Machine Learning For Credit Risk: From Model to Business Value

Xiangyu Wang
Boston University – MSMFT Program   xywk2024@bu.edu

Introduction

This project uses real loan application data from Home Credit, a global consumer finance company that provides credit to people with limited borrowing history. The goal is to help lenders assess a client's default risk more reliably and sustainably, ensuring stable performance across their consumer-loan portfolio. The business problem is to increase loan approval rates while keeping portfolio default risk under control.

The final model shows high and consistent accuracy over time, meaning its predictions remain reliable as new data arrives. Most of the main factors influencing default risk are stable, while a few show early signs of drift and are being monitored to prevent future degradation.

Data

Source. *Home Credit – Credit Risk Model Stability* (Kaggle). The dataset contains detailed loan application records where each client is labeled by payment outcome (1 = default, 0 = non-default). Each record represents one borrower's credit application and combines both internal behavioral data and external bureau information.

Feature groups.

- **P (behavioral):** delinquency and repayment patterns (Days-Past-Due history)
- **A (financial):** credit limits, balances, and amount-based metrics
- **D (date-derived):** time intervals such as days since approval or last payment
- **M (masked categorical):** anonymized identifiers and encoded categorical data
- **T/L (transformed or legacy):** engineered ratios and legacy fields

These variables jointly describe a client's financial situation, credit usage, and repayment behavior, forming the input for default-risk modeling.

Preprocessing

Stage 1 – Before Merge.

- Missing-value imputation.
- Row-level combo features such as delinquency ratios, DTI, utilization, repayment progress, overdue mix.
- Aggregation by suffix: `P/A` – max/mean/last; `D` – days from decision; `M/T/L` – max/last.
- Merge; extract decision month / weekday; convert dates to days from decision.

Stage 2 – After Split.

- Imputation on aggregated data; remove high-nan / high-cardinality features.
- Interaction features such as recent delinquency or limit change, no recent installment.
- Winsorize numeric variables (99.8th percentile, train-based).
- Encoding: one-hot (20 levels); temporal target encoding (>20 levels).
- Drop highly correlated features ($|\rho| > 0.97$), keep mean > max > last.
- Feature selection: LightGBM gain + SHAP → union of top drivers.

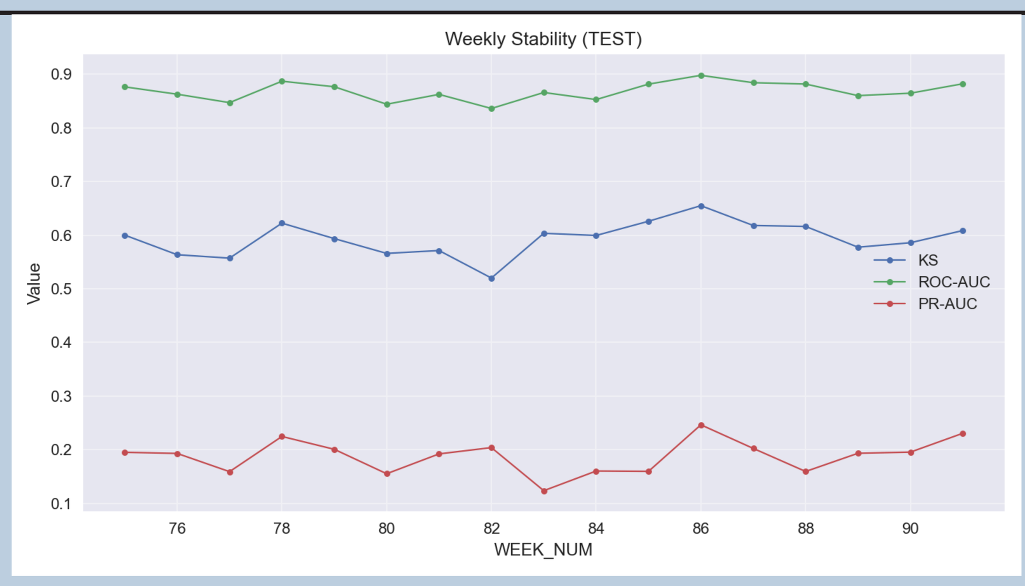
Model

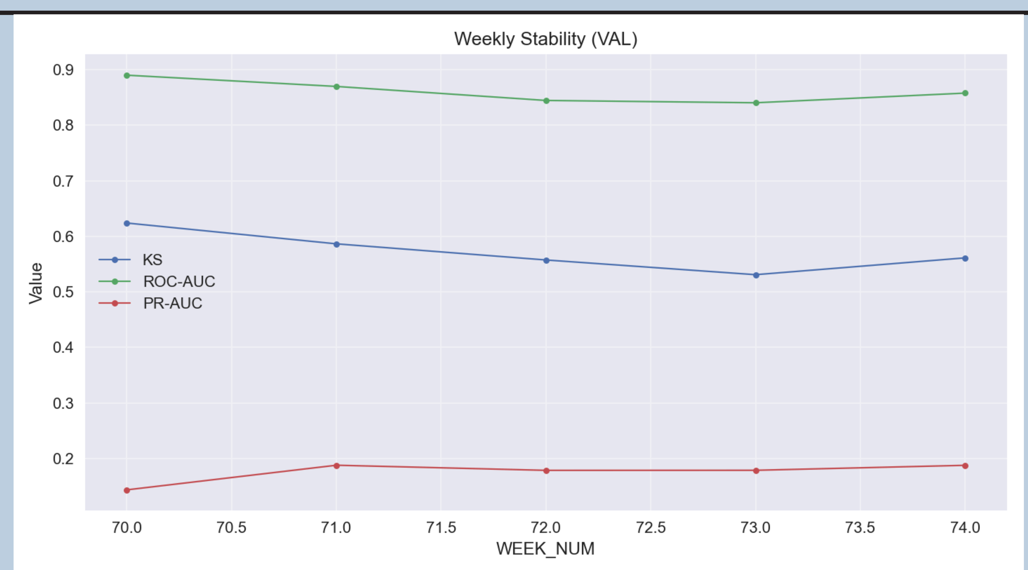
A time-series expanding-window cross-validation was used to mimic real-world deployment, training on past data and validating on future unseen periods. Each validation window covered four weeks and moved forward sequentially, ensuring all prior data were included in the training window. The model was trained with Optuna-tuned hyperparameters, optimizing for PR-AUC to emphasize ranking quality under severe class imbalance, which was addressed through weighted samples. Finally, a LightGBM gradient boosting classifier was trained on the full training period (Weeks 0–69) with the selected features and hyperparameters, and used for subsequent stability and business evaluations.

