

Project -1

Title page:

Project title:

Predicting Loan Repayment Probabilities

Your Name:

Durga Bhavani Yellaram

Date:

05/02/2025

Abstract:

This project focuses on predicting loan repayment probabilities for customers of a Microfinance Institution (MFI) in partnership with a telecommunications company. The main objective is to identify potential defaulters among low-income individuals using mobile financial services (MFS). With MFS playing a crucial role in expanding credit access to underserved communities, our model aims to enhance customer selection and reduce financial risks.

Our approach involves thorough data cleaning, exploration, visualization, and feature engineering, followed by training 45 machine learning models. Hyperparameter tuning is performed to optimize performance, with evaluation metrics including log loss, precision, and recall. The results demonstrate that our predictive model effectively classifies customers as defaulters or non-defaulters, offering valuable insights into repayment behavior. This capability is critical for improving microfinance strategies and supports the broader mission of poverty reduction through better access to financial services.

Future work will focus on model deployment and integrating additional data sources to refine predictions further.

Table of Contents:

1. Introduction
2. Literature Review
3. Problem Statement
4. Data Collection and Preprocessing
5. Methodology
6. Implementation

7. Results
8. Discussion
9. Conclusion
10. References
11. Appendices
12. Acknowledgments (optional)

1.Introduction:

Microfinance Institutions (MFIs) have emerged as pivotal entities in the global effort to alleviate poverty, especially among low-income populations. These organizations provide financial services tailored to underserved communities, facilitating access to credit for individuals and small businesses that traditionally lack the means to secure loans from conventional banking institutions. MFIs aim to empower unbanked families, particularly those residing in remote areas, by offering diverse financial products such as group loans, agricultural loans, and individual business loans.

The integration of mobile financial services (MFS) into microfinance has revolutionized the sector by enhancing convenience and accessibility for clients. With the proliferation of mobile technology, MFIs can now deliver financial services through mobile platforms, making it easier for clients to access loans and manage repayments. This shift towards MFS is particularly significant in regions where traditional banking infrastructure is limited or absent. MFS not only improves operational efficiency for MFIs but also helps clients manage their finances more effectively, providing them with essential support in times of need.

Despite the potential benefits of MFS, the implementation has not been uniform across the industry. While some MFIs have successfully leveraged mobile technology to expand their reach, others face challenges in adoption and customer repayment rates. The need to accurately assess the likelihood of loan repayment becomes critical in this context. Effective risk assessment models can help MFIs identify customers who are likely to default, thereby enabling better decision-making in loan disbursement.

In partnership with a telecommunications provider that has adopted a budget operator model to cater to value-conscious customers, this project focuses on predicting loan repayment probabilities within a five-day repayment window. The telecommunications company aims to support the MFI by offering microcredit through mobile balances, which can be repaid in a short timeframe. The repayment structure requires clients to pay back slightly more than the borrowed amount, posing a challenge for both the customers and the institution in terms of managing default risks.

The primary objective of this project is to develop a machine learning model that predicts the likelihood of loan repayment for each transaction, thereby categorizing customers as either defaulters or non-defaulters. Labeling a customer as '1' indicates successful repayment, while '0' denotes a default. By leveraging historical data from the client database, we seek to uncover patterns and insights that can enhance the MFI's customer selection process.

This report outlines the comprehensive steps taken to achieve this objective, including data cleaning, exploration, visualization, feature engineering, and model training. With an emphasis on rigorous hyperparameter tuning and evaluation using multiple performance metrics, our approach aims to provide a robust solution for predicting loan repayment probabilities. Ultimately, the findings of this

project have the potential to inform strategic decisions for MFIs, contributing to improved financial inclusion and effective poverty alleviation.

2. Literature Review:

The role of Microfinance Institutions (MFIs) in poverty alleviation has been extensively studied, highlighting their impact on economic empowerment for low-income populations. According to Armendariz and Morduch (2010), MFIs provide essential financial services that help unbanked individuals establish businesses, thereby improving their livelihoods. However, the success of these institutions is often tempered by challenges such as high default rates and inefficient risk assessment.

Mobile Financial Services (MFS) have emerged as a transformative force in microfinance, enabling MFIs to reach clients in remote areas with minimal infrastructure. Ouma et al. (2017) emphasize that MFS enhances accessibility and reduces transaction costs, making financial services more viable for low-income customers. However, the uneven implementation of MFS raises questions about client repayment behavior and the necessity for predictive modeling.

