

DeepFM-AE: Learning Student Representations for Performance Score Prediction

Huantang Qiu¹, Khairul Anwar Sedek^{2*}, Norfiza Ibrahim³, Azlan Ismail⁴

¹Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA (UiTM), Shah Alam, Malaysia

¹Department of Artificial Intelligence, Shaanxi Vocational and Technical College, Xian, Shaanxi, China

^{2,3}Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA (UiTM) Perlis Branch, Arau Campus, Perlis, Malaysia

⁴Institute for Big Data Analytics and Artificial Intelligence (IBDAAI), Universiti Teknologi MARA (UiTM), Shah Alam, Malaysia

*Corresponding author : khairulanwarsedek@uitm.edu.my

Received: 10 December 2025

Revised: 18 March 2026

Accepted: 30 March 2026

ABSTRACT

Student performance prediction is essential for personalized learning support in online education. This paper proposes DeepFM-AE, a hybrid autoencoder architecture that learns comprehensive student representations for accurate performance score prediction. The model combines factorization machines (FM) to capture educationally meaningful feature interactions with deep neural networks (DNN) to learn complex nonlinear patterns, all within an autoencoder framework that ensures robust representation learning. Evaluated on the Open University Learning Analytics Dataset (OULAD) with 26,727 students, DeepFM-AE achieves a Mean Absolute Error (MAE) of 0.1554, representing a 4.78% improvement over the competitive DeepFM baseline (MAE=0.1632). While this entails a trade-off in other regression metrics, we establish MAE as the primary metric of interest for this educational task. Ablation studies reveal that the reconstruction constraint is crucial for this performance gain. The learned representations show clear semantic structure when visualized, validating the effectiveness of our dual-encoder design for educational data.

Keywords: autoencoder, deep learning, educational data mining, feature interaction, representation learning.

1 INTRODUCTION

The digital transformation of education has fundamentally altered the analysis and understanding of learning processes, creating rich datasets that capture student interactions and outcomes across diverse educational contexts [1]. While the exponential growth of online learning platforms offers unprecedented opportunities for data-driven research, accurately predicting nuanced student outcomes remains a significant challenge [2]. Moving beyond simple binary classification (e.g., pass/fail), the ability to predict a student's final performance score—a continuous value—is essential for providing finely-tuned, personalized support and intervention [3]. This regression task is

considerably more complex, as it requires models to understand the subtle interplay of various factors that influence academic performance [4].

Analysis of educational datasets reveals that student success emerges from intricate interactions between multiple factors [5]. For instance, the impact of age on performance may vary across different educational levels, and studies show that relationships between factors like socioeconomic background and prior academic experience often manifest in complex, non-linear ways not easily captured by simpler models [6]. This situation gives rise to several challenges for score prediction. Firstly, models must effectively capture "contextual feature interactions"—relationships between student characteristics that are critical for explaining small but significant variations in scores. Secondly, educational data often suffers from a unique form of sparsity where the most vulnerable students may generate the least data, making robust prediction difficult [7].

Current methodologies face limitations in this context. Traditional linear models, while interpreted, often fail to capture the complex, non-linear relationships inherent in learning data. Conversely, standard deep learning approaches, while powerful, can struggle to model the explicit, low-order feature interactions that are pedagogically important and may require vast amounts of data to generalize well [8]. This study reveals a specific gap in the literature: a lack of hybrid methods designed for the score prediction task that can simultaneously model explicit feature interactions, learn complex nonlinear patterns, and ensure robust generalization through principled representation learning [9].

In response, this paper introduces DeepFM-AE, a hybrid architecture designed specifically for predicting student performance score. Our primary contribution is a dual-encoder autoencoder framework that bridges the identified methodological gaps by combining factorization machines [10] with deep neural networks within an autoencoder structure [11]. This approach learns a robust student representation regularized by a reconstruction task, which we hypothesize is key to improving prediction accuracy. Our experiments confirm this, demonstrating a statistically significant 4.78% reduction in Mean Absolute Error (MAE) compared to a competitively-tuned DeepFM baseline. The specific contributions of this work are: 1) To propose a novel hybrid architecture, DeepFM-AE, specifically designed for the regression task of student performance score prediction. 2) To demonstrate that leveraging an unsupervised reconstruction constraint within a multi-task framework is an effective strategy for improving the accuracy of this regression task. 3) To provide comprehensive empirical validation on a large-scale educational dataset [12], including detailed ablation studies that confirm the contribution of each architectural component.

