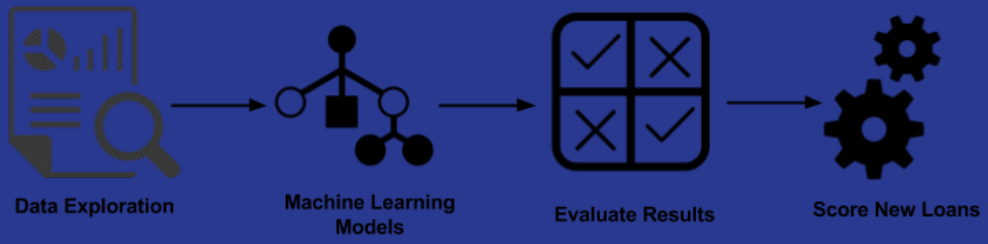
AI Based Risk Assessment

Lending Club



Torben Holst, Arne Scheunemann 3/2020

Agenda

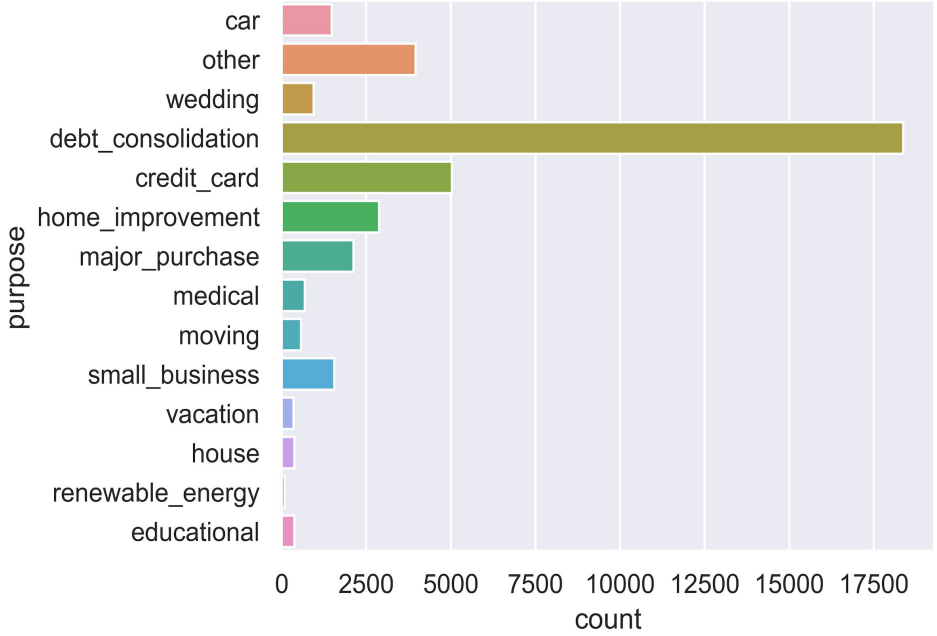
1. About Peer-to-Peer-lending
2. About Lending Clubs Business in 2007-2011
3. Charge-Off Risk: Explaining the Model
4. Conclusion & Future Work

About Peer-to-Peer

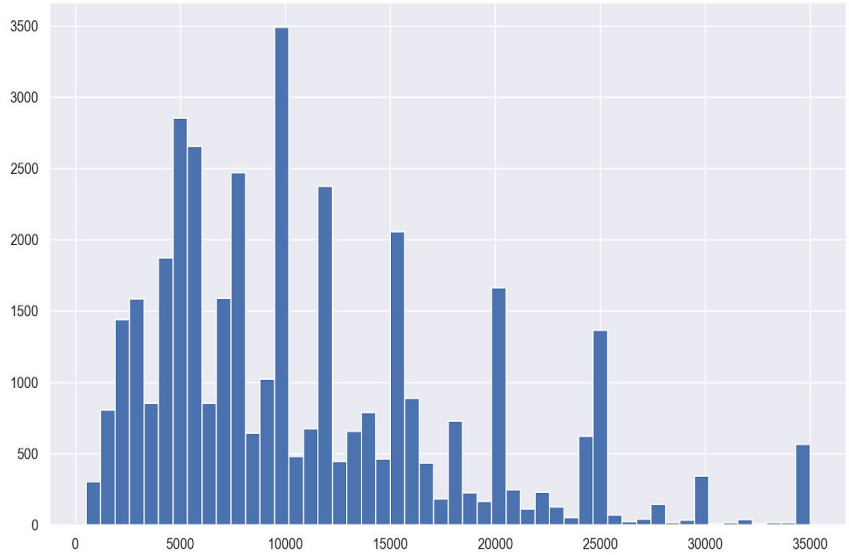
About Lending Club Borrowers

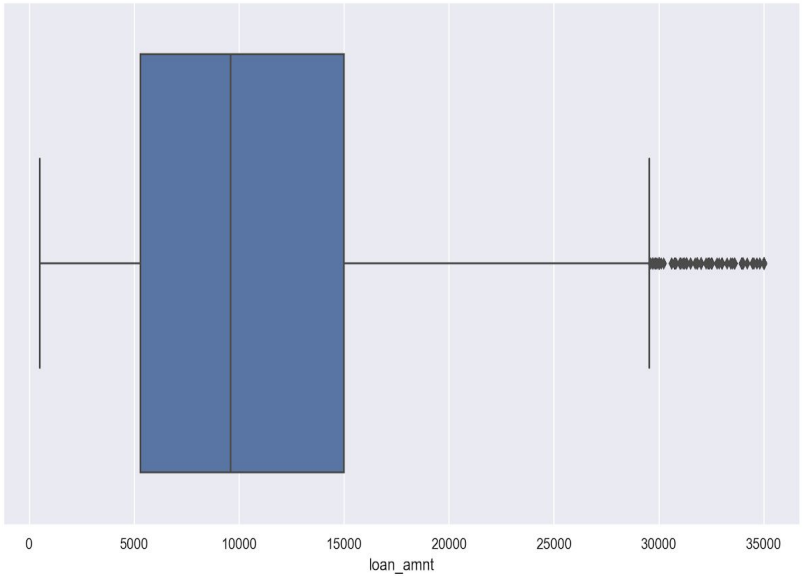
What can we say about Customer

- often low credit rating
- for debt consolidation and credit card
- need loan fast



