Rakamin Academy

HOME CREDIT
Anda Bisa!

Approve This Loan?

Reducing Defaults & Boost Revenue with Predictive Models

Presented by
Robert Gunanta Bukit

Rakamin Academy

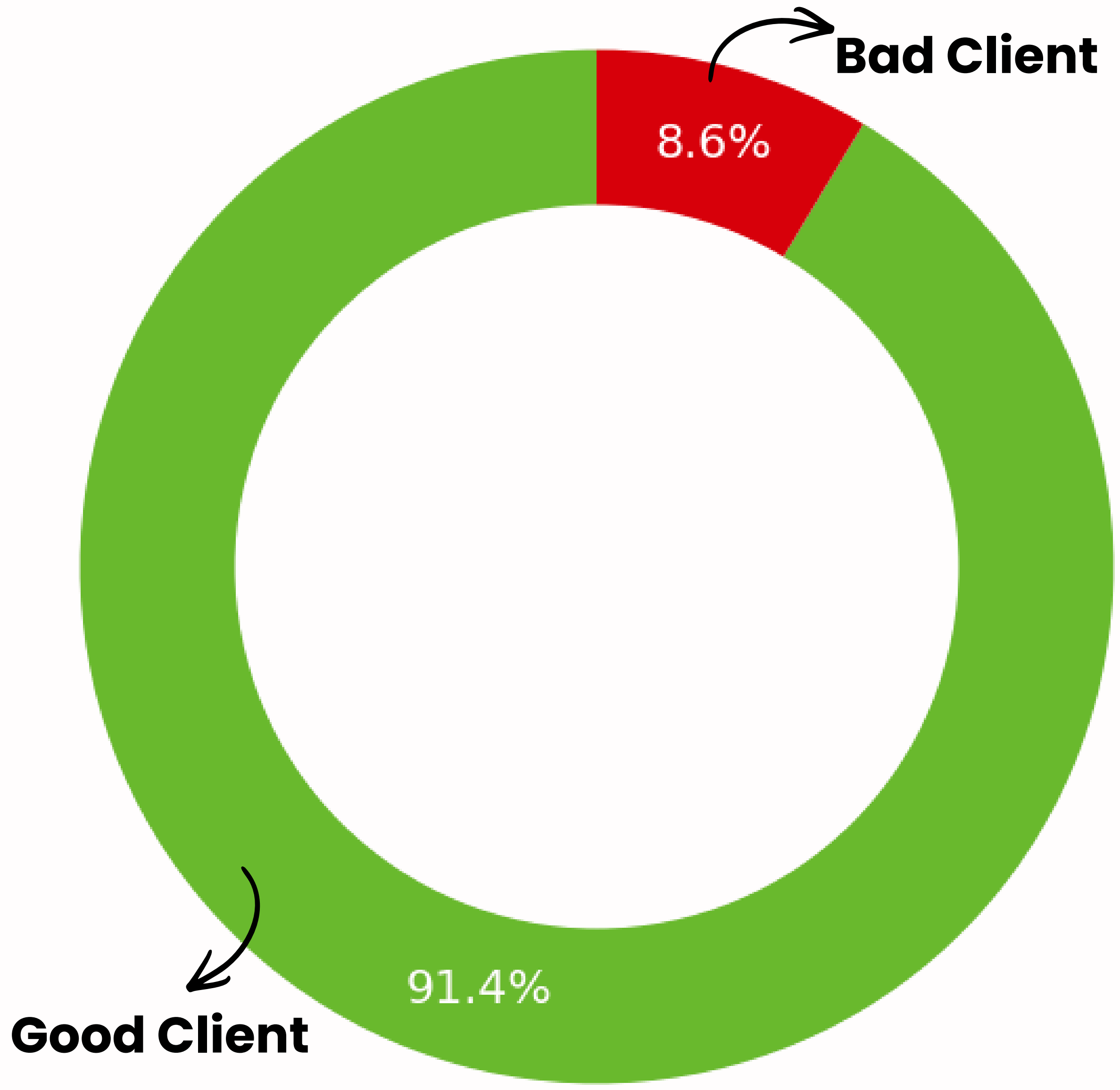
HOME CREDIT Anda Bisa!

