Personalized Evaluation of Online Predictive Systems through Learner Profiles

Fatna Ennibras
L2IACS Laboratory, ENSET Mohammedia, Hassan II University of Casablanca, Morocco
ennibrasfatna18@gmail.com (corresponding author)

Bouchra Bouihi
L2IACS Laboratory, ENSET Mohammedia, Hassan II University of Casablanca, Morocco
bouihi@enset-media.ac.ma

Rachid El Aasri
L2IACS Laboratory, ENSET Mohammedia, Hassan II University of Casablanca, Morocco
rachidelaasri@enset-media.ac.ma

Es-Saadia Aoula
L2IACS Laboratory, ENSET Mohammedia, Hassan II University of Casablanca, Morocco
aoula@enset-media.ac.ma

Abdelmajid Bousselham
L2IACS Laboratory, ENSET Mohammedia, Hassan II University of Casablanca, Morocco
bousselham@enset-media.ac.ma

Received: 7 October 2025 | Revised: 21 November 2025 and 8 December 2025 | Accepted: 19 December 2025

 DOI: https://doi.org/10.48084/etasr.15384

ABSTRACT

Distance education institutions continue to face high rates of failure and dropout each year, mainly due to the autonomy of the learner and the absence of continuous monitoring. Numerous predictive systems have been developed to identify students at risk; however, their evaluation often relies on global performance metrics that overlook the heterogeneity of learner profiles. This paper introduces a personalized evaluation methodology for predictive models, applied to real data from Loghate, an online multilingual learning platform implemented at Hassan II University of Casablanca to promote language acquisition among students. After thorough preprocessing, engagement and performance indicators were extracted, and learners were clustered into homogeneous profiles using the K-means algorithm. A stacking ensemble model, combining several supervised learning techniques, was then trained and assessed both globally and within each profile. The model achieved strong global performance (82% accuracy), yet clear disparities appeared across learner profiles, with one group reaching 88% accuracy while another dropped to 75%. These variations show that global metrics can mask subgroup-specific biases and reduce the reliability of early warning systems. This approach reinforces the need for differentiated evaluation to build fairer and more reliable predictive systems that adapt interventions to learners' needs. In the long term, it could support the development of adaptive and explainable dashboards for more personalized and equitable learning environments.

Keywords-online learning; at-risk students; prediction; personalized evaluation; stacking; clustering; learner profiles

I. INTRODUCTION

With the rapid growth of digital technologies and the widespread adoption of online learning, higher education has undergone a profound transformation, offering more flexible and accessible formats for distance education [1, 2]. Traditional geographical and financial barriers associated with face-to-face instruction have gradually diminished, allowing institutions worldwide to adopt online courses as a viable model [3]. This shift was further accelerated during the COVID-19 pandemic, which expanded access to education and increased the need for effective digital pedagogies [1, 2].

Despite these advantages, online learning brings specific challenges. The lack of direct interaction makes it difficult to detect struggling students early [4], highlighting the need for predictive systems that can support timely and personalized interventions [5, 6]. A previous work [7] explored the prediction of dropout using regularized regression models, emphasizing the importance of generalization in small educational datasets. Unlike previous studies, this work evaluates predictive reliability across learner profiles, revealing limitations of global metrics when learner variability is not considered.

The prediction of online learning students at-risk relies mainly on supervised models that use activity and performance traces to detect learners who are likely to fail [8]. Common approaches include decision trees, random forests, SVM, logistic regression, and neural networks, typically assessed using global metrics such as accuracy, precision, recall, and F1-score [9, 10]. However, these aggregated indicators overlook the heterogeneity of learners' behaviors, meaning a model may achieve a high average score while still failing to predict outcomes for atypical or minority student groups [11, 12].

The lack of attention to learner profile diversity raises issues of reliability and fairness, as global models tend to favor the most represented students while neglecting others. These concerns have fueled interest in differentiated approaches that identify learner profiles and evaluate predictive models within each group [13]. To address this, unsupervised methods, such as K-means, hierarchical clustering, fuzzy c-means, and DBSCAN, have been introduced to detect learner profiles. These studies show that students can be grouped by engagement and achievement patterns, each reflecting distinct pedagogical needs [14, 15]. However, most works stop at describing these profiles and rarely assess the reliability of predictive systems across different learner groups.

