Case Study: Banking & Loans in Ghana

Focus: Reducing Non-Performing Loans (NPLs)

GhanaLoanConnect

February 12, 2025

Introduction

GhanaLoanConnect is a Non-Banking Financial Institution (NBFI) that facilitates peer-to-peer lending. By connecting individuals who need financial support (borrowers) with those who have excess funds and are willing to invest (lenders), the platform plays a crucial role in the financial ecosystem. However, like many lending institutions, GhanaLoanConnect faces the challenge of non-performing loans (NPLs) — loans where the borrower defaults or fails to make timely repayments.

1 Business Challenge

Objective: Develop a supervised machine learning model to predict whether a borrower is likely to default or successfully repay a loan. This predictive model will enable the institution to be more selective in approving loans, thus reducing the risk of NPLs and ensuring a healthier loan portfolio.

Key Considerations:

- **Minimizing Financial Risk:** The primary goal is to identify high-risk borrowers in advance, thereby reducing the rate of default.
- **Informed Lending Decisions:** A robust model will help the institution make data-driven decisions when approving loan applications.

2 Dataset Overview

The dataset for this case study is provided in the file `loan_borrower_data.csv` and contains approximately 9,578 records. The key attributes include:

- **credit.policy:** Indicates whether a borrower meets GhanaLoanConnect's credit approval standards. **Values:** 1 (eligible) and 0 (ineligible).

- **purpose:** The intended use of the loan. Categories include `credit_card`, `debt_consolidation`, `educational`, `major_purchase`, `small_business`, and `all_other`.
- **int.rate:** The interest rate charged on the loan, represented as a proportion (e.g., 0.11 for an 11% rate). Higher-risk borrowers are typically charged higher interest rates.
- **installment:** The fixed monthly repayment amount required from the borrower if the loan is approved.
- **log.annual.inc:** The natural logarithm of the borrower's self-reported annual income. This transformation helps manage skewness in income distribution.

3 Project Requirements

Task: Build a machine learning model to predict the likelihood of a borrower defaulting on a loan.

Key Steps:

1. **Data Exploration and Preprocessing:**
   - Load and inspect the dataset.
   - Handle missing or anomalous data.
   - Perform necessary transformations (e.g., converting categorical variables using one-hot encoding, scaling numerical features if required).
2. **Feature Engineering:**
   - Investigate which features are most indicative of loan performance.
   - Consider creating additional features (e.g., interaction terms or derived features) that might improve model performance.
3. **Model Selection and Training:**
   - Experiment with various supervised learning algorithms (e.g., logistic regression, decision trees, random forests, gradient boosting machines, etc.).
   - Use appropriate metrics (accuracy, precision, recall, F1-score, ROC-AUC) to evaluate model performance.
4. **Validation and Testing:**
   - Split the dataset into training and testing sets.
   - Use cross-validation techniques to ensure model robustness.

- Evaluate model performance on unseen data.

5. **Model Interpretation and Insights:**
   - Identify the key predictors of loan default.
   - Provide business insights that can inform GhanaLoanConnect's lending policies.
6. **Reporting:**
   - Document the model development process.
   - Present findings in a clear and concise report, highlighting how the model can be used to reduce NPLs.

4 Deliverables

Students are expected to provide the following:

1. **Code Repository:** A well-documented codebase (preferably in Jupyter Notebook, RMarkdown, or a similar environment) detailing each step of the project.
2. **Final Report:** A comprehensive document (in PDF format) that includes:
   - Introduction and Problem Statement.
   - Data Exploration and Preprocessing Steps.
   - Methodology and Model Building.
   - Evaluation Metrics and Results.
   - Interpretation of Findings.
   - Recommendations for GhanaLoanConnect.
3. **Presentation:** A slide deck summarizing the project approach, key insights, and final recommendations.

5 Evaluation Criteria

Projects will be evaluated based on:

- **Understanding of the Business Problem:** Clear articulation of the challenges and the rationale behind the modeling approach.
- **Data Handling and Preprocessing:** Effective use of data cleaning, transformation, and feature engineering techniques.
- **Modeling Approach:** Justification for the chosen models and the application of sound machine learning principles.

- **Evaluation and Validation:** Robustness of the model using appropriate metrics and cross-validation techniques.
- **Interpretability and Business Insights:** Clarity in presenting key features influencing loan default and actionable recommendations.
- **Clarity and Presentation:** Quality of documentation and presentation of findings.

6 Additional Considerations

- **Ethical Considerations:** Discuss potential ethical implications, including fairness and bias in the model predictions. Ensure that the model does not inadvertently discriminate against any borrower group.
- **Scalability and Future Work:** Consider how the model might be updated with new data and the steps necessary to scale the solution for real-world application.
- **Limitations:** Reflect on any limitations of the current approach and suggest possible improvements.

Good luck, and we look forward to your innovative solutions that could help shape the future of lending at GhanaLoanConnect!