Fighting Predatory Lending

Final Report

Tilina Alzaben
Elizabeth Morris
Hamza Zia

Sewanee DataLab 2022

In collaboration with BetterFi

Table of Contents

Problem Statement

Predatory payday, title, and flex loans are a growing affliction in low-income communities that lead people to pay hundreds in interest and fees every month yet never pay down their loans. Our non-profit partner BetterFi offers those borrowers a pathway out of debt traps and toward financial fitness by refinancing their loans at reasonable rates, as well as offering new loans to those who qualify. Using past client data, we will develop predictive models which will create a more efficient and scalable way for BetterFi to service the community and end the cycle of debt.

Partner Perspective

Scope

The United States Department of Justice loosely defines predatory lending as "the fraudulent, deceptive, and unfair tactics some people use to dupe us into loans that we can't afford." Most of these strategies focus on people from lower socioeconomic communities who are already vulnerable to financial hardships or have poor credit.

If you have bills, rent, or a financial emergency– payday and title loans can act as a quick fix. However, these loans have astonishingly high-interest rates bundled in a deal that sounds too good to be true. They often trap people in a cycle of debt that can last for years, wreaking havoc on families, neighborhoods, and lifestyles. Several of these shops can be found in local communities around Tennessee, and according to new evidence, these shops have now surpassed McDonald's in terms of the number of locations in the United States.

Background

BetterFi, a CDFI and 501(c)(3), is an economic justice organization that offers low-interest payday loans to people with bad credit in order to combat predatory lending and help low-income households get out of debt. It currently has two full-time employees and a $120,000 annual budget, which is funded through donations, grants, and loan repayments.

Problem

As of now, the lending procedure involves obtaining as much information about the client before agreeing on whether or not to provide the client with a loan. The loan committee makes lending decisions on a case-to-case basis; however, because of the limited available data and the labor-intensive nature of the process, this method often results in inconsistent decisions and bottlenecks the number of loans that can be approved.

Solution

BetterFi hopes to develop a metric based on prior lending data to enhance and evaluate the lending policy. The approach is to quantify the measure of reliability in hopes of minimizing the

delinquency rate by predicting whether or not a customer would repay their loans. As a result, a machine learning model will be produced to generate a loan repayment score using data obtained from clients and prospective clients.

Outcome

If DataLab's study is successful, BetterFi will be able to use the machine learning model to produce more accurate loan repayment predictions, as well as lend to more people to help them escape predatory lending.

Goal Setting

Bare Minimum (MVP):
- Building a predictive model that takes in client data and returns the likelihood of a client either repaying their loan in full or charging off
- Improving the model to continuously take in new clients
- Creating data quality guidelines to standardize the data being inputted into the model

Aspirations:
- Visualizing areas where people are vulnerable to predatory lending to identify prospective clients through an interactive, educational map
- Finding additional variables to incorporate into the BetterFi application

Pre Mortem

Not Finishing All Tasks
- Problem: Not finishing all desired tasks by the end of the program
- Why: Poor time management and task allocation
- Solution: Making use of project management programs such as Trello

Overfitting the Model
- Problem: Delivering a model that can't generalize with new client data
- Why: Working with a small dataset
- Solution: Creating a simple predictive model with 10-fold cross-validation

Misinformed Decisions
- Problem: Implementing a model with weak accuracy
- Why: The current client interface used to gather client data is tricky to navigate which results in messy data being inputted
- Solution: A data quality guideline to standardize the data

Literature Review

It is simple to envision situations in which a person would require quick cash– perhaps your car broke down and you are unable to travel to work; maybe you don't have enough cash on hand to cover your rent or expenses at the end of the month; maybe you experienced a medical emergency. With two-thirds of American households living paycheck to paycheck, there are endless reasons as to why a loan might be needed to help make ends meet. Residents are in the unfortunate position of being exposed to the most title and payday loan businesses, with the state of Tennessee being top five in the country of brick-and-mortar shops. Predatory lending has been the subject of previous research, and we will examine that research as well as solutions in this review.