What Lending Club offers



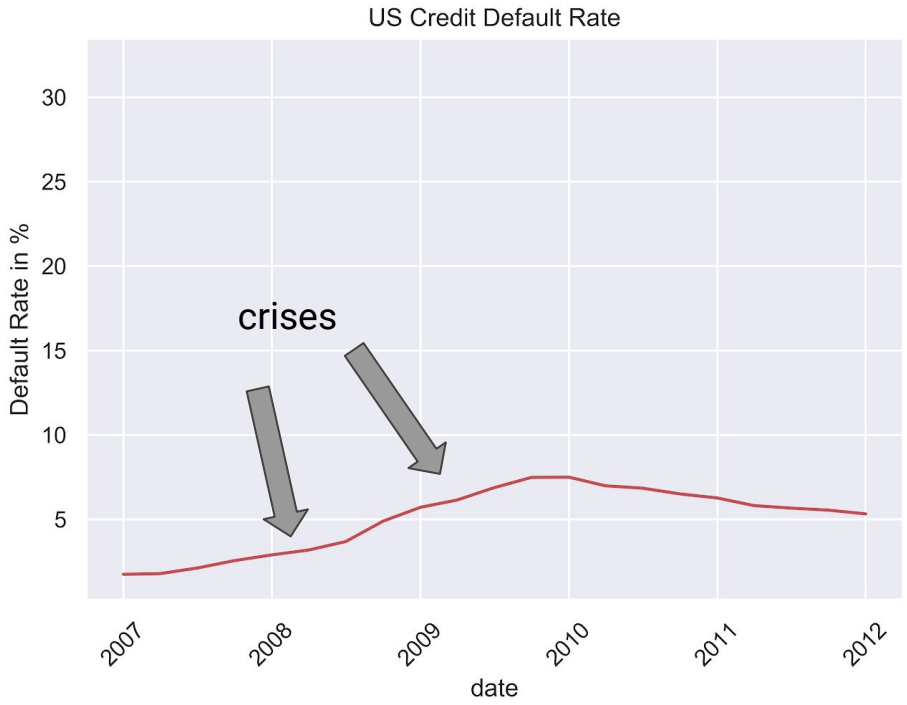


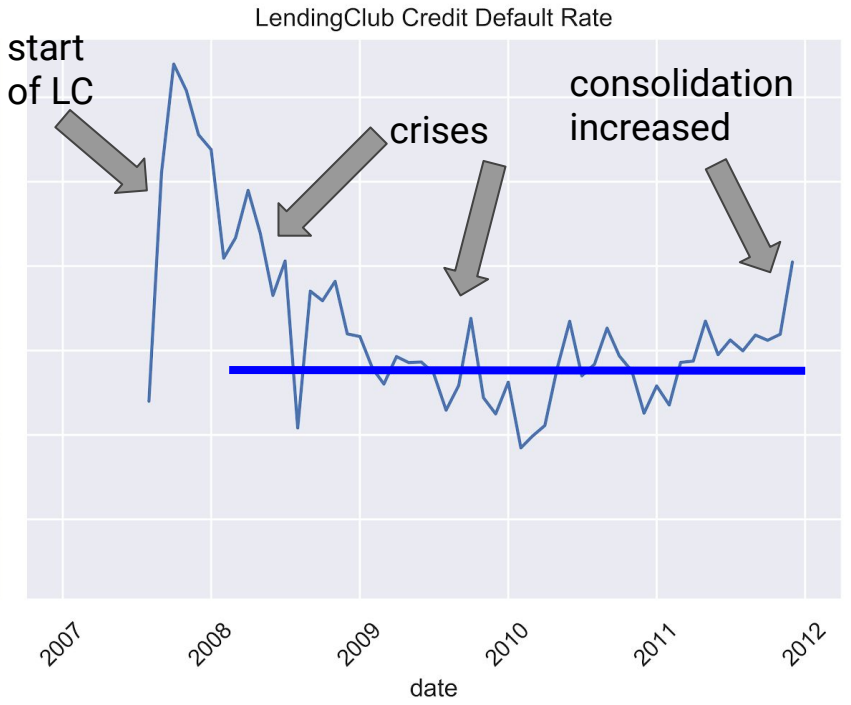
Loan Amount: 500 $ - 40,000 $

Development of Lending Club Loans

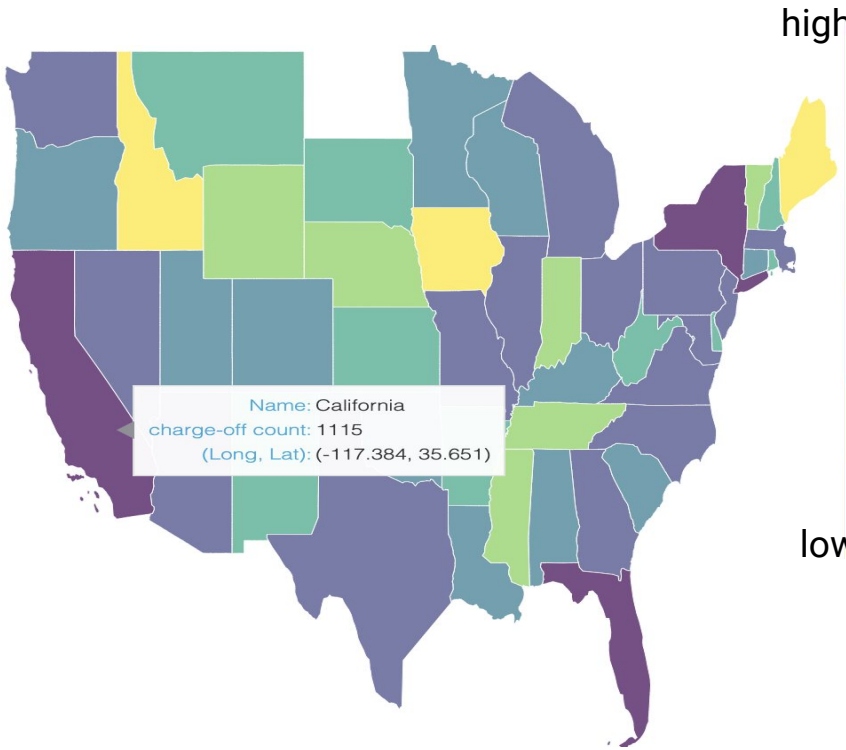
→ **Demand is high so we need to attract the investors!**

