Proceedings of Machine Learning Research **318** (2026)
The 39th Canadian Conference on Artificial Intelligence
Canadian AI 2026
25-29 May 2026,Vancouver,BC

Personalized Stability Triggers: A Model-Agnostic Framework for Adaptive Early Prediction of At-Risk Students in Education

Qixiang Pang[†,*], Linrui Zhang[†], Shan Du[◇]
[†] Department of Computer Science and Cybersecurity, University of Central Missouri
[‡] Department of Computer Science and Cybersecurity, University of Central Missouri
[◇] Department of Computer Science, Mathematics, Physics and Statistics, The University of British Columbia

Abstract

Early prediction of student performance and timely intervention are important in education. A fundamental challenge is the trade-off between temporal earliness (intervening sooner) and predictive reliability (waiting for sufficient data). Conventional machine learning models typically impose static observation windows that do not account for the heterogeneous behaviors of individual students. In this paper, we propose the Personalized Stability Trigger (PST), a novel dynamic framework that identifies the optimal inference moment for each student based on the stochastic convergence of model confidence. By leveraging the ensemble variance of a Random Forest estimator, PST detects an "information plateau" where further data collection yields marginal information gain. We validate this framework on two disparate educational datasets: ASSISTments (micro-scale problem solving) and OULAD (macro-scale course engagement). Experimental results demonstrate that PST reduces observation latency by up to 20% compared to fixed-window baselines while preserving $>$99.5% of prediction accuracy. These findings indicate that stability-driven triggers offer a scalable, robust, and model-agnostic solution for early prediction of at-risk students.

Keywords: Applied Machine Learning, Educational Data Mining, Early Prediction, Sequential Decision-Making, Personalized Stability, Adaptive Triggers.

1. Introduction

Early identification of students at risk of failure or dropout is a major challenge in learning analytics [1]. Timely intervention can improve learning outcomes, reduce attrition, and support more personalized learning strategies. However, there is a trade-off between earliness (intervening as soon as possible) and accuracy (waiting for enough data to make reliable predictions). Most existing models rely on fixed-length observation windows [2], assuming that all students can be evaluated within a predefined period. In practice, learning trajectories vary widely: some students show clear signs of mastery or struggle early, while others take longer to develop observable patterns. As a result, fixed-window approaches may be too late for some students or too early for others.

To address this issue, we introduce the Personalized Stability Trigger (PST), a framework that determines the prediction point for each student dynamically. PST monitors the model's predicted probability for a target outcome and triggers a prediction once the probability stabilizes within a defined tolerance. This stability indicates that additional data is unlikely to significantly change the prediction, helping balance earliness and accuracy [3].

2. Related Work

2.1. Early Prediction in Learning Analytics

Early prediction of student performance has been widely studied in Educational Data Mining (EDM) and Learning Analytics (LAK), with approaches differing in data granularity and prediction targets. At the micro level, research typically focuses on predicting student

---

[*] qpang@ucmo.edu

performance on individual problems [4], while macro-level studies aim to predict course-level outcomes such as dropout or failure [5].

2.1.1. Micro-level prediction

A substantial body of work uses Knowledge Tracing (KT) models to estimate student mastery and predict future correctness [4][6]. Bayesian and deep KT variants mainly focus on improving next-step prediction accuracy, rather than determining when enough evidence is available to make an early and reliable prediction. As a result, most KT-based approaches implicitly assume continuous observation instead of early stopping [7].

2.1.2. Macro-level prediction

For course-level datasets such as OULAD [5], prior studies commonly use classifiers such as logistic regression, Random Forests, and gradient boosting to predict final outcomes [1][3]. These models typically rely on fixed prediction checkpoints (e.g., Day 30), which fail to capture individual differences in engagement patterns and learning trajectories [2][8].

In contrast to both strands, our work treats early prediction as a personalized temporal decision problem, explicitly addressing the trade-off between earliness and accuracy.

2.2. Adaptive and Optimal Stopping Approaches

Beyond education, adaptive stopping and early classification have been studied in sequential decision-making and online learning settings. Prior work has explored probability stabilization as a convergence criterion for stopping predictions once uncertainty drops below a threshold [9]. Similarly, early classification frameworks balance the cost of delayed decisions against prediction accuracy [10], often using dynamic window strategies [7].

However, these methods are rarely adapted to educational data, where learner behavior is highly diverse and observation frequencies vary widely. Moreover, existing approaches often depend on model-specific confidence estimates, which limits their practical use.

The proposed Personalized Stability Trigger (PST) adapts these ideas to learning analytics by using ensemble-based probability stability, enabling a model-agnostic and robust stopping mechanism that works for both high-frequency and low-frequency educational data.