Increasing Predatory Lending Locations
Predatory lending's past and present statistics, particularly in Tennessee, are examined in-depth in a report by Metro Ideas Project (MIP). According to the report, "the payday industry emerged in the 1990s and would allow loans to be made against a postdated check in exchange for triple-digit APRs, with effective annual percentage rates sometimes exceeding 1,826 percent. The number of payday brick-and-mortar locations grew from virtually zero in 1990 to over 10,000 locations across the United States by 1999." To give scale to these institutions' operations they reference, "the Consumer Finance Protection Bureau estimates that in 2015 there were 15,766 payday loan stores across 36 states—compare that to the 14,350 McDonald's fast-food outlets in all of the United States in 2014." The placement of these establishments, which is often in poorer neighborhoods where individuals are more likely to require loans, has had detrimental effects both nationally and in Tennessee. The conclusions below are based on data gathered by MIP from authorized payday loan outlets in Tennessee. However, it is still helpful to view licensed lenders by county in order to determine what counties may need the greatest assistance. It is important to note that this is not a complete depiction because many of these institutions are not regulated by the government.

Findings

MAP 1. Payday Lending Locations by County

In total, there were 1,233 payday lending locations in Tennessee in November 2017. Compare that to the 155 total retail locations of Walmart or the 393 McDonald's franchises or even the 3,350 gas stations serving customers in the state.

TABLE 1. Tennessee Counties with the Most Predatory Lenders

County	Predatory Lenders	Population
Shelby County	232	937,750
Davidson County	109	658,506
Hamilton County	71	348,121
Knox County	68	444,348
Rutherford County	50	282,558

Policy
Considering historical laws is a good method to see how lending impacts people. In the 1990s and 2000s, interest rates skyrocketed thanks to loopholes in several policies, one of them being the Depository Institutions Deregulation and Monetary Control Act of 1980. This was originally implemented to uncap mortgage rates during a time of high inflation and savings, yet twenty years later it became means for unusually high interest rates. These were some of the early beginnings of the subprime lending sector; with a market that was rapidly expanding and few legal safeguards in place, loan recipients were in fact the true victims.

On another note, research done in 1999 by Deborah Goldstein for the Joint Center for Housing Studies of Harvard University gives insight into the beginning of these lending institutions' growth. Goldstein discusses some policies that have remedies and weaknesses surrounding this issue, with the Fair Trade Commision Act and the Truth in Lending Act being some of the most notable to offer protection to lendees. These policies are still very broad, and those gaps allow predatory lending institutions to still exist, as well as not release complete disclosure to potential clients.

Remedies for Abusive Terms and Practices

Statute	The Depository Institutions Deregulation and Monetary Control Act	Alternative Mortgage Transaction Parity Act	Equal Credit Opportunity Act	Fair Trade Commission Act	Fair Credit Reporting Act	Fair Debt Collection Practices Act
Remedy For			Targeting minorities	Upselling, collection	Collection	Collection
Weakness	Limits state regulatory power.	Limits state regulatory power.			No provision against failure to report credit history.	

In modern day, the issue is just now being recognized as a systemic issue that affects banking as a whole.

Solutions
As a solution, nearly all research indicates to revise and enact new policies to help combat predatory lending. Nevertheless, this takes a lot of time, and most likely the predatory lenders would retaliate by any medium possible including gaining access to powerful lobbyists. Another solution that has been proposed in the meantime is to create lending institutions that can compete with them, but for far lower interest rates. This is where BetterFi comes in– a non-profit, Community Development Financial Institution that offers loans at lower interest rates for those who would otherwise have nowhere else to turn to. There have been a couple of start-ups similar to BetterFi, for example one in Minnesota called Exodus Lending who has had great success saving clients over $1,000,000. If more and more of these location-based lending institutions are constructed with solutions specific to their community, it will introduce a competitive market that can combat predatory lending.

Data Description

- Three datasets shared with us by BetterFi
- External data was collected to create an interactive map
- Historical data collected from previous loan approval decisions

Summary of BetterFi's Datasets
Client Information
- Dataset collected from 85 clients
- Includes information such as a client's income amount, ethnicity, age, employer, area code, zip code, county, etc.

Loan Information
- Dataset regarding 197 loans given out between 2018 to 2022
- Broken down by each client and includes information regarding a loan's purpose, the amount approved, the interest rate, and the current applicant's status (either paid in full or defaulted on the loan)

Transaction Information
- Broken down by each client and includes the total disbursement amount and every transaction made towards the loan

Summary of External Datasets
To create a map, a dataset including the location of all predatory lenders in Tennessee. Data from the U.S. Census Bureau and the Housing of Urban Development Department was also utilized to obtain ranges of income levels per county.

Product Proposal

Final Product
Our final product will consist of an internal and external dashboard. The internal dashboard includes a predictive model that will help determine whether or not a loan will be paid in full or charged off. It also provides insights into the current existing client, loan and transaction data. In the external dashboard there is a map showing potential predatory lending locations in BetterFi's area of operation and respective public APRs, as well as median household income and population of each county.