TABLE OF CONTENT

PROBLEM RESEARCH

Rakamin Academy | HOME CREDIT Anda Bisa!

Proportion of Good and Bad Clients



Total Clients: 1430155

"High loan default rate (8.6%) significantly reduces net revenue by over $61.67B annually."

1. Revenue
$179.6 M

Total good client		Revenue/Avg Loan
1306815	X	$500.000

2. Total Bad Client Debt
$61.67 M

Total good client		Revenue/Avg Loan
123340	X	$500.000

3. Net Revenue
$591.74 M

PROBLEM STATEMENT

The current loan approval process lacks effective risk assessment, resulting in a high number of bad clients and significant revenue loss.

Goal

Reduce default rate through datadriven client assessment using machine learning.

Objective

Build and deploy a predictive model to identify high-risk clients before loan approval.

Business Metrics

- Default Rate Reduction (%)
- Increase in Net Revenue ($)
- Model Accuracy / Recall on Bad Clients (%)
- Cost Saved from Bad Loans ($)

DATA PREPARATION

Rakamin Academy

HOME CREDIT
Anda Bisa!

Credit_card_Balance.csv

- No duplicate
- Handling missing value using median
- Agg 7 features (sum) by SK_ID_PREV

Installments_payment.csv

- No duplicate
- Handling missing value using drop and fill 0
- Agg instalment & payment and create new feature

POSH_CASH_Balance.csv

- No duplicate
- Handling missing value using ratio and median
- Agg SK_DPD by SK_ID_PREV (mean)

Previous_Application.csv

bureau_balance.csv

- No duplicate
- No missing value
- Create new features

bureau.csv

- No duplicate
- Handling missing value; > 50% drop, 10-50% impute median, <10% impute mode.

application_train.csv

Main Dataset

1.430.155 x 112

DATA PRE-PROCESSING

Rakamin Academy

HOME CREDIT Anda Bisa!

1

Feature Engineering

- **Age grouping:** Binned DAYS_BIRTH into age group
- **Handled Invalid Values:** Replaced 'NAME_FAMILY_STATUS' = 'Unknown' to NaN. And replaced 'XNA','XAP with NaN.
- **Value Correction:** Fixed CNT_FAM_MEMBERS anomalies (0,5 → 1, & 4,5 → 5)