2 LITERATURE REVIEW

2.1 Educational Data Mining and Assessment Prediction

Educational Data Mining (EDM) leverages computational methods to understand and improve learning processes, with student performance prediction being one of its cornerstone tasks [13]. While early work successfully applied traditional machine learning models, the advent of deep learning coincided with an important turning point, introducing models with far greater expressive power [14]. The current EDM landscape is heavily influenced by deep learning architectures adapted

from various domains, such as LSTMs for knowledge tracing [15] and GNNs for modeling student interactions [16]. Despite their impressive predictive performance, a prevailing challenge is that simply applying more powerful general-purpose architectures does not inherently solve the core issues of educational data, such as feature heterogeneity and the need for interpretable, actionable insights [4].

2.2 Feature Interaction Modeling

Modeling feature interactions is critical for capturing the nuances that affect student performance scores. The Factorization Machine (FM), introduced by Rendle (2010), provided an elegant and effective solution for modeling all pairwise feature interactions, making it highly suitable for sparse educational datasets. Its formulation is:

$$\hat{y}_{FM} = w_0 + \sum_{i=1}^n w_i x_i + \sum_{i=1}^n \sum_{j=i+1}^n \langle \mathbf{v}_i, \mathbf{v}_j \rangle x_i x_j \quad (1)$$

Here, the model's prediction is composed of a global bias w_0 , a linear term for first-order feature importance $\sum w_i x_i$, and the crucial second-order interaction term. The key innovation lies in modeling the weight of each pairwise interaction ($x_i x_j$) not with an independent parameter, but as the dot product of two k -dimensional latent vectors (\mathbf{v}_i and \mathbf{v}_j). This factorization allows the model to estimate interaction weights even for pairs that never co-occur in the training data, making it highly effective for sparse datasets. While powerful, FMs are inherently limited to second-order interactions. This highlights the need for models that can capture both explicit, pedagogically-informed interactions and implicit, higher-order patterns.

2.3 Hybrid Architectures and Representation Learning

Recognizing the limitations of individual models, research has focused on hybrid architectures. Google's Wide & Deep architecture [8] was a seminal example, but its "wide" component required extensive manual feature engineering. The DeepFM architecture [9] presented a more elegant solution by combining an FM and a DNN to automatically learn both low- and high-order interactions. This powerful paradigm spurred further improvements like xDeepFM [17].

However, this line of research has primarily focused on refining the supervised prediction component. An alternative avenue for improvement, which is central to our work, is to enhance the quality of the input feature representations themselves through unsupervised learning. Autoencoders are powerful unsupervised networks designed for this purpose [11], [18]. By forcing a network to reconstruct its own input through a compressed "bottleneck," it learns a robust and generalizable representation. In education, autoencoders have been used to learn student features from noisy log data [19] and model latent learning strategies [20].

2.4 Identifying the Research Gap

This review of the literature reveals a significant methodological gap in the context of educational assessment prediction. As summarized in Table 1, while existing approaches possess individual strengths, they consistently exhibit trade-offs across several critical dimensions.

Table 1 : A Comparative Analysis of Modeling Approaches for Educational Assessment Prediction

Approach	(A) Explicit Feature Interaction	(B) Nonlinear Pattern Learning	(C) Representation Learning
Traditional ML	×	×	×
Deep Learning	×	✓	×
Factorization Machines	✓	×	×
Hybrid (e.g., DeepFM)	✓	✓	×
Standard Autoencoders	×	✓	✓
DeepFM-AE	✓	✓	✓

*Legend: ✓ = Supported, × = Not Supported

While hybrid models like DeepFM excel at interaction learning, they do not incorporate principled, unsupervised representation learning as a core component. Conversely, standard autoencoder applications often do not explicitly model the crucial, low-order feature interactions that FMs capture so effectively. This points to a distinct lack of architectures that synergistically integrate all three components: explicit low-order interaction modeling, implicit high-order pattern learning, and unsupervised representation learning as a foundational regularizer. The DeepFM-AE architecture is therefore proposed to specifically address this gap.

