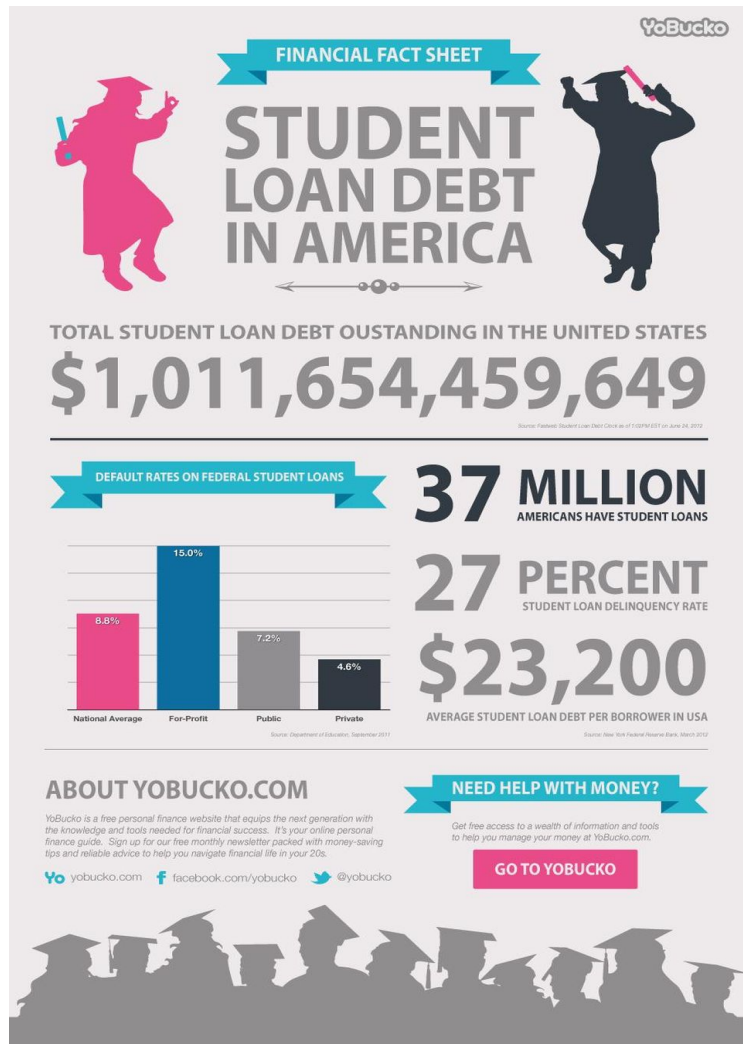


Final Project Prestation Student Loan Debt Analysis

Kelly Chen

Background Information

Affordability of student loans is a rising topic in the United States in recent years. Many argue that it has jeopardized many families' financial situation.

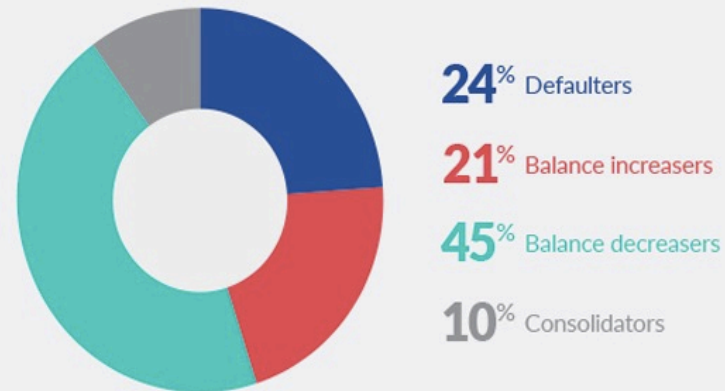


The Categories of Texas Borrowers

Figure 1

Texas Student Loan Borrowers Divided Into Repayment Categories

Borrowers by loan status after five years in repayment



Note: This analysis followed borrowers for five years from the end of their six-month grace periods. In addition to defaulters, balance increasers, and balance decreasees, the Texas dataset included borrowers who consolidated their loans within five years of entering repayment.

Source: The Trellis Company's administrative data

- Skyrocketing student debts doesn't seem to stop families apply for student loan. With more families put great emphasis on education, and supportive policy, many families are willing to take a risk. In fact, a survey showed that 90% of private student loans now have an adult co-signer.
-
- However,
- Do families have an clear idea about their financial situation to take a student loan?
- Do they really know how the educational loan is going to affect their finance ?

