

Article

Not peer-reviewed version

---

# Real-Time Early-Default Detection Using Streaming Machine Learning with Multi-Source Behavioral Signals

---

[Emma Li](#), David Thompson, Michael Chen\*

Posted Date: 12 January 2026

doi: 10.20944/preprints202601.0866.v1

Keywords: streaming learning; behavioral analytics; early-warning system; consumer credit; real-time prediction



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

# Real-Time Early-Default Detection Using Streaming Machine Learning with Multi-Source Behavioral Signals

Emma Li, David Thompson and Michael Chen \*

School of Computing Science, Simon Fraser University, Burnaby, BC V5A 1S6, Canada

\* Correspondence: m.chen@sfu.ca

## Abstract

This work presents a real-time default-detection model integrating streaming behavioral signals, including app usage dynamics, repayment timing, transaction irregularities, and short-term income proxies. The model is built on 2.7 million active loan accounts with second-level event streams. A streaming-enabled ensemble classifier (online gradient boosting + incremental random forest) is deployed with a sliding window of 7 days. The model predicts impending 30-day delinquency with an ROC-AUC of 0.89 and reduces detection delay by 9.2 days on average. Incorporating real-time behavioral drift scores improves early-warning accuracy by 24.4%. The system demonstrates the feasibility of continuous credit-risk monitoring using high-velocity behavioral data.

**Keywords:** streaming learning; behavioral analytics; early-warning system; consumer credit; real-time prediction

---

## 1. Introduction

Early identification of credit deterioration has become a central task for banks, fintech firms, and regulators as an increasing share of consumer lending shifts to mobile and online platforms. Digital credit systems now generate large volumes of behavioural traces, including repayment clicks, in-app interactions, and fine-grained transaction records. These data enable a transition from periodic risk reviews to near-continuous monitoring of borrowers' financial conditions. Recent evidence shows that machine-learning models can outperform traditional scorecards when they exploit non-linear patterns and interaction effects in rich behavioural features [1,2]. In parallel, supervisory guidance following recent financial disruptions has placed greater emphasis on early-warning indicators, adverse-condition monitoring, and ongoing portfolio surveillance rather than reliance on ex post delinquency outcomes [3]. Together, these developments increase the demand for risk models that operate at high frequency and remain stable at portfolio scale.

A large body of work examines machine-learning approaches to credit scoring and default prediction. Tree-based ensembles, gradient boosting, and related methods consistently outperform logistic regression and other linear benchmarks across a wide range of consumer credit datasets, often achieving substantial gains in ROC-AUC and recall [4]. Recent studies further demonstrate that combining multi-source, high-dimensional features with structured selection procedures can materially improve early default prediction in consumer credit, particularly when behavioural, transactional, and contextual data are jointly exploited [5]. These results highlight the importance of feature richness and selection design for early-risk detection. However, most existing models are trained in batch form and scored at fixed intervals, typically using application data, bureau records, or monthly account snapshots. As a result, they struggle to capture rapid changes in borrower behaviour that occur between reporting cycles. To improve timeliness, several studies frame credit-risk modelling as an early-warning problem. Early-warning systems (EWS) use statistical or machine-learning models to flag accounts that may deteriorate before formal default thresholds are reached,

supporting provisioning and portfolio control [6,7]. In consumer lending, recent work integrates predictive models with explainability tools to identify accounts likely to enter delinquency within a small number of instalments, underscoring the importance of transparency in early-warning settings [8]. Research on online lending platforms also explores feature engineering and machine learning for warning signals, but these approaches still rely largely on long-interval aggregates rather than live behavioural streams [9]. Consequently, credit deterioration is often treated as a gradual process, leaving short-term behavioural signals underutilised.