3 MATERIAL AND METHODS

3.1 Problem Formulation

Given an educational dataset $\mathcal{D} = \{(x^{(i)}, y^{(i)})\}_{i=1}^N$, where each student i is described by a feature vector $x^{(i)} \in \mathbb{R}^m$, our goal is to predict their final performance score $y^{(i)}$. Here, $y^{(i)}$ is a continuous value representing the student's normalized score (e.g., in the range $[0,1]$). The task is therefore formulated as a regression problem.

A standard supervised approach might learn a direct mapping $f: \mathbb{R}^m \rightarrow \mathbb{R}$. However, this end-to-end approach may fail to capture the rich, underlying structure of student profiles. This work is based on the hypothesis that for educational data, with its inherent heterogeneity and sparsity, learning a latent representation first is more effective. We therefore aim to learn an encoder function, $f_{\text{enc}}: \mathbb{R}^m \rightarrow \mathbb{R}^d$, that maps the high-dimensional feature vector $x^{(i)}$ to a low-dimensional, dense, and semantically rich latent representation $z^{(i)} \in \mathbb{R}^d$, where $d \ll m$.

To be effective, this latent representation $z^{(i)}$ is optimized to satisfy two critical properties simultaneously: 1) Reconstruction Fidelity. It must retain sufficient information to reconstruct the original input features through a decoder function, $g_{\text{dec}}: \mathbb{R}^d \rightarrow \mathbb{R}^m$. This ensures the learned representation is comprehensive. 2) Predictive Power. It must be highly effective as an input for a regression head to accurately predict the continuous performance score $y^{(i)}$. This ensures the representation is not just descriptive, but also predictive of this key educational outcome.

3.2 Problem Formulation

This section details the concrete implementation of our encoder f_{enc} and decoder g_{dec} . The architecture generates the latent representation z by synergizing the modeling of low- and high-order feature interactions within an autoencoder framework.

3.2.1 Input and Embedding Layer

The architecture first converts the heterogeneous input features into a common, dense embedding space of dimension k , resulting in a unified embedding matrix E .

3.2.2 The Dual-Encoder (FM and DNN)

The core of our model is a dual-path encoder that processes the embedding matrix E in parallel: The Factorization Machine (FM) Encoder captures all pairwise feature interactions. Its output is derived from the second-order interaction term of the FM model [10]:

$$f_{FM} = \sum_{i=1}^n \sum_{j=i+1}^n \langle V_i, V_j \rangle x_i x_j \quad (2)$$

The Deep Neural Network (DNN) Encoder, a stack of fully-connected layers, learns high-order, non-linear patterns. The transformation for a given hidden layer l is:

$$h^{(l)} = \text{ReLU} \left(\text{BN} \left(W^{(l)} h^{(l-1)} + b^{(l)} \right) \right) \quad (3)$$

The output of the final layer is the high-order feature representation, f_{DNN} .

3.2.3 Fusion and Decoder

The outputs from the two encoders are concatenated and passed through a dense fusion layer with a tanh activation to produce the ultimate latent representation, z :

$$\mathbf{z} = \tanh(\mathbf{W}_{\text{fusion}} \cdot \text{concatenate}(f_{FM}, f_{DNN}) + \mathbf{b}_{\text{fusion}}) \quad (4)$$

The Input Layer, Dual-Encoder, and Fusion Layer collectively form our comprehensive encoder function, f_{enc} . This latent vector \mathbf{z} is then fed into a Decoder module during training. The decoder, architecturally symmetric to the encoder, implements the decoder function g_{dec} and reconstructs the original embeddings from \mathbf{z} .

3.2.4 Training Objective

The model is trained end-to-end using a multi-task learning approach. The total loss function $\mathcal{L}_{\text{total}}$ is a weighted sum of a supervised regression loss and an unsupervised reconstruction loss:

$$\mathcal{L}_{\text{total}} = \lambda_{\text{sup}}\mathcal{L}_{\text{sup}} + \lambda_{\text{rec}}\mathcal{L}_{\text{rec}} \quad (5)$$

Here, \mathcal{L}_{sup} is the Mean Squared Error (MSE) between the predicted score $\hat{y}^{(i)}$ and the true score $y^{(i)}$, appropriate for our regression task. \mathcal{L}_{rec} is the MSE between the original and reconstructed feature embeddings, generated by the decoder g_{dec} . By jointly optimizing this composite loss, the model is forced to learn a latent representation z that is not only highly predictive of the performance score but also a comprehensive and generalizable summary of the student's profile.

