Proceedings of the 11th North American International Conference on Industrial Engineering and Operations Management, Milwaukee, Wisconsin, USA, June 9-11, 2026

Proceedings of the International Conference on Industrial Engineering and Operations Management

Publisher: IEOM Society International, USA DOI: 10.46254/NA11.20260080
Published: June 9, 2026

Short-Term Medicaid Dropout Prediction Using Longitudinal Weekly Eligibility Data and Interpretable Machine Learning Models

Majed Aljamrah
Graduate Research Associate
School of Systems Science and Industrial Engineering
Binghamton University, State University of New York
Binghamton, NY, USA
maljamrah@binghamton.edu

Yong Wang
Associate Professor
School of Systems Science and Industrial Engineering
Binghamton University, State University of New York
Binghamton, NY, USA
yongwang@binghamton.edu

Abstract

Maintaining continuous Medicaid eligibility is essential for ensuring access to healthcare services, continuity of care, and effective administration of public health programs. Despite ongoing policy efforts to improve retention, short-term eligibility loss remains frequent and often occurs with little advance notice. This study presents an interpretable machine learning framework for predicting next-week Medicaid dropout using longitudinal weekly eligibility records. Leveraging 52 weeks of binary eligibility indicators, the problem is formulated as a one-week-ahead prediction task, and transparent temporal features are constructed to summarize recent participation intensity, continuity, and instability. These longitudinal signals are integrated with demographic characteristics, including age group, sex, primary language category, and county of residence. Multiple ensemble learning models are assessed under extreme class imbalance using predictive performance and outreach-oriented ranking metrics, including ROC-AUC, PR-AUC, recall@10%, and lift@10%. The results indicate that incorporating demographic context substantially enhances predictive performance. The strongest model, CatBoost with demographic features, achieves a recall@10% of 0.65 and a lift of 6.5×, supporting effective prioritization for targeted outreach. Overall, the findings highlight the benefit of combining eligibility behavior over time with demographic context to enable proactive Medicaid retention strategies.

Keywords
Medicaid Retention; Dropout Prediction; Longitudinal Panel Data; Class Imbalance; Interpretable Machine Learning.

1. Introduction
Continuous Medicaid enrollment is a key determinant of healthcare access, continuity of care, and population health outcomes. Even brief interruptions in eligibility can lead to missed preventive services, delays in treatment, and increased downstream costs for both beneficiaries and providers. Although recent policy initiatives have sought to

stabilize enrollment, short-term eligibility loss remains prevalent, particularly among populations facing administrative complexity and communication barriers.

Administrative eligibility systems generate high-frequency longitudinal data that record enrollment status at weekly or monthly intervals. While these data are routinely collected, they are most often used for retrospective reporting rather than proactive risk identification. Existing analytic approaches frequently focus on long-horizon churn or cumulative disenrollment, limiting their usefulness for near-term intervention. In operational settings, care managers and Medicaid administrators must allocate limited outreach resources and therefore require tools that identify members at imminent risk of losing coverage.

This study addresses this gap by framing Medicaid eligibility loss as a short-horizon, one-week-ahead prediction problem using weekly eligibility histories. By aligning the prediction task with operational workflows, the proposed framework supports proactive outreach before eligibility disruption occurs.

This study makes three key contributions. First, it formulates Medicaid dropout as a next-week prediction problem using high-frequency administrative data. Second, it proposes an interpretable feature engineering framework that integrates longitudinal eligibility patterns with demographic context. Third, it evaluates predictive performance using ranking-based metrics aligned with outreach capacity constraints and demonstrates substantial gains in recall and lift when demographic information is incorporated.

1.1 Objectives

This study aims to develop an interpretable and operationally useful framework for predicting short-term Medicaid eligibility loss using longitudinal weekly data. The focus is on identifying members who remain eligible in a given week but face an increased likelihood of losing coverage in the following week, allowing outreach efforts to be initiated before eligibility disruption occurs. In support of this goal, the study constructs transparent temporal features that summarize recent engagement patterns, including participation intensity, continuity, and instability. In addition, multiple machine learning models are evaluated and compared under conditions of extreme class imbalance using ranking-based performance measures that align with real-world outreach capacity. The study further examines the contribution of demographic characteristics to predictive performance and investigates how these factors interact with longitudinal eligibility histories in shaping short-term dropout risk.

2. Literature Review

2.1 Medicaid Churn and Coverage Instability

Medicaid coverage instability, commonly referred to as churn, has been extensively documented as a source of disrupted access and administrative burden. Studies show that Affordable Care Act (ACA) expansions reduced churn but did not eliminate eligibility instability, particularly among working-age adults (Goldman and Sommers 2020; Daw et al. 2020). Evidence from the COVID-19 continuous coverage policy further demonstrates that administrative processes continue to generate disenrollment despite reduced transitions out of Medicaid (Nelson et al. 2023). During the post–public health emergency unwinding, coverage losses varied substantially across states and demographic groups, reinforcing the need for targeted outreach strategies (McIntyre et al. 2025; Soni and Blackburn 2025; Rumalla et al. 2024).

