Early Detection of Struggling Learners in Online Professional Training: A Data-Driven Approach

Mohamed Mouaici*
mouaici.med@gmail.com, mohamed.mouaici@logipro.com

***Abstract*— In online professional training, early identification of learners encountering difficulties is critical for enabling prompt, targeted pedagogical interventions. This study aims to develop a solution to predict at-risk learners at the earliest possible stage by leveraging comprehensive behavioral data extracted from a Learning Management System (LMS) over a seven-year period. We extract 12 quantitative behavioral indicators from detailed LMS logs, capturing diverse facets of learner engagement, such as time spent on activities, frequency of content interactions, and evaluation performance. By integrating these indicators, we develop a predictive framework that evaluates five different models. We conduct a misclassification analysis to identify and understand the sources of erroneous predictions. Based on these insights, we augment the feature set by introducing new indicators to better contextualize quiz-related interactions. Comprehensive hyperparameter optimization further refines the model parameters, balancing sensitivity and specificity effectively. Collectively, these steps significantly improve the model's robustness, culminating in an optimized Random Forest classifier that achieves an F1 score of 82 %.**

I. Introduction

Professional learning encompasses all activities through which individuals acquire, refine, and apply skills and knowledge in their careers. It includes formal courses, on-the-job experiences, and peer exchanges that together drive workforce growth and adaptability [1]. The rise of digital learning technologies has transformed professional learning, enabling employees and job seekers to develop new skills remotely and at their own pace [2]. Online professional training plays a critical role in workforce development, providing continuous education opportunities across various industries [2]–[5]. Learning Management Systems (LMS) are widely used to deliver these programs, offering structured learning experiences through digital platforms. Despite these advances, ensuring the effectiveness of online training remains challenging, particularly when it comes to identifying learners who struggle and require additional support.

In this context, our research is conducted in collaboration with a French training organization that specializes in online professional training across multiple sectors. The organization delivers its courses through an LMS, catering primarily to adult learners, including employees and job seekers. To maintain the quality of its programs, the organization relies on a dedicated pedagogical team composed of educators—who design the training content—and instructional designers—who implement the programs within the LMS. Together, they support learners throughout their training, ensuring that educational objectives are reached.

The stakes in professional training are high, as these programs are funded by various entities such as the State, regional authorities, companies, joint collector organizations, and employment agencies. Consequently, learners must adhere to strict requirements, including mandatory program completion and minimum training durations corresponding to the allocated funding. Failure to meet these standards can lead to significant financial penalties, such as the reimbursement of training costs or the loss of unemployment benefits, and learners must comply fully to obtain their training certificates.

Given the high stakes involved in professional training, it is critical to detect learners who may be at risk of not meeting these requirements early on. Therefore, the early identification of learners facing difficulties in online professional training is essential for ensuring effective support and improving learning outcomes. However, as the number of participants and learning activities in such programs grows, manual detection of struggling learners becomes increasingly complex and resource-intensive.

This paper addresses this challenge by proposing an analytical approach to predict learners in difficulty as early as possible, using data collected from an LMS in an online professional learning context.

II. Related Work and Contributions

Professional learning spans a continuum from structured training programs to informal, on-the-job learning and peer interactions [2]–[5]. Over the past decade, Learning Analytics (LA) techniques have been increasingly adopted to support and enhance professional development. Much of this work has focused on the social dimensions of learning—often termed Social Learning Analytics (SLA)—and on the analysis of social networks in professional settings, such as eTwinning communities [2]–[6]. SLA studies typically examine interactions, collaboration patterns, and knowledge sharing, but they pay less attention to the formal, course-based contexts in which many organizations deliver continuing professional education.

Only recently have researchers begun to target formal professional training programs. Mouaici et al. [7] investigate early identification of learners at risk of failure in online professional training. They use a dataset of 13,719 observations from 912 learners across 182 LMS-delivered courses between 2017 and 2022, define failure risk in three levels—low, medium, and high—and evaluate nine predictive models. To boost prediction accuracy, they introduce a novel ensemble based on the Borda count method. This ensemble yields an average F1-score improvement of 1.2% over the baseline models, demonstrating both the promise and the challenges of enhancing predictive performance in this domain. Their work is one of the few that specifically targets the formal online professional training context and focuses on detecting learners who are struggling—a comparatively underexplored area in Learning Analytics.