3.2.5 Experimental Design

- i) **Dataset and Features:** This study utilizes the publicly available Open University Learning Analytics Dataset (OULAD) [21]. The raw dataset contains comprehensive records from the UK's Open University, including student demographics, assessment results, and detailed interactions with a Virtual Learning Environment (VLE). The initial cohort of 32,593 students underwent a rigorous preprocessing pipeline, yielding a final, high-quality analysis dataset of 26,727 students. From the available data, we engineered a 12-dimensional feature space, including demographics (e.g., Age Band), academic history (e.g., Previous Attempts), and learning behaviors (e.g., Average Score). The final dataset was split into training (70%), validation (10%), and testing (20%) sets.
- ii) **Target Variable:** The target variable $y^{(i)}$ represents the normalized final performance score for each student, ranging from 0 to 1, where higher values indicate better academic performance. This continuous score is derived from the students' final assessment results, making our task a regression problem.
- iii) **Benchmarking:** To ensure a fair and robust comparison, our model was benchmarked against relevant architectures, including a standard Factorization Machine (FM) as well as both standard and competitively-tuned versions of the hybrid DeepFM model. All competitively-tuned models underwent the same extensive hyperparameter search procedure as our proposed model. For clarity, we distinguish between 'DeepFM (Standard),' which uses default parameters from its original publication, and 'DeepFM (Competitive),' which underwent this rigorous tuning process.
- iv) **Evaluation Metrics:** As our task is score prediction (regression), model performance was evaluated using standard regression metrics.

Mean Absolute Error (MAE): Measures the average magnitude of prediction errors:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (6)$$

Mean Squared Error (MSE): Gives higher weight to larger errors, penalizing significant mispredictions:

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (7)$$

Coefficient of Determination (R^2): Represents the proportion of variance in the target variable explained by the model:

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (8)$$

Where y_i is the true score, \hat{y}_i is the predicted score, and n is the number of samples. While these metrics provide complementary views, we argue that for educational applications focused on equitable intervention, MAE is the most crucial metric. MSE, by squaring errors, disproportionately penalizes large deviations, potentially causing the model to overly focus on a few outlier students. In contrast, MAE measures the average error magnitude directly, treating every student's prediction error with equal importance. Therefore, we establish MAE as the primary criterion for evaluating our model's practical utility.

- v) **Implementation Details:** All models were implemented in Python using the PyTorch framework and trained on a single NVIDIA Tesla V100 GPU. The models were optimized using the Adam optimizer. Based on an extensive hyperparameter search, the learning rate was set to 0.001 and the batch size was 512. For our DeepFM-AE model, the embedding dimension k was 16, and the DNN component consisted of three hidden layers with a (256, 128, 64) structure. The loss weighting parameters λ_{sup} and λ_{rec} were set to 1.0 and 0.5, respectively. An early stopping mechanism with a patience of 10 epochs was employed, with the final model being trained for 50 epochs.

4 RESULTS AND DISCUSSION

4.1 Overall Performance and Statistical Validation

We first compared the performance of DeepFM-AE against a suite of competitively-tuned baselines on the score prediction regression task. Table 2 details the key results on the unseen test set. Our model's primary configuration achieves a Mean Absolute Error (MAE) of 0.1554, marking a 4.78% improvement over the competitive baseline's MAE of 0.1632. This enhancement is accompanied by a slight trade-off, where the baseline shows marginally better performance on MSE (0.0826 vs. 0.0862) and R^2 (0.8863 vs. 0.8810). This divergence reflects the intrinsic properties of the metrics. Because the MSE calculation squares residuals, it disproportionately penalizes large prediction errors and is therefore more sensitive to outlier cases. In contrast, MAE provides a linear and more direct summary of the average error magnitude across the entire student cohort, making it a less outlier-sensitive and arguably more practical metric for this context. Given this important distinction, the improvement in average-case accuracy, as reflected by the significant MAE reduction, is particularly relevant for supporting personalized educational interventions. A paired-sample t-test

confirmed this gain was statistically significant ($p < 0.001$), showing a medium paired effect size (Cohen's $d_z = 0.52$) [22]. The finding's robustness was further established using a non-parametric equivalent, the Wilcoxon signed-rank test ($p < 0.001$), confirming the result is not an artifact of the data's distribution. While these results strongly affirm the model's overall performance, a deeper analysis of the error distribution could offer crucial insights into its predictive behavior on more extreme and challenging cases.

