Analysis of 2018 LendingClub Loan Data

Statistical Data Analysis & Machine Learning Models

Ngoc Phan | M.S. Business Analytics | nphan20181@gmail.com | Github

Agenda

- Project Background
- Mission Statement
- Dataset
- Data Wrangling
- Exploratory Data Analysis
- Machine Learning Models
- Recommendations

Project Background

LendingClub

Source: https://www.pymnts.com/news/alternative-financial-services/2016/lending-club-timeline/

Mission Statement

- Assist LendingClub investors in the loan selection process
  - Exploring…
    - Fully paid vs. default loans
    - Return on investment (ROI) and loss of investment.
  - Estimating number of defaults for a selected sample
  - Developing machine learning models
  - Providing recommendation on loan selection

Dataset

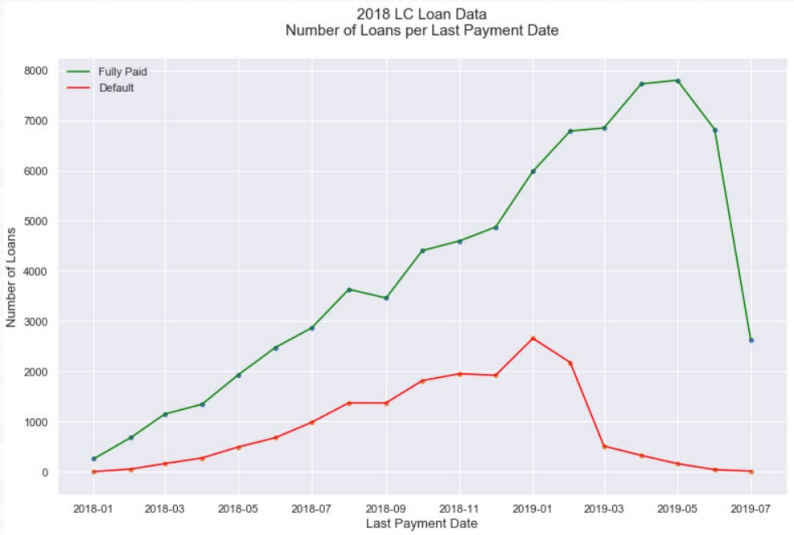
2018 LendingClub Loan Data	
Rows	495,242
Columns	144
Column Title	Loan Amount, Interest Rate, Term, Annual Income, etc.

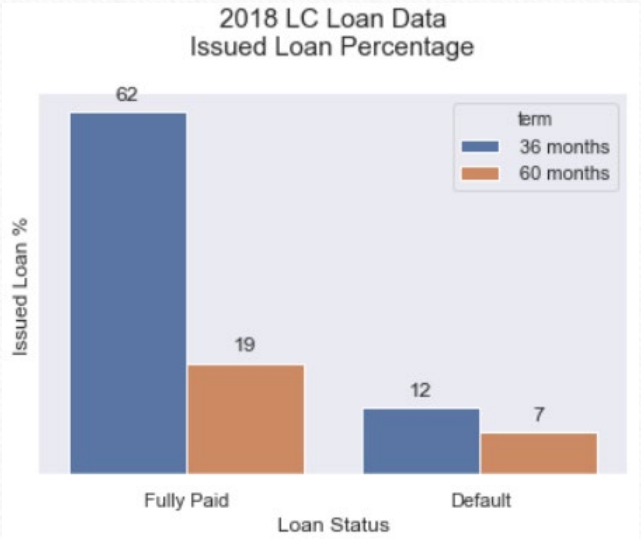
Data Wrangling

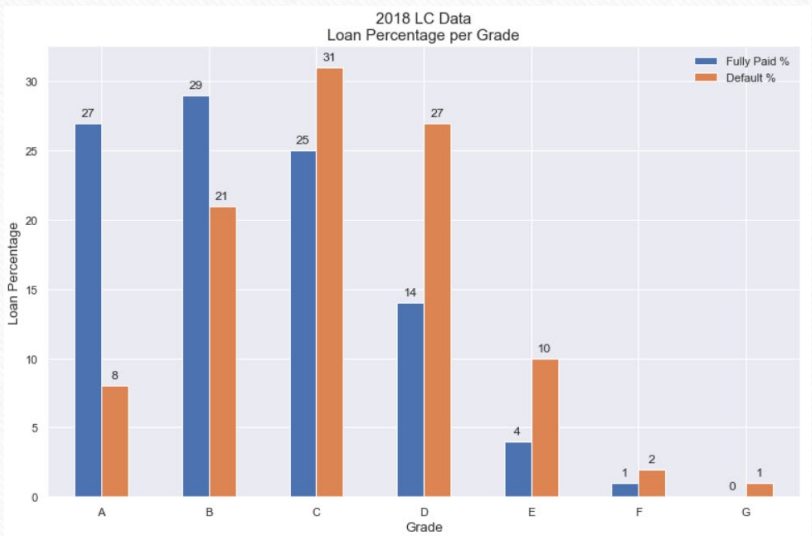
- Remove columns
  - Empty Columns
  - Columns with unimportant info
- Keep rows without status of fully paid / charged off / default
- Outliers
  - 64 columns and 36,355 rows
- Missing Values
  - Columns with > 25% missing values
  - Columns with < 25% missing values
- New Columns
  - Loan Status Flag: default, fully-paid
  - Return on Investment (ROI)
  - Months-in-Loan