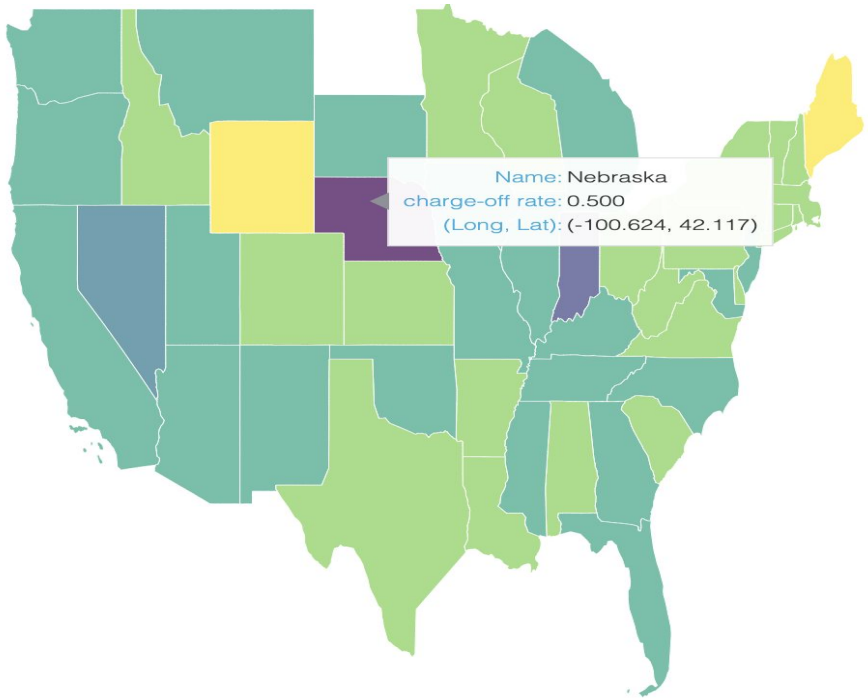
US Credit Default vs. Lending Club Default