Goal of the analysis

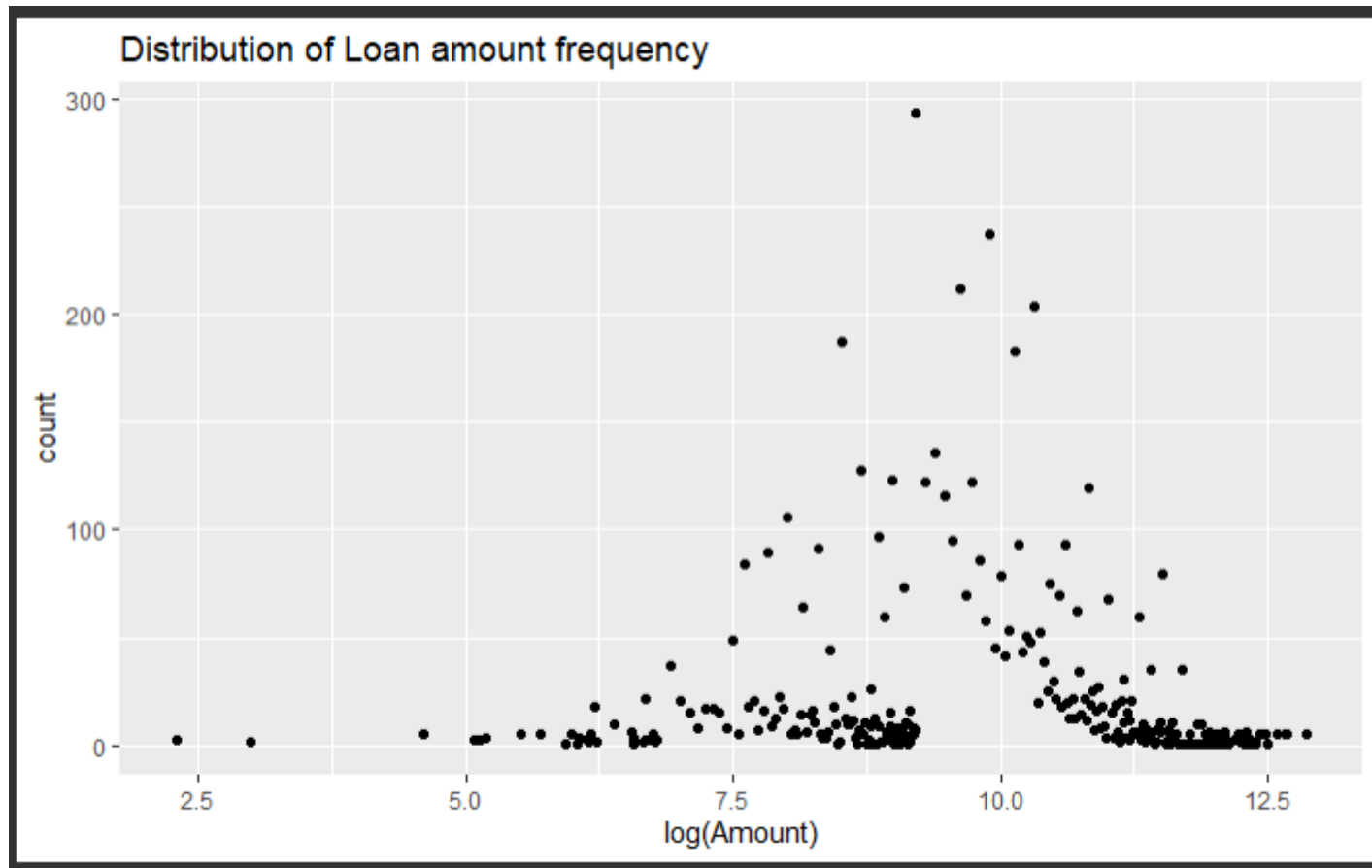
- Determine which characteristics of a family have correlations with education loan amount if they decide to apply for a loan.
- Determine how amount of education loan can exacerbate financial struggle of a family.

