05_FinalModelSelection

January 29, 2026

1 Final Model Selection & Recommendation

1.1 Selected Model:

Random Forest Classifier

Selected over Logistic Regression based on: - Higher ROC-AUC (stronger ranking ability) - Superior decision-level performance after threshold tuning - Ability to capture non-linear relationships and feature interactions

While Logistic Regression provided a strong and interpretable baseline, the Random Forest demonstrated meaningful performance gains that justify its additional complexity.

1.2 Decision Threshold

Selected threshold: t = 0.55

- Chosen based on F1-score maximization, balancing precision and recall
- Avoids overly aggressive targeting while maintaining strong churn detection

This operating point reflects a balanced retention strategy, suitable for scenarios where missing churners is costly, but intervention resources are not unlimited.

1.3 Business Interpretation

At the selected threshold, the final model: - Identifies a substantial proportion of churners (high recall) - Reduces wasted retention actions compared to lower thresholds - Provides actionable churn probabilities rather than binary predictions

This enables the business to: - prioritize high-risk customers, - scale interventions according to available budget, - and adjust the threshold dynamically as business constraints change.

1.4 Trade-offs and Considerations

Interpretability:

Logistic Regression remains more transparent; however, feature importance analysis confirms that the Random Forest's decisions align with known behavioral drivers (tenure, product category, payment behavior).

Complexity:

The Random Forest introduces higher computational and conceptual complexity, which is acceptable given the observed performance gains.

Data Leakage:

The model remains leakage-safe, excluding features such as DaySinceLastOrder that could compromise real-world deployment.

1.5 Recommendation

- Deploy the Random Forest model with a decision threshold of 0.55 as the primary churn prediction system.
- Use the Logistic Regression model as:
  - a benchmarking reference,
  - an interpretability aid,
  - or a fallback in low-complexity environments.

1.6 Future Improvements

- Cost-based threshold optimization using explicit business costs
- Model monitoring and recalibration over time
- Feature expansion with time-aware or behavioral aggregates
- Advanced models (e.g. Gradient Boosting) if further performance gains are required