Data-Driven Intelligent Model for Learning Analytics: Bridging Secondary and Undergraduate Vocational Education

Yikun Xu, Sha Xian, and Linjie Xiang*
School of Culture and Management, Chengdu Vocational University of the Arts, Chengdu, China
Email: 18780168819@163.com (L.X)
*Corresponding author

***Abstract*—This article undertakes a rigorous examination of the learning analytics model, with a particular focus on its application in the context of clustering algorithms. The integration of students' academic performance data enables the implementation of hierarchical tagging and trend tracking of learning situations. The model's initial phase involves the construction of a data layer, integrated with machine learning mechanisms, which are capable of efficiently identifying potential student groups that are at risk in terms of academic performance. Subsequently, an analytics layer is developed using Power BI, and the designed dynamic signage enables educational administrators to comprehensively understand the learning situation of students in the transition stage in real time. Ultimately, the DeepSeek API is integrated into the application layer, thereby providing precise teaching intervention suggestions. The model for teaching reform was successfully implemented during a one-year study encompassing over 1,500 students from three institutions. It provides a management framework for the transition from secondary to undergraduate vocational education.**

***Keywords*—secondary vocational education, undergraduate vocational education, learning analytics model, clustering algorithms**

I. INTRODUCTION

The articulation between vocational education in secondary and undergraduate schools has emerged as a key issue in China. Challenges related to articulation and integration within the nation's education system are now a major concern. There are discrepancies between secondary and undergraduate vocational education, resulting in secondary graduates struggling to switch to vocational training. These challenges often go unnoticed by traditional management frameworks.

These challenges hinder students' academic progress and constrain the quality and efficiency of vocational training. Previous literature has explored this, but not proposed effective solutions, especially in articulation between secondary and undergraduate vocational education. This study aims to address this through academic articulation analytics. The research employs a scientific approach and builds a numerical intelligence model, exploring ways to effectively articulate learning between secondary and vocational levels, with the aim of improving vocational training. Additionally, the study will furnish valuable references and materials for other educational articulation, contributing to both theoretical and practical significance.

II. RELATED WORKS

In recent years, secondary bridging undergraduate vocational education has become a subject of considerable interest within academic discourse, recognized as a pivotal element in the ongoing modernization of the vocational education system. The extant research in this area has focused on the design of curriculum systems, the innovation of evaluation mechanisms, and the development of data-driven analytics models. The objective of these endeavours is to address the issue of disconnection between secondary and undergraduate vocational education, thereby promoting the continuous cultivation of talent.

In the connection of curriculum system, the research generally emphasizes the collaborative construction of professional chain and industrial chain. For instance, Li and Suman [1] proposed a seven-year undergraduate curriculum framework via vocational groups to address curriculum misalignment. Feng, Li, and Pan [2] restructured tourism management programmes to seamlessly connect theoretical and practical modules.

From the perspective of the application of educational data mining and machine learning technology, existing studies have used a variety of linear and nonlinear models to analyze educational data. Li, Duan, and Yue [3] used random forest-Adaboost algorithm to improve the diagnostic accuracy of educational information. Zhang [4] supported dynamic evaluation through multi-modal data (behavior, emotion, and cognition).

III. MODEL CONSTRUCTION

A. Model Framework

Recent progress in VET reform and big data has been made in policy articulation, optimization and dynamic

Manuscript received March 3, 2025; accepted April 7, 2025; published May 22, 2025.

modelling, but data silos, long-cycle analysis and insufficient sentiment intelligence remain challenges [5].

In terms of methodology, following the excellent experience of machine learning, the heterogeneous performance data from multiple sources are downgraded and mapped into standard learning labels, and clustering algorithms are introduced to realize the accurate classification and growth path visualization of student groups. This process not only breaks through the data barriers, but also encourages teachers to dynamically adjust the difficulty of courses and teaching strategies based on the clustering results through the “data-teaching” dual-wheel-drive mechanism, forming a closed-loop governance system of “monitoring-diagnosis-improvement”.

This paper proposes a collaborative governance framework for addressing these issues. The methodology involves mapping multi-origin heterogeneous achievement data into standard chemical sentiment labels using machine learning, then clustering to visualize student growth. As illustrated in Fig. 1, the model architecture comprises three layers: data, analytics, and application.

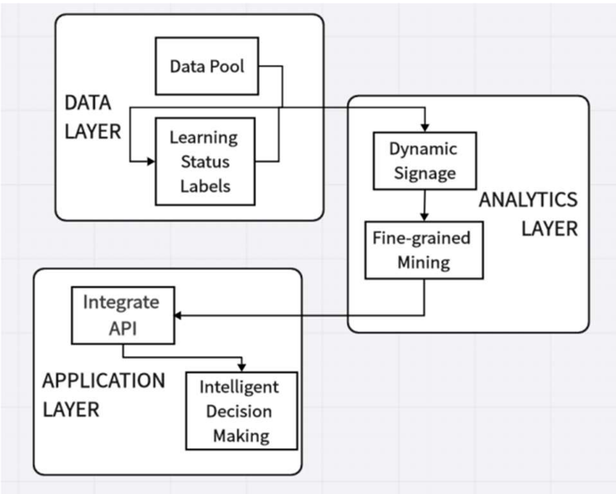

Fig. 1. The learning analytics framework for secondary-undergraduate vocational education.

