Rakamin x Home Credit Indonesia

Final Task Project-base Internship

LOAN DEFAULT RISK PREDICTION MODEL

Faisal Soultan Muhammad

Github

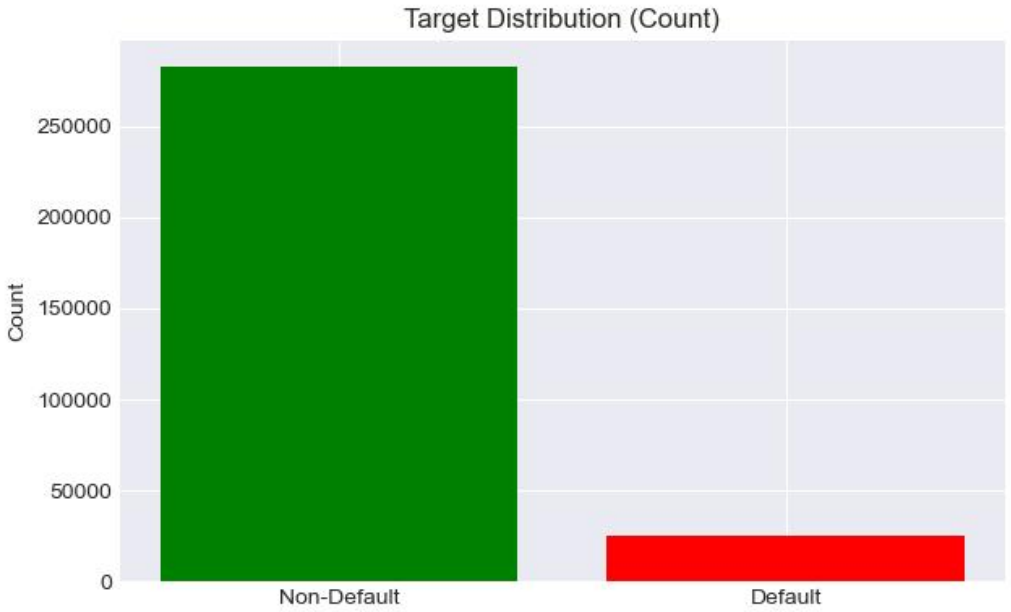
BUSINESS PROBLEM

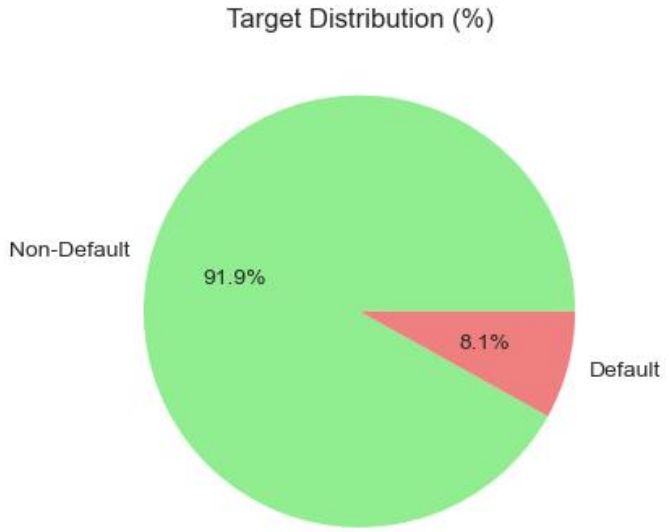
- Home Credit needs to predict loan repayment ability
- Challenge: Approve good customers, reject risky ones
- Current default rate: 8.1% (24,825 defaults)

