HARVARD | BUSINESS | SCHOOL

9-119-020

REV: SEPTEMBER 24, 2018

SRIKANT M. DATAR

CAITLIN N. BOWLER

LendingClub (A): Data Analytic Thinking (Abridged)

Emily Figel had just signed with a small wealth management firm and looked forward to beginning work in Chicago in the fall. As an MBA student she had become fascinated by alternative banking models and she accepted this offer for the opportunity to research the current state of the peer-to-peer lending platform LendingClub. Her first project would be to evaluate whether it was a reasonable vehicle for investment on behalf of some of the firm's more aggressive clients.

LendingClub and other peer-to-peer lending platforms like Zopa (U.K.) and Prosper (U.S.) emerged in 2005 and 2006 to connect individual borrowers to individual investor-lenders through online platforms. These companies used online platforms to bypass the traditional banking infrastructure and higher operations costs to offer borrowers lower interest rates and investors higher returns. The platforms collected a variety of personal and financial information from borrowers and assigned grades and sub-grades to each loan application for potential investor-lenders to review. However, they played no role in granting decisions to fund a loan application. That decision was left entirely to the qualified individual investor-lenders.[1]

Transparency was key to developing the trust investors demanded in exchange for investment. LendingClub anonymized its historical data—not just new lending opportunities—and made it available for download from its website. Interested parties could download data on loans granted from 2007 through Q1 2018.[a] These tables contained many features per application and tens of thousands of applicant entries.[b] One of Figel's former colleagues had become an investor-lender and used LendingClub data to build models to predict (*classify*) whether particular loan applicants would default or repay their loans. This prediction became one of several pieces of data he then used to evaluate a loan's risk and decide whether to fund it himself.

Figel still had a few months before she began her new position and was eager to explore the loan data on the LendingClub platform more carefully. She wanted to learn the basics of predictive modeling and get a nuanced feel for the opportunities and risks of investing on the platform before advising her employer on much larger investments.

---

[a] LendingClub regularly released new data from the previous quarter.

[b] "These files contain complete loan data for all loans issued through the time period stated, including the current loan status (Current, Late, Fully Paid, etc.) and latest payment information." Source: LendingClub, "LendingClub Statistics," https://www.lendingclub.com/info/download-data.action, accessed April 2018.

Professor Srikant M. Datar and Research Associate Caitlin N. Bowler prepared the original verson of this case, "Lending Club: Predicting Default," HBS No. 518-010, which is being replaced by this version prepared by the same authors with the assistance of Rashmi Banthia. This case was developed from published sources. Funding for the development of this case was provided by Harvard Business School and not by the company. Emily Figel is a fictional character. HBS cases are developed solely as the basis for class discussion. Cases are not intended to serve as endorsements, sources of primary data, or illustrations of effective or ineffective management.

LendingClub Basics

Figel had first become aware of the LendingClub platform in 2014 when it became the largest technology initial public offering (IPO) of that year. In the December offering, it raised nearly $900 million and was viewed as an emblem of the future of "fintech."[2] Its peer-to-peer model and customer orientation, for both lender and borrower, positioned it as the forward looking antithesis of the large traditional banks that had been hammered from all sides since 2008. These qualities had attracted investors to the IPO and also attracted Figel.

LendingClub had come a long way since its comparatively humble beginning in 2007. In 2012, after five years of operation, it reached a milestone of issuing $1 billion in loans (cumulative). At the time of the IPO it was issuing $1 billion in loans per quarter and $1.5 billion per quarter in 2016.[3]

The company facilitated these loans through three different vehicles: notes, certificates, and whole loan sales.[4] LendingClub offered these three- or five-year unsecured personal loans to borrowers with a FICO score of at least 660 who met other credit criteria through its online *Standard Program Loans* division. Applicants stated the amounts they wished to borrow and submitted information about their financial situation, employment, and intended purposes of the loans. The most common purposes were debt consolidation, credit card debt refinance, and home renovation. Investors could browse notes through the online platform and decide which ones to fund. When an investor funded a loan she really purchased a "note," which could be traded on the platform.[5] (See **Exhibit 1** for data included in a loan application as shown on the LendingClub investor dashboard.)

After the IPO, LendingClub enlarged its market by opening the platform to larger, institutional style investors. As a public company facing quarterly earnings targets, it was a predictable and reasonable strategic shift. In its 2015 10-K filing with the SEC, the company stated:

> We do not assume credit risk or use our own capital to invest in loans facilitated by our marketplace, except in limited circumstances and in amounts that are not material. The capital to invest in the loans facilitated through our marketplace comes directly from a wide range of investors, including retail investors, high-net-worth individuals and family offices, banks and finance companies, insurance companies, hedge funds, foundations, pension plans and university endowments, and through a variety of channels, such as borrower payment dependent investment securities and whole loan purchases.[6]

As of 2016, only 15% of loans came from small individual investors. The interest from investment houses and other sophisticated investors signaled an increase in credibility that had eluded the industry in its early days.[7]

The lending process

The phrase "peer-to-peer" connoted a simple and direct process that belied the complex reality of this alternate model for lending. The LendingClub model contained two parts: (1) maintenance of the online platform that allowed borrowers and lenders from the LendingClub marketplace to connect, and (2) facilitation of the process between the borrower, the investor (lender), and partner banks that actually originated the personal and business loans.

Part 1 was a standard platform operation. Part 2 had four distinct moments, labeled *a* through *d* in **Exhibit 2**. LendingClub (a) directed a *partner bank* to (b) initiate a loan to the *borrower*, which the partner bank immediately (c) sold to LendingClub, which the company paid for using funds an individual *investor* (d) paid to LendingClub for the note.[8]

Mechanics of investing

LendingClub assigned grades to every loan application; these grades ranged from A to G. LendingClub broke the grades down further into sub-grades of 1-5. LendingClub assigned different interest rates to each sub-grade. (See **Exhibit 3** for table of interest rates by sub-grade.) However, the *features*, grade and sub-grade, were just two of over 100 the company made available to investors for each loan application.

