Churn Algorithm Variable Architecture

Formal Proposal for Predicting Investor Subscription Risk

Juan Arroyo & Martine Dirkzwager

March 26, 2026

1. Objective

The purpose of this document is to define a first formal architecture for the variables that could feed a future **Churn Algorithm** focused on predicting the probability that an investor stops being subscribed to the platform.

The goal is not to explain churn only after it happens, but to identify **leading indicators** that can anticipate it early enough to enable intervention, expectation management, or retention actions.

This document is intended to support implementation by the Data team and should be understood as a first structured inventory of variables, modules, and predictive hypotheses.

2. Core Predictive Principle

Investor churn should not be interpreted only as a commercial or relational issue. In many cases, churn emerges from a combination of:

- weak marketplace reciprocity;
- repeated friction during the search process;
- low speed of counterpart reaction;
- low density of relevant opportunities;
- poor perceived value during the first weeks of subscription;
- deterioration in engagement and willingness to collaborate.

The Churn Algorithm should therefore focus on variables that capture:

- **frustration**;
- **lack of perceived progress**;
- **mismatch between investor thesis and available supply**;
- **declining interaction quality over time**;
- **risk accumulation during the subscription lifecycle**.

3. Proposed Churn Architecture

A first formal architecture for churn prediction may be organized into five modules:

1. Sentiment Module
2. Response Friction Module
3. Dealflow Fit / Reciprocity Module
4. Resilience Module
5. Search and Thesis Complexity Module

Each module captures a different predictive dimension of investor attrition.

4. Sentiment Module

4.1. Predictive role

Sentiment is relevant because churn often has an emotional and experiential component before it becomes a behavioral fact.

An investor may still be subscribed while already showing signs of:

- anxiety,
- impatience,
- disappointment,
- or loss of confidence in the platform.

For this reason, Sentiment should be modeled as an evolving indicator of the investor's perceived experience.

4.2. Main inputs

- calls and emails from the FU Account function;
- early platform receptivity during the first weeks after subscription;
- repeated absence of meaningful responses after investor actions;
- repeated lack of progress across multiple opportunities.

4.3. Possible labels

- Very Anxious
- Anxious
- Neutral
- Calm
- Very Calm

4.4. Predictive hypothesis

A deterioration in Sentiment, especially early in the subscription lifecycle, may serve as one of the clearest forward-looking indicators of churn risk.

In practice:

- persistent *Anxious* or *Very Anxious* states should correlate with higher churn probability;
- stable *Calm* states should correlate with stronger perceived value and lower churn probability.

5. Response Friction Module

5.1. Predictive role

Many churn cases are not driven by explicit rejection, but by repeated friction.

What tends to matter is not only whether the investor is active, but whether the platform and the counterpart ecosystem respond in a timely and credible way to that activity.

The Response Friction Module should measure how often investors act meaningfully and then encounter delays or non-resolution.

5.2. Core waiting-time variables

5.2.1. Deal Interest → Deal Accepted

Measures how quickly counterparties react after the investor sends an initial expression of interest.

Predictive logic: if an investor repeatedly sends connections and the transition to Deal Accepted is slow or weak, perceived responsiveness decreases and churn risk rises.

5.2.2. Q&A Sent → Q&A Answered

Measures the time required for companies to answer investor questions.

Predictive logic: when investors invest time in analysis and receive delayed feedback, the platform may be perceived as low quality or operationally slow.

5.2.3. Meeting Created → Meeting Accepted

Measures how quickly meeting requests translate into actual confirmation.

Predictive logic: this is one of the clearest signals of whether the marketplace is converting interest into concrete interaction.

5.3. Friction at scale

The strongest churn signal is often not one isolated delay, but repeated delay across volume.

Examples:

- many connections sent + slow Deal Accepted;
- many Q&As sent + slow Q&A Answered;
- many meeting requests + weak acceptance conversion.

5.4. Predictive hypothesis

The probability of churn should increase significantly when:

- investor effort is high;
- reciprocal progress is low;
- this pattern repeats across several deals.

6. Dealflow Fit / Reciprocity Module

6.1. Predictive role

One of the strongest structural causes of churn is simply that the platform does not contain enough relevant opportunities for a given investor thesis.

For this reason, Dealflow Fit should be treated as a predictive variable of subscription sustainability.

6.2. Formal definition

Dealflow Fit (DF) estimates the real offer the platform can provide to a specific investor, based on the investor's thesis and the weighted quality of matching companies.

The complete universe of companies should include:

- companies with active subscription;
- companies without subscription but with complete registration;
- companies without subscription and with incomplete registration.

