INSURANCE DATA SCIENCE CONFERENCE 2026 · HANNOVER

Early Warning Systems

Optimizing default risk with AI — turning probabilities into decisions and opportunities

Raúl Cancino R.
Actuary at AXA
SCOR International Actuarial Thesis Prize for Iberia

CONTENTS

Index

CHAPTER I

The Decision Problem

Why a probability is not yet a decision

A probability is what models produce.

A probability is what models produce.

A decision is what institutions need.

SCALING FINANCIAL RISK DECISIONS

Consumer credit as empirical application

Scaling exposure means scaling uncertainty.

OPPORTUNITIES

- Millions of decisions at scale
- New efficiency in underwriting
- Data-driven growth potential

CHALLENGES

- Increasing uncertainty per decision
- Higher pressure for actionable systems
- Risk compounds with volume

Millions of decisions · Increasing uncertainty · Higher pressure for actionable risk systems

THE CORE TENSION

The industry does not struggle with prediction. It struggles with actionability.

OPERATIONAL FRICTION

8/10 EWS alerts in production are false positives.

PwC & GALYTIX, 2024

CALIBRATION GAP

Models rank clients correctly but mis-estimate absolute adverse outcome by 2–5 pts.

REGULATORY OPACITY

Solvency II / IFRS 17 require defensibility, not just performance.

A probability is not an explanation.

THE COST OF BEING WRONG

The pain is not only what you missed

Both errors carry a cost. A false negative is not a modeling error — it is unpriced risk sitting in your portfolio.

FALSE POSITIVE

Client flagged high-risk → actually repays

- Loss of business
- Over-provisioning
- Client relationship friction
- Capital over-allocation
- Operational overhead

FALSE NEGATIVE

Client classified low-risk → adverse outcome

- Solvency problem
- Unexpected loss
- Inadequate pricing
- IFRS 17 under-reserving impact
- Potential SCR breach

How do you design a system that transforms data into risk decisions that are accurate, calibrated, and regulatory-defensible?

Not a model. ***A system.***

The answer is not a better model.

The answer is a better architecture.

CHAPTER II

The Architecture of a Decision

From raw data to an actionable risk decision

DATA › PREPROCESSING › CLASSIFICATION › EXPLAINABILITY › SCORING › REGRESSION › EWS

STAGE 1

The Problem at Scale

THE CONTEXT

- Consumer credit exposure
- Personal loans
- Cash advances
- Early deterioration signals
- Regulatory aligned

THE PORTFOLIO

- 307,511 credit operations
- 122 raw variables
- Financial + behavioral signals
- 8.2% default rate

COMPLEXITY

- Class imbalance → accuracy trap
- Regulatory pressure
- Missing values
- High-cardinality categoricals
- Non-linear risk interactions

The empirical application comes from credit risk, but the architectural contribution is insurance-native.

STAGE 2

Preparing Data for Inference

INPUT

Raw dataset

307,511 × 122

Anomaly treatment › **Missing-value imputation** › **Feature engineering** › **One-hot encoding**

SMOTE — Synthetic Minority Over-sampling, applied ***after the split*** to avoid leakage.

OUTPUT

Preprocessed dataset — *ready for modeling.*

STAGE 3

What Drives the Default?

MODELS EVALUATED

Logistic Regression
Naive Bayes
Decision Tree
Random Forest
Artificial NN
XGBoost
LightGBM
CatBoost
AdaBoost

SELECTION CRITERIA

Not just F1.

✓ **Predictive performance**
F1, AUC, Recall, Precision, Specificity

✓ **Stability**
Cross-validation consistency

✓ **Regulatory fitness**
Auditability

SELECTED

CatBoost

F1 **99.9%** CV **99.4%**

Native categorical encoding eliminates a class of leakage risk affecting every other model here.

…but we keep an eye on XGBoost.

The model is not the probability. It is the inferential engine.

STAGE 4

What Did the Model Actually Learn?

Mean Decrease Gini Importance

MDGI

SHAP Values

Shapley attribution

We keep the top 11 variables confirmed as genuine predictors.

TOP VARIABLES — CatBoost

Income & Credit

AMT_INCOME_TOTAL · AMT_GOODS_PRICE

Behavioral Signals

DAYS_EMPLOYED · DAYS_LAST_PHONE_CHANGE

External Funding Sources

EXT_SOURCE_1 / 2 / 3

Socioeconomic Context

REGION_RATING_CLIENT · EDUCATION / OCCUPATION
REGION_RATING_W_CTY · GUARANTOR_EMPLOY

STAGE 5 · WHERE AI MEETS ACTUARIAL JUDGMENT

Expert-Weighted Scoring

STEP 1 — TRAMIFIED: 11 variables transformed into risk bands — Deciles, Quartiles, Terciles.