2.2 Short-Horizon Dropout Prediction and Rare Events

While Medicaid churn has a substantial policy literature, fewer studies frame coverage loss as a short-horizon prediction task aligned with operational workflows. In adjacent healthcare settings, such as loss to follow-up in HIV and tuberculosis care, machine learning models are used to identify near-term disengagement risk under limited outreach capacity (Endebu et al. 2025; Chen et al. 2024; Rodrigues et al. 2024). These studies highlight challenges shared with Medicaid dropout prediction, including outcome rarity and temporal dependence.

2.3 Longitudinal Feature Engineering

Across healthcare retention problems, translating longitudinal histories into interpretable temporal features such as recency, streak length, switching behavior, and volatility has been shown to improve prediction performance (Chen et al. 2024; Rodrigues et al. 2024). In missed-appointment prediction, engineered temporal features tied to recent attendance behavior and continuity are particularly effective for operational targeting (Tuan et al. 2025; Yang et al. 2024).

2.4 Machine Learning and Evaluation under Imbalance

Tree-based ensemble models frequently outperform linear baselines in nonlinear settings while retaining interpretability via feature importance analysis (Chen and Guestrin 2016; Ke et al. 2017; Prokhorenkova et al. 2018). In rare-event prediction, precision–recall metrics and ranking-based evaluation are preferred over accuracy due to extreme class imbalance (Saito and Rehmsmeier 2015). Collectively, this literature motivates an approach that integrates interpretable temporal features with demographic context and evaluates performance using outreach-oriented metrics.

3. Methods

This section describes the modeling framework used to predict short-term member dropout from longitudinal weekly binary panel data. The workflow includes data preprocessing, label construction, temporal feature engineering, model development, and evaluation using both discrimination and outreach-oriented ranking metrics.

3.1 Problem Formulation

Let $Y_{i,t} \in \{0,1\}$denote the eligibility status of member $i$in week t, where $Y_{i,t} = 1$ indicates active eligibility and $Y_{i,t} = 0$ indicates inactive eligibility. Short-term dropout is defined as:

$$D_{i,t+1} = 1(Y_{i,t} = 1 \text{ and } Y_{i,t+1} = 0)$$

This equation defines dropout as a transition from active eligibility to inactive eligibility. The indicator $D_{i,t+1}$equals 1 when member $i$is active in week $t$and becomes inactive in week $t + 1$. Otherwise, it equals 0.
The prediction task is to estimate:

$$P(D_{i,52} = 1 \mid Y_{i,1}, \dots, Y_{i,51}, X_i)$$

This equation represents the probability that member $i$will drop out in Week 52, given the member's eligibility history from Week 1 through Week 51 and demographic attributes X_i. In other words, the model uses all available information before Week 52 to estimate the risk of next-week dropout. All predictors are constructed using information available up to Week 51 to prevent label leakage.

3.2 Feature Engineering

Eligibility histories are transformed into interpretable temporal features summarizing recent participation behavior. All features are computed using weeks 1–51 (Table 1).

Table 1. Temporal feature definitions used for model training

Feature	Type	Description
weeks_present_last_4	Numeric	Number of active weeks in the most recent 4-week window
weeks_present_last_8	Numeric	Number of active weeks in the most recent 8-week window
weeks_present_last_12	Numeric	Number of active weeks in the most recent 12-week window
weeks_present_last_26	Numeric	Number of active weeks in the most recent 26-week window
total_weeks_active	Numeric	Total active weeks across the full 51-week history
consecutive_weeks_present	Numeric	Length of the uninterrupted active streak ending at Week 51

weeks_since_last_active	Numeric	Number of weeks elapsed since the most recent active week
ever_dropped	Binary	Indicator of at least one prior eligibility loss (1→0 transition)
num_dropout_events	Numeric	Total number of 1→0 eligibility transitions observed
switches_last_12	Numeric	Number of active–inactive transitions in the last 12 weeks
switches_last_26	Numeric	Number of active–inactive transitions in the last 26 weeks
trend_4_vs_12	Numeric	Short-term activity rate minus long-term rate (4-week vs 12-week)
trend_4_vs_26	Numeric	Short-term activity rate minus long-term rate (4-week vs 26-week)
half_trend	Numeric	Active weeks in second half (weeks 27–51) minus first half (weeks 1–26)
AgeGroup	Categorical	Binned age category (Child / Young_Adult / Adult / Middle_Aged / Senior / Elderly)
MbrSexCd	Categorical	Member sex (M / F)
LangGroup	Categorical	Grouped primary spoken language (8 categories)
MbrLvlResCountyCd	Categorical	County of residence (New York State county code)