B. *Data Layer: Data Pool and Algorithm-Driven Learning Status Labels*

The data repository of the model encompasses the achievement data of all students enrolled in a range of secondary and vocational undergraduate programmes across three vocational colleges. The data was optimized to retain 12 key attributes. The process created a time-series panel data pool of 20,297 valid records.

This paper proposes a novel approach to the classification of learning status labels using students’ overall assessment scores for cluster analysis [6]. The elbow rule is utilized to determine the optimal number of clusters, which is set at K = 5. Following the selection of the optimal number of clusters, machine learning is implemented through code to categorize students with similar characteristics into one group and to automatically calculate the cluster centre mean.

TABLE I. THE OUTCOMES OF THE CLUSTER ANALYSIS

Label	Quantity	Clustering Centre Mean
Dormant	2691	44.8
Risk	406	64.1
Balanced	5181	74.9
Robust	6758	83.4
Excellence	5031	92.1

The outcomes of the cluster analysis are presented in Table I. Following the implementation of machine learning and model training, five cluster centres are ultimately generated, designated as Dormant, Risk, Balance, Robust, and Excellence labels, respectively, in conjunction with the students’ psychological and behavioural characteristics [7].

The term “Dormant label” is employed to denote a student’s failing overall assessment grade in a course. Such a grade is typically indicative of a chronic lack of motivation, interest, or appropriate study methods. The “Risk label” refers to a student who is barely passing the overall assessment and whose academic performance is inconsistent, and who may be improvising or relying on unannounced revision. The “Balanced label” is allocated to students whose overall assessment scores are in the lower-middle range, and whose academic performance lacks the stability to ensure timely completion of learning tasks. The “Robust label” is allocated to students whose overall assessed performance is in the upper-middle range, exhibiting academic performance that is stable but not particularly noteworthy. The “Excellence label” is reserved for students whose overall assessment scores are at the top level, who have demonstrated exceptional academic performance, a profound understanding of the course content, and the capacity for innovative thinking.

C. *Analytics Layer: Dynamic Signage and Fine-Grained Mining*

Following the construction and refinement of the data layer, Python will be utilized to develop automated export scripts for the data pool. These scripts will then be delivered to the Power BI platform [8]. In addition, the development of a dynamic signage will proceed, with the objective of facilitating real-time feedback and in-depth insights into the state of learning across school years.

Dynamic signage can extract secondary and undergraduate articulation learning data from the data pool. It employs a hierarchical filtering architecture, with the table header featuring a multi-level slicer, the grade dimension presented in a tile matrix, the class displayed in a vertical drop-down menu, combined with gender and semester-assisted filtering, and the DAX function facilitating real-time synchronization of the data with a response time of less than 0.5 seconds. The core area focuses on the mean values of usual, final, and overall assessment grades through card charts to assess the effectiveness of the teaching evaluation system and compare the differences in the distribution of grades among gender groups. The multi-dimensional analysis module contains the following: First, percentage stacked bar charts are used to reveal gender differences in learning labels. Second, ribbon charts are used to demonstrate the overall distribution of learning labels. Third, tree clustering is used to present class heterogeneity and support the location of classes with excellence rates that significantly deviate from their peers for traceability analysis. The system provides a comprehensive view of teaching data by incorporating rapid screening and in-depth correlation analysis, facilitating multi-faceted visualization support for the assessment of learning status.

Accounting and Big Data and Financial Management (Secondary and Undergraduate) can be considered. Core course pairs are screened and analyzed. Intermediate Tax Accounting and Intelligent Filing; Undergraduate Intelligent Costing and Management. The percentage of excellence labels for the secondary level is 18.2%, while for the undergraduate level, it is 9.7%. This suggests a significant articulation fault, indicating that the knowledge of the cost management module is inherently difficult and extensive, and that it is necessary to strengthen the foundation of management accounting at the intermediate level, exposing the problem of irrational design of the course content articulation gradient.

D. *Integrate the API to Support Intelligent Decision Making*

The application layer of the model is designed to be coupled with the DeepSeek education model through the Power BI embedded API interface with a view to building an AI intelligent decision-making assistant [9]. In the front-end of Power BI, the demand for learning situation analysis can be input through natural language, and the system is based on sending structured query requests to the DeepSeek cloud, such as course codes, learning status labels, etc. Subsequently, DeepSeek conducts semantic parsing and attributional reasoning on the cross-section learning situation data, generating dynamic decision-making suggestions and returning statistical and visualization results.

In the application layer, teachers are able to obtain personalized intervention plans in real time through natural language Q&A, and the system automatically invokes the reinforcement learning module of DeepSeek to recommend optimal decision-making in combination with historical data. This enhances the efficiency of teaching management by tailoring the teaching to the students' needs.