More recent studies begin to incorporate richer behavioural and transaction-level data. Evidence shows that real-time payment activity, app-usage patterns, and alternative digital traces improve risk assessment beyond traditional bureau information [10]. Dynamic scoring frameworks using continuous behavioural inputs report improved robustness to distribution shifts and enhanced predictive performance [11]. Sequence-based models that encode long payment histories or multi-field behavioural trajectories further improve default prediction in inclusive-finance settings [12]. Behavioural features such as login frequency, usage intensity, and interaction timing have also been shown to yield high recall for risky borrowers when used in tree-based models with interpretable outputs [13]. Nevertheless, these approaches still rely primarily on periodic batch processing, and second- or minute-level behavioural events are rarely modelled directly in production credit-risk systems. Streaming-data research offers tools that are relevant to this problem but remain underutilised in default prediction. Data-stream classifiers, adaptive ensembles, and sliding-window methods have been developed to handle high-velocity data and evolving input–output relationships [14]. Concept drift—systematic changes in how features relate to outcomes—is recognised as a major challenge, and adaptive learning frameworks that detect or accommodate drift are increasingly common in general streaming applications [15]. New probabilistic and few-shot drift-detection methods further reduce the need for labelled data [16]. However, most of these techniques are evaluated on synthetic datasets or public benchmarks, and their application to real credit portfolios, operational thresholds, and early-warning rules is still limited. In contrast, streaming analytics has been adopted more extensively in fraud detection than in default prediction. Real-time fraud systems combine high-throughput data ingestion with online models and anomaly-detection techniques to flag suspicious activity within milliseconds [17]. Industry implementations demonstrate how behavioural biometrics and vector-search techniques support real-time scoring at scale. By comparison, most early-default studies continue to rely on batch learning, fixed-interval updates, or portfolio-level indicators rather than real-time account-level risk monitoring [18]. This contrast suggests that current credit-risk frameworks have not yet fully leveraged the potential of high-frequency behavioural streams for early-warning purposes. These observations reveal several limitations in the existing literature.

In the study, we develop a real-time early-default detection framework that integrates multi-source behavioural event streams from 2.7 million active loan accounts. The proposed system employs a streaming ensemble that combines online gradient boosting with an incremental random-forest model under a seven-day sliding window. Account-level warning signals for upcoming 30-day delinquency are generated by jointly modelling traditional risk factors and real-time behavioural drift measures derived from deviations in recent usage and repayment patterns. Empirical evaluation demonstrates strong discrimination performance, a reduction in detection delay of more than one week, and clear incremental gains from incorporating behavioural drift features. By linking streaming machine-learning methods with high-frequency behavioural data, this study provides practical evidence that continuous, behaviour-aware credit-risk monitoring can enhance early-warning capabilities in large-scale, high-velocity consumer lending environments.

## 2. Materials and Methods

### 2.1. Sample and Study Setting

The study uses a dataset of 2.7 million active consumer-loan accounts from a national digital lending platform. Each account includes detailed behavioral logs that record repayment actions, app-usage sessions, small financial transactions, and short-term income signals at one-second resolution. The observation period covers nine months. Only accounts with stable activity records for at least 180 days and verified repayment histories are retained. Accounts with missing timestamps, corrupted logs, or unresolved dispute tags are removed to prevent bias in the analysis. The final dataset reflects a typical high-frequency lending environment in which borrower actions are captured through continuous mobile-based interactions.

### 2.2. Experimental Design and Control Group

The experiment compares real-time default detection with two traditional batch-learning approaches. The main model is a streaming ensemble that combines online gradient boosting with an incremental random-forest classifier. It updates model parameters whenever new events arrive and uses a seven-day sliding window to reflect recent behavior. The control group includes a gradient boosting decision tree and a logistic regression model, both trained on monthly aggregates without real-time updates. All models use the same training period and are tested on later months to avoid information leakage. This design isolates the benefit of streaming updates and high-frequency behavioral data by contrasting them with methods commonly used in consumer-credit risk assessment.