3.3 Models and Evaluation

Seven machine learning models are evaluated: Logistic Regression, Random Forest, AdaBoost, XGBoost, LightGBM, Linear Support Vector Machine (LinearSVM), and CatBoost. All tree-based models receive one-hot-encoded categorical features, while linear models additionally receive standardized numeric features. CatBoost is the only model that handles categorical variables natively without preprocessing. Class imbalance (923:1 negative-to-positive ratio) is addressed through class-weighted loss functions and scale_pos_weight for gradient-boosted models. Performance is assessed using ROC-AUC, PR-AUC, Recall@K, and Lift@K, where K denotes the top fraction of members ranked by predicted risk. Ranking metrics are the primary evaluation criterion because outreach capacity is limited and only a small fraction of members can be contacted each week.

$$\text{ROC-AUC} = \int_0^1 \text{TPR}(\text{FPR})\, d(\text{FPR})$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

$$\text{Recall@10\%} = \frac{|\{i \in \mathcal{S}_{10\%}: D_i = 1\}|}{|\{i: D_i = 1\}|}, \text{Lift@10\%} = \frac{\text{Recall@10\%}}{0.10}$$

4. Data Collection

The dataset is collected in a healthcare organization in New York State, and it consists of weekly Medicaid eligibility records covering a 52-week observation period. Each row represents a unique member, with binary indicators reflecting weekly eligibility status, where 1 denotes active eligibility, and 0 denotes inactive eligibility. In this context, inactivity may reflect true coverage loss or administrative gaps but remains operationally meaningful as a loss of active status within the eligibility system.

The final analytic dataset contains N = 429,875 members, with a short-term dropout base rate of approximately 0.11%. Demographic variables associated with each member include age group, sex, primary language group, and county of residence. These attributes are treated as categorical features, with missing values handled through explicit "Unknown" categories to preserve information without introducing bias.

Table 2 presents an illustrative example of the weekly eligibility/activity panel structure used in this study, including both longitudinal eligibility indicators and demographic attributes. All predictive features are constructed using information available up to Week 51 to prevent label leakage.

Table 2. Example of weekly Medicaid eligibility panel data

Member_ID	**AgeGroup**	**MbrSexCd**	**LangGroup**	**MbrLvlResCountyCd**	**Week_1**	**..**	**Week_52**
A0	B	M	A	01	1	..	1
A1	A	M	B	02	1	..	0
A2	C	F	C	03	0	..	0

5. Results and Discussion

5.1 Numerical Results

The final analytic dataset contains N = 429,875 Medicaid members observed over 52 consecutive weeks. Of these, 465 members (0.11%) experienced next-week dropout defined as transitioning from active eligibility in Week 51 to inactive status in Week 52. This yields a severe class imbalance ratio of 924:1, which is typical of administrative healthcare datasets and necessitates outreach-oriented evaluation metrics rather than conventional accuracy measures.

All models were trained on a stratified 70% partition (300,912 members; 325 dropout events) and evaluated on the held-out 30% test set (128,963 members; 140 dropout events). Table 3 presents the full comparison of seven machine learning models across four primary metrics: ROC-AUC, PR-AUC, Recall at the top 10% of predicted risk (Recall@10%), and Lift at the top 10% (Lift@10%). Additionally, Recall and Lift are reported at the top 1% and top 5% thresholds to reflect stricter outreach capacity constraints.

Table 3. Model Performance Comparison on the 30% Held-Out Test Set (N = 128,963; 140 Positive Cases; Base Rate = 0.11%)

Model	**ROC-AUC**	**PR-AUC**	**Recall@1%**	**Recall@5%**	**Recall@10%**	**Lift@1%**	**Lift@5%**	**Lift@10%**
CatBoost	0.866	0.084	0.321	0.479	0.679	32.2×	9.6×	6.79×
LinearSVM	0.864	0.106	0.286	0.479	0.671	28.6×	9.6×	6.71×
AdaBoost	0.837	0.077	0.207	0.371	0.614	20.7×	7.4×	6.14×
Logistic Regression	0.843	0.035	0.221	0.421	0.593	22.2×	8.4×	5.93×
Random Forest	0.812	0.009	0.079	0.350	0.457	7.9×	7.0×	4.57×
XGBoost	0.757	0.052	0.171	0.279	0.400	17.2×	5.6×	4.00×

LightGBM	0.597	0.004	0.114	0.271	0.350	11.4×	5.4×	3.50×

CatBoost achieved the strongest overall performance, attaining a ROC-AUC of 0.866, a Recall@10% of 0.679, and a Lift@10% of 6.79×. Operationally, this means that by contacting only the 10% of members with the highest predicted risk (approximately 12,896 members per week), outreach teams can expect to identify 67.9% of all imminent dropouts representing a 6.8-fold improvement over random selection. At the stricter top-1% threshold (1,289 members), CatBoost achieves a lift of 32.2×, capturing 32.1% of all dropouts while contacting fewer than 1.3% of the population. These figures highlight the substantial efficiency gains achievable through risk-stratified outreach.

