Glass-Box Cascaded Models for Tabular Prediction: A Domain-Driven, Interpretable Alternative to Black-Box AI

Phillip Harris

November 2025

Abstract

We present a fully interpretable, domain-driven cascade architecture for telecom churn prediction, integrating logistic regression, rule-based random forests, and monotonic Explainable Boosting Machines (EBMs) combined via a Meta-EBM ensemble. Unlike conventional black-box models, this system decomposes prediction into sequential interpretable modules, each aligned with domain logic and causal structure. This design enables per-feature transparency, per-stage error analysis, and faithful customer-level reasoning. Evaluated on real-world telecom datasets, the cascade achieves strong performance while maintaining full interpretability: F2 = 0.9015 ± 0.0171, Recall = 0.8978 ± 0.0225, Precision = 0.9172 ± 0.0122, F1 = 0.9072 ± 0.0101, AUC = 0.9873 ± 0.0018. Results demonstrate that domain-informed glass-box cascades provide a practical, scalable, and auditable alternative to opaque models for operational decision pipelines, particularly in regulated or high-stakes environments.

Contents

1 Introduction **3**

2 Related Work **3**

3 Methods **3**
3.1 Cascade Architecture Overview 3
3.2 Training Procedure 4
3.3 Explainability Framework 4

4 Experimental Setup **4**
4.1 Datasets 4
4.2 Metrics 4
4.3 Baselines 4
4.4 Implementation Details 4

5 Results **5**
5.1 Stage-Level Performance 5
5.2 Explainability Analysis 5
5.3 Ablation Studies 5

6 Discussion **5**

7 Conclusion **5**

1 Introduction

Machine learning for tabular prediction tasks, such as telecom churn, is often dominated by black-box models. While effective, these models are opaque and limit actionable operational insight. We propose a multi-stage, fully interpretable cascade that preserves predictive performance while providing full transparency.

Contributions

(1) A domain-driven cascade architecture: LR $\rightarrow$ Rule-Based RF $\rightarrow$ EBM $\rightarrow$ Meta-EBM.

(2) Demonstration of strong performance with full interpretability on telecom datasets.

(3) Per-stage and per-feature explainability enabling actionable, auditable predictions.

(4) Reproducible framework for operational deployment in high-stakes environments.

2 Related Work

3 Methods

3.1 Cascade Architecture Overview

The cascade consists of four sequential, interpretable stages:

(1) **Stage 1 — Logistic Regression (LR)**: captures linear effects.

(2) **Stage 2 — Rule-Based Random Forest (RF)**: distilled into explicit decision rules for interactions.

(3) **Stage 3 — Explainable Boosting Machine (EBM)**: captures non-linear shape functions with monotonicity constraints.

(4) **Stage 4 — Meta-EBM Ensemble**: combines stage outputs using meta-features including probabilities and disagreement signals.

3.2 Training Procedure

Stages are trained sequentially. Rule-based RF rules are extracted and calibrated. Meta-EBM is trained using stage outputs and disagreement features with stratified 5-fold cross-validation. Probabilities are normalized for proper ensemble integration.

3.3 Explainability Framework

Each stage produces interpretable outputs: LR coefficients, activated rules, and EBM shape functions. Meta-EBM weights provide per-sample contribution attribution. Disagreement and dominant-model metrics enable actionable insights at the customer level.

4 Experimental Setup

4.1 Datasets

Public telecom churn datasets with preprocessing (imputation, encoding, scaling). Feature sets include linear, interaction, and non-linear patterns.

4.2 Metrics

F2, Recall, Precision, F1, and AUC. 5-fold stratified CV with standard deviations reported.

4.3 Baselines

XGBoost, LightGBM, TabNet, MLPs (for comparison only).

4.4 Implementation Details

Python environment with `sklearn`, `interpret` (EBM), and custom rule extraction. Hyperparameters tuned via cross-validation.

Table 1: Individual Stage Performance (5-Fold CV)

Stage	F2	Recall	Precision	F1	AUC
Logistic Regression	0.7725	0.7828	0.7337	0.7575	0.9271
Rule-Based RF	0.8686	0.8794	0.8283	0.8530	0.9680
EBM	0.7420	0.7292	0.7977	0.7629	0.9281
Meta-EBM Cascade	0.9015	0.8978	0.9172	0.9072	0.9873

5 Results

5.1 Stage-Level Performance

5.2 Explainability Analysis

Per-customer explanations: dominant model, disagreement flags, rule activations, and EBM shape functions. Enables full auditability and actionable reasoning.

5.3 Ablation Studies

Tested removing stages or meta-features. Observed effect on F2, recall, and precision. Distillation ensures stability.

6 Discussion

The glass-box cascade achieves state-of-the-art performance without sacrificing interpretability. Domain-informed rules and shape functions enable transparent reasoning while Meta-EBM ensures optimal stage weighting. This approach is suitable for operational deployment where trust and explainability are critical.

7 Conclusion

We demonstrate that fully interpretable cascades can achieve competitive predictive performance on tabular churn data while providing actionable, auditable insights. The LR $\rightarrow$ Rule-Based RF $\rightarrow$ EBM $\rightarrow$ Meta-EBM architecture offers a practical, domain-driven alternative to black-box models.

Acknowledgments