Method
To optimize accuracy and interpretability, both a logistic regression and a decision tree were trained. A decision tree was selected as the final model since it consistently outperformed the logistic regression.

Since the dataset included less than 200 loans, but more than 20 features, the decision tree selected specific variables to avoid overfitting; however, due to anti-discriminatory laws, some variables were excluded from the final model.

The final variables included in the model include employer, street zip, annual income, income level, total expenses, income source, payment period, monthly payments towards the loan, and the total amount of the loan approved. Furthermore, 10-fold cross validation was performed to better evaluate the model's performance.

Reasoning
The model will act as a tool to supplement BetterFi's current time-consuming decision making. Our dashboard also offers insights and shows trends in their data to help decide which people to reach out to. With the map, it can act as a resource to the community that shows potential predatory lending locations and annual percentage rates to customers to help stay informed.

Mockup
Predictive Model:
  - Includes text boxes that will take in client information
  - Will generate a prediction of a customer either charging off or paying in full

Data Analytics:
  - Categories broken down into: Demographics, Income and Expenses, and Loan Details
  - Significant trends that influence the probability of repayment
  - Relationship between variables and loan repayment

Map:
  - Potential predatory lending locations and their respective APR
  - Household median income and population broken down by county
  - Facts about predatory lending and explaining details about the map

Data Quality Guidelines

Below, each data quality dimension is highlighted and addressed to ensure standardized data which will help with future model development.

Accuracy
Accuracy refers to the degree in which the collected data is correct and measures what it was originally designed to measure. For an accurate analysis, data must be meaningful which means that it must conform to a consistent defined format.

Properties
There are three methods to ensure data accuracy moving forward:

1. Data Collection: the method in which data is collected could contribute to error. Currently, BetterFi takes in client information in 4 ways: in-person, over the phone, through a partner organization, and through an online form that the client fills out. This is what the breakdown of the number of applications per application method looks like:
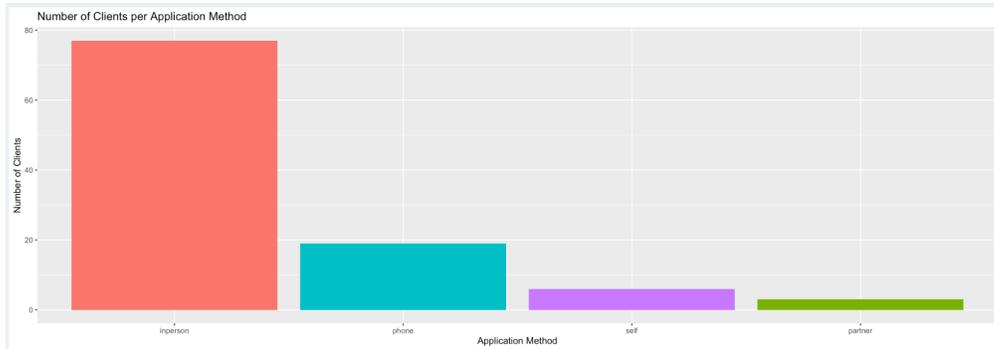
We can see here that the majority of applicants fill out their applications in person compared to the other methods. We can also analyze data quality by looking at the proportion of data with missing values.
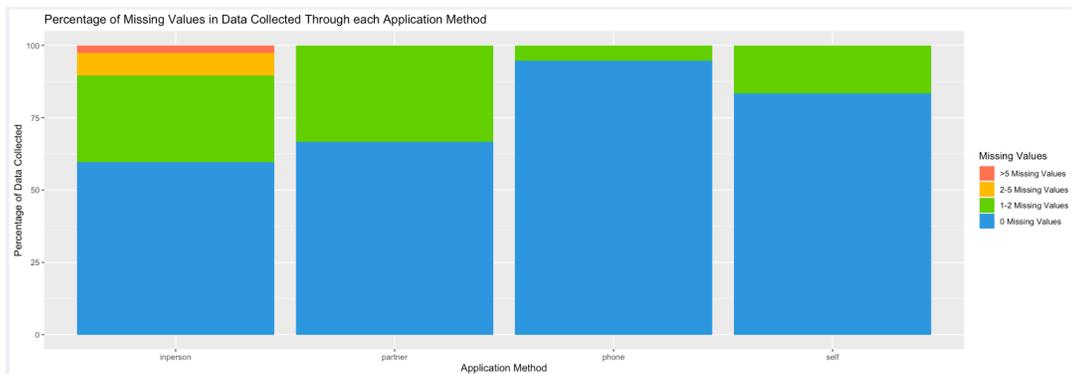
Here we can see that in person applications also had the highest percentage of missing values. This does not mean that in person applications should be avoided, especially considering the higher complexity of the new data collection system, but quality checks need to be added that verify and review each in person application. It is also important to consider that applications from partner organizations accounted for the fewest number of clients, but had the second highest percentage of missing data. Perhaps better communication about data collection guidelines with partner organizations will help.

