

Predicting Loan Defaults with Logistic Regression

DS705

The Data

This dataset includes 30 variables for 50,000 loans. You can grab the csv file at

<https://datascienceuwl.github.io/Project2018/loans50k.csv> (<https://datascienceuwl.github.io/Project2018/loans50k.csv>)

Details of the variables are given below.

The variables in the data are fully described in a table below. The response variable will be built from the loan status variable **status**. The total amount repaid to the bank, **totalPaid**, cannot be used as a predictor variable because it is information that cannot be known before the loan is issue so be sure not to include it as a predictor in your models. All other variables can be used as predictors, but some are not useful.

variable	description
amount	loan amount in dollars
term	loan term is 36 or 60 months
rate	interest rate as a decimal
payment	monthly payment amount
grade	grade of loan: A is least risk, G is most risk
employment	job title of applicant
length	time continuously employed
home	home ownership: rent, own, mortgage
income	annual income in dollars
verified	verification status of annual income
status	loan status: DEFAULT, CURRENT, CHARGED OFF, etc.
reason	applicants purpose for the loan
state	two letter state code of applicant
debtIncRat	ratio monthly non-mortgage debt payment to monthly income
delinq2yr	number of 30+ day late payments in last two years
inq6mth	number of credit checks in the past 6 months
openAcc	number of open credit lines
pubRec	number of derogatory public records including bankruptcy filings, tax liens, etc.
revolRatio	proportion of revolving credit in use
totalAcc	total number of credit lines in file, includes both open and closed accounts

variable	description
totalPaid	total amount repaid to bank (THIS IS NOT A PREDICTOR SINCE IT CAN ONLY BE DETERMINED AFTER A LOAN IS ISSUED)
totalBal	total current balance of all credit accounts
totalRevLim	sum of credit limits from all credit lines
accOpen24	how many accounts were opened in the past 24 months
avgBal	average balance per account
bcOpen	total unused credit on credit cards
bcRatio	ratio of total credit card balance to total credit card limits
totalLim	total credit limits
totalRevBal	total credit balance except mortgages
totalBcLim	total credit limits of credit cards
totalIllLim	total of credit limits for installment accounts