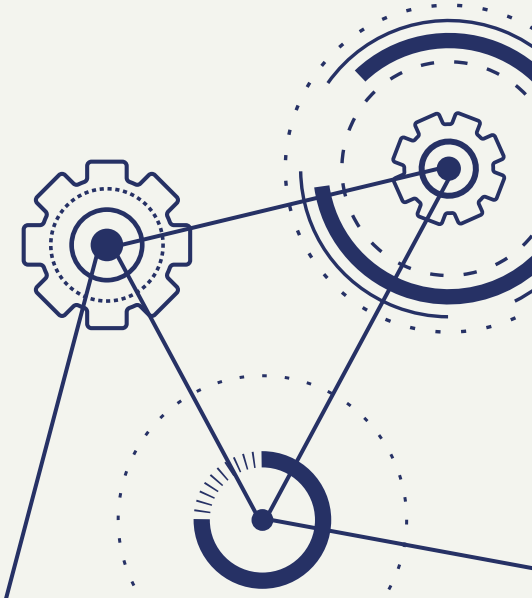
Moto Hope AI Lending System

Project Proposal

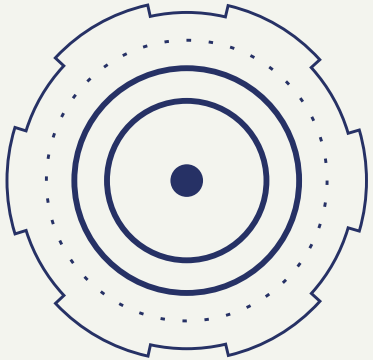
Risha Desai, Kate Picca, Kai Wong, Tess Wei



Current Lending Challenges

- Loan decisions currently depend on several sources of borrower information.
- Repayment risk includes both default and late repayment.
- Evaluating borrower risk manually can be a time-consuming and inconsistent.
- Moto Hope already has a useful historical borrower and repayment data that can support a more data driven approach.





What will the Model **Predict**?



Proposed primary target: Credit / repayment risk for a new loan

Primary candidate: Default risk

- Probability that a loan will automatically default

Late repayment risk:

- Probability that the borrower will pay significantly late

Expected Lateness:

- Days past due
- Repayment Completion
- How long repayment will take relative to the planned schedule



What data has been provided?



Client summary:

- M-Pesa activity
- CRB status
- Fuliza usage
- Occupation

Loan-Level Data:

- Requested loan
- Loan type
- Repayment terms
- Previous Moto Hope history
- Eventual loan outcome

Installment Schedule:

- Amount due
- Payment date
- Amount paid
- Days late
- Missed/partial payments

Fee Breakdown:

- Processing fees
- CRB fees
- Bank fees
- Insurance
- Other loan charges

Process analysis and **optimization**



Turning the Data Into One Model Dataset

Borrower Information

- M-PESA inflow
- CRB status
- Fuliza usage

Current Loan

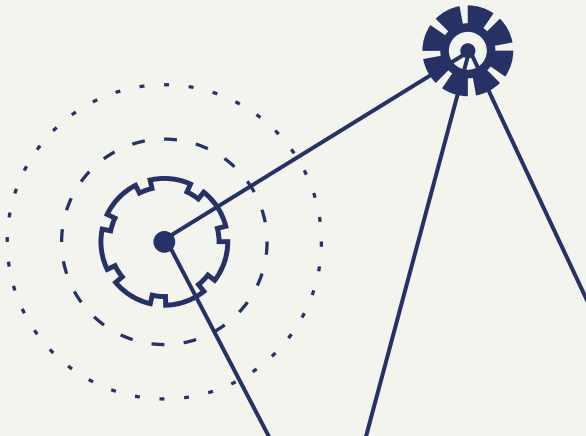
- Requested amount
- Product type
- Term

Previous History

- Previous loans
- Previously borrowed
- Previous repayment

Outcome

- Default



Data Cleaning & Preparation

Missing Data:

- Some Fuliza values
- Previous-loan fields are naturally blank for first-time borrowers

Identifier inconsistency:

- Example: MB2028 vs. MB/2028

Different data levels:

- Customer data = one row/customer
- Loan data = one row/loan
- Installment data = many rows/loan

Potential future-information leakage:

- Late payments, DPD, default date, and missed installments happen after the loan is approved
    - Therefore should not be used to predict approval-time risk

Proposed Model Pipeline

01	02	03	04	05
Moto Hope borrower data	Data cleaning & feature engineering	ML model	Risk prediction	Moto Hope lending workflow



Workflow

Model Details

Inputs

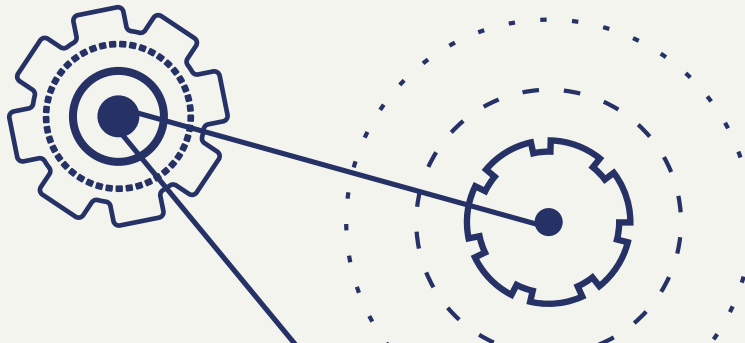
- M-Pesa
- CRB
- Fuliza
- Existing lenders
- Requested loan
- Previous Moto Hope history

Models

- Logistic Regression
- Random Forest
- Tree-Based Models

Output

- Predicted repayment risk
- Estimated default probability



AI Modeling Approaches

Logistic Regression

- Simple
- Fast
- Easy to interpret
- Strong baseline for binary default prediction

Random Forest

- Combines many decision trees
- More robust than one tree
- Allows feature-importance analysis

Decision / Tree-Based Models

- Captures nonlinear relationships
- Can show understandable decision patterns



How We Will Measure Performance

- Train/Test Split:
  - Model learns from previous complete data on clients and their loans.
  - Test with data the model hasn’t seen before, check predictions against known results.
  - Current plan is to ignore any in-progress loans.
    - Too time-consuming to wait for real results from clients.



Metrics



Accuracy

How many predictions were correct overall?

Precision

When the model flags a borrower as risky, how often is it correct?

Recall

Of borrowers who actually become risky/default, how many did we catch?

F1 Score

Balance between precision and recall.

Progress + Next Steps

Completed:

- ❑ Received and reviewed final Moto Hope dataset
- ❑ Identified relationships between the four tables
- ❑ Identified possible model inputs and outcomes
- ❑ Built one-row-per-loan working dataset
- ❑ Identified missing/inconsistent values
- ❑ Began feature engineering
- ❑ Established baseline modeling approach

Currently working on:

- ❑ Finalizing cleaned dataset
- ❑ Confirming prediction target
- ❑ Exploratory data analysis
- ❑ Finalizing feature set

Proposed Timeline:

Week 3-4: Finish data cleaning and preparation

Week 5: Data exploration ~ 1 week

Weeks 5-8: Build training + evaluation pipeline and initial models

Weeks 8-10: Model evaluation and comparison

Winter Quarter:

Refinement, deployment integration, testing

Questions for Moto Hope

- Which outcome would be most valuable for the model to predict: default, late repayment, or both?
- Should the model only help determine whether a loan should be approved, or should it eventually help recommend the amount/terms as well?
- At what stage in the loan process should the prediction be generated?
- Are the M-Pesa, CRB, Fuliza, and active-lender values historical values from each application or the customer's latest values?

Thank you!