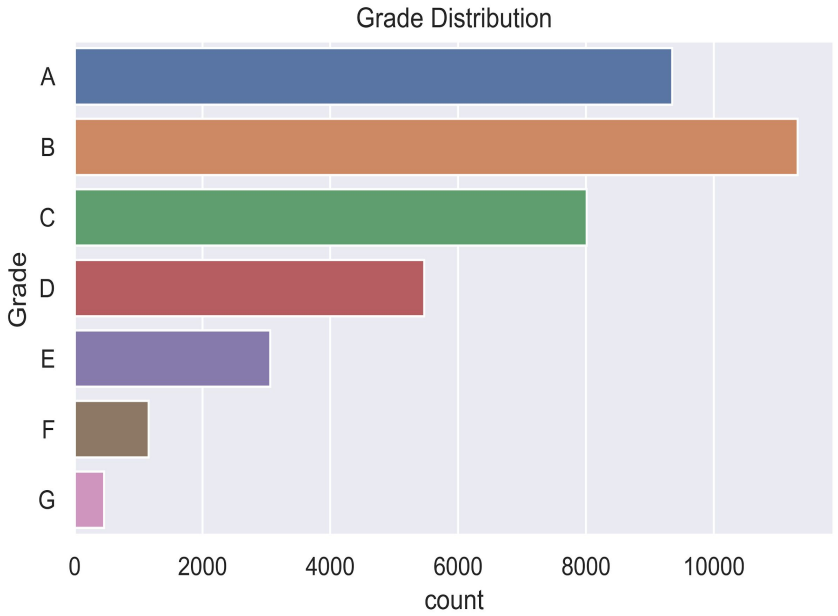


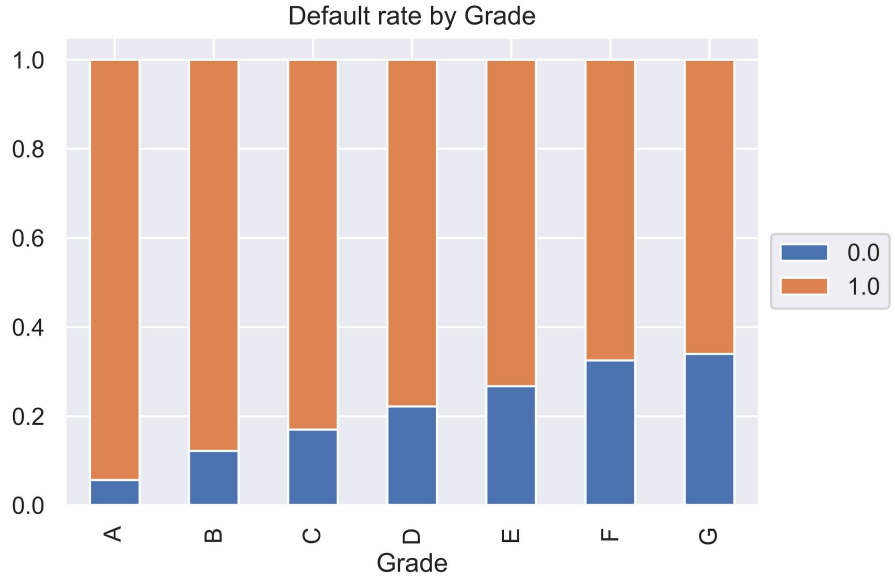
Default Count vs. Default Rate



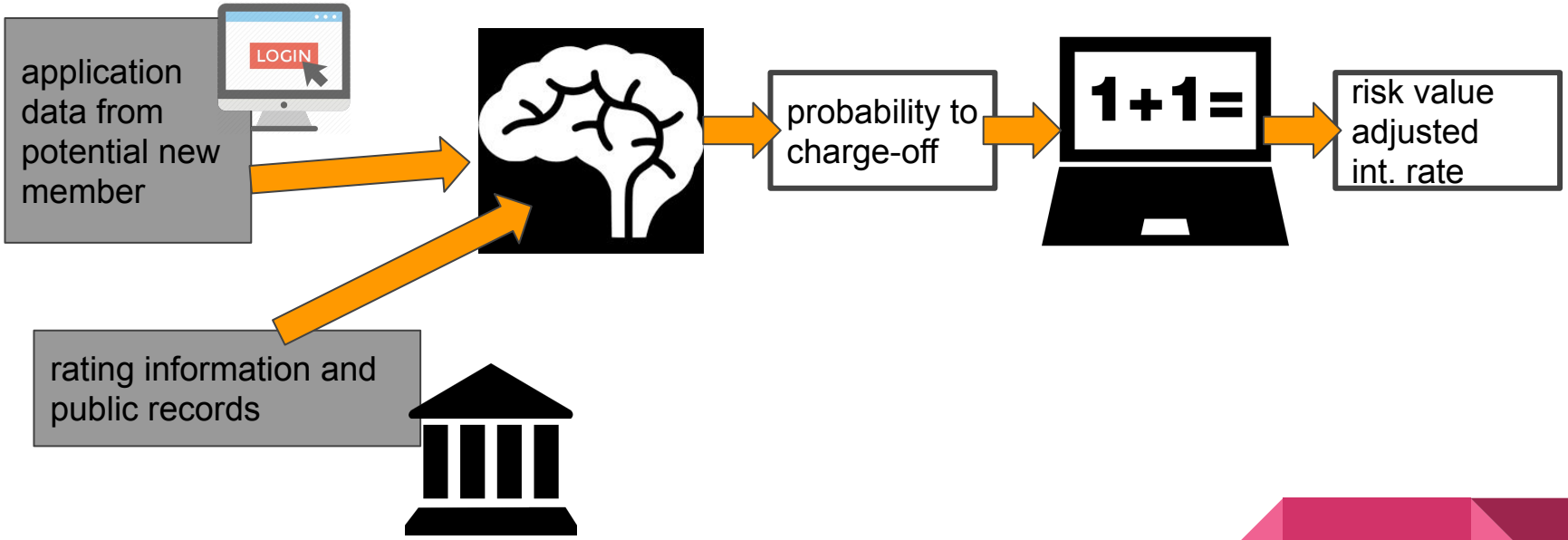


Lending Club Grade System





Prediction Model - Model use



Prediction Model - Focus

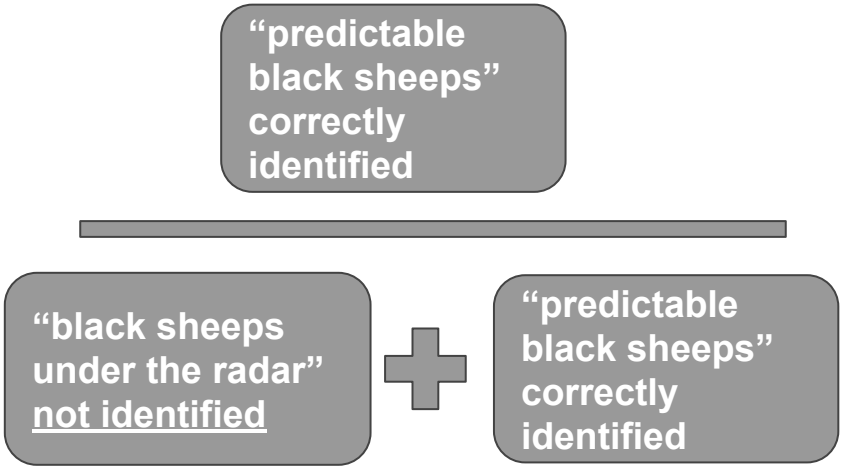
What is important - getting the money into the system

How to achieve - convincing investors that we can do good credit scoring

some borrowers can be lost demand is high

	AI Prediction Model	
	towards 0% - low risk prediction	towards 100% high risk prediction
recorded as actually **fully paid**	good borrowers and approved	lost borrowers
recorded as actually **charge-off (default or very late)**	**“black sheeps under the radar” not identified**	**“predictable black sheeps” correctly identified**