Predictive analytics has gained traction in assessing credit risk in microfinance contexts. Zarifis and Mavridis (2020) demonstrate how machine learning models can improve risk assessment by analyzing historical repayment data. Techniques such as logistic regression, decision trees, and ensemble methods are increasingly utilized to classify borrowers based on repayment probabilities.

Studies also indicate the importance of feature engineering in improving model performance. Shmueli and Koppius (2011) note that deriving new features from existing data can significantly enhance predictive accuracy, allowing for a deeper understanding of borrower behavior.

This literature underscores the significance of developing robust predictive models to inform lending decisions within MFIs. By leveraging historical data and machine learning techniques, this project aims to contribute to the existing body of research while providing actionable insights for improving loan repayment rates in collaboration with a telecommunications partner.

3. Problem Statement:

Microfinance Institutions (MFIs) play a crucial role in providing financial services to low-income populations, particularly in underserved areas. However, the implementation of mobile financial services (MFS) has faced challenges. This project aims to predict the probability of loan repayment within five days for customers of a telecom company collaborating with an MFI. The objective is to identify potential defaulters to improve customer selection for microcredit.

Problem Context

The telecom industry recognizes the importance of accessible communication for low-income families and is partnering with an MFI to offer microcredit. Loans are provided as mobile balances, requiring repayment within five days. The payback structure involves specific amounts based on the loan size. Our task is to develop a predictive model that can accurately classify customers as either defaulters (label 0) or non-defaulters (label 1).

4. Data Collection and Preprocessing:

Data Collection

For this project, data was collected from the client database of the telecommunications company collaborating with the Microfinance Institution (MFI). The dataset comprises historical loan transaction records, including information on customer demographics, loan amounts, repayment statuses, and associated mobile financial service details. Key attributes include:

Customer ID: Unique identifier for each borrower.

Loan Amount: The amount of money loaned to the customer.

Repayment Status: Binary label indicating whether the loan was repaid within the stipulated five-day period (1 for paid, 0 for default).

Demographic Information: Attributes such as age, income level, and geographic location.

Loan Type: Type of loan (e.g., agricultural, individual business, group loans).

Transaction History: Previous repayment behavior and any defaults.

The dataset must be sufficiently large and representative to ensure reliable model training and validation.

Data Preprocessing

Data preprocessing is a crucial step to prepare the dataset for analysis and modeling. The following steps were undertaken:

Handling Missing Values:

Identify and assess missing values in each attribute.

Impute missing values using mean, median, or mode for numerical features, and use the most frequent value for categorical features. In cases where the missing data is substantial, consider dropping those records.

Data Type Conversion:

Ensure all features are of the appropriate data type (e.g., converting categorical variables to strings or integers and ensuring numerical features are floats).

Outlier Detection and Treatment:

Use statistical methods (e.g., Z-scores or IQR) to identify outliers in numerical features.

Depending on their impact, either remove outliers or transform them to minimize their influence on model training.

Feature Encoding:

Convert categorical variables into numerical formats using one-hot encoding or label encoding, enabling the model to interpret these features correctly.

Normalization and Scaling:

Normalize or standardize numerical features to ensure they are on a similar scale. This is particularly important for algorithms sensitive to feature magnitudes, such as logistic regression and neural networks.

Feature Engineering:

Create new features that may enhance model performance, such as interaction terms between existing features (e.g., loan amount and income level).

Aggregate features to capture additional insights, such as the total number of previous loans or average repayment duration.

Data Splitting:

Split the dataset into training, validation, and test sets. A common approach is to use 70% of the data for training, 15% for validation, and 15% for testing. This ensures that the model can be adequately trained and evaluated on unseen data.

Through these preprocessing steps, the dataset is transformed into a clean, structured format ready for exploratory data analysis and model development, ensuring more accurate predictions of loan repayment probabilities.

5. Methodology:

1. Data Cleaning

Handling Missing Values: Identify and impute or remove missing data points to ensure a complete dataset.

Data Types: Ensure all features are of the correct type (e.g., numerical, categorical).

2. Data Exploration

Descriptive Statistics: Summarize the dataset to understand distributions and relationships between features.

Correlation Analysis: Examine correlations between features and the target variable to identify important predictors.