Salary → (Salary10, Salary20, Salary30, … , Salary100)

Result: **84 binary indicators**

"Continuous variables become interpretable risk signals"

▼

STEP 2 — DIRECTIONAL WEIGHTING: Economic logic assigned to each signal.

↑ Income → **↓ Risk**

↓ Employer-guarantor → **↑ Risk**

"Every weight has an economic interpretation"

▼

STEP 3 — EWS COMPOSITE SCORE: Σ (binary indicator × weight).

Final score range: **2 → 11**

Higher score = **higher default risk**

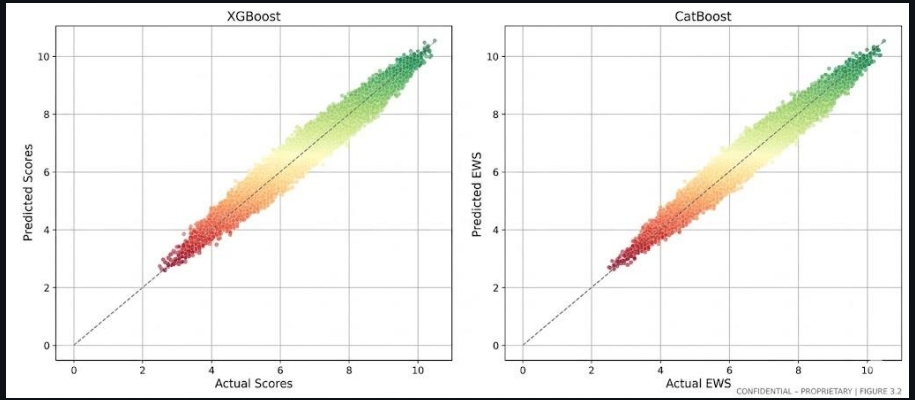
"Comparable, interpretable and auditable"

STAGE 6

Calibrating the Decision System

Objective: predict the EWS composite score per client using the 84 binary indicators as input features — a scoring (regression) problem, no longer classification.

	CatBoost	XGBoost
NSE	19.69%	19.05%
RMSE	23.02%	24.86%
Variance	98.25%	93.97%



The question is which risk tier this client belongs to. Insurance operations rarely act on raw probabilities — they act on calibrated risk tiers.

STAGE 7

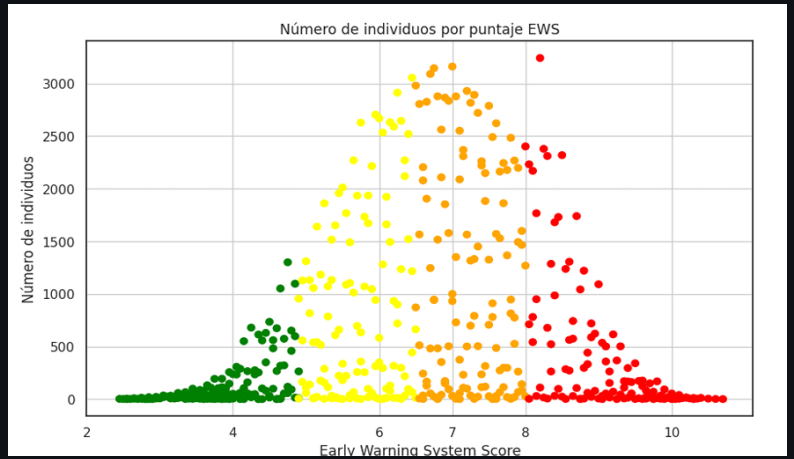
Early Warning System — from score to decision

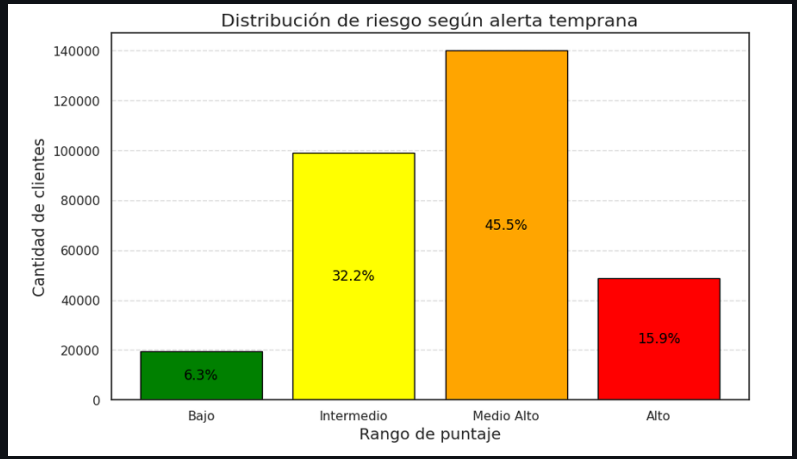
LOW RISK
Score < 4.9
Portfolio share 31.4%

