Leveraging Artificial Intelligence for Cross-Disciplinary Student Performance Prediction a Framework for Personalized Education and Academic Success

Bo Hou[1], Cong Zhou[2], Yahan Liu[3,*], Wei Xu[1] and Junting Zhang[4]

[1]Universiti Malaya, Kuala Lumpur, 50603, Malaysia
[2]Chongqing Normal University, Chongqing, 401331, China
[3]Tianjin University of Commerce, Tianjin, 300134, China
[4]Universiti Putra Malaysia, Selangor, 43400, Malaysia

*934868109@qq.com

Abstract. In the era of interdisciplinary education, personalized learning has become essential for fostering student success. However, accurately predicting student performance across diverse disciplines remains a significant challenge. This study explores the integration of Artificial Intelligence (AI) to develop a framework for cross-disciplinary student performance prediction, aiming to enhance educational resource allocation and support personalized education. Leveraging open-source datasets, the proposed framework incorporates machine learning models and Explainable AI (XAI) techniques, such as Shapley Additive Explanations (SHAP), to identify key factors influencing academic outcomes. Results demonstrate the framework's high accuracy in predicting performance and its ability to provide interpretable insights for educators, facilitating tailored interventions. This research highlights the potential of AI-driven approaches to optimize educational practices and promote academic success in interdisciplinary learning environments.

Keywords: Artificial Intelligence (AI), Cross-Disciplinary Education, Personalized Learning, Explainable AI (XAI)

1 Introduction

With the increasing diversity in education and the rise of interdisciplinary learning, predicting students' academic performance has become an increasingly complex challenge. These challenges stem from various factors, including imbalances in curriculum design, disparities in access to learning resources, and the diverse backgrounds of students. Against this backdrop, Artificial Intelligence (AI) technology, with its unique capabilities in data analysis, pattern recognition, and predictive modeling, has shown tremendous potential in the field of education. By leveraging AI, it is possible not only to enhance the accuracy of student performance predictions but also to provide robust support for the development of personalized education. However, traditional

R. Koul et al. (eds.), *Proceedings of the 2nd International Conference on Educational Development and Social Sciences (EDSS 2025)*, Advances in Social Science, Education and Humanities Research 924,
https://doi.org/10.2991/978-2-38476-400-6_94

educational research methods often struggle to fully integrate the interdisciplinary strengths of AI, pedagogy, and data science, thereby limiting the precision and effectiveness of educational interventions and student support strategies. To address this gap, this study proposes an AI-based interdisciplinary predictive framework that combines theories and techniques from multiple disciplines to optimize student performance prediction models and provide scientific evidence for educational interventions, ultimately fostering personalized education and academic success.

Recent advancements in Artificial Intelligence (AI) have provided new opportunities to address this issue. AI algorithms, particularly machine learning and deep learning methods, excel at processing high-dimensional and multivariate data. For example, XGBoost has been widely applied in predictive tasks due to its efficient feature selection and strong generalization capabilities. Meanwhile, Explainable AI (XAI) techniques, such as Shapley Additive Explanations (SHAP), enhance interpretability, allowing educators to understand the rationale behind model predictions.

This study proposes an AI-based interdisciplinary student performance prediction framework, integrating XGBoost and SHAP to improve predictive accuracy and interpretability. The key contributions of this research are as follows:

1. The development and implementation of an interdisciplinary student performance prediction framework that integrates machine learning and interpretability techniques.

2. The application of XGBoost for predictive modeling and SHAP for identifying key influencing factors, supporting educational interventions.

3. Experimental validation of the proposed framework's effectiveness in interdisciplinary student performance prediction and discussion of its potential applications in personalized education.

2 Related Work

Artificial Intelligence (AI) offers immense potential in addressing the complexities of predicting student performance in cross-disciplinary education. Challenges such as curriculum inconsistencies, resource disparities, and diverse student backgrounds necessitate advanced predictive tools. AI models like Artificial Neural Networks (ANNs), Random Forests, and Convolutional Neural Networks (CNNs) have demonstrated significant accuracy in forecasting academic outcomes. For example, ANNs have outperformed traditional methods in accuracy and reliability [1], while CNNs have proven effective in early-stage performance predictions [2].

These AI-driven frameworks also enable personalized education by adapting to individual needs. Explainable AI (XAI) techniques, such as SHAP, enhance transparency, allowing educators to provide tailored interventions [3]. Additionally, studies highlight the transformative potential of AI in personalizing learning pathways, improving engagement and academic outcomes by aligning content delivery with students' unique abilities [4].

Moreover, AI-enabled systems, AIEd in education offers key advantages such as 24/7 access to training, adaptation of content to students' needs, real-time feedback, and improvements in the educational process [5]. However, integrating cross-disciplinary data while addressing algorithmic bias and data privacy remains a challenge [6].