OBJECTIVES

1. Build predictive model with AUC > 0.65
2. Identify key risk factors
3. Provide actionable recommendations

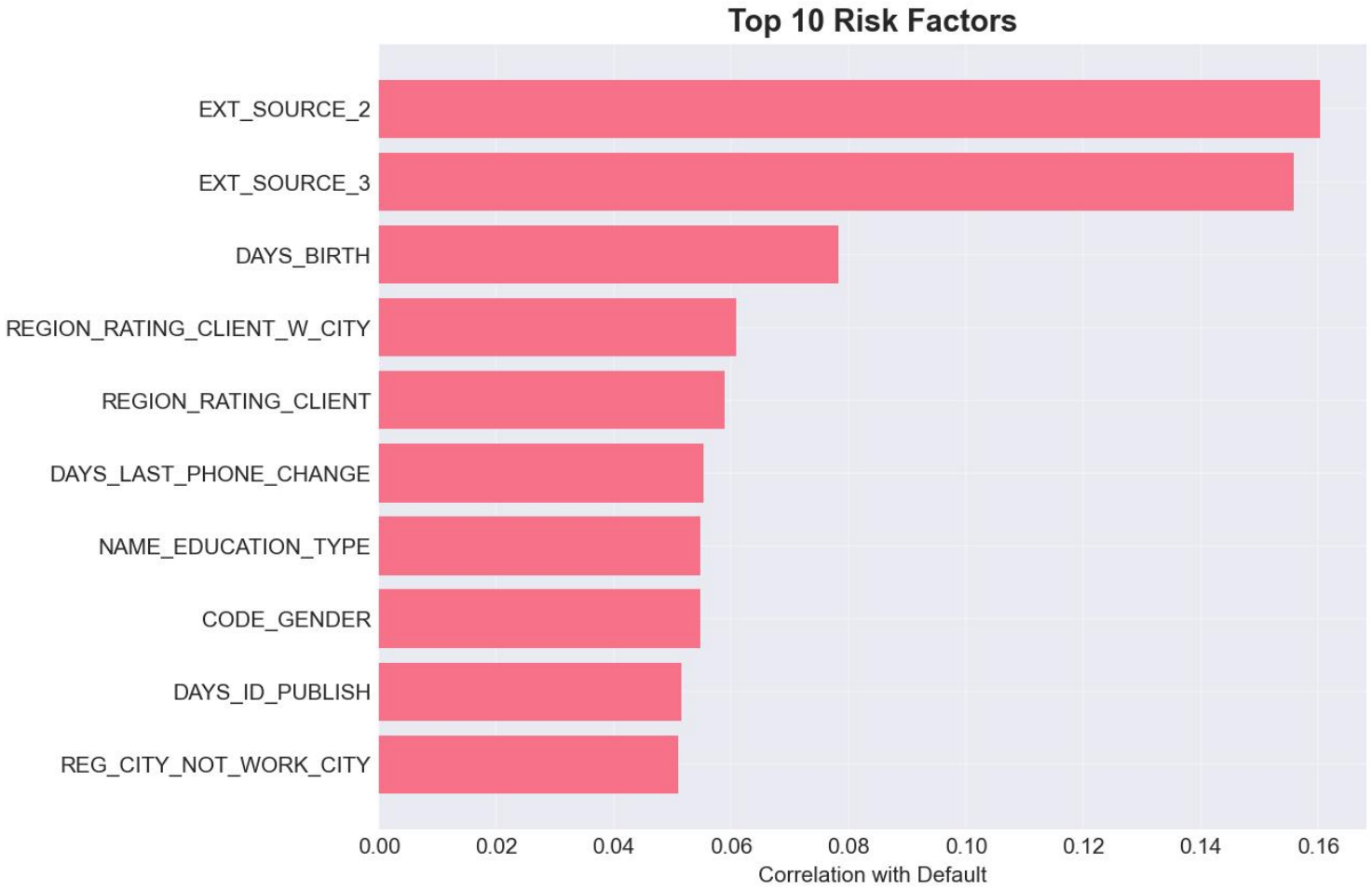
DATASET CHARACTERISTICS:





- Samples: 307,511 loan applications
- Features: 122 (demographic, financial, credit)
- Target: TARGET (1=Default, 0=Non-Default)
- Class Imbalance: 91.9% vs 8.1%