To broaden the scope, studies from broader online learning environments are instructive [8]. Using the UCI Machine Learning Repository, one investigation compared 14 tree- and rule-based classifiers for student performance prediction, applying feature selection and SMOTE oversampling; Random Forest emerged as the top performer in accuracy and stability [9]. Another study combined LMS interaction logs, internet access records, and demographic attributes, selected features via Pearson correlation, and tested five classifiers—decision trees, Naive Bayes, logistic regression, Random Forest, and neural networks—identifying Random Forest as most effective at predicting at-risk learners [10]. Additionally, research in [11] has focused on identifying meaningful behavioral indicators from LMS data to predict success in online courses. Through correlation and hierarchical regression analyses, the research examined the relationship between behavioral indicators—such as regular study habits, late assignment submissions, the number of sessions, and engagement with course materials—and final course scores, confirming that these indicators are significant predictors of course achievement.

Collectively, these studies underscore the potential of LA techniques for monitoring learner performance in online environments while highlighting the need for further research within formal professional training contexts. Indeed, despite this body of work, three gaps remain. Few studies target structured competency-based professional courses. Iterative model refinement techniques—such as misclassification analysis and targeted feature engineering—are not enough explored. Methods for detecting at-risk learners at the earliest possible stage, before significant disengagement in professional learning contexts, are lacking. Therefore, this paper aims to address these gaps by:

- Conducting a benchmark of classification methods applied to data collected from a Learning Management System within a formal professional e-learning context, providing a comparative evaluation of their effectiveness to predict at-risk learners as early as possible.
- Implementing an incremental optimization approach through misclassification analysis, feature engineering, and hyperparameter tuning to refine predictive models and improve their accuracy.

III. Learning Context and Predictive Data Modeling

A. E-Learning Activity Structure and Design

Training in our context is delivered through a series of e-learning activities designed according to the SCORM standard [12]. Each training program is designed to address a specific set of skills. For each target skill, the training organization provides a diverse range of e-learning activities, including video, textual, and interactive content. This approach allows learners to choose from the available activities based on their individual preferences and learning styles, meaning that even for the same target skill, learners may follow different learning paths. Some e-learning activities incorporate quiz evaluations, while others do not. When quizzes are included, they serve as formative assessments intended to provide feedback and reinforce learning rather than to act as gatekeepers. In these cases, successful quiz completion is not required for learners to progress through the training. Finally, each e-learning activity is associated with **a minimum required duration** that ensures the learner accumulates the total training time mandated by the course.

B. Definition of Learners in Difficulty

While learners have the option to exceed the minimum required duration of an e-learning activity (to consolidate their understanding), the pedagogical team has established that **exceeding twice the required duration is indicative of potential difficulty**. In other words, if a learner spends at least twice the prescribed time on an activity, it is likely that they are struggling with one or more of the concepts, which may hinder the acquisition of the targeted skills.

C. Learners Behavior Indicators from the LMS

Learning traces are extracted from the LMS employed by the training organization, with data collected continuously from January 2017 to December 2023. The extraction process involves interfacing with the LMS database and retrieving comprehensive log files that capture every user interaction during e-learning activities. To ensure the quality and reliability of the data, a dedicated pre-processing tool is developed. This tool performs key functions: it extracts raw log data, filters out system events and incomplete records, and standardizes timestamps for consistency. Additionally, it converts time metrics into seconds, enabling the calculation of continuous variables. From the pre-processed traces, twelve indicators are derived to capture learners' behavior:

- I_1: Total time spent by the learner on an e-learning activity.
- I_2: Total number of visualizations of the activity's content.
- I_3: Total number of days during which the learner performed an action on the activity content.
- I_4: Average time spent per day on the activity.
- I_5: Average time spent on each individual content.
- I_6: Score obtained in the evaluation of the activity.
- I_7: Total number of attempts made in the quiz evaluation.
- I_8: Total time spent on the quiz evaluation.
- I_9: Average time per attempt during the quiz evaluation.
- I_{10}: Number of days during which the learner was active.
- I_{11}: Standard deviation of daily time spent on the activity.
- I_{12}: Total number of revisits to the activity content.

