

ICS v2.6 Technical Architecture

Investor Classification System at Deale

Juan Arroyo & Martine Dirkzwager

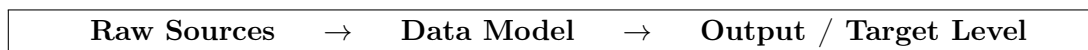
April 10, 2026

1. System Purpose

The Investor Classification System (ICS) is a taxonomic model designed to classify investors into business-relevant target levels based on structural quality, observed behaviour, and execution/supply fit inside Deale's ecosystem.

2. System Overview

ICS transforms heterogeneous investor signals into a consistent operational output for internal business teams.



3. ICS Model Architecture

ICS v2.6 is organized into three main layers:

1. **Structural Layer**
2. **Behavioural Layer**
3. **Execution & Supply Layer**

Each layer captures a different dimension of investor quality:

- **Layer 1** captures profile structure and declared investment capacity.
- **Layer 2** captures actual post-subscription behaviour and progression through the funnel.
- **Layer 3** captures whether the investor's evolving real preferences fit Deale's available or realistically sourceable supply.

4. Layer 1: Structural Layer

Layer 1 is intended to evaluate the structural quality of the investor using registration, onboarding, and explicitly declared profile information.

4.1. Variables and weights

Variable	Weight
TicketScore	0.30
InvestorTypeScore	0.25
MAExperienceScore	0.25
ThesisCompletionScore	0.20

4.2. Investor Type

Investor type is treated as a structural proxy for expected professional sophistication.

Investor Type	Score
INDIVIDUAL	0.50
SEARCH_FUND	0.60
INDUSTRIAL	0.70
FAMILY_OFFICE	0.80
PRIVATE_EQUITY	0.90
ADVISOR	0.50

4.3. TicketScore

The proposed ticket mapping is:

Declared Ticket Range	TicketScore
$\leq 250k$	0.00
250k – 300k	0.08
300k – 500k	0.18
500k – 1M	0.35
1M – 2M	0.55
2M – 5M	0.72
5M – 10M	0.85
> 10M	1.00

4.4. M&A Experience

Closed Deals	MAExperienceScore
0 deals	0.00
1 deal	0.35
2–3 deals	0.65
4+ deals	1.00

4.5. Thesis Completion

The structural thesis signal should reflect whether the investor has provided enough explicit information to support a meaningful classification.

4.5.1. Sub-weights

Sub-variable	Weight
SavedSearchesScore	0.60
ThesisTabsCompletedScore	0.30
FavCompaniesScore	0.10

4.5.2. Formal representation

$$\begin{aligned} ThesisCompletionScore = & 0.60 \cdot SavedSearchesScore \\ & + 0.30 \cdot ThesisTabsCompletedScore \\ & + 0.10 \cdot FavCompaniesScore \end{aligned}$$

4.6. L1 formal representation

$$\begin{aligned} L1 = & 0.30 \cdot TicketScore \\ & + 0.25 \cdot InvestorTypeScore \\ & + 0.25 \cdot MAExperienceScore \\ & + 0.20 \cdot ThesisCompletionScore \end{aligned}$$

4.7. Layer outputs

$$L1_Score = L1$$

$$L1_eligibility_flag = \begin{cases} 1 & \text{if minimum Layer 1 information is available} \\ 0 & \text{otherwise} \end{cases}$$

This structural model is especially relevant for leads and registered investors that are not yet subscribed. If the minimum structural information is not available, then the target level should not be displayed in internal tools. For subscribed users, Layers 2 and 3 should progressively take precedence as new evidence accumulates.

5. Layer 2: Behavioural Layer

Layer 2 measures how the investor behaves in practice after subscription. It is intended to capture actual usage, responsiveness, and funnel progression.

5.1. Layer 2 structure

Layer 2 is composed of four main components:

1. EngagementScore
2. AMScore
3. EarlyFunnelScore
4. AdvancedFunnelScore

5.2. Engagement

This component captures actual interaction with the platform.