This research proposes a comprehensive AI-based framework, leveraging interdisciplinary data and advanced predictive analytics, to optimize educational practices. By combining AI's precision with educational insights, the framework aims to foster personalized learning and academic success in diverse, interdisciplinary environments.

3 Proposed Framework

3.1 Data Description

The Student Performance Dataset is a privately maintained dataset designed for research in educational data mining and machine learning. It includes two subsets: the Computer Science Programming dataset and the Computer Science Theory dataset. The Programming dataset contains 565 valid records, while the Theory dataset comprises 326 valid records. Both datasets feature 30-dimensional attribute characteristics, including two prior grades (G1: Midterm Exam 1, G2: Midterm Exam 2) and the final grade (G3: Final Exam). These datasets provide academic and demographic information about students, enabling the prediction and analysis of their academic performance in the field of computer science.

3.2 Methodology

Artificial Intelligence (AI) is revolutionizing education by enabling personalized learning and fostering academic success. This framework leverages XGBoost for cross-disciplinary student performance prediction, valued for its scalability and robust predictive capabilities in handling complex educational datasets. To enhance interpretability, SHAP (Shapley Additive Explanations) is integrated, providing clear insights into feature contributions and ensuring that educators can understand the factors influencing student outcomes.

By combining XGBoost's predictive power with SHAP's explainability, this AI-driven framework bridges advanced AI techniques with practical educational applications. It supports personalized education by identifying at-risk students and tailoring interventions to their needs, empowering educators with data-informed strategies to enhance academic achievement across diverse educational settings.

Data Preprocessing: Handle missing values, normalize continuous variables, and encode categorical features.

Model Training: Train the XGBoost model using the processed data.

Interpretability Analysis: Apply SHAP to the trained model to evaluate feature importance.

Evaluation: Compare predictions with actual outcomes using metrics like MAE, RMSE and R^2.

To enhance the performance of the XGBoost model, we employ hyperparameter tuning techniques:

Grid Search: Optimizes learning rate, maximum depth, and other parameters.

Bayesian Optimization: Further fine-tunes hyperparameters to identify the optimal configuration.

Additionally, to improve generalization, we adopt k-fold cross-validation (k=3) and evaluate model stability using metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R^2.

3.3 Predictive Modelling Design

Model construction and validation: The model construction and validation process begins with dividing the dataset into training and testing subsets, typically following an 80/20 split, to ensure sufficient data for training while retaining a representative sample for evaluation. The training data is used to train the XGBoost model, leveraging its efficiency and scalability to handle high-dimensional educational datasets. Cross-validation techniques, such as k-fold cross-validation, are applied to mitigate overfitting and enhance the model's generalizability. The model's performance is evaluated using metrics such as Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE), which quantify prediction accuracy, along with accuracy scores for assessing categorical outcomes.

Model Transparency Output and Interpretive Analysis: To enhance the model's interpretability and provide actionable insights, SHAP (Shapley Additive Explanations) is employed as an explanation tool. SHAP assigns contribution values to each feature for individual predictions, enabling a clear understanding of how different variables influence the model's decisions. Visualizations such as summary plots are generated to highlight the overall impact and importance of features across the dataset, while decision plots and force plots offer instance-specific explanations, showing how individual features contribute to specific predictions. These transparent and interpretable outputs bridge the gap between complex machine learning algorithms and practical applications, empowering educators and stakeholders to trust the model's predictions and apply them effectively in designing personalized educational strategies.

4 Experimental Results

4.1 Model Performance Evaluation

After extensive tuning and adjustment, the optimal prediction model and parameters were finally achieved. The performance of the XGBoost model, evaluated using key metrics, demonstrates its effectiveness. The model achieved an R^2 of 0.89, indicating a strong correlation between predicted and actual outcomes. Additionally, the Mean Absolute Error (MAE) is 0.87, reflecting relatively small deviations from the actual values, while the Root Mean Squared Error (RMSE) is 1.43, indicating a low level of error in the model's predictions. These results highlight the robustness of the model in predicting student performance.

4.2 SHAP interpretation model

As shown in Fig. 1, SHAP values offer clear insights into how each feature contributes to individual predictions. The top 10 features that have the greatest influence on the model's predictions are ranked based on their average absolute SHAP values. The most significant feature is the interaction between the first and second exam scores (Feature 133: G1 G2).