In this context, in [13], a profile-based evaluation methodology involved three steps: (i) identifying personas from engagement and activity indicators, (ii) assessing model performance for each profile, and (iii) integrating temporal indicators such as earliness and stability. The results showed that system quality can vary significantly across learner profiles, and that a high global score does not guarantee reliable or fair alerts for all. These findings support the need for differentiated evaluation methods and the inclusion of time-sensitive indicators aligned with early intervention.

This study introduces a personalized evaluation of a stacking-based predictive system by integrating learner profiles derived from the Loghate platform. Students are clustered according to engagement and performance indicators, and model performance is examined within each subgroup to identify potential biases and limitations in generalization. The aim was to provide a more fine-grained understanding of the reliability of online predictive systems and contribute to the development of strategies to personalize pedagogical support.

II. RESEARCH OBJECTIVES

The primary objective of this study is to examine how a personalized evaluation of predictive systems, grounded in learner profiles, can enrich the interpretation of results and uncover biases or limitations that remain hidden in traditional global evaluations (such as accuracy, F1-score, precision, or recall). By focusing on the heterogeneous behaviors of students in an online learning environment, this approach seeks to identify for which profiles a predictive model performs reliably, and for which it fails. More specifically, the aims of this study are as follows:

- Define learning indicators from digital traces that can be used to characterize students' learning behaviors.
- Identify learner profiles from the data collected on the Loghate platform by clustering students according to similar patterns of learning behavior.
- Evaluate the performance of a stacking prediction model within each of the identified groups to detect differences in reliability across profiles.
- Analyze potential biases in the model according to learner categories and propose recommendations for a more equitable and personalized use of early warning systems.

In line with these objectives, this study hypothesizes that the stacking model will not perform uniformly across all learner groups, and that performance disparities will emerge based on differences in engagement and learning behaviors.

III. EVALUATION METHODOLOGY

A. Adopted Methodology

To achieve these objectives, the adopted methodology is structured into three main stages, inspired by profile-based analysis approaches in educational prediction systems. Figure 1 illustrates these stages.

1) Identification of Learner Profiles

This study begins with an analysis of the dataset obtained from the Loghate platform to identify relevant learning indicators, including engagement variables (total connection time, number of completed activities) and performance measures (placement test scores, overall results).APreprocessing included handling missing values, removing anomalies using z-score thresholds, and normalizing numerical variables through Min–Max scaling. Key hyperparameters for the machine learning models (e.g., number of trees for Random Forest, C parameter for SVM) were tuned through grid search within cross-validation to ensure optimal performance.

The learners were then segmented into homogeneous groups using the K-means clustering algorithm. The choice of $k = 4$ was supported by empirical validation: the elbow method showed a clear inflection between $k = 3$ and $k = 4$, and silhouette scores computed for $k = 2$ to 7 indicated comparable cohesion–separation quality (0.46–0.47), with $k = 4$ offering a good balance between statistical validity and interpretability. To ensure robustness, the clustering was repeated with multiple initializations ($n\_init = 10$), and the resulting profiles remained stable across runs. Each student was thus assigned to a consistent and meaningful profile reflecting their general learning behavior.

2) Predictive Modeling and Initial Evaluation

At this stage, a supervised prediction model is applied based on the stacking of several machine learning algorithms. The base models are Random Forest (RF), Support Vector Machine (SVM), Logistic Regression (LR), Decision Tree (DT), Naive Bayes (NB), and Artificial Neural Network (ANN), with predictions aggregated through an SVM meta-learner. Given that 73% of the learners were certified, class imbalance was addressed by using stratified sampling during cross-validation and by applying class weights within the SVM meta-learner to reduce bias toward the majority class. The model's performance was first assessed globally using standard metrics such as accuracy, precision, recall, and F1-score. To ensure robustness, a stratified 10-fold cross-validation procedure was applied. All experiments were conducted in Python 3.10 using Scikit-learn 1.4 for machine learning models, Pandas and NumPy for data processing, and Matplotlib/Seaborn for visualizations, ensuring full reproducibility.