Prediction Model - Metric Recall

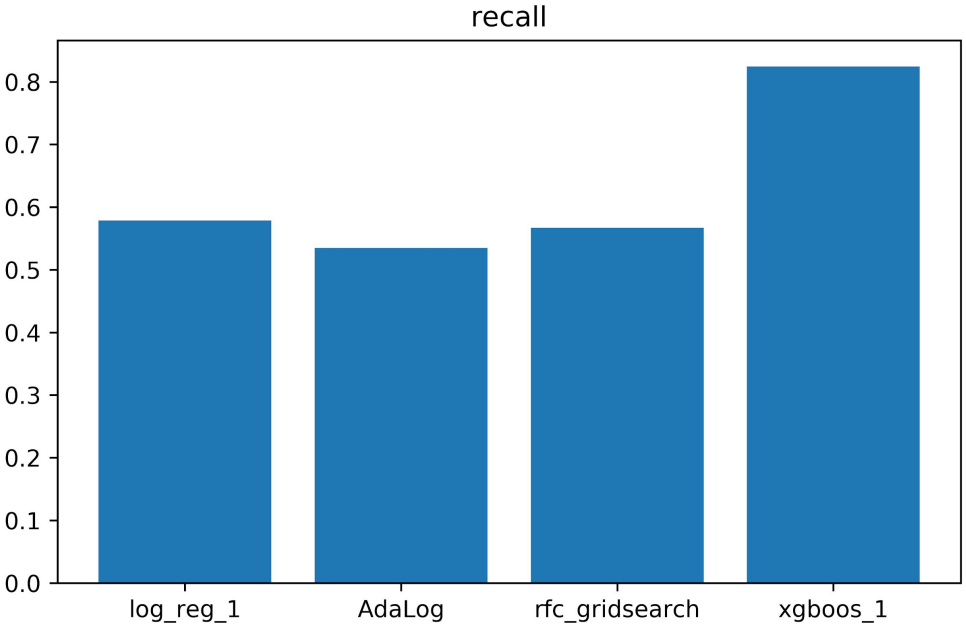


Best Model gets to score of 83%
this comes with penalty of lost borrowers

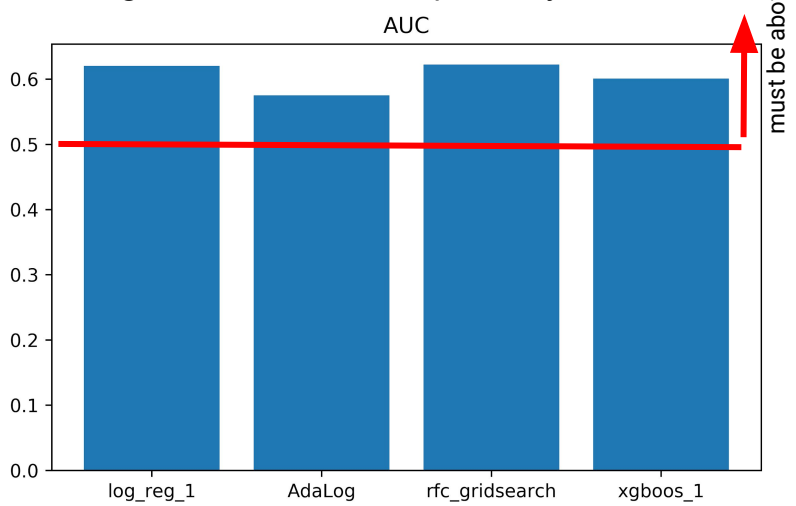
The model is based on modern Machine Learning algorithm :

XGBoost

Prediction Model - Variations / Status



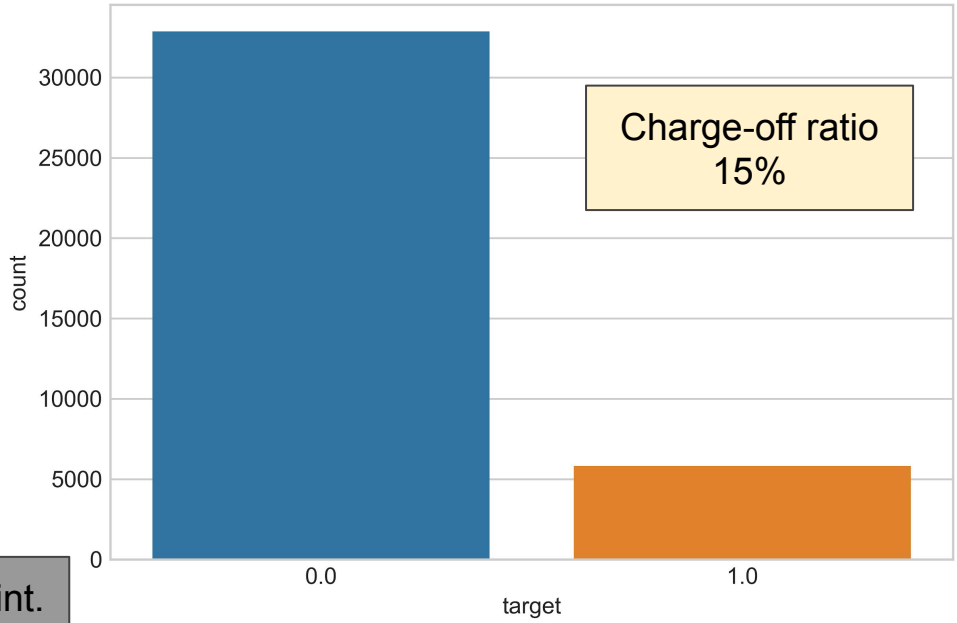
degree or measure of separability:



Prediction Model - Difficulties when training

Distribution the information about charged-off are naturally low in contrast to the sum of data points.