To invest in LendingClub notes through the online platform, Figel signed up for an account. When she was ready, she would fund the account and select notes to purchase. (See **Exhibit 4** for a view of the LendingClub investor dashboard.) The lowest fraction of a note in which Figel could invest was $25. Many investors liked the fractional model because it allowed them to build a portfolio that specifically matched their own risk profiles. Figel's goal was to use LendingClub's data to build a model that would allow her to create a portfolio of notes that produced returns commensurate with the risks she was taking. In 2017 average annual returns for equities (S&P 500), corporate bonds, and treasuries were 10.44%,[c,9] 5.98%,[d,10] and 2.80%,[e,11] respectively.

Fees, interest, and payments: A simple example

To understand the business model and lending process Figel considered a simple example: a borrower who applied for a loan for $10,000 through the LendingClub platform. To begin, LendingClub would evaluate the borrower and assign him a grade/sub-grade of, say A1. LendingClub offered the borrower an interest rate, 6.03%, according to its grade-based schedule. (See **Exhibit 3**.) If the borrower accepted the offer terms, LendingClub would list the loan on the platform where an investor could see the loan and offer to fund it at that rate for a 36-month term.

To facilitate the lending process, LendingClub directed the originating bank to lend to the borrower. LendingClub would then immediately purchase the note from the bank using funds provided by the investor. LendingClub charged the borrower a 3.5% origination fee ($350), and disbursed $9,650 to the borrower. Going forward, the borrower paid 6.03% on the value of the full $10,000 value of the loan amount. The borrower's monthly loan payment to the investor would be $304.36, of which LendingClub kept $3.046 as the 1% servicing fee for "maintaining investor accounts, collecting and processing principal and interest payments net of fees to investors." By the end of payment 36, the borrower would have repaid $10,956.79, of which $956.79 was interest.[f,12] If a borrower went into default, LendingClub charged investors collection fees for their efforts to bring the loan back to current status.

Like a traditional housing mortgage, the amount of a borrower's monthly payment that went to interest and principal changed over time. In this example, 16.5% of the first payment would go toward interest ($50.22), while only 0.50% would go toward interest in the last payment ($1.52). (See **Exhibit 5**

[c] This annualized return is for the period of January 1, 2014 to December 31, 2017, which is equal to the 36-month term of LendingClub's shortest loan. It is adjusted for inflation, but does not include dividends.

[d] This annualized return is for the Bank of America Merrill Lynch U.S. Corporate Master Total Return Index.

[e] Ten year T-bond.

[f] The borrower's stated interest rate was 6.03%. However, because LendingClub disbursed the value of the loan *less* the origination fee, the *effective interest rate* (i.e. annual percentage rate or what the loan actually cost the borrower) was higher. A "back of the envelope" calculation of the EIR was simple. First, Figel used the stated interest rate to calculate the interest paid in year 1 ($10,000 × .0603 = $603). Second, she calculated the funds actually disbursed ($9,650). Third, she divided the interest paid in year 1 by the amount disbursed ($603/$9,650 = .0624). So the effective interest rate on the $10,000 loan was actually 6.24%. Source: https://pocketsense.com/calculate-effective-interest-rate-including-discount-points-1295.html, accessed May 2018.

for the amortized payment schedule.) This repayment structure had implications for evaluating LendingClub loans as an investment opportunity.

The Data

Before Figel could build a model, she needed to get a feel for the LendingClub data. She downloaded the three .csv (comma separated value) files that contained data on all the loans originated in 2013 through Q1 2015 and merged them into a single .csv file. She also downloaded the *data dictionary*, which described all features included in the table (see **Exhibit 6**). (This .csv file with the aggregated loan records is referred to hereon as "the data" or "the table.")

After a cursory review, Figel settled in to review the descriptive features more carefully. She was aware that data scientists took a different approach from traditional academic researchers to decide which features to include in a dataset. Academic researchers considered inclusion of any feature about which the researcher did not have a reasonable hypothesis to be an example of data mining. Data scientists broke dramatically from this long-held position in academia; they were more interested in whether a particular combination of features could correctly predict an outcome for a target feature above some pre-determined threshold than they were in proving causal relationships. Data scientist Eric Siegel described the contrast in approaches:

> Causality is elusive, tough to nail down. …Data scientists have it easier with predictive analytics (PA). It just needs to work; prediction trumps explanation. PA operates with extreme solution-oriented intent. The whole point, the 'ka-ching' of value, comes in driving decisions from many individual predictions, one per patient, customer, or person of any kind. And while PA often delivers meaningful insights akin to those of various social sciences, this is usually a side effect, not the primary objective.[13]

Understanding the data

For each entry in the table (row) there were over 100 descriptive *features* (columns). These features included basics such as applicant ID, credit score, loan amount, purpose of loan, as well as more detailed data on financial activity and credit history. Most importantly, the file included the components of what Figel could transform into a *target* feature, *loan status*, which had two classes, default (denoted by 1) and repay (denoted by 0).[g] The *target* was what Figel eventually wanted her model to predict on *new* data with a reassuring level of confidence. Using different prediction models, she would train each model to classify a new loan application on this target feature as either default or repay.

After clearly defining a new target feature and its contents, Figel would work with all descriptive features in the data set. Of course, she would review the feature set for any extremely puzzling or inappropriate features, but her approach would be to whittle down, not build up the feature set, and consider ways to amplify important variables based on her domain knowledge.

[g] Different fields that contribute to data science activity use different vocabulary to refer to the same term. For instance, statistics uses the term *variable* to refer to anything with a quantity or quality that varies, while machine learning uses the term *feature*. The variable of interest in a predictive model is the *dependent variable* in statistics and the *target feature* in machine learning. Other such synonyms exist across these fields.

A spectrum of repayment behavior

Clarifying exactly which repayment behavior would constitute the two classes of her target feature, *loan status*, was critical to think through carefully. Figel cared about whether a borrower defaulted or repaid and the data included those classes for the target feature. However, it also included several other classes, such as *current* and *late*, and Figel had to sort out how to deal with them. **Table 1** shows all possible classes for *loan status* contained in the dataset.

Table 1 *Loan status* classes

Value	Definition
Fully paid	The loan has been repaid in full.
Current	The loan is still being repaid.
Late 16-30 days	The loan has not been current for 16-30 days.
Late 31-120 days	The loan has not been current for 31-120 days.
Default	The loan has not been paid in 121 days or more.
Charged Off	The loan has been in Default for 30 days or more and no more future payments are expected.

Source: LendingClub, "LendingClub Statistics," https://www.lendingclub.com/info/download-data.action, accessed April 2018.

