Journal of Contemporary Educational Research, 2025, Volume 9, Issue 5
http://ojs.bbwpublisher.com/index.php/JCER
ISSN Online: 2208-8474
ISSN Print: 2208-8466

Analysis of Influencing Factors of Academic Warning in Higher Vocational Colleges Based on the Importance of Machine Learning Features and Paths to Improve Learning Ability

Meimei Huang[1], Lei Zhang[1,2]*, Xifeng Fan[1]

[1]Nanning Normal University, Nanning 530001, Guangxi, China
[2]Guangxi Police College, Nanning 530028, Guangxi, China

Author to whom correspondence should be addressed.

Abstract: The traditional academic warning methods for students in higher vocational colleges are relatively backward, single, and have many influencing factors, which have a limited effect on improving their learning ability. A data set was established by collecting academic warning data of students in a certain university. The importance of the school, major, grade, and warning level for the students was analyzed using the Pearson correlation coefficient, random forest variable importance, and permutation importance. It was found that the characteristic of the major has a great impact on the academic warning level. Countermeasures such as dynamic adjustment of majors, reform of cognitive adaptation of courses, full-cycle academic support, and data-driven precise intervention were proposed to provide theoretical support and practical paths for universities to improve the efficiency of academic warning and enhance students' learning ability.

Keywords: Academic warning; Pearson correlation coefficient; Random forest variable importance; Permutation importance

Online publication: June 5, 2025

1. Introduction

With the popularization of higher education and the promotion of the policy of expanding enrollment, the source structure of students in higher vocational colleges has become increasingly diversified. Some students have fallen into academic difficulties due to a weak academic foundation, a lack of self-control, and environmental interference. The *Outline of the National Medium- and Long-Term Education Reform and Development Plan (2010–2020)* clearly requires the establishment of a support mechanism for students with learning difficulties [1], highlighting the urgency of building an academic early warning mechanism. As an innovative means of teaching management, academic early warning helps students get back on track through dynamic monitoring and graded intervention [2], and the penetration of big data technology provides a new path for accurately identifying early warning factors.

In recent years, machine learning (ML) technology has shown significant potential in biology, materials science, education, and other fields. Jia and Sun [3] established a model of mutant protein amino acid sequence and spectral maximum absorption wavelength through ML and obtained a prediction coefficient R^2 of 0.944. He *et al.* [4] used feature quantity screening to identify key alloy factors affecting alloy properties and established an alloy prediction model to achieve rapid design of alloy composition. Wang [5] constructed an academic warning model through ensemble learning, with an accuracy rate of 80.96%; variable importance analysis verified its ability to explain complex associations in fields such as biology and materials science. In the field of education, existing research focuses on undergraduate education, and quantitative analysis of warning factors for the characteristics of higher vocational and technical college students is still insufficient.

Based on the academic data of the former Guangxi College of Education from 2016 to 2021, this study integrates ML variable importance evaluation methods to analyze the influencing factors of academic warning for higher vocational and technical colleges. Compared with traditional statistical methods, this technology can capture nonlinear relationships and quantify the contribution of each factor, helping to improve the effectiveness of teaching management and providing a certain scientific basis and practical foundation for improving learning ability.

2. Evaluation indicators of factors affecting academic warning

The Pearson correlation coefficient (PCC) quantifies the strength and direction of the linear association between two variables by measuring the ratio of the covariance of two variables to their respective standard deviations. The value (r) range of the correlation coefficient is between -1 and 1. When the two variables are positively correlated, r is between 0 and 1; when they are negatively correlated, r is between -1 and 0; if $r = 0$, it indicates that there is no linear correlation between x and y, that is, linear independence. The larger the absolute value of the correlation coefficient, the stronger the correlation; the closer the absolute value of the correlation coefficient is to 0, the weaker the correlation [6]. This coefficient eliminates the influence of dimension through standardization and is suitable for evaluating the closeness of linear patterns between variables, but it cannot capture nonlinear relationships or causal relationships.

Random forests variable importance (RFI) [7] is based on the Gini impurity reduction method. They measure the contribution of features to model prediction by calculating the average change in the Gini coefficient when the feature is split across all decision tree nodes. Their advantages include resistance to overfitting, high stability, and suitability for nonlinear relationship modeling [8].

Permutation importance (PI) [9] evaluates the contribution of a feature to model prediction by destroying the original association between the feature and the target variable. Specifically, this method assumes that if a feature is critical to the model, the model's prediction performance will drop significantly when its value is randomly perturbed; conversely, if the performance does not change much, it means that the feature is less important. The benefit of this technique is that it does not depend on the model and can be calculated multiple times with different permutations of the feature.

