Loan Default Prediction

Report 1: Project Approach & Team Contributions
Project Title: Loan Default Prediction for GhanaLoanConnect

Organized by Thrive Africa
Machine Learning 2025 Cohort

Submitted by: Patrick Ayeh Ayisi (PM)
Date: June 02, 2025

TEAM (ALPHA-2)

Patrick Ayeh Ayisi

Umar Zaidu Yakubu

David Wilson

Owiredu-Amoh Baffuor Etto Jnr.

Marvin Dawson

Jemima Ama Fiapemetsi

Iwikotan Oreofe Gloria

Maya Leotina Agebedekpui

Objective

To build a machine learning model that can predict loan default (i.e., whether a borrower will not fully pay back a loan), helping GhanaLoanConnect minimize Non-Performing Loans (NPLs) and improve loan approval decisions.

Team Structure & Roles

Part A: Data Preprocessing & Exploration
Team Members: *Maya, Baffuor Jnr*
Tasks:

- Loaded the dataset
- Assessed data types, missing values, and statistical summaries
- Visualized the distribution of key features and the target variable
- Ensured a clear understanding of data structure before modeling

Part B: Feature Engineering & Data Splitting
Team Members: *Gloria, Umar*
Tasks:

- Performed one-hot encoding of the `purpose` feature
- Created a new binary feature `long_credit_history` based on credit length
- Discussed potential derived features (e.g., high DTI, low FICO)
- Split data into training and test sets for modeling

Part C: Model Building & Baseline Evaluation
Team Members: *David, Jemima*
Tasks:

- Trained initial models: Logistic Regression and Random Forest
- Evaluated models using Accuracy, Precision, Recall, and F1-Score
- Identified Logistic Regression as a simple but limited model

Part D: Advanced Models & Feature Importance
Team Members: *Patrick, Marvin*
Tasks:

- Implemented Gradient Boosting Classifier
- Tuned hyperparameters using GridSearchCV
- Applied SMOTE for class imbalance
- Used Random Forest to generate and visualize feature importances

Part E: Report Writing & Presentation
Team Member: *Patrick*
Tasks:

- Compiled all findings and visualizations
- Drafted the technical report and summary presentation
- Coordinated communication among sub-teams for cohesion

Workflow Summary
1. Team brainstorming: Defined goal and divided work
2. Collaborative coding: Shared Google Colab notebook
3. Progress check-ins: Each part was reviewed by Patrick
4. Integration: All parts combined and validated
5. Final review: Code cleanup, visualization tuning, report draft

Tools Used

- Python (Pandas, Scikit-learn, Matplotlib, Seaborn)
- Google Colab
- SMOTE (via imblearn) for handling class imbalance

Outcome
A collaborative and structured approach ensured all key tasks were covered with clear ownership and accountability. Each contributor's section directly improved the final model's quality and interpretability.