

REFACTORY CAPSTONE PROJECT PROPOSAL

LOAN DEFAULT PREDICTION CHALLENGE

BY

NANSUBUGA NORAH CATHERINE
AMBALE TARA
STELLA ADONGO
PATIENCE RUGYENDO

April 23, 2024

1 Introduction

SuperLender, a local digital lending company, prides itself on its effective credit risk models that drive profitable and impactful lending decisions. The company's assessment strategy revolves around two key risk drivers: the borrower's willingness to pay and their ability to pay. To maintain its competitive edge, SuperLender invests in experienced data scientists to develop robust models that accurately predict the odds of loan repayment. This proposal outlines our approach to building predictive models for loan default prediction, focusing on assessing both new business risk and repeat customer behavior risk.

2 Data Description

The dataset provided contains comprehensive information on loan applications, including customer/borrower demographics, financial history/performance, previous loans for customers who have been clients before, loan terms, and repayment outcomes. It encompasses both new business and repeat customer data, enabling us to build separate models for each scenario. The dataset is structured with a binary outcome variable indicating whether a loan was "Good" (1) or "Bad" (0) based on repayment status.

3 Challenge Overview

The primary challenge is to develop robust predictive models that accurately classify loan applications as either good or bad, based on the provided data and predict the likelihood of loan repayment at the point of each application, enabling the credit grantor to make informed decisions regarding loan approval, size, price, and tenure. The models must effectively assess the borrower's willingness to pay and ability to pay. Additionally, the challenge entails addressing the specific requirements for new business risk models for assessing first-time loan applicants and repeat or behavior risk models for existing customers applying for repeat loans. These models will accurately predict the binary outcome variable, where a "Good" loan is denoted by 1 and a "Bad" loan by 0.

4 Proposed Approach

Our proposed approach involves the following key steps:

4.1 Data preprocessing

We will conduct thorough preprocessing of the data provided on Zindi to handle missing values, outliers, and feature engineering. This step is crucial for ensuring the quality and relevance of the input data for model training.

4.2 Model Development

- **New Business Risk Model:** For first-time loan applicants, we will develop predictive models using a variety of machine learning algorithms such as logistic regression, decision trees, and ensemble methods. These models will assess the borrower's likelihood of repayment based on application information.
- **Repeat Customer Behavior Model:** For existing customers applying for repeat loans, we will incorporate historical performance data from their prior loans. This information will be used to enhance the prediction accuracy by capturing the borrower's repayment behavior over time.

4.3 Evaluation Metrics

The performance of the developed models will be evaluated using metrics such as accuracy, precision, recall and F1-score. These metrics will provide insights into the models' ability to correctly classify loan applications and effectively differentiate between good and bad loans.

5 Expected Outcomes

Our anticipated outcomes include:

1. Development of accurate and robust predictive models for loan default prediction, tailored to both new business and repeat customer scenarios.
2. Insights into key factors influencing loan repayment behavior, facilitating better risk assessment and decision-making for lenders.
3. Documentation of the methodology, experimental results, and recommendations for deploying the models in real-world lending scenarios.

6 Conclusion

In conclusion, our proposed project aims to address the critical need for accurate loan default prediction models at SuperLender. By leveraging advanced machine learning techniques and comprehensive data analysis, we aim to empower the company to make informed lending decisions, minimize default risks, and drive sustainable growth.