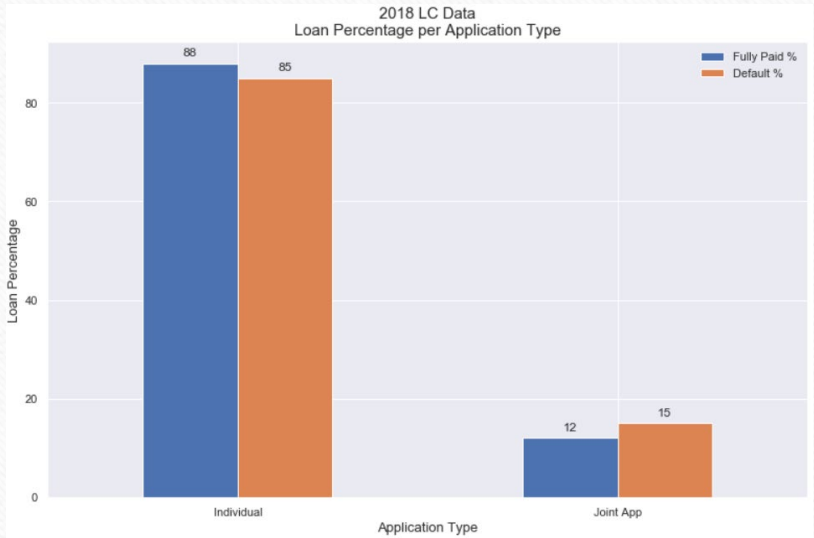
Exploratory Data Analysis

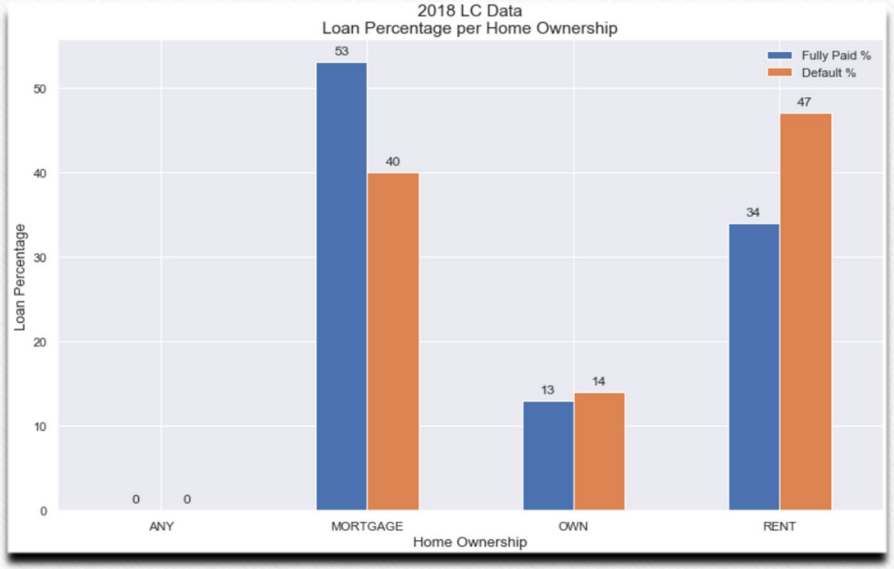
Descriptive Statistics & Inferential Statistics

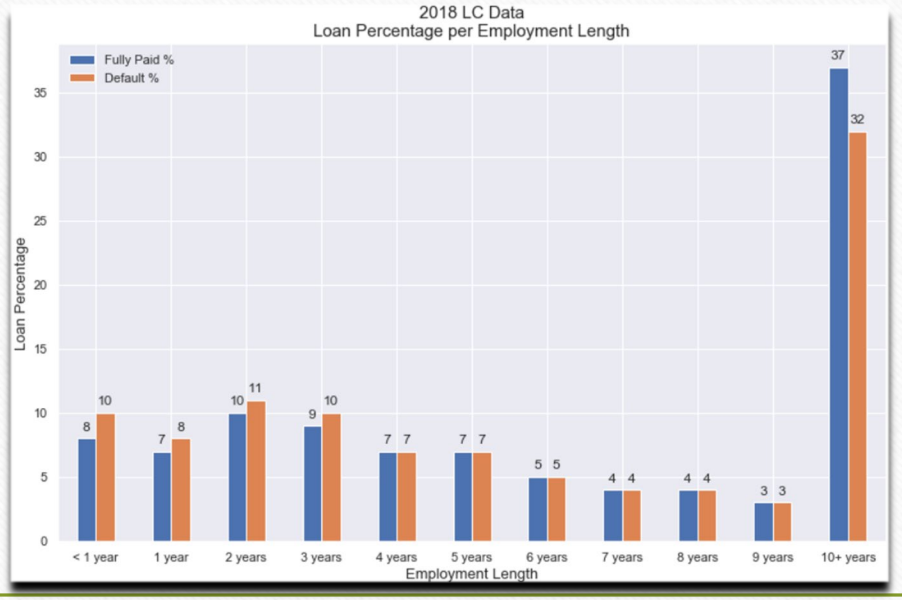

2018 LC Loan Data
Number of Loans per Last Payment Date
Fully Paid
Default
Number of Loans
Last Payment Date
2018-01
2018-03
2018-05
2018-07
2018-09
2018-11
2019-01
2019-03
2019-05
2019-07