To improve reproducibility, the main hyperparameters of the base models were tuned using a grid-search procedure combined with stratified 10-fold cross-validation. Random Forest was evaluated with different values of $n\_estimators$ and max_depth, SVM with several C values and a linear kernel, and Logistic Regression with L2 regularization. The Decision Tree was constrained through a maximum depth to reduce overfitting, while the ANN meta-learner used two dense layers (64 and 32 units) with ReLU activations and the Adam optimizer. All models were trained and validated under the same stratified cross-validation scheme to ensure stable performance despite class imbalance.

3) Profile-based Evaluation

The final step focuses on analyzing model performance within each learner profile identified in Step 1. Specifically, the metrics obtained for each subgroup were compared to determine potential variations in predictive performance across profiles. This analysis allows identifying groups of learners for which the model is less reliable, thereby improving the interpretation of its outcomes and anticipating its limitations. This step provides a critical foundation for the design of a more equitable and accurate early-warning system.

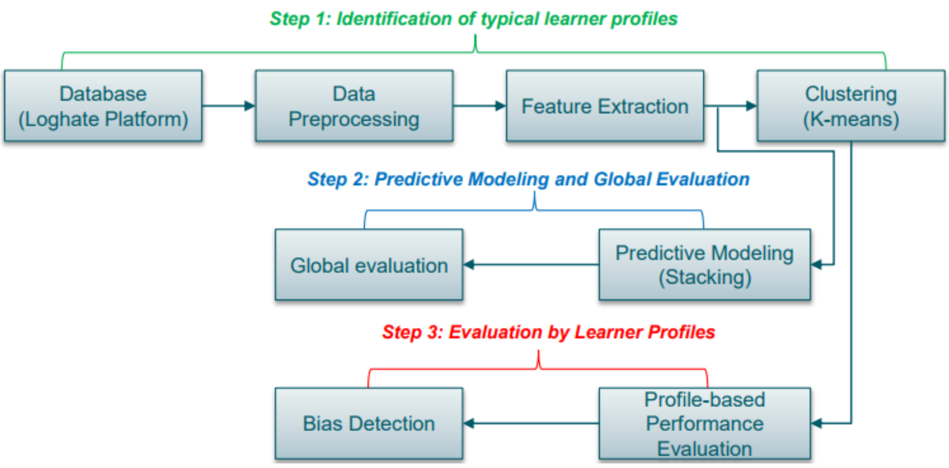

Fig. 1. Proposed methodology for learner profile-based predictive evaluation.

B. Data Description

This study relies on a dataset collected from the Loghate platform (ALTISSIA) at Hassan II University, covering the period from January 2022 to February 2023. Loghate is an online learning environment, open to the entire university community, that supports the learning of seven languages. Although its use is not mandatory, it provides valuable data on student engagement and performance.

The dataset includes several interconnected sources (engagement metrics, test results, certification records, and general information), allowing us to combine demographic, behavioral, and academic indicators. A binary target variable at-risk was defined: students who obtained certification were labeled 1 (not at risk), while those without certification were labeled 0 (at risk). Among the 471 students included, 73.67% achieved certification and 26.33% did not.

To ensure data quality, missing values were handled through imputation when possible and removal when necessary, and anomalies were detected using z-score thresholds. All data were anonymized before processing, and the study was in accordance with institutional privacy and ethical guidelines for the use of educational data.

These data constitute the foundation for the definition of learner profiles and for conducting a profile-based evaluation of the predictive system's performance.

C. Identification of Learning Indicators

The digital traces collected from the Loghate platform dataset combine both engagement data and academic performance scores. These elements were used to define two main learning indicators, calculated individually for each student:

- Engagement: measured by the total number of hours connected to the platform, reflecting the learner's level of involvement in their language learning activities.
- Performance: assessed through the results obtained in intermediate tests (listening, reading, grammar, vocabulary) as well as the final placement test, indicating the degree of mastery of the targeted competencies.

These two indicators were selected because an exploratory correlation analysis showed that engagement and performance variables were the strongest predictors of certification status, while also being the most consistently available and reliable across the dataset.

To account for individual differences, these absolute indicators were complemented with relative ones, calculated by comparing each student's value against the group average. This allows each learner to be positioned above or below the mean, providing a comparative perspective on their learning behavior.

Together, these absolute and relative indicators form the foundation for learner profile characterization and enable an exploration of how predictive model performance varies across the identified groups.