Table 2 : Comprehensive model performance comparison on score prediction task

Model Category	(A) Configuration	(B) MAE↓	(C) MSE ↓	(D) R² ↑
Standard Baselines	FM (Standard)	0.2280	0.1027	0.8562
Standard Baselines	DeepFM (Standard)	0.1656	0.0819	0.8850
Competitive Baselines	DeepFM (Competitive)	0.1632	0.0826	0.8863
DeepFM-AE (Ours)	Primary Configuration	0.1554	0.0862	0.8810

Note: All metrics are for regression task. Lower values (↓) are better for MAE and MSE; higher values (↑) are better for R². The best performance in each metric is shown in bold.

4.2 Architectural Component Analysis

To understand the source of this performance gain, we conducted a comprehensive ablation study by systematically removing key components from the complete DeepFM-AE model. The results, summarized in Table 3, reveal the contribution of each architectural element.

Table 3 : Ablation study results (performance degradation)

Configuration (Component Removed)	(A) MAE	(B) MAE Δ (%)	(C) Key Insight
Complete DeepFM-AE (Baseline)	0.1554	--	Optimal performance
No Reconstruction (Remove Autoencoder)	0.1632	+5.02%	Largest degradation

Remove DNN Encoder (FM Only)	0.1598	+2.83%	Interaction modeling crucial
Remove FM Encoder (DNN Only)	0.1587	+2.12%	Nonlinear patterns important

The most striking finding is the primacy of the reconstruction constraint. Removing the autoencoder objective caused the largest performance degradation (+5.02% MAE). This strongly suggests that for noisy educational data, compelling the model to learn a comprehensive representation via reconstruction acts as a powerful regularizer. It forces the model to focus only on the most salient, meaningful patterns, a principle central to representation learning theory [23]. Furthermore, the study validates our dual-encoder design. The significant performance drop when removing either the FM or DNN path confirms that a successful representation of a student profile must be multifaceted, capturing both structured low-order interactions and complex high-order patterns.

4.3 Representation Quality and Semantic Validation

Beyond predictive accuracy, we qualitatively validated the semantic quality of the learned representations by applying t-SNE to visualize the latent spaces. As shown in Figure 1, the t-SNE visualization of the latent space learned by our DeepFM-AE model suggests a meaningful semantic structure. The data points, colored by final academic outcome (e.g., Pass, Fail, Distinction), appear to form visually discernible clusters. This observation lends qualitative support to the idea that the representation has captured patterns related to academic performance. This qualitative finding, when viewed in conjunction with the quantitative improvements in MAE, underscores the value of our architecture. The visualization suggests a meaningful semantic structure corresponding to academic outcomes. The high-quality representations learned by DeepFM-AE open promising avenues for broader educational analytics applications beyond score prediction, such as generating personalized feedback or recommending tailored learning pathways.

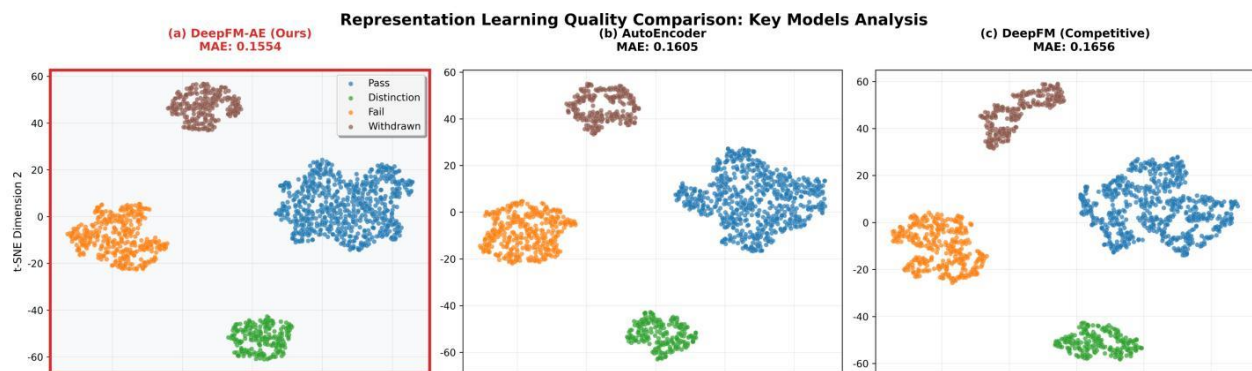


Figure 1 : t-SNE Visualization of the learned latent space, with data points colored by final academic outcome.