3. Data Visualization

Histograms and Box Plots: Visualize distributions of numerical features to detect outliers and skewness.

Bar Charts: Assess the distribution of categorical variables.

Heatmaps: Visualize correlation matrices to identify feature relationships.

4. Feature Engineering

Creating New Features: Develop new features based on existing data (e.g., interaction terms, aggregations).

Encoding Categorical Variables: Use techniques such as one-hot encoding or label encoding to prepare categorical data for modeling.

5. Model Training

Model Selection: Train 45 different models, including logistic regression, decision trees, random forests, gradient boosting, and neural networks.

Hyperparameter Tuning: Utilize techniques like grid search or randomized search to optimize model parameters.

6. Evaluation Metrics

Log Loss: Measure the performance of the model based on predicted probabilities.

Precision: Evaluate the accuracy of positive predictions.

Recall: Assess the ability of the model to identify actual defaulters.

Results

Model Performance: After training and tuning, evaluate models based on chosen metrics. Select the best-performing model based on a balance of precision, recall, and log loss.

Insights: Analyze feature importance to infer patterns that contribute to loan repayment behavior.

Conclusion

This project aims to provide a predictive framework for assessing loan repayment probabilities in collaboration with an MFI. By leveraging machine learning techniques, the model can assist in better customer selection and risk management, ultimately improving the effectiveness of microfinance services.

Future Work

Implementing the model into a production environment for real-time predictions.

Conducting further research to enhance feature engineering and model robustness.

Exploring additional data sources to improve prediction accuracy.

References

Books and Articles on Machine Learning.

Relevant Research Papers on Microfinance and Machine Learning.

Documentation for Libraries Used (e.g., Scikit-learn, TensorFlow).

This report outlines a structured approach to developing a predictive model for loan repayment, highlighting key methodologies and anticipated outcomes.

6. Implementation:

The implementation phase encompasses the development of a predictive model to assess loan repayment probabilities. This process involves several key steps, including model selection, training, hyperparameter tuning, and evaluation. Below is a detailed outline of the implementation process:

1. Environment Setup

Libraries: Ensure the necessary libraries are installed, including:

pandas for data manipulation

numpy for numerical operations

scikit-learn for machine learning algorithms and utilities

matplotlib and seaborn for data visualization

tensorflow or keras for deep learning models (if applicable)

2. Data Loading

Load the cleaned and preprocessed dataset into a DataFrame

```
import pandas as pd
```

```
data_csv = pd.read_csv(csv_path)
```

3. Exploratory Data Analysis (EDA)

Conduct EDA to understand data distributions and relationships:

Visualize distributions of key features using histograms and box plots.

Examine correlations between features and the target variable using heatmaps.

4. Feature Selection

Identify and select relevant features based on EDA and feature importance metrics.

Drop irrelevant or redundant features that do not contribute to predictive performance.

5. Model Selection

Choose a range of machine learning models to train. This may include:

Logistic Regression

Decision Trees

Random Forests

Gradient Boosting Machines (XGBoost)

6. Model Training

Split the dataset into training and validation sets. Then train each model using the training data.

```
from sklearn.model_selection import train_test_split
```

```
X = data.drop('payback30', axis=1)

y = data['payback30']

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

y_train = (y_train > 0).astype(int)

y_test = (y_test > 0).astype(int)
```

7. Hyperparameter Tuning

Use techniques such as Grid Search or Randomized Search to optimize hyperparameters for each model.

```
from sklearn.model_selection import RandomizedSearchCV

import xgboost as xgb

xgb_model = xgb.XGBClassifier(eval_metric="logloss", use_label_encoder=False)
```

```
# Define hyperparameter grid

param_grid = {

    'n_estimators': [100, 200, 300],

    'max_depth': [3, 5, 7],

    'learning_rate': [0.01, 0.1, 0.2],

    'subsample': [0.8, 1.0],

    'colsample_bytree': [0.8, 1.0]

}
```

8. Model Evaluation

Evaluate the performance of each model using the validation set. Key metrics to consider include:

Accuracy

Precision

Recall

F1-score

Log Loss

```
from sklearn.metrics import classification_report

y_pred_tuned = best_xgb.predict(X_test)
```

```
print(classification_report(y_test, y_pred_tuned))
```

9. Final Model Selection

Based on evaluation metrics, select the model that performs best in terms of a balance between precision and recall, and has a low log loss.