5.2.1. Inputs

Variable	Description
SessionTime	Total recent session time
WeeklyRecurrence	Number of active weeks with at least one session
SpecificProjectsAbsoluteTime	Absolute time on accepted projects with >2 views
ProjectViews $\times$ AvgProjectAnalysisTime	Crossed attention variable over different projects

5.2.2. Weights

Variable	Weight
SessionTime	0.10
WeeklyRecurrence	0.20
SpecificProjectsAbsoluteTime	0.30
ProjectViews $\times$ AvgProjectAnalysisTime	0.40

5.2.3. Formal representation

$$\begin{aligned} \text{EngagementScore} = & 0.10 \cdot \text{SessionTimeScore} \\ & + 0.20 \cdot \text{WeeklyRecurrenceScore} \\ & + 0.30 \cdot \text{SpecificProjectsAbsoluteTimeScore} \\ & + 0.40 \cdot \text{CrossProjectAttentionScore} \end{aligned}$$

Where:

$$\text{CrossProjectAttentionScore} = f(\text{ProjectViews} \times \text{AvgProjectAnalysisTime})$$

Normalization for this crossed term should be defined with the Data team.

5.3. Responsiveness to AM Team

This component measures responsiveness to AM / CS activity.

5.3.1. Inputs

Variable	Description
CallsResponseRatio	Recent responsiveness to calls
EmailReplyRatio	Recent responsiveness to emails
AMResponseTime	Delay buckets: 2d, 7d, 15d, >15d

5.3.2. Response-time scoring

$$AMResponseTimeScore = 1.0 \cdot R_{2d} + 0.7 \cdot R_{7d} + 0.3 \cdot R_{15d} + 0.1 \cdot R_{>15d}$$

5.3.3. Crossed recent responsiveness

$$RecentResponsivenessCross = f(\text{CallsResponseRatio}, \text{EmailReplyRatio})$$

Calls should be considered over recent windows of 2 and 7 days. Emails should be considered over 2, 7, and 15 days. Exact normalization and weights should be defined with the Data team.

5.3.4. Formal representation

$$AMScore = f(\text{RecentResponsivenessCross}, \text{AMResponseTimeScore})$$

5.4. Early Funnel Score

This component captures early-stage action when the investor has not yet accumulated advanced funnel evidence.

5.4.1. Inputs

Variable	Unit
Deal Interest Sent	Count / week
Q&A Sent	Count / week
Meeting Requests	Count / week

5.4.2. Applicability caveat

This module is only applicable when the top 3 maximum deal stages of the investor do **not** reach at least the Analysis stage or beyond. If the investor has already reached Analysis or a more advanced phase, then the early-stage component should be temporarily blocked.

5.4.3. Formal representation

$$EarlyFunnelScore = w_1 \cdot DealInterestSentScore + w_2 \cdot QASentScore + w_3 \cdot MeetingRequestsScore$$

5.5. Advanced Funnel Score

This component captures stronger evidence of actual buying behaviour.

5.5.1. Inputs

Variable	Temporal logic
AnalysisCount	Intemporal
QualifiedCount	Intemporal
PreLOICount	Intemporal
LOICount	Intemporal
DDCount	Intemporal
DDFinalizedCount	Intemporal
FundedCount	Intemporal

These variables should be treated as historical evidence that remains positively informative regardless of recency. Time-window logic may be useful for churn prediction, but not for the classification model itself.

5.5.2. Formal representation

$$\begin{aligned} AdvancedFunnelScore = & 0.07 \cdot AnalysisScore \\ & + 0.08 \cdot QualifiedScore \\ & + 0.10 \cdot PreLOIScore \\ & + 0.25 \cdot LOIScore \\ & + 0.25 \cdot DDScore \\ & + 0.05 \cdot DDFinalizedScore \\ & + 0.20 \cdot FundedScore \end{aligned}$$

5.6. L2 formal representation

5.6.1. Case A: if AdvancedFunnelScore exists

$$L2 = 0.20 \cdot EngagementScore + 0.15 \cdot AMScore + 0.25 \cdot EarlyFunnelScore + 0.40 \cdot AdvancedFunnelScore$$