D. Identification of Typical Learner Profiles

To better understand the differentiated performance of the predictive system, students were grouped according to similar learning behaviors. This identification was based on two key indicators extracted from the Loghate platform: engagement (measured by the number of connection hours) and academic performance (scores from placement and intermediate tests).

Before clustering, the data were normalized to ensure fair comparisons among students. The K-means clustering algorithm was applied to form homogeneous groups, each representing a typical learner profile. K-means is a partitioning method that organizes individuals into clusters by minimizing within-group variance and maximizing separation between groups [16]. In practice, each student is assigned to the cluster whose centroid is closest, calculated on the basis of the engagement and performance indicators.

Each group was then represented as a persona, that is, a simplified profile designed to highlight and better visualize the dominant learning behaviors within the group. Table I provides examples of the identified profiles. This categorization makes it possible to conduct a more fine-grained analysis of the prediction system's performance by assessing its accuracy across different profiles, while also enabling more targeted pedagogical intervention strategies.

TABLE I. TYPOLOGY OF LEARNER PROFILES AND PEDAGOGICAL IMPLICATIONS

Learner profile	Behavior	Pedagogical implication
Highly engaged, low performance	Frequent logins but weak results	Requires targeted support despite strong effort
Low engagement, high performance	Limited participation but good achievement	Autonomous learner, no immediate follow-up needed
Low engagement, low performance	Rare logins and poor results	High-risk profile, urgent intervention recommended
Highly engaged, high performance	Strong commitment and good results	Autonomous learner, no alert required

IV. RESULTS AND DISCUSSION

This section presents the results obtained from applying the stacking model to the entire dataset, followed by a discussion of the observed performance across the different learner profiles identified through clustering. The aim is to examine the extent to which the model remains reliable and fair with respect to the learning behaviors detected.

A. Identification of Learner Profiles through Clustering

To gain a deeper understanding of learning behaviors within the Loghate platform, an unsupervised clustering approach was applied using engagement indicators (total connection time) and performance measures (overall test results). As detailed previously, the K-means algorithm identified four learner profiles based on engagement and performance indicators. These profiles represent distinct behavioral patterns in the Loghate platform. The distribution of students across these groups was:

- Profile 0 (high engagement, high performance): 164 learners
- Profile 1 (high engagement, low performance): 122 learners
- Profile 2 (low engagement, low performance): 77 learners
- Profile 3 (low engagement, high performance): 108 learners

These profiles are used as the basis for the differentiated evaluation of the predictive system. This segmentation goes beyond a global assessment of performance by revealing meaningful differences in learners' behaviors and effectiveness in online learning, providing a relevant foundation for evaluating the robustness of predictive models in a differentiated manner.

B. Overall Performance of the Stacking Model

The stacking model, which combined several supervised algorithms (Random Forest, Decision Tree, Support Vector Machine, Naive Bayes, Logistic Regression, and Artificial Neural Network), was tested on the entire dataset to predict at-risk students. These algorithms were selected for their complementarity; Random Forest and Decision Tree for their ability to handle heterogeneous data and their robustness to noise [17, 18], Naive Bayes for its simplicity and classification speed [19], Logistic Regression for its interpretability and efficiency in modeling linear relationships [20], SVM for its ability to separate classes even with complex decision boundaries [21], and Artificial Neural Networks (ANN) for their capacity to model more intricate non-linear relationships [22].

Stacking was chosen because it leverages the complementarity of these algorithms. By aggregating the strengths of linear, nonlinear, and probabilistic models, it reduces the biases of individual learners, improves the stability of predictions, and enhances generalization [23, 24]. The architecture follows a two-level design: at the first level, each base learner produces intermediate predictions, which are then consolidated at the second level by the meta-learner. SVM was selected as the meta-model due to its observed performance, strong generalization ability, and resilience to class imbalance through implicit weighting [25, 26]. As illustrated in Figure 2, this configuration optimizes the integration of the base models while mitigating their respective weaknesses.

To ensure robust evaluation, a stratified 10-fold cross-validation was applied, maintaining balanced class distributions in each fold. The results show an average precision of 81%, recall of 100%, F1-score of 89%, and accuracy of 82%. For completeness, the stacking ensemble was compared with each base model. The stacking method achieved the highest global performance (accuracy of 82%), slightly outperforming Random Forest (80%), Logistic Regression (81%), SVM (81%), Naive Bayes (73%), Decision Tree (72%), and the Artificial Neural Network (81%). This comparison confirms the added value of combining heterogeneous learners within a unified meta-model and the capacity of stacking to outperform individual models and effectively identify at-risk students.

