

Canadian Banking Analytics Business Case

Overview

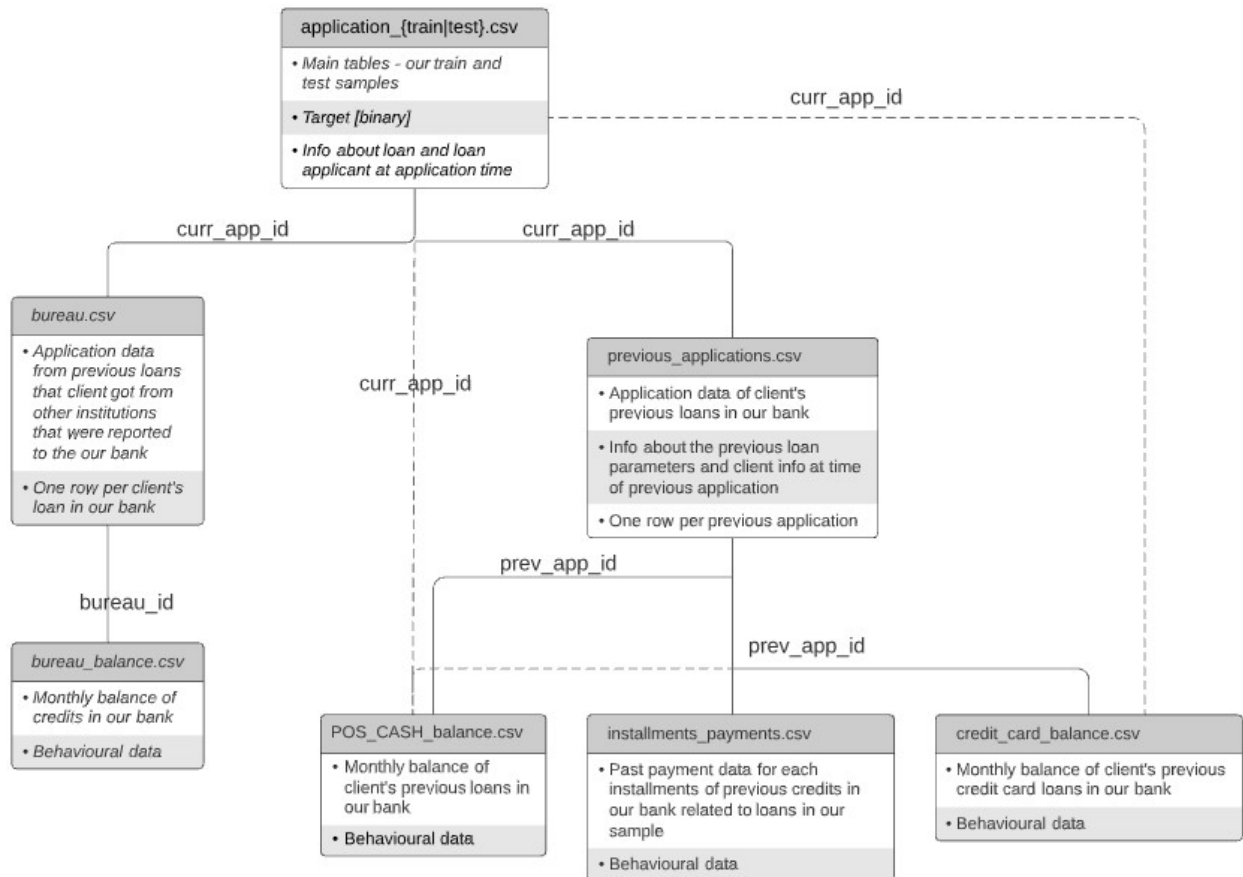
You are provided with customer credit history data and static information. The objective is to predict the likelihood of default for the current loan applications in our bank. Your solution should apply to customers both with and without sufficient credit history with our bank. You will submit your predictions for the loans in `application_test.csv`.

The datasets do not contain any real bank data. The description of the datasets are as follows.

- `application_{train|test}.csv`
 - This is the main table, broken into two files for Train (with TARGET) and Test (without TARGET).
 - Static data for all applications. One row represents one loan in our data sample.
- `bureau.csv`
 - All client's previous credits provided by other financial institutions that were reported to Credit Bureau (for clients who have a loan in our sample).
 - For every loan in our sample, there are as many rows as number of credits the client had in Credit Bureau before the application date.
- `bureau_balance.csv`
 - Monthly balances of previous credits in Credit Bureau.
 - This table has one row for each month of history of every previous credit reported to Credit Bureau – i.e the table has (#loans in sample * # of relative previous credits * # of months where we have some history observable for the previous credits) rows.
- `POS_CASH_balance.csv`
 - Monthly balance snapshots of previous POS (point of sales) and cash loans that the applicant had with the bank.
 - This table has one row for each month of history of every previous credit in our bank (consumer credit and cash loans) related to loans in our sample – i.e. the table has (#loans in sample * # of relative previous credits * # of months in which we have some history observable for the previous credits) rows.
- `credit_card_balance.csv`
 - Monthly balance snapshots of previous credit cards that the applicant has with our bank.
 - This table has one row for each month of history of every previous credit in our bank (consumer credit and cash loans) related to loans in our sample – i.e. the table has

(#loans in sample * # of relative previous credit cards * # of months where we have some history observable for the previous credit card) rows.

- previous_application.csv
 - All previous loan applications of clients who have loans in our sample.
 - There is one row for each previous application related to loans in our data sample.
- installments_payments.csv
 - Repayment history for the previously disbursed credits in our bank related to the loans in our sample.
 - There is a) one row for every payment that was made plus b) one row each for missed payment.
 - One row is equivalent to one payment of one installment OR one installment corresponding to one payment of one previous credit related to loans in our sample.



The columns description can be found in `columns_description.csv`.

Submission Files

1. Prediction: For each curr_app_id in application_test.csv, you will predict a probability for the TARGET variable. The file should contain a header and have the following format:

```
curr_app_id,TARGET
870323,0.1
883226,0.9
634374,0.2
etc.
```

Please name the file: '{your_name}_prediction.csv'.

2. Code: You can include comments on the approach taken.
3. Slides: You can assume this deck is designed for stakeholders. If you are selected for the next round, you can also leverage this deck to help us follow along in the interview. (In the deck, you can describe your methodology, additional steps you could have taken to increase your model(s) ability, business applications for this case, etc.)
Please name the deck: '{your_name}_presentation'

Evaluation

Submissions are evaluated on [area under the Precision-Recall curves](#) between the predicted probability and the observed target. It will be evaluated both on the entire test population and separately on the customers with no previous credit with the bank (i.e. with only Credit Bureau information), and those with previous credit.

Suggestions

While obtaining a good evaluation is important, we are more interested in the approach that you take to make your predictions. Writing detailed explanations along with your code will allow for a more fruitful discussion during the interview process.