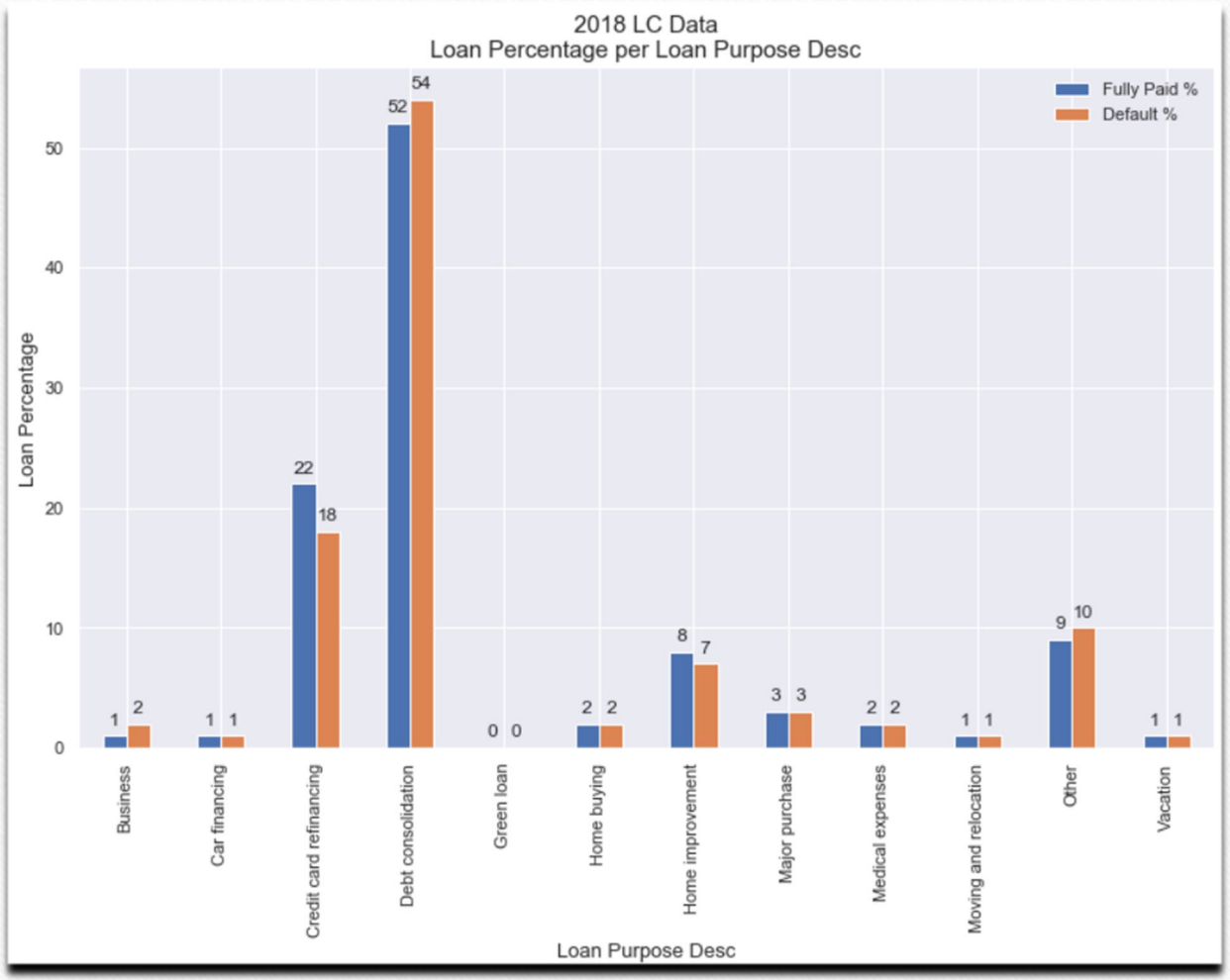

2018 LC Loan Data
Issued Loan Percentage
term
36 months
60 months
Issued Loan %
62
19
12
7
Fully Paid
Default
Loan Status

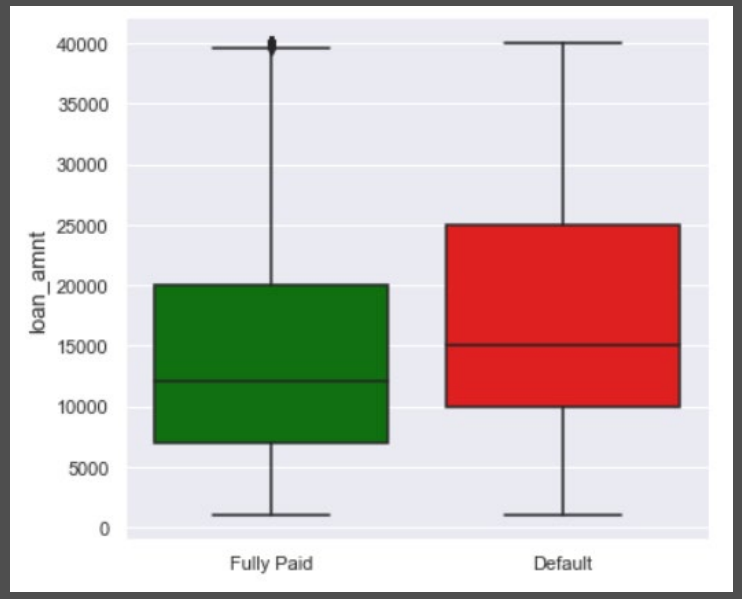

2018 LC Data
Loan Percentage per Grade
Fully Paid %
Default %
Loan Percentage
27
8
29
21
25
31
14
27
4
10
1
2
0
1
A
B
C
D
E
F
G
Grade

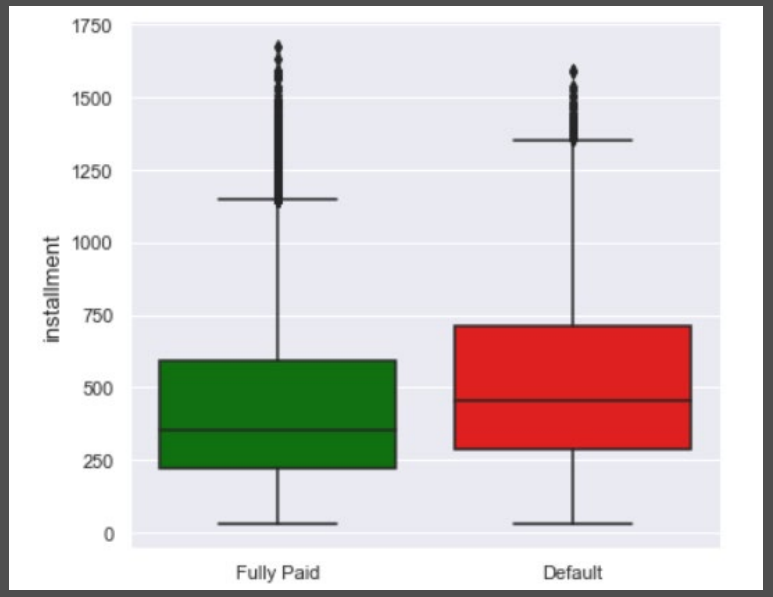

2018 LC Data
Loan Percentage per Application Type
Fully Paid %
Default %
Loan Percentage
88
85
12
15
Individual
Joint App
Application Type

