

From Loan Books to Leading Indicators: Uncovering Market Signals & Predicting Delinquency

nCino Sponsor & Team: Bill Stuart, Nolan Frost, Tyler Held
MIT Faculty Advisor: Allan Telio



Justine Murdock



Judy Zhu

Taking Lending Data from Noise to Signal

Problem:

- Lending patterns could reveal **broad market trends**, but this signal goes untapped today.
- Current early-warning system relies on fixed, **overly sensitive** weights so reviewers get **flooded** with **false alarms**.

Solution:

- Use nCino's lending data to surface **early** signals of market trends.
- Update the early-warning system to catch risky loans **more accurately**, with **less review volume**.

Impact Across All Clients



45%
Reduction in review volume



~2.5 Million
Potential people-hours reallocated to closing new deals quarterly (~1 week/person/quarter)

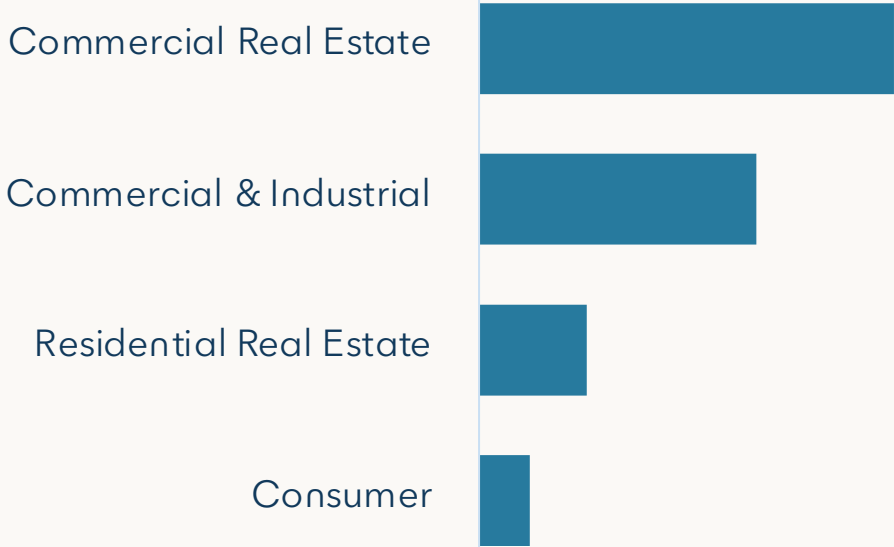


~\$1.15 Billion