2

Data Splitting

Train 75 & Test 25, stratify y

3

Feature Selection

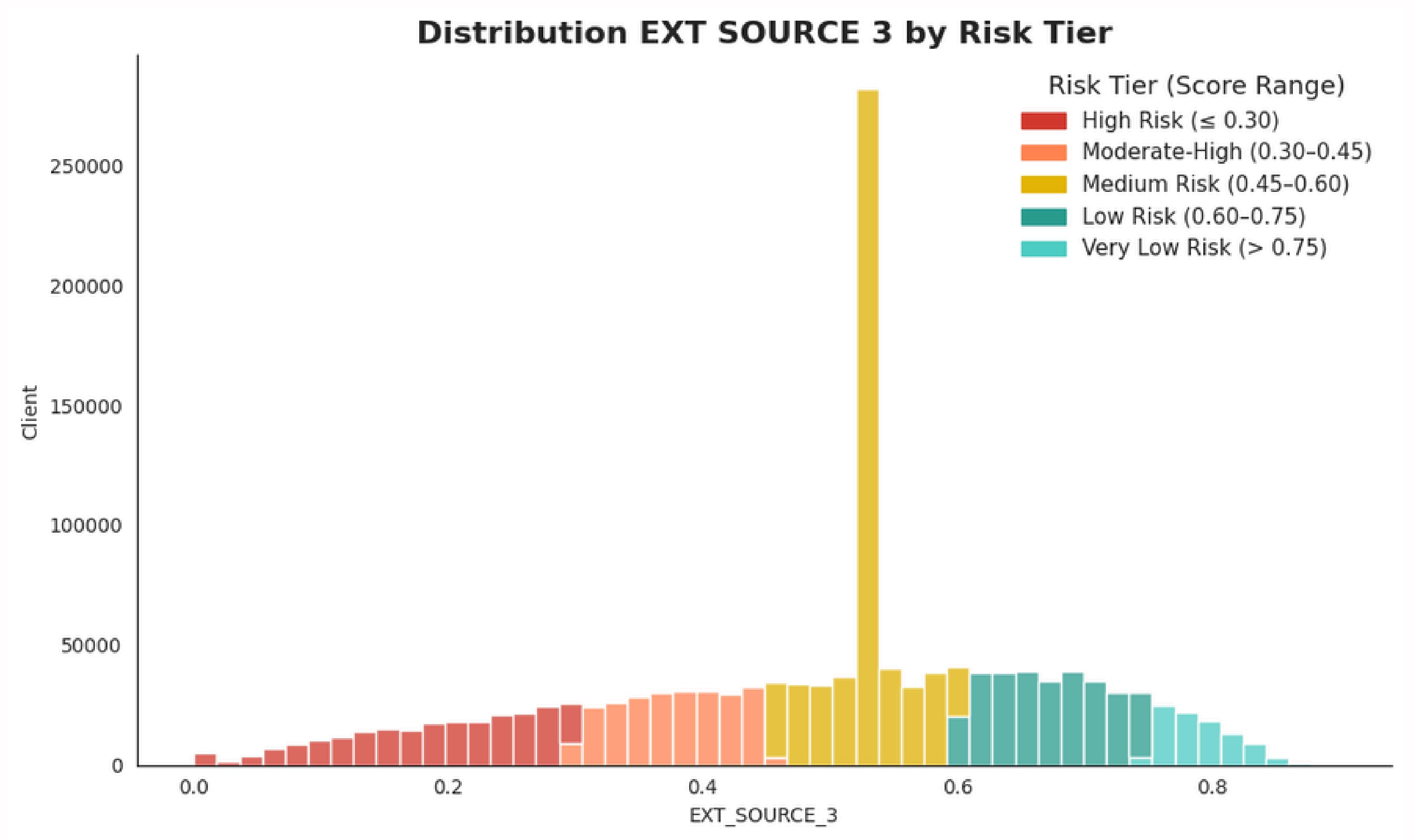
- Removed high-missing-value features
- Selected variable with strong and useful
- Dropped multicollinear features

4

Handling Outlier

IQR Winsorization: Capped Outliers at IQR boundaries

BUSINESS INSIGHT



Insight:

- Clients with higher EXT_SOURCE_3 scores (above 0.45) tend to have lower default rates.
- The majority of clients fall into the medium to low risk category (score 0.45 – 0.75).
- This indicates a strong positive correlation between EXT_SOURCE_3 and client creditworthiness.

Recommendation:

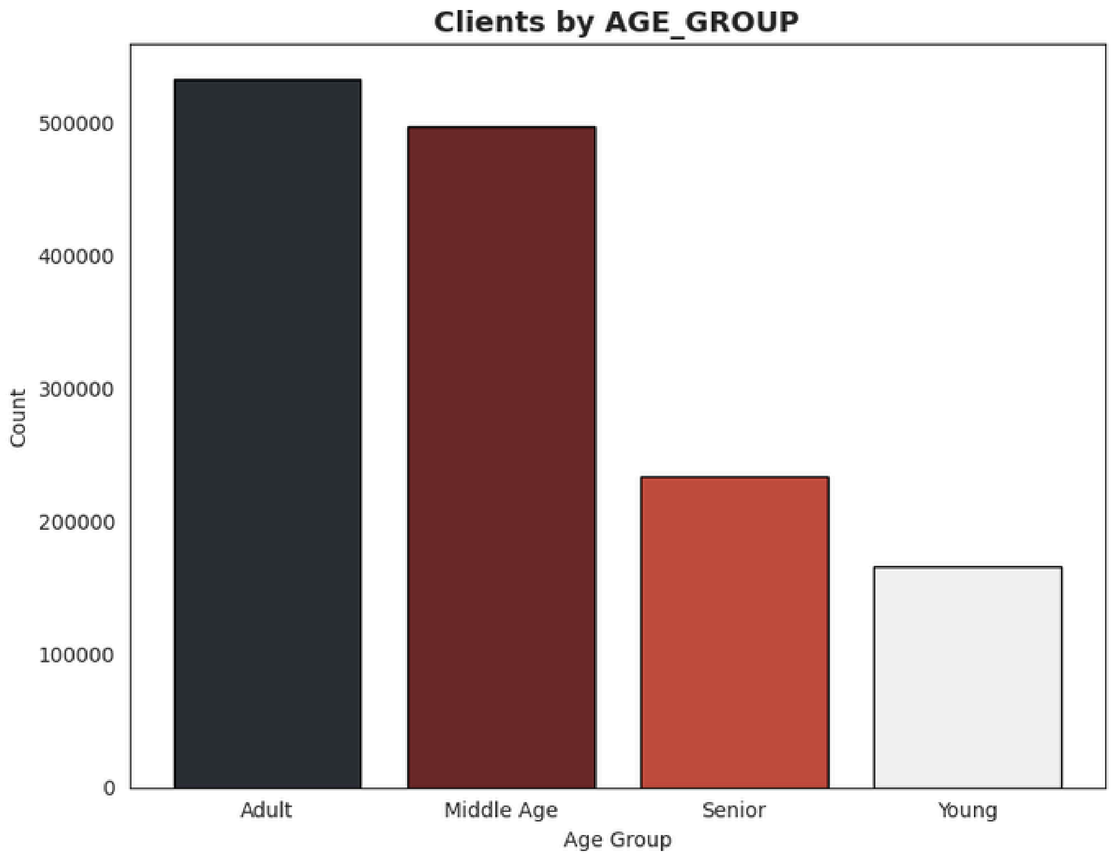
- Target segments with EXT_SOURCE_3 ≥ 0.45.
- Use AI-powered digital ads & lookalike audiences.
