M4: Engagement tracking

Omar Boudarka, Valentine Casalta, Nino Gerber

1 Research question

Online learning platforms face significant challenges in keeping learners engaged over time. In this project, we investigate the question: *How does user behavior influence retention rates on the Lernnavi platform?*
We hypothesize that students' engagement can be predicted using behavioral signals reflecting effort, self-regulation, and performance.
Specifically, we aim to explore how various dimensions of student behavior—such as the amount of effort invested, the consistency and regularity of study habits, the frequency of proactive learning actions, the degree of control over the learning process, and assessment performance—affect students' continued use and engagement with the platform.

Our approach builds on prior work in learning analytics that emphasizes multi-dimensional modeling of student behavior. Mejia-Domenzain et al. (1) introduced an interpretable profiling framework across six dimensions of self-regulated learning, which we adapt to our weekly engagement context. Urrutia Cordero et al. (2) and Khalil et al. (3) have demonstrated the value of visualizing learner profiles and modeling future engagement, respectively. We extend these ideas by combining behavior profiling with next-week engagement forecasting.

2 Methodology

2.1 Feature engineering

To prepare the dataset for analysis, we aggregated event and transaction logs by calendar week, creating a weekly granularity for each user. Missing values were filled with zeros under the assumption that no recorded activity indicates genuine inactivity during that week.
Following the framework introduced by Mejia-Domenzain et al. (1), we organized student behavior into six interpretable dimensions of self-regulated learning summarized in 1. The features were tailored to reflect indicators of skill proficiency and engagement patterns relevant to our research question.
Users with insufficient data—fewer than 10 recorded events or activity in less than two weeks—were excluded from further analysis.

Dimension	Feature	Description
Effort	Weekly events	Total number of platform actions
	Weekly clicks	Number of interaction clicks
	Time spent	Total minutes spent online
Consistency	Mean session duration	Avg. length of user sessions
	Std of activity	Std. dev. of weekly events
Regularity	Days between sessions	Avg. gap between study days
Proactivity	Go to theory	Visits to theory pages
	Early sessions	Sessions before 8 AM
Control	Next freq.	Frequency of clicking "next"
	Skip freq.	Frequency of skipped tasks
Assessment	% Correct	Share of correct answers
	Challenges done	Completed challenges

Table 1: Behavioral features grouped by self-regulated learning dimensions.

2.2 Engagement labeling

To capture user engagement, we defined a set of *meaningful actions* on the platform: `SUBMIT_ANSWER`, `REVIEW_TASK`, `NEXT`, `SKIP`, and `GO_TO_THEORY`. These events reflect both task completion and active navigation behavior. For each user-week pair, we assigned a binary label, $engaged\_next\_week$, indicating whether the user performed at least one meaningful action in the following week $(t + 1)$.
The labeling process used ISO weeks as a time index and was implemented via a self-join on the activity table, shifting the week forward for engagement target generation.

2.3 Behavioral profiling

To enhance the interpretability of behavioral dynamics, we clustered users within each self-regulated learning dimension—Effort, Consistency, Regularity, Proactivity, Control, and Assessment—using K-Means ($k = 3$) applied separately to the respective weekly features.
Clustering was conducted offline on training weeks only (up to week `2022-09`) to prevent label leakage. Features were standardized using `StandardScaler`, and centroids were ordered post hoc to produce ordinal profile labels: *Low*, *Mid*, and *High*. This profiling proce-

dure was stored using `joblib` and repeated once per semester.
During feature generation, each week's behavioral vector was transformed using the saved scalers and assigned to its nearest cluster centroid for each dimension. The resulting profile labels were one-hot encoded and appended to the main feature matrix X.

2.4 Modeling

We used a time-aware cross-validation strategy (`TimeSeriesSplit`) to ensure realistic evaluation that respects the temporal nature of user behavior. We evaluated several classification models, each offering strengths well-suited to our engagement prediction task:

- **Logistic Regression**: Serves as a strong baseline and is useful for identifying linear trends between behavioral features and future engagement.
- **Random Forest**: Captures complex, non-linear relationships in user behavior and is robust to noise—ideal for heterogeneous activity logs.
- **Gradient Boosting**: Builds on previous mistakes to improve predictions, which is effective for subtle patterns in weekly engagement signals.
- **LightGBM**: Efficient for large datasets and high-dimensional behavioral features, making it suitable for rapid experimentation and tuning.
- **CatBoost**: Especially effective with categorical data and works well even with limited preprocessing—helpful for behavioral labels like profile clusters.