LinearSVM ranked a close second (ROC-AUC = 0.864, Lift@10% = 6.71×) and achieved the highest PR-AUC of 0.106, indicating superior precision among the top-ranked members. This metric is particularly relevant when outreach budgets are highly constrained, as it reflects the proportion of true dropouts among the most confidently flagged cases. AdaBoost ranked third (ROC-AUC = 0.837, Lift@10% = 6.14×), followed by Logistic Regression (ROC-AUC = 0.843, Lift@10% = 5.93×), which serves as a transparent and deployable baseline. Random Forest and XGBoost showed moderate but practically useful performance (Lift@10% = 4.57× and 4.00×, respectively). LightGBM produced the weakest results (ROC-AUC = 0.597, Lift@10% = 3.50×) despite imbalance-aware configuration, suggesting that its leaf-splitting strategy does not generalize effectively under the 923:1 imbalance present in this dataset.

5.2 Graphical Results

Figure 1 shows the aggregate weekly Medicaid eligibility rate across the study population over the full 52-week observation window. The rate remains broadly stable between approximately 78% and 83% from Weeks 1 through 47, with short-term week-to-week fluctuations consistent with administrative processing cycles. A pronounced decline is visible in Weeks 48–52, reflecting an increase in eligibility disruptions toward the end of the observation year. This late-period dip coincides with the prediction window (Week 51 → Week 52) and confirms the operational relevance of the one-week-ahead framing, as dropout risk intensifies near annual recertification deadlines. The shaded region in Figure 1 highlights the prediction window from which the dropout label is derived.

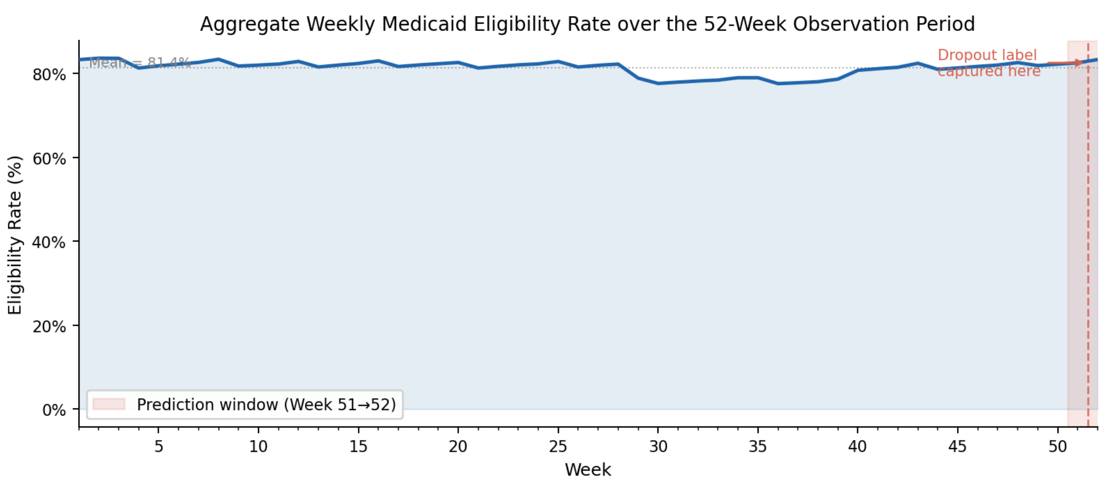

Figure 1. Aggregate weekly Medicaid eligibility rate over the 52-week observation period. The shaded region marks the prediction window (Week 51 → Week 52) from which the dropout label is constructed.

Figure 2 provides a side-by-side visual comparison of all seven models across four performance metrics: ROC-AUC, PR-AUC, Recall@10%, and Lift@10%. CatBoost and LinearSVM dominate across nearly all metrics, with the gap between them small enough to suggest that the linear boundary imposed by LinearSVM is approximately sufficient for this problem at these capacity thresholds. The large drop in PR-AUC for Random Forest (0.009) and LightGBM (0.004) relative to their ROC-AUC values reveals that these models produce many false positives among their top-ranked predictions a critical limitation when outreach budgets are constrained to the top 1–5% of the population.