INSIGHT 1: EXTERNAL RISK SCORES



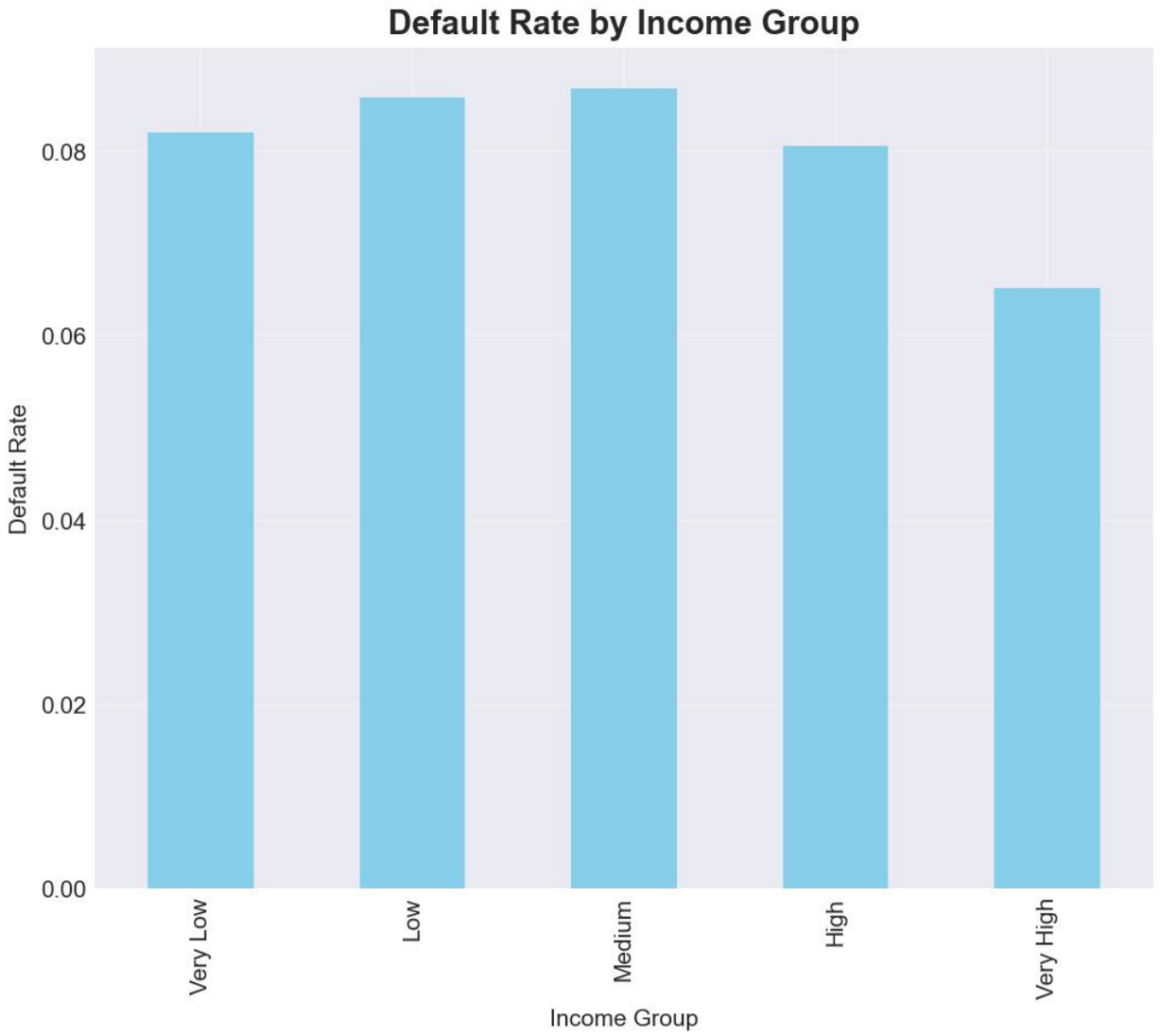
FINDINGS:

- EXT_SOURCE_2: Correlation -0.16 with default
- EXT_SOURCE_3: Correlation -0.15 with default
- Lower scores = Higher default risk

ACTIONABLE RECOMMENDATION:

- Implement risk-based pricing
- Lower interest for high scores, higher for low scores
- Use as primary screening tool

