GLASS-BRW

Precision-Gated Rule Ensemble with Abstention

Disclaimer

This document outlines a theoretical framework and initial system design. Parameter values and constraints are subject to change as experimentation progresses; however, the overarching methodology and design philosophy will be preserved.

1 Problem Setting

- $X \in \mathbb{R}^d$ denotes customer feature vectors.
- $Y \in \{0, 1\}$ denotes the target variable (NO SUBSCRIBE, SUBSCRIBE).
- $S = \phi(X)$ denotes a deterministic segment assignment function.

$$S = \phi(X)$$

- The function $\phi(\cdot)$ maps each customer to a finite categorical lattice.
- Segment assignments are deterministic and fixed at inference time.
- GLASS-BRW does **not** learn a monolithic classifier $f : X \to Y$.
- Instead, it learns a **sequential, partial decision system** composed of two ordered rule passes with abstention:

$$f_{\text{GLASS}} = f^{(2)} \circ f^{(1)}$$

- $f^{(1)}$ is a **false-negative risk router**: it emits a routing decision (ROUTE or ABSTAIN) when a rule identifies elevated risk of missing a true subscriber.
- $f^{(2)}$ is a **high-precision subscriber detector**: it emits a class prediction ($Y = 1$) when a rule fires, and otherwise abstains.
- Neither stage emits a forced negative prediction; abstained samples are resolved by the Meta-EBM or left abstained when no model is sufficiently certain.

2 Rule Representation

- A rule r is defined as an ordered pair:

$$r = (C_r, y_r)$$

- C_r is a conjunction of segment conditions.
- $y_r \in \{0, 1\}$ is the predicted class *when applicable*.

$$C_r = \{(f_1 = \ell_1), \ldots, (f_k = \ell_k)\}$$

- Each $(f_i = \ell_i)$ corresponds to a deterministic segment constraint.
- Rule complexity is bounded by $k \leq 3$.
- A customer x matches rule r if and only if:

$$\forall (f, \ell) \in C_r : \phi_f(x) = \ell$$

- For Pass 1 routing rules, y_r is unused and the rule encodes a routing decision.
- For Pass 2 decision rules, $y_r = 1$ denotes a SUBSCRIBE prediction.

3 Candidate Rule Generation

3.1 Feature Pre-Filtering

Let:

- F denote the full set of segment features
- L_f denote the discrete levels of feature f
- allowed_features $\subseteq F$ be a pre-selected subset chosen via permutation importance or mutual information (typically top 20–40 features)

3.2 Lattice Enumeration with Quality Gates

Candidate rules are generated via bounded lattice enumeration using only features in allowed_features:

$$\mathcal{R} = \bigcup_{k=1}^{3} \left\{ r \;\middle|\; C_r \subseteq \prod_{i=1}^{k} (f_i, \ell_i),\ f_i \in \text{allowed\_features},\ |\text{supp}(r)| \geq n_{\min} \right\}$$

where $n_{\min}$ is a minimum support threshold (e.g. 50 - 150 samples).

3.3 Rule Evaluation Context

Each candidate rule is evaluated according to its intended role in the GLASS-BRW cascade:

- Pass 1 rules are evaluated as **false-negative risk routers**
- Pass 2 rules are evaluated as **high-precision subscriber detectors**

3.4 Enhanced Quality Metrics

For each candidate rule r, metrics are computed on held-out validation data. Metrics are interpreted relative to the role of the rule.

Base Statistics

- True positives TP_r, false positives FP_r, true negatives TN_r, false negatives FN_r
- Rule quality score:

$$p_r = \begin{cases} \frac{\mathrm{TP}_r}{\mathrm{TP}_r+\mathrm{FP}_r}, & \text{if } r \text{ is a Pass 2 decision rule} \\ \frac{\mathrm{TP}_r}{\mathrm{TP}_r+\mathrm{FN}_r}, & \text{if } r \text{ is a Pass 1 routing rule} \end{cases}$$

- Coverage: $c_r = \frac{|\{x : x \models r\}|}{|X|}$

Information-Theoretic Metrics

- Information gain: $\mathrm{IG}_r = H(Y) - H(Y \mid r)$
- Segment entropy: $H_{\text{seg}} = -p_r \log_2(p_r) - (1-p_r)\log_2(1-p_r)$

Business-Aligned Metrics

- Lift (Pass 2 decision rules only):

$$\text{lift}_r = \frac{p_r}{\text{base\_rate}_{\text{subscribe}}}$$

- Relative recall (Pass 2 decision rules only):

$$\text{rr} = \frac{\mathrm{TP}_r}{\text{total\_subscribers}}$$

- Absolute recall: TP_r

3.5 Multi-Stage Filtering

Rules must pass all of the following gates to enter the candidate pool $\mathcal{R}$.

