Solution S49^x Technical Assesement 2

Dr. Jennifer Piscionere

What are the questions?

What's the best predictor for loan defaults?

Identify at risk loans and give incentives for customers to repay loan.

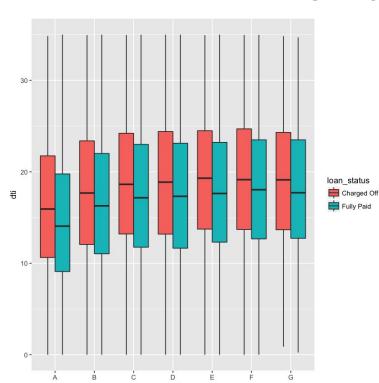
Late payment forgiveness

Reduction of interest rates

Outline of approach

- We have historical data on which loans were paid off and which ones were charged-off.
- Create a model that quantifies likelihood of charge-off based on financial factors
 - Basic machine learning algorithm
- Apply that model to current ('17) data to see which customers are at risk of not repaying loan

The Data

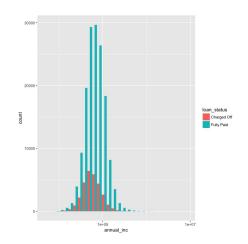


We are using the 2012-2013 data set as a training set because it has more customers.

Interesting columns:

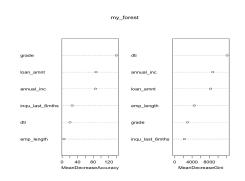
- DTI
- Credit Inquiries past 6 months
- Delinquencies

Interesting tidbits



- Most people are using their loan as debt consolidation
- All loan classes are charged off at a roughly equal rate
- The mean DTI in the rejected group is higher, but the median is about the same
- We have credit score in the rejected group but not the approved.

Taking a crack at a ML approach



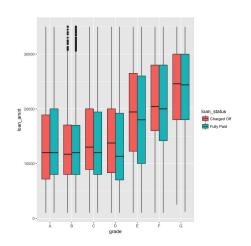
- In an ideal world, this approach would help to narrow down the variables that most strongly correlate with loan status.
- Since this is not an ideal world, I cherry picked some variables to work with

Loan status = F(loan_amnt + grade + annual_inc +dti + inqu_last_6mths + emp_length)

Building a better future model

- Computational power limited the scope of the ML
 - The next step is to apply it to the current data
 - Clean out all the na's
- Why didn't the decision tree work?
 - Investigate how random loan repayment failure is
- Taking a closer look at the rejected loan applications

To the client...?



It looks like you're doing a good job limiting your exposure to risk across all loan grades. However, you might want to take a closer look at a client's DTI in higher loan classes and offer incentives for loan repayment.