- Partner with relevant digital platforms (Socmed, E-commerce, etc.)

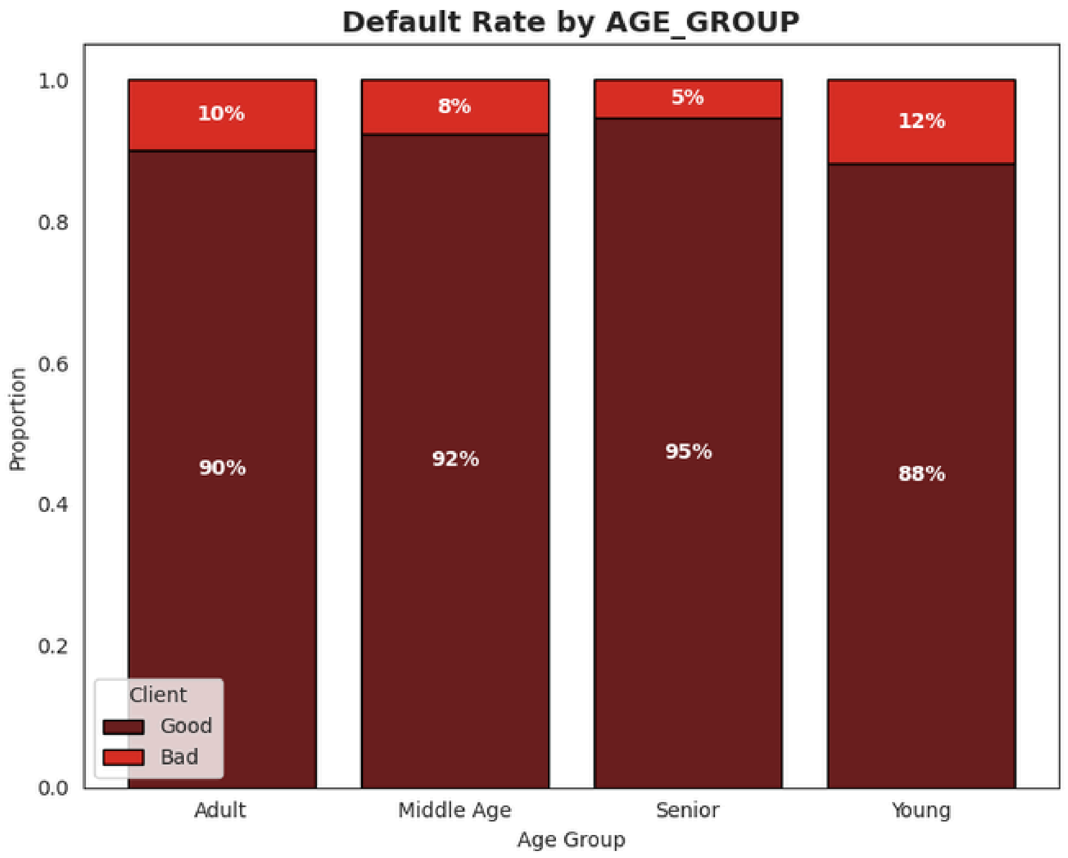
Insight:

- Most clients are in the Adult and Middle Age groups.
- Seniors show the lowest default rate (5%).
- Young clients have the highest credit risk (12% default rate).

Recommendation:

- Focus on retaining Adult and Middle Age clients.
- Use financial education & low-risk products for Young clients
- Collaborate with platforms popular among younger demographics (social media, fintech apps, etc).





IMPLEMENTATION & EVALUATION

Rakamin Academy

HOME CREDIT
Anda Bisa!

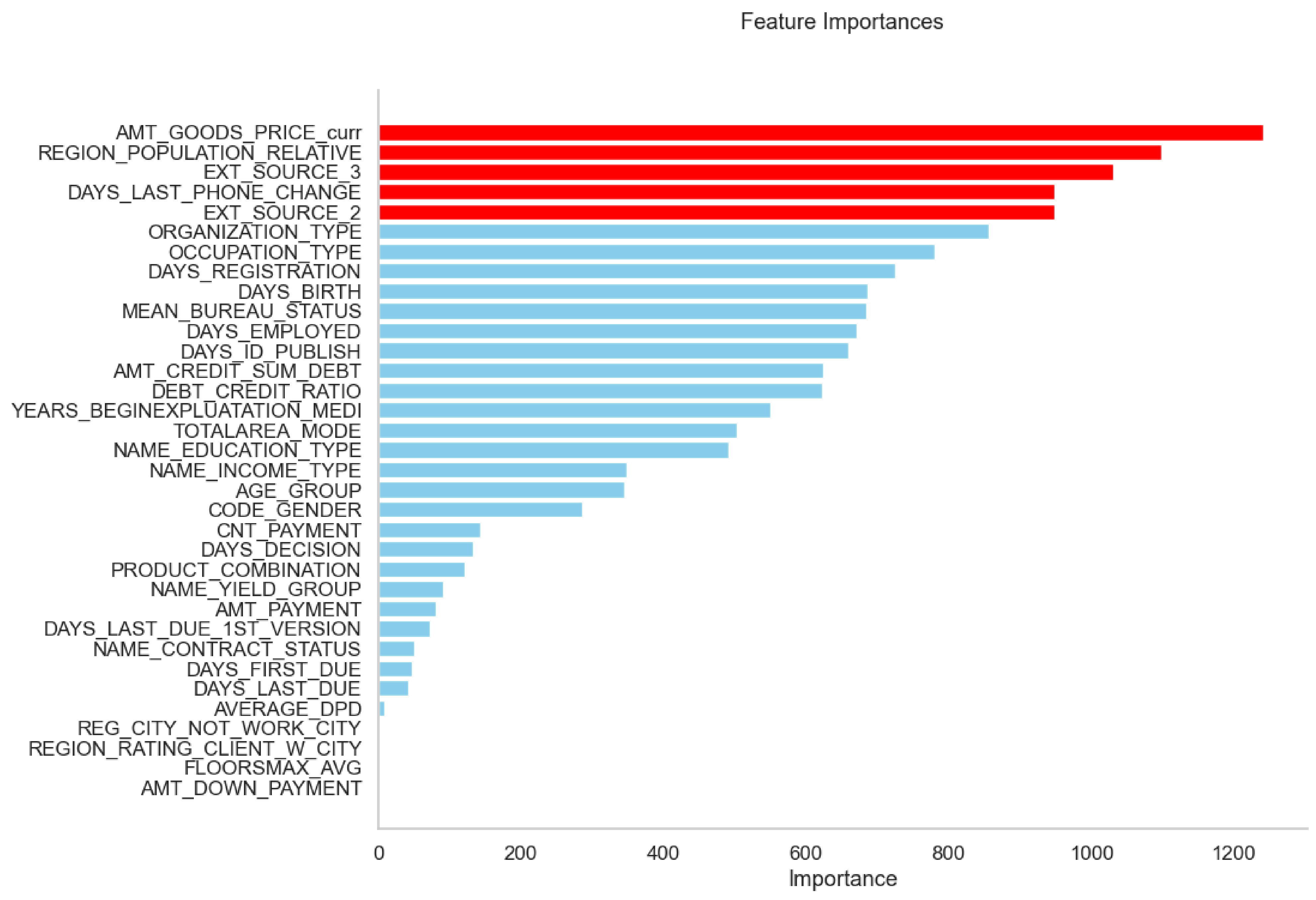
Metric Evaluation: Recall

Recall measures the model's ability to detect high-risk clients (actual positives). A high recall score indicates better performance in capturing most default cases.

Features Importance: LightGBM

Best Model: LightGBM

Best Parameters: 'num_leaves': 120, 'n_estimators': 500, 'min_child_samples': 60, 'max_depth': -1, 'learning_rate': 0.1