The experiment is as follows:

In order to provide a comprehensive analysis of the reasons and intervention strategies for the 23% increase in dormant tags after a student's transition to undergraduate school, the following questions are hereby posed to the reader:

The intelligent decision recommendation that emerges from a thorough data analysis is as follows: the student exhibits a disconnection from a Financial Big Data Analytics course, a potential deficiency in Python fundamentals at their mid-career level, and a concurrent decline in group project participation. In order to address these challenges, it is recommended that a 12-lesson Python Fundamentals compensation micro-course be embedded and a cross-grade project mentorship be initiated.

IV. Results

Following implementation in the relevant academic disciplines at the experimental institutions, the model was found to optimize the teaching process evaluation mechanism to a significant degree and effectively address the issue of an irrational arrangement of teaching content. Generally, teachers recognize the accuracy of the model in data identification, with 87% of respondents believing that the model significantly improves the efficiency of academic diagnosis, especially the ability to accurately capture the characteristics of the distribution of student performance. In combination with the attribution report generated by DeepSeek, the model can target the adjustment of teaching content [10]. Furthermore, AI supports the dynamic generation of personalized intervention plans, which can help teachers complete teaching decisions efficiently.

However, certain pedagogues have indicated the model's deficiencies with regard to its capacity for dynamic adaptability. Primarily, the inability to encompass unstructured learning performances, such as classroom interactions and group practices, results in an incomplete depiction of the learning environment. This incomplete representation is a significant factor in the assessment of students' learning situations. A significant proportion of respondents, 87%, attested to the substantial enhancement in the efficiency of diagnosing learning situations that the model has brought about. Illustrative feedback included the following statement: "Previously, the manual analysis of assignment data required two weeks; presently, the model enables a comprehensive understanding of defective issues within ten minutes." The enhancement of data literacy through the model is also evident, as evidenced by the increase in young teachers conducting academic analysis independently, which in turn signifies the transformation of the teaching paradigm of mathematical intelligence.

V. Conclusion

In this paper, we propose a data-driven intelligent model for vocational education articulation. This model utilizes dynamic signage connected to multi-source heterogeneous data pools, with AI providing intelligent decision-making assistance. The model aims to achieve early warning of academic risk and quantitative diagnosis of cross-section course articulation. However, the model's small sample size and time limit mean it has not yet comprehensively portrayed the learning situation.

Future research will explore more unstructured academic data to achieve full-dimensional academic tracking and develop big data analysis in vocational education.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

Yikun Xu's responsibilities encompassed data analysis and paper writing; Sha Xian's responsibilities entailed data organization and visualization; Linjie Xiang was tasked with validating the effect of teaching reform in the pilot institutions; all authors had approved the final version.

FUNDING

2025 Chengdu Vocational University of the Arts Education and Teaching Reform Research project "Demand analysis for accounting talent and exploration of talent cultivation adaptation programs in the digital age".

REFERENCES

[1] L. Li and H. Suman, "Research on "Connecting with Secondary Vocational and Undergraduate Transition Examination" of preschool education specialty based on ability orientation," *Mobile Information Systems*, 2022.

[2] W. Feng, Z. Li, and Y. Pan, "Study on the "Secondary and Undergraduate" comprehensive training curriculum system of tourism management vocational education," *Tourism Management and Technology Economy*, vol. 5, no. 1, 2022.

[3] H. Li, L. X. Duan, and X. J. Yue, "Research on the integrated training path of secondary vocational, vocational, and undergraduate talents based on modern vocational education group," *Journal of Contemporary Educational Research*, vol. 7, no. 11, pp. 75–83, 2023.

[4] W. Zhang, "Dynamics of artificial intelligence educational applications in the context of education digital transformation – An analysis based on text data mining," *China Education Science (in English and Chinese)*, vol. 6, no. 3, pp. 52–60, 2023. doi: 10.13527/j.cnki.educ.sci.china.2023.03.005

[5] X. Yikun, X. Lizhu, and X. Fan, "A study on the educational model of connecting vocational and undergraduate education in finance and accounting majors," *Journal of Contemporary Educational Research*, vol. 8, no. 10, pp. 207–213, 2024.

[6] W. Wang and X. Huang, "Transitioning from secondary vocational school to university: A case study of first-year students from two Chinese universities," *Asia Pacific Journal of Education*, vol. 45, no. 2, pp. 614–627, 2025.

[7] Z. Hong, T. Shufen, Z. Peiyan, *et al.*, "Research on the construction system of teachers' ethics and style in secondary vocational schools," *International Education Forum*, vol. 3, no. 1, pp. 1–6, 2025.

[8] L. Wang, "Analysis of learning conditions in the context of big data for the integration of industry and education in private vocational colleges: A case study of the big data and accounting major in J College," *Public Relations World*, no. 22, pp. 106–108, 2024.

[9] C. Quanjie and N. Yanmin, "Research on student learning situation data analysis based on educational data mining," *Journal of Xinjiang Production and Construction Corps Education Institute*, vol. 33, no. 3, pp. 62–67+80, 2023.

[10] C. Ting, "Research on personalized diagnostic analysis and application based on student learning data," doctoral thesis, Hainan Normal University, 2023. doi: 10.27719/d.cnki.ghnsf.2023.000519