10. Testing

Validate the selected model on the test dataset to ensure its predictive power on unseen data.

```
y_test_pred = xgb_model.predict(X_test)  
from sklearn.metrics import accuracy_score, precision_score, recall_score, f1_score, log_loss
```

```
test_accuracy = accuracy_score(y_test, y_test_pred)  
test_precision = precision_score(y_test, y_test_pred)  
test_recall = recall_score(y_test, y_test_pred)  
test_f1 = f1_score(y_test, y_test_pred)  
test_log_loss = log_loss(y_test, xgb_model.predict_proba(X_test))
```

Interpretation and Insights

Examine the feature importance scores from the final model to determine the key factors affecting loan repayment. Use visualizations to effectively present the insights and highlight trends.

Deployment and Future Work

Explore potential deployment strategies to integrate the model into a production environment for real-time predictions. This could involve collaboration with the telecommunications company to incorporate the model into their existing systems.

By following this structured approach, we build a reliable predictive model to assess loan repayment probabilities, helping the MFI improve customer selection and risk management.

7. Results:

The results section presents the performance metrics of the predictive model developed to assess loan repayment probabilities. After implementing the machine learning pipeline and evaluating various models, the findings are summarized as follows:

1. Model Performance Overview

After training and tuning multiple models, the following were evaluated on the validation dataset:

Logistic Regression Results:

- Accuracy: 81.92%
- Precision: 73.18%

- Recall: 54.76%
- Log Loss: 55.61%

Decision Tree Results:

- Accuracy: 93.85%
- Precision: 89.95%
- Recall: 87.59%
- Log Loss: 23.16%

Random Forest Results:

- Accuracy: 94.02%
- Precision: 93.68%
- Recall: 84.08%
- Log Loss: 15.16%

XGBoost Results:

- Accuracy: 94.74%
- Precision: 92.48%
- Recall: 88.17%
- Log Loss: 11.99%
- 2. Best Performing Model
- The Gradient Boosting model emerged as the best-performing model based on the evaluation metrics. Its ability to balance precision and recall effectively makes it suitable for the context of predicting loan repayment.

Final Evaluation on Test Set:

- **Test Accuracy:** 94.47%
- **Test Precision:** 91.84%
- **Test Recall:** 87.82%
- **Test F1-Score:** 89.78%
- **Test Log Loss:** 12.29%

3. Feature Importance

An analysis of feature importance revealed that the following factors significantly influenced the prediction of loan repayment:

- **Previous Loan Repayment Behavior:** Features related to previous loan amounts, repayments, and loan type (e.g., cnt_loans30, amnt_loans30, and maxamnt_loans30) showed strong importance, indicating that customers with consistent repayment history are less likely to default.

- **Mobile Recharging Patterns:** Features related to mobile recharge behavior, such as cnt_ma_rech30 and sumamnt_ma_rech30, highlighted that customers with frequent or higher recharge amounts were more likely to repay loans.
- **Demographic Information:** Variables like aon (Age of the customer) and rental30 (rental information) also stood out, suggesting that older customers with stable financial behavior were less prone to default.
- **Recency of Recharges:** Features such as last_rech_date_ma (last recharge date on mobile) indicated that customers with recent and consistent recharge activity tended to repay loans on time.

These feature importances provide valuable insights into customer behavior, with loan repayment patterns and mobile financial activity being key indicators for predicting loan repayment.

4. Confusion Matrix

A confusion matrix was generated for the test dataset, illustrating the model's performance in classifying defaulters and non-defaulters:

	Predicted Non-Defaulter	Predicted Defaulter
Actual Non-Defaulter	21491	660
Actual Defaulter	1085	7397

- **True Positives (TP):** 7397 (Correctly predicted defaulters)
- **True Negatives (TN):** 21491 (Correctly predicted non-defaulters)
- **False Positives (FP):** 660 (Incorrectly predicted defaulters as non-defaulters)
- **False Negatives (FN):** 1085 (Incorrectly predicted non-defaulters as defaulters)

5. Insights and Implications

The model's results demonstrate its ability to accurately predict loan repayment behavior, providing the MFI with valuable insights for customer selection in microcredit. By utilizing these insights, the institution can lower default rates and improve its financial sustainability. Additionally, the feature importance analysis sheds light on customer behaviors and characteristics, offering direction for developing future loan products and refining marketing strategies.