5.6.2. Case B: if no AdvancedFunnelScore exists

$$L2 = 0.25 \cdot EngagementScore + 0.25 \cdot AMScore + 0.50 \cdot EarlyFunnelScore$$

5.6.3. Case C: if neither AdvancedFunnelScore nor EarlyFunnelScore exists

$$L2 = 0.60 \cdot EngagementScore + 0.40 \cdot AMScore$$

6. Layer 3: Execution & Supply Layer

Layer 3 condenses the investor's evolving real preferences and their fit with Deale's available supply.

6.1. Core logic

In v2.6, Layer 3 is intentionally simplified at score level:

$$L3 = DealflowFit$$

This does **not** mean that Layer 3 only uses one raw piece of information. Instead, it means that the final scoring variable of this layer is a synthetic fit measure that condenses information from the investor's evolving real thesis.

6.2. Real thesis and inferred ticket

Real Thesis should progressively infer what the investor actually seems interested in, based on behavioural evidence and real company interaction. It may include:

- sectors;
- revenue;
- EBITDA;
- location;
- other filters;
- inferred ticket.

The key idea is that ticket is one dimension of the real thesis. Therefore, instead of modelling ticket and thesis separately at Layer 3, the system should infer a broader preference vector.

6.3. Real Ticket Inference

6.3.1. Calculation basis

Real Ticket Inference should be estimated from actual company evidence. Companies with stronger progression through the pipeline should carry more importance than companies with weak or superficial evidence.

Sector multiple and latest-year EBITDA may be used as proxies for implied market valuation and, therefore, actual ticket appetite.

6.3.2. Weighting logic for company evidence

Each company in the inference set should receive a composite evidence weight:

$$\begin{aligned} \textit{CompanyEvidenceWeight} = & \alpha \cdot \textit{StageWeight} \\ & + \beta \cdot \textit{AttentionWeight} \\ & + \gamma \cdot \textit{RecencyWeight} \end{aligned}$$

Where:

- **StageWeight** increases with deeper progression through the pipeline;
- **AttentionWeight** captures views, reading time, shortlist attention, and related evidence;
- **RecencyWeight** helps the inference remain updated when behaviour shifts over time.

6.3.3. Output structure

The inferred output should be visible operationally as a monthly updated investor card or baseball card including:

- inferred sectors;
- inferred revenue / EBITDA preferences;
- inferred location;
- inferred ticket;
- degree of consistency with the originally declared thesis.

6.4. Why Dealflow Fit is the only Layer 3 score

Dealflow Fit is the protagonist and only scoring variable of Layer 3 because it condenses the most relevant execution-side information: whether the investor's **real thesis** — including inferred ticket — matches Deale's realistically available or sourceable company universe.

The purpose of preserving the more granular components (real thesis and inferred ticket) is not to score them independently at this stage, but to:

- maintain an updated operational view of the investor's actual preferences;
- make these preferences visible to internal teams;
- update them monthly once sufficient behaviour exists;
- monitor whether investor behaviour is drifting away from the declared onboarding thesis.

This layer becomes especially useful from month 3 onward.

6.5. Confidence and consistency metrics

Layer 3 should also produce two supporting metrics:

- **Inference Confidence Ratio / Index:** how reliable the inferred thesis appears to be.
- **Declared vs. Inferred Consistency Metric:** how consistent the current inferred thesis is relative to the thesis declared at the beginning of the collaboration.

6.6. Formal representation

$$L3 = DealflowFit(\text{Real Thesis})$$

7. Fixed Rules

The model should incorporate hard business rules that override ambiguity in the continuous score:

- If Ticket $\geq 5\text{M}$ and Experience ≥ 1 deal closed, then minimum classification is **Whale**.
- If Ticket $\leq 250\text{k}$ and Experience = 0, then classification is **Very Low**.
- If Ticket $\leq 250\text{k}$ and Experience ≥ 1 , then classification is **Low**.