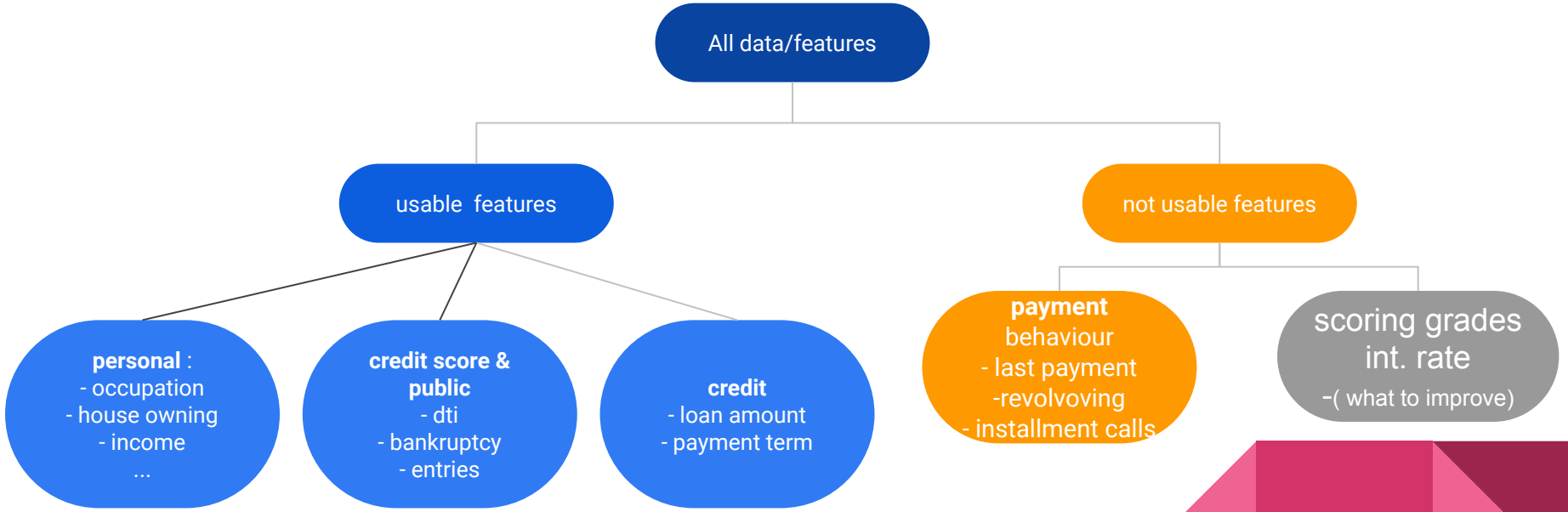
- we can prognose a risk score from the model per borrower
- the risk score then can be combined with the knowledge of the charge-off rate



Adjusted int. Ratio

Prediction Model - what can be used

for express prediction of new borrower:

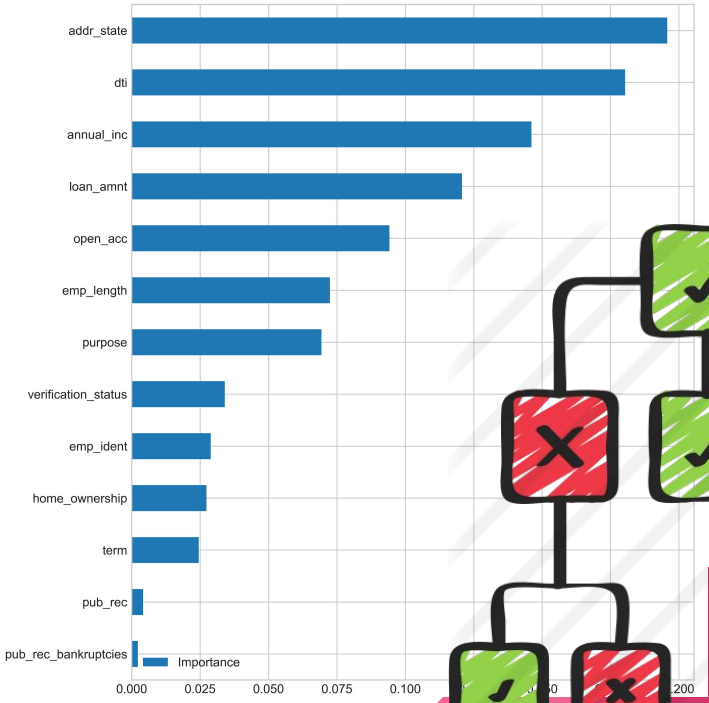


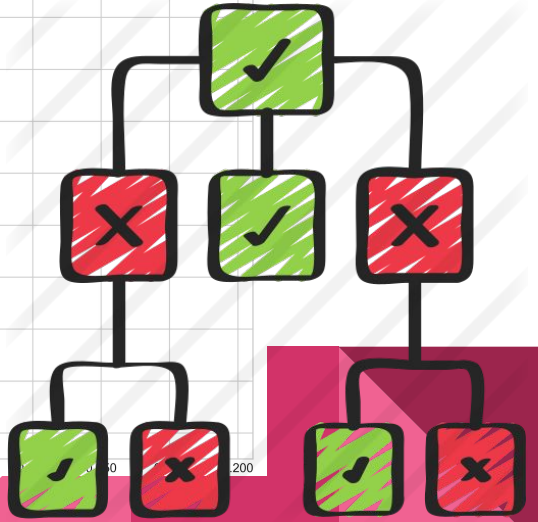
Prediction Model - Input Information value

many small hints form a picture ...

computer modeled decision tree - can determine the importance of available input for making prediction:

- **location** (personal - verification easy)
- **debt to income ratio** (credit score / public)
- **income** (personal - verification harder)
- **loan size / amount** (credit)
- **employment length** (personal)
- **all open credit accounts** (credit score / public





Conclusion

The business **is growing** - there are **many borrowers**
to keep **money flowing -> investor need security** on borrowers that are mostly in consolidation

<u>We think reliable express scoring can be done !</u>

- correctly finding “black Sheeps” recall now = 83%
- default rate in the population can be used
- more training time and tuning can improve the penalty
  penalty = potential to high int. rates or rejected good borrowers