Potential revenue increase from new deals closed quarterly

Market Context, Metric Construction, & Signal Analysis

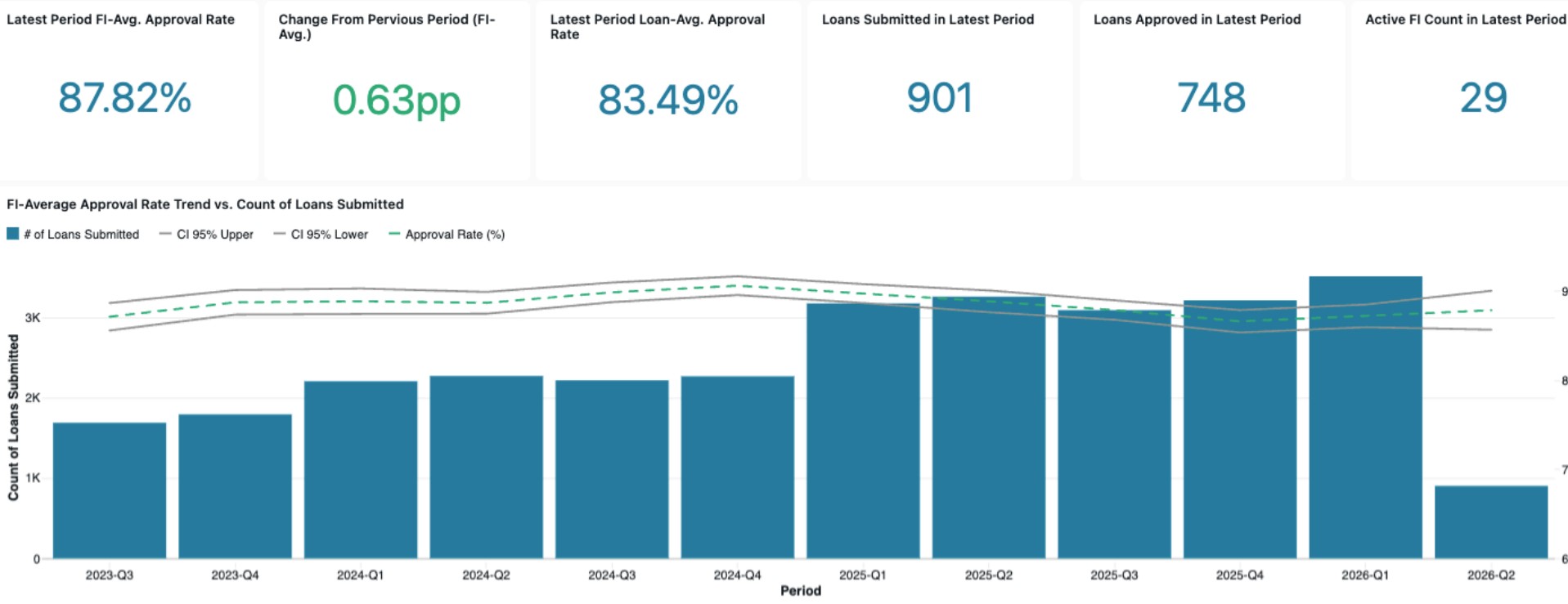
nCino Market Share

Due to privacy, this is relative share of U.S. loan market in each sector



nCino's presence is strongest in Commercial Real Estate and Commercial & Industrial Loans

Approval Rate Live Monitoring Dashboard



Automated pipeline constructed and metric validated as series for lead-lag test for predictiveness

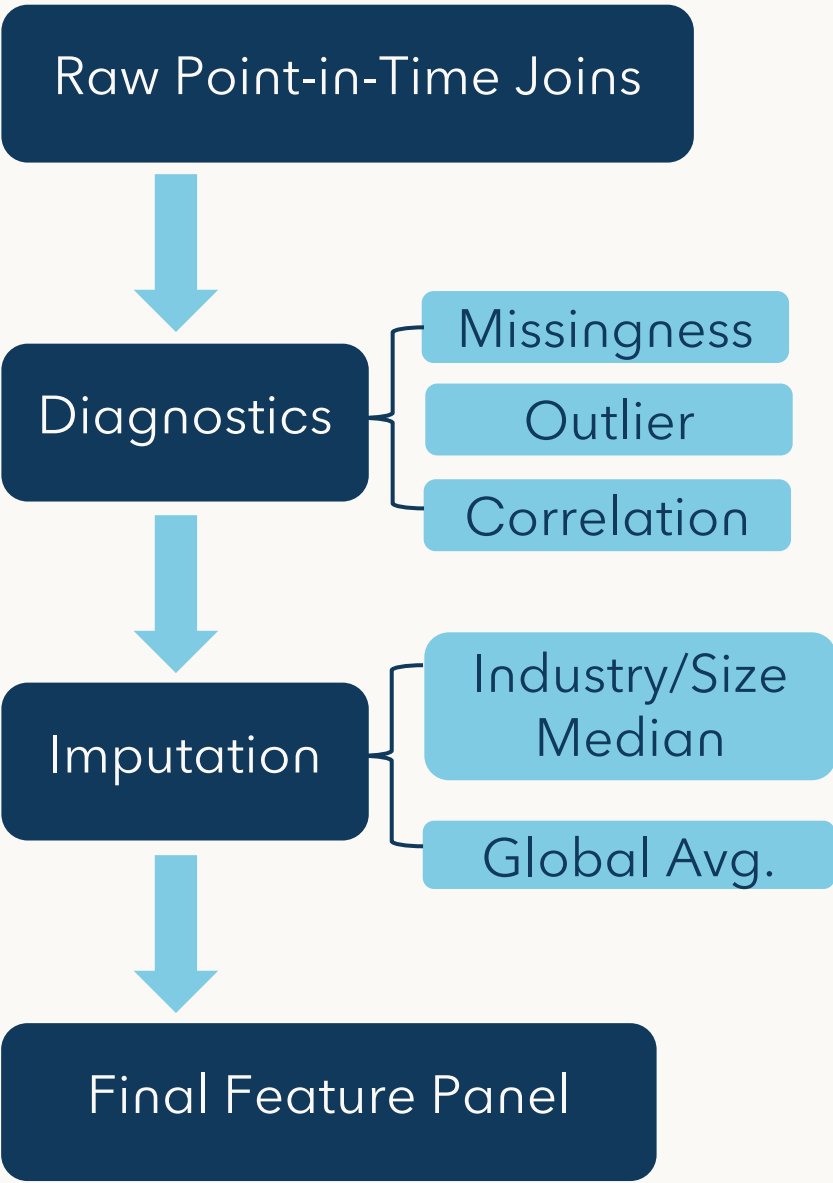
Lead-Lag Statistical Testing Pipeline

- Visual Test**
Screen for missing periods, outliers, and short samples
- Stationarity**
Test and transform non-stationary series using ADF & KPSS
- Pre-whitened Cross-correlation**
Remove autocorrelation and test for lead-lag correlation
- Granger causality**
Test for one-directional predictive power (nCino -> macro)
- Robustness Test**
Stress-test across transformations, rolling windows, etc.

