

Dual-Path Transformer: A Novel Model for Education Resource Analysis and Optimization

Yajun Yuan¹, Dou Wen², Qinghui Xu^{3*}

¹INTI International University, Nilai, Negeri Sembilan, Malaysia

²Southwest Jiaotong University Hope College, Chengdu, China

³Faculty of Data Science and Information Technology, INTI International University, Nilai, Negeri Sembilan, Malaysia

Abstract. Reasonable distribution of educational resources is an important foundation for realizing educational equity and improving educational quality. However, analyzing educational data effectively remains challenging due to its heterogeneous nature, comprising both numerical metrics (e.g., study time, attendance) and categorical features (e.g., school type, family income). This paper proposes the Dual-Path Transformer model, an innovative educational resource analysis framework that addresses these challenges through two key innovations: (1) A dual-path feature processing architecture that separately processes numerical and categorical features through dedicated paths, then intelligently fuses them using an adaptive fusion strategy to capture their intrinsic relationships; (2) A specialized multi-attention mechanism that models complex educational patterns from multiple perspectives, enabling the model to understand intricate relationships such as the coupling between family economic status and learning resource access. Experiments on public educational datasets demonstrate that our model achieves superior performance with 90% accuracy and 88% F1 score, significantly outperforming both traditional machine learning approaches (Logistic Regression, SVM, Random Forest) and classical deep learning models (Neural Network). The model's interpretable nature provides educators with actionable insights into key factors affecting resource allocation, enabling data-driven decisions for promoting educational equity. These results demonstrate the Dual-Path Transformer's effectiveness in educational data analysis and its practical value for optimizing resource allocation strategies.

* Corresponding Author: Qinghui Xu (qinghui.xu.sg@gmail.com)

Keywords: educational resource analysis, Transformer, dual-path feature processing, multi-attention mechanism, educational equity, machine learning

1 Introduction

With the deepening global concern about the equity and quality of education, the rational distribution of educational resources has gradually become an important issue in both academia and practice. The distribution and utilization of educational resources directly affects students' learning effectiveness and the overall equity of society [1]. However, the diversity and complexity of educational data, such as student background, family income, and school quality, make it a great challenge to analyze and optimize the allocation of educational resources. Therefore, how to effectively mine and model the potential patterns of educational data, and then provide support for the scientific allocation of educational resources has become an urgent problem [2].

Educational resource analysis encompasses multiple critical dimensions that directly impact learning outcomes and institutional effectiveness. Physical infrastructure resources include classroom capacity optimization, library resource allocation, and laboratory equipment distribution, which form the foundation of educational delivery. Human capital resources involve strategic teacher-student ratio management, professional development investment, and instructional quality enhancement programs that significantly influence student engagement and academic achievement. Digital learning resources have become increasingly important, encompassing online platform utilization, educational software licensing, and network infrastructure development to support modern pedagogical approaches. Additionally, student support services such as tutoring programs, counseling resources, and financial aid distribution play crucial roles in ensuring equitable access to quality education. The complexity of these interconnected resource types creates multidimensional optimization challenges that traditional analytical approaches struggle to address effectively [3]. For instance, the relationship between teacher quality and student motivation, the interaction between family economic status and access to learning resources, and the impact of

digital infrastructure on learning outcomes all represent complex, non-linear relationships that require sophisticated modeling techniques for effective analysis and visualization. Current educational data mining approaches have shown promise in understanding these patterns, but often fail to capture the full complexity of educational ecosystems due to their reliance on conventional processing methods that cannot adequately handle the heterogeneous nature of educational data, thus limit the accuracy of educational data analysis and make it difficult to provide sufficient insights for educational decision makers [4].

To address these issues, this paper proposes an innovative model for analyzing educational resources, the Dual-Path Transformer model, which aims to build an analytical framework with both high performance and high interpretability by combining the diversity and complexity characteristics of educational data. The model in this study adopts a dual-path feature processing architecture, where numerical features (e.g., study time, sleep duration, historical grades, etc.) and categorical features (e.g., school type, family income, gender, etc.) are separately streamed, and intelligent combinations among features are realized through embedding and fusion strategies. In addition, based on Transformer's multi-attention mechanism, the model is able to model complex nonlinear relationships in education data from multiple perspectives, thus comprehensively improving the accuracy and robustness of the analysis [4].