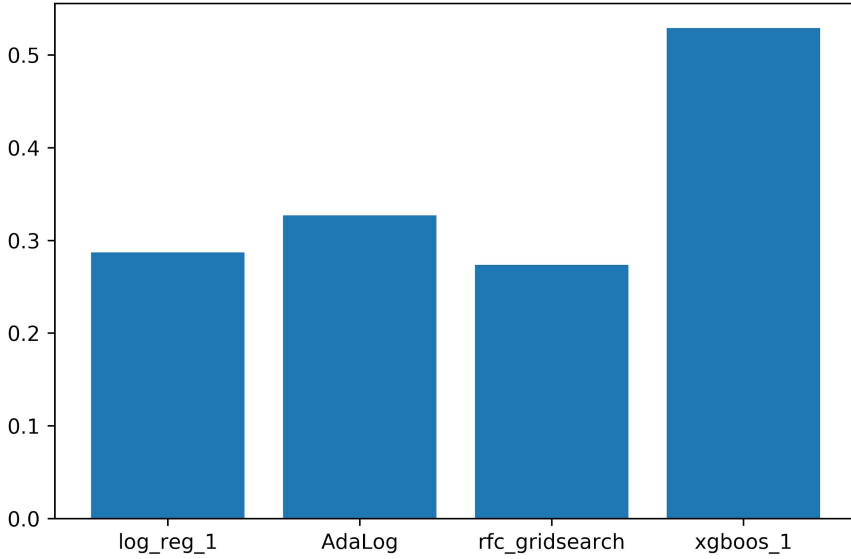
- we know top influencing data and their origin
  - can be used to define effort to make evaluation more express..

Future Work - Prediction Model - member loss

The downside of the model that we were able to create is that the model is

a little to suspicious

here test result for lost borrowers rate



- we are on good path, but more time needed to improve model
- We have also hope that more incoming data can compensate (make the model learn more)
- adding more input features need to be explored

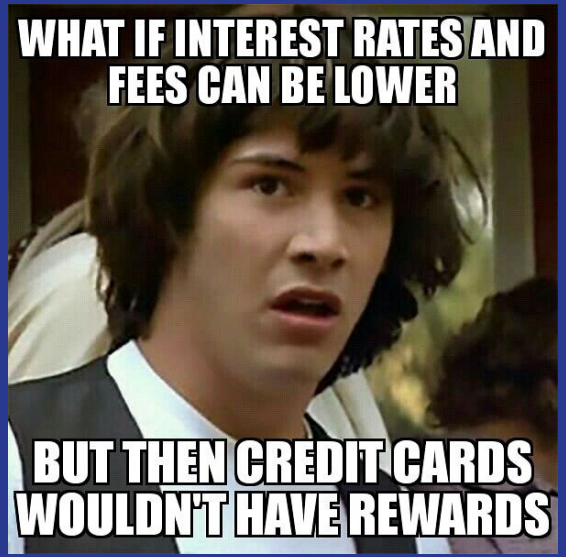
Future Work - post process model output

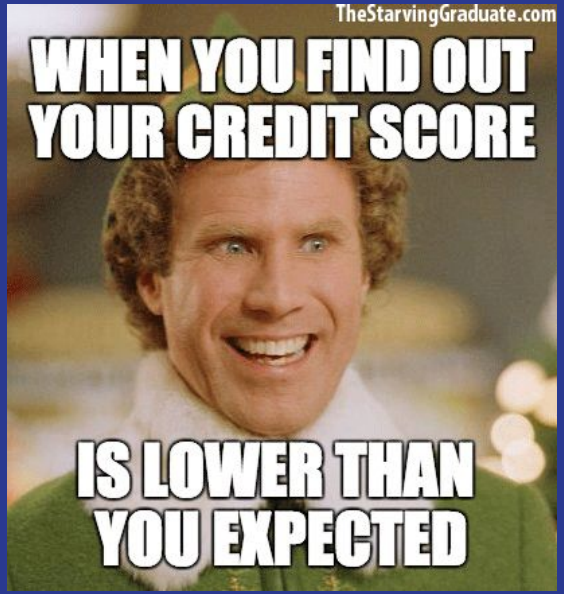
- Compensate the model Charge-off probability by the entropy
  - meaning to involve information about the distribution in population
  - - > way lesser charge-off then payed off (only not good for the model training)
- Calculate the interest rate and combine with investment strategies
- Compare the grade system to the new scoring system
- Overall risk of loss in correlation with the money in the platform
  - calculate a risk in $
- Risk compensation based on mean last default rate change
  - in example last 3 Month
- Two stage express prediction (fast borrower experience)
  - with and without verification

Future Work - data preprocessing

- Employment title processing can be improved
  - currently simple text compare and to little amount of classes
    - public - like state, county, school, sheriff
    - army, self employment, bank, …
    - known company (here size and position matter …)
- location details can be more detailed
  - for now the zip code data has not been processed completely
  - state location could be improved by using population density in account
- historical data of the transaction can be used to form digital twins
  - recovery rate
  - payment periods
  - installment collection call count ...
  - - - > sell the data -- use to correct default rate base -- evaluate risk in system