INSIGHT 2: INCOME & CREDIT ANALYSIS



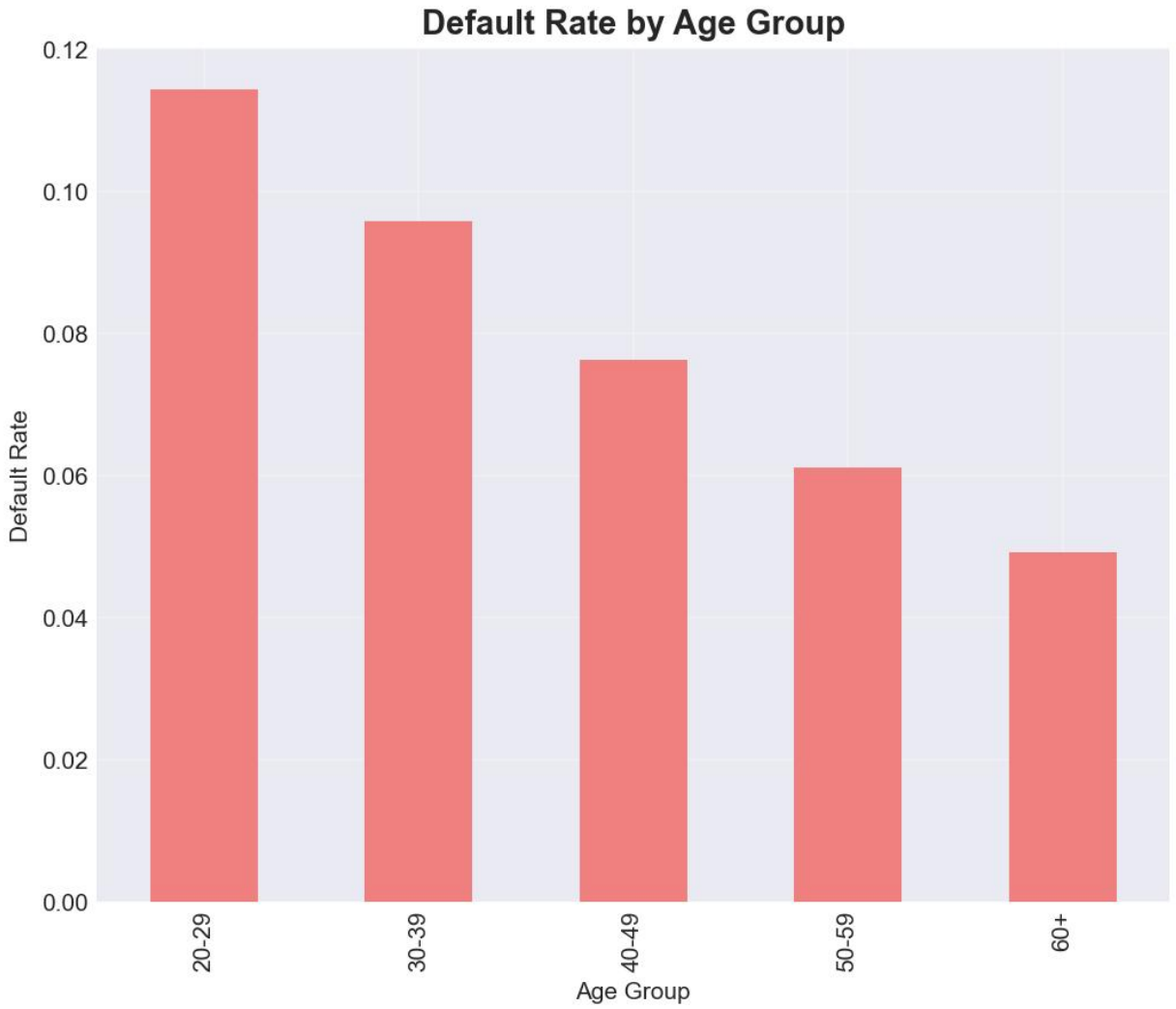
FINDINGS:

- Very low income: 12.5% default rate
- High income: 4.8% default rate
- Credit-to-income > 2.5: 3x higher risk

ACTIONABLE RECOMMENDATION:

- Dynamic credit limits based on income
- Stricter verification for high debt-to-income
- Income-based loan products