Pass 2 (Subscriber Detection) Rules

1. $p_r \geq p_{\min}$ (high-precision requirement)
2. $\mathrm{IG}_r \geq \mathrm{min\_info\_gain}_{\mathrm{subscribe}}$
3. $H_{\mathrm{seg}} \leq \mathrm{max\_entropy}_{\mathrm{subscribe}}$
4. $\mathrm{lift}_r \geq \mathrm{min\_lift}_{\mathrm{subscribe}}$
5. $\mathrm{rr} \geq \mathrm{min\_recall}_{\mathrm{subscribe}}$ and $\mathrm{TP}_r \geq \mathrm{min\_tp}_{\mathrm{subscribe}}$

Pass 1 (FN-Risk Routing) Rules

1. $p_r \geq p_{\min}$ (minimum sensitivity to subscriber presence)
2. $\mathrm{IG}_r \geq \mathrm{min\_info\_gain}_{\mathrm{route}}$
3. $H_{\mathrm{seg}} \leq \mathrm{max\_entropy}_{\mathrm{route}}$
4. Support and stability thresholds to ensure reliable routing

The resulting filtered rule pool typically satisfies $|\mathcal{R}| \approx 200$–$500$.

3.6 Ensemble-Aware Metrics

The following metrics will assess complementarity with the Random Forest and will not affect rule firing directly.

Given a Random Forest with output $\hat{p}_{\mathrm{RF}}(x)$:

Boundary Ambiguity

$$a_r = \mathbb{E}_{x \models r}\left[|\hat{p}_{\mathrm{RF}}(x) - 0.5|\right]$$

- Lower values will indicate regions where RF exhibits higher uncertainty
- Such regions will be preferred for rule-based handling

RF Overlap

$$o_r = \mathbb{P}_{x \models r}\left(|\hat{p}_{\mathrm{RF}}(x) - 0.5| > 0.20\right)$$

- Higher values will indicate redundancy with confident RF predictions

4 Rule Metrics (Validation-Based)

All metrics are computed on held-out validation data and are interpreted relative to the role of the rule (routing or detection).

4.1 Precision (Primary Gate)

$$p_r = \begin{cases} \frac{\text{TP}_r}{\text{TP}_r+\text{FP}_r}, & \text{for Pass 2 (decision) rules} \\ \frac{\text{TP}_r}{\text{TP}_r+\text{FN}_r}, & \text{for Pass 1 (routing) rules} \end{cases}$$

- Pass 2 precision measures subscriber detection accuracy
- Pass 1 precision measures sensitivity (recall) for capturing true subscribers at risk of false-negative error
- Rules with $p_r < p_{\min}$ (typically 0.75) are ineligible

4.2 Coverage

Let

$$\text{Survivors}_{(1)} = \{x : \exists r \in \mathcal{R}^{(1)} \text{ s.t. } x \models r\}$$

denote the set of samples routed by Pass 1.

$$c_r = \begin{cases} \frac{|\{x : x \models r\}|}{|X|}, & \text{for Pass 1 rules} \\ \frac{|\{x : x \models r\}|}{|\text{Survivors}_{(1)}|}, & \text{for Pass 2 rules} \end{cases}$$

- Measures the fraction of the relevant population covered by rule r

4.3 Interpretability

$$i_r = \frac{1}{|C_r| + 1}$$

- Penalizes rule complexity
- Biases selection toward simpler symbolic explanations

4.4 Stability

$$\text{stability}_r = 1 - \min\left(1, \frac{\text{std}(p_r^{(j)})}{\mathbb{E}[p_r^{(j)}]}\right)$$

- $p_r^{(j)}$ denotes precision across cross-validation folds
- Bounded to $[0, 1]$ to handle low-precision edge cases
- Low stability indicates non-stationarity or sample noise

5 Rule Selection as Integer Optimization

Rule selection is performed **independently per pass**, with pass-specific objectives and constraints. The formulation below applies generically to either pass.

5.1 Decision Variables

$$x_r \in \{0,1\} \quad \forall r \in \mathcal{R}$$

5.2 Objective Function

$$\max \sum_{r \in \mathcal{R}} x_r \left(p_r^2 c_r i_r - \lambda_1 a_r - \lambda_2 o_r \right)$$

- Quadratic precision enforces superlinear preference for high-purity rules
- Penalties encode complementarity with downstream models

5.3 Constraints

Cardinality

$$8 \leq \sum_r x_r \leq 10$$

Minimum Coverage

$$\sum_r x_r c_r \geq \tau_{\text{cov}}$$

where τ_{cov} is pass-dependent (e.g. 0.60 for Pass 1, survivor-relative for Pass 2).