Figel had to determine which of these classes should be included in her target feature, *loan status* (default is 1), and which might not be appropriate for the analysis. After making these decisions she finalized a data set of approximately 413,000 records (numbers are presented in thousands to ease presentation). The number of *defaulters* (1) in the set was approximately 75,000. The number of *repayers* (0) in the set was approximately 338,000.

How to evaluate performance?

Figel had some research to do to square the default rates she was calculating from the LendingClub data (see **Table 2**) with the performance details LendingClub reported (see **Table 3**). At first glance, the default rates in **Table 2** led her to believe that it would not be possible to make a profit on LendingClub loans. However, it was clear from **Table 3** that investors had earned reasonable returns on the loans.

Table 2 Percentage of defaulted LendingClub loans (loans originated 2013 – Q1 2015) in each Grade

Grade	A	B	C	D	E	FG	All
Default Rate	5.29%	11.21%	19.63%	27.23%	36.01%	42.24%	17.97%

Source: LendingClub, "LendingClub Statistics," https://www.lendingclub.com/info/download-data.action, accessed April 2018.

Table 3 Net return (after defaults and charge offs) by grade for Loans issued between 2013 – Q1 2015

Grade	A	B	C	D	E	FG	All
Adjusted Net Annualized Return	5.24%	7.10%	7.81%	8.08%	7.49%	8.01%	7.39%

Source: LendingClub, "LendingClub Statistics," https://www.lendingclub.com/info/demand-and-credit-profile.action, accessed June 2018.

She suspected that the source of the discrepancy lay in a mismatch between the discrete classes of "default" and "repay" and the payments made over time, which had a changing composition of interest and principal from payment to payment. Even when loans defaulted, borrowers had started making payments of interest and principal before the default occurred. Therefore, investors did not lose all the money they had lent.

Building a Predictive Model

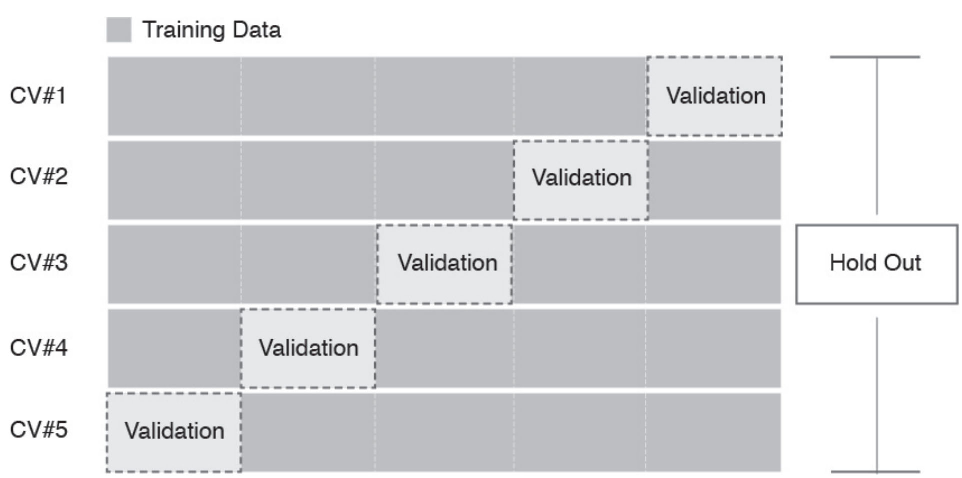
Dividing the data – training, validation, and holdout sets

Before building a model Figel opened a cleaned .csv file with the nearly 413,000 records—75,000 defaulters and 338,000 repayers. Each record was described by over 100 descriptive features about the borrowers. She started by separating the available data into two discrete sets: (1) the *training* set and (2) *holdout* set. These two sets would facilitate three distinct phases of model building.

First, she randomly assigned 20% of the data set from defaulters (15,000) and just over 20% of repayers (68,000) for a total of 83,000 borrowers to the *holdout* set. She then assigned the other 80% of defaulters (60,000) and 80% of repayers (270,000) to the *training* set. She would not touch the holdout set until the very end of the process.

She then turned her attention to the training set. She divided it into five equal, smaller subsets that each constituted approximately 16% of the original training data set. Each subset, or *fold*, now contained 12,000 (60,000 ÷ 5) defaulters and 54,000 (270,000 ÷ 5) repayers for a total of 66,000 borrowers. These five folds would facilitate a *cross-validation* (and *cross-testing*) process during the model building process. **Figure 1** shows the relationship between the training and validation folds and the holdout set. **Table 4** provides a summary of Figel's subdivision of the data.

Figure 1 Cross-validation model with five folds



Source: Casewriter.

Table 4 Summary of data sub-division

Sub-set	Total	Loan is Bad: Default (1)	Loan is Good: Repay (0)
Total	413,000	75,000	338,000
Holdout Set (20%)	83,000	15,000	68,000
Training Set (80%)	330,000	60,000	270,000
Four folds used for estimating the model	264,000	48,000	216,000
Cross-validation fold	66,000	12,000	54,000

Source: Casewriter.

Using cross-validation to train models

In the first iteration of the cross-validation process (CV#1) Figel designated sets 1 through 4 as the *training data* and designated set 5 as the *validation* or *testing* set. She would *train* several different models (using techniques such as logistic regression, decision trees, gradient boosting and their variants) using subsets 1-4, *validate* the models using subset 5, and then calculate a measure of model quality based on how well the model *correctly identified defaulters* and *repayers* in the validation set. The performance on the validation set would help her select the model that most effectively separated defaulters from repayers.

She repeated this process four more times, training each model using 5 different combinations of the folds (e.g. training set: 1,2,3,5; validation or test set: 4). This process created five different versions of the various models that she had run previously on data sets 1-4 and *validation set* 5. She averaged the performance of the different models on each of the validation sets and chose the model with the best accuracy in correctly predicting defaulters and repayers. Once she selected the type of model (e.g. logistic regression or decision trees) she would then further refine it for use in the rest of her work.

Validating the model on the holdout sample

She then reran the preferred model on all 5 sub-segments (80% of the data) to come up with a final model. She used the last 20% of the data, the *holdout set*, to *evaluate* the model. Although it came from the same pool of data, technically the hold out set was brand new data that the model had not yet encountered. The integrity of the holdout set, for which she knew the outcome, was critical to her ability to properly evaluate how well the model would perform on new data for which an outcome was *not* known.