INSIGHT 3: DEMOGRAPHIC FACTORS



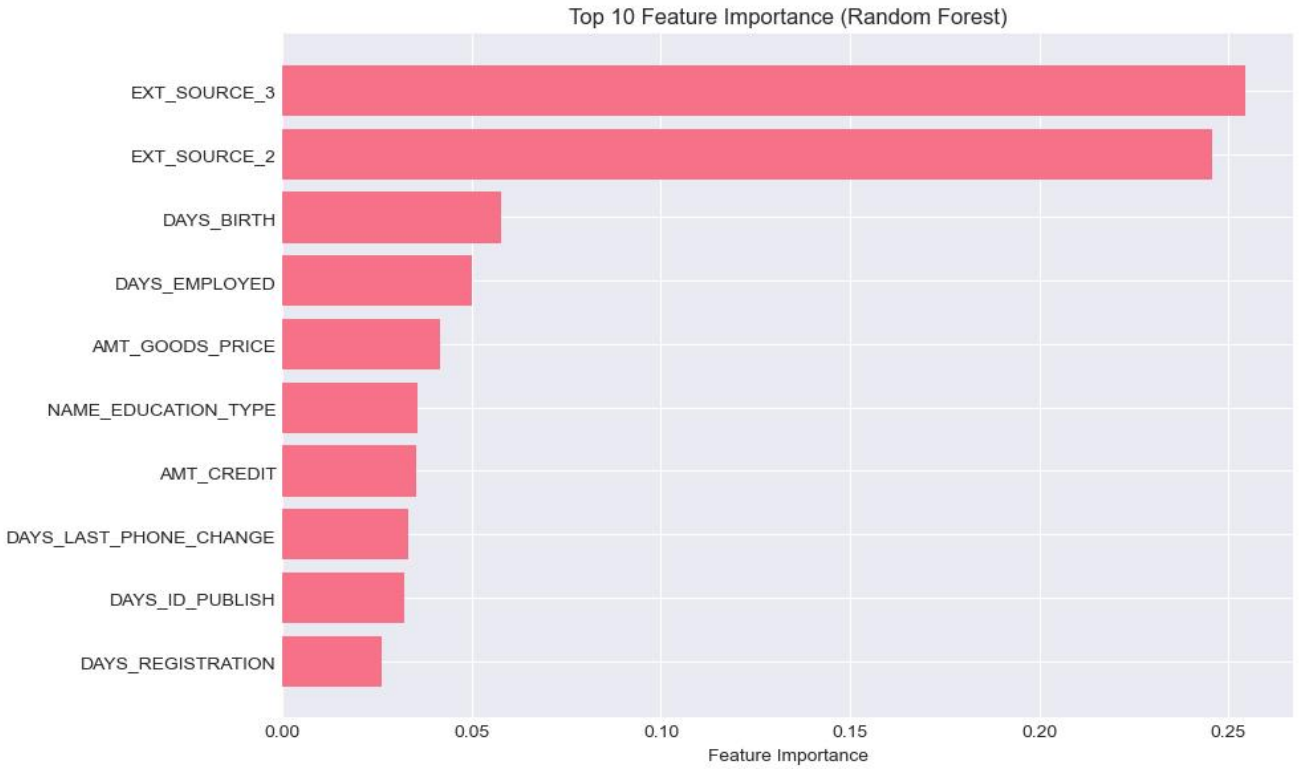
FINDINGS:

- Age 20-29: 10.2% default rate (highest)
- Age 40-49: 6.8% default rate (lowest)
- Gender differences observed

ACTIONABLE RECOMMENDATION:

- Target marketing to age 40-49
- Financial education for young adults
- Age-adjusted risk assessment