MODERATE
4.9 – 6.4
Portfolio share 23.2%

CONSIDERABLE
6.4 – 8.1
Portfolio share 30.0%

HIGH RISK
> 8.1
Portfolio share 15.4%





CHAPTER III

What the System Actually Delivers

Calibrated, auditable, and deployable risk decisions

DECISION IMPACT

Where the Cost Actually Lives

FLAGGED HIGH-RISK → CLIENT REPAYS · TYPE I

Segment	% Type I Error
High segment	**78.80%**
Considerable segment	**74.70%**

CLASSIFIED SAFE → ADVERSE OUTCOME · TYPE II

Segment	% Type II Error
Low segment	**2.30%**
Moderate segment	**4.20%**

A Type I error leads to lost business.

Industry context: 8/10 EWS alerts are false positives (PwC & GALYTIX, 2024).

A Type II error implies additional capital consumption.

≈ 4,200 operations in the Moderate segment alone.

Unpriced. Unprovisioned. Unexpectedly risky.

This is where the unseen risk actually lives.

FROM PROBABILITY TO DECISION

The Architecture of Action

ML OUTPUT	PD SCORE	ML EWS	ACTION	GOVERNANCE
PD = 0.73	EWS 9.2	HIGH · DR 14.8%	Provisioning ↑ · Pricing review	Auditable · Traceable · Human-validated

FROM PROBABILITY TO DECISION

The Architecture of Action

ML OUTPUT	PD SCORE	ML EWS	ACTION	GOVERNANCE
PD = 0.73	EWS 9.2	HIGH · DR 14.8%	Provisioning ↑ · Pricing review	Auditable · Traceable · Human-validated

SEGMENT	OBJECTIVE	ACTIONS
● LOW	Maximize growth efficiency	Streamlined approval · Baseline pricing · Reduced manual review · Standard loss monitoring
● MODERATE	Controlled expansion	Premium sensitivity adjustment · Exposure resizing · Conditional underwriting · Enhanced monitoring
● CONSIDERABLE	Risk containment	Mandatory underwriter review · Dynamic reserve review · Coverage reassessment · Behavioral escalation
● HIGH	Capital protection	Underwriting escalation · Solvency II / IFRS 17 reserving trigger · SCR alert · Executive escalation

FROM PROBABILITY TO DECISION

The Architecture of Action

Human-in-the-loop:
humans supervise the architecture, not just the final decision.
The AI infers — humans decide.

S		
● CONSIDERABLE	Risk containment	Mandatory underwriter review · Dynamic reserve review · Coverage reassessment · Behavioral escalation
● HIGH	Capital protection	Underwriting escalation · Solvency II / IFRS 17 reserving trigger · SCR alert · Executive escalation

SYNTHESIS

The Contribution

Not a new prediction model — an architecture that turns probabilistic outputs into calibrated, auditable, actionable risk decisions.

Interpretable by design

Machine learning combined with actuarial expertise.

Deployability

No core-system replacement required.

Governance

Auditability, recalibration, and human oversight.

Transferability

From credit risk to insurance risk monitoring.

Not a model. ***A system.***

A probability is what models produce.

A decision is what institutions need.

Thank you

Insurance Data Science Conference 2026 · Hannover

Raúl Cancino R. · Actuary at AXA

SOURCES

References

Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer.

Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., & Gulin, A. (2018). CatBoost: unbiased boosting with categorical features. Advances in Neural Information Processing Systems, 31.

Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.

Baesens, B., Van Gestel, T., Stepanova, M., Van den Poel, D., & Vanthienen, J. (2003). Neural network survival analysis for personal loan data. Journal of the Operational Research Society, 54(12), 1289–1298.

PwC & GalytiX (2024). The Future of Early Warning Systems in Financial Services.

Bräuning, M., Malikkidou, D., Scalone, S., & Scricco, G. (2019). A new approach to early warning systems for small European banks. ECB Working Paper Series, European Central Bank.

European Central Bank. (2026). Consumer credit – ECB Data Portal. ECB Data Portal.

.