Results

1. Model Evaluation.

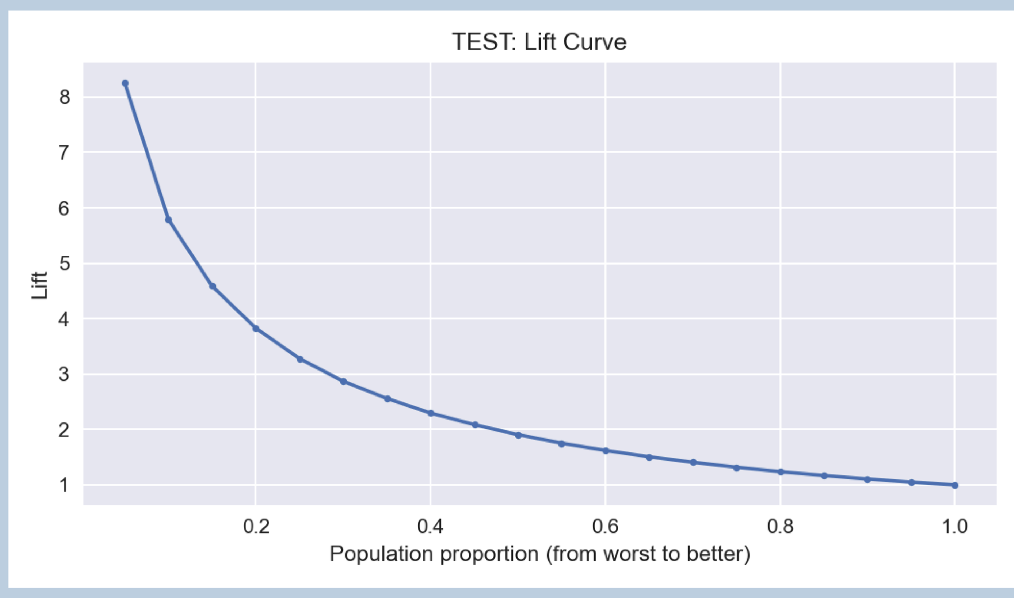
Split	#Cases	Positive Rate	PR-AUC	ROC-AUC	KS	Brier	ECE
TRAIN	1,302,727	0.0331	0.2593	0.9013	0.6495	0.1429	0.2682
VAL	40,554	0.0236	0.1756	0.8569	0.5550	0.1165	0.2380
TEST	183,378	0.02158	0.1845	0.8718	0.5855	0.1488	0.2869





High accuracy and strong ranking ability mean the model can clearly distinguish reliable borrowers from risky ones, supporting safer lending decisions and steadier portfolio growth. Lower Brier/ECE(expected calibration error) indicate well-calibrated probabilities, easing threshold setting across periods.

2. Business Evaluation.



Some business performance indicators on `test set`*:*

bad_cap	threshold	approve rate	bad rate in approved	bad capture rate	weeks over cap	worst week	std
0.016	0.7318	0.9267	0.0116	0.5021	0	0.0158	0.0021
0.018	0.7780	0.9547	0.0138	0.3907	0.0588	0.0181	0.0023
0.020	0.8201	0.9756	0.0162	0.2661	0.1765	0.0214	0.0027

PSI computed on approved population: Test vs Validation.

bad_cap	threshold	overall_PSI	median_week_PSI	max_week_PSI
0.016	0.7318	0.0277	0.0480	0.0873
0.018	0.7780	0.0310	0.0518	0.0895
0.020	0.8201	0.0346	0.0544	0.0928

Stable model performance across time means the lender can apply consistent approval thresholds without sudden increases in bad loans, ensuring predictable portfolio quality and smoother risk management.

3. Feature Analysis.

- *Top drivers (SHAP):* `Average Origination Date of Closed Loans`, `Average Payment Delay (Last 24 Months)` are predictive and stable.
- *Unstable signals:* `Average Date of Last Credit Bureau Update`, `Maximum Tax Deduction` show high PSI (>1) → monitor, not drop.
- *Interpretability:* SHAP trends align with credit logic (older refresh dates, higher DTI ↑ risk).

Feature stability metrics across splits (PSI, KS, and IV) for top 5 drivers.

feature	psi(train-val)	psi(train-test)	ks(train-test)	iv(test-train)
Average Origination Date of Closed Loans	0.0332	0.0448	0.0730	0.0705
Gender of the Client	5.0558	0.0031	0.0269	0.0233
Average Date of Last Credit Bureau Update	11.1747	11.1946	0.9589	-0.0128
Average Payment Delay (Last 24 Months)	0.0018	0.0020	0.0274	0.0954
Number of People Sharing the Same Mobile Number	0.0234	0.0229	0.0675	0.0076