8. ICS Orchestration Layer

The orchestration layer combines the outputs of the three layers and applies business logic depending on information completeness and investor maturity.

$$ICS = f(L1, L2, L3, \text{Fixed Rules}, \text{Eligibility Rules})$$

A practical operating logic is:

- If only structural information exists, the model relies on Layer 1.
- If subscription and behavioural evidence exist, Layer 2 should materially influence classification.
- If at least 3 months of behavioural history exist and preferences can be inferred with sufficient confidence, Layer 3 should also affect classification.

9. Target Level Mapping

9.1. Proposed initial score bands

The initial mapping may be implemented as:

ICS score band	Target Level
0.00 – 0.24	Very Low
0.25 – 0.44	Low
0.45 – 0.64	Medium
0.65 – 0.84	High
0.85 – 1.00	Whale / Top Tier

These score bands are placeholders and should be calibrated against AM judgment and empirical distribution.

A. Appendix A: Canonical Entities

Entity	Description
Investor	Core investor profile
Subscription	Commercial relationship status and start date
Investor Thesis	Declared investment preferences
Saved Search	Search configurations stored by the investor
Company	Company object and metadata
Deal / Deal Stage Event	Investor-company process evidence
Shortlist / AM Interaction	AM-led outreach and shortlist evidence
Session / Activity Log	Product usage and behavioural signals

B. Appendix B: Layer Summary Tables

B.1. Layer 1 summary

Variable	Weight	Comment
TicketScore	0.30	Structural declared ticket signal
InvestorTypeScore	0.25	Initial sophistication proxy
MAExperienceScore	0.25	Closed-deal evidence declared in onboarding
ThesisCompletionScore	0.20	Explicit thesis completion and basic clarity

B.2. Layer 2 summary

Component	Weight logic	Comment
EngagementScore	Dynamic	Platform usage quality and recurrence
AMScore	Dynamic	Responsiveness to AM / CS interactions
EarlyFunnelScore	Dynamic	Early commercial action when advanced evidence does not exist
AdvancedFunnelScore	Dynamic	Strongest behavioural proof of buying intent

B.3. Layer 3 summary

Component	Role	Comment
Real Thesis	Supporting structure	Inferred investor preference vector
Inferred Ticket	Part of Real Thesis	Not scored independently in L3
DealflowFit	Main scoring variable	Final Layer 3 score
Confidence Ratio	Supporting metric	Reliability of inference
Declared vs Inferred Consistency	Supporting metric	Behavioural drift vs onboarding thesis

C. Appendix C: Dealflow Reciprocity

Author: Martine Dirkzwager

Dealflow Reciprocity is a supply-demand fit metric designed to anticipate churn risk caused by lack of matching opportunities.

C.1. Concept

Before an investor subscribes, or during the first month, Deale estimates how many companies from its catalog match that investor's thesis. The goal is to quantify the real offer that the platform can provide.

C.2. Portfolio fit logic

Step 1 consists of calculating, over the total company universe, those companies that match the investor's criteria.

Step 2 consists of weighting those companies by quality:

$$DR = (n_{premium} \times 1.0) + (n_{platform_complete} \times 0.6) + (n_{registration} \times 0.2)$$

Term	Meaning
DR	Dealflow Reciprocity
$n_{premium}$	Number of premium / subscribed companies
$n_{platform_complete}$	Number of companies with full company profile and additional documentation
$n_{registration}$	Number of companies with only basic registration

C.3. Interpretation

DR level	Interpretation
High	Strong supply fit, lower probability of churn due to no-match-found
Medium	Partial fit, may require AM intervention or sourcing support
Low	Weak fit, higher probability of frustration and churn

D. Appendix D: Fixed Rules Summary

Rule	Output
Ticket $\geq$ 5M and Experience $\geq$ 1 closed deal	Whale floor
Ticket $\leq$ 250k and Experience = 0	Very Low
Ticket $\leq$ 250k and Experience $\geq$ 1	Low