Predicting student loan

Data wrangling

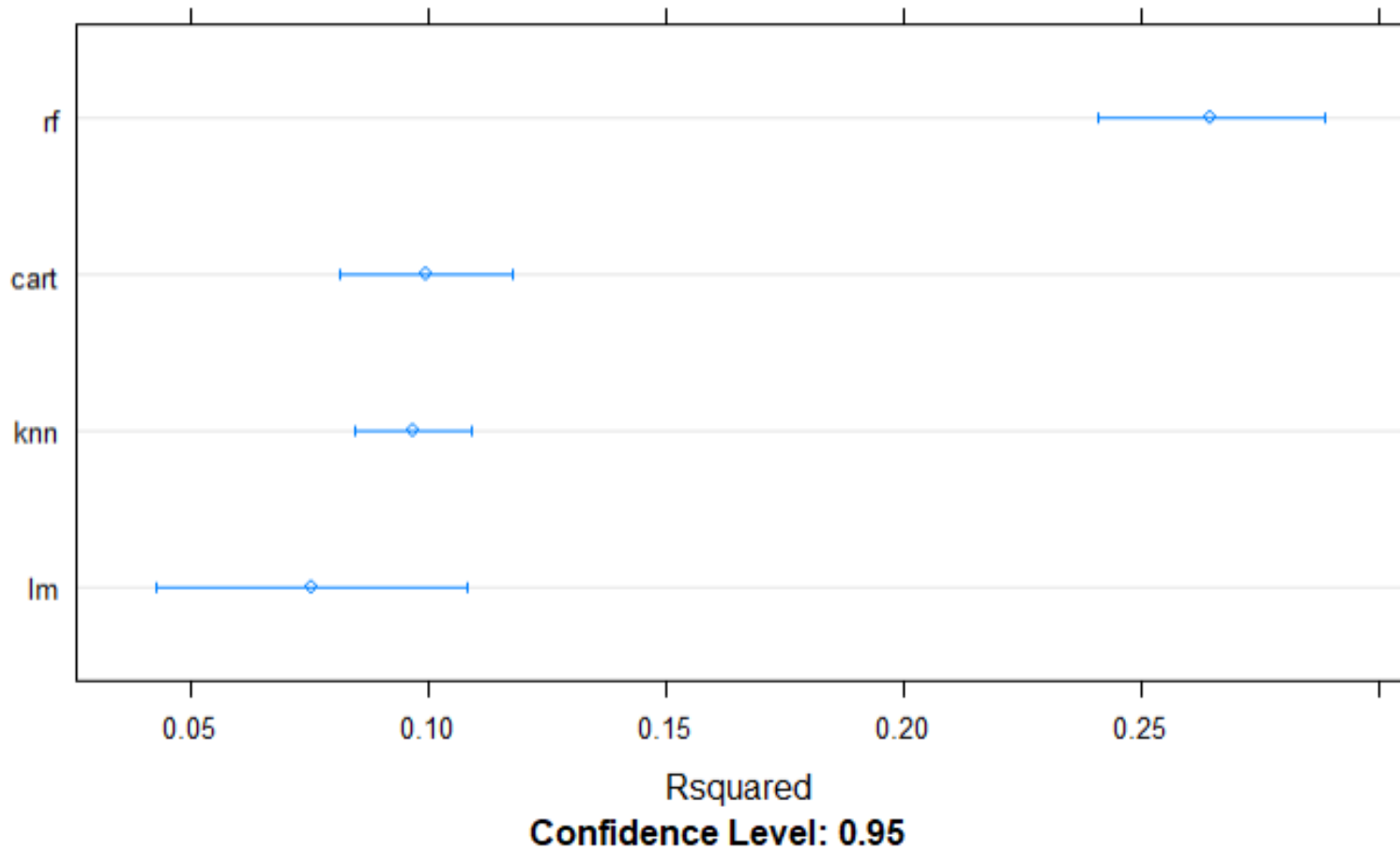
X7489	Confidence in US economy
X7556	Do you think you're knowledgeable about your own finances (-1 to 10)
X7557	Are you willing to take financial risks? (-1 to 10)
X1032	Do you owe money to purchase property?
X6111	Highest education received
X7660	Do you have any trust fund?
X3506	How much is on the main bank account?
X501	What kind of house?
X5702	In total, what was your (family's) annual income from wages and salaries in 2015, before deductions for taxes and anything else?
X6809	Race
X4001	Do you have life insurance?
X3526	How much is on the saving account?
X3721	What is the total \$ value of your certificate of deposit
X3902	Total value of saving bonds
X8022	Age
X3822	Total market value of stock neutral funds
X8021	Gender
X8023	Marital status
X5931	Highest received education/ degree
X7824	How much is still owed on this loan?
X7179	How much in total is owed on all the remaining loans?

