IS LENDERS IN DANGER?

Genevieve Vallee
Department of Economics
Genevieve.vallee@carleton.ca

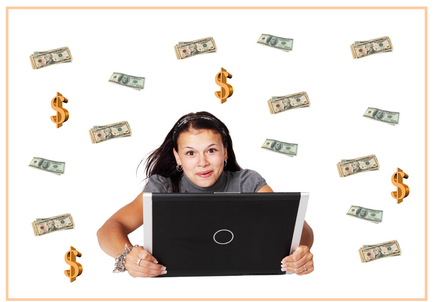
Siyu Zhou
School of Computer Science
siyuzhou3@carleton.ca

Motivation

With the growing popularity of peer-to-peer (P2P) lending and the availability of the data from the largest provider of this type of lending, we aim to determine the probability that current recipients of loans from crowd-funding sources will payback or default on their loans within one year from receiving the loan.

Knowing if current holders of loans are more likely to payback their loans early could provide lenders with information to maximize their returns and minimize their losses by take measures to protect themselves.

Our model can be used to do this prediction with high accuracy!



Data

Using data from Lending Club

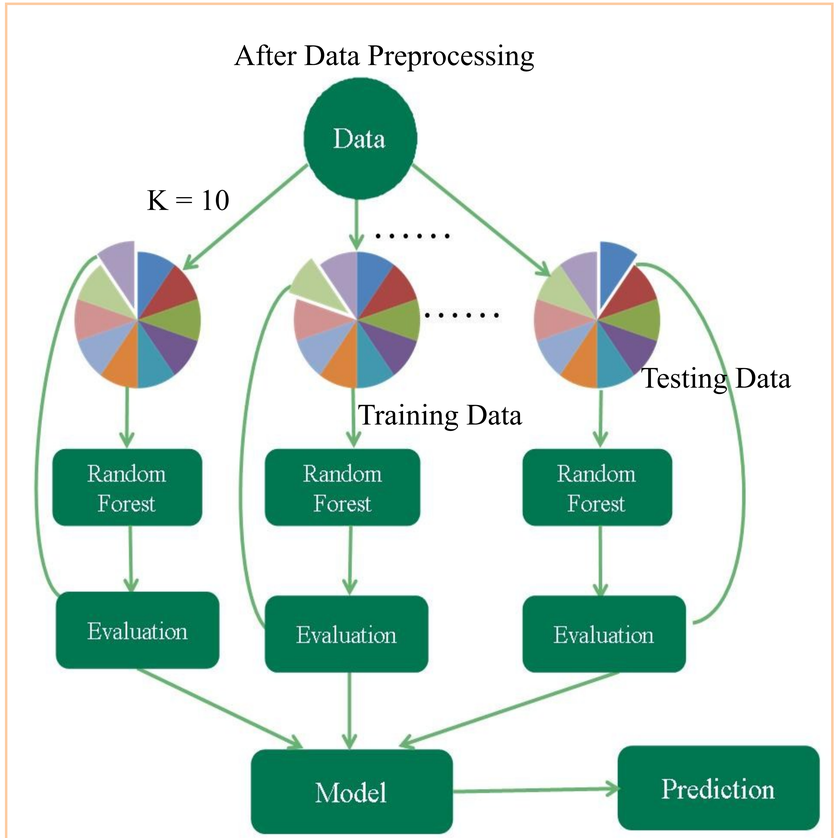
We have two datasets

- December 2015 loan issue data — Training & Testing datasets
  - * Covers 1 year period(Dec,2015 — Dec,2016)
  - ** 44,344 observations & 116 variables*
  - * Used to construct models
- February 2016 loan issue data — Prediction dataset
  - * Covers 1 year period(Feb,2016 — Feb,2017)
  - ** 2288 observations*
  - * Used to assess the predictive capabilities of the model

Methodology — Data Preprocessing

- Using R to deal with missing values and do data transform
- Using step-wise to select ***29 significant attributes*** from 116 attributes.
- List 4 of 29 attributes as example
  - Loan_amnt
  - term
  - installment
  - int-rate

Methodology — Model



- Using ***Cross-Validation*** method to generate testing data.
- Select best algorithm from ***seven candidates*** to construct final model
- Use final model to do ***prediction***

Result

Details Of Final Model

Correctly Classified Percentage	99.2473 %
Incorrectly Classified Percentage	0.7527 %

=== Detailed Accuracy By Class ===

TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
1.000	0.000	1.000	1.000	1.000	1.000	Curr&Late
0.984	0.003	0.993	0.984	0.989	1.000	Fully Paid
0.994	0.008	0.984	0.994	0.989	1.000	Charged Off
0.992	0.004	0.993	0.992	0.992	1.000	Weighted Avg

Details Of Prediction Result

Correctly Classified Instances	2282	99.7378 %
Incorrectly Classified Instances	6	0.2622 %

=== Confusion Matrix ===

a	b	c	<-- classified as
1917	5	0	a = Curr&Late
0	303	1	b = Fully Paid
0	0	62	c = Charged Off

After applying our final model to raw data to do prediction. It is only 6 out of 2288 instances be wrongly classified. Correct-Prediction probability is ***99.7378%.***

Future Work

- Lending Club (or other crow-funding lenders) can use this frame work to make accurate predictions
- With enough data, we could predict pay-back rates for the terms of Loans and determine who pays back early during that time framework.

References

[1]S.Chang; S,Dae-Oong Kim; *G. Kondo. Predicting Default Risk of Lending Club Loans pp.1-5* 2015: Research Project completed for Machine Learning Class(CS229) at Stanford University.
Image source: https://pixabay.com/p-1966550/?no_redirect
Image source: https://pixabay.com/p-1635464/?no_redirect