Precision Gate

$$x_r = 0 \quad \text{if } p_r < p_{\min}$$

Information Gain Floor

$$x_r = 0 \quad \text{if } \mathrm{IG}_r < \text{min\_info\_gain}_{\text{pass}}$$

Entropy Ceiling

$$x_r = 0 \quad \text{if } H_{\text{seg}} > \text{max\_entropy}_{\text{pass}}$$

Decision-Rule Constraints (Pass 2 Only) The following constraints apply only to subscriber detection rules:

$$x_r = 0 \quad \text{if } \text{lift}_r < \text{min\_lift}_{\text{subscribe}}$$

$$x_r = 0 \quad \text{if } \left(\mathrm{TP}_r < \text{min\_tp}_{\text{subscribe}} \ \text{ or } \ \mathrm{rr} < \text{min\_recall}_{\text{subscribe}}\right)$$

where

$$\mathrm{rr} = \frac{\mathrm{TP}_r}{\text{total\_subscribers}}$$

is relative recall.

- Absolute thresholds prevent spurious rules from small folds
- Relative thresholds ensure population-level impact

Routing-Rule Constraints (Pass 1 Only) Routing rules are subject to stricter stability and support requirements but do not emit class predictions so that we avoid any possible false negatives in Pass 1.

Feature Diversity

$$\sum_{r: f \in C_r} x_r \leq 3 \quad \forall f \in F$$

6 Execution Semantics (Sequential First-Match-Wins)

Selected rules are applied sequentially by pass.

1. **Pass 1 (Routing)**: Routing rules are evaluated in descending precision order. If a routing rule fires, the sample is forwarded to Pass 2; otherwise the sample abstains from the GLASS-BRW system.

2. **Pass 2 (Detection)**: Subscriber detection rules are evaluated only on routed samples. If a rule fires, a SUBSCRIBE prediction is emitted; otherwise the sample abstains.

Let

$$\text{Survivors}_{(1)} = \{x : \exists r \in \mathcal{R}^{(1)} \text{ s.t. } x \models r\}.$$

The GLASS-BRW output is defined as:

$$f_{\text{GLASS}}(x) = \begin{cases} 1, & x \in \text{Survivors}_{(1)} \ \wedge \ \exists r \in \mathcal{R}^{(2)} \text{ s.t. } x \models r \\ -1, & \text{otherwise (abstain)} \end{cases}$$

- A sample abstains if either Pass 1 does not route it or Pass 2 does not match it
- Mutual exclusivity within each pass
- Deterministic evaluation order
- Single-rule attribution for all non-abstained predictions

7 Probabilistic Output Encoding

GLASS-BRW emits calibrated probabilities only for positive predictions. Abstention explicitly encodes uncertainty and defers resolution to downstream models.

$$P(Y = 1 \mid x) = \begin{cases} p_r, & f_{\text{GLASS}}(x) = 1 \\ \texttt{NULL}, & f_{\text{GLASS}}(x) = -1 \end{cases}$$

- Positive predictions use the validation precision p_r of the firing Pass 2 rule
- Abstained samples output `NULL` (no probability assigned)

- The Meta-EBM will resolve `NULL` cases using its own confidence estimates
- This avoids the semantic error of encoding abstention as $P(Y = 1) = 0.5$

8 Guarantees

Theorem 1 (Pass 2 Precision Lower Bound)

$$\mathbb{P}(Y = \hat{Y} \mid f_{\text{GLASS}}(x) = 1) \geq p_{\min}$$

- Every non-abstained prediction corresponds to exactly one Pass 2 rule with precision $\geq p_{\min}$

Theorem 2 (Interpretability Bound)

- Every GLASS-BRW prediction is explained by at most three human-interpretable segment conditions from the firing Pass 2 rule, plus implicit routing constraints from Pass 1

Theorem 3 (Abstention-Aware Complementarity)

- For $\lambda_1 > 0$, the optimizer prefers rules that avoid internally ambiguous GLASS decision regions, as measured by proximity to the abstention boundary $p_r \approx 0.5$.
- This promotes complementary decision-making: GLASS-BRW captures high-precision symbolic patterns, while downstream models (e.g., the Meta-EBM) resolve abstained or ambiguous cases.
- The boundary ambiguity metric a_r quantifies this effect as the expected distance from the decision threshold within a rule's support.