Miles McCunniff, Sebastian Perez Parra, Sina Odejinmi

Predicting Loan Default Risk for Home Credit

MSBA Practice Capstone | Fall 2025

AGENDA

Problem Overview

Home Credit wants to extend loans to underbanked customers

Challenge: Many applicants **lack traditional credit histories** → high uncertainty → financial risk

Goal: Build a model that predicts default risk to guide smarter approvals

HOME CREDIT

Modeling Approach

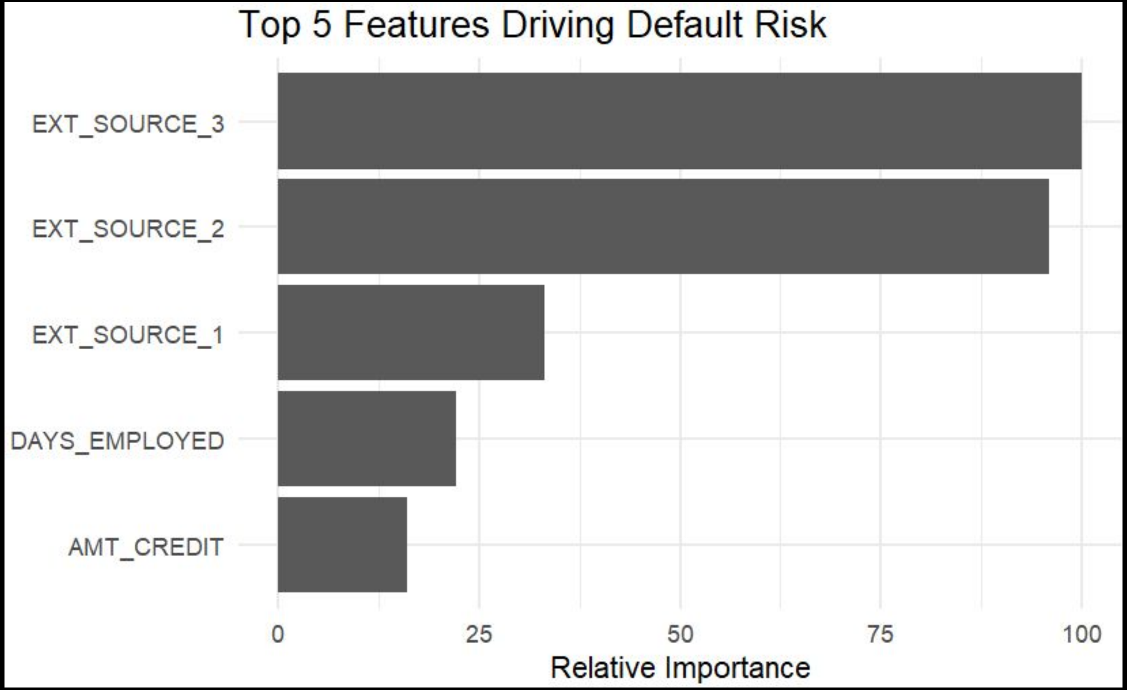
Model	Description	AUC
Logistic Regression (Baseline)	Benchmark	0.709
Regularized Logistic	LASSO/Ridge tuned	0.721
Random Forest	Nonlinear ensemble	**0.73399**
Neural Network	Basic feedforward	0.730
Logistic + CV	Cross-validated tuned version	0.722

Best model: Random Forest – strong balance of accuracy and interpretability.