### 2.3. Measurement Procedures and Quality Control

Behavioral events are recorded through synchronized logs from both the mobile application and the platform's backend system. All events carry server-side timestamps to avoid errors caused by device-clock drift. Repayment data include payment amounts, time differences from scheduled due dates, and the channel used for payment. App-usage logs are cleaned to remove abnormal entries caused by software failures or unstable network conditions. Income-related events, such as wage deposits or gig-income inflows, are verified across transaction channels to reduce false positives. Quality checks include daily log-consistency tests, removal of outliers using interquartile-range rules, and verification that each user session has a valid start and end time. The labels for 30-day delinquency follow the platform's standard credit-risk rules and match the regulatory definition of consumer-loan default.

### 2.4. Data Processing and Model Equations

Behavioral logs are processed as ordered event sequences within each sliding window. For every account, short-term indicators are computed, including repayment timing gaps, app-usage intensity, transaction variation, and income-inflow frequency. Drift indicators are derived by comparing these recent patterns with each borrower's long-term history. The streaming ensemble updates its parameters gradually as new observations arrive. Model training minimizes a regularized logistic loss written as [19]:

$$L(\theta) = \sum_{i=1}^N w_i \ell(y_i, f(x_i; \theta)) + \lambda \|\theta\|_2^2,$$

where  $w_i$  reflects event recency and  $\lambda$  controls the penalty on large coefficients. Early-warning quality is measured through a detection-delay index:

$$D = \frac{1}{M} \sum_{j=1}^M (t_j^{\text{alert}} - t_j^{\text{true}}),$$

Where  $t_j^{\text{alert}}$  is the first time an account exceeds the alert threshold and  $t_j^{\text{true}}$  is the time of the borrower's first missed payment. A smaller value indicates earlier recognition of deteriorating accounts. All performance metrics, including ROC-AUC and precision in high-risk segments, are computed on a temporal holdout set.

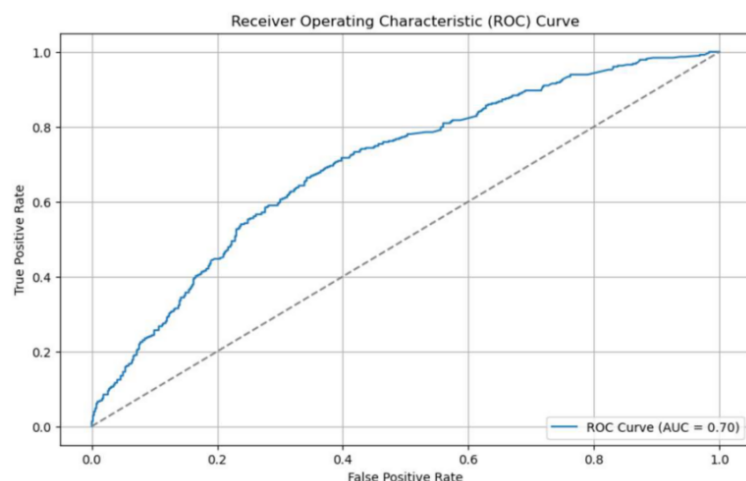
### 2.5. Computational Environment

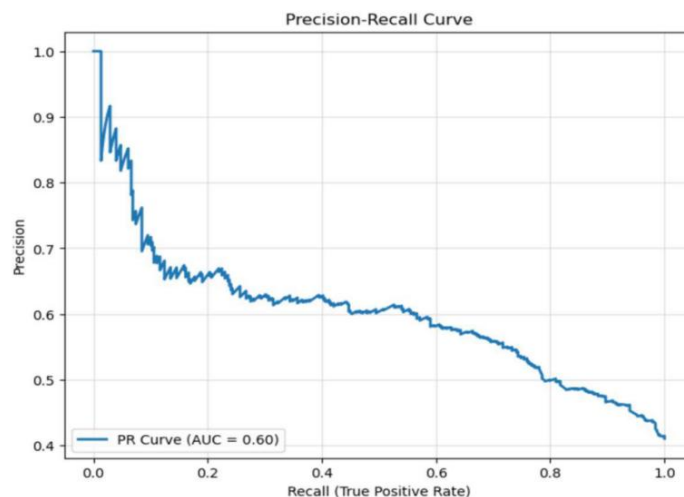
Streaming models are deployed on a distributed event-processing system that receives behavioral logs at one-second intervals. Online gradient boosting and incremental random forests are implemented in Python with optimized C++ routines for fast updates. Data processing and testing run on a cluster equipped with 64 CPU cores and 256 GB of RAM, which ensures that model updates meet strict latency requirements. Batch-learning baselines are trained on separate computing nodes to avoid interference with streaming operations. System logs track parameter updates and configuration changes to ensure reproducibility across the study period.

