

Tilina Alzaben
The University of Maryland
Baltimore County

Elizabeth Morris
The University
of the South

Hamza Zia
The University
of the South

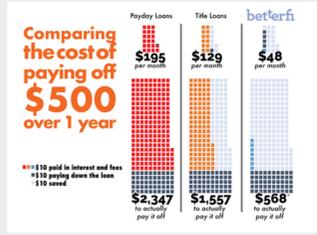
Mentor
Huarui Jing
Professor of
Finance

Partners
Spike Hosch
Executive Director
and Founder

Sam Shaw
Director of
Operations

Introduction

For millions of Americans with bad credit or no credit, taking out a loan means paying 400% or higher to a lender in both interest and fees, trapping people in a payday debt cycle for years. BetterFi is a 501(c)(3) non-profit economic justice enterprise that offers low-interest rates to help low-income households get out of debt. BetterFi is looking for a data-driven method to enhance their current lending decisions and scale their business.



Approach

We used BetterFi's historical data collected from previous loan approval decisions and applied machine learning methods to predict the likelihood that an individual will either pay off a loan or default in paying. We also included insightful graphs, and an educational map that highlights all predatory loan shops in counties that BetterFi currently operates in, along with their APR.

Data

Model and Insights

BetterFi has provided us with anonymized data, broken into three categories.



CLIENT INFORMATION

- Dataset collected from 85 clients
- Includes information such as a client's income amount, ethnicity, sex, age, employer, area code, street city, 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 including principle and interest

External Data

To create the educational map, a dataset including the location of predatory lenders and regional geographic data was used. Data from the U.S. Census Bureau and the Housing of Urban Development Department was also utilized to obtain ranges of income level per county.

Method

Baseline

BetterFi's current decision process is highly labor intensive, and their average default rate is about 16% of all loans to date.

Outcome of Interest

Our model generates the probability of an individual either paying a loan in full or defaulting on in.

Model Summary

To optimize precision 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.

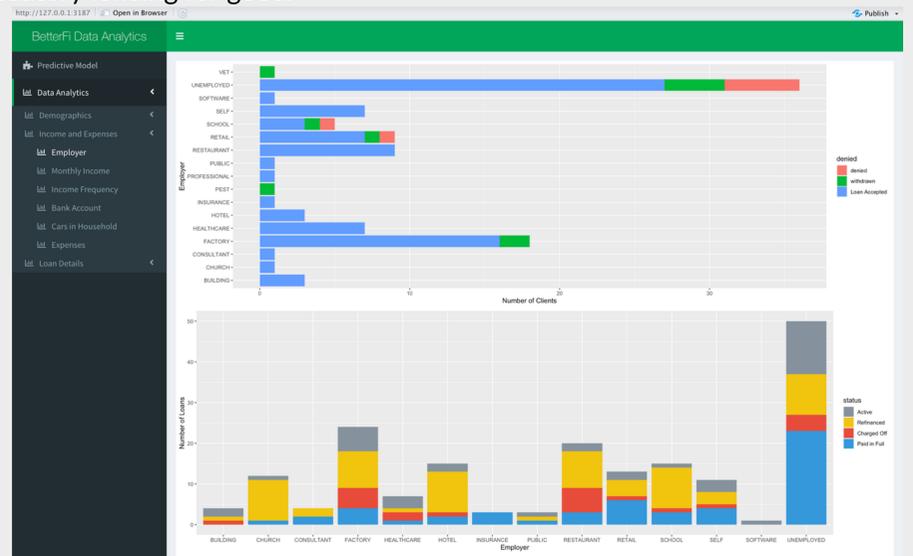
Results

Internal Dashboard: Predictive Model and Insights

Our decision tree can correctly predict whether an applicant will pay in full or default on a loan in 86% of the cases, comparable to the current BetterFi rate of 84%. 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.

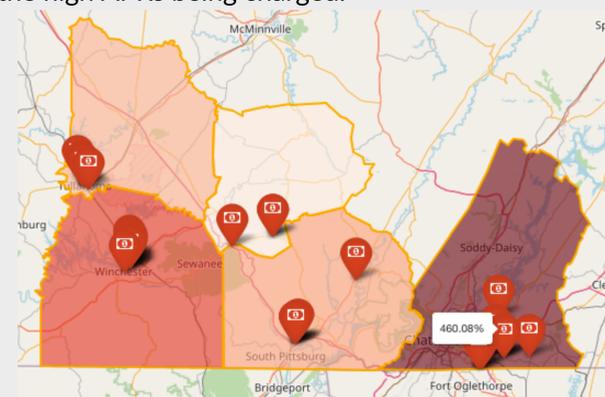
According to the model, five of the most important variables in the prediction process are the client's street zip, annual income, employer, total amount of loan received, and the amount of monthly payments. Overall, due to the limited training data, and especially defaulting data, the model does a better job at predicting loans that will be paid off compared to loans that will default.

Since the data is sensitive, the dashboard that contains both the model and analysis of BetterFi's data will only be available to our partners. Our analysis provides insights into BetterFi's historical data and repayment patterns, to help them understand their clients better and be able to offer more loans to the people who both need them and are able to repay them in full. This will allow BetterFi to scale their business and give out more loans, helping combat predatory lending for good.



External Dashboard

We also created a map for informational purposes, which will allow consumers to identify predatory payday and title loan shops located in their neighborhoods, as well as look up the high APRs being charged.



Impact

BetterFi

With the model and analysis provided, BetterFi will be able to supplement and scale their current decision making process, while maintaining their low default rate. Their goal is to continue expanding into more communities where there is a need for their services, so they can continue pulling more people out of the cycle of debt.

Community

The educational dashboard and map highlight the loan shops with predatory APRs and aim to raise awareness of the issue not only among those possibly considering such loans, but also among the general public.

We also hope it could inspire more people and organizations to fight predatory lending and pave the way to new policies being created, so that more low-income households can escape debt and more communities can be uplifted.