Built and ready, re-runs automatically as more data history accrues

Feature Pipelines & Predictive Modeling

Feature Pipeline



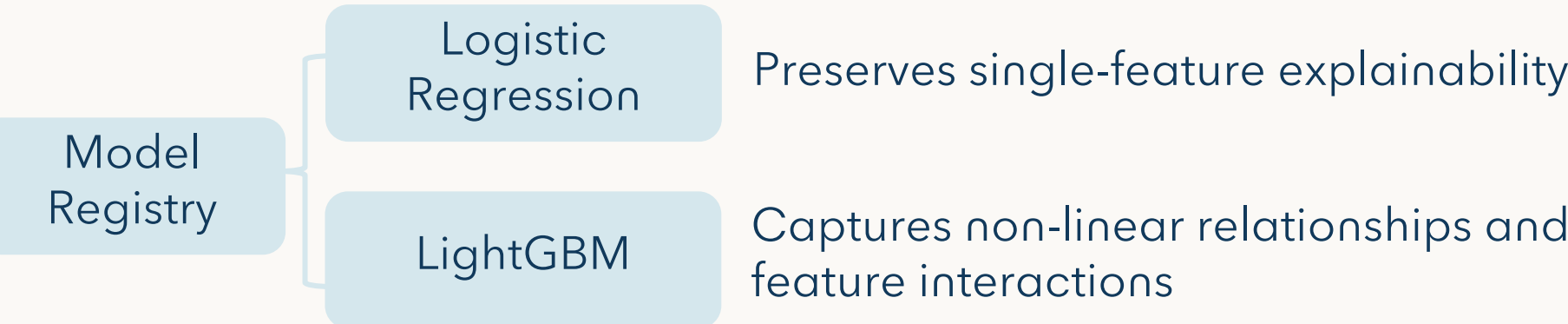
Discrete-Time Hazard Framework

Monthly Hazard Score:

h_t = Probability of 30+ days-past-due on payment next month

Multi-month Compound through Month T :

$$S(T) = \prod_{t=1}^T (1 - h_t)$$



- Trained on the full loan population, adaptable to **every** FI client
- Iterated on different feature sets; **Class-weighted** and tuned