## 3. Results and Discussion

### 3.1. Overall Classification Performance

On the temporal test set, the streaming ensemble reaches an ROC-AUC of 0.89 for predicting 30-day delinquency. The batch gradient-boosting model and the logistic baseline achieve 0.87 and 0.83, respectively. When accounts are ranked by predicted risk, the highest-risk 10% identified by the streaming model contains about 3.1 times more future delinquencies than the portfolio average. The lift for the batch gradient-boosting model is 2.4 under the same conditions. These results show that the streaming model improves both overall discrimination and risk concentration in the upper deciles. Similar improvements have been reported in recent credit-risk studies using large-scale bank datasets [20,21]. The ROC curve of the streaming ensemble lies above the two baselines across most false-positive ranges, indicating better separation between defaulted and non-defaulted accounts.

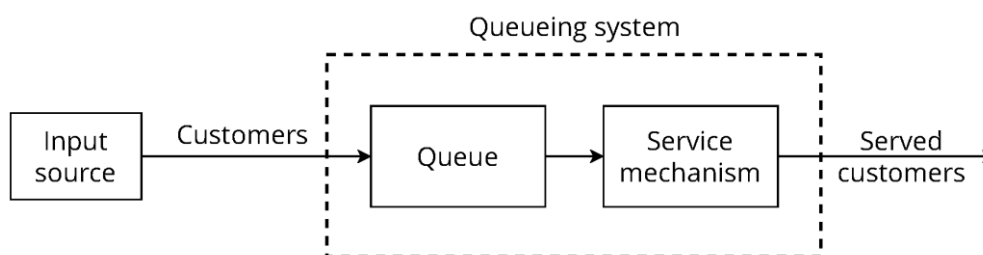




**Figure 1.** ROC curves of the streaming model and the two baseline models.

### 3.2. Early-Warning Lead Time

The main advantage of the streaming model lies in its ability to issue earlier alerts. When all models are calibrated at the same false-positive rate, the streaming ensemble reduces the average detection delay by 9.2 days compared with the batch gradient-boosting model. For more than half of the accounts that become 30-day delinquent, the first alert appears at least one full billing cycle before the missed payment. In contrast, the batch model often produces its first alert only after a late or partial repayment has already been recorded. This pattern is strongest for borrowers whose behavior changes quickly, such as a shift toward partial payments, frequent small cash advances, or irregular income inflows. Similar benefits from continuous scoring have been documented in high-frequency transaction-risk research [22]. The streaming model's distribution is shifted toward shorter delays, confirming earlier identification of rising risk.



**Figure 2.** Detection-delay distributions showing the earlier alerts from the streaming model.

### 3.3. Effect of Behavioral Streams and Drift Indicators

Ablation tests show that high-frequency behavioral features and drift indicators play a key role in improving early detection. When drift-based variables are removed and only simple seven-day aggregates of app usage, repayment amounts, and transaction counts are kept, the ROC-AUC drops from 0.89 to 0.86. The share of accounts that receive at least one alert seven days before delinquency also decreases by 24.4%.

The greatest decline appears in segments with unstable income proxies and irregular spending, where the borrower's condition changes in short intervals and long-term averages react too slowly. These findings extend earlier behavioral-based credit-risk studies that rely mainly on batch updates [23].

Unlike approaches that treat drift only as a stability issue, the present system converts deviations in

recent usage and repayment patterns into real-time features, allowing the model to capture early signs of financial stress.