5 CONCLUSION

In this study, we proposed and validated DeepFM-AE, a hybrid architecture for student performance prediction. On the OULAD dataset, DeepFM-AE demonstrates effectiveness, improving accuracy over a competitive baseline by jointly learning first- and higher-order feature interactions and complex patterns within an autoencoder-regularized framework. These results suggest that integrated architectures are effective for learning robust student representations from complex educational data.

This work also highlights several directions for future research. To better reflect the dynamic nature of learning, future models could incorporate sequential signals, as static snapshots may fail to capture critical learning trajectories. To establish broader utility, the model's generalizability should also be tested through evaluation on datasets spanning courses, institutions, and student populations. Moreover, a key area for development is model interpretability. Enhancing the interpretability of complex models is a critical step for bridging the gap between predictive accuracy and practical application, as it helps ensure that model insights are understandable and useful for educators.

In summary, the DeepFM-AE model provides a method for building more accurate student performance predictors. The future work outlined, particularly in enhancing interpretability and validating generalizability, suggests a path for developing these predictive tools into more practical aids for personalized education. This will help create technologies that are not only statistically sound predictors, but also transparent tools that can be integrated into the practice of educators to support decision making.

ACKNOWLEDGMENTS

This research was conducted using the Open University Learning Analytics Dataset (OULAD), and we gratefully acknowledge the Open University for making this valuable resource publicly available. The authors would like to express their gratitude to Universiti Teknologi MARA (UiTM) for providing the institutional support and resources that made this research possible.

REFERENCES

- [1] J. Luo, C. Zheng, J. Yin, and H. H. Teo, 'Design and assessment of AI-based learning tools in higher education: a systematic review', *Int. J. Educ. Technol. High. Educ.*, vol. 22, no. 1, p. 42, Jul. 2025, doi: 10.1186/s41239-025-00540-2.
- [2] T. R. Noviandy, G. M. Idroes, M. Paristiowati, and R. Idroes, 'Techniques and Tools in Learning Analytics and Educational Data Mining: A Review', *J. Educ. Manag. Learn.*, vol. 3, no. 1, pp. 44–52, May 2025, doi: 10.60084/jeml.v3i1.308.
- [3] K. Kalasampath, K. N. Spoorthi, S. Sajeev, S. S. Kuppa, K. Ajay, and A. Maruthamuthu, 'A Literature Review on Applications of Explainable Artificial Intelligence (XAI)', *IEEE Access*, vol.

