of the Random Forest (RF) model is 90%, the recall rate is 88%, and the F1 score is 89%. The indicator values of SVM algorithm and LR algorithm are relatively low. From the data results, it can be seen that the overall performance of the GBDT model is excellent, which is helpful in improving the accuracy and decision-making efficiency of the student aid management system. The specific situation is shown in Figure 1:

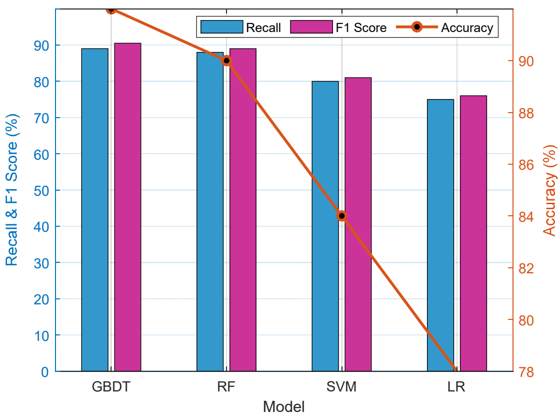

Fig. 1. Model prediction performance evaluation

4.2 Model Robustness Testing and Feature Analysis

In the robustness testing experiment of the model, the sensitivity of the GBDT model to noise in processing student funding data was evaluated. In the experiment, different levels of noise were introduced to test the performance of the GBDT model in the face of changes in data quality. We set five different noise levels of 0, 0.01, 0.05, 0.1, and 0.2 in the experiment, and plotted the accuracy of the GBDT model at different noise levels.

From Figure 2, it can be seen that as the noise level increases from 0 to 0.2, the accuracy of the model gradually decreases from the original 95% to 85%. Especially when the noise level exceeds 0.1, the decreasing trend of accuracy is more pronounced. From the data conclusion, it can be seen that the GBDT model performs well in low noise environments, but its adaptability to high noise situations is weak, highlighting the need for strict control of data quality in practical applications to improve its robustness. The specific data situation is shown in Figure 2:

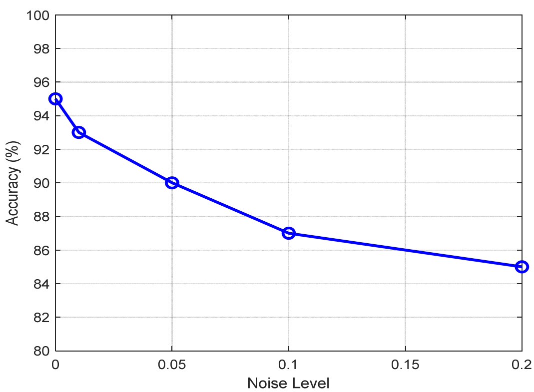

Fig. 2. Model robustness evaluation

The specific data details of feature importance are shown in Table 1:

Table 1. Details of feature importance analysis

Family Income	GPA	Attendance Rate	Parental Education
0.4252	0.2387	0.7744	0.1999
0.9710	0.8069	0.9231	0.6822
0.1337	0.7758	0.5954	0.9902

From the data in Table 1, it can be seen that in the three tests, the average importance score of household income is about 0.5, while the average GPA score is about 0.3. From the data conclusions, it can be seen that these features play a crucial role in predicting student funding needs, and we should prioritize these data features when further developing GBDT models.

5 Conclusion

In this article, we adopted a multidimensional data-driven approach and successfully developed a system that uses GBDT model technology to predict student funding needs. The experimental results in the experimental stage show that the GBDT model performs significantly better in prediction accuracy than other algorithm models, especially in dealing with complex scenarios containing diverse data such as economic status, academic performance, and mental health. In addition, through feature importance analysis, we have revealed to policy makers the key factors that play a decisive role in the student funding process. Although some achievements have been made, the GBDT model still heavily relies on data quality, which may limit its popularity in wider practical applications. Future work can consider exploring more data sources and adopting advanced algorithms to improve the robustness and adaptability of the model. Meanwhile, enhancing the interpretability of the GBDT model will also help policy makers and education managers better understand and trust the predicted results, thereby promoting the practical application of the model in the field of education funding.

Acknowledgement

This paper is the research result of the project of Innovation Ability Cultivation Project (Funding education) of Zhuhai College of Science and Technology, "Research on the precise funding of students in private universities". Project Number: 023XJCQYR004

References

1. Chen Yiting, Yang Xiangdong. The impact of family socioeconomic status on mathematical performance: a chain mediation model of parent-child communication and academic self-efficacy [J]. Applied Psychology, 2020, 26 (1): 66-74.

2. Wang Min, Li Juan, He Zihong, et al. The relationship between physical activity level and academic performance of high school students in Beijing: an analysis based on family socio-economic conditions [J]. China Sports Science and Technology, 2021, 57 (5): 7-9.
3. Hascot M, Giaconi V, Jamain L. Family socioeconomic status and parental expectations affect mathematics achievement in a national sample of Chilean students: [J]. International Journal of Behavioral Development, 2021, 45(2):122-132.
4. Zhao J, Bodovski K.The Effect of Family Background and Parental Expectations on Eighth Graders' Academic Achievement in Rural and Urban China[J].Frontiers of Education in China, 2020, 15(4):647-677.
5. Marley S C, Wilcox M J.Do family and peer academic social supports predict academic motivations and achievement of first-year college students?[J].Journal of Applied Research in Higher Education, 2022, 14(3):958-973.
6. Nishina A, Witkow M, Zhang F, et al.Family Socioeconomic Status, Parental Involvement, and Academic Achievement: The Moderating Role of Adolescents' Subjective Social Mobility:[J].The Journal of Early Adolescence, 2021, 41(9):1425-1454.
7. Liu J, Shi G, Zhou J, et al.Prediction of College Students' Psychological Crisis Based on Data Mining[J].Mobile Information Systems, 2021, 2021(23):1-7.
8. Li Yongming. A Psychological Disorder Prediction Model Based on Data Mining Technology [J]. Application of Microcomputers, 2021, 37 (11): 161-164.
9. Tian Wenxiang. Research on Academic Early Warning of College Students Based on Outlier Data Mining Technology [J]. Heilongjiang Science, 2021, 012 (007): 54-56.
10. Ye Yuping. Research on the Application of Data Mining Technology in the Academic Early Warning System of Vocational Colleges [J]. Computer and Information Technology, 2023, 31 (5): 94-96.