Matching criteria should be derived from:

- Sector (CNAE)
- Investment range / ticket
- Geography / location

The proposed formula is:

$$DF = (n_{premium} \times 1.0) + (n_{plataforma\_completa} \times 0.6) + (n_{registro} \times 0.2)$$

Where:

- DF: Dealflow Fit
- $n_{premium}$: number of Premium subscribed companies
- $n_{plataforma\_completa}$: number of complete-profile companies without active subscription
- $n_{registro}$: number of basic-registration companies without active subscription

6.3. Predictive hypothesis

A low DF suggests that:

- the investor will see too few relevant opportunities;
- perceived platform value will likely decline early;
- churn risk due to *no match found* is elevated.

A high DF suggests that:

- the investor has enough relevant supply to explore;
- the marketplace can sustain perceived value longer;
- churn risk from lack of fit is lower.

6.4. Operational hypothesis

As an initial working assumption, before internal calibration:

- DF > 8: low risk
- DF between 4 and 8: medium risk
- DF < 4: high risk

7. Resilience Module

7.1. Predictive role

Not all investors react equally to friction. Some tolerate temporary scarcity, slow answers, or low marketplace activity. Others disengage quickly.

Resilience is therefore a useful predictor of churn because it measures the investor's **tolerance to adverse operating conditions**.

7.2. Definition

Platform Resilience should measure how consistently the investor remains active despite:

- delayed responses,
- imperfect supply,
- temporary inactivity in the funnel,
- or repeated weak progress.

7.3. Possible outputs

- Resilience Index
- Resilience band (low / medium / high)
- Resilience evolution over time

7.4. Predictive hypothesis

A low-resilience investor is more likely to churn under conditions of:

- slow counterpart reaction;
- poor dealflow fit;
- repeated dead ends in the early process.

A high-resilience investor may still remain subscribed despite temporary weakness in one or more of these factors.

8. Search and Thesis Complexity Module

8.1. Predictive role

The complexity and articulation of an investor's search behavior can help predict churn because they reveal how precise, narrow, or demanding the investor's expectations may be.

An investor with highly specific search behavior may be more exposed to disappointment if the available supply does not match those constraints.

8.2. Main variables

- number of saved searches;
- number of filters per search;
- diversity of search filters;
- evidence of search refinement over time;
- thesis complexity proxies derived from structured fields.

8.3. Predictive hypothesis

Higher search / thesis complexity may correlate with:

- stronger specificity of expectations;
- lower tolerance to generic supply;
- higher sensitivity to no-match situations;
- and therefore higher churn risk when Dealflow Fit is weak.

9. Post-Meeting Experiential Outcome Signals

9.1. Predictive role

The investor's perception of value is strongly affected by what happens after the first real interactions.
Outcomes after Meeting Completed may function as highly informative experiential signals.

9.2. Suggested state transitions

- Meeting Completed → Analysis: Sentiment becomes *Neutral*
- Meeting Completed → Qualified: Sentiment becomes *Calm*
- Meeting Completed → Lost: Sentiment becomes *Anxious*

9.3. Predictive hypothesis

A user whose post-meeting outcomes accumulate as:

- repeated *Lost*,
- low conversion into *Qualified*,
- or weak post-meeting continuity

is likely to perceive lower value and face higher risk of churn.

10. Suggested Variable Inventory

Variable family	Suggested formal variables
Dynamic Sentiment	sentiment label, sentiment trend, sentiment volatility
Counterpart waiting times	deal interest to accepted time, Q&A sent to answered time, meeting created to accepted time
Friction at scale	delayed response rate weighted by action volume, repeated no-progress rate
Platform Resilience	resilience index, resilience band, resilience trend
Search / Thesis Complexity	saved searches count, filters per search, search refinement depth, thesis complexity proxy
Dealflow Fit / Reciprocity	DF score, DF risk band, matching premium count, matching complete count, matching basic count
Post-meeting experiential outcomes	analysis outcome state, qualified outcome state, lost outcome state

11. Suggested Modeling Strategy

11.1. Stage 1: First predictive baseline

Start with the variables that are both highly informative and easier to operationalize:

- Dealflow Fit / Reciprocity
- counterpart waiting times
- post-meeting experiential states

11.2. Stage 2: Expand to richer behavioral risk logic

Then add:

- friction-at-scale composites;
- resilience signals;
- recency-weighted deterioration measures.

11.3. Stage 3: Add precision layers

Finally add:

- search / thesis complexity;
- richer sentiment state transitions;
- longitudinal churn-risk trajectories.

12. Recommended Outputs

A future Churn Algorithm should ideally return:

- churn probability score;
- churn risk band (low / medium / high);
- dominant churn driver;
- recency-adjusted risk trend;
- suggested intervention priority.

13. Conclusion

A strong Churn Algorithm should not only identify investors who are at risk of leaving, but also explain **why** that risk is building.

The variables defined in this document are especially useful because they capture:

- weak perceived value,
- insufficient relevant supply,
- repeated friction,
- low marketplace reciprocity,
- and deteriorating investor experience.

In practical terms, these are precisely the dimensions most likely to predict whether an investor will stop being subscribed to the platform.

This document should therefore serve as the first formal variable architecture for future Churn Algorithm implementation with the Data team.