Tupande Data Scientist - R3 Exercise

October 2023 | Digital Team

ONE ACRE FUND

*** Disclaimer: The following exercise is based on real context and business considerations, and uses a project that was already run by an associate at this level. The quantitative and qualitative data included here was created for the purposes of this exercise, and is not related to real trial results or to the real decision that was made. No real client data has been shared.*

Time allowed: You have 3 days to complete this exercise

Context

One Acre Fund offers loans to hundreds of thousands of smallholder farmers. Farmers repay those loans over the course of the season in small installments. One Acre Fund wants to know at different points in the season which farmers might be at risk of defaulting so we can direct the field team to follow up with them to prevent that outcome from happening. For this exercise, you are expected to complete the three part deliverables highlighted below.

Deliverables

Part 1.1: Create a model that will predict repayment behavior of the customers and ensure we achieve a 98% repayment target.

1. We expect to see a report including description/ codes of your EDA steps, statistical tests, feature engineering, etc. We want to see you work in Python and report in Google Colab. We are not only evaluating the performance of your model. We are also evaluating your ability to communicate clearly and your ability to write good code so please be clear!
2. Predicted values for the holdout set

Task: Predict the cumulative_amount_paid (total paid on first of November 2023)

Data and variable description: The data we provide you is in a Google Sheet.

The training data set contains 36,434 instances. There are 10 input features. The dependent variable to be predicted is labeled cumulative_amount_paid. Below are selected variables description

- Client ID. A unique identifier of the client.
- Loan type. The type of loan (either Group Loan or Individual Loan).
- Region. A region is a topographical area in Kenya.
- Duka. A client is linked to a duka (store) where he collects his inputs.
- Nominal Contract Value. The total value of the inputs the client received.
- Deposit Amount. The total amount deposited at the contract start date.
- Contract start date. The start of the contract.
- Area: A region is divided into multiple area's (53 total).
- Sales territory: An area is divided into multiple sales territories.
- Cumulative amount paid start. The total amount that was paid before the first of September 2023.

Tupande Data Scientist - R3 Exercise

October 2023 | Digital Team

ONE ACRE FUND

What should be in the report?

1. Description of your methodology, solution path, and any assumptions you made.
2. List of techniques you attempted.
3. Results and evaluation of your models - Please specify the metrics you'd use to evaluate your model(s) and a brief explanation of your metric selection and how to interpret them.
4. Interpretation of your model including visual interpretations - What recommendations can you make to 1AF based on your analysis?
5. The code that was used to develop your model

Note on hold out set:

1. Create a separate dataset with predicted cumulative_amount_paid for the hold out set.

Part 1.2: If you had access to text-based field agents reports or voice transcriptions from farmer interactions, please illustrate how you could use AI to extract meaningful features that improve repayment predictions.

Part 2: Assume that you are presenting your findings to the Tupande Steering Committee and need to share recommendations to guide decision making on the lending strategy. Tupande has the below objectives for their lending work:

- Achieve a repayment rate of 98%
- Reduce the default rate to 1%
- Expand the farmers served by accurately targeting those most likely to repay their loan on time

Draft a **1-2 page** memo to the Steering Committee laying out the task, your findings and your recommendations. Please include at a minimum the following elements:

- Context for the task
- Well articulated findings from your model
- Clear recommendation(s) to the committee using the insights gained from your model
- Description of the next steps required to implement and achieve your recommendation i.e. what are the next technical, business and operational steps.

Note that you will be communicating to non-Tech stakeholders and therefore, you will be required to make your memo easily understandable to all teams across the organization.

Part 3: Assume that your recommendation(s) were approved by the Steering Committee, you implemented a 6 month trial of your recommendation(s) in one region, and the trial was successful in achieving the organization's objectives. You now need to scale it countrywide and are assigning an analyst to plan a 2-year countrywide scale up of the model. The analyst has the following objectives, resources, and timelines:

- Increasing the number of low-risk farmers served for the next two years while ensuring 98% repayment rate is achieved.
- In addition to your team's analyst, the program has a team on the ground who can help with this project. This team consists of one manager, a data analyst, several survey enumerators, and a government relations liaison.
- We expect scale up to start in about 4-6 months.

Tupande Data Scientist - R3 Exercise

October 2023 | Digital Team

Draft a **1-2 page** email to the analyst laying out the project and explaining clearly what the final deliverable should be. Please include at a minimum the following elements:

- Context for the project (Note, for this exercise, feel free to fill any information gaps as you see fit)
- Clear scope/ direction of what the analyst should do when planning the expansion
- Description of the final deliverable you expect to see, and what major topics the deliverable should cover. This could be a bulleted outline, a paper-and-pen sketch, etc. We have no predetermined format for this piece, so use whatever format you think best conveys the goals of the work so that the analyst is well-equipped to begin.

Remember, you are **not** being asked to craft the actual scale up plan. Rather, give clear, comprehensive instructions to an analyst so *they* can craft the expansion plan.