As Siegel writes, "Since the [holdout] set was not used to create the model, there's no way the model could have captured its esoteric aspects, its eccentricities."[14] This final step approximated the model's ability to classify a *new* loan application as default/not default based on data it had never before seen. Figel checked to see that the prediction results she observed in the validation sample were roughly similar to the prediction results on the holdout sample. (See **Exhibit 7** for a summary of this model building process.)

Balanced or unbalanced samples

Figel had grappled with how she should deal with the small number of defaulters compared to repayers in this dataset. She was confused about whether she should use all 264,000 observations in subsets 1-4 (48,000 defaulters and 216,000 repayers) to prepare the model. She was really interested in understanding defaulters and this data would skew the analysis toward repayers.

Alternatively, she could use a balanced data set of 48,000 defaulters and 48,000 repayers to allow the model to focus equally on defaulters and repayers. Her concern in doing so was that the model would no longer be representative of the actual data. If she did decide to use a balanced data set, she also had to decide what she should do with regard to the validation sample of subset 5. Should she use a balanced data set of 12,000 defaulters and 12,000 repayers to test the model? Or should she use all 66,000 borrowers (12,000 defaulters and 54,000 repayers) to test her results?

Making predictions

To start with, Figel trained a regularized logistic regression model. She was interested in learning how well the model did in the cross-validation samples.

First Attempt Figel's first attempt resulted in log losses close to zero in the cross validation tests. The predictions were almost perfect! The first three columns of **Exhibit 8** show 20 observations from the cross-validation sample. The model had predicted all defaulters with a probability of 1 and had predicted all repayers to have a zero probability of default.

Doubt quickly crowded out her initial sense of awe. She had a hunch that she had not properly prepared the data set.

Second Attempt Figel made some adjustments to the features she had included in her training data. **Exhibit 9** presents the features excluded from the dataset in "Attempt 2." She then re-ran the model. It now produced much different results. **Exhibit 8** presents the prediction probabilities for the subset of 20 observations from the cross-validation sample using the adjusted features (see column 4). The results were dramatically different. The predictions were far from perfect.

Risk and Return

Figel was well aware that the opportunities for earning higher returns came with additional risk factors.

Quality of borrowers

One concern that emerged in mid-2016 was the quality of LendingClub borrowers. In the summer of 2016, *The Wall Street Journal* reported that between 2013 and 2016 loan charge offs[h] had risen as much as 38%.

Some critics pointed to LendingClub's loan approval process as a weak point in the system. The company screened all applicants using automated credit checks and credit histories, but it only verified income through review of tax returns or pay stubs some of the time. Income verification rates reached a high of 49% in 2013, but had dropped to 26.8% as of Q1 in 2016. (The company verified *sources* of income for another 30% of applicants.) The charge off rate among lower-graded loans also rose during this period, from 4.58% to 6.31%. This was in contrast to high-graded loan charge offs, which rose from just 1.46% to 1.51%. For comparison, delinquency rates for consumer loans (including credit cards) across all U.S. banks hovered between 1.98% and 2.56% from Q1 2013 through Q1 2018, as tracked by the Federal Reserve.[15]

Some analysts cited irrational behavior on the part of borrowers for some of the problems. The research showed that many borrowers cited "debt consolidation" in their loan applications, but that nearly half of those who borrowed for debt consolidation "actually started carrying 10% more in credit-card debt after getting the loan." It appeared that for some, LendingClub loans served as temporary stop gaps to support a longer pattern of overspending and not as permanent solutions to elimination of personal debt.

[h] Refer to **Table 1** for the definition of loan charge off.

Other analysts pointed to misaligned incentives between LendingClub and investor-lenders. LendingClub generated revenue through loan origination fees. These analysts argued there was some incentive (however subtle) to approve an applicant for inclusion on the platform along with other more credit worthy applicants.[16] However, the delinquency rates LendingClub reported (and third parties confirmed) for any given cohort of loans it originated were generally in line with rates for consumer loans issued by banks around the U.S. as tracked by Federal Reserve.

Transparency

LendingClub's transparent approach to data was impressive. It was also crucial to the company's business model. Potential investors had to be able to assess the quality of a loan by reviewing data about it to determine whether the potential return aligned with their individual risk profiles. They needed to trust that the data LendingClub was providing was correct. The presence of sophisticated institutional investors had provided confidence to investors of all sizes, including the smaller investment firms. Their engagement signaled that the organization was trustworthy and that the data the company released about individuals loans, including key features like grade assignments, was accurate.

A 2016 incident underscored the importance of trust to the LendingClub operation, as well as its fragility. Jefferies was a global investment bank with offices across the Americas, Asia, Europe, and the Middle East that offered to buy loans of a certain grade from LendingClub. After completing the transaction, executives at Jeffries discovered that at least $22 million of the loans purchased did not meet the specified grade levels. They also learned that several LendingClub executives had known about the mistake and said nothing.[17] Founder and CEO Renaud Laplanche resigned, the executives were fired, and the company redoubled efforts to improve and implement effective controls to ensure and maintain the veracity of its data.

LendingClub as corporate entity

A last concern was the relationship between the borrowers, lenders, and LendingClub the corporate entity. A critical implication of investing in an unsecured note was that investors were exposed, however minimally, to risk that they might lose part of their investments should LendingClub the corporate entity face serious financial trouble. For example, Figel could purchase the note that represented an individual's entire loan. That borrower might be diligently paying month after month with minimal chance or intention of default, but if LendingClub filed for bankruptcy Figel would not be entitled to continue receiving the income stream of monthly payments from "her" borrower. Rather, she would be just one of many of unsecured creditors standing in line to see what portion of her investment she might be able to get back. And there was no guarantee that she would get any back.[18]

Overall, Figel was confident in her abilities to leverage a model that could predict default with high accuracy. With enough loan data, Figel believed she could build a model to predict which applicants were likely to default. That was the benefit of having access to all of LendingClub's application data—she could use it to train models to answer the nuanced questions that mattered to her.

