

Credit Card Behavior Score

Introduction

Bank A issues Credit Cards to eligible customers. The Bank deploys advanced ML models and frameworks to decide on eligibility, limit, and interest rate assignment. The models and frameworks are optimized to manage early risk and ensure profitability.

The Bank has now decided to build a robust risk management framework for its existing Credit Card customers, irrespective of when they were acquired. To enable this, the Bank has decided to create a “Behavior Score”. A Behavior Score is a predictive model. It is developed on a base of customers whose Credit Cards are open and are not past due. The model predicts the probability of customers defaulting on the Credit Cards going forward.

This model will then be used for several portfolio risk management activities.

Problem statement

Your objective is to develop the Behavior Score for Bank A.

Datasets

(https://drive.google.com/drive/folders/1EgCaB1o9-2Pw9U4x_Z7ypC0f-Lai6fL-?usp=drive_link)

You have been provided with a random sample of 77,444 Credit Card details in “train_data.zip”, along with a flag (bad_flag) – henceforth known as “development data”. This is a historical snapshot of the Credit Card portfolio of Bank A. Credit Cards that have actually defaulted have bad_flag = 1. You have also been provided with several independent variables. These include:

- On us attributes like credit limit (variables with names starting with onus_attributes)
- Transaction level attributes like number of transactions / rupee value transactions on various kinds of merchants (variables with names starting with transaction_attribute)
- Bureau tradeline level attributes (like product holdings, historical delinquencies) – variables starting with bureau
- Bureau enquiry level attributes (like PL enquiries in the last 3 months etc) – variables starting with bureau_enquiry

You have also been provided with another random sample of 19,362 Credit Card details in “test_data.zip” with the same set of input variables, but without “bad_flag”. This will be referred to going forward as “validation data”.

Requirements

Using the data provided, you will have to come up with a way to predict the probability that a given Credit Card customer will default. You can use the development data for this purpose.

You are then required to use the same logic to predict the probability of all the Credit Cards which are a part of the validation data. Your submission should contain two columns – the Primary key from the validation data (account_number), and the predicted probability against that account.

You are also required to submit a detailed documentation of this exercise. A good document should contain details about your approach. In this section, you should include a write up on any algorithms that you use. You should then cover each of the steps that you have followed in as much detail as you can. You should then move on to any key insights or observations that you have come across in the data provided to you. Finally, you should write about what metrics you have used to measure the effectiveness of the approach that you have followed.

Evaluation

As detailed in the previous section, you are required to submit the Primary key and predicted probabilities of all the accounts provided to you in the validation data, as well as a documentation.

We will only evaluate submissions that are complete and pass sanity checks (probability values should be between 0 and 1 for example).

Submissions will be evaluated on the basis how close the predicted probabilities are to the actual outcome. We will also evaluate the documentation on the basis of its completeness and accuracy. Submissions must include interesting insights/observations on the data provided. (Note: screenshots, text, etc. accepted). Extra points will be granted to submissions that include more number of insights.