2. Data Understanding: a common understanding of the meaning of the data being collected will help reduce error in the collection phase. This property will benefit BetterFi internally, and make the application process easier to follow for the clients:

    a. Having a data dictionary or glossary will allow BetterFi to have a consistent document with formats and standards that would allow future data scientists to understand the information being collected. This can be a document that lists all

the variables being collected, their meaning, and why they might be relevant. This will also help with data consistency and lay an important foundation for data quality as betterfi expands and collects more data.

    b. Having extra text on the user interface that explains each question and lists example answers, using drop down menus instead of open input for basic questions, and having one question on the screen at a time will allow clients to understand which information is needed to be inputted and increase data quality

3. Validation: continuously verifying that the data being collected fits the criteria specified in the data dictionary for each input the applicant makes will restrict messy data from being entered and will lead to a higher data quality.

New Data Entries
Potential new entries to include on application
1. <u>Education level</u>: to show a client's highest level of education
2. <u>Marital Status</u>: this can include length of marital status as well
3. <u>Number of Dependents in Household</u>: could include age of children
4. <u>Number of Providers in Household</u>
5. <u>Employment Duration</u>: How long they have worked at their current job
6. <u>Number of Past Employments</u>
7. <u>Employment Location</u>: How much do they have to travel to go to work? How do they travel to work?

Dashboard Usage

1. Open up the dashboard file in R, make sure your working directory is set to the right file.
2. Click "Run App"
3. The Dashboard should now appear, if not click "show in new window" in the viewer.
4. You can input all the details you need into the dropdown boxes, and click "Generate Prediction"
5. There is now a percentage showing "Paid in full" and "Charged Off", which shows the predictability of whether or not the person will pay back. (Also, the "Paid in full" takes into account paid in full via refinance as paying in full).
6. For the data analytics tabs, you can click on the side tab and it will show the graphs

Model Usage

About the model:
To optimize interpretability and accuracy, a decision tree was implemented. The variables used to train the model were streetcity, streetzip, income_source, employer, annualincome, incomelevel, referenceperson, status_code, purpose, payment_periods and amount approved. Other variables used that were created by DataLab interns were incomeamt, which is the monthly income per customer, totalExpense, which is the sum of a client's housing, gas, car

insurance, electric, and water expenses, and monthlyPayments, the monthly payments towards the loan which is the quotient of amount_approved and payment_periods.

Our decision tree can correctly predict whether an applicant will pay in full or default on a loan in 86% of the cases. The operational cost of mistakenly giving a loan to someone who will default on it is higher than the cost of mistakenly denying a loan to someone would have repaid it, so we selected for models that minimized overall cost.

According to the model, five of the most important variables in the prediction process are the streetzip, annual income, employer, amount_approved, and the monthlyPayments. Overall, due to the limited training data, and especially defaulting data, the model does a better job at predicting loans that will be paid off compared to loans that will default.

Using the model:

1. Open R Studio, and make sure to have the 'app.R' file open.
2. Click on the 'Run' dropdown, then select 'Run All'
3. The dashboard will load, automatically showing the 'Predictive Model' tab
4. Input a client's information by filling in the text boxes and selecting the appropriate choices (calculations are required for incomeamt, totalExpense and monthlyPayments)
5. Click on 'Generate Prediction' to see the probabilities
6. The information will be saved onto a csv file on your laptop called 'newClientData.csv'

Future work:
Currently, the model will only generate a prediction with data that it has seen. Trying to input unknown categorical values will break the model, so the model has pre-set choices.

To improve the generalizability of the model, it would be best to turn payment_periods and amount_approved into buckets rather than factors and a numerical value, respectively. With the data quality guidelines, we're expecting to have more standardized data.

Map Usage

Using the map:
1. Open R-Studio, and make sure to have the 'mapshiny.Rmd' file open.
2. Install any missing packages, and library all packages. If a question asks about
3. Load in tn_county shapefile, maploan.csv
4. Select 'Run Document' at the top of the shiny document
5. The map should load, you can hover over the county and locations.