3. Personalized Stability Trigger Algorithm

The Personalized Stability Trigger (PST) algorithm determines when enough evidence has been observed to confidently predict a student's risk status, balancing early prediction and reliability. Instead of using a fixed observation window, PST continuously tracks both the stability and confidence of model outputs, and triggers a decision once additional data is unlikely to change the result.

To ensure contextual adequacy [11], PST enforces an information floor ($\boldsymbol{F}$), which corresponds to a minimum number of observations (e.g., a fixed number of problems or days). After this threshold is reached, the model produces a risk probability $\boldsymbol{P_t}$ at each time step $\boldsymbol{t}$. Prediction stability is then assessed by tracking how recent probability values vary over a patience window of length $\boldsymbol{K}$. Stability is declared when the standard deviation within this window falls below a tolerance parameter ($\boldsymbol{\epsilon}$), indicating that an information plateau has been reached.

To further improve reliability and avoid premature triggers during periods of high model uncertainty [12], we introduce an Uncertainty Hold based on a confidence margin ($\boldsymbol{M}$). This secondary gate ensures that the algorithm only stops when the probability $\boldsymbol{P_t}$ has moved far enough away from the point of maximum uncertainty (0.5).

Formally, the dual-trigger condition is satisfied at time $\boldsymbol{t}$ if:

$$(t \geq F) \ and \ (\text{std}(P_{t-k:t}) < \epsilon) \ and \ (|P_t - 0.5| \geq M) \tag{3.1}$$

Once these conditions are met at the same time, the algorithm stops observing and makes a prediction. If this convergence is not reached before the sequence ends, the model instead predicts at the final observation.

The PST framework introduces four interpretable hyperparameters that provide fine-grained control over the trade-off between earliness and reliability.

- **Stability Tolerance** (ϵ): controls how much fluctuation is allowed before predictions are considered converged. Lower values (e.g., 0.001 to 0.01) require a flatter "information plateau" before triggering.
- **Patience** (k): specifies the number of consecutive stable steps required to trigger a prediction, providing robustness against transient noise and temporary performance plateaus.
- **Information floor** (F): guarantees a minimum observation period, preventing premature decisions based on insufficient initial data.
- **Confidence Margin** (M): governs Uncertainty Hold, requiring the model's output to diverge from the decision boundary (0.5) by a specific threshold. This ensures that the algorithm only triggers when the prediction is both stable and statistically decisive.

Together, these components allow PST to function as a lightweight, model-agnostic stopping mechanism that adapts to individual learning patterns. By combining stability and confidence checks, the framework remains robust to noise in later stages while still improving time efficiency.

4. Datasets and Experiments

4.1. Datasets

We evaluate the PST framework on two complementary datasets representing micro- and macro-level learning contexts [4][5]:

- **ASSISTments 2009-2010**: A micro-level dataset of 12,311 students completing approximately 278,000 problem attempts. Features include correctness, hint usage, response time, and prior skill mastery, capturing fine-grained, skill-level interactions.
- **OULAD** (Open University Learning Analytics Dataset): A macro-level dataset of 32,593 students containing clickstream and demographic information. Features include total clicks, video views, assessment submissions, and enrollment data.

These datasets together enable cross-domain validation of PST, demonstrating its effectiveness across diverse temporal and feature granularities.

4.2. Experimental Setup

- **Classifier**: Random Forest (n=100 trees). Gini impurity is used as the split criterion to balance interpretability and strong baseline performance. The Random Forest was chosen because of its proven effectiveness on tabular educational datasets and its ability to produce stable probability estimates, which are important for evaluating the convergence of the PST stability trigger [3]. It is worth noting that the PST framework is model-agnostic; although this study uses Random Forest, the method can also be applied to other models (e.g., Deep Knowledge Tracing or Gradient Boosting), which we leave for future work.
- **Baselines**:
  - ASSISTments: Fixed-window prediction at T = 25 problems

  - OULAD: Fixed-window prediction at T = 60 days

  These benchmarks represent conventional fixed-window early prediction approaches.
- **Metrics**:
  - Balanced Accuracy: Unlike standard accuracy, Balanced Accuracy calculates the arithmetic mean of class-specific recall (sensitivity and specificity) [8]. This accounts for the inherent class imbalance in educational datasets, where the number of successful students often significantly outweighs those at-risk.

$$\text{Balanced Accuracy} = \frac{\text{Sensitivity} + \text{Specificity}}{2} \tag{4.1}$$

  - Lead-Time Gained (LTG): The absolute number of problems (ASSISTments) or days (OULAD) saved compared to the baseline
  - Earliness Improvement (%): The relative percentage of time or problems saved compared to the total fixed-baseline observation window (T).

$$\text{Earliness} = \frac{(T - \text{PST}_{\text{step}})}{T} \times 100 \tag{4.2}$$

  - Accuracy Retention (%): The ratio of PST accuracy to baseline accuracy, quantifying the trade-off between speed and performance.
- **Evaluation Procedure**: We evaluated the PST framework using a stratified hold-out validation strategy. By splitting the data based on unique student IDs rather than individual interactions, we ensured that all interactions from a single student were kept entirely in either the training or testing set. This prevents information leakage and helps ensure that the model generalizes to unseen student profiles.