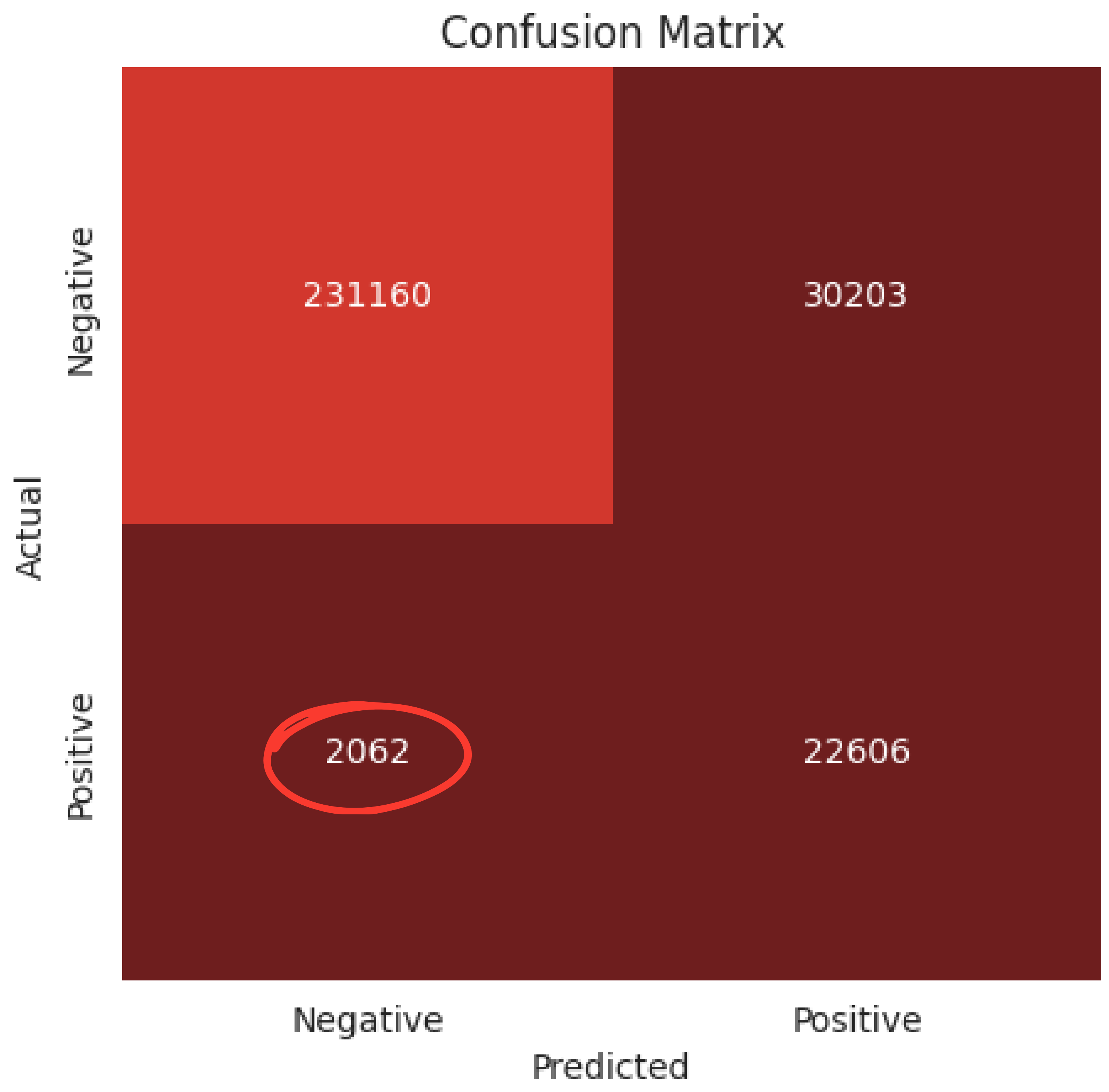


Summary Evaluation Report

Classification Report (Test):

	precision	recall	f1-score	support
0	0.99	0.88	0.93	261363
1	0.42	0.91	0.58	24668
accuracy			0.89	286031
macro avg	0.71	0.90	0.76	286031
weighted avg	0.94	0.89	0.90	286031