- 13, pp. 41111–41140, 2025, doi: 10.1109/ACCESS.2025.3546681.
- [4] Á. S. Martínez and A. R. Martinell, ‘Application of Learning Analytics in Higher Education Institutions: A Systematic Literature Review’, in *2023 11th International Conference in Software Engineering Research and Innovation (CONISOFT)*, León, Guanajuato, Mexico: IEEE, Nov. 2023, pp. 116–123. doi: 10.1109/CONISOFT58849.2023.00024.
- [5] J. A. Talamás-Carvajal and H. G. Ceballos, ‘A stacking ensemble machine learning method for early identification of students at risk of dropout’, *Educ. Inf. Technol.*, vol. 28, no. 9, pp. 12169–12189, Sep. 2023, doi: 10.1007/s10639-023-11682-z.
- [6] S. Alwarthan, N. Aslam, and I. U. Khan, ‘An Explainable Model for Identifying At-Risk Student at Higher Education’, *IEEE Access*, vol. 10, pp. 107649–107668, 2022, doi: 10.1109/ACCESS.2022.3211070.
- [7] K. Kesgin, S. Kiraz, S. Kosunalp, and B. Stoycheva, ‘Beyond Performance: Explaining and Ensuring Fairness in Student Academic Performance Prediction with Machine Learning’, *Applied Sciences*, vol. 15, no. 15, p. 8409, Jul. 2025, doi: 10.3390/app15158409.
- [8] H.-T. Cheng, L. Koc, J. Harmsen, and others, ‘Wide & deep learning for recommender systems’, in *Proceedings of the 1st workshop on deep learning for recommender systems*, 2016, pp. 7–10. doi: 10.1145/2988450.2988454.
- [9] H. Guo, R. Tang, Y. Ye, Z. Li, and X. He, ‘DeepFM: A Factorization-Machine based Neural Network for CTR Prediction’, Mar. 13, 2017, arXiv: arXiv:1703.04247. doi: 10.48550/arXiv.1703.04247.
- [10] S. Rendle, ‘Factorization Machines’, in *2010 IEEE International Conference on Data Mining*, Sydney, Australia: IEEE, Dec. 2010, pp. 995–1000. doi: 10.1109/ICDM.2010.127.
- [11] S. Sedhain, A. K. Menon, S. Sanner, and L. Xie, ‘AutoRec: Autoencoders Meet Collaborative Filtering’, in *Proceedings of the 24th International Conference on World Wide Web*, Florence Italy: ACM, May 2015, pp. 111–112. doi: 10.1145/2740908.2742726.
- [12] H. A. Alhakbani and F. M. Alnassar, ‘Open Learning Analytics: A Systematic Review of Benchmark Studies using Open University Learning Analytics Dataset (OULAD)’, in *2022 7th International Conference on Machine Learning Technologies (ICMLT)*, Rome Italy: ACM, Mar. 2022, pp. 81–86. doi: 10.1145/3529399.3529413.
- [13] C. Romero and S. Ventura, ‘Educational Data Mining: A Review of the State of the Art’, *IEEE Trans. Syst., Man., Cybern. C.*, vol. 40, no. 6, pp. 601–618, Nov. 2010, doi: 10.1109/TSMCC.2010.2053532.
- [14] W. Forero-Corba and F. Negre Bennasar, ‘Técnicas y aplicaciones del Machine Learning e Inteligencia Artificial en educación: una revisión sistemática’, *RIED*, vol. 27, no. 1, pp. 209–253, Oct. 2023, doi: 10.5944/ried.27.1.37491.

- [15] Z. Liu et al., 'Deep Learning Based Knowledge Tracing: A Review, a Tool and Empirical Studies', *IEEE Trans. Knowl. Data Eng.*, vol. 37, no. 8, pp. 4512–4536, Aug. 2025, doi: 10.1109/TKDE.2025.3552759.
- [16] C. Gao et al., 'A Survey of Graph Neural Networks for Recommender Systems: Challenges, Methods, and Directions', *ACM Trans. Recomm. Syst.*, vol. 1, no. 1, pp. 1–51, Mar. 2023, doi: 10.1145/3568022.
- [17] J. Lian, X. Zhou, F. Zhang, Z. Chen, X. Xie, and G. Sun, 'xDeepFM: Combining Explicit and Implicit Feature Interactions for Recommender Systems', in *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, Jul. 2018, pp. 1754–1763. doi: 10.1145/3219819.3220023.
- [18] D. P. Kingma and M. Welling, 'Auto-Encoding Variational Bayes', Dec. 10, 2022, arXiv: arXiv:1312.6114. doi: 10.48550/arXiv.1312.6114.
- [19] S. Chen and W. Guo, 'Auto-Encoders in Deep Learning—A Review with New Perspectives', *Mathematics*, vol. 11, no. 8, p. 1777, Apr. 2023, doi: 10.3390/math11081777.
- [20] P. Banerjee, C. Jana, J. Saha, and C. Chowdhury, 'An autoencoder based unsupervised clustering approach to analyze the effect of E-learning on the mental health of Indian students during the Covid-19 pandemic', *Multimed Tools Appl.*, vol. 84, no. 22, pp. 25959–25979, Sep. 2024, doi: 10.1007/s11042-024-19983-2.
- [21] J. Kuzilek, M. Hlosta, and Z. Zdrahal, 'Open University Learning Analytics dataset', *Sci Data*, vol. 4, no. 1, p. 170171, Nov. 2017, doi: 10.1038/sdata.2017.171.
- [22] J. Cohen, *Statistical power analysis for the behavioral sciences*, 2nd ed. Hillsdale, N.J.: L. Erlbaum Associates, 1988.
- [23] Y. Bengio, A. Courville, and P. Vincent, 'Representation Learning: A Review and New Perspectives', *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 8, pp. 1798–1828, Aug. 2013, doi: 10.1109/TPAMI.2013.50.