Ablation studies compared performance with and without behavioral profile features. Feature importance was analyzed using gain-based metrics and SHAP values to interpret model decisions and understand which behaviors most strongly predict future engagement.

3 Results and Insights

Across all models, tree-based methods consistently outperformed baselines. The best-performing model—Gradient Boosting without profile features—achieved an ROC-AUC of 0.752 and an F1 score of 0.55. Including profile features did not lead to a significant performance improvement, yielding an ROC-AUC of 0.751 and an F1 score of 0.54. However, the addition of profile features did not degrade performance and provided interpretable groupings of learner behaviors. Performance metrics remained stable across cross-validation folds, suggesting that engagement patterns are consistent over time and generalizable.

Features from the *Effort* dimension emerged as dominant predictors. Notably, the feature

Model	Variant	AUC	F1 Score
Gradient Boosting	no_profile	0.7520	0.5506
Gradient Boosting	with_profile	0.7515	0.5452
Random Forest	no_profile	0.7498	0.5568
Random Forest	with_profile	0.7488	0.5558
CatBoost	no_profile	0.7473	0.5646
CatBoost	with_profile	0.7472	0.5627
LightGBM	with_profile	0.7387	0.5786
LightGBM	no_profile	0.7378	0.5786
Logistic Regression	no_profile	0.5793	0.4714
Logistic Regression	with_profile	0.5728	0.4661
Dummy	with_profile	0.5000	0.0000
Dummy	no_profile	0.5000	0.0000

Table 2: Performance comparison of models with and without profile features (best scores in bold).

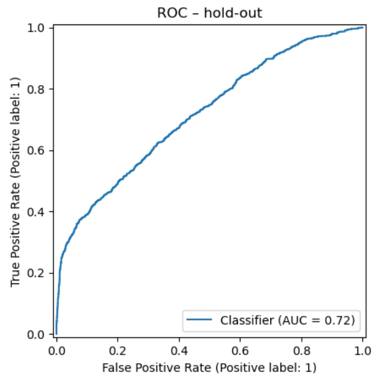

Figure 1: Enter Caption

`weekly_time_spent` alone accounted for approximately 80% of the model's total gain-based feature importance, indicating that it played a central role in decision-making across the ensemble of trees. Furthermore, the top eight features all originated from the Effort, Consistency, or Regularity dimensions.

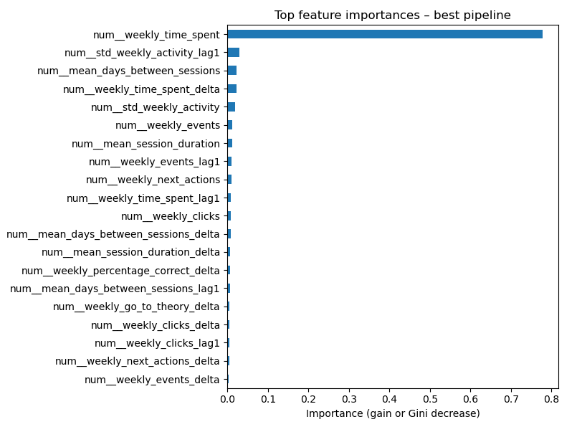

Figure 2: Enter Caption

These results support our hypothesis that student engagement can be effectively predicted using behavioral data, particularly features that capture effort and consistency over time. The stability of performance across time-based folds, and the dominance of effort-related predictors, suggest that retention is closely tied to regular and sustained platform usage. While the inclusion of profile labels did not significantly enhance predictive accuracy, it did offer interpretability that could support more personalized interventions.

In future work, it would be valuable to investigate how dynamic adaptations—such as personalized content delivery or targeted nudges might influence the behavioral dimensions identified as critical to retention. This could inform proactive strategies to support at-risk learners and improve long-term engagement outcomes.

References

[1] P. Mejia-Domenzain, M. Marras, C. Giang, and T. Käser, "Identifying and comparing multi-dimensional student profiles across flipped classrooms," *Proceedings of the 15th International Conference on Educational Data Mining (EDM)*, 2022.

[2] D. Urrutia Cordero, M. Scheffel, E. Ternieden, R. Pursian, and M. A. Chatti, "Visualizing self-regulated learner profiles in dashboards: Design insights from teachers," *arXiv preprint arXiv:2305.16851*, 2023.

[3] M. Khalil, D. Urrutia Cordero, and M. A. Chatti, "Student answer forecasting: Transformer-driven answer choice prediction for next-week engagement," *arXiv preprint arXiv:2405.20079*, 2024.