In this paper, a comprehensive experimental validation of the proposed model is carried out on the task of educational resource allocation, which is compared and analyzed with traditional machine learning models (Logistic Regression, SVM, Random Forest) as well as classical deep learning models (Neural Network) [5]. The experimental results show that the Dual-Path Transformer model achieves significant performance improvements in both accuracy and F1 scores, proving its effectiveness and advantages in the field of educational resource analysis. More importantly, the model possesses good interpretability, which can help educators to deeply understand the key factors affecting the allocation of educational resources, thus providing a scientific basis for achieving educational equity and optimizing resource allocation [6].

2 Methodology

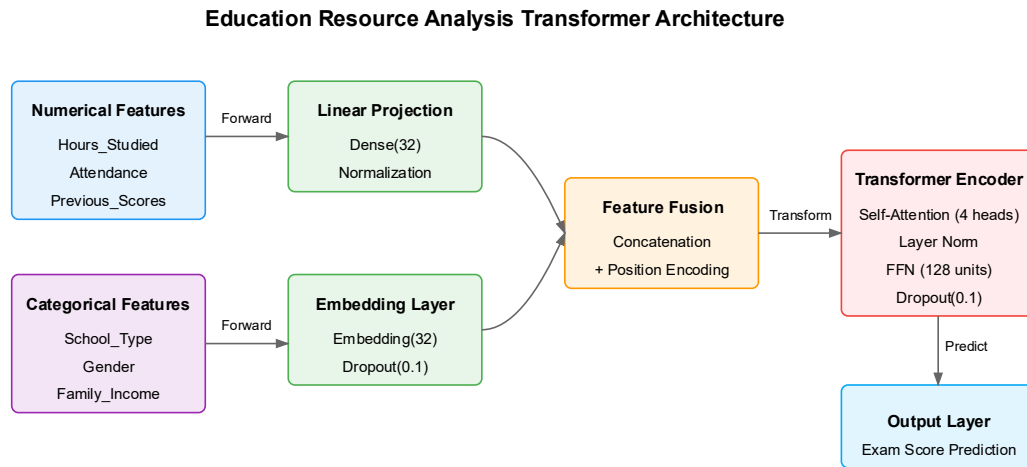


Figure 1. Overall architecture of the model

As shown in Figure 1, this Transformer-based education resource analysis model adopts an innovative two-way feature processing architecture that fully considers the diversity and complexity of education data. In the input layer, the model processes numerical features (e.g., learning time, attendance, sleep time, historical grades, etc.) and categorical features (e.g., school type, gender, family income, teacher quality, etc.) separately. Numerical features are mapped to a 32-dimensional space and normalized through a linear projection layer to ensure that features of different magnitudes can be compared within the same numerical range; category features are converted to a dense representation of the same dimension through an embedding layer, which not only captures the semantic relationships between categories, but also reflects the potential connections between different educational context factors [7]. This triage design of feature processing takes full account of the characteristics of educational data and lays a solid foundation for subsequent in-depth analysis.

In the feature fusion stage, the model adopts a unique feature integration strategy to intelligently combine the representations of numerical and categorical features through the feature fusion layer. This process can be expressed by the following formula:

$$X_{fused} = Concat(Linear(X_{numerical}), Embedding(X_{categorical})) + PE(pos)$$

where PE represents positional encoding.

