Muhammad Abdillah Caesar
Data Scientist Project Based Intern

Home Credit x Rakamin Academy

Home Credit Indonesia
Score Card Modeling

Problem Statement and Proposed Solutions

Home Credit Indonesia Score Card Modeling

Loan Approval Accuracy
Ensure deserving customers are not denied credit due to flawed assessments, preserving profitability.

Risk Identification
Detect applicants with high default risk to limit exposure and reduce potential financial losses.

Balanced Strategy
Create a robust scoring model balancing precision and recall for optimal loan decision-making.

Dataset, Goals, Objectives, and Metrics

Home Credit Indonesia Score Card Modeling

Dataset Overview

307,511 loan recipients, 122 attributes including demographics, income, and IDs.

Modeling Goal

Build predictive model to assess repayment capacity and flag potential defaulters.

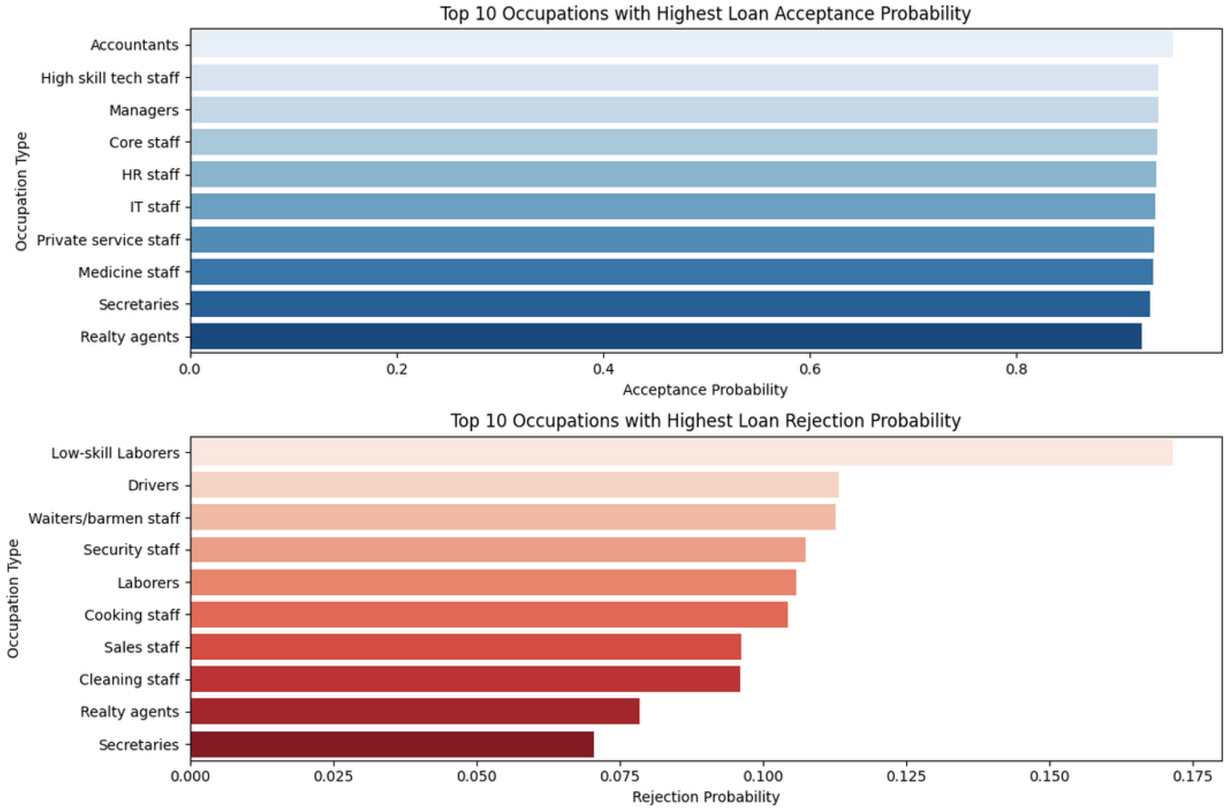
Key Metrics

Precision, Recall, and Balanced Accuracy to evaluate model performance.

Data Visualization and Business Insights

Home Credit Indonesia Score Card Modeling

- **Occupational Insights:** Identify professions with high loan approval rates using data visualizations.
- **Profitability Optimization:** Target segments with strong repayment behavior for marketing campaigns.
- **Data-Driven Strategies:** Leverage insights to refine loan offerings and risk profiling.



Data Preprocessing

Home Credit Indonesia Score Card Modeling

- **Null Value Handling:** Drop columns with >20% missing data; impute others using median/mode.
- **Feature Selection:** Remove irrelevant fields like ID and FLAG_DOCUMENT for model clarity.
- **Encoding & Balancing:** Apply label/one-hot encoding and use SMOTE for class balance.

Machine Learning Implementation and Evaluation

Home Credit Indonesia Score Card Modeling

Model Choice

1. Logistic Regression
2. Random Forest Classifier

Performance Metrics

Logistic Regression
Balanced Accuracy: 0.75;
Precision: 0.71;
Recall: 0.71

Random Forest Classifier
Balanced Accuracy: 0.90; Precision: 0.98; Recall: 0.82

Interpretation

Logistic Regression offers clear, audit-friendly coefficients but, with balanced accuracy 0.75 and precision/recall 0.71, still misclassifies nearly a third of cases; by contrast, Random Forest's balanced accuracy 0.90, precision 0.98, and recall 0.82 sharply cut false approvals while retaining most credit-worthy clients

Business Recommendations

Home Credit Indonesia Score Card Modeling

Sector Focus
Prioritize lending to high-approval occupations: Accountants, Managers, Tech Staff.

Demographic Targeting
Promote loans to married women with good histories using tailored incentives.

Model Enhancement
Improve recall, add external data, and build ML-based pre-approval systems.