

Group_4_(Foxtrot_2)_Project_Work

May 21, 2025

[]:

1 Case Study: Banking & Loans in Ghana

1.1 Project Plan Report

1.1.1 Introduction

In today's competitive and digitally driven financial environment, the ability to predict credit risk is a strategic advantage. GhanaLoanConnect, a digital lending platform connecting borrowers to investors, is actively seeking data-driven solutions to reduce its rate of non-performing loans (NPLs). Rising default rates not only impact profitability but also weaken lender trust and reduce platform credibility.

To address this challenge, this project proposes the development of a machine learning model that can predict the likelihood of a borrower defaulting on a loan. With accurate predictions, GhanaLoanConnect can proactively minimize risk, prioritize low-risk borrowers, and enhance portfolio performance.

1.1.2 Objectives

This project aims to apply machine learning techniques to build a binary classification model for predicting borrower default risk. Using a real-world dataset from GhanaLoanConnect, the objectives are as follows:

1. Conduct exploratory data analysis (EDA) to understand variable distributions and relationships.
2. Perform feature engineering to optimize model performance.
3. Implement, train, and evaluate supervised learning algorithms using metrics such as ROC-AUC, F1-score, and precision-recall.
4. Interpret model outputs to identify significant predictors of loan default.
5. Establishing a feedback loop for continuous improvement through model retraining and performance monitoring.

1.1.3 Methodological Approach

This project will follow a structured machine learning workflow to ensure rigor, reproducibility, and actionable results. The process includes the following phases:

1. **Problem Definition and Goal Setting** Understand the business context of non-performing loans (NPLs) and frame the loan default problem as a binary classification task.

2. **Data Exploration and Preprocessing**

- a. Load and inspect the dataset to assess completeness and structure.
- b. Conduct exploratory data analysis (EDA) to examine feature distributions, detect outliers, and explore relationships between variables.
- c. Handle missing values, normalize skewed data, and convert categorical features (e.g., purpose) using one-hot encoding.

3. **Feature Engineering**

Analyze correlations and domain relevance to select informative features.

Create derived variables (e.g., income-to-installment ratio) to improve model signal.

4. **Model Development**

Apply multiple supervised learning algorithms (Logistic Regression, Decision Tree, Random Forest, XGBoost).

Use grid search and cross-validation to optimize hyperparameters.

5. **Evaluation**

Assess performance on validation and test sets using metrics such as Accuracy, Precision, Recall, F1-score, and ROC-AUC.

Compare models to select the best-performing one.

6. **Model Interpretation and Business Insights**

Analyze feature importance to understand drivers of default.

Translate findings into policy recommendations for credit risk management.

7. **Ethical Review**

Examine potential biases in features or outcomes.

Ensure fair treatment across borrower groups.

8. **Reporting and Presentation**

Compile findings in a formal report and deliver a summary presentation to stakeholders.

1.1.4 Project Roles and Responsibilities

To ensure efficient teamwork, the project has been divided into key phases, with responsibilities evenly distributed among all group members based on workload and collaboration requirements. Each member contributes to both technical tasks and project deliverables.

| Name | Role/Title | Responsibilities |
|-----------------------------|----------------------------------|---|
| Aminatu Yusif | Group Leader / Product Manager | Project oversight, coordination, final review of deliverables |
| Evans Kabu Akrofi | Data Cleaning & Feature Engineer | Data integrity, missing value handling, feature creation. |
| Andrews Wese Nasara | EDA & Visualization Lead | Performs and presents exploratory analysis using plots and dashboards. |
| Emmanuel Doe-Afealete | Machine Learning Engineer (1) | Logistic Regression & Decision Tree implementation. |
| Aboagye Jonathan | Machine Learning Engineer (2) | Random Forest & XGBoost implementation; model tuning |
| Enchill Delsie Dwomoh Fokuo | Evaluation & Metrics Analyst | Leads model evaluation, error analysis, and performance benchmarking. |
| Azure Jeremy | Ethics & Fairness Lead | Oversees bias detection and fairness audits in model behavior |
| Godly Tettey Kumah | Reporting & Documentation Lead | Assembles final PDF report, writes business insights, coordinates formatting. |
| Aminatu Yusif (dual role) | GitHub Repository Manager | Creates and maintains the GitHub repo, uploads notebooks/code, manages version control, issues, and README updates. |

Collaborative Tasks (All Members Contribute)

Initial problem analysis and case study review.

Peer review of code and report drafts.

Final review and polish of deliverables.

Group discussions and decision-making throughout project stages.

This structure ensures that:

Each member contributes significantly.

Leadership and final accountability rest with the Group Leader.

Workload is balanced across analysis, modeling and documentation.

1.1.5 Conclusion

This project is a strategic response to the pressing issue of loan defaults at GhanaLoanConnect. By employing a rigorous, end-to-end machine learning approach, our team seeks to transform raw borrower data into actionable intelligence. The proposed model will not only enable early identification of high-risk loan applicants but also support more consistent and data-driven credit decisions.

Beyond technical implementation, the project emphasizes business relevance, ethical responsibility, and long-term scalability. Through effective collaboration, thoughtful analysis, and clear communication of insights, we aim to deliver a solution that strengthens financial risk management and enhances the overall quality of GhanaLoanConnect's lending portfolio.