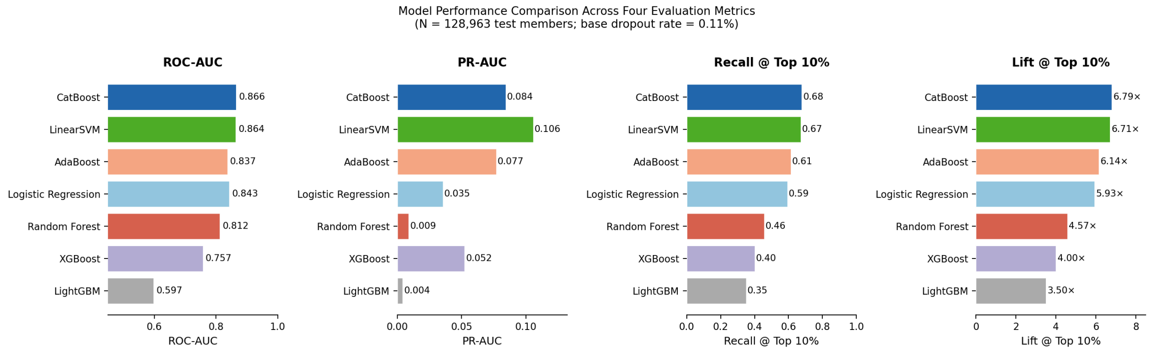

Figure 2. Side-by-side comparison of all seven models across four performance metrics. Models are sorted by Lift@10% (descending). ROC-AUC and PR-AUC measure global discrimination; Recall@10% and Lift@10% measure outreach efficiency under a 10% contact budget.

Figure 3 plots the lift curves for all seven models as a function of the percentage of the population contacted (ranked by predicted risk). The shaded region highlights the operationally relevant zone the top 1–10% of members by predicted risk which corresponds to practical outreach capacity in most Medicaid programs. CatBoost and LinearSVM maintain the highest lift throughout this zone, achieving lifts of 32.2× and 28.6× at the top 1% threshold, respectively. The steep decline in lift from top 1% to top 10% is characteristic of high-imbalance problems where the most predictable cases cluster sharply at the top of the risk distribution. Beyond the top 30%, most models converge toward the random baseline (lift = 1×), indicating that the predictive signal is concentrated in the highest-risk members.

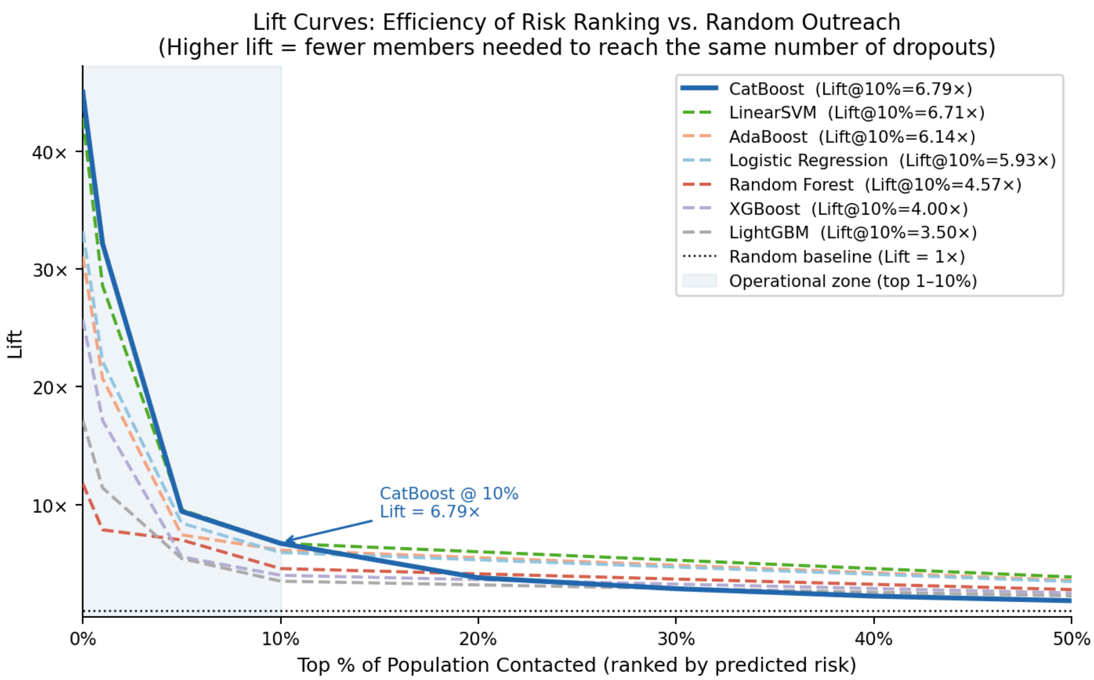

Figure 3. Lift curves for all seven models as a function of the top percentage of members contacted by predicted risk. The shaded region (0–10%) marks the operational zone corresponding to realistic outreach capacity. The dashed horizontal line represents the random baseline (Lift = 1×).

Figure 4 presents the cumulative gains chart for CatBoost alongside the top three competing models (LinearSVM, AdaBoost, Logistic Regression) and the random outreach baseline. The vertical axis reports the cumulative percentage of dropout events captured, and the horizontal axis represents the percentage of the population contacted. CatBoost's curve lies consistently above all competing models, capturing 32.1% of dropouts by contacting just 1% of the population, 47.1% by 5%, and 67.1% by 10%. By contrast, the random baseline requires contacting 67.1% of the population to achieve the same 67.1% recall that CatBoost achieves with only 10%. The shaded region between the CatBoost curve and the random baseline represents the cumulative outreach efficiency gain attributable to the risk-stratification model.