All these indicators are continuous variables, and the time-related metrics are uniformly expressed in seconds. These indicators form the foundation for our predictive modeling efforts aimed at early detection of learners in difficulty.

D. Prediction Problem Definition

The goal of this research is to develop a solution that predicts learners at risk of difficulty during their e-learning activities. Predictions rely on the behavioral indicators described above.

In discussion with the pedagogical team, we define the prediction problem by:

- **Indicator set (I)**: the twelve indicators identified and validated with the pedagogical team (see Section C).
- **Difficulty classification (Y)**: learners are deemed **"in difficulty (Y = 1)"** if they exceed **twice the required duration** of an e-learning activity; otherwise, they are considered **"not in difficulty (Y = 0)"**.
- **Learning Duration Threshold (LDT)**: Predictions are made at three different points relative to the required activity duration: 60%, 80%, and 100%. These thresholds allow for early detection of struggling learners before they reach critical learning delays (twice the required duration). For example, in an activity with a required duration of 5 hours, predictions would be made after 3 hours (60%), 4 hours (80%), and 5 hours (100%) to detect potential learning difficulties.

Therefore, the primary objective is to accurately predict learner in difficulty (Y) based on the set of 12 behavioral indicators (I). To enable early detection, predictions are experimented with data extracted at 60%, 80%, and 100% of the required duration of each e-learning activity to determine the optimal threshold (LDT) for timely intervention.

E. Dataset Description

The predictive models in this study rely on the previous 12 behavioral indicators. These indicators are computed from LMS data collected over a seven-year period (January 2017 to December 2023). To facilitate model training and evaluation, the collected dataset is divided into two subsets:

- **Training and testing dataset (DS1):** Data collected between **January 2017 and December 2022** are used to build and test predictive models.
- **Final evaluation dataset (DS2):** Data from **January to December 2023** are reserved to evaluate the model's generalization on unseen data.

TABLE 1 summarizes these subsets in terms of the number of learners, number of e-learning activities, total observations, and the percentage of learners in difficulty.

TABLE 1: DATASETS DESCRIPTION

Dataset	Learner	activity	Observation	%difficulty
DS1 (Build & Test Models)	407	377	15,360	12.45%
DS2 (Final Evaluation)	78	102	4,750	13.72%

It is important to note that DS1 and DS2 are constructed separately for each Learning Duration Threshold (LDT) by filtering the LMS data based on timestamps. As a result, distinct datasets are created with the 12 indicators calculated at LDT= 60%, LDT= 80%, and LDT = 100%, each containing the same number of observations as indicated in TABLE 1. The percentage of learners in difficulty remains unchanged across LDTs, as difficulty is defined retrospectively based on the total time spent on an activity, ensuring consistency in class distribution across datasets.

IV. IN-DIFFICULTY LEARNER PREDICTION

Given that our target variable (Y) is binary, several supervised classification approaches are suitable for predicting learners in difficulty. In this study, we focus on five well-established binary classification algorithms selected for their interpretability and their ability to assess the contribution of each indicator. The models under consideration include:

- Logistic Regression (**LR**)
- Decision Trees (**DT**)
- Random Forest (**RF**)
- Extreme Gradient Boosting (**XGBoost**)
- Artificial Neural Networks (**ANNs**)

The choice of these models is driven by the pedagogical team's requirement for explicit insights into the most influential predictors. Tree-based models, such as DT, RF, and XGBoost, inherently provide feature importance measures that help clarify the decision process. Meanwhile, Logistic Regression offers a probabilistic interpretation of predictor impact, and ANNs excel at modeling complex, non-linear relationships in the data.

A. Predictive Performance Metrics

To assess the performance of the selected models, we construct a confusion matrix that compares the models' predictions with the actual outcomes. Then, this matrix is used to evaluate predictions based on key performance metrics:

- **Accuracy**: Overall correctness calculating the proportion of correctly classified instances among all predictions.
- **Precision**: Proportion of correctly identified learners in difficulty among all those classified as in difficulty by the model. It reflects the reliability of positive predictions.
- **Recall** (Sensitivity): Proportion of actual learners in difficulty that were correctly identified by the model. It measures the model's ability to detect at-risk learners.
- **F1-score:** A harmonic mean of precision and recall, providing a single metric that balances both aspects.