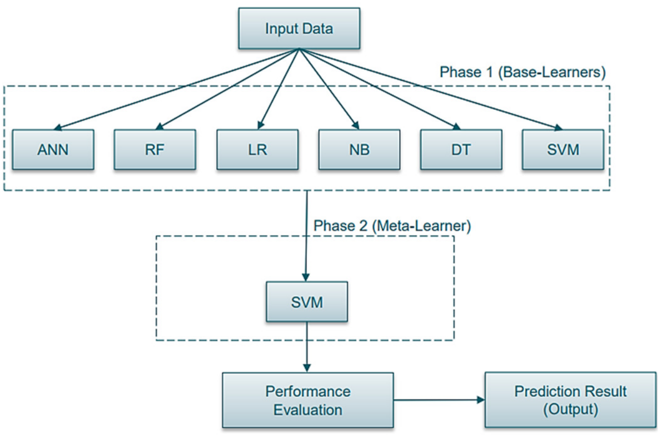

Fig. 2. Architecture of the stacking ensemble model used for prediction.

C. *Stacking Model Results by Profile*

In order to analyze the predictive system's performance more precisely, the stacking ensemble model was applied to each learner profile identified by the K-means clustering algorithm. Model performance was then evaluated within each of the previously defined profiles. For each group, a specific subset was extracted from the original dataset, and a stacking model was trained and evaluated separately. This approach makes it possible to observe performance variations according to the characteristics specific to each profile. The evaluation was carried out using standard metrics: accuracy, precision, recall, and F1-score.

TABLE II. MODEL PERFORMANCE BY LEARNER PROFILE

Profile	Accuracy	Precision	Recall	F1-Score
Profile 0	0.75	0.76	1.00	0.86
Profile 1	0.88	0.88	1.00	0.94
Profile 2	0.87	0.88	1.00	0.93
Profile 3	0.81	0.71	1.00	0.83

The results in Table II show that the model achieves generally satisfactory performance across all profiles. However, precision varies between profiles, with a relatively lower value for Profile 3 (0.71), which may suggest that the model has more difficulty correctly distinguishing classes in this group. In contrast, Profile 1 demonstrates particularly high performance, with a precision of 0.88 and an F1-score of 0.94, suggesting that the model performs optimally for this group.

Although recall reaches 100% across all profiles, this result should be interpreted with caution. The high recall is partly influenced by class imbalance (73% certified learners) and by the class-weighting strategy used in the SVM meta-learner, which prioritizes avoiding false negatives. This design choice ensures that at-risk students are not missed, but it may also increase false positives in some profiles. Moreover, this perfect recall should be interpreted as a sensitivity-oriented behavior rather than as evidence of perfect predictive accuracy. Such outcomes are common in early-warning systems trained on imbalanced data, where maximizing the detection of at-risk learners can slightly reduce specificity. This does not indicate model degradation but reflects a deliberate trade-off that favors minimizing false negatives while maintaining acceptable discrimination across profiles.

These trends are also illustrated in Figure 3, which visually compares the performance metrics for each profile. It can be observed that profiles 1 and 2 show consistently high results, while profiles 0 and 3 stand out with lower precision, confirming the previously mentioned limitations.

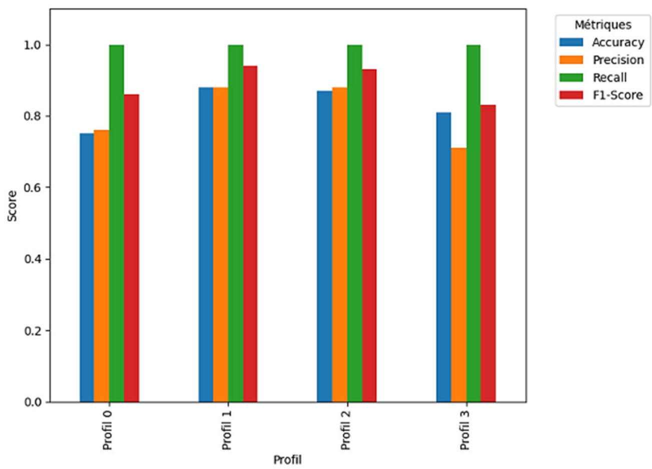

Fig. 3. Visualization of performance metrics by learner profile, highlighting variations in precision across groups.