Exhibit 1 Individual borrower profile as shown on the LendingClub investor dashboard

LOAN ID:

Loan Submitted:	**4/30/18 10:06 AM**	Review Status:	**Approved**
Monthly Payment:	**$980.75 (36)**	Member Loan#:	Prospectus

Profile *(all information unverified unless otherwise denoted)*

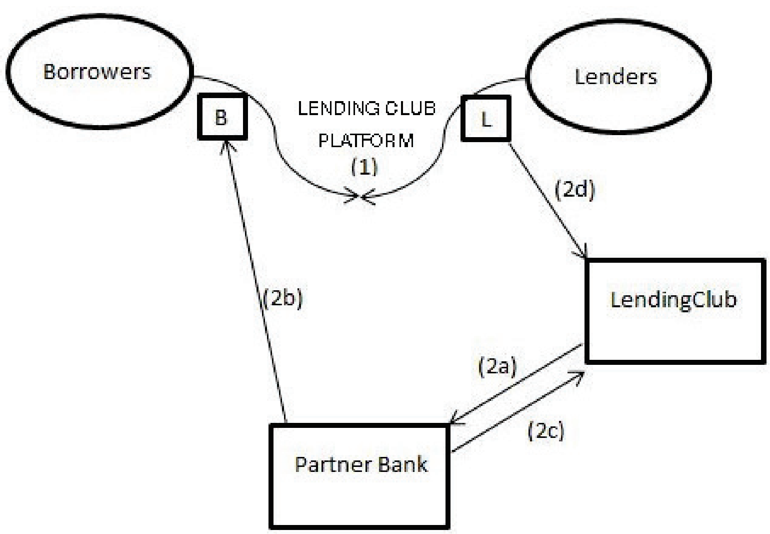
Home Ownership:	**MORTGAGE**	Job Title:	**Manager**
Length of Employment:	**5 years**	Location:	**446xxOH**
Individual Gross Income:	**$5,000 / month**	Individual Debt-to-Income:	**28.86%****

Credit History *(as reported by credit bureau on 4/25/18)*

Credit Score Range:	**720-724**	Delinquent Amount:	**$0.00**
Earliest Credit Line:	**08/2000**	Delinquencies (last 2 yrs):	**0**
Open Credit Lines:	**20**	Months Since Last Delinquency:	**80**
Total Credit Lines:	**47**	Public Records on File:	**0**
Revolving Credit Balance:	**$32,423.00**	Months Since Last Record:	**n/a**
Revolving Line Utilization:	**52.50%**	Months Since Last Major Derogatory:	**n/a**
Inquiries in Last 6 Months:	**1**	Collections Excluding Medical:	**0**
Accounts Now Delinquent:	**0**		

Source: LendingClub, "Manual Investing," www.lendingclub.com/browse/browse.action, accessed June 2018.

Exhibit 2 Lending process diagram



Source: LendingClub Securities Litigation, 99 *F. Sec*. 730, N.D.C, (2017).

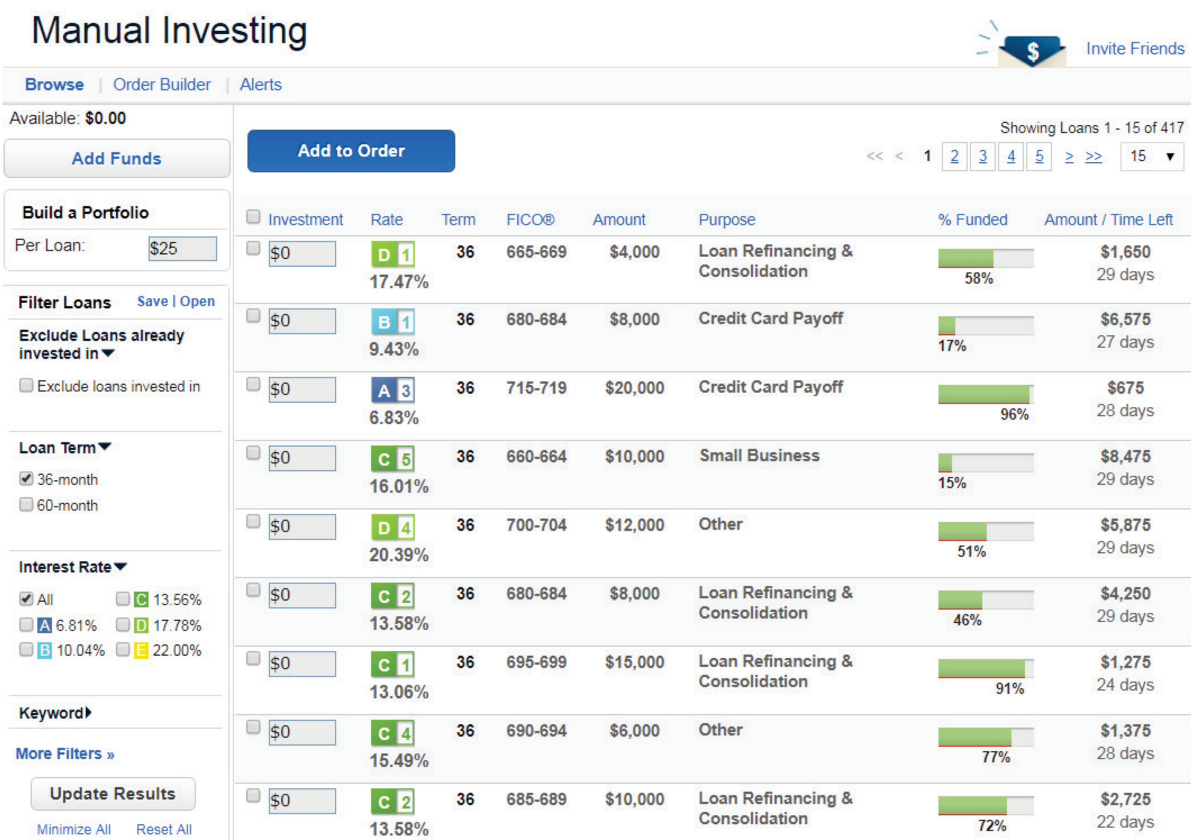
Note: LendingClub (a) directed a *partner bank* to (b) initiate a loan to the *borrower*, which the partner bank immediately (c) sold to LendingClub, which the company paid for using funds an individual *investor* (d) paid to LendingClub for the note.