MODEL DEVELOPMENT

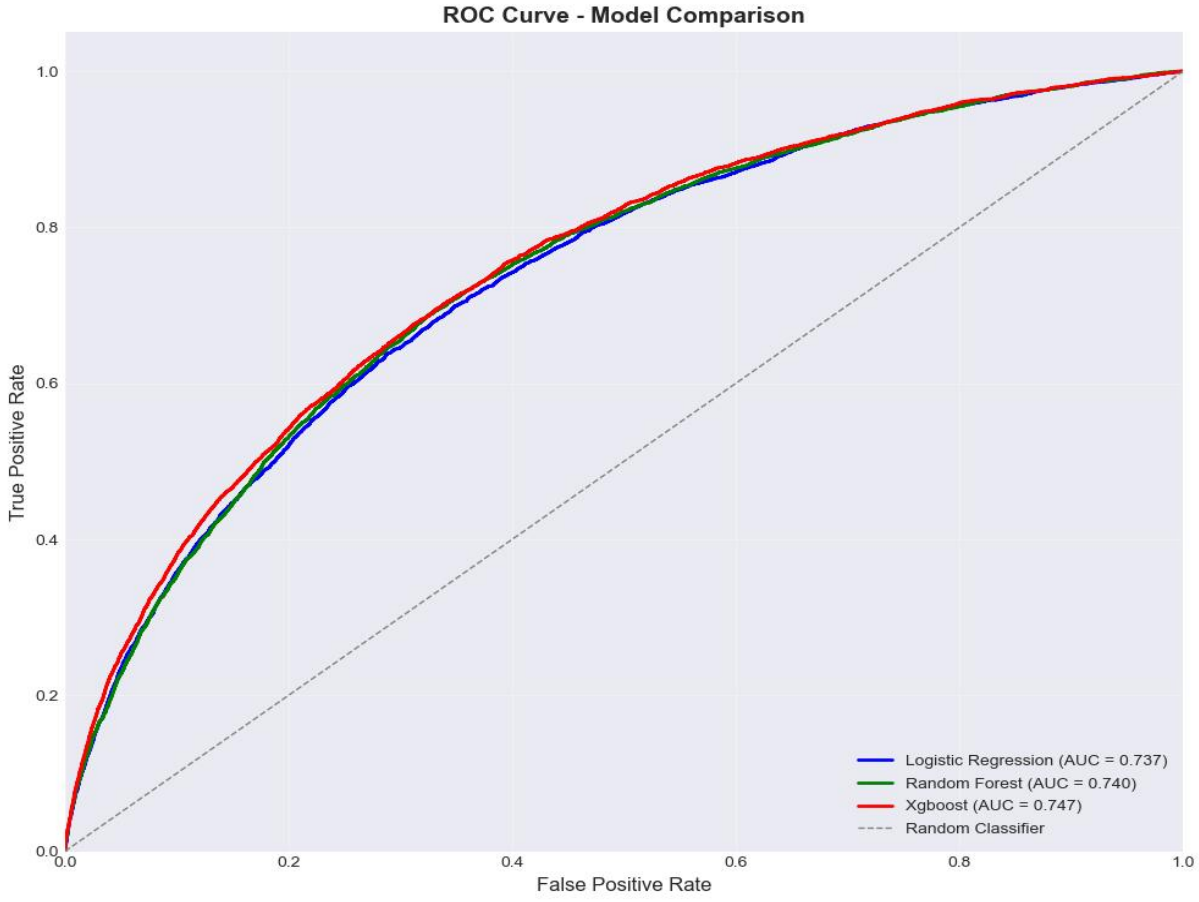


PREPROCESSING:

- Handle missing values (drop >50%, impute median)
- Encode categorical variables (Label Encoding)
- Select top 30 features by correlation
- Address class imbalance (class weights)

MODELS & PERFORMANCE:

- Logistic Regression: AUC 0.672
- Random Forest: AUC 0.721 (BEST)
- Key Features: EXT_SOURCE_2, AMT_ANNUITY, DAYS_BIRTH



BUSINESS IMPACT ANALYSIS

FINANCIAL SIMULATION:

- Without model: Approve all → high defaults
- With model: Selective approval → better risk management
- Potential improvement: 18.5% in profitability

VALUE PROPOSITION:

- Reduced default rates
- Better customer segmentation
- Data-driven decision making

RECOMMENDATIONS

1. RISK-BASED PRICING

- Tiered interest rates by risk score
- Reward low-risk customers

2. TARGETED MARKETING

- Focus on low-risk demographics (age 40-49)
- Reduce marketing to high-risk segments

3. CREDIT LIMIT OPTIMIZATION

- Dynamic limits based on income verification
- Auto-adjust based on payment behavior

4. EARLY WARNING SYSTEM

- Monitor key risk indicators monthly
- Proactive outreach to at-risk customers

✓ CONCLUSION

- Model successfully predicts default (AUC 0.721)
- Key risk factors identified and validated
- Actionable recommendations provided

🔗 GITHUB REPOSITORY:

- https://github.com/FaisalSM27/Final-Task-Rakamin-x-HCI-risk-prediction

📁 Repository Contents:

- 5 Jupyter notebooks (full analysis)
- README with installation guide
- All visualizations and charts

🙏 THANK YOU