D. *Discussion of Results and Recommendations*

The analysis of stacking model performance across learner profiles reveals notable disparities, underscoring the importance of differentiated evaluation. Although the model achieves generally satisfactory results on the overall dataset (average precision of 81%, recall of 100%, F1-score of 89%, and accuracy of 82%), a profile-level examination shows that some groups benefit from stronger predictions than others.

The learners in Profile 3 reflect patterns commonly described in Self-Regulated Learning (SRL) models, where students rely on prior knowledge or selective engagement strategies to maintain satisfactory performance. Their irregular interaction patterns reduce the predictive value of the engagement indicators, making their behavior harder for supervised models to capture. In contrast, students in Profile 1 invest substantial effort but struggle to progress, a pattern aligned with Cognitive Load Theory, which posits that excessive mental effort can impair learning when cognitive resources are overwhelmed. Such mismatches between effort and outcomes are difficult for machine learning models to distinguish from genuine improvement, contributing to the lower precision observed for these profiles.

Profiles 0 and 2 exhibit more consistent relationships between engagement and performance, which aligns with traditional behavioral learning patterns and makes them easier to classify. However, the lower accuracy observed for Profile 0 still highlights the model's sensitivity to atypical or mixed behaviors. More broadly, the variability observed across the four profiles reflects established behavioral and motivational models in educational psychology, where differences in persistence, autonomy, and engagement shape learning trajectories and influence predictive reliability.

It is important to note that the feature set used in this study is limited to engagement and performance indicators. Although these indicators are among the strongest predictors in this dataset, incorporating additional variables such as motivational, temporal, or interaction-based features could further improve the robustness and interpretability of the model across learner profiles.

These findings are consistent with recent fairness-focused studies in educational AI. In [27], predictive models were shown to often produce uneven performance across learner subgroups even when global metrics appear strong, reinforcing the need for bias-aware evaluation. Similarly, in [13], it was shown that profile-based assessment reveals hidden disparities that remain invisible in aggregated evaluations. The results of this study align with this literature, confirming that learner-profile analysis is a necessary step toward ensuring fair and trustworthy early-warning systems. To improve equity and reliability, this study recommends that:

- Alert thresholds should be adapted according to identified profiles in order to reduce false positives and false negatives.
- The model should be reinforced on minority or less accurately predicted profiles, for example, through oversampling or weighting techniques.
- Educators should be provided with profile-based, interpretable information to guide targeted pedagogical interventions.

In summary, integrating learner profiles into evaluation not only helps to better detect the limitations of current predictive models but also offers concrete pathways for designing early-warning systems that are more equitable and aligned with the actual needs of learners.

V. CONCLUSION

This study is part of a personalized evaluation approach for predictive systems aimed at identifying at-risk students in online learning. Unlike traditional global assessments, this study proposes a differentiated analysis method that considers the diversity of learning behaviors by identifying representative learner profiles, derived from data collected on the Loghate platform.

By combining engagement and performance indicators, four distinct profiles were identified using K-means clustering. A stacking model, applied both globally and within each profile, allowed the comparison of predictive performance, revealing significant variations. Although the overall results are satisfactory (accuracy of 82% and F1-score of 89%), some profiles—particularly those less represented or exhibiting atypical behaviors—show lower precision, pointing to potential biases in the predictions. These findings underscore the value of integrating learner profiles into early-warning systems to enhance prediction reliability and support more targeted interventions.

However, some limitations of this study should be acknowledged. First, the features extracted and used in the predictive model were relatively limited, which restricts the generalizability of the findings to other disciplines and educational contexts. Therefore, a more comprehensive predictive model that integrates a wider range of learner characteristics deserves further investigation.

Looking ahead, this methodology could be enhanced through the temporal analysis of learning behaviors and the integration of additional contextual factors (such as motivation and external constraints). Moreover, the provision of explainable results at the profile level represents a powerful lever to foster teachers' trust in these tools and promote a more equitable personalization of learner support. Beyond methodological contributions, this profile-based evaluation framework could be directly integrated into institutional learning dashboards or automated alert systems, helping universities deploy early-warning mechanisms that are both more transparent and better aligned with the heterogeneous needs of their students.