Exhibit 3 Interest rates by grade and sub-grade (percent)

Sub-grade	A	B	C	D	E
1	6.03	9.43	13.06	17.47	22.90
2	6.14	10.07	13.58	18.45	23.87
3	6.83	10.56	14.52	19.42	24.84
4	7.46	11.05	15.49	20.39	25.81
5	8.08	12.13	16.01	21.85	26.77

Source: LendingClub, "Rates and Fees," https://www.lendingclub.com/public/rates-and-fees.action, accessed May 2018.

Exhibit 4 Investor dashboard



Source: LendingClub, "Manual Investing," https://www.lendingclub.com/browse/browse.action, accessed May 2018.

Exhibit 5 Amortization schedule for a $10,000 loan over 36 months (6.03% interest rate)

Payment #	Date	Payment	Principal	Interest	Remaining	% PMT Interest
1	7/1/2018	$304.36	$254.11	$50.25	$9,745.89	16.51%
2	8/1/2018	$304.36	$255.38	$48.97	$9,490.51	16.09%
3	9/1/2018	$304.36	$256.67	$47.69	$9,233.85	15.67%
4	10/1/2018	$304.36	$257.96	$46.40	$8,975.89	15.25%
5	11/1/2018	$304.36	$259.25	$45.10	$8,716.64	14.82%
6	12/1/2018	$304.36	$260.55	$43.80	$8,456.09	14.39%
7	1/1/2019	$304.36	$261.86	$42.49	$8,194.22	13.96%
8	2/1/2019	$304.36	$263.18	$41.18	$7,931.04	13.53%
9	3/1/2019	$304.36	$264.50	$39.85	$7,666.54	13.09%
10	4/1/2019	$304.36	$265.83	$38.52	$7,400.71	12.66%
11	5/1/2019	$304.36	$267.17	$37.19	$7,133.54	12.22%
12	6/1/2019	$304.36	$268.51	$35.85	$6,865.03	11.78%
13	7/1/2019	$304.36	$269.86	$34.50	$6,595.18	11.34%
14	8/1/2019	$304.36	$271.21	$33.14	$6,323.96	10.89%
15	9/1/2019	$304.36	$272.58	$31.78	$6,051.38	10.44%
16	10/1/2019	$304.36	$273.95	$30.41	$5,777.44	9.99%
17	11/1/2019	$304.36	$275.32	$29.03	$5,502.11	9.54%
18	12/1/2019	$304.36	$276.71	$27.65	$5,225.41	9.08%
19	1/1/2020	$304.36	$278.10	$26.26	$4,947.31	8.63%
20	2/1/2020	$304.36	$279.50	$24.86	$4,667.81	8.17%
21	3/1/2020	$304.36	$280.90	$23.46	$4,386.91	7.71%
22	4/1/2020	$304.36	$282.31	$22.04	$4,104.60	7.24%
23	5/1/2020	$304.36	$283.73	$20.63	$3,820.87	6.78%
24	6/1/2020	$304.36	$285.16	$19.20	$3,535.72	6.31%
25	7/1/2020	$304.36	$286.59	$17.77	$3,249.13	5.84%
26	8/1/2020	$304.36	$288.03	$16.33	$2,961.10	5.37%
27	9/1/2020	$304.36	$289.48	$14.88	$2,671.62	4.89%
28	10/1/2020	$304.36	$290.93	$13.42	$2,380.69	4.41%
29	11/1/2020	$304.36	$292.39	$11.96	$2,088.30	3.93%
30	12/1/2020	$304.36	$293.86	$10.49	$1,794.44	3.45%
31	1/1/2021	$304.36	$295.34	$9.02	$1,499.10	2.96%
32	2/1/2021	$304.36	$296.82	$7.53	$1,202.28	2.47%
33	3/1/2021	$304.36	$298.31	$6.04	$903.97	1.98%
34	4/1/2021	$304.36	$299.81	$4.54	$604.15	1.49%
35	5/1/2021	$304.36	$301.32	$3.04	$302.83	1.00%
36	6/1/2021	$304.36	$302.83	$1.52	$0.00	0.50%

Source: Casewriter using LendingMemo, "Peer to peer loan calculator for LendingClub and Prosper," https://www.lendingmemo.com/amortization-calculator/, accessed May 2018.

Exhibit 6 Data dictionary

This dictionary provides a representative sample of those features.

APPLICATION	
Basic Data	
id	A unique LC assigned ID for the loan listing.
zip_code	The first 3 numbers of the zip code provided by the borrower in the loan application.
addr_state	The state provided by the borrower in the loan application.
home_ownership	The home ownership status provided by the borrower during registration or obtained from the credit report. Our values are: Rent, Own, Mortgage, Other.
emp_title	The job title supplied by the Borrower when applying for the loan.*
emp_length	Employment length in years. Possible values are between 0 and 10 where 0 means less than one year and 10 means 10 or more years.
annual_inc	The self-reported annual income provided by the borrower during registration.
annual_inc_joint	The combined self-reported annual income provided by the borrower during registration.
mort_acc	Number of mortgage accounts.
Loan Request	
desc	Loan description provided by the borrower.
purpose	A category provided by the borrower for the loan request.
application_type	Indicates whether the loan is an individual application or a joint application with two co-borrowers.
Loan Terms	
loan_ammt	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.
installment	The monthly payment owed by the borrower if the loan originates.
int_rate	Interest rate on the loan.
term	The number of payments on the loan. Values are in months and can be either 36 or 60.
LendingClub Actions	
verification_status	Indicates if income was verified by LC, not verified, or if the income source was verified.
verified_status_joint	Indicates if the co-borrowers' joint income was verified by LC, not verified, or if the income source was verified.
CREDIT PROFILE	
Debt Load	
dti	A ratio calculated using the borrower's total monthly debt payments on the total debt obligations, excluding mortgage and the requested LC loan, divided by the borrower's self-reported monthly income.
Score	
fico_range_high	The upper boundary range the borrower's FICO at loan origination belongs to.
fico_range_low	The lower boundary range the borrower's FICO at loan origination belongs to.
mths_since_recent_inq	Months since most recent inquiry.
Credit Lines	
earliest_cr_line	he month the borrower's earliest reported credit line was opened.
mo_sin_old_rev_tl_op	Months since oldest revolving account opened.
mo_sin_rcnt_rev_tl_op	Months since most recent revolving account opened.
num_tl_op_past_12m	Number of accounts opened in past 12 months.
Credit Balance	
total_acc	The total number of credit lines currently in the borrower's credit file.
total_bal_ex_mort	Total credit balance excluding mortgage.
tot_cur_bal	Total current balance of all accounts.
avg_cur_bal	Average current balance of all accounts.
total_acc	The total number of credit lines currently in the borrower's credit file.
total_bc_limit	Total bankcard high credit/credit limit.
tot_hi_cred_lim	Total high credit/credit limit.
open_acc	The number of open credit lines in the borrower's credit file.
percent_bc_gt_75	Percentage of all bankcard accounts > 75% of limit.
Credit Inquiries	
inq_last_12m	Number of credit inquiries in past 12 months.
inq_last_6mths	The number of inquiries in past 6 months (excluding auto and mortgage inquiries).