Data visualization

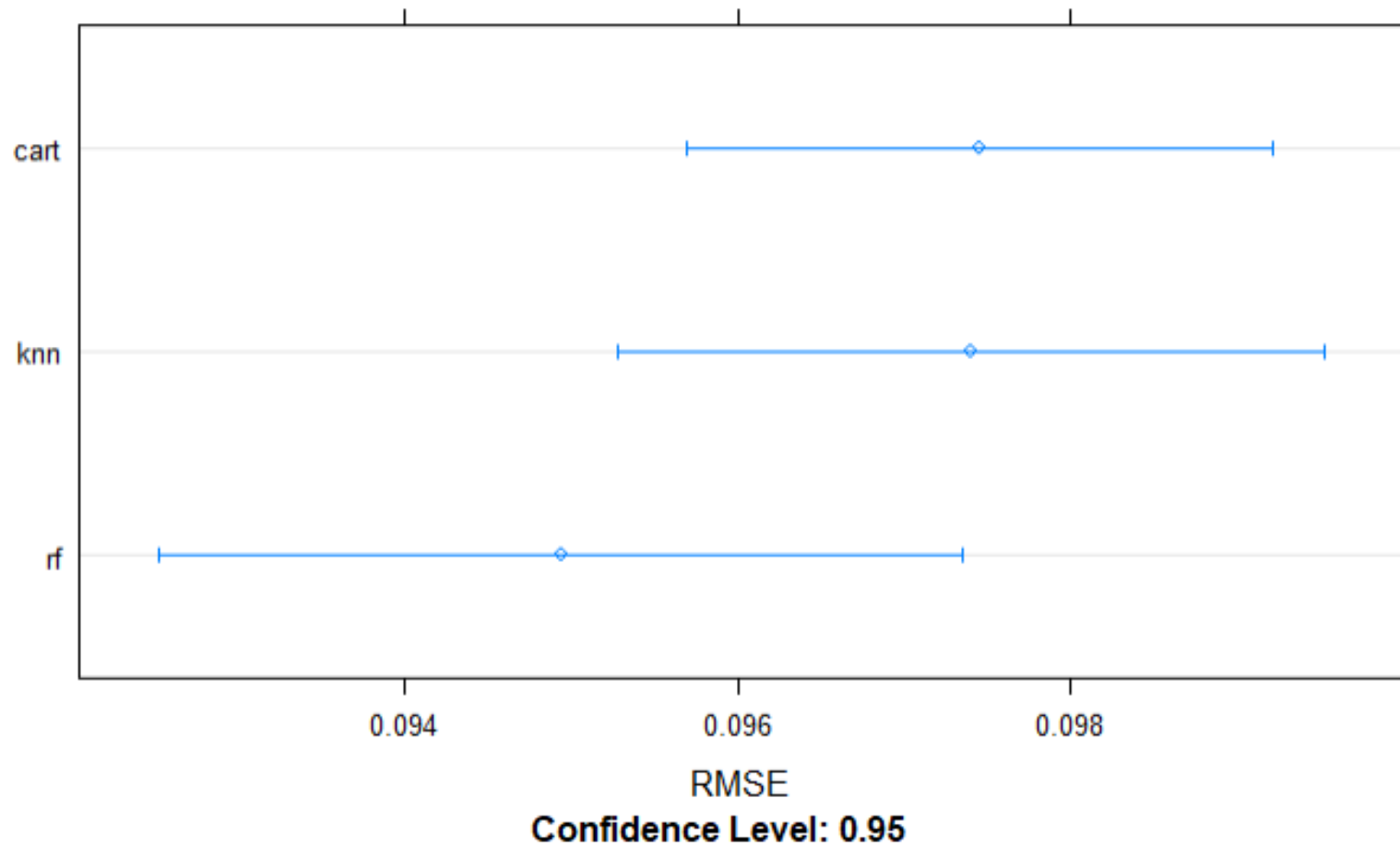


Loan amount-frequency distribution

Model comparison



Mean Squared Error



Mean Squared Error

```
## {r}  
pred <- predict(mod_rf_1, newdata = testing2)  
mse = sum(testing2$Loan_amount - pred^2) / nrow(testing2)  
mse  
##
```

