

# Project 3 Classification Proposal: Predicting Consumer Loan Defaults

Linda Ju

## Domain

### *Motivation*

Peer-to-peer lending ('P2P') is a relatively new form of financing in which individuals loan money directly to other individuals. Since its inception in 2005, the P2P industry has grown to an over \$3 billion industry. The most popular P2P platforms include Prosper, Lending Club, and Upstart with Prosper being the first ever peer-to-peer lending marketplace in the U.S. I have an account with Prosper and I'd like to determine which loans are credit-worthy, as in, will not default.

### *How does P2P work?*

To take out a loan, a borrower typically fills out an online application describing how much they want to borrow (up to \$40,000 on Prosper), for how long, (3 or 5 years), and for what purpose. Additionally, the lending platform will often perform a credit check. During the loan funding stage, investors review the listing information such as a borrower's interest rate, amount, FICO range, number of past delinquencies, debt-to-income ratio, employment status, and income and decide which loans they want to fund (or invest in). Investments can be made in small increments, as low as \$25, so investors can diversify their investments and put their money in many loan listings. A loan is considered funded once investors fund "enough" of the requested amount, as defined by the lending platform (70% for Prosper).

When a loan is fully funded, the borrower receives their funds and begin the loan repayment stage. The borrower makes fixed regular payments of principal and interest which is split proportionally among the investors and a small percentage in fees. If a borrower pays in full and on the schedule as given from the beginning, the investor will realize their expected return on investment. If, however, a borrower is late on payments or, worse, defaults on a payment, the investor will realize a lower, zero, or negative return.

## Project Details

The goal of my project is to help me choose a loan to invest in by predicting its eventual status, fully paid, defaulted, or charged off. I will use classification techniques based on listing information loan amount, interest rate, term, annual income, and FICO score.

### *Data*

The initial dataset I will use from Lending Club's database from 2007 to 2018 is described below. As I work on my project, the actual final dataset may differ as I discard less relevant features and add more relevant features.

Variable	Description
Loan Status	Status of the loan (Fully Paid, Charged Off, or Default)

Loan amount	The listed amount of the loan applied for by the borrower. If at some point in time, the credit department reduces the loan amount, then it will be reflected in this value
Term	The number of payments on the loan. Values are in months and can be either 36 or 60
Interest Rate	Interest Rate on the loan
Grade	Lending Club assigned loan grade
Employment 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
Home Ownership	The home ownership status provided by the borrower during registration or obtained from the credit report (rent, own, mortgage, or other)
Annual Income	The self-reported annual income provided by the borrower during registration.
Verification Status	Indicates if income was verified by LC, not verified, or if the income source was verified
Purpose	A category provided by the borrower for the loan request
State	The state provided by the borrower in the loan application
Delinquency in Last 2 Years	The number of 30+ days past-due incidences of delinquency in the borrower's credit file for the past 2 years
Earliest Credit Line	The month the borrower's earliest reported credit line was opened
FICO range	The upper and lower boundary range the borrower's FICO at loan origination belongs to
Months Since Last Delinquency	The number of months since the borrower's last delinquency
Revolving balance	Total credit revolving balance

### *Considerations*

- Relatively new investment product; haven't been through many cycles
  - May not be reflective of future performance in a different economic cycle