In this study, special emphasis is placed on the F1-score, as the training dataset DS1 is imbalanced, with significantly fewer learners in difficulty (12.45%). In this case, accuracy alone can be misleading, as a model could achieve high accuracy simply by predicting the majority class (learners not in difficulty) while failing to detect learners in difficulty. The F1-score provides a more meaningful evaluation by ensuring that both precision and recall are considered, making it particularly well-suited for handling class imbalances and optimizing the detection of learners in difficulty.

B. Oversampling and Cross Validation with Stratification

The training dataset (DS1) is highly imbalanced, with only 12.45% of learners classified as in difficulty. This imbalance poses a challenge for predictive models, as they may become biased toward the majority class, leading to poor detection of learners in difficulty. To address this issue, SMOTE (Synthetic Minority Oversampling Technique) is applied to rebalance DS1 before training the models. SMOTE [13] generates synthetic instances in the minority class, ensuring an equal number of observations in each class, thereby improving model performance and reducing bias.

Then, to ensure model reliability, stratified k-fold cross-validation is applied to the training dataset (DS1) after oversampling with SMOTE, preserving the class distribution across subsets [13]. At each LDT (with the DS1 extracted at that LDT) models are trained k times, each time using k–1 subsets for training and the remaining subset for validation, which enhances the assessment of generalization performance. Given the variability of cross-validation, a repeated cross-validation approach is employed, where performance metrics are averaged over multiple runs to improve stability.

C. *First Results and Analysis*

As mentioned earlier, separate training datasets (DS1) are extracted at each Learning Duration Threshold (LDT = 60%, 80%, and 100%) based on the timestamps in the LMS data. The predictive models are trained on the oversampled version of DS1 (DS1-OS) at each threshold using repeated stratified k-fold cross-validation. After training, the models are evaluated on the corresponding DS2 dataset extracted at the same LDT without oversampling. TABLE 2 summarizes the evaluation results for the five models obtained on DS2.

TABLE 2: MODELS PERFORMANCE EVALUATION ON DS2. MODELS BUILT USING K-FOLDS CROSS VALIDATION REPEATED 10 TIMES WITH K = 3.

LDT	Model	Accuracy	Precision	Recall	F1 Score
60	Logistic Regression	0.65	0.50	0.30	0.38
	Decision Tree	0.72	0.55	0.48	0.51
	Random Forest	0.78	0.62	0.70	**0.66**
	XGBoost	0.75	0.60	0.65	0.62
	ANN	0.73	0.58	0.63	0.60
80	Logistic Regression	0.67	0.52	0.35	0.42
	Decision Tree	0.74	0.57	0.50	0.53
	Random Forest	0.80	0.65	0.74	**0.69**
	XGBoost	0.77	0.62	0.68	0.65
	ANN	0.75	0.60	0.66	0.63
100	Logistic Regression	0.69	0.55	0.38	0.45
	Decision Tree	0.76	0.60	0.53	0.56
	Random Forest	0.82	0.68	0.76	**0.72**
	XGBoost	0.79	0.65	0.70	0.67
	ANN	0.77	0.63	0.71	0.67

The evaluation results reveal several important trends across the different learning duration thresholds. At LDT = 60%, all models exhibit lower performance metrics compared to those at higher thresholds. This is likely because, at 60% of the required duration, the behavioral indicators have not yet fully matured, providing less discriminative information for predicting learner difficulty. As a result, distinguishing between at-risk and non-at-risk learners is more challenging at this early stage, which is reflected in lower recall and F1 scores. Notably, among the models tested at LDT = 60%, Random Forest stands out as the best performer, achieving an F1 score of 0.66. In contrast, Logistic Regression consistently struggles, with much lower recall and overall F1 scores, highlighting its limitations in capturing the complex, non-linear patterns present in the learner behavior data.

At LDT = 80%, the Random Forest model's performance further improves, benefiting from the extended learning duration that allows it to capture more distinctive features associated with learner difficulty. While both XGBoost and the Artificial Neural Network models also demonstrate competitive performance—with robust accuracy and balanced precision and recall—they still fall slightly short of the superior results achieved by the Random Forest model with its F1 score of 0.69.

