

Topics in Quantitative Finance - ML: Project Proposal

Michal Topinka

March 27, 2017

Loan Default Prediction

In this project, I will use data from a three-year-old competition on *kaggle.com*. (<https://www.kaggle.com/c/loan-default-prediction>) The dataset contains a set of standardized, de-trended and anonymized financial transactions with over two hundred thousand observations and 778 features. For each observation, it was recorded whether a default was triggered. In case of a default, the loss was measured. The task of the competition was to predict whether a loan will default, as well as the loss incurred if it does.

In my project, I will try to find identify the most important features and predict whether a particular loan will default or not. I will use methods covered in this course, such as logistic regression, support vector machine, decision trees or k-neares neighbors.

The full dataset is available on *kaggle.com* (login required for download). <https://www.kaggle.com/c/loan-default-prediction/data>