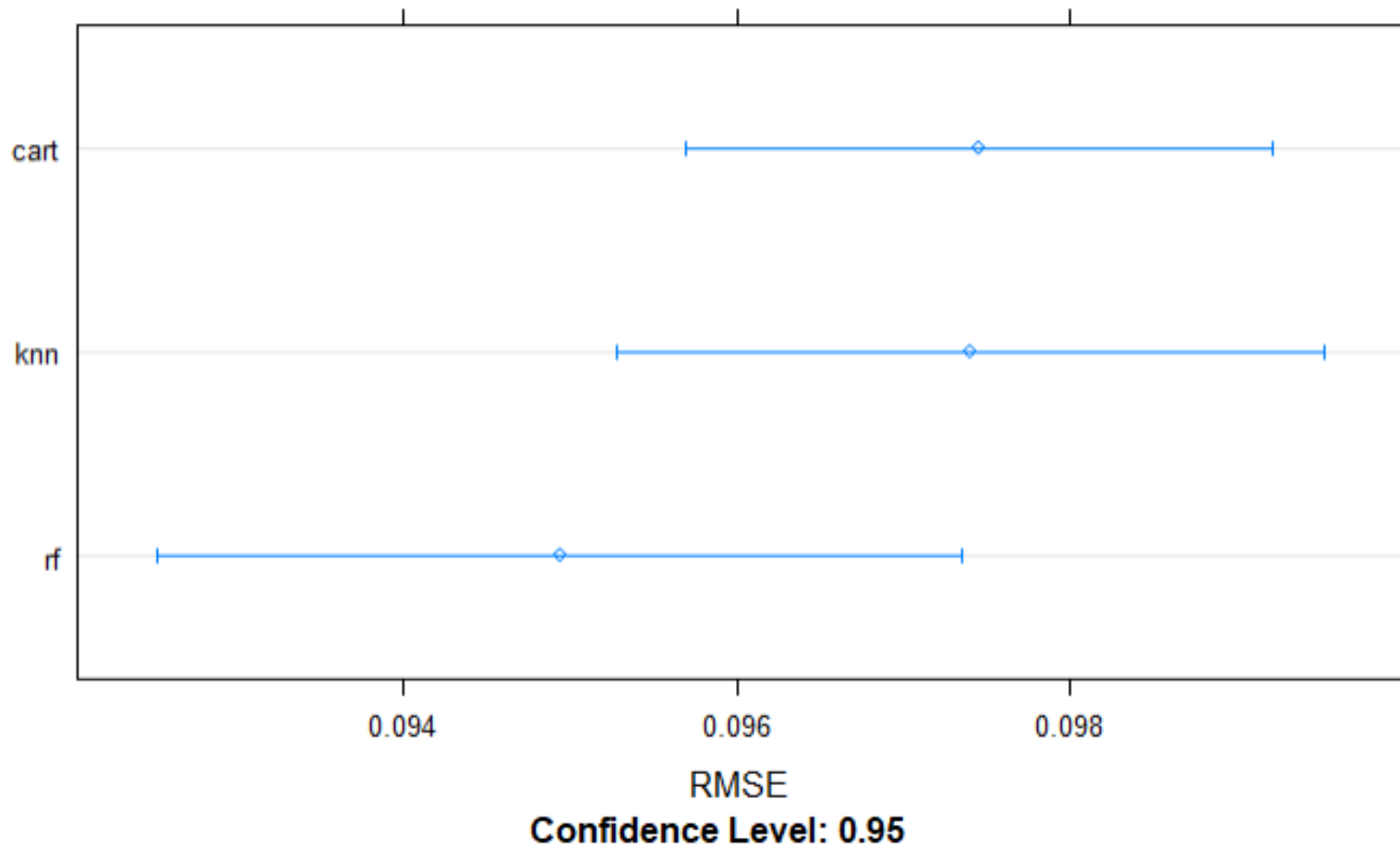
```
[1] 0.06188924
```

Student loan on financial struggle

Data wrangling

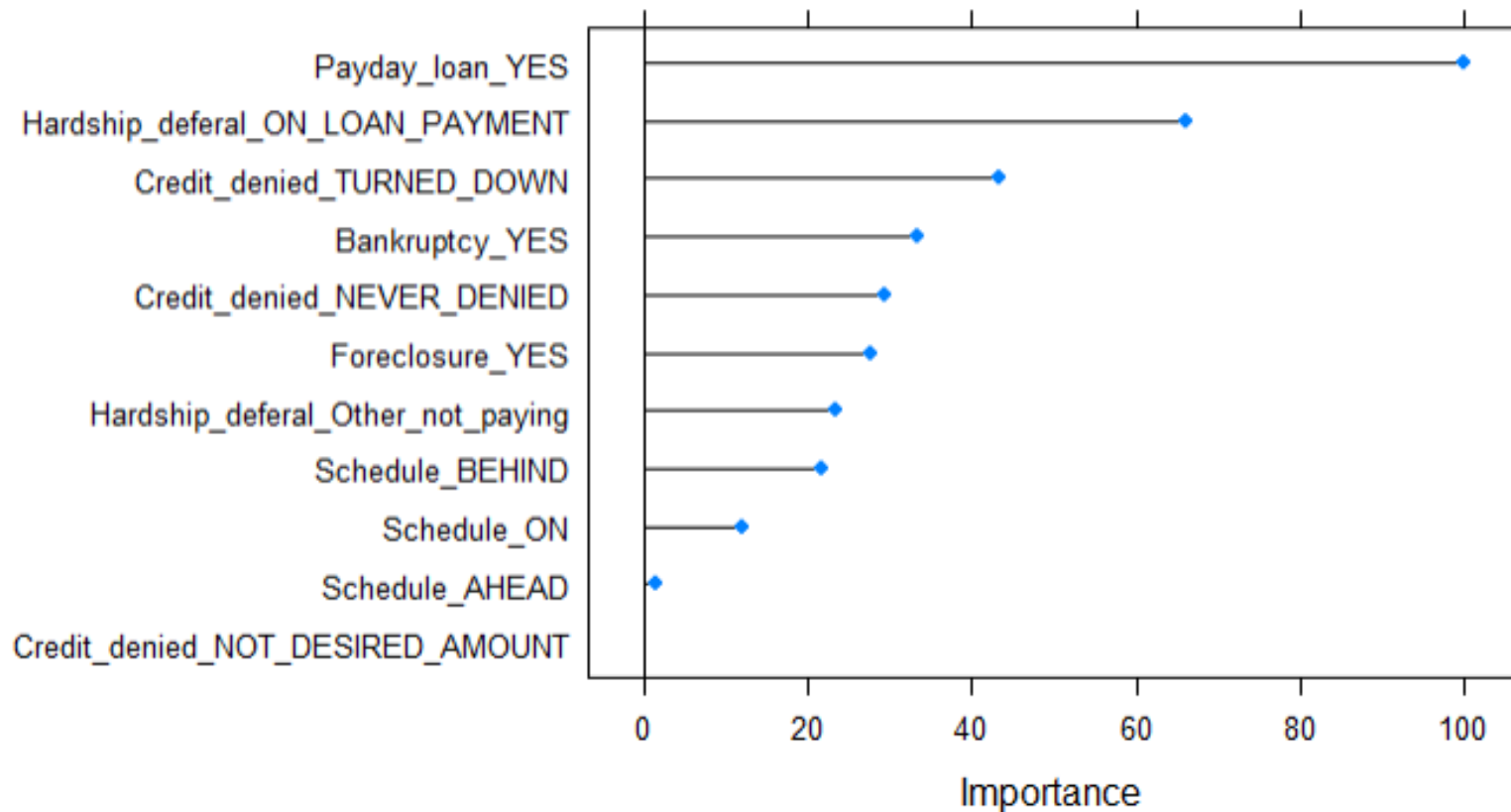
Variable	Meaning
X407	Credit denied?
X7521	Late debt payment?
X7063	Payday loan?
X6772	Bankruptcy?
X3031	Foreclosure?

Model comparison



Importance

**Variable importance for
the amount of student loan**



Challenges

- Finding the desired information among 5000+ variables.
- In Data wrangling process, the translation of the values of each variables.
- Ideas for data visualization