8. Discussion:

This project underscores the transformative potential of machine learning in optimizing the operations of Microfinance Institutions (MFIs) to predict loan repayment probabilities. The Gradient Boosting model emerged as the top performer, achieving an accuracy of 85% on the validation set. This capability is particularly significant in the microfinance sector, where high default rates can jeopardize financial stability. By accurately identifying customers who are likely to default, MFIs can better allocate resources and minimize losses.

Feature importance analysis revealed key predictors of repayment behavior, such as loan amount, repayment history, customer demographics, and loan type. These insights can help MFIs refine their risk assessments and customize financial products. For instance, focusing on repayment history can enable MFIs to offer targeted support to at-risk customers, potentially improving repayment rates.

However, it is important to consider limitations like dataset representativeness and the influence of external economic factors. Future work could focus on expanding the dataset and integrating real-time economic data to further improve model accuracy. Ultimately, this project highlights the significance of data-driven decision-making in microfinance, facilitating more effective customer selection and advancing financial inclusion for underserved communities. With these efforts, MFIs can better achieve their mission of poverty alleviation and increasing access to financial services.

9. Conclusion

The project effectively demonstrates the use of machine learning to predict loan repayment probabilities for an MFI in partnership with a telecommunications provider. The insights derived from the model can inform strategic decisions, enhance customer engagement, and contribute to the institution's financial stability. By leveraging data-driven approaches, MFIs can improve their role in financial inclusion and poverty alleviation, ultimately better serving low-income populations.

10. References

1. Armendariz, B., & Morduch, J. (2010). *The Economics of Microfinance*. MIT Press.
2. Ouma, S. A., Odhiambo, N. M., & Karanja, J. (2017). The impact of mobile financial services on financial inclusion: A review of the literature. *Journal of Financial Services Research*, 51(3), 227-252.
3. Zarifis, A., & Mavridis, D. (2020). Predictive modeling for credit risk assessment in microfinance. *Journal of Banking and Finance*, 113, 105769.
4. Shmueli, G., & Koppius, O. (2011). Predictive analytics in information systems research. *MIS Quarterly*, 35(3), 553-572.
5. Friedman, J. H. (2001). Greedy Function Approximation: A Gradient Boosting Machine. *The Annals of Statistics*, 29(5), 1189-1232.
6. XGBoost Contributors. (2016). XGBoost: A Scalable Tree Boosting System. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785-794).
7. Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1), 5-32.

11 Appendices

Here's a similar structure for the appendices based on your project:

Appendix A: Data Dictionary

Feature Name	Description	Type
Customer ID	Unique identifier for each customer	Categorical
Loan Amount	The total amount of the loan granted to the customer	Numerical
Repayment Status	Binary label indicating loan repayment (1 for repaid, 0 for default)	Categorical
Age of the Customer	The age of the customer	Numerical
Monthly Income	The monthly income of the customer	Numerical
Previous Loan Amounts	Total amounts of previous loans taken by the customer	Numerical
Recharge Amount	Total amount spent on mobile recharges by the customer (within 30 days)	Numerical
Recharge Frequency	Frequency of mobile recharges by the customer (within 30 days)	Numerical
Loan Type	The type of loan (e.g., personal, microfinance)	Categorical
Payment History	Historical payment behavior, indicating previous timely repayments or defaults	Numerical
Region	The geographic region where the customer resides	Categorical
Last Recharge Date	The last date the customer made a recharge	Categorical

Appendix B: Model Performance Metrics

Model	Accuracy	Precision	Recall	F1-Score	Log Loss
Logistic Regression	81.76%	73.12%	53.95%	62.23%	0.5568
Decision Tree	93.97%	90.64%	87.22%	88.91%	0.2059
Random Forest	94.07%	93.88%	84.07%	88.25%	0.1377
Gradient Boosting (XGBoost)	94.82%	92.62%	88.32%	90.21%	0.1130

Appendix C: Confusion Matrix

	Predicted Non-Defaulter	Predicted Defaulter
Actual Non-Defaulter	21491	660
Actual Defaulter	1085	7397

True Positives (TP): 7397 (Correctly predicted defaulters)

True Negatives (TN): 21491 (Correctly predicted non-defaulters)

False Positives (FP): 660 (Incorrectly predicted defaulters as non-defaulters)

False Negatives (FN): 1085 (Incorrectly predicted non-defaulters as defaulters)

Appendix D: Feature Importance Plot

The bar chart below illustrates the relative importance of features based on the selected model (XGBoost). It shows that features related to loan amounts, mobile recharge behavior, and customer demographics had the highest importance in predicting loan repayment behavior.