3. Analysis of factors affecting academic warning

The data set consists of the grade warning information of the four schools of the former Guangxi College of

Education from the second semester of the 2016–2017 school year to the first semester of the 2021–2022 school year. Only the grade warning data in the academic warning information is analyzed. The academic warning mentioned below refers to the grade warning. The data set contains 2,022 samples, and the feature variables are warning level, school, major, and grade. The schools are the School of Mathematics & Information Sciences, the School of Educational Sciences, the School of Languages & Literature, and the School of Arts. The warning levels are divided into yellow warning, orange warning, and red warning from low to high according to the number of students who fail the exams, covering three grades of higher vocational college students. The following analyzes the relationship between feature variables and warning levels through traditional statistical methods and ML variable importance evaluation.

3.1. Comparison of methods and differences in explanatory power

Figure 1 shows the PCCs between the warning level and the characteristic variables. It can be seen that the PCC between the major and the school is 0.394, close to 0.5, indicating that there is a certain correlation between the two, which is basically consistent with the division principle of majors and schools. However, the linear correlation between each characteristic variable and the warning level is lower than 0.5, and the linear model is difficult to capture complex relationships. In view of this, the RFI and PI methods based on machine learning technology are used for analysis, and the results are shown in **Figure 2**.

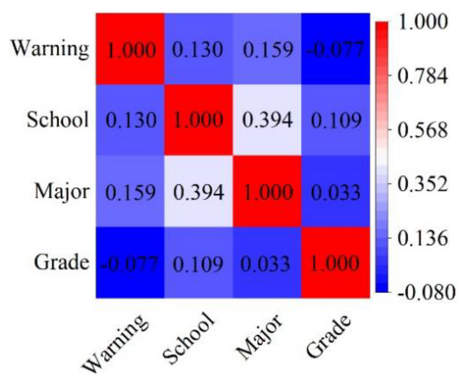

Figure 1. PCCs between warning level and characteristic variables

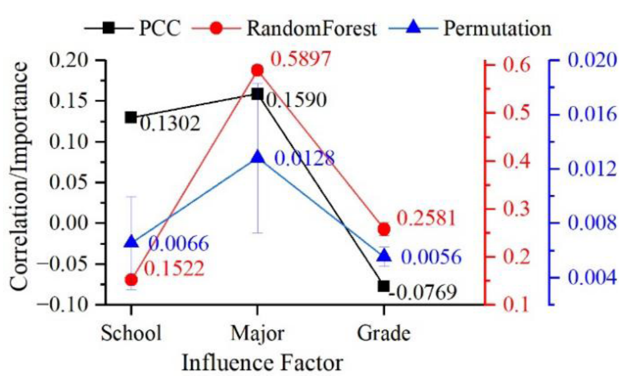

Figure 2. Evaluation of the importance of warning level characteristics

The larger the value of feature importance evaluation, the greater the impact of the feature on the student warning level. As shown in Figure 2, the importance score of the major feature (0.5897) is significantly higher than other variables, which verifies the advantage of variable importance analysis in processing highdimensional heterogeneous educational data. It is worth noting that the consistency of the results of random forest and permutation importance methods (major > grade > school) shows that the nonlinear model has a higher stability in feature sorting.

3.2. Analysis of the characteristics of the major

For art majors, students majoring in music, dance, etc., have a high failure rate in theoretical courses due to long-term art training (average professional training time per day ≥ 4 hours), and their awareness of rules is weak [10]. There is a general cognitive bias of "focusing on skills rather than theory," and there is a structural imbalance in learning input.

For the School of Educational Sciences, the number of students majoring in primary school general education accounts for the highest proportion. This major is a five-year junior high school-based higher vocational education. The high school management method is fully adopted in terms of academic style construction, teacher-student communication, etc., so that students' learning status is significantly better than that of students in other majors. They have stronger learning ability, are more likely to accept teacher guidance, have a low failure rate, and naturally have fewer academic warnings. However, we need to be vigilant against excessive management that weakens students' independent learning ability.

For science and engineering majors, the knowledge density of the courses in the School of Mathematics and Information Sciences is high, and there is a strong dependence between courses. There are a lot of calculations and knowledge points to remember. The combination of these two aspects leads to a significant increase in the difficulty of the courses, which easily leads to an increase in the number of failures in professional courses.

For liberal arts majors, students in the School of Language and Literature need to cope with both theoretical recitation and practical assessments, such as teacher qualification certificates, and their cognitive load index is significantly higher than that of other majors. This has resulted in the college's failure rate being second only to that of the School of Arts, and showing a "polarization" feature.

4. Optimization strategies for academic warning and paths to improve learning ability

Traditional post-event warnings have poor timeliness and need to shift to pre-event prediction and intervention, and improve academic performance and learning ability with the concept of "student-oriented." The specific countermeasures are as follows.

Building a three-dimensional model of "regional industry demand–professional competitiveness–warning rate," dynamically adjusting the course structure of high-warning majors such as arts, piloting the mechanism of replacing general credits with art competitions, and promoting the establishment of interdisciplinary integration courses in liberal arts majors, transforming recitation tasks into project practice, and alleviating the conflict between professional training and theoretical learning.

