

Assignment

The objective is to use the loan data to build a model that can predict the loan status (whether or not a loan will be fully paid or charged off) and interpret the results to find the factors which affect the result. The document you received earlier, “An introduction to credit scoring models”, provides some of the basics of scoring models, however all other sources of information may be used.

At the end of the case you are asked to share your results with us and elaborate on them. At the start and midway the case you have the opportunity to ask questions according to the following planning.

13:00	13:15	Introduction & plan of approach
13:15	13:45	Development time
13:45	14:00	Q&A session
14:00	15:30	Development time
15:30	16:00	Final presentation
16.15	16.45	Market sizing

Description of data set

This data set contains information on loan payments, demographic factors, credit history, history of the borrowers, and many more variables from loan data provided by one of our clients in the period from 2007Q1 to 2020Q1. The dataset contains almost sixty thousand observations; each record with a different loan, representing a unique borrower. Note that all variables contain historical information, including the target variable (“loan_status”). This target variable indicates whether the loan of a borrower is charged off (default) or fully paid.

Data description

The dataset contains 25 variables. The default variable (“loan_status”) is the target variable. All other variables may be used as discriminatory variables. The description of the variables and an additional dictionary is given below.

#	Variable	Description
1	loan_amnt	The listed amount of the loan applied for by the borrower. Charged off (default) or Fully paid (no default)
2	term	The number of payments on the loan
3	int_rate	Interest rate on the loan
4	installment	The monthly payment owed by the borrower
6	sub_grade	Assigned loan subgrade
7	emp_length	Employment length in years. Possible values are between 0 and 10 where 0 means less than one year and 10 means ten or more years.
8	home_ownership	The home ownership status provided by the borrower. Our values are: RENT, OWN, MORTGAGE, OTHER
9	annual_inc	The self-reported annual income provided by the borrower.

10	verification_status	Indicates if income was verified or not verified, or if the source of the income was verified
11	Issue_d	The month in which the loan was funded to borrower
12	purpose	A category provided by the borrower for the loan request.
13	addr_state	The state provided by the borrower in the loan application
14	dti	A ratio calculated using the borrower's total monthly debt payments on the total debt (including other debts) obligations, divided by the borrower's self-reported monthly income.
15	fico_range_low	The lower boundary range the borrower's FICO at loan origination belongs to.
16	fico_range_high	The upper boundary range the borrower's FICO at loan origination belongs to.
17	open_acc	The number of open credit lines in the borrower's credit file.
18	pub_rec	Number of derogatory public records
19	revol_bal	Total credit revolving balance
20	revol_util	Revolving line utilization rate, or the amount of credit the borrower is using relative to all available revolving credit.
21	mort_acc	Number of mortgage accounts
22	pub_rec_bankruptcies	Number of public record bankruptcies
23	age	The age of the borrower at the time of application
24	pay_status	Last known repayment status (-2 and -1=pay duly, 1=payment delay for one month, 2=payment delay for two months, ... 8=payment delay for eight months, 9=payment delay for nine months and above)
25	loan_status	Current status of the loan (Charged off = DEFAULT (1) or Fully paid = No default (0))

Additional dictionary

#	Dictionary variable	Description
1	Credit Line	The amount of money that can be charged to a credit card account
2	FICO score	One of the most well-known types of credit score. Range from 660 to 850 in the dataset
3	Derogatory public record	A derogatory public record is negative information on a credit report that is of a serious nature and has become a matter of public record.
4	Credit revolving balance	the portion of credit card spending that goes unpaid at the end of a billing cycle
5	Revolving utilization rate	It measures how much of your credit limits are in use on each of your credit card accounts. Displayed as percentage.
6	Installment account	Like a auto loan. It comes in the form of a loan that you pay back in steady payments every month. The amount of the loan is determined at the time you're approved.
7	Revolving account	Like a credit card. You'll have a limit on how much you're able to borrow, but the amount you use within that limit is up to you.
8	Credit inquiry	A credit inquiry is a request by an institution for credit report information from a credit reporting agency.