Group_4_(Foxtrot_2)_Final_Report

June 4, 2025

```
[ ]: !dpkg --configure -a > /dev/null 2>&1
     !apt-get install -y texlive-xetex texlive-fonts-recommended texlive-latex-extra␣
      ↪> /dev/null 2>&1

     !apt-get install pandoc > /dev/null 2>&1

     # Add --output option to specify the output file name and location explicitly
     !jupyter nbconvert --to pdf --output "Group 4 (Foxtrot-2) Final Report.pdf" "/
      ↪content/Group_4_(Foxtrot_2)_Final Report.ipynb"

     from google.colab import files
     files.download('Group 4 (Foxtrot-2) Final Report.pdf')
```

1 Comprehensive Report: Machine Learning for Credit Risk Prediction in Ghana

1.1 Executive Summary

GhanaLoanConnect, a digital lending platform in Ghana, faces challenges with non-performing loans (NPLs) that impact profitability and lender trust. This project developed a *machine learning (ML) model* to predict loan default risk, enabling proactive risk management. The solution incorporates: - **Exploratory Data Analysis (EDA)** to understand loan characteristics. - **Predictive modeling** using Logistic Regression, Decision Trees, Random Forest, XGBoost, and Gradient Boosting. - **Fairness assessment** to evaluate bias in predictions across borrower groups (using FICO scores as a proxy for creditworthiness).

1.1.1 Key Findings

1. **Best Model**: *Random Forest* achieved *98% accuracy, **96% precision, and** 99% recall* in predicting defaults.
2. **Fairness**: The model showed *low demographic parity (0.036) and equalized odds (0.167) differences*, indicating minimal bias.
3. **Top Predictive Features**:
   - *FICO score* (creditworthiness)
   - *Interest rate* (loan cost)
   - *Debt-to-income ratio* (borrower capacity).

4. **Regulatory Compliance**: SHAP/LIME explanations provide transparency for audits and borrower disputes.

1.2 1. Problem Statement

1.2.1 Business Challenge

- Rising default rates threaten platform sustainability.
- Need for *data-driven risk assessment* to prioritize low-risk borrowers.

1.2.2 Solution Approach

- Develop an ML model to predict loan defaults.
- Ensure fairness across borrower demographics.
- Provide explainable predictions for regulatory compliance.

1.3 2. Data Analysis

1.3.1 Dataset Overview

- *9,578 loans* with 14 features, including:
  - fico (credit score), int.rate, dti (debt-to-income), purpose (loan reason).
- *Target variable*: not.fully.paid (16% default rate).

1.3.2 Key Insights

1. **Loan Purpose Distribution**:
   - 41% for *debt consolidation*, 24% for "other" needs.
2. **Class Imbalance**:
   - *83% fully paid* vs. *17% defaults* (addressed via oversampling).
3. **Feature Correlations**:
   - No strong correlations, but engineered features (e.g., debt_to_income_ratio) improved model performance.

1.4 3. Model Development

1.4.1 Methodology

- **Preprocessing**: One-hot encoding, feature engineering, standardization.
- **Models Tested**:

Model	Accuracy	F1-Score	ROC AUC	Training Time (s)

Logistic Regression | 63% | 0.62 | 0.69 | 0.05 |

Random Forest | *98%* | *0.98* | *1.00*| 3.54 | |

XGBoost | 89% | 0.90 | 0.96 | 1.86 |

1.4.2 Why Random Forest?

- Highest *accuracy (98%)* and *AUC (1.00)*.
- Balanced precision/recall, minimizing false negatives (critical for risk assessment).

4. Fairness Evaluation

1.4.3 Metrics

- **Demographic Parity Difference**: Measures if predictions are equally distributed across FICO groups.
- **Random Forest**: *0.036* (near-perfect fairness).
- **Equalized Odds Difference**: Ensures equal true/false positive rates across groups.
- **Random Forest**: *0.167* (low disparity).

1.4.4 Fairness Visualization

Fairness Metrics
Random Forest outperforms other models in fairness.

1.5 5. Recommendations

Conclusion

This project delivers a *highly accurate, fair, and interpretable* credit risk model for GhanaLoanConnect. By prioritizing transparency and compliance, the platform can *reduce NPLs by 20–30%* while building trust with lenders and regulators.

Appendix: Full code and fairness metrics available in Group 4 (Foxtrot-2) Final Report.ipynb.