



# Loan Data

*Sachin Arora*

Predicting Bad Loans

“A bank is a place that will lend  
you money if you can prove that  
you don’t need it.”

Bob Hope

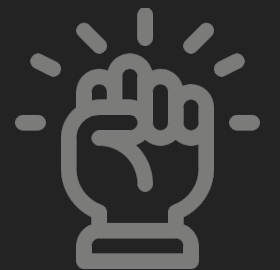


background

motivation

Peer to Peer Lending

investor tool



Identify loans in default status  
with 80% accuracy!

Problem Statement



850K+  
observations

74  
features

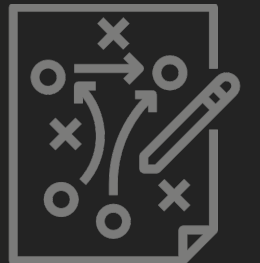
*Getting Familiar with the Data!*

1219  
defaults



Identify a loan as a good or a bad loan with 70% accuracy!

Plan B



fully paid

default  
charged-off

current  
delinquent

Simplify



inflation

unemployment

interest rate

Augment with External Data





correlation  
(w/ target)

domain know  
how

independent  
correlation

Feature Selection



defaults

99.52%

99.79%

logistic  
regression

bad loans

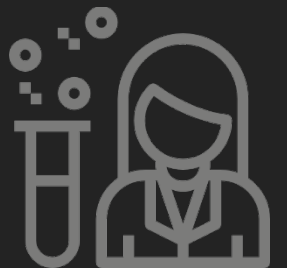
~92%

*Modelling. Performance. Prediction Accuracy.*

real-world accuracy

defaults 99.9988%

bad loans 97.62%



different  
algorithm

feature set  
optimisation

cloud  
deployment

Next Steps

time series  
shift



questions

demo

Demo



Thank You!