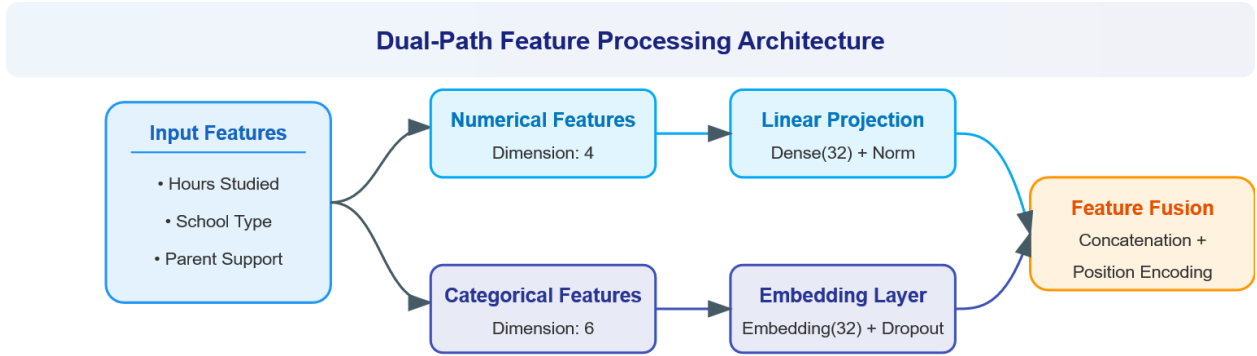


Figure 2. Dual path feature processing architecture

As shown in Figure 2, the core of the model is a specially optimized Transformer encoder structure containing two layers of encoder blocks, each equipped with a 4-head self-attention mechanism and a 128-cell feedforward network. The multi-head attention mechanism operates through parallel attention heads that capture different types of relationships within the educational data. Specifically, each attention head computes attention weights using the formula:

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

where Q , K , and V represent query, key, and value matrices respectively, and d_k is the dimension of the key vectors. The 128-cell feedforward network consists of two linear transformations with ReLU activation:

$$FFN(x) = MAX(0, xW_1 + b_1)W_2 + b_2$$

which enables the model to learn complex non-linear transformations. During the training process, the model adopts a dropout rate of 0.1 applied after each sub-layer and employs Layer Normalization technique using the formula:

$$LayerNorm(x) = \gamma \frac{x - \mu}{\sigma} + \beta$$

where μ and σ are the mean and standard deviation of the input, and γ and β are learnable parameters to ensure training stability and generalization ability.

This sophisticated architecture design allows the model to capture complex patterns in the allocation of educational resources, such as the relationship between household economic status and access to learning resources, the interaction between teacher quality and student motivation, and the impact of extracurricular activity participation on learning outcomes [9]. Through the mechanism of multiple attention heads, the model can simultaneously understand the interactions between these features from multiple perspectives and form a comprehensive knowledge of the educational process. Each attention head focuses on different aspects of the feature relationships, enabling the model to identify both direct correlations and subtle interdependencies that traditional methods might overlook. This well-designed architecture not only provides excellent prediction performance, but more importantly, it provides educators with an interpretable and trustworthy decision-support tool to help them better understand and optimize educational resource allocation strategies, which ultimately promotes educational equity and improves educational quality.

3 Experiments

3.1 Data Preprocessing

The experimental data in this study is sourced from the Student Performance Dataset, a publicly available educational dataset on Kaggle platform. This dataset contains comprehensive educational records of 480 students, including various types of features such as demographic information, academic performance indicators, behavioral metrics, and family background factors. The experiments were conducted using PyTorch framework with standard hyperparameter settings: batch size of 32, learning rate of 0.001 with Adam optimizer, and 100 training epochs. Five-fold cross-validation was employed to ensure the reliability and robustness of the experimental results. In this study, data preprocessing is considered as a key step in order to ensure that the model can extract high-quality features from educational data. First, numerical features (e.g., study time, attendance, sleep time, historical grades, etc.) were standardized to eliminate the influence of different feature scales so that they could be compared within the same numerical range. The

metric for the task of evaluating a relatively balanced distribution of data. F1-Score: the F1-score is the reconciled average of Precision and Recall, which is used to balance the performance of the model in the prediction of positive and negative samples, and is especially suitable for datasets with unbalanced categories. Its calculation formula is:

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

By combining precision and recall, the F1 score is able to reflect the actual effectiveness of the model more comprehensively.

The above two evaluation metrics complement each other and together provide the core basis for quantitative analysis of model performance. Among them, the precision rate measures the overall prediction accuracy, while the F1 score focuses on the robustness and generalization ability of the model in complex data distribution.

