Transaction Categorization Technical Assessment

Welcome to the Rocket Money Data Science team technical assessment! The goal of this exercise is for you to become familiar with the type of data our team uses and highlight some of the challenges we currently face in classifying it. There is not one correct answer. Be creative! We very much evaluate ***approach and communication*** *over outputs.*

The Dataset

1902 transaction descriptions with eliminated PII. PII, account numbers, and order numbers have been replaced with {CUSTOMER NAME, NUMBER, HASH} tokens. Observations are representative of the the type of data we get from Plaid, a service we use to retrieve our users' transaction data from their banks. We strive to build useful abstraction layers (categories/subscriptions/services) on top of this raw data for the benefit of our members. For reference, we sampled this data from a set of internally assigned classes.

Task 1

Consider a handful of methodologies one might use to classify these transactions into "groups" that might be useful to an end user of Rocket Money. What "group" means is flexible - use your best judgment about what type of groups might be helpful for users to explore in our app. One simple place to start might be to classify transactions user deterministic tokens or partial string matches. **Resist the urge to simply label the data by hand as you work on this exercise — this is a toy dataset that is representative of a much larger *real* dataset, and we'd like to get a feel for how you'd work with the larger *real version*.**

1. Implement a handful of classification methods - how do you assess how well each method works? How might your assessment methodology differ if you had labeled data that assigned each transaction to a category? Implement your assessment methodology and show us how you assess the different classification methods. (Tip: At minimum, you probably want implement your methodology on a subset of the data, retaining a small subset for testing.)
2. From a qualitative perspective, how might these categories be used in-app to help our users understand their finances?
3. Choose your favorite or best performing classification method you have implemented. How might this method perform as we scale up the number of transactions (and the diversity of the groups)?

4. For the same method, in a production setting, how would you deal with changes to underlying transaction data over time? (Ex: a cool new competitor to Netflix arises and lots of people start paying for it).
5. Are there any product enhancements or features that might be useful in measuring classifier performance or to maintain a production version of your model over time?

Task 2 (bonus)

Propose (describe) a high level architecture for a production classification pipeline for the classifier you've created. Assume that we ingest transactions from Plaid into a production database that is used to surface transaction data to user devices. How might you design a backend system to interact with this data, assign categories, and surface those categories to user devices? Feel free to propose use of tools provided by any major cloud-based production environments. State any assumptions you might make about how our systems are structured. (This is open ended so don't worry, we aren't asking you to guess how our system works. We are seeking to understand how you think about ML system architecture.)

Limit yourself to roughly 3-4 hours for the exercise (we'll understand if you lose track of time 😉). We appreciate the time you'll invest in the problem set and in getting to know Rocket Money. Please return to us any code, notebooks and write-ups you feel represent your solution to the tasks above - whatever you feel best represents your data science skills. If you have any questions, please reach out (or simply clearly state your assumptions). We're excited to see what you come up with.

Good luck!

Rocket Money Data Science Team