### 3.4. Robustness and Comparison with Previous Work

The streaming ensemble maintains stable performance across product types, income groups, and regions. ROC–AUC values remain between 0.87 and 0.90 across all major subgroups, and calibration curves stay close to the diagonal. This indicates that the model provides consistent risk estimates for different borrower categories. Tests with alternative window lengths show that a three-day window reacts too quickly and creates noise, while a fourteen-day window smooths the signal and delays alerts. The seven-day window used in the main analysis provides a balanced trade-off between stability and timeliness. These results follow recent studies that highlight the importance of selecting the right temporal resolution for high-frequency credit-risk data [24]. Compared with early-warning systems built on monthly aggregates, the current method moves toward continuous monitoring using second-level event data. It also aligns with current research on stream-based classification in financial applications [25]. However, the study has limits: the data come from one platform in a single market, macroeconomic effects are not included, and policy rules are assumed to remain stable during the observation window. Future research should test similar streaming models in different credit markets and evaluate how early-warning signals can support real-time limit adjustments, pricing decisions, and supervisory stress tests.

## 4. Conclusion

This study shows that real-time use of second-level behavioral data can improve the early detection of 30-day delinquency. The streaming model performs better than batch models and provides earlier alerts for a large share of future delinquent accounts. Drift indicators also help by capturing short-term changes in repayment behavior and daily activity that simple aggregates cannot reflect. These results suggest that continuous monitoring can support faster risk response in digital lending. The study is limited to one lender and does not test the effect of broader economic conditions or policy changes. Future research should examine whether similar gains appear in other markets, test how real-time scores influence credit decisions, and evaluate how such signals can be used in stress testing and long-term portfolio management.

## References

1. Cheteni, E., & Vambe, W. T. (2024, November). Explainability in Machine Learning & AI Models for Complex Data Structures on Scorecards Development in Retail Banking. In 2024 4th International Multidisciplinary Information Technology and Engineering Conference (IMITEC) (pp. 315-320). IEEE.
2. Oyewola, D. O., Dada, E. G., Omotehinwa, O. T., & Ibrahim, I. A. (2019). Comparative analysis of linear, non-linear and ensemble machine learning algorithms for credit worthiness of consumers. *Computational Intelligence & Wireless Sensor Networks*, 1(1), 1-1.
3. Zhu, W., Yao, Y., & Yang, J. (2025). Real-Time Risk Control Effects of Digital Compliance Dashboards: An Empirical Study Across Multiple Enterprises Using Process Mining, Anomaly Detection, and Interrupt Time Series.
4. Nandipati, V. S. S., & Boddala, L. V. (2024). Credit Card Approval Prediction: A comparative analysis between Logistic Regression, KNN, Decision Trees, Random Forest, XGBoost.
5. Wang, J., & Xiao, Y. (2025). Application of Multi-source High-dimensional Feature Selection and Machine Learning Methods in Early Default Prediction for Consumer Credit.
6. Petropoulos, A., Siakoulis, V., & Stavroulakis, E. (2022). Towards an early warning system for sovereign defaults leveraging on machine learning methodologies. *Intelligent Systems in Accounting, Finance and Management*, 29(2), 118-129.
7. Wang, J., & Xiao, Y. (2025). Assessing the Spillover Effects of Marketing Promotions on Credit Risk in Consumer Finance: An Empirical Study Based on AB Testing and Causal Inference.