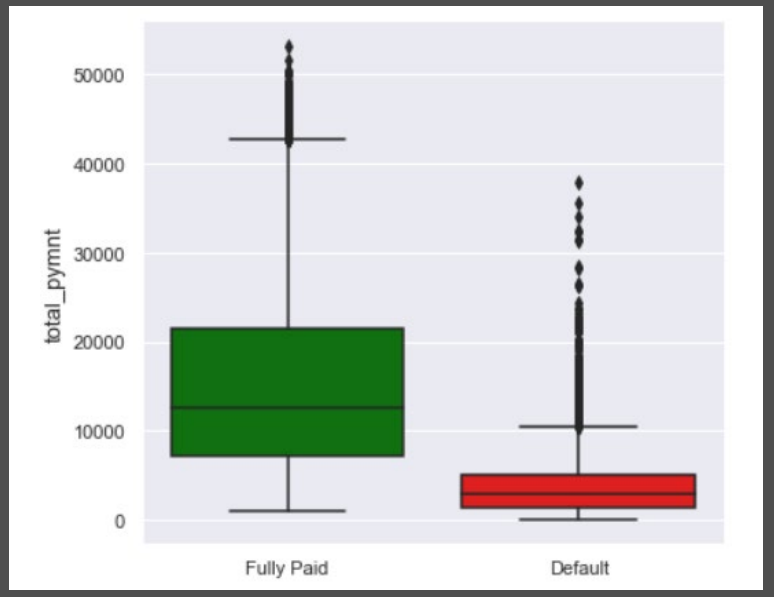

2018 LC Data
Loan Percentage per Home Ownership
Fully Paid %
Default %
Loan Percentage
Home Ownership
ANY
MORTGAGE
OWN
RENT
0
0
53
40
13
14
34
47

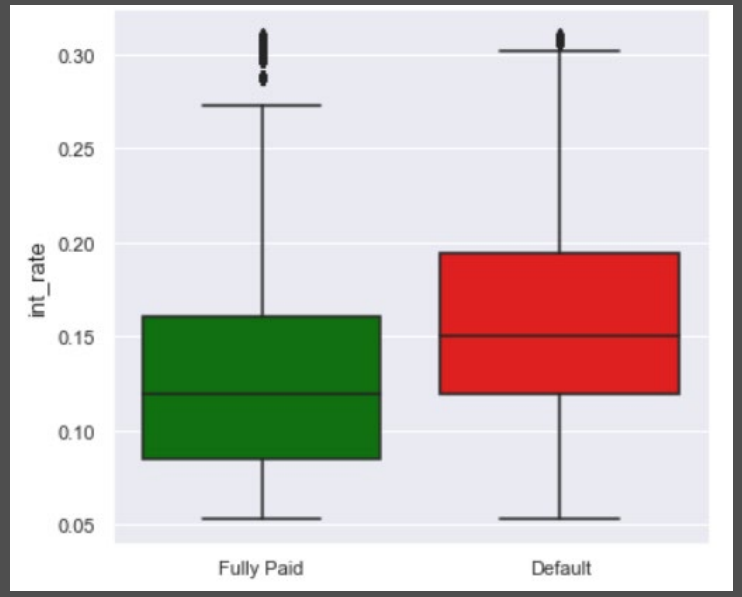

2018 LC Data
Loan Percentage per Employment Length
Fully Paid %
Default %
Loan Percentage
Employment Length
< 1 year
1 year
2 years
3 years
4 years
5 years
6 years
7 years
8 years
9 years
10+ years
8
10
7
8
10
11
9
10
7
7
7
7
5
5
4
4
4
4
3
3
37
32

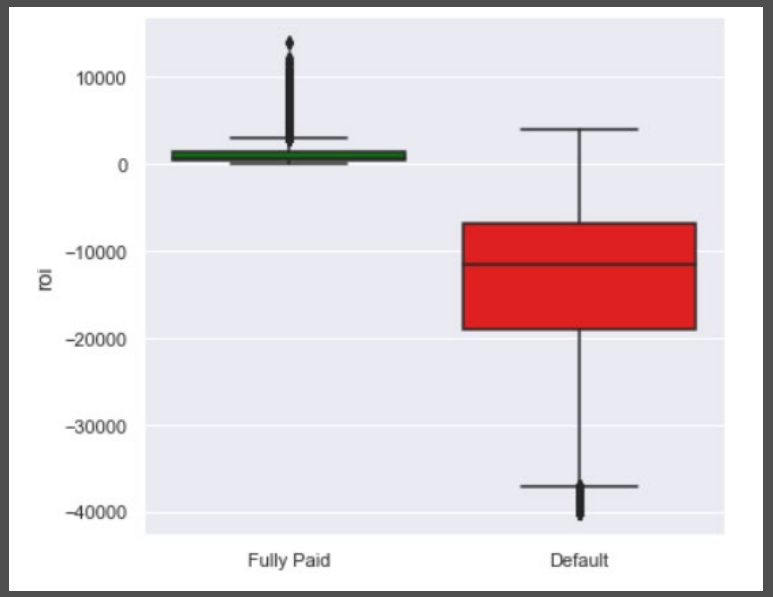

2018 LC Data
Loan Percentage per Loan Purpose Desc
Fully Paid %
Default %
Loan Percentage
Loan Purpose Desc
Business
Car financing
Credit card refinancing
Debt consolidation
Green loan
Home buying
Home improvement
Major purchase
Medical expenses
Moving and relocation
Other
Vacation
1
2
1
1
22
18
52
54
0
0
2
2
8
7
3
3
2
2
1
1
9
10
1
1