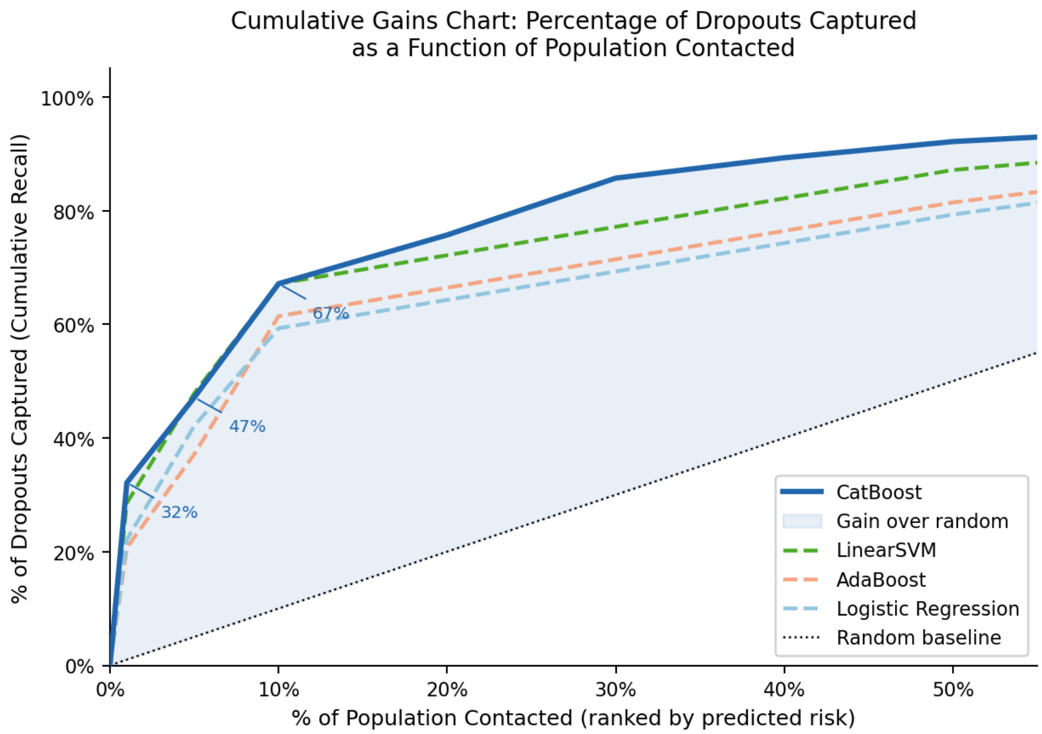

Figure 4. Cumulative gains chart showing the percentage of next-week dropouts captured as a function of the percentage of the population contacted. Key operating points for CatBoost are annotated. The diagonal line represents the random outreach baseline.

Figure 5 displays the CatBoost feature importance scores, with features grouped and color-coded by category: Intensity (recent activity volume), Recency (time since last activity), Continuity (streak length), Instability (switching and dropout history), Trend (directional change), and Demographic. The most important predictor is the number of active weeks in the most recent 4-week window (20.4%), confirming that very recent engagement is the dominant signal for imminent dropout. AgeGroup ranks second (18.1%), indicating that demographic characteristics particularly age-related enrollment patterns carry substantial independent predictive power beyond behavioral history. Sex (8.1%), language group (5.4%), and county of residence (3.9%) further contribute as demographic predictors. Among temporal features, weeks_since_last_active (4.9%) and total_weeks_active (4.8%) are the next most informative signals after the 4-week activity window, reflecting the combined importance of recency and cumulative engagement. Instability measures (switches and dropout events) contribute modestly (2.2–3.1%), suggesting that behavioral irregularity has secondary predictive value relative to recent activity level and demographic context.