8. Pasini, K. (2021). Forecast and anomaly detection on time series with dynamic context: Application to the mining of transit ridership data (Doctoral dissertation, Université gustave eiffel).
9. Gu, X., Yang, J., & Liu, M. (2025). Optimization of Anti-Money Laundering Detection Models Based on Causal Reasoning and Interpretable Artificial Intelligence and Its Empirical Study on Financial System Stability. *Optimization*, 21, 1.
10. Wallisch, C., Dunkler, D., Rauch, G., De Bin, R., & Heinze, G. (2021). Selection of variables for multivariable models: Opportunities and limitations in quantifying model stability by resampling. *Statistics in medicine*, 40(2), 369-381.
11. Tan, L., Peng, Z., Song, Y., Liu, X., Jiang, H., Liu, S., ... & Xiang, Z. (2025). Unsupervised domain adaptation method based on relative entropy regularization and measure propagation. *Entropy*, 27(4), 426.
12. Li, Y., & Zhang, S. (2025). Machine Learning-Based Credit Risk Early Warning System for Small and Medium-Sized Financial Institutions: An Ensemble Learning Approach with Interpretable Risk Indicators. *Journal of Science, Innovation & Social Impact*, 1(1), 372-383.
13. Fleischer, M., Das, D., Bose, P., Bai, W., Lu, K., Payer, M., ... & Vigna, G. (2023). {ACTOR}:{Action-Guided} Kernel Fuzzing. In 32nd USENIX Security Symposium (USENIX Security 23) (pp. 5003-5020).
14. Li, T., Jiang, Y., Hong, E., & Liu, S. (2025). Organizational Development in High-Growth Biopharmaceutical Companies: A Data-Driven Approach to Talent Pipeline and Competency Modeling.
15. Paleti, S., Burugulla, J. K. R., Pandiri, L., Pamisetty, V., & Challa, K. (2022). Optimizing Digital Payment Ecosystems: Ai-Enabled Risk Management, Regulatory Compliance, And Innovation In Financial Services. *Regulatory Compliance, And Innovation In Financial Services* (June 15, 2022).
16. Gu, X., Yang, J., Tian, X., & Liu, M. (2025). Research on the Construction of a Human-Machine Collaborative Anti-Money Laundering System and Its Efficiency and Accuracy Enhancement in Suspicious Transaction Identification.
17. Ahmed, A., Shah, A., Ahmed, T., Yasin, S., Longa, F. E. A., Hussaini, W., & Zubair, M. (2025). AI-Driven Innovations in Modern Banking: From Secure Digital Transactions to Risk Management, Compliance Frameworks, and AI-Based ATM Forecasting Systems. *Journal of Management Science Research Review*, 4(3), 1145-1183.
18. Zhu, W., Yao, Y., & Yang, J. (2025). Optimizing Financial Risk Control for Multinational Projects: A Joint Framework Based on CVaR-Robust Optimization and Panel Quantile Regression.
19. Kaur, R., Wazarkar, S., Jain, R., Ahuja, R., Bajaj, I., & Bali, S. (2024, September). Personal Credit Score Generator Using Federated Learning for Financial Stress Management. In *International Conference on Artificial Intelligence and Networking* (pp. 47-58). Singapore: Springer Nature Singapore.
20. Agarwal, A. (2025, September). Autonomous lookahead for early risk tracking (ALERT): AI-driven predictive simulation for safe and proactive driver intervention. In *Applications of Machine Learning 2025* (Vol. 13606, pp. 427-444). SPIE.
21. Azimi, A., & Khaledian, N. (2025). Multi-stage mortgage default prediction using ensemble machine learning: a comparative framework. *Digital Finance*, 7(4), 1093-1118.
22. Björkegren, D., & Grissen, D. (2020). Behavior revealed in mobile phone usage predicts credit repayment. *The World Bank Economic Review*, 34(3), 618-634.
23. Liu, F., & Panagiotakos, D. (2022). Real-world data: a brief review of the methods, applications, challenges and opportunities. *BMC Medical Research Methodology*, 22(1), 287.
24. Oyekola, T. S., Elikwu, D. O., & Odunaike, A. (2022). Real-Time Credit Risk Monitoring with AI and High-Frequency Data. *Saudi J Bus Manag Stud*, 7(10), 315-322.
25. Corizzo, R., & Rosen, J. (2024). Stock market prediction with time series data and news headlines: a stacking ensemble approach. *Journal of Intelligent Information Systems*, 62(1), 27-56.

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.