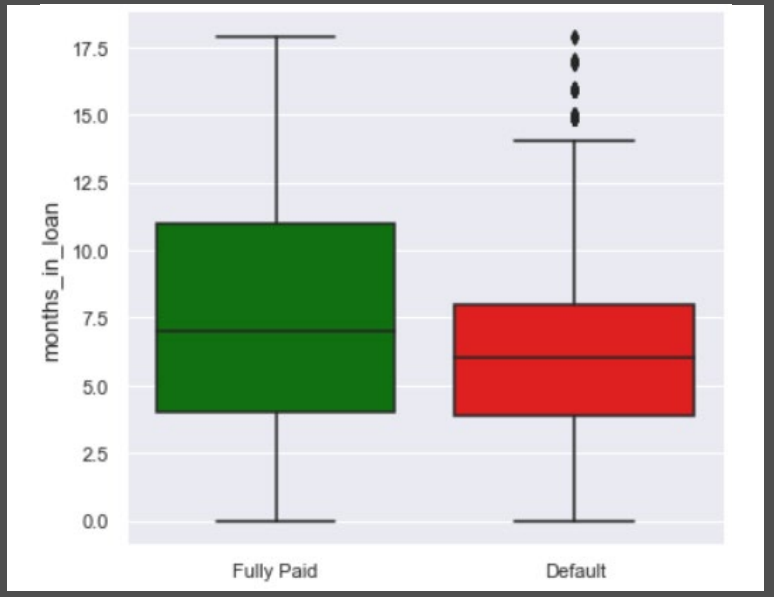

loan_amnt
40000
35000
30000
25000
20000
15000
10000
5000
0
Fully Paid
Default


installment
1750
1500
1250
1000
750
500
250
0
Fully Paid
Default

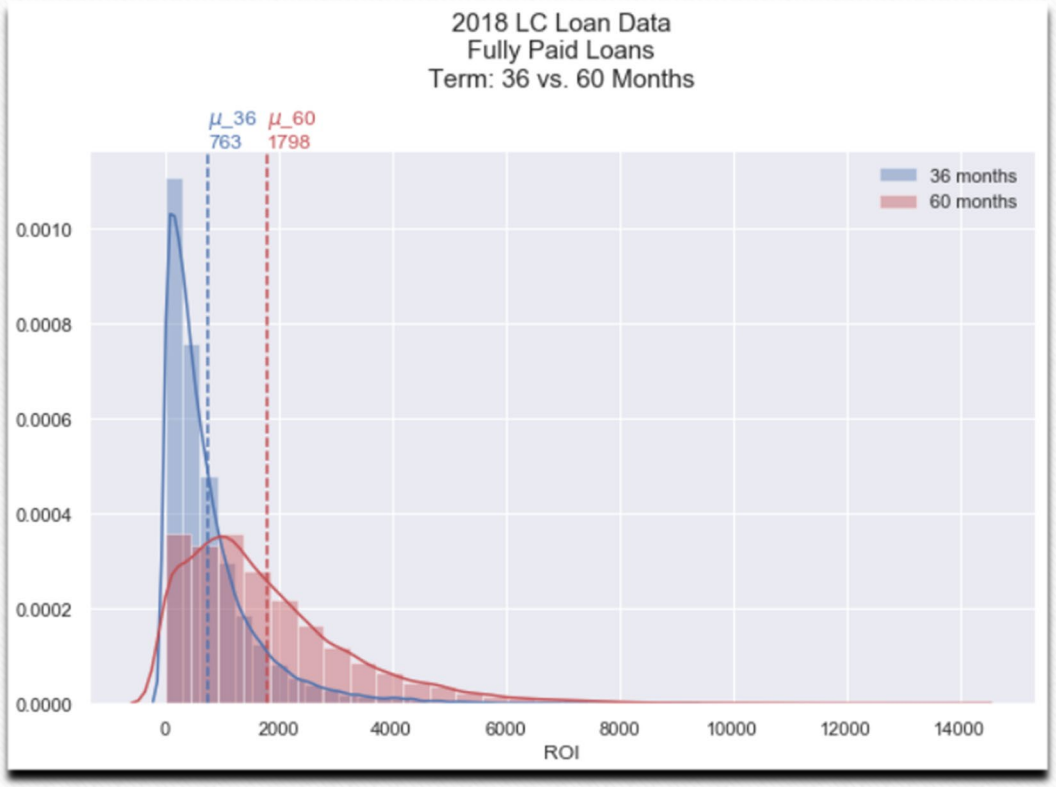

total_pymnt
50000
40000
30000
20000
10000
0
Fully Paid
Default

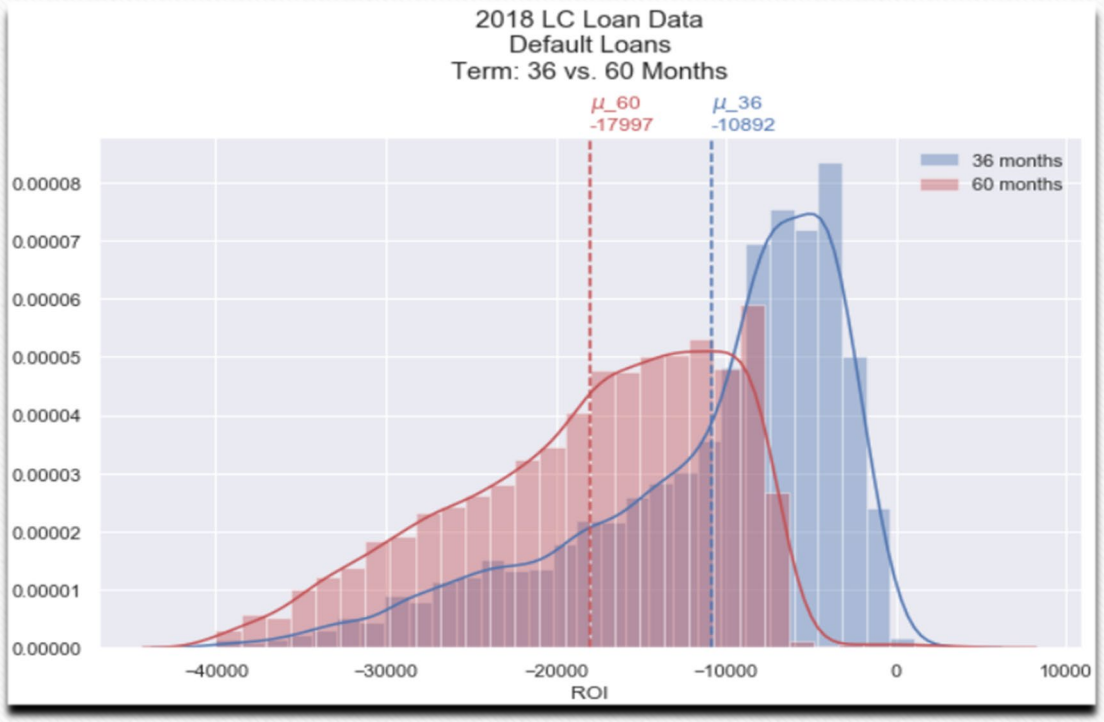

int_rate
0.30
0.25
0.20
0.15
0.10
0.05
Fully Paid
Default


roi
10000
0
−10000
−20000
−30000
−40000
Fully Paid
Default


months_in_loan
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0.0