Carrying out cognitive adaptation reform of the curriculum system, building a knowledge map navigation system for science and engineering, and monitoring the risk of knowledge gaps in real time; implementing "theory + practice" modular assessment for liberal arts, allowing students to choose the difficulty of tasks independently, and balance the fairness and challenge of learning.

Establishing a full-cycle academic support mechanism with three-level intervention of "entrance assessment–behavior monitoring during the semester–targeted counseling before the exam," identifying high-risk students through the "Academic Adaptability Scale," and dynamically adjusting the warning threshold based on self-study time and test data; building a student portrait platform that integrates academic affairs, behavior, and social data to achieve automatic warning and precise resource push.

Integrating multi-source data, such as the teaching system and course platform, to build an early warning information management platform. Generating personalized learning plans through time management diagnosis and cognitive style assessment to cultivate self-regulation ability, gradually improve learning ability, and reduce the repeated warning rate. Studies by Zhao and Zhao [11], Lv and Xia [12] have confirmed that multidimensional data fusion can significantly improve the explanation of early warnings and the effectiveness of interventions.

5. Conclusion

This study established a data set for academic early warning information and revealed the deep mechanism of academic early warning in higher vocational colleges through PCC, RFI, and PI analysis methods. Data analysis found that machine learning methods broke through the limitations of traditional statistics and effectively captured nonlinear relationships. The majority of students in the academic early warning data was the core driving factor of academic early warning, with an importance score of 0.5897. Among them, art majors had the highest failure rate due to "focusing on skills and neglecting theory," while the School of Education Science controlled the warning rate to the lowest through high school-style refined management of students in general classes, but it is also necessary to be vigilant about the suppression of student autonomy by management rigidity. Based on this, countermeasures such as dynamic adjustment of majors, reform of cognitive adaptation of courses (such as knowledge graph navigation, modular assessment, etc.), full-cycle academic support of the three-level intervention system and data-driven precise intervention were proposed to form a closed loop of "problem identification–technology empowerment–strategy iteration" and promote the transformation of academic early warning from "after-the-fact remediation" to "pre-emptive prevention." In the future, it is necessary to conduct an in-depth analysis on data richness, combining multi-center data with dynamic models, etc. This study provides guidance for colleges and universities that offer higher vocational and technical majors to improve academic warning methods and enhance learning ability, and provides a data-driven decision-making framework for optimizing teaching management, which has important practical value for implementing the goal of "precision education."

Funding

This work was financially supported by the Basic Ability Improvement Project of Young and Middle-Aged Teachers in Colleges and Universities of Guangxi (2022KY1922, 2021KY1938).

Disclosure statement

The authors declare no conflict of interest.

References

[1] Hu JT, 2010, Speech at the National Education Work Conference. People's Daily, 9(09): 2.

[2] Zhang BZ, Li H, Hou HX, et al., 2021, Design of Early Warning System of Academic Performance at Colleges

and Universities Based on Database and Machine Learning Technology. Education of Chinese Medicine, 40(3): 63–67.

[3] Jia LL, Sun TT, 2022, A Machine Learning Study on Gloeobacter violaceus Rhodopsin Spectral Properties. Journal of Zhejiang University (Science Edition), 49(3): 280–286.

[4] He XQ, Fu HD, Zhang HT, et al., 2022, Machine Learning Aided Rapid Discovery of High-Performance Silver Alloy Electrical Contact Materials. Acta Metallurgica Sinica, 58(6): 816–826.

[5] Wang GT, 2021, Academic Early Warning Research Based on Machine Learning, dissertation, Xi'an Shiyou University, 1–52.

[6] Xue X, Zhou J, 2017, A Hybrid Fault Diagnosis Approach Based on Mixed-Domain State Features for Rotating Machinery. ISA Transactions, 66: 284–295.

[7] Lee WM, 2019, Python Machine Learning, John Wiley & Sons, New Jersey.

[8] Li ZG, 2014, Several Research on Random Forest Improvement, dissertation, Xiamen University, 1–46.

[9] Radivojac P, Obradovic Z, Dunker AK, et al., 2004, Feature Selection Filters Based on the Permutation Test, European Conference on Machine Learning, Springer Berlin Heidelberg, Berlin, Heidelberg, 334–346.

[10] Zhang LJ, Yao Qi, 2019, Research on the Characteristics and Educational Management Strategies of Higher Vocational Art Students—Taking Jiangsu Information Vocational and Technical College as an Example. Think Tank Era, (10): 164–165.

[11] Zhao XM, Zhao KY, 2018, A Design Research About Big Data of Education Applied to Academic Warning: An Example of the Early-Warning System for Children Left Behind in Rural Areas. Research in Educational Development, 38(12): 64–71.

[12] Lv L, Xia ZH, 2020, Cause Analysis and Countermeasure Research of College Academic Warning Based on Random Forest Algorithm. Journal of Nanchang Institute of Technology, 39(6): 81–86.

Publisher's note

Bio-Byword Scientific Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.