**Quantitative analysis of credit**

The Great Recession was due in a sizeable degree to the process of extending credit to people who defaulted on their loans (typically mortgages for their houses) as they were not able to repay them. Combined with decreasing real estate prices, many of the institutions that extended the loans ended up owning property that had decreased in value, and therefore lost significant amount of money.

In the spreadsheet credit3.xls under the tab “Data”, you will find Data pertaining to 1000 personal loan accounts at a bank. The tab “Data Dictionary” contains a description of what the various variables mean.

When a new applicant applies for credit, as a part of the application, the company collects information which is available in the form of Variables 2 to 21. The company then decides an amount to be credited (the variable CREDIT\_EXTENDED.) For these 1000 accounts, we also have information on how profitable each account turned out to be (variable NPV). A negative value indicates a net loss and this typically happens when the debtor defaults on his/her payments or if the loan is renegotiated to prevent default.

The goal in this case is to investigate how one can use this data to better manage the bank's credit extension program. **Specifically, our goal is to develop a classification regression model to classify a new account as “profitable” or “not profitable”.**

IMPORTANT: In this assignment it will be important for categorical predictors to be numerical dummy variables for KNN (so that distances can be computed). For Naïve Bayes, you will need categorical predictors to be factors (otherwise the algorithm will fit a Normal distribution to the data for the conditional probabilities).

Creating dummy variables using ifelse(…) can be tedious. You may use the fastDummies package to create dummy variables.