

Capstone Idea 1: Credit Risk Prediction

P. A. Ortiz Otalvaro

June 2019

I. Business Problem: Targeted Review of Internal Models (TRIM)

Pillar 1 funds requirements is the minimum amount of capital a bank must hold by law. Banks can choose to use internal models to set out their own funds requirements. In 2015, the European Central Bank (ECB) decided to start a project to check Pillar 1 models at all directly supervised banks that use them. In this way the ECB aims to assess whether the internal models currently used by banks comply with regulatory requirements, and whether their results are reliable and comparable.

One of TRIM's major goals is to increase consistency and stability when banks calculate risk-weighted assets through their own internal models. This is because the evaluation of risk assets is an important metric that contributes in determining a bank's Pillar 1 funds requirements as it weighs its assets according to their riskiness.

All banks under the supervision of the ECB are therefore following a transformation in their models and have a clear need for accurate models that best match their in-house framework and the complexity of their investments.

<https://www.bankingsupervision.europa.eu/about/ssmexplained/html/trim.en.html>

II. Business questions

Main questions/tasks:

1. Predict credit default: whether a loan will default. In other words: what is the credit risk (credit default) of a service given to a specific customer? (Credit risk or credit default indicates the probability of non-repayment of bank financial services that have been given to customers)
2. Predict: How much is the loss incurred when a loan defaults?

Other questions:

3. How can we measure the credit worthiness of the loan application over a time period? In other words, which are good techniques/algorithms/models to predict credit default ? (logistic regression, discriminant analysis methods, neural networks)
4. Which is the best model to predict credit default for a given institution? How can this model be selected? Which criteria can be used to select the best model for a given institution?
5. Which parameters/criteria are crucial in a model that predicts credit default? Which are the most important? Why?

III. Data: Links and descriptions

Option 1: Kaggle, lendingClub loan data

Description:

- These files contain complete loan data for all loans issued through the 2007-2018.
- The file is a matrix of about 890 thousand observations and 75 variables.
- A data dictionary is provided in a separate file.

Some features:

- current loan status (Current, Late, Fully Paid, etc.)
- latest payment information.
- credit scores
- number of finance inquiries
- address including zip codes, and state

... all other fields are described in the data dictionary (in the official website)

<https://www.kaggle.com/wordsforthewise/lending-club> <https://www.lendingclub.com/info/download-data.action>

Option 2: Kaggle, Loan Default Prediction, Imperial College London

Description:

- Set of financial transactions associated with individuals.
- The data has been standardized, de-trended, and anonymized.
- It contains over two hundred thousand observations and nearly 800 features.
- Each observation is independent from the previous.

NOTES:

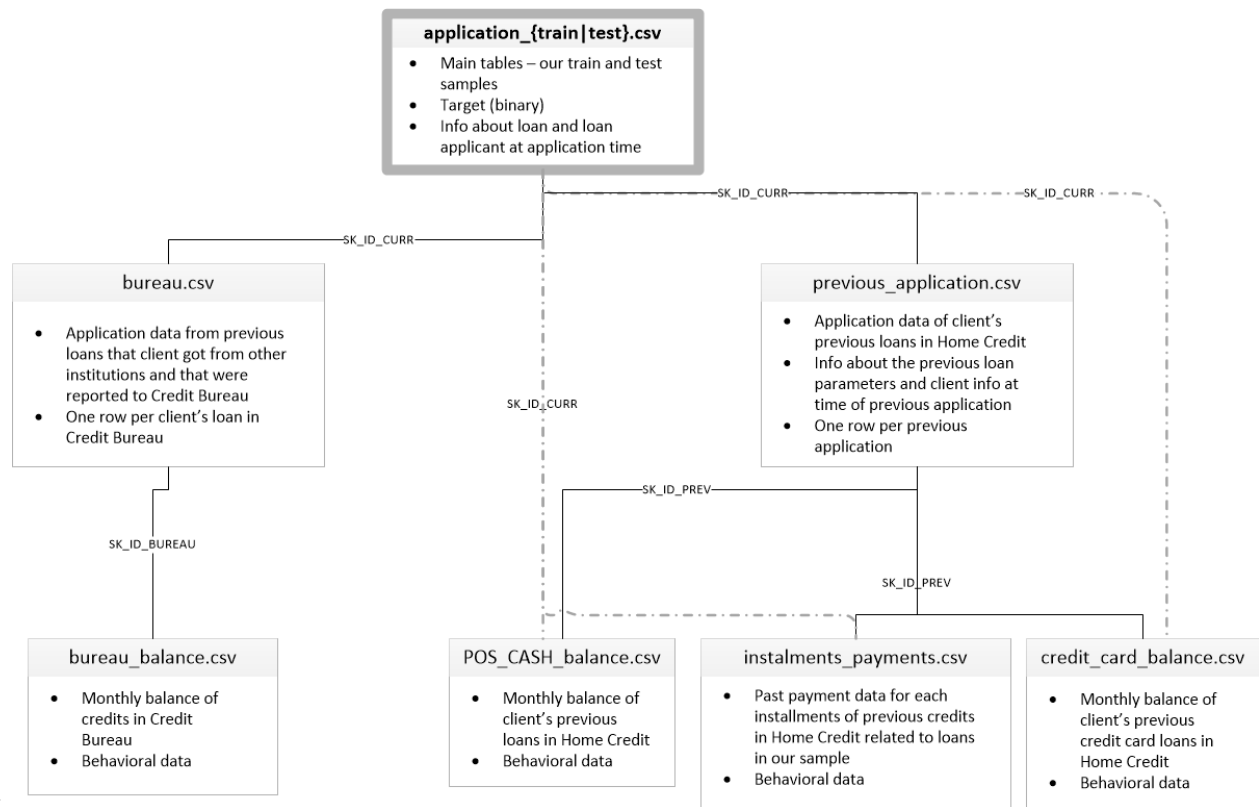
- **“Variable meanings are and will remain anonymized for this competition. Consider it part of the challenge.”**
- For each observation, it was recorded whether a default was triggered. In case of a default, the loss was measured. This quantity lies between 0 and 100. It has been normalised, considering that the notional of each transaction at inception is 100. For example, a loss of 60 means that only 40 is reimbursed. If the loan did not default, the loss was 0.
- Some variables may be categorical (e.g. f776 and f777).
- Observations are listed in order from old to new in the training set. In the test set they are in random order.

<https://www.kaggle.com/c/loan-default-prediction/data> <https://www.kaggle.com/c/loan-default-prediction/discussion/6978>

Option 3: Kagge data, Home Credit default risk

10 csv files with information related to customer's past financial data. Detailed description can be found on Kaggle's website.

<https://www.kaggle.com/c/home-credit-default-risk/data>



In a diagram:

Option 4: Kaggle, German credit risk

Features:

- Age (numeric)
- Sex (text: male, female) Job (numeric: 0 - unskilled and non-resident, 1 - unskilled and resident, 2 - skilled, 3 - highly skilled)
- Housing (text: own, rent, or free)
- Saving accounts (text - little, moderate, quite rich, rich)
- Checking account (numeric, in DM - Deutsch Mark)
- Credit amount (numeric, in DM)
- Duration (numeric, in month)
- Purpose (text: car, furniture/equipment, radio/TV, domestic appliances, repairs, education, business, vacation/others)

<https://www.kaggle.com/uciml/german-credit>