3.3 Comparison Experiments

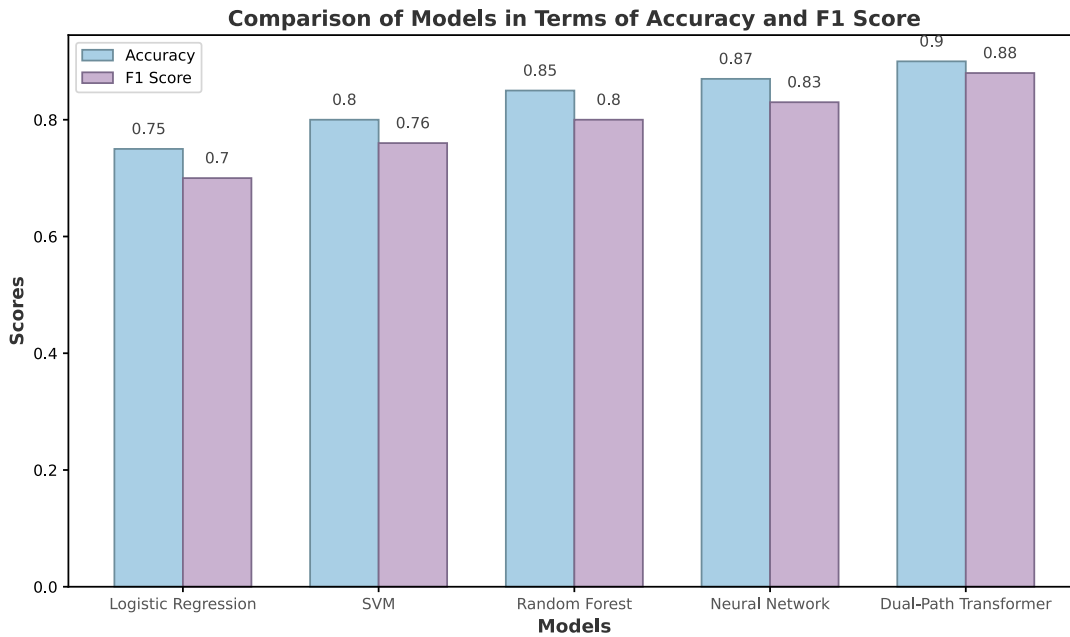


Figure 3. Comparison of experimental plots

In order to validate the effectiveness of the Dual-Path Transformer model in the task of analyzing

educational resources, this study conducted comparative experiments with traditional machine learning models (Logistic Regression, SVM, Random Forest) as well as classical deep learning models (Neural Network). The evaluation metrics are Accuracy (Accuracy) and F1-Score (F1-Score) to comprehensively measure the performance of the models in the classification task. The experimental results show that the Dual-Path Transformer model achieves the best performance on both metrics, with an Accuracy of 0.90 and an F1-Score of 0.88, which significantly outperforms the other compared models. This result indicates that the Dual-Path feature processing and feature fusion strategy proposed in this study can effectively mine the potential patterns of educational data and provide a more accurate analysis tool for the scientific allocation of educational resources [10].

In terms of the performance of traditional machine learning models, Logistic Regression is difficult to model complex educational data features due to its linear nature, with an accuracy of 0.75 and an F1 score of 0.70, which is the lowest among all the models. SVM improves its ability to model nonlinear features, with an accuracy and an F1 score of 0.80 and 0.76, respectively, but its ability to Random Forest performs relatively better, thanks to its decision tree integration-based approach, with an accuracy of 0.85 and an F1 score of 0.80 [10]. However, the overall performance of these traditional methods is still inferior to that of deep learning methods, as they are unable to adequately capture the diversity of the educational data and the deep interactions between features.

In contrast, deep learning models show stronger performance in the educational resource analysis task. neural Network outperforms all the traditional methods with its powerful feature extraction capabilities, achieving an accuracy and F1 score of 0.87 and 0.83, respectively. The Dual-Path Transformer model proposed in this study outperforms Neural Network in accuracy and F1 score by 3% and 5%, respectively, through the innovative dual-path feature processing architecture and multi-head attention mechanism [11]. The dual-path architecture effectively combines the characteristics of numerical and categorical features to accurately model the relationship between complex factors such as family economic status, teacher quality, and students' historical performance, while the multi-head attention mechanism is able to capture the interactions between

features from multiple perspectives. Taken together, the Dual-Path Transformer model not only significantly improves the performance of educational resources analysis, but also provides a scientific basis for educational equity and optimal allocation of resources [12].