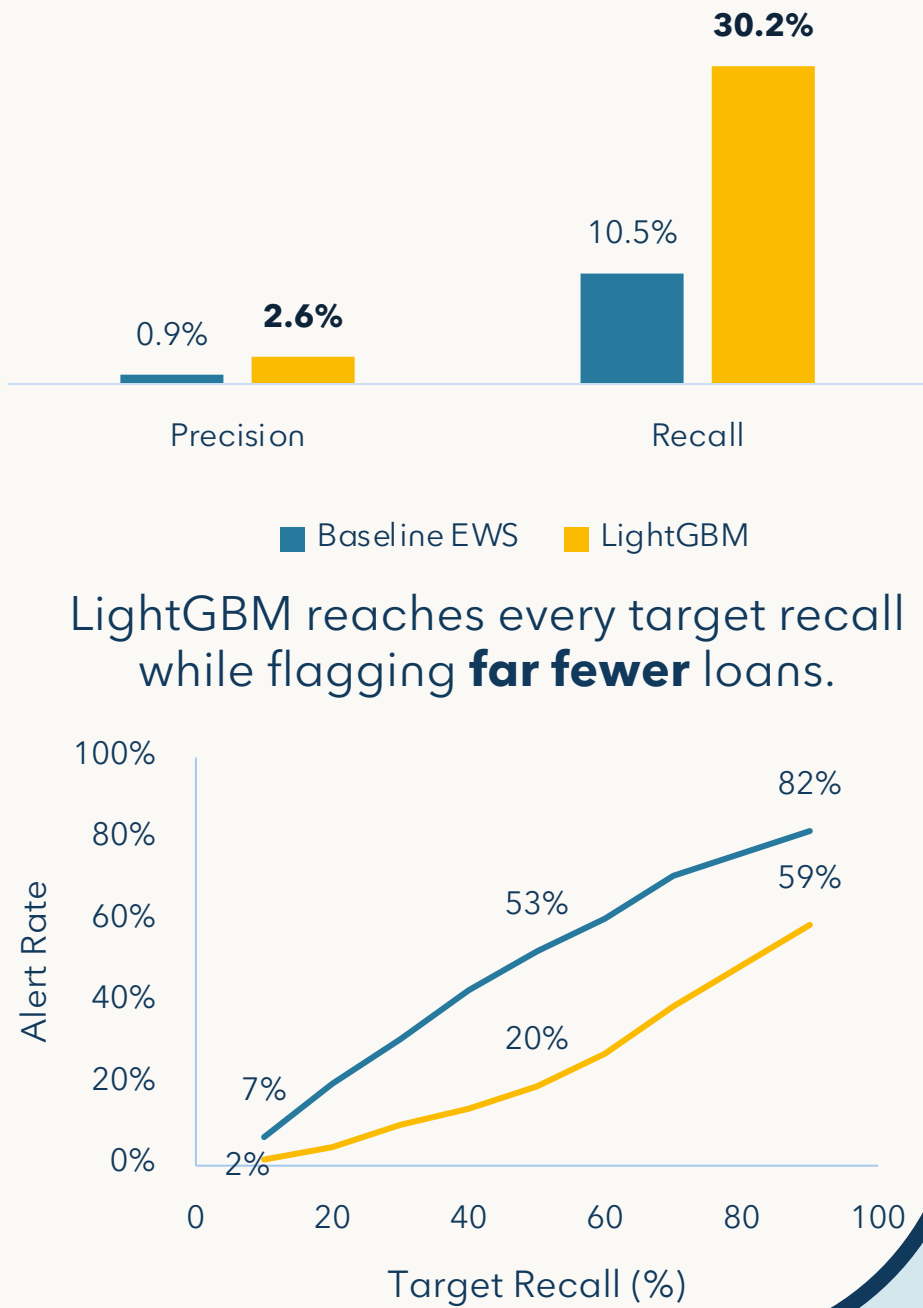
Results

Model Choice:
LightGBM

Evaluated on
Held-Out Set:
200,756
loan-months
Jan to Jun 2026

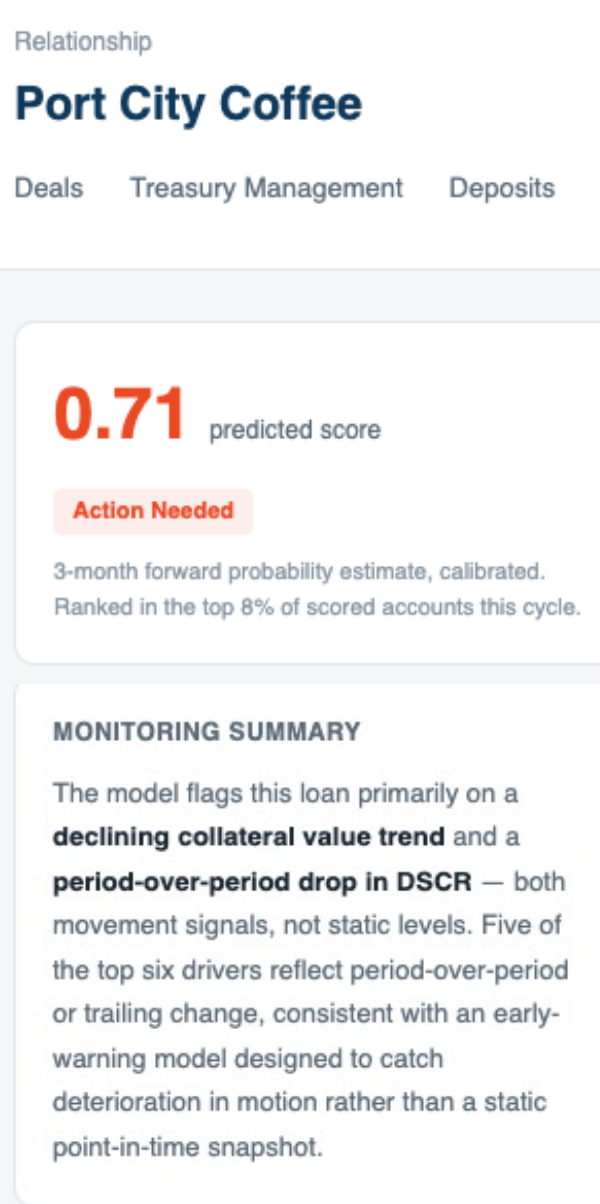
Given everything known up until now, what is the probability a loan goes 30+ days-past-due on payment sometime in the next **3 months?**

