Technical Test, Data Science at Lendable

1. Introduction

At Lendable, customers take on loans with a fixed term, e.g. they might take out a £10k loan, with 24 month term.

An “early settlement” is when a customer pays off this loan before the initially-agreed term.

“Credit-bureaus” provide a “credit file” which summarises the financial profile of the applicant, and customers also enter data in an application form, as part of their application.

2. The task

We want you to build a model to predict the probability that a customer “early settles”. **Please predict a probability, rather than a class-label**

We’ve provided you with a sample dataset containing some dummy data representing underwritten loans, along with some credit features. These include a small selection of attributes from credit reports, as well as “tradeline” data, which is a more granular monthly snapshot of different financial ‘accounts’ they have had.

If you have promising ideas, but are running out of time, please add some comments at the end of your submission with what you would have done next, if allowed more time.

3. Data

Here is a **link to the data**.

- loans.csv: Contains some basic information about loans that we have underwritten, such as the loan amount, term, the interest rate, etc. One row, per booked loan. This also includes the early-settlement flag to use as a target **early_settled**
- attributes.csv: A selection of credit attributes for each customer. One row, per booked loan. Can be joined on loan using the loan_id column.
- tradeline.csv: Contains detailed information about the customer’s history with different accounts through time. E.g. if they had a credit card from 2023-10 to 2023-12, there would be one row for each month. Can be joined on ‘loans’ using loan_id, but be careful about ‘fan-out’ due to the multiple accounts and snapshot dates.

test-set/*: A folder that contains information about a different set of loans, along with the same data—but without the early-settlement flag. You should be able to score all of these loans using the model you build

4. What we’re looking for

We’re mostly interested in your thinking through the modelling process, so please include any thoughts or comments you have while building it.

Your model can be as simple or as complex as you like, but we’d like you to be able to explain your choices at each stage.

5 Submission format

Please submit your work as a Jupyter notebook or python script, along with any other code, scripts, or environment files you use.

The unseen test data (from test-set/*) should be scored using your model, and the results should be submitted as a CSV file called **[your_name]**_predictions.csv in the format

- Id
- early_settlement_probability