Fully Paid
Default

Return on Investment (ROI)



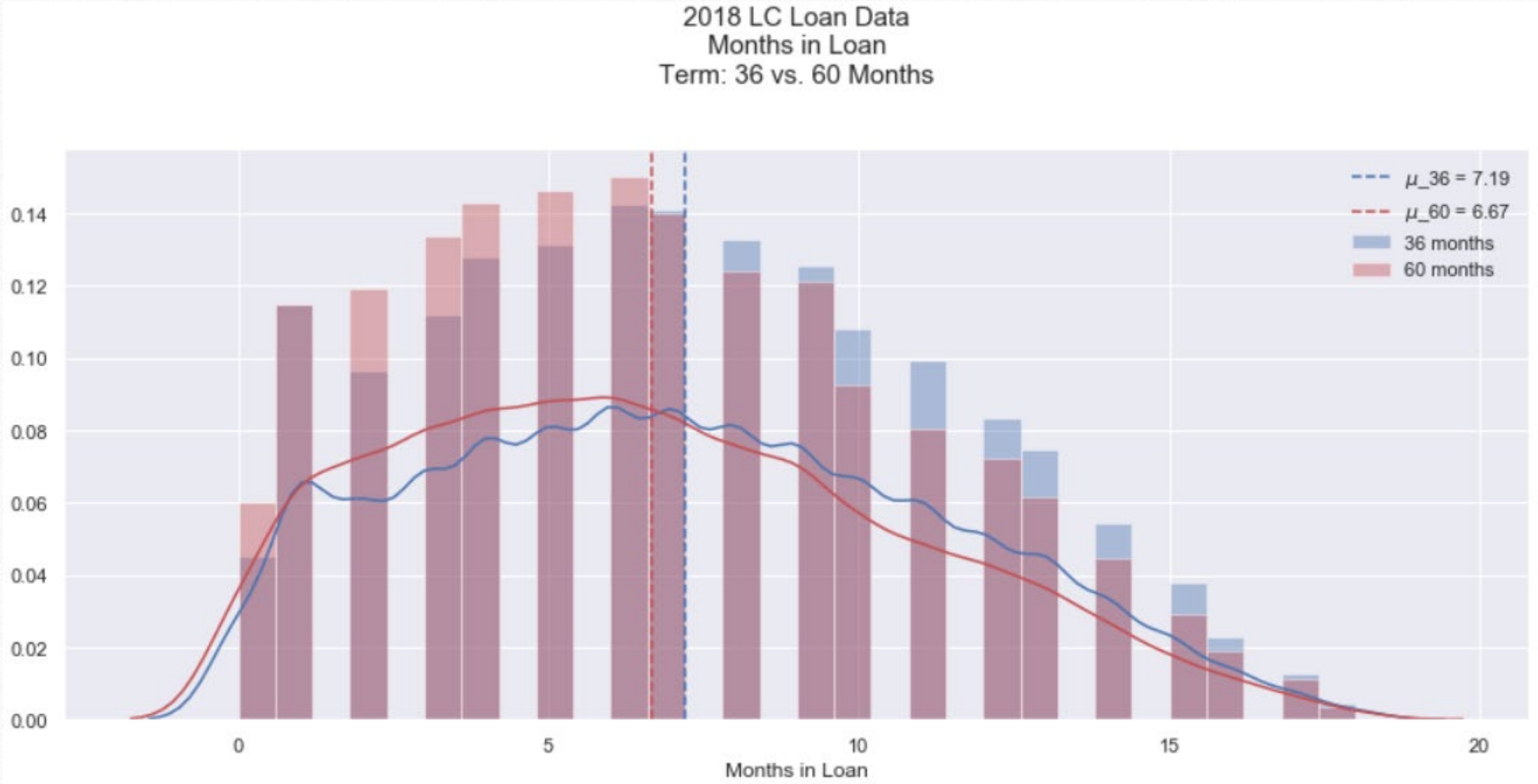


95% Confident Interval for ROI

- 36 months: between $756 and 770
- 60 months: between $1,776 and $1,820

95% Confident Interval for Lost of Investment

- 36 months: between $10,744 and $11,040
- 60 months: between $17,810 and $18,185


2018 LC Loan Data
Months in Loan
Term: 36 vs. 60 Months
μ_36 = 7.19
μ_60 = 6.67
36 months
60 months
0.14
0.12
0.10
0.08
0.06
0.04
0.02
0.00
0
5
10
15
20
Months in Loan