REFERENCES

[1] Y. H. Hu, C. L. Lo, and S. P. Shih, "Developing early warning systems to predict students' online learning performance," *Computers in Human Behavior*, vol. 36, pp. 469–478, July 2014, https://doi.org/10.1016/j.chb.2014.04.002.

[2] S. Rizvi, B. Rienties, and S. A. Khoja, "The role of demographics in online learning; A decision tree based approach," *Computers & Education*, vol. 137, pp. 32–47, Aug. 2019, https://doi.org/10.1016/j.compedu.2019.04.001.

[3] R. Al-Shabandar, A. J. Hussain, P. Liatsis, and R. Keight, "Detecting At-Risk Students With Early Interventions Using Machine Learning Techniques," *IEEE Access*, vol. 7, pp. 149464–149478, 2019, https://doi.org/10.1109/ACCESS.2019.2943351.

[4] H. Karalar, C. Kapucu, and H. Gürüler, "Predicting students at risk of academic failure using ensemble model during pandemic in a distance learning system," *International Journal of Educational Technology in Higher Education*, vol. 18, no. 1, Dec. 2021, Art. no. 63, https://doi.org/10.1186/s41239-021-00300-y.

[5] J. L. Hung, M. C. Wang, S. Wang, M. Abdelrasoul, Y. Li, and W. He, "Identifying At-Risk Students for Early Interventions—A Time-Series Clustering Approach," *IEEE Transactions on Emerging Topics in Computing*, vol. 5, no. 1, pp. 45–55, Jan. 2017, https://doi.org/10.1109/TETC.2015.2504239.

[6] R. Alshabandar, A. Hussain, R. Keight, A. Laws, and T. Baker, "The Application of Gaussian Mixture Models for the Identification of At-Risk Learners in Massive Open Online Courses," in *2018 IEEE Congress on Evolutionary Computation (CEC),* Rio de Janeiro, Brazi, July 2018, pp. 1–8, https://doi.org/10.1109/CEC.2018.8477770.

[7] B. Bouihi, A. Bousselham, E. Aoula, F. Ennibras, and A. Deraoui, "Prediction of Higher Education Student Dropout based on Regularized Regression Models," *Engineering, Technology & Applied Science Research*, vol. 14, no. 6, pp. 17811–17815, Dec. 2024, https://doi.org/10.48084/etasr.8644.

[8] D. Bañeres, M. E. Rodríguez, A. E. Guerrero-Roldán, and A. Karadeniz, "An Early Warning System to Detect At-Risk Students in Online Higher Education," *Applied Sciences*, vol. 10, no. 13, June 2020, Art. no. 4427, https://doi.org/10.3390/app10134427.

[9] M. Yağcı, "Educational data mining: prediction of students' academic performance using machine learning algorithms," *Smart Learning Environments*, vol. 9, no. 1, Dec. 2022, Art. no. 11, https://doi.org/10.1186/s40561-022-00192-z.

[10] J. B. Diekuu, M. S. Mekala, U. S. Abonie, J. Isaacs, and E. Elyan, "Predicting student next-term performance in degree programs using AI-based approach: a case study from Ghana," *Cogent Education*, vol. 12, no. 1, Dec. 2025, Art. no. 2481000, https://doi.org/10.1080/2331186X.2025.2481000.

[11] W. Holmes *et al.*, "Ethics of AI in Education: Towards a Community-Wide Framework," *International Journal of Artificial Intelligence in Education*, vol. 32, no. 3, pp. 504–526, Sept. 2022, https://doi.org/10.1007/s40593-021-00239-1.

[12] S. Slade and P. Prinsloo, "Learning Analytics: Ethical Issues and Dilemmas," *American Behavioral Scientist*, vol. 57, no. 10, pp. 1510–1529, Oct. 2013, https://doi.org/10.1177/0002764213479366.

[13] A. Ben Soussia, C. Treuillier, A. Roussanaly, and A. Boyer, "Learning Profiles to Assess Educational Prediction Systems," in *Artificial Intelligence in Education*, vol. 13355, M. M. Rodrigo, N. Matsuda, A. I. Cristea, and V. Dimitrova, Eds. Springer International Publishing, 2022, pp. 41–52.