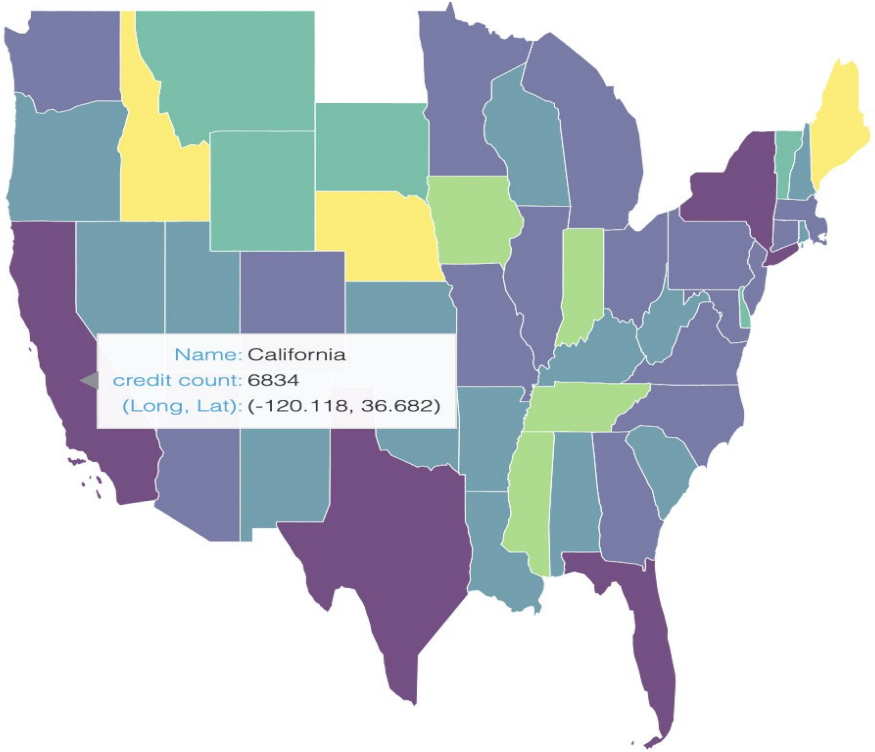
Thank you for your attention

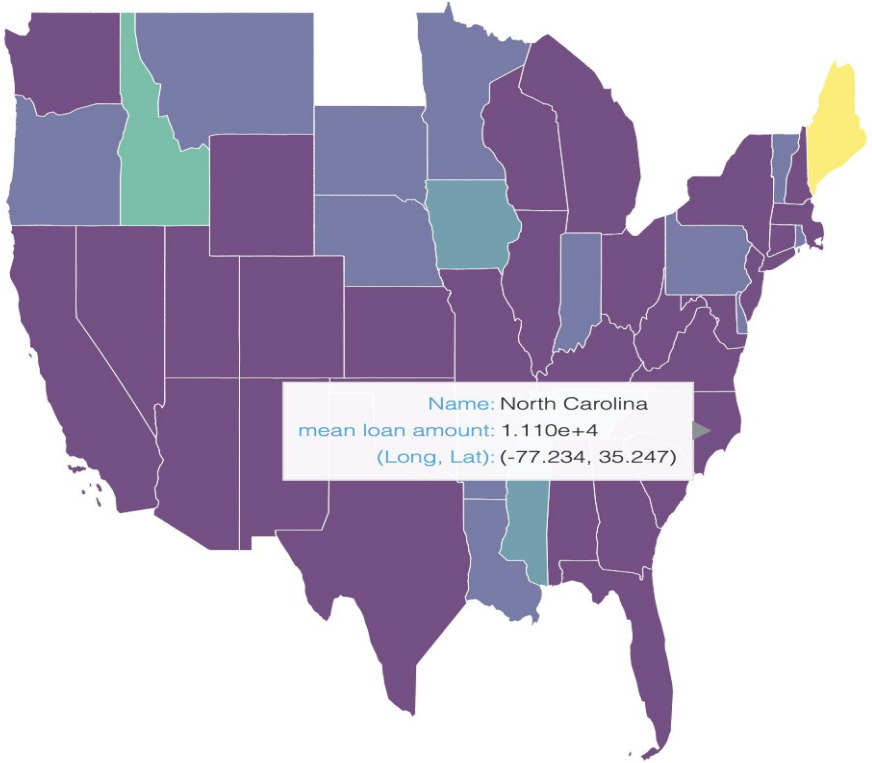




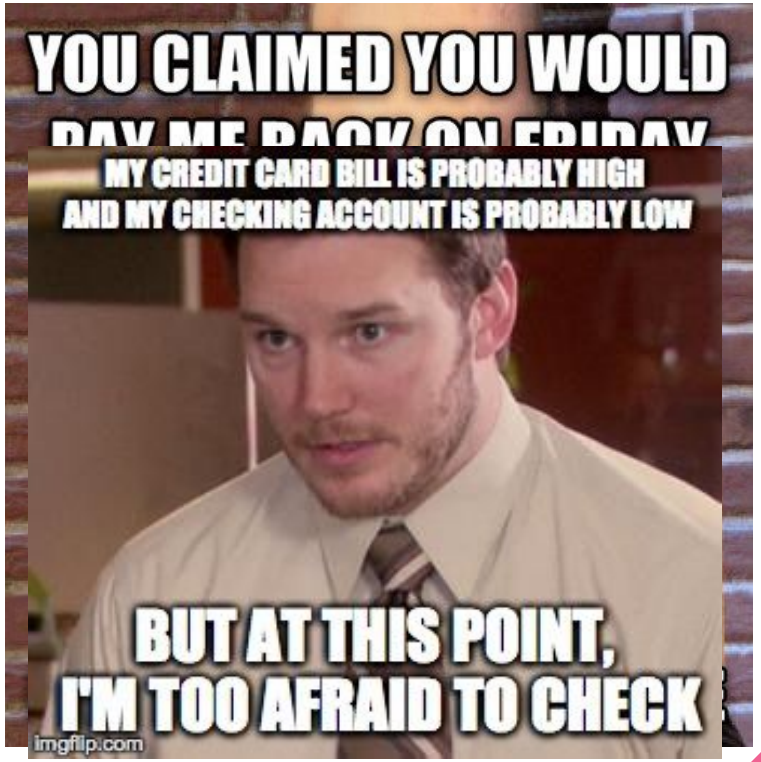
Backup

Credit Count vs. mean loan amount






LOAN - LOAN - LOAN
YOU ALL GET LOANS!!!
memegenerator.net


YOU CLAIMED YOU WOULD
MY CREDIT CARD BILL IS PROBABLY HIGH
AND MY CHECKING ACCOUNT IS PROBABLY LOW
BUT AT THIS POINT,
I'M TOO AFRAID TO CHECK
imgflip.com


LOW CREDIT? NO CREDIT?
NO PROBLEM!

Background rise of Peer-to-Peer Lending

- in the US starting in 2006
- result of the financial crisis of in late 2000s:
  - 2007-2009 Housing bubble leads to severe financial crisis
  - 2008 housing and Economic Recovery Act
  - 2009 Dow Jones hit low
- Banks very reluctant on providing loans
- very low interest rates for investors and savers

→ Connect People who need money with people who are interested in higher interest rates

1. crisis-related
2. competition-related
3. Internet-related