Exhibit 6 (continued)

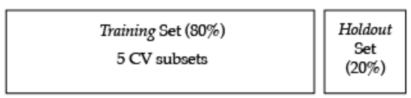
Active Accounts	
num_sats	Number of satisfactory accounts.
revol_bal	Total credit revolving balance.
revol_util	Revolving line utilization rate, or the amount of credit the borrower is using relative to all available revolving credit.
Delinquency	
tot_coll_amt	Total collection amounts ever owed.
chargeoff_within_12_mths	Number of charge-offs within 12 months.
collections_12_mths_ex_med	Number of collections in 12 months excluding medical collections.
delinq_2yrs	The number of 30+ days past-due incidences of delinquency in the borrower's credit file for the past 2 years.
delinq_amnt	The past-due amount owed for the accounts on which the borrower is now delinquent.
mths_since_last_delinq	The number of months since the borrower's last delinquency.
num_accts_ever_120_pd	Number of accounts ever 120 or more days past due.
acc_now_delinq	The number of accounts on which the borrower is now delinquent.
num_tl_120dpd_2m	Number of accounts currently 120 days past due (updated in past 2 months).
num_tl_30dpd	Number of accounts currently 30 days past due (updated in past 2 months).
num_tl_90g_dpd_24m	Number of accounts 90 or more days past due in last 24 months.
tax_liens	Number of tax liens.
pub_rec	Number of derogatory public records.
pub_rec_bankrupticies	Number of public record bankruptcies.
LOAN	
Funding	
funded_amnt	The total amount committed to that loan at that point in time.
last_fico_range_high	The most recent month LC pulled credit for this loan.
last_fico_range_low	The upper boundary range the borrower's last FICO pulled belongs to.
last_credit_pull_d	The lower boundary range the borrower's last FICO pulled belongs to.
Loan Payments	
loan status	Current status of the loan.
total_pymnt	Payments received to date for total amount funded.
total_rec_prncp	Principal received to date.
total_rec_int	Interest received to date.
last_pymnt_amnt	Last total payment amount received.
last_pymnt_d	Last month payment was received.
out_prncp	Remaining outstanding principal for total amount funded.
out_prncp_inv	Remaining outstanding principal for portion of total amount funded by investors.
pymnt_plan	Indicates if a payment plan has been put in place for the loan.
Hardship Payment	
hardship_flag	Flags whether or not the borrower is on a hardship plan.
hardship_type	Describes the hardship plan offering.
hardship_reason	Describes the reason the hardship plan was offered.
hardship_status	Describes if the hardship plan is active, pending, canceled, completed, or broken.
hardship_amount	The interest payment that the borrower has committed to make each month while they are on a hardship plan.
Settlement Plan	
debt_settlement_flag	Flags whether or not the borrower, who has charged-off, is working with a debt-settlement company.
debt_settlement_flag_date	The most recent date that the Debt_Settlement_Flag has been set.
settlement_percentage	The settlement amount as a percentage of the payoff balance amount on the loan.

Source: LendingClub, "LendingClub Statistics," https://www.lendingclub.com/info/download-data.action, accessed April 2017.

Exhibit 7 Summary of model building process

1 Build five different models.

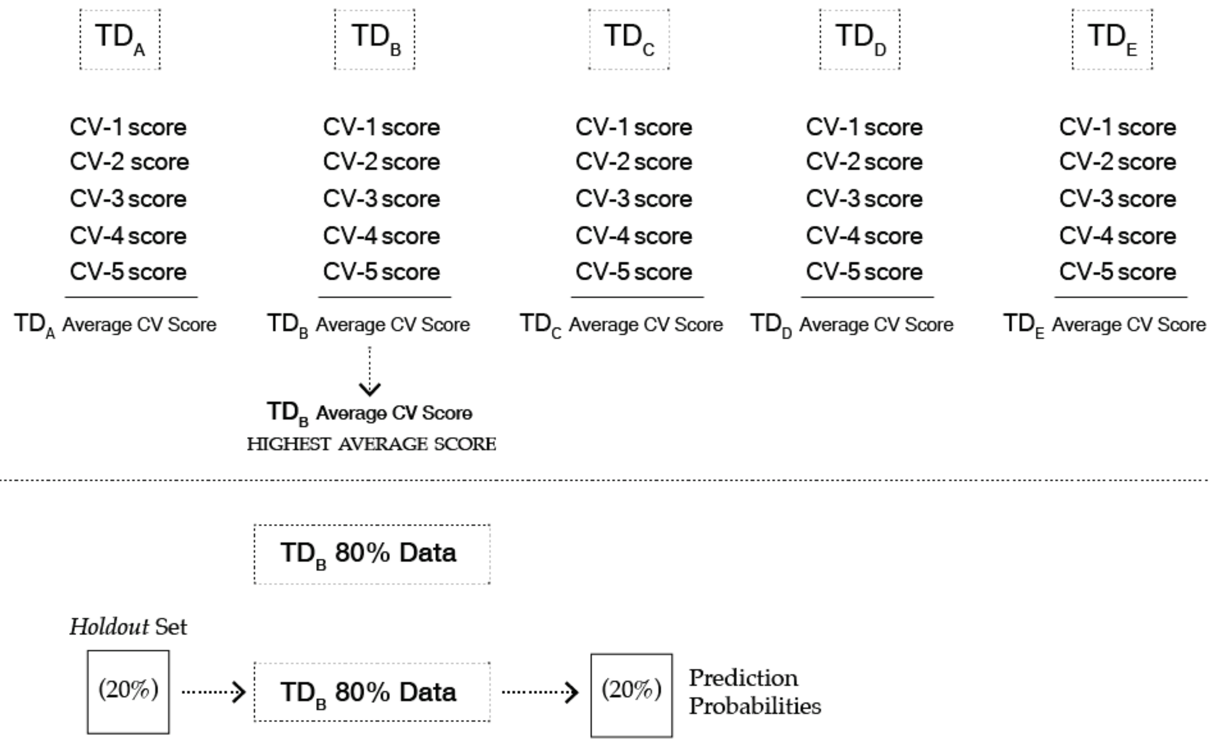
2 Use the cross-validation method to produce 5 different CV scores for each model.
(Training set: 1,2,3,4; Validation set: 5).