At LDT = 100%, the trend remains consistent, with Random Forest continuing to outperform the other models. Its accuracy, precision, and recall are superior, resulting in the highest F1 score (0.72). Although XGBoost and ANN yield promising results, they do not match the performance of Random Forest, which more effectively captures the nuanced patterns indicative of learner difficulties. At this threshold, the behavioral indicators have matured, providing a clearer distinction between the learners' behaviors and enhancing the model's performance.

V. PREDICTIONS IMPROVEMENTS

Since our goal is to predict learners in difficulty as early as possible and Random Forest already performs well, the next step is to further enhance its performance. We considered three potential improvement strategies: expanding the training dataset, enriching the feature set with additional indicators, and analyzing misclassifications to devise targeted solutions. Due to limitations in the available LMS data, we focused exclusively on misclassification analysis.

A. *Misclassification Analysis and Indicators Augmentation*

To enhance the model's predictive performance, we conduct an in-depth analysis of misclassifications, focusing on observations where learners were incorrectly flagged as "in difficulty". Using the Random Forest model built at LDT= 100% as a reference, we examine the decision rules and computed feature importance via the Gini index. This analysis reveals that several quiz-related indicators, particularly the total time spent on the quiz (I8), are disproportionately influencing predictions—even though quizzes are optional and not present in every e-learning activity. When a quiz is absent or not attempted, the corresponding features are either null or imputed with zero values, which can mislead the model.

To address this issue, we propose a feature augmentation strategy that introduces two binary indicators to better contextualize quiz-related data. The first indicator, I13 ("**hasQuiz**"), is set to 1 if the activity includes a quiz and 0 otherwise, allowing the model to differentiate between activities with and without quizzes. The second indicator, I14 ("**quizAttempted**"), reflects whether a learner attempted the quiz (1 if attempted, 0 if not). This dual approach enables the model to more accurately interpret quiz-related features—recognizing when they are not applicable or when a lack of attempt should be considered a signal of difficulty.

We then rebuild the random forest model at the three learning duration thresholds (LDT =60%, 80%, and 100%) using the oversampled training datasets at each LDT with these enhanced indicators. The evaluation of the new model on the independent dataset (DS2) shows a marked improvement, with reduced false positives and an overall enhancement in the F1 score. TABLE 3 presents the evaluation results and highlights the five most important indicators from the new model at each LDT.

TABLE 3: RANDOM FOREST RESULTS AFTER ADDING I13 AND I14

LDT	RF evaluation results				Best five indicators
(%)	Acc.	Pre.	Rec.	F1	(Gini index)
60	0.83	0.72	0.78	0.75	I2,I1,I4,I5,I8
80	0.85	0.75	0.81	0.78	I2,I5,I1,I4,**I13**
100	0.87	0.78	0.83	**0.80**	I2,I1,**I13,I14**,I8

The new RF model shows a steady improvement as the LDT increases. At LDT = 60%, the model achieves an F1 score of 0.75, marking a significant increase from the 0.66 obtained without I_{13} and I_{14}. At LDT = 80%, performance further improves, reaching an F1 score of 0.78, compared to 0.69 before incorporating the new indicators. The highest performance is observed at LDT= 100%, where the model attains an F1 score of 0.80—an 8% improvement over the previous score of 0.72 (without I_{13} and I_{14}).

These improvements highlight the effectiveness of the additional quiz-related indicators and our refined misclassification analysis in boosting the model's overall predictive ability. Indeed, the enhanced model's performance can be largely attributed to the integration of both traditional activity indicators and the newly introduced quiz-related features. At LDT= 80%, the importance of I13 ("hasQuiz") becomes evident, ranking among the top predictors alongside I2, I5, I1, and I4. The inclusion of I13 allows the model to distinguish between activities that include a quiz and those that do not. This differentiation is critical because, at this intermediate threshold, the presence or absence of a quiz significantly impacts the interpretation of related engagement metrics. When I13 indicates that a quiz is available, the model adjusts its expectations regarding quiz-related behaviors, leading to a more accurate assessment of learner difficulty.

At LDT=100%, both I13 ("hasQuiz") and I14 ("quizAttempted") emerge as key indicators, alongside I2, I1, and I8. At this stage, the model benefits from a clearer picture of the learner's interaction with the quiz component. I13 confirms the existence of a quiz, while I14 reveals whether the learner has engaged with it. This combined insight enables the model to more accurately differentiate between learners who are struggling and those who are not, as a lack of quiz attempts in an activity with a quiz can be a strong indicator of disengagement or difficulty.