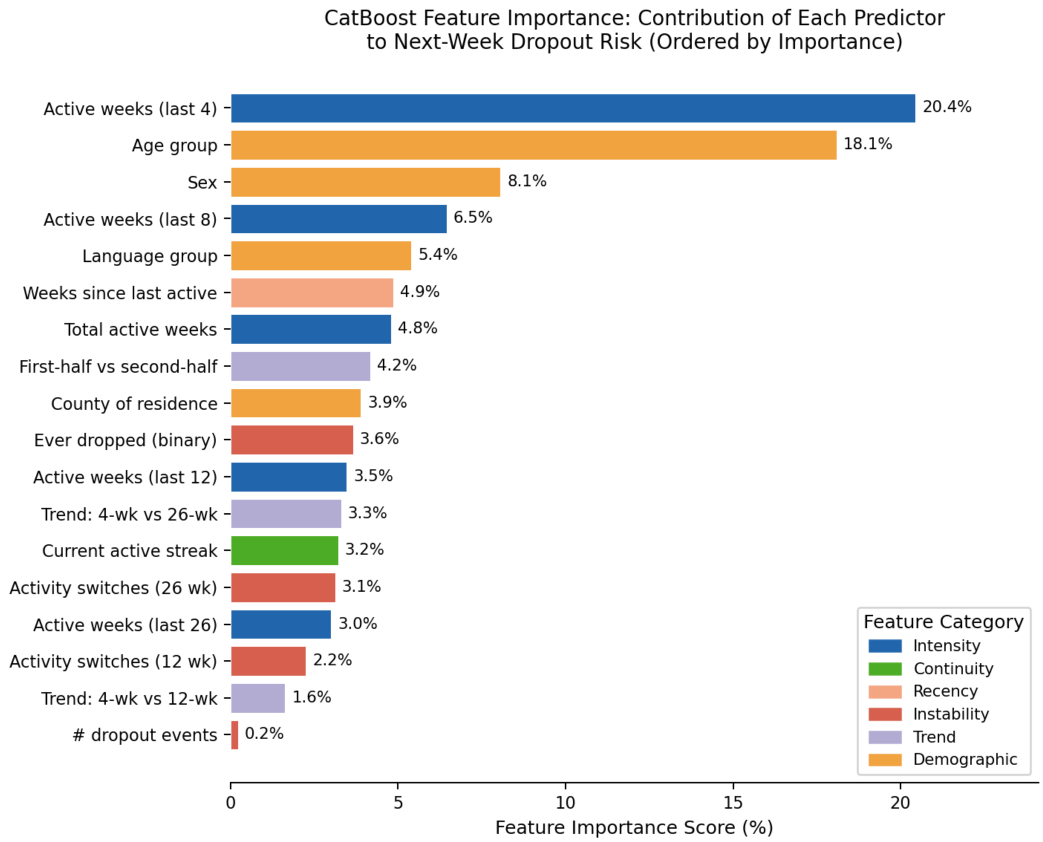

Figure 5. CatBoost feature importance scores for all 18 predictors, grouped and color-coded by feature category. Importance values sum to 100%. Temporal features are divided into Intensity, Continuity, Recency, Instability, and Trend categories; demographic features are shown separately.

5.3 Discussion

Several patterns across models and metrics warrant further discussion. First, the divergence between ROC-AUC and PR-AUC rankings reveals an important distinction for practice. Random Forest has a ROC-AUC of 0.812 but a PR-AUC of only 0.009, whereas LinearSVM achieves a ROC-AUC of 0.864 and a PR-AUC of 0.106 more than ten times higher. This discrepancy arises because ROC-AUC measures ranking quality across all decision thresholds, whereas PR-AUC penalizes heavily for false positives among the most confidently predicted cases. In outreach programs where staff can realistically contact only 1–5% of the enrolled population per week, PR-AUC is a more operationally meaningful criterion than ROC-AUC alone.

Second, CatBoost's advantage over XGBoost and Random Forest is attributable in part to its native handling of categorical features. County of residence and language group carry meaningful interaction structure that one-hot encoding does not efficiently represent at the tree depths used in this study. CatBoost's ordered target-encoding avoids this limitation, allowing the model to learn county- and language-specific dropout risk patterns directly from the data. LinearSVM's strong performance despite using one-hot encoding suggests that a well-calibrated linear decision boundary can partially compensate for this representation gap.

Third, LightGBM consistently underperformed despite the application of imbalance-aware settings (is_unbalance=True, min_child_samples=1). With only 325 positive training examples across 300,912 instances, LightGBM's gradient-based leaf-splitting may not generate sufficient signal for the minority class even after reweighting. Practitioners applying this framework at comparable imbalance levels (>500:1) should treat LightGBM with caution and prioritize CatBoost or LinearSVM as primary model candidates.

5.4 Validation
Model robustness is supported by several design choices. All temporal features are constructed exclusively from Weeks 1 through 51, with Week 52 status reserved for label construction, ensuring no temporal leakage. The train–test split is stratified on the dropout label to preserve the 0.11% positive rate in both partitions, and performance is computed solely on the held-out test set with no reuse of test data during training or model selection. CatBoost's early stopping was triggered at iteration 201 out of a maximum of 3,000, with no further improvement on the validation partition, confirming that the reported performance reflects a well-regularized model rather than an overfit one. The consistency of model rankings across ROC-AUC, PR-AUC, Recall@K, and Lift@K metrics provides additional confidence in the relative comparisons reported in Table 3 and Figures 2–3. Class imbalance was addressed through cost-sensitive training objectives for all models, and all metrics are interpreted using ranking-based measures that are appropriate for severe-imbalance settings.

6. Conclusion
This study demonstrates that short-term Medicaid eligibility loss can be predicted with meaningful accuracy using longitudinal weekly eligibility data augmented by interpretable temporal and demographic features. By formulating dropout as a next-week prediction problem, the proposed framework aligns closely with real-world operational workflows, enabling timely and targeted outreach before coverage disruption occurs. The results show that recent engagement patterns as short-term activity intensity, continuity, and instability provide strong predictive signals, and that incorporating demographic context further enhances model performance.

