Data-Driven Investment Strategies for Peer-to-Peer Lending
Student Handout

In this case, we follow Jasmin Gonzales, a young professional looking to diversify her investment portfolio. Jasmin graduated from a Masters in Data Science program, and after four successful years as a product manager in a tech company, she had managed to save a sizable amount of money. She now wants to start diversifying her savings portfolio. So far, she has focused on traditional investments (stocks, bonds, etc.) and she now wants to look further afield. One asset class she is particularly interested in is peer-to-peer loans issued on online platforms. The high returns advertised by these platforms seem to be an attractive value proposition, and Jasmin is especially excited by the large amount of data these platforms make publicly available. With her data science background, she is hoping to use machine learning tools on this data to come up with lucrative investment strategies.

1. Background on Peer-to-Peer Lending

Peer-to-peer lending refers to the practice of lending money to individuals (or small businesses) via online services that match anonymous lenders with borrowers. Lenders can typically earn higher returns relative to savings and investment products offered by banking institutions. However, there is of course the risk that the borrower defaults on his or her loan. Interest rates are usually set by an intermediary platform on the basis of analyzing the borrower's credit (using features such as grade, employment status, annual income, debt-to-income ratio, number of open credit lines). The intermediary platform generates revenue by collecting a one-time fee on funded loans (from borrowers) and by charging a loan servicing fee to investors.

The peer-to-peer lending industry in the U.S. started in February 2006 with the launch of Prosper, followed by LendingClub. In December 2015, LendingClub reported that $15.98 billion in loans had been originated through its platform. With very high year-over-year growth, peer-to-peer lending has been one of the fastest growing investments. According to InvestmentZen, as of May

2017, the interest rates range from 6.7%-22.8%, depending on the loan term and the rating of the borrower, and default rates vary between 1.3% and 10.6%.

LendingClub issues loans between $1,000 to $40,000 for a duration of either 36 months or 60 months. As mentioned, the interest rates for borrowers are determined based on personal information such as credit score and annual income. LendingClub categorizes its loans using a grading scheme (grades A, B, C, D, E, and F, where grade A corresponds to the loans judged to be “safest" by LendingClub).

One of the interesting features of the peer-to-peer lending market is the richness of the historical data available. The two largest U.S. platforms (LendingClub and Prosper) have chosen to give free access to their data to potential investors. This then raises a whole host of questions for investors like Jasmin: Is this data valuable when selecting loans to invest in? How could an investor use this data to develop machine learning tools to guide investment decisions?

The goal of this case study is to provide answers to the questions above. In particular, we investigate how data analytics and machine learning tools can be used in the context of peer-to-peer lending investments. We will use the historical data from loans that were issued on LendingClub between January 2014 and December 2015.

2. Datasets and Descriptive Statistics

The data records 38 features including the following, for each loan: Interest rate, Loan amount, Monthly installment amount, Loan status (e.g., fully-paid, charged-off). And several additional attributes related to the borrower such as type of home ownership, annual Income etc. If the payment is delayed by more than 121 days, the loan is considered as being in Default. If LendingClub has decided that the loan will not be paid off then it is given the status of Charged-Off. Five months after the term of each loan has ended, every loan end in one of two LendingClub states - fully paid or charged-off. We can refer to a loan that has reached one of these statuses as expired. Current refers to a loan that is still being paid back in a timely manner.

3. Investment Strategies and Portfolio Construction

Making predictions and constructing a portfolio in the context of online peer-to-peer lending can be challenging. The volume of data available provides an opportunity to develop sophisticated data- driven methods to increase Jasmin’s portfolio performance.