Overall, the introduction of I13 and I14 refines the model's understanding of quiz-related behaviors, particularly as learners progress further into the activity. Their inclusion helps to reduce misclassification by providing contextual clarity, ultimately contributing to an improvement of 7% to 10% in performance metrics.

B. Hyperparameter Tuning and Model Optimization

To further enhance prediction performance, we refine the Random Forest model by optimizing its hyperparameters using the oversampled training datasets. We employ a two-step tuning approach that combines an initial random search with a subsequent grid search refinement for each hyperparameter. This strategy efficiently explores the hyperparameter space and converges on the best configuration based on cross-validated F1 scores. Our optimization focuses on the key hyperparameters of the Random Forest model [14]:

- **NT (Number of Trees):** Tested over a range from 250 to 2500. Our analysis revealed that as the LDT increases, the model benefits from a larger forest.
- **NI (Number of Indicators):** Evaluated for values between 2 and 14 (number of indicators after adding I13 and I14). The optimal value consistently converged to 5.
- **DT (Tree Depth):** We explored depths from 1 to 30, including an unbounded ("max") option. In all cases, allowing the trees to grow to their maximum depth yielded better performance.
- **SZ (Sample Size):** Ranged from 500 to 4500, with 500 emerging as optimal, indicating that a smaller sample size per split suffices for maximizing node purity.

The tuning process is carried out separately at three LDTs, resulting in the following optimal settings:

- LDT = 60%: {NT = 1000, NI = 4, DT = max, SZ = 500}
- LDT = 80%: {NT = 1250, NI = 4, DT = max, SZ = 500}
- LDT = 100%: {NT = 1750, NI = 4, DT = max, SZ = 500}

These results indicate that while the optimal values for NI, DT, and SZ remain constant, the optimal number of trees (NT) increases as the behavioral indicators mature at higher LDTs. This suggests that a larger ensemble is needed to balance the voting process and capture the complete complexity of learner behavior at higher LDTs.

Following hyperparameter tuning, the optimized Random Forest models are evaluated on the independent test dataset (DS2 without oversampling), demonstrating an approximate 3% improvement across key performance metrics. TABLE 4 summarizes the improved results for each LDT, along with the top five indicators ranked by the Gini index.

TABLE 4: EVALUATION RESULTS AFTER TUNING HYPERPARAMETERS

LDT	RF evaluation results				Best five indicators
(%)	Acc.	Pre.	Rec.	F1	(Gini index)
60	0.82	0.77	0.79	0.78	I2,I5,I1,I4, I8
80	0.85	0.76	0.85	0.80	I2,I5,I1,**I13, I14**
100	0.84	0.85	0.80	**0.82**	I1,**I13,I14**, I5, I8

The improvements in performance—reflected in higher accuracy, precision, recall, and F1 scores—validate the efficacy of our tuning approach. The F1 score is improved by 3 % at LDT= 60 % and by 2% at LDT= 80% and LDT=100%.

These results underscore the importance of both thorough hyperparameter optimization and the integration of context-specific features quiz-related indicators in enhancing the predictions of our Random Forest model.

C. *Discussion and Final Model Choice*

Our experiments demonstrate that the Random Forest model is highly effective in predicting learners in difficulty within the online professional training context, particularly after applying the proposed iterative optimization approach, which includes misclassification analysis, feature engineering, and hyperparameter tuning. Prediction performance improves consistently across the Learning Duration Thresholds (LDTs). At LDT = 100%, the model achieves an accuracy of 0.84, a precision of 0.85, a recall of 0.80, and an F1 score of 0.82—outperforming all other configurations. While the results at LDT= 80% are promising, the superior performance and more stable balance between precision and recall at LDT = 100% led us **to select this threshold as the final configuration for deployment**. This means that predictions will be generated once a learner **reaches the full mandatory duration** of an e-learning activity (at LDT = 100%).