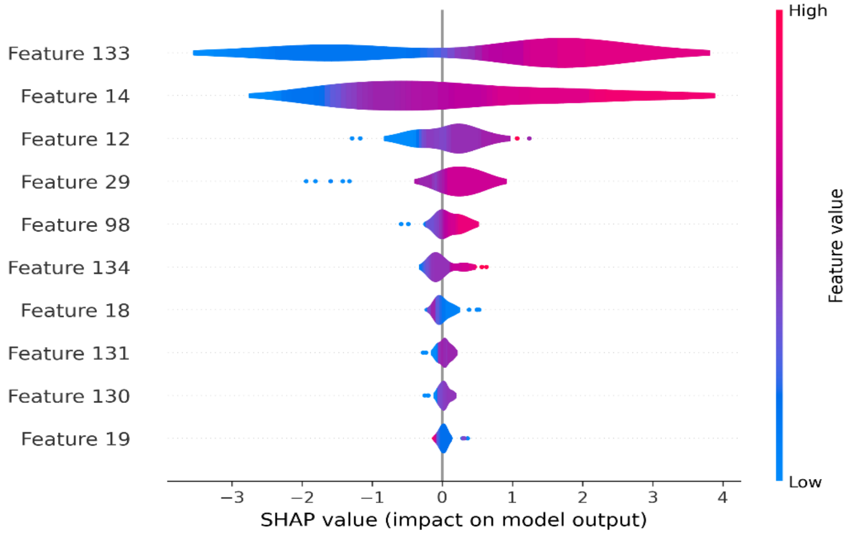

Fig. 1. SHAP Value Distribution: Violin Chart for Features

5 Conclusion

The integration of Artificial Intelligence (AI) into education addresses cross-disciplinary challenges by enabling accurate performance predictions and personalized learning. AI models like Artificial Neural Networks (ANNs) and Convolutional Neural Networks (CNNs) achieve high predictive accuracy, helping educators identify at-risk students and provide timely interventions. Explainable AI (XAI) techniques, such as SHAP, enhance transparency and trust, offering actionable insights for effective use. These advancements highlight AI's transformative role in reshaping education, promoting equity, and improving learning outcomes across disciplines.

The proposed AI-based framework aligns with the goals of personalized education by tailoring learning paths and resources to individual student needs. This adaptability not only boosts academic success but also fosters engagement and motivation among learners. However, challenges such as algorithmic bias, data privacy concerns, and the integration of interdisciplinary data must be addressed to fully realize AI's potential. Future research should focus on refining AI algorithms for greater fairness and scalability, exploring the long-term impacts of AI-driven education, and enhancing collaboration between AI developers, educators, and policymakers to bridge technological and pedagogical gaps.

In conclusion, leveraging AI for cross-disciplinary student performance prediction represents a significant leap toward personalized education and academic success. The proposed framework highlights the synergies between AI, pedagogy, and data science, offering a scalable solution for modern education. By addressing current limitations and fostering an interdisciplinary approach, AI-driven educational frameworks can unlock new possibilities for inclusive, effective, and dynamic learning environments, ultimately preparing students for the demands of a rapidly evolving world.

Future research should focus on integrating multimodal data sources, such as textual, audio, and video data, to enhance the predictive capabilities of AI-driven educational models. Additionally, the development of adaptive learning systems that leverage real-time data for personalized interventions could significantly improve student engagement and learning outcomes. Furthermore, addressing ethical considerations, such as algorithmic fairness and bias mitigation, is crucial to ensuring that predictive models do not inadvertently reinforce disparities based on gender, ethnicity, or socioeconomic status. By advancing these areas, future studies can contribute to the development of more equitable, effective, and scalable AI-powered education frameworks.

References

1. Agarwal, M., & Agarwal, B.: Mini Agarwal and Bharat Bhushan Agarwal. Predicting Student Academic Performance Using Neural Networks: Analyzing the Impact of Transfer Functions, Momentum and Learning Rate. *International Journal of Experimental Research and Review* 40, (2024)
2. Alshaikh, F., & Hewahi, N.: Convolutional neural network for predicting student academic performance in intelligent tutoring system. *International Journal of Computing and Digital Systems* 15(1), 239-258 (2024).
3. Adnan, M., et al., Earliest Possible Global and Local Interpretation of Students' Performance in Virtual Learning Environment by Leveraging Explainable AI. *IEEE Access* 10, 129843-129864 (2022).
4. Thimmanna, A.V.N.S.Sharma, Mahesh Sudhakar Naik, S. Radhakrishnan, Aarti Sharma, Personalized Learning Paths: Adapting Education with AI-Driven Curriculum, *EEL*, 14(1), 31–40 (2024).
5. O. Tapalova et al. Artificial Intelligence in Education: AIEd for Personalised Learning Pathways. *Electronic Journal of e-Learning* 20(5), 639-653 (2022).
6. Castro, G. P. B., Chiappe, A., Rodríguez, D. F. B., & Sepulveda, F. G. . Harnessing AI for Education 4.0: Drivers of Personalized Learning. *Electronic Journal of e-Learning*, *22*(5), 01-14 (2024).