Evaluation using outreach-oriented ranking metrics highlights the practical value of the approach under severe class imbalance, where only a limited proportion of members can realistically be contacted. In particular, the strong recall and lift achieved by the best-performing model indicate that a substantial share of imminent dropouts can be identified within a small high-risk segment, supporting efficient allocation of outreach resources. Importantly, the use of transparent feature engineering and interpretable models facilitates adoption by healthcare administrators and policy stakeholders, while also enabling monitoring of demographic patterns in predicted risk.

Overall, this work provides a scalable and operationally relevant foundation for proactive Medicaid retention strategies. The framework can be readily extended to incorporate additional administrative signals, such as recertification timelines or outreach history, and offers a generalizable approach for short-horizon risk prediction in other public health and social service programs where continuity of coverage is critical.

References

Chen, J., Jiang, Y., Li, Z., Zhang, M., Liu, L., Li, A. and Lu, H., Predictive machine learning models for anticipating loss to follow-up in tuberculosis patients throughout anti-TB treatment journey, *Scientific Reports*, vol. 14, no. 1, Article 24685, 2024.

Chen, T. and Guestrin, C., XGBoost: A scalable tree boosting system, *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 785-794, 2016.

Daw, J. R., Winkelman, T. N., Dalton, V. K., Kozhimannil, K. B. and Admon, L. K., Medicaid expansion improved perinatal insurance continuity for low-income women, *Health Affairs*, vol. 39, no. 9, pp. 1531-1539, 2020.

Endebu, T., Taye, G. and Deressa, W., Development of a machine learning prediction model for loss to follow-up in HIV care using routine electronic medical records in a low-resource setting, *BMC Medical Informatics and Decision Making*, vol. 25, no. 1, Article 192, 2025.

Goldman, A. L. and Sommers, B. D., Among low-income adults enrolled in Medicaid, churning decreased after the Affordable Care Act, *Health Affairs*, vol. 39, no. 1, pp. 85-93, 2020.

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W. and Liu, T. Y., LightGBM: A highly efficient gradient boosting decision tree, *Advances in Neural Information Processing Systems*, vol. 30, 2017.

McIntyre, A., Morein, M., Kim, J., Figueroa, J. F. and Sommers, B. D., US coverage changes during Medicaid unwinding in 2023, *JAMA Health Forum*, vol. 6, no. 10, Article e253887, 2025.

Nelson, D. B., Goldman, A. L., Zhang, F. and Yu, H., Continuous Medicaid coverage during the COVID-19 public health emergency reduced churning, but did not eliminate it, *Health Affairs Scholar*, vol. 1, no. 5, Article qxad055, 2023.

Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V. and Gulin, A., CatBoost: Unbiased boosting with categorical features, *Advances in Neural Information Processing Systems*, vol. 31, 2018.

Rodrigues, M. M., Barreto-Duarte, B., Vinhaes, C. L., Araújo-Pereira, M., Fukutani, E. R., Bergamaschi, K. B. and Andrade, B. B., Machine learning algorithms using national registry data to predict loss to follow-up during tuberculosis treatment, *BMC Public Health*, vol. 24, no. 1, Article 1385, 2024.

Rumalla, K. C., Nelson, D. B., McConnell, K. J. and Zhu, J. M., Racial and ethnic disparities in Medicaid disenrollment after the end of the COVID-19 public health emergency, *JAMA Internal Medicine*, vol. 184, no. 8, pp. 987-989, 2024.

Saito, T. and Rehmsmeier, M., The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets, *PLOS ONE*, vol. 10, no. 3, Article e0118432, 2015.

Soni, A. and Blackburn, J., Health characteristics of adults unable to complete Medicaid renewal during the unwinding period, *JAMA Health Forum*, vol. 6, no. 3, Article e250092, 2025.

Tuan, W. J., Yan, Y., Abou Al Ardat, B., Felix, T. and Chen, Q., Predicting missed appointments in primary care: A personalized machine learning approach, *The Annals of Family Medicine*, vol. 23, no. 4, pp. 294-301, 2025.

Wang, Z., Guo, Z., Wang, W., Zhang, Q., Song, S., Xue, Y. and Wang, J., Prediction of tuberculosis treatment outcomes using biochemical makers with machine learning, *BMC Infectious Diseases*, vol. 25, no. 1, Article 229, 2025.

Yang, Y., Madanian, S. and Parry, D., Enhancing health equity by predicting missed appointments in health care: Machine learning study, *JMIR Medical Informatics*, vol. 12, no. 1, Article e48273, 2024.

Biographies

Majed Aljamrah is a graduate student in the School of Systems Science and Industrial Engineering at Binghamton University, State University of New York. His research interests include healthcare systems, machine learning, operations research, and data science.

Dr. Yong Wang is an Associate Professor and Associate School Director in the School of Systems Science and Industrial Engineering at Binghamton University, New York, USA. His research interests include autonomous systems, energy systems, manufacturing systems, healthcare systems, operations research, and data science.