4 Conclusion

In this paper, an innovative analysis model based on the Transformer architecture, the Dual-Path Transformer model, is proposed for the complexity and diversity problems faced in the task of educational resource allocation. Through the organic combination of Dual-Path feature processing architecture and multi-attention mechanism, the model can not only fully explore the potential relationship between numerical and categorical features in educational data, but also effectively capture the complex interaction patterns between different educational context factors. The experimental results show that the Dual-Path Transformer model significantly outperforms the traditional machine learning models (Logistic Regression, SVM, Random Forest) and the classical deep learning model (Neural Network) in terms of accuracy and F1 scores, reaching 90% and 88% respectively of the performance level. This result validates the effectiveness and superiority of the model in the task of educational resource analysis, especially its ability to model complex educational data shows significant advantages.

Despite the promising results achieved in this study, several limitations warrant consideration. The current model primarily analyzes structured educational data and has not yet incorporated unstructured data sources such as classroom recordings and textual materials, which may contain valuable educational insights. Additionally, the computational efficiency in high-dimensional scenarios and robustness to noisy data require further optimization for complex real-world environments. While our experiments validate the model's superiority using public datasets, its generalizability across different regional contexts and educational systems needs extensive validation. Future research should address these limitations through key directions: first, developing multimodal data fusion capabilities to integrate diverse data types for comprehensive resource analysis; second, incorporating transfer learning techniques to maintain stable

performance under varying data conditions; third, designing lightweight architectures for deployment on resource-constrained devices. Critically, future work must integrate real-time data processing capabilities to enable dynamic educational resource management, while addressing essential ethical considerations including data privacy protection and algorithmic fairness in educational decision-making processes. Overall, this research provides innovative tools for educational data analysis and resource optimization, contributing both theoretical advances and practical foundations for evidence-based educational policy making.

REFERENCES

- [1] Wu W. We know what attention is![[J]]. Trends in Cognitive Sciences, 2024, 28(4): 304-318.
- [2] Rabelo A, Rodrigues M W, Nobre C, et al. Educational data mining and learning analytics: A review of educational management in e-learning[[J]]. Information Discovery and Delivery, 2024, 52(2): 149-163.
- [3] Hashmi S J, Alabdullah B, Al Mudawi N, et al. Enhanced data mining and visualization of sensory-graph-Modeled datasets through summarization[[J]]. Sensors, 2024, 24(14): 4554.
- [4] Shao Y, Geng Z, Liu Y, et al. Cpt: A pre-trained unbalanced transformer for both chinese language understanding and generation[[J]]. Science China Information Sciences, 2024, 67(5): 152102.
- [5] Wang S, Li B Z, Khabisa M, et al. Linformer: Self-attention with linear complexity[[J]]. arXiv preprint arXiv:2006.04768, 2020.
- [6] Nemkul K. Use of bidirectional encoder representations from transformers (BERT) and robustly optimized BERT pretraining approach (RoBERTa) for Nepali news classification[[J]]. Tribhuvan University Journal, 2024, 39(1): 124-137.
- [7] Hassan W M. INTELLIGENT TUTORING SYSTEM ONTOLOGY[[J]]. 2024.
- [8] Holstein K, McLaren B M, Aleven V. Student learning benefits of a mixed-reality teacher awareness tool in AI-enhanced classrooms[C]//Artificial Intelligence in Education: 19th International Conference, AIED 2018, London, UK, June 27–30, 2018, Proceedings, Part I 19. Springer International Publishing, 2018: 154-168.
- [9] Reich J, Ruipérez-Valiente J A. The MOOC pivot[[J]]. Science, 2019, 363(6423): 130-131.
- [10] Chen T, Guestrin C. Xgboost: A scalable tree boosting system[C]//Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining. 2016: 785-794.
- [11] Lam P X, Mai P Q H, Nguyen Q H, et al. Enhancing educational evaluation through predictive student assessment modeling[[J]]. Computers and Education: Artificial

Intelligence, 2024, 6: 100244.

- [12] Alsariera Y A, Baashar Y, Alkawsi G, et al. Assessment and evaluation of different machine learning algorithms for predicting student performance[J]. Computational intelligence and neuroscience, 2022, 2022(1): 4151487.
- [13] A. Nursikuwagus, R. Munir, M. L. Khodra, and D. A. Dewi, "Model Semantic Attention (SemAtt) With Hybrid Learning Separable Neural Network and Long Short-Term Memory to Generate Caption," IEEE Access, vol. 12, pp. 154467-154481, 2024, doi: 10.1109/ACCESS.2024.3481499.