Estimating Months-in-Loan for Selected Sample

Binomial Distribution

- Probability of Default: 18.74%

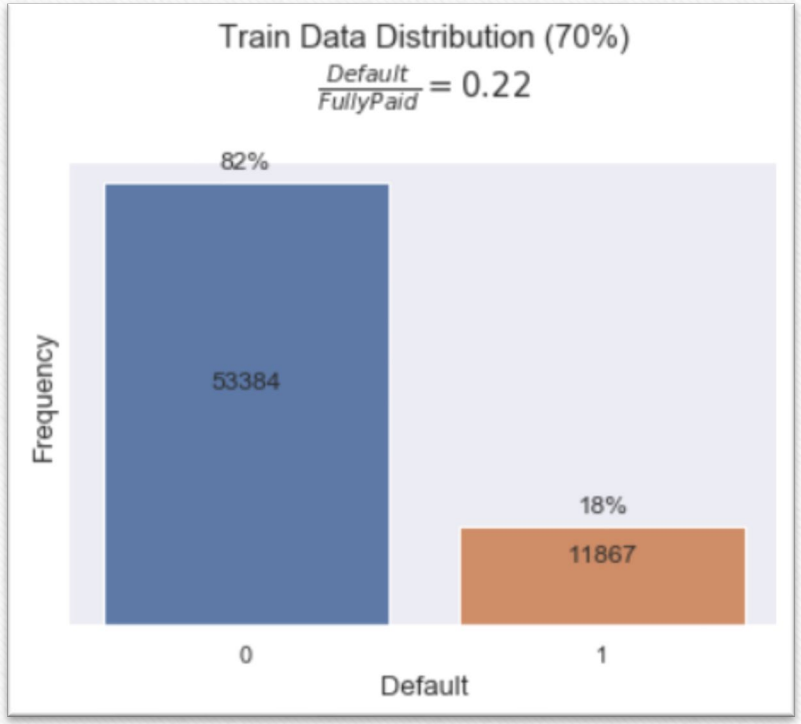
Number of Loans	5	10	15	20	25	30	35	40	45	50
Expected No. of Defaults	1	2	3	4	5	6	7	7	8	9

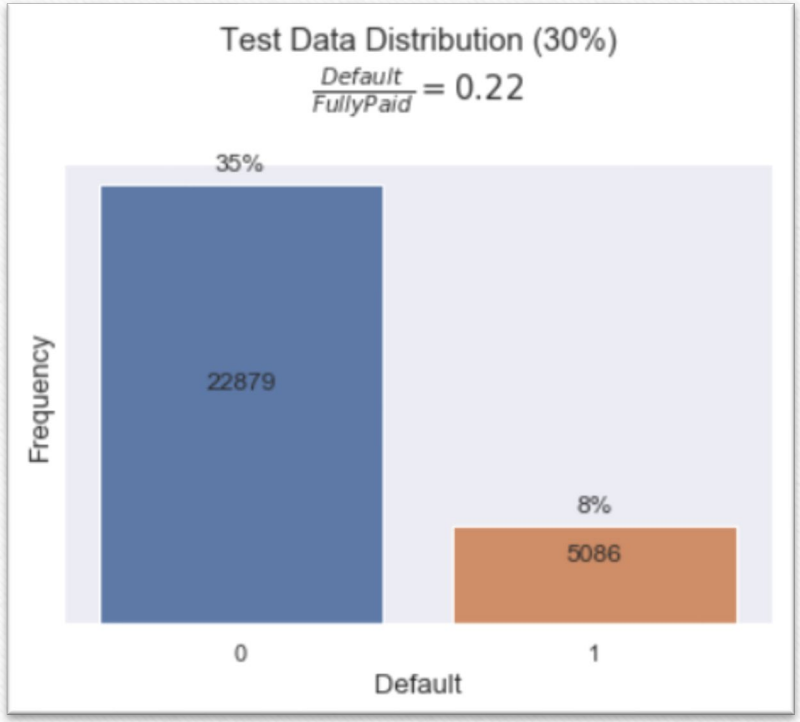
Machine Learning Models

Random Forest Classifier: Classify Loan

Random Forest Regressor: Estimate Months-in-Loan

Train Data vs. Test Data





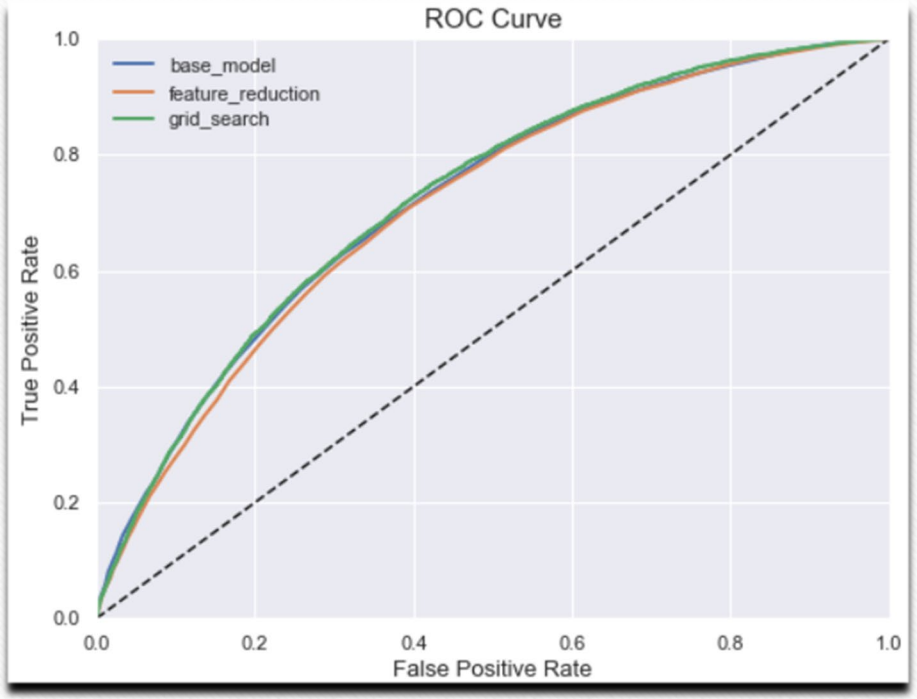
Model Development

Base Model	Model 1	Model 2
• Use all features • Use the following parameter: param = {'bootstrap': [True], 'n_estimators': [100]}	• Use same parameter as base model. • Use features with important score greater than 0.2. • Exclude some features with strong correlation.	• Use same features are model 1. • Use the best parameters from grid search.