IMPLEMENTATION & EVALUATION

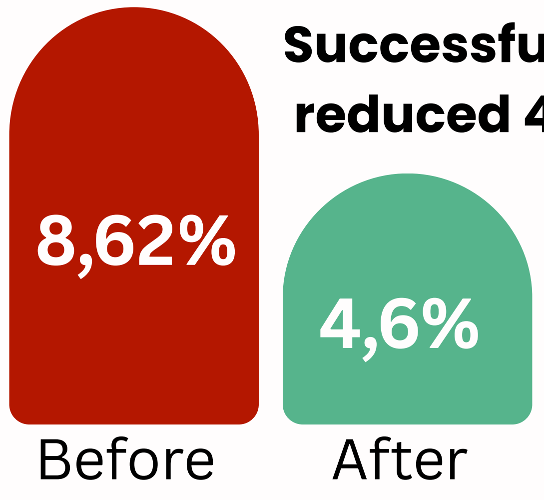


Positive: Bad Client
Negative: Good Client

Impact Simulation

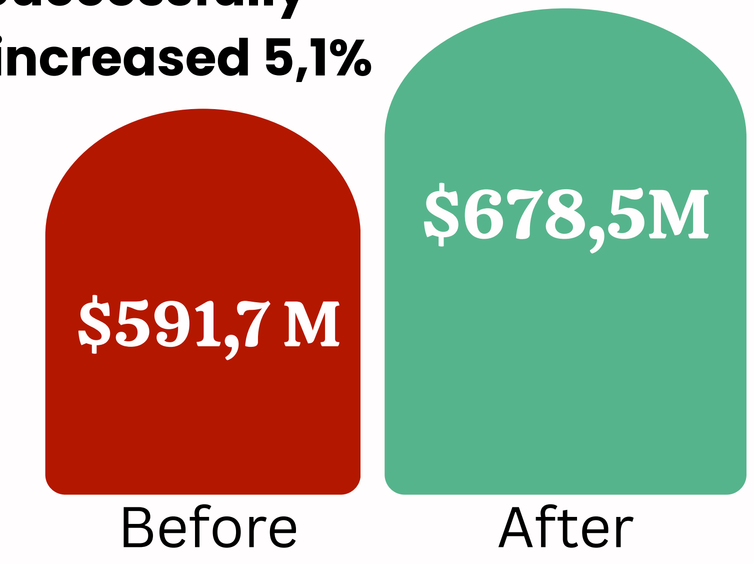
Default Rate

Successfully reduced 4%



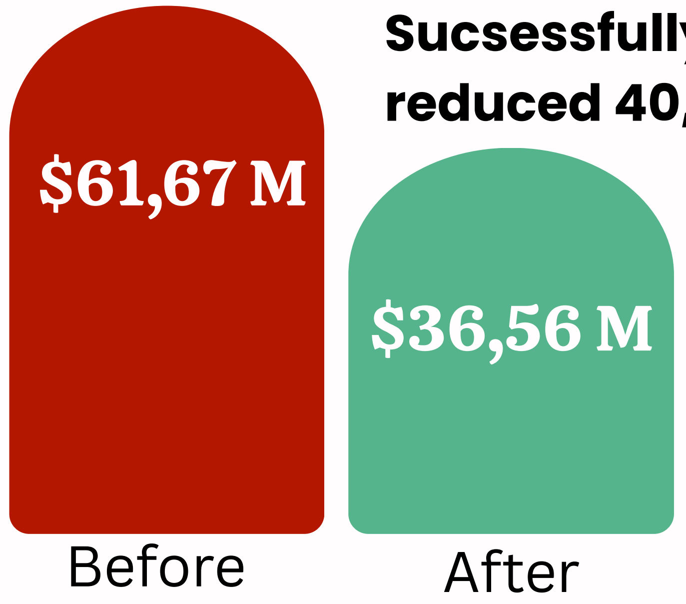
Increase in Net Revenue ($)

Successfully increased 5,1%



Saved Cost from Bad Loans ($)

Sucsessfully reduced 40,4%



BUSINESS RECOMMENDATION

Rakamin Academy

HOME CREDIT
Anda Bisa!

📍 Offer Credit Schemes Based on Risk Score & Product Value

Insight: Clients with EXT_SOURCE_3 ≥ 0.45 show significantly lower default rates

Strategy:

- Leverage ML-driven credit scoring (EXT_SOURCE_3 & EXT_SOURCE_2) to personalize loan limits, interest rates, and tenors.
- Automate installment offers based on product price (AMT_GOODS_PRICE) and customer risk profile.
- Implement real-time loan approval with transparent interest rates tailored to individual risk scores.

📍 Target High-Populations Areas with Localized Campaigns

IInsight: Customers from densely populated regions (REGION_POPULATION_RELATIVE) tend to be more digitally active and lower-risk.

Strategy:

- Use geo-AI and population heatmaps to prioritize marketing and digital expansion areas
- Launch hyper-local campaigns with customized content (e.g., regional languages, local influencers).
- Deploy digital financial literacy programs in high-potential but underbanked regions.

📍 Monitor Digital Behavior Changes as Risk Signals

Insight: Customers from densely populated regions (REGION_POPULATION_RELATIVE) tend to be more digitally active and lower-risk.

Strategy:

- Implement early warning systems based on changes in digital identity (phone, email, device).
- Trigger verification or alerts when sudden digital behavior shifts are detected.

Thank You

GitHub

LinkedIn

By Robert Gunanta Bukit