5. Results and Discussion

5.1. Quantitative Performance Analysis

The PST framework shows a strong ability to reduce time overhead while maintaining prediction accuracy. As shown in Table 1, it achieves almost no loss in accuracy across both micro-level (ASSISTments) and macro-level (OULAD) educational settings.

Metric	**ASSISTments (Micro)**	**OULAD (Macro)**
Parameters	F=15, $\varepsilon = 0.005$, K=7, M=0.1	F=30, $\varepsilon = 0.005$, K=7, M=0.1
Baseline Balanced Accuracy	0.8745	0.7037
PST Balanced Accuracy	0.8724	0.7005
Average Lead Time Gained	4.94 problems	9.63 days
Earliness Improvement	**19.76%**	**16.05%**
Accuracy Retention vs. Baseline	99.76%	99.55%

Table 1. Performance of PST Adaptive Triggering vs. Baselines.

The results show that on the OULAD dataset, PST identified at-risk students nearly 10 days earlier than traditional methods. In the ASSISTments setting, it enabled intervention about 5 problems earlier. Importantly, the accuracy drop in both cases was under 0.5%, confirming the robustness of the stability criterion.

5.2. Analysis of the Information Plateaus

The success of PST provides empirical evidence for the “Information Plateau” in learning analytics—the point at which the added value of additional student interaction data becomes negligible. This phenomenon is consistent with findings in Retrieval-Augmented Generation (RAG) systems [13] and studies in explainable AI [14].

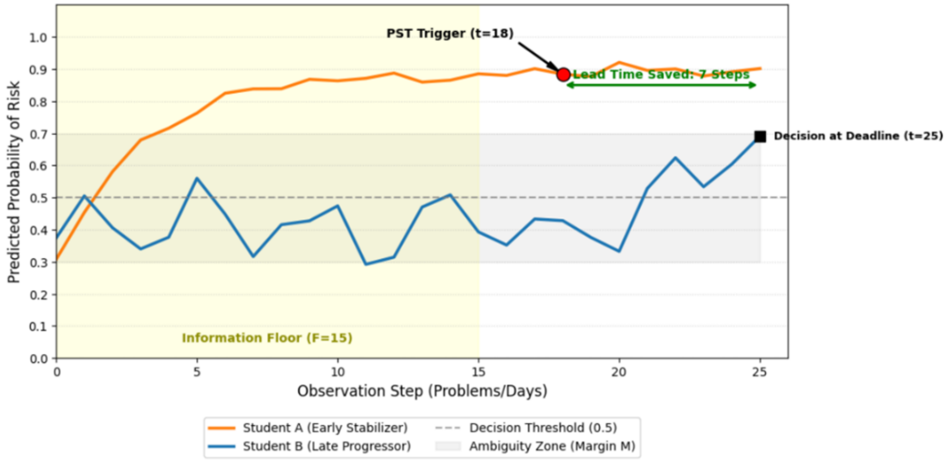

Figure 1. Probability Trajectory (Visualizing the PST Adaptive Trigger).

By identifying this plateau at the student level, PST enables more “titrated” educational interventions: systems can respond once prediction stability is reached, rather than waiting for a fixed administrative deadline.

5.3. Trajectory Visualization and Lead-Time Gain Analysis

Visual inspection of prediction trajectories reveals two representative archetypes:

- Student A (Stable): The predicted risk probability rapidly converges to 0.9. PST detects stability at $t=18$, yielding a 7-step lead-time gain relative to the fixed 25 problems.
- Student B (Unstable): The predicted probability fluctuates between 0.3 and 0.7. PST refrains from triggering early and instead waits until $t=F$(the final observation), avoiding premature and potentially incorrect classification.

Figure 1 illustrates how PST adapts to the temporal behavior of each student. Unlike fixed-window methods that use a single, uniform decision point, PST tracks how prediction confidence evolves and only triggers when it becomes stable. This adaptive behavior helps explain its better balance between earliness and reliability: cases that stabilize quickly allow earlier intervention, while uncertain cases avoid accuracy loss caused by making decisions too early.