3 Select model with best accuracy.
(As measured by CV score.)

4 Train a new model with the same tree depth as model TD_B.
Use all 80% of the data to train.

5 Test the model on brand new data: the *holdout* set.
(The holdout set is 20% of original set.)



Source: Casewriter.

Note: TD stands for "training data."

Exhibit 8 Model scoring set prediction probabilities

Record #	Actual Loan is Bad = 1	Model Attempt 1	Model Attempt 2
1	1	1.0	86.96%
2	1	1.0	9.47%
3	1	1.0	72.75%
4	1	1.0	47.14%
5	1	1.0	95.05%
6	1	1.0	77.27%
7	1	1.0	81.48%
8	1	1.0	49.76%
9	1	1.0	72.95%
10	1	1.0	8.64%
11	0	0.0	3.07%
12	0	0.0	79.58%
13	0	0.0	1.45%
14	0	0.0	17.90%
15	0	0.0	1.55%
16	0	0.0	0.75%
17	0	0.0	0.15%
18	0	0.0	0.20%
19	0	0.0	21.65%
20	0	0.0	0.67%

Source: Casewriter.

Exhibit 9 Data set features excluded from attempt 2

LOAN	
Funding	
funded_amnt	The total amount committed to that loan at that point in time.
last_fico_range_high	The most recent month LC pulled credit for this loan.
last_fico_range_low	The upper boundary range the borrower's last FICO pulled belongs to.
last_credit_pull_d	The lower boundary range the borrower's last FICO pulled belongs to.
Loan Payments	
loan status	Current status of the loan.
total_pymnt	Payments received to date for total amount funded.
total_rec_prncp	Principal received to date.
total_rec_int	Interest received to date.
last_pymnt_amnt	Last total payment amount received.
last_pymnt_d	Last month payment was received.
out_prncp	Remaining outstanding principal for total amount funded.
out_prncp_inv	Remaining outstanding principal for portion of total amount funded by investors.
pymnt_plan	Indicates if a payment plan has been put in place for the loan.
Hardship Payment	
hardship_flag	Flags whether or not the borrower is on a hardship plan.
hardship_type	Describes the hardship plan offering.
hardship_reason	Describes the reason the hardship plan was offered.
hardship_status	Describes if the hardship plan is active, pending, canceled, completed, or broken.
hardship_amount	The interest payment that the borrower has committed to make each month while they are on a hardship plan.
Settlement Plan	
debt_settlement_flag	Flags whether or not the borrower, who has charged-off, is working with a debt-settlement company.
debt_settlement_flag_date	The most recent date that the Debt_Settlement_Flag has been set.
settlement_percentage	The settlement amount as a percentage of the payoff balance amount on the loan.

Source: LendingClub, "LendingClub Statistics," https://www.lendingclub.com/info/download-data.action, accessed May 2018.

Endnotes

[1] Noah Buhayar, "Peer-to-peer Lending," *Bloomberg*, June 6, 2016, https://www.bloomberg.com/quicktake/peer-peer-lending, accessed June 20, 2017.

[2] Simon Cunningham, "IPO: The Day the Nation Met Lending Club," LendingMemo, December 11, 2014, https://www.lendingmemo.com/lending-club-ipo/, accessed May 2018.

[3] Ibid.

[4] LendingClub, 2014 10-K p. 4, http://ir.lendingclub.com/Cache/28067148.pdf?IID=4213397&FID=28067148&O=3&OSID=9, accessed May 2018.

[5] Ibid.

[6] Ibid.

[7] Ian Salisbury, "Lending Club's CEO Has Left and Its Stock Has Plunged. Should Lenders Bail Out?" Money, June 13, 2016, http://time.com/money/4340588/peer-to-peer-lending-club-woes/.

[8] LendingClub Securities Litigation, 99 *F. Sec.* 730, N.D.C, (2017).

[9] "Compound Annual Growth Rate (Annualized Return)," Money Chimp, http://www.moneychimp.com/features/market_cagr.htm, accessed May 2018.

[10] "Annual asset performance comparison, 1978-2017," BullionVault, https://www.bullionvault.com/gold-guide/annual-asset-performance-comparison, accessed May 2018.

[11] Aswath Damodaran, "Annual Returns on Stock, T.Bonds and T.Bills: 1928 – Current," NYU Stern School of Business, January 5, 2018, http://pages.stern.nyu.edu/~adamodar/New_Home_Page/datafile/histretSP.html, accessed May 2018.

[12] "Peer to peer loan calculator for LendingClub and Prosper," LendingMemo, https://www.lendingmemo.com/amortization-calculator/, accessed May 2018.

[13] Eric Siegel, *Predictive Analytics: The Power to Predict Who Will Click, Buy, Lie, Or Die* (Hoboken: Wiley, 2013), p. 132.

[14] Ibid, p. 173.

[15] Board of Governors of the Federal Reserve System, "Charge-Off and Delinquency Rates on Loans and Leases at Commercial Banks," Federal Reserve System Website, https://www.federalreserve.gov/releases/Chargeoff/delallsa.htm, accessed June 2018.

[16] Telis Demos and Peter Rudegeair, "LendingClub's Newest Problem: Its Borrowers; Charge-off rates have risen as much as 38% since 2013 at the largest online lender," *Wall Street Journal* (online), July 12, 2016.

[17] Ian Salisbury, "Lending Club's CEO Has Left and Its Stock Has Plunged. Should Lenders Bail Out?" Money, June 13, 2016, http://time.com/money/4340588/peer-to-peer-lending-club-woes/.

[18] Ibid.