VI. Conclusion

This study focuses on the identification of learners in difficulty within online professional training, a key factor in enabling timely and targeted pedagogical interventions. Conducted in collaboration with a French training organization, our work focuses on leveraging Learning Analytics to enhance learner support. To this end, we propose a comprehensive predictive approach that integrates machine learning models, feature engineering, and hyperparameter tuning to improve the detection of learners at risk.

A benchmark of classification models is performed, with Random Forest emerging as the best-performing model. Through an iterative optimization process, we refine the model using misclassification analysis and introduce new indicators to better account for quiz-related interactions. Further hyperparameter tuning enhances model performance, leading to a reliable and interpretable classification system. The final model, selected to generate predictions at LDT = 100% (when learners reach the full mandatory duration of an e-learning activity) demonstrates strong predictive capabilities with an F1 score of 0.82, making it well-suited for early intervention strategies in professional e-learning environments. Despite these advances, our work has limitations. First, the analysis is confined to one organization's LMS data, which may limit generalizability. Second, we rely on predefined behavioral indicators and may miss other informative signals. Third, our approach predicts at fixed LDT and does not adapt to individual learner pacing.

Future work will integrate additional behavioral metrics, such as temporal engagement patterns, to refine predictions. We will also explore deep learning techniques to improve model robustness and adaptability.

VII. Acknowledgement

This work was financed by the company Logipro (Le Puy-en-Velay, France). We gratefully thank them for their support.

References

[1] M. Mouaici, "Aide au suivi des apprenants dans le contexte de la formation professionnelle en ligne : méthodes, algorithmes et outils," Thèse CIFRE, SYMME-USMB 2020.

[2] M. Lambert, J. Vero, and B. Zimmermann, "Vocational training and professional development: A capability perspective," *Int. J. Train. Dev.*, vol. 16, no. 3, pp. 164–182, 2012.

[3] M. Mouaici, L. Vignollet, C. Galez, and M. Etienne, "Learning analytics dashboards for professional training - Challenges and proposal," in *CEUR Workshop Proceedings*, 2018, vol. 2294.

[4] A. Littlejohn, "Learning and Work: Professional Learning Analytics," in *Handbook of Learning Analytics*, Society for Learning Analytics Research (SoLAR), 2017, pp. 269–277.

[5] A. Ruiz-Calleja, L. P. Prieto, T. Ley, M. J. Rodríguez-Triana, and S. Dennerlein, "Learning Analytics for Professional and Workplace Learning: A Literature Review," Springer, Cham, 2017, pp. 164–178.

[6] C. Ping, G. Schellings, and D. Beijaard, "Teacher educators' professional learning: A literature review," *Teach. Teach. Educ.*, vol. 75, pp. 93–104, Oct. 2018.

[7] M. Mouaici, "Early Prediction of Learners At-Risk of Failure in Online Professional Training Using a Weighted Vote," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2023, vol. 14200 LNCS, pp. 246–262.

[8] A. ; Alhothali *et al.*, "Predicting Student Outcomes in Online Courses Using Machine Learning Techniques: A Review," *Sustain. 2022, Vol. 14, Page 6199*, vol. 14, no. 10, p. 6199, May 2022.

[9] S. Senthil and W. M. Lin, "Applying classification techniques to predict students' academic results," in *2017 IEEE International Conference on Current Trends in Advanced Computing, ICCTAC 2017*, 2017, vol. 2018-Janua, pp. 1–6.

[10] R. Trakunphutthirak, Y. Cheung, and V. Lee, "Detecting student at risk of failure: A case study of conceptualizing mining from internet access log files," in *IEEE International Conference on Data Mining Workshops, ICDMW*, 2019, vol. 2018-Novem, pp. 365–371.

[11] J. W. You, "Identifying significant indicators using LMS data to predict course achievement in online learning," *Internet High. Educ.*, vol. 29, pp. 23–30, Apr. 2016.

[12] "What is SCORM and How it Works." [Online]. Available: https://scorm.com/. [Accessed: 05-Apr-2023].

[13] E. Ramentol, Y. Caballero, R. Bello, and F. Herrera, "SMOTE-RSB *: A hybrid preprocessing approach based on oversampling and undersampling for high imbalanced data-sets using SMOTE and rough sets theory," *Knowl. Inf. Syst.*, vol. 33, no. 2, pp. 245–265, Nov. 2012.

[14] L. Breiman, "Random Forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.