Executive Brief

Glass-Box Cascaded Models for Telecom Churn Prediction
Phillip Harris – November 2025

---

1. Executive Summary

Customer churn is a major challenge for telecom operators, directly affecting revenue and growth. Traditional machine learning models often provide strong predictive performance but are opaque, making it difficult for business teams to understand, trust, or act on predictions.

We developed a fully interpretable, domain-driven cascade architecture that predicts churn with high accuracy while providing clear explanations for each prediction. This approach enables actionable insights, auditability, and operational trust.

Note on Data: All experiments use the IBM synthetic telecom dataset, which is publicly available and widely used for research. Real telecom datasets are rarely accessible due to privacy and regulatory constraints.

---

2. Problem Statement

- Churn leads to lost revenue and increased acquisition costs.

- Existing black-box predictive models are effective but lack transparency.

- Business teams need both accurate predictions and explainable reasoning to act effectively.

---

3. Solution Overview

The solution uses a four-stage interpretable cascade:

1. **Logistic Regression (LR)** – captures linear patterns in customer behavior.

2. **Rule-Based Random Forest (RF)** – identifies interactions and patterns as explicit rules.

3. **Explainable Boosting Machine (EBM)** – captures non-linear relationships with monotonicity constraints.

4. **Meta-EBM Ensemble** – combines outputs from previous stages into a final optimized prediction.

Every stage produces interpretable outputs, allowing stakeholders to understand why a customer is predicted to churn.

[Insert Diagram: Cascade Architecture Overview]

---

4. Results & Impact

- High predictive performance on the IBM synthetic telecom dataset.

- Per-customer explanations enable actionable insights for retention campaigns.

- Domain-driven approach ensures alignment with operational logic and strategy.

Metrics	Value
F2 Score	0.9015 ± 0.0171
Recall	0.8978 ± 0.0225
Precision	0.9172 ± 0.0122
F1 Score	0.9072 ± 0.0101
AUC	0.9873 ± 0.0018

[Insert Chart: Stage-Level Performance Comparison]
[Insert Chart: Feature Importance / Rule Activation Examples]

---

5. Limitations

While the glass-box cascade approach achieves high predictive performance and interpretability, it is **dataset-dependent**. Churn datasets vary significantly across domains and even within telecom, which makes creating a single universal model challenging.

Currently, the model is validated on known datasets (e.g., IBM synthetic telecom data) and can generalize to other datasets with clear provenance. However, we avoid using datasets where the source or data quality is uncertain, as unknown sources may introduce hidden biases or inconsistencies that could compromise interpretability and reliability.

Example future datasets include:

- **Cell2Cell telecom churn datasets**

- **Business-to-business (B2B) churn datasets**

- **Banking churn datasets**

This demonstrates that this approach can deliver **interpretable, high-performance predictions** without relying on black-box models, while acknowledging the limitation that results are dataset-dependent.

---

6. Benefits for the Business

- **Transparency & Trust:** Decisions are fully explainable.

- **Operational Efficiency:** Targeted interventions based on interpretable predictions.

- **Scalable & Reproducible:** Can be adapted to other datasets and telecom environments.

- **Actionable Insights:** Rules and feature contributions directly inform retention strategies.

---

7. Next Steps

- Integrate the cascade into existing analytics pipelines.

- Test this approach on additional datasets for further validation.

- Develop dashboards for operational teams to explore predictions and explanations.

Contact: Phillip Harris
Email: phillipjerald23@gmail.com