Key Drivers of Default

Top 5 Features:

1. EXT_SOURCE_3 – external credit score proxy
2. EXT_SOURCE_2 – external credit score proxy
3. EXT_SOURCE_1 – external credit score proxy
4. Days Employed
5. Amount of Credit



John vs. Jane

John (Rejected)

- **Income:** $61,650
- **Credit Amount:** $997,335 **→ Credit/Income Ratio = 16.18×**
- **Income Type:** Pensioner
- **Education:** Secondary / secondary special
- **External Credit Score (EXT_SOURCE_3):** Missing (no credit history)
- **Children:** 0
- **Age:** ≈ 59 years
- <u>**Model Target:** 1 → **Defaulted**</u>

Jane (Approved)

- **Income:** $90,000
- **Credit Amount:** $251,091 **→ Credit/Income Ratio = 2.79×**
- **Income Type:** Pensioner
- **Education:** Secondary / secondary special
- **External Credit Score (EXT_SOURCE_3):** 0.894 (very high)
- **Children:** 0
- **Age: ≈** 56 years
- <u>**Model Target: 0 → Did not default**</u>

Projected Results

We want to Reject the riskiest 5% of applicants

Using a cutoff of 0.49

- Removes ~20% of defaulters
- Reduces default rate from 8% to ~6.9%
- Avoids ~$16M in annual losses per 100k loans
- Approves 95% of applicants

Confusion matrix from our validation predictions

	Actual Default (1)	Actual Non-Default (0)
Predicted Default (1)	1022 (TP)	2054 (FP)
Predicted No Default (0)	4044 (FN)	54382 (TN)

Implementation & Monitoring

Implementation

- Integrate model in application review workflow
- Use probability threshold to approve/deny - riskiest 5% rejected (0.49 cutoff)

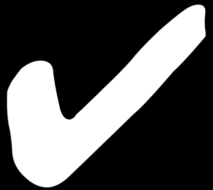
Monitoring

- Track AUC, drift, false positives quarterly
- Re-train every 6 months with new applicants

Evaluate performance by gender, age, and income groups

Next Steps

1

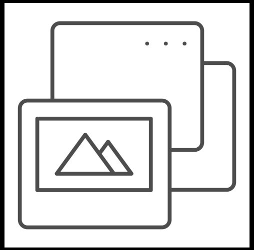


Conduct pilot with 10K new applicants

2

Test ensemble stacking for future improvement

3



Build dashboard to monitor default trends

4

Collaborate with risk team for operational rollout

“Our model helps Home Credit extend loans confidently, balancing inclusion with financial sustainability”

THANK YOU

Any questions?

Appendix

Metric	**Reject Top 5% (Cutoff = 0.494)**	**Reject Top 10% (Cutoff = 0.432)**
Applicants Rejected	5%	10%
Defaulters Removed (TPR)	**~20%**	**~33%**
Default Rate After Model	**~6.9%** (down from 8%)	**~6.2%** (down from 8%)
Approved Applicants	95%	90%
Annual Loss Avoided *(per 100k loans @ $10k)*	**~$16M**	**~$26M**
False Positives (Good borrowers rejected)	~3.6%	~8%

10% Cutoff confusion matrix

	Actual Default (1)	**Actual Non-Default (0)**
Predicted Default (1)	**1654 (TP)**	**4497 (FP)**
Predicted No Default (0)	**3412 (FN)**	**51939 (TN)**

Appendix Cont.

5% cutoff math explained

Total defaulters: TP+FN=1022 + 4044 = 5066

Captured defaulters = 1022

Capture rate: 1022 / 5066 = 0.2018 ≈ 20.2%

Rejecting the riskiest 5% removes ~20% of defaulters

False positives (FP): 2054

Total non-defaulters: 2054 + 54382 = 56436

FP rate: 2054 / 56436 = 0.036 = 3.6%

Only ~3.6% of safe applicants get rejected

Approvals = Pred 0 group = 4044 + 54382 =58426

Defaults remaining = FN = 4044

New effective default rate: 4044 / 58426 ≈ 0.0692 = 6.9%

Default rate goes from 8% to ~6.9%

	Actual Default (1)	Actual Non-Default (0)
Predicted Default (1)	1022 (TP)	2054 (FP)
Predicted No Default (0)	4044 (FN)	54382 (TN)

Baseline loss (8% default):

100,000 loans ×8% × $10,000 = $80M

Reduction: ≈ **20% of defaults removed**

Loss avoided:

80M × 0.20 = $16M

Savings ≈ $16M per 100,000 loans