[14] K. R. Premlatha, B. Dharani, and T. V. Geetha, "Dynamic learner profiling and automatic learner classification for adaptive e-learning environment," *Interactive Learning Environments*, vol. 24, no. 6, pp. 1054–1075, Aug. 2016, https://doi.org/10.1080/10494820.2014.948459.

[15] N. Zhang, G. Biswas, and Y. Dong, "Characterizing Students' Learning Behaviors Using Unsupervised Learning Methods," in *Artificial Intelligence in Education*, vol. 10331, E. André, R. Baker, X. Hu, Ma. M. T. Rodrigo, and B. Du Boulay, Eds. Springer International Publishing, 2017, pp. 430–441.

[16] M. Goyal, "k-mean clustering algorithm and its real use case in the security domain," *Medium*, Aug. 24, 2021. https://monilgoyal.medium.com/k-mean-clustering-algorithm-and-its-real-use-case-in-the-security-domain-fb97f23113a5.

[17] R. Katuwal, P. N. Suganthan, and L. Zhang, "Heterogeneous oblique random forest," *Pattern Recognition*, vol. 99, Mar. 2020, Art. no. 107078, https://doi.org/10.1016/j.patcog.2019.107078.

[18] T. Lan, H. Hu, C. Jiang, G. Yang, and Z. Zhao, "A comparative study of decision tree, random forest, and convolutional neural network for spread-F identification," *Advances in Space Research*, vol. 65, no. 8, pp. 2052–2061, Apr. 2020, https://doi.org/10.1016/j.asr.2020.01.036.

[19] I. Rish, "An empirical study of the naive Bayes classifier," in *Proceedings of the IJCAI 2001 Workshop on Empirical Methods in Artificial Intelligence*, 2001, pp. 41–46.

[20] N. Šarlija and M. Stanić, "Logistic regression modelling: procedures and pitfalls in developing and interpreting prediction models," *Croatian Operational Research Review*, vol. 8, no. 2, pp. 631–652, Dec. 2017, https://doi.org/10.17535/crorr.2017.0041.

[21] J. Hamidzadeh and S. Moslemnejad, "Identification of uncertainty and decision boundary for SVM classification training using belief function," *Applied Intelligence*, vol. 49, no. 6, pp. 2030–2045, June 2019, https://doi.org/10.1007/s10489-018-1374-0.

[22] M. J. Somers and J. C. Casal, "Using Artificial Neural Networks to Model Nonlinearity: The Case of the Job Satisfaction—Job Performance Relationship," *Organizational Research Methods*, vol. 12, no. 3, pp. 403–417, July 2009, https://doi.org/10.1177/1094428107309326.

[23] X. Chang, X. Jia, and C. Fan, "A probabilistic prediction model based on stacking ensemble learning for completion time in flexible shop-floor," *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture*, vol. 239, no. 10, pp. 1458–1471, Aug. 2025, https://doi.org/10.1177/09544054241277581.

[24] Ç. Demirel, "RocketStack: Level-aware deep recursive ensemble learning framework with adaptive feature fusion and model pruning dynamics." arXiv, 2025, https://doi.org/10.48550/ARXIV.2506.16965.

[25] R. Z. Pek, S. T. Ozyer, T. Elhage, T. Ozyer, and R. Alhajj, "The Role of Machine Learning in Identifying Students At-Risk and Minimizing Failure," *IEEE Access*, vol. 11, pp. 1224–1243, 2023, https://doi.org/10.1109/ACCESS.2022.3232984.

[26] L. A and S. M, "An Efficient Android Malware Detection Framework with Stacking Ensemble Model," *International Journal of Engineering Trends and Technology*, vol. 70, no. 4, pp. 294–302, Apr. 2022, https://doi.org/10.14445/22315381/IJETT-V70I4P226.

[27] G. Raftopoulos, G. Davrazos, and S. Kotsiantis, "Evaluating Fairness Strategies in Educational Data Mining: A Comparative Study of Bias Mitigation Techniques," *Electronics*, vol. 14, no. 9, May 2025, Art. no. 1856, https://doi.org/10.3390/electronics14091856.