LightGBM catches **2.9x** as many delinquencies as the baseline. (Alert rate = 10%)



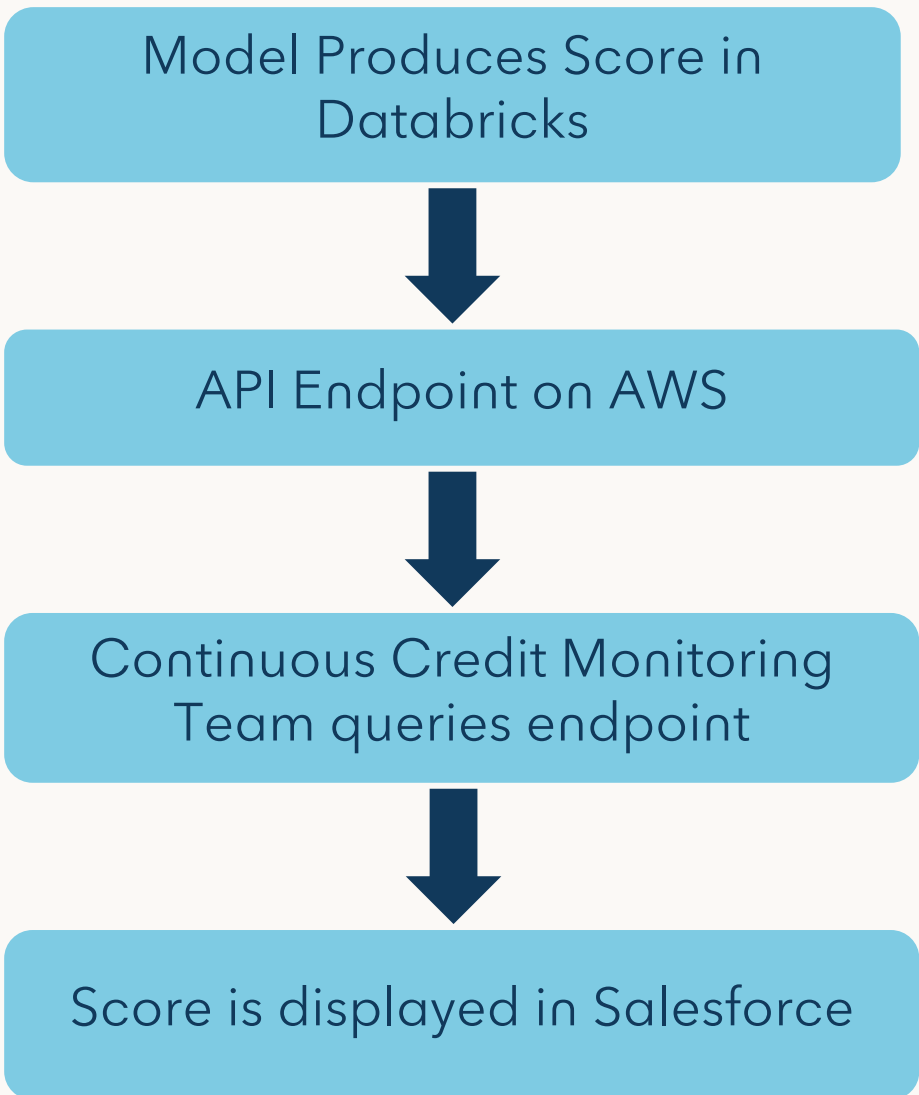
Projected Early-Warning Implementation

Real-Time Risk flag learned from outcomes, replacing the fixed-weight baseline.



Top Risk Drivers	SHAP contribution to predicted score — period-over-period and trend signals
Collateral value, 3-mo change	+0.184
DSCR, period-over-period change	+0.142
Risk rating, period-over-period change	+0.098
Revenue, period-over-period change	+0.069
Current ratio, trend change	-0.081
Deposit balance, 30-day trend	-0.072

Implementation Workflow



Updated live scoring lives inside the operating system that loan officers use daily.

nCino's Data

- 348** Financial Institution **Clients**
- 20 Million** Loans in nCino's database across time
- 30 Months** of clean panel history for modeling approach
- 550 Thousand loan-months** joined to multiple data sources (Accounts, Spreads, Covenants, Collateral)