Random Forest Classifier: Classify Loan

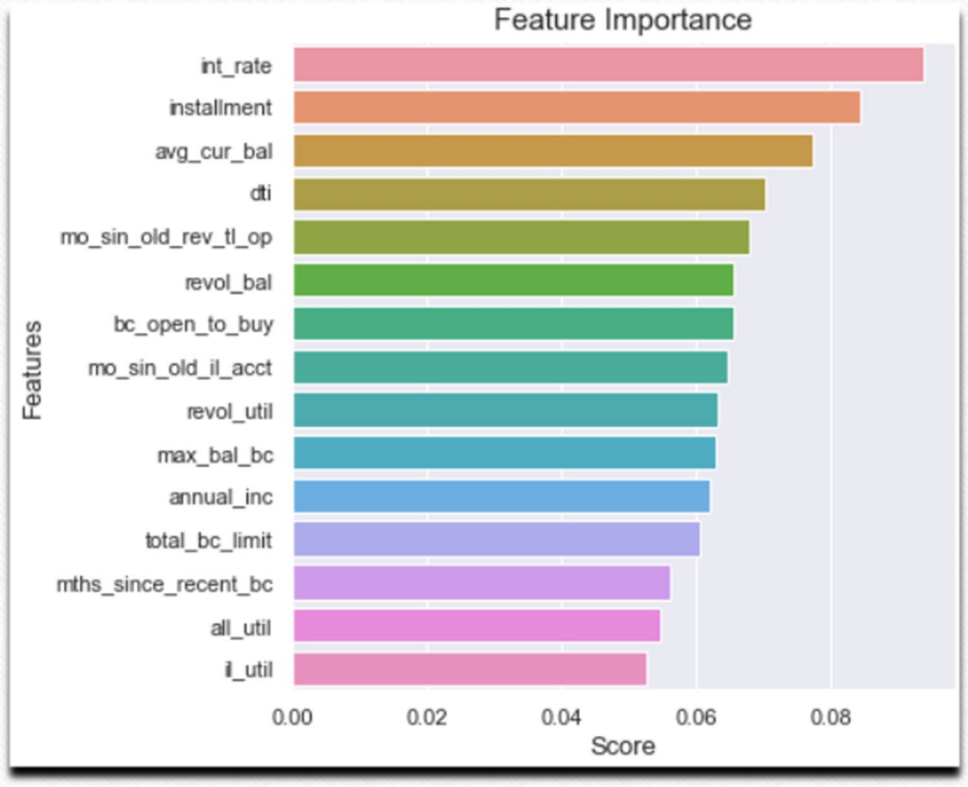
Model Evaluation & Selection | Model Prediction

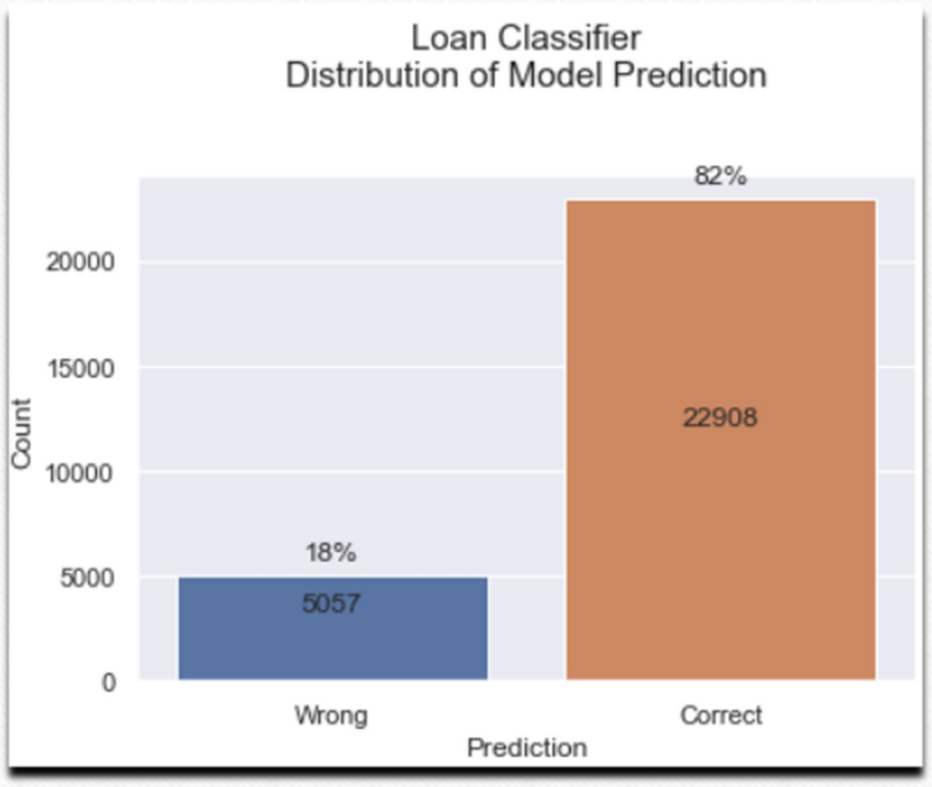
Loan Classifier: Model Evaluation