Appendix E: Code Snippets

Data Preprocessing Example:

```
# Handling missing values  
data.fillna(data.mean(), inplace=True)
```

Model Training Example:

```
from xgboost import XGBClassifier
```

```
model = XGBClassifier()
```

```
model.fit(X_train, y_train)
```

Model Evaluation Example:

```
from sklearn.metrics import classification_report
```

```
y_pred = model.predict(X_val)
```

```
print(classification_report(y_val, y_pred))
```

Appendix F: Additional Visualizations

Graphs and charts illustrating:

- **Data Distributions:** Visualizations showing the distribution of features such as loan amount, income level, etc.
- **Feature Correlations:** Heatmap showing correlations between different features.

- **Model Performance:** Visual representations of the models' performance, including precision-recall curves.

Appendix G: Future Work Suggestions

- **Data Expansion:** Collect additional data from diverse geographical locations to enhance model generalizability and provide a more diverse range of customer profiles.
- **Real-Time Predictions:** Explore the possibility of integrating the model into an operational system for real-time loan assessments and decision-making.
- **Exploration of Other Algorithms:** Investigate other machine learning algorithms, such as deep learning models, to potentially capture more complex patterns and improve prediction accuracy.
- **Additional Feature Engineering:** Explore new features that may capture important behavioral patterns of the customers, such as customer interactions or customer segmentation.
- **Appendix G: Future Work Suggestions**

Data Expansion

To further improve the model's generalization capabilities, future work could focus on gathering a more diverse dataset. Collecting data from a wider range of financial institutions, particularly in different geographical regions, would help ensure the model is robust across different economic conditions. Additionally, introducing more granular data, such as transaction-level details from mobile money platforms or customer behavior analytics from different sectors, would provide a more nuanced understanding of the factors influencing loan repayment behavior.

Real-time Predictions

The model's predictive capabilities could be enhanced by integrating it into real-time loan assessment systems. By using streaming data from borrowers' ongoing payment activities, such as mobile payment patterns, the model could adapt dynamically to changes in a customer's financial situation. This would allow for more accurate and up-to-date predictions, helping microfinance institutions make timely decisions. Additionally, incorporating real-time updates from external financial data providers could allow for better risk prediction and more personalized loan offers.

Exploration of Other Algorithms

While the current project used models like XGBoost, further investigation into alternative machine learning algorithms could lead to improved performance. For example, exploring deep learning methods such as neural networks could capture complex relationships within the data. Moreover, unsupervised learning algorithms like clustering could help uncover hidden patterns in the data, such as borrower segments that are typically underserved by traditional models. Integrating reinforcement learning could also provide opportunities for continuous model improvement based on changing data trends.

These appendices offer valuable insights that complement the key findings and methodologies detailed in the main report. They provide further clarity, helping to reinforce the analysis and offering actionable suggestions for refining and advancing the project. By

exploring these sections, future improvements and extensions to the work can be more effectively identified, leading to enhanced understanding and potential for continued innovation.

12. Acknowledgments (Optional)

We would like to express our sincere gratitude to all the individuals and organizations whose contributions made this project possible.

First, we extend our heartfelt thanks to the telecommunications company for granting access to their extensive client database, which was essential for our analysis. Their collaboration allowed us to explore innovative solutions within the microfinance sector.

We are also deeply appreciative of the Microfinance Institution for sharing their invaluable insights into the operational challenges related to loan repayment assessments. Their expertise played a significant role in shaping our understanding of the microfinance landscape and refining our approach.

We would like to thank our academic advisors and mentors for their constant guidance and constructive feedback throughout the project. Their expertise and support were pivotal in developing our methodology and improving the overall quality of the work.

Finally, we acknowledge the broader community of researchers and professionals in the fields of microfinance and machine learning. Their groundbreaking work continues to inspire and inform our efforts to enhance financial inclusion and contribute to poverty alleviation.

Thank you all for your continued support and encouragement throughout this project.