6. Conclusion and Future Work

The Personalized Stability Trigger (PST) framework shows that stability-based adaptive early prediction performs better than traditional fixed-window methods in educational data mining. Across both micro- and macro-level datasets, PST provides significant gains in lead time while preserving predictive accuracy for identifying at-risk students, giving educators earlier and more reliable opportunities for intervention. The main contributions of this paper are as follows:

- Personalized Stability Trigger (PST) - A novel, model-agnostic stopping criterion that determines the optimal prediction time for each student based on probability stabilization, reducing reliance on fixed observation windows.
- Cross-domain validation: PST demonstrates robustness across both micro- and macro-level educational datasets.

- Earliness–accuracy balance: PST achieves significant lead-time gains with minimal performance loss.

Future work includes incorporating Gradient Boosting and SHAP explainability [14] to enhance feature interpretability; and extending PST to other models (LSTM, Transformer or reinforcement learning) [7].

Acknowledgements

Funding provided by a UCM Alumni Foundation Opportunity Grant Program. The authors also acknowledge the use of ChatGPT (free version) and Grammarly (free version) for proofreading the manuscript.

References

[1] Y. Yang, Z. Wen, J. Cao, J. Shen, H. Yin, and X. Zhou. "EPARS: Early prediction of at-risk students with online and offline learning behaviors". In: *International Conference on Database Systems for Advanced Applications*. Springer. 2020, pp. 3–19.

[2] W. Dai, J. Lin, F. J.-Y. Jin, Y.-S. Tsai, N. Srivastava, P. Le Bodic, D. Gašević, and G. Chen. "Learning Analytics for Early Identification of At-Risk Students and Feedback Intervention". In: *Journal of Learning Analytics* 12.3 (2025), pp. 102–125.

[3] Z. Azizah, T. Ohyama, X. Zhao, Y. Ohkawa, and T. Mitsuishi. "Predicting at-risk students in the early stage of a blended learning course via machine learning using limited data". In: *Computers and Education: Artificial Intelligence* 7 (2024), p. 100261.

[4] N. T. Heffernan and C. L. Heffernan. "The ASSISTments ecosystem: Building a platform that brings scientists and teachers together for minimally invasive research on human learning and teaching". In: *International journal of artificial intelligence in education* 24.4 (2014), pp. 470–497.

[5] J. Kuzilek, M. Hlosta, and Z. Zdrahal. "Open university learning analytics dataset". In: *Scientific data* 4.1 (2017), p. 170171.

[6] C. Piech, J. Bassen, J. Huang, S. Ganguli, M. Sahami, L. J. Guibas, and J. Sohl-Dickstein. "Deep knowledge tracing". In: *Advances in neural information processing systems* 28 (2015).

[7] Y. Lin, H. Chen, W. Xia, F. Lin, Z. Wang, and Y. Liu. "A Comprehensive Survey on Deep Learning Techniques in Educational Data Mining: Y. Lin et al." In: *Data Science and Engineering* (2025), pp. 1–27.

[8] B. Sonnleitner, T. Madou, M. Deceuninck, F. Theodosiou, and Y. R. Sagaert. "Evaluation of early student performance prediction given concept drift". In: *Computers and Education: Artificial Intelligence* 8 (2025), p. 100369.

[9] X. Chen, H. Dai, Y. Li, X. Gao, and L. Song. "Learning to stop while learning to predict". In: *International conference on machine learning*. PMLR. 2020, pp. 1520–1530.

[10] A. Gupta, H. P. Gupta, B. Biswas, and T. Dutta. "Approaches and applications of early classification of time series: A review". In: *IEEE Transactions on Artificial Intelligence* 1.1 (2020), pp. 47–61.

[11] A. N. Rafferty, E. Brunskill, T. L. Griffiths, and P. Shafto. "Faster teaching via pomdp planning". In: *Cognitive science* 40.6 (2016), pp. 1290–1332.

[12] B. Ujkani, D. Minkovska, and N. Hinov. "Course success prediction and early identification of at-risk students using explainable artificial intelligence". In: *Electronics* 13.21 (2024), p. 4157.

[13] M. Hoq, P. Brusilovsky, and B. Akram. "Explaining explainability: Early performance prediction with student programming pattern profiling". In: *Journal of Educational Data Mining* 16.2 (2024).

[14] A. Abukader, A. Alzubi, and O. R. Adegboye. "Intelligent system for student performance prediction: An educational data mining approach using metaheuristic-optimized lightgbm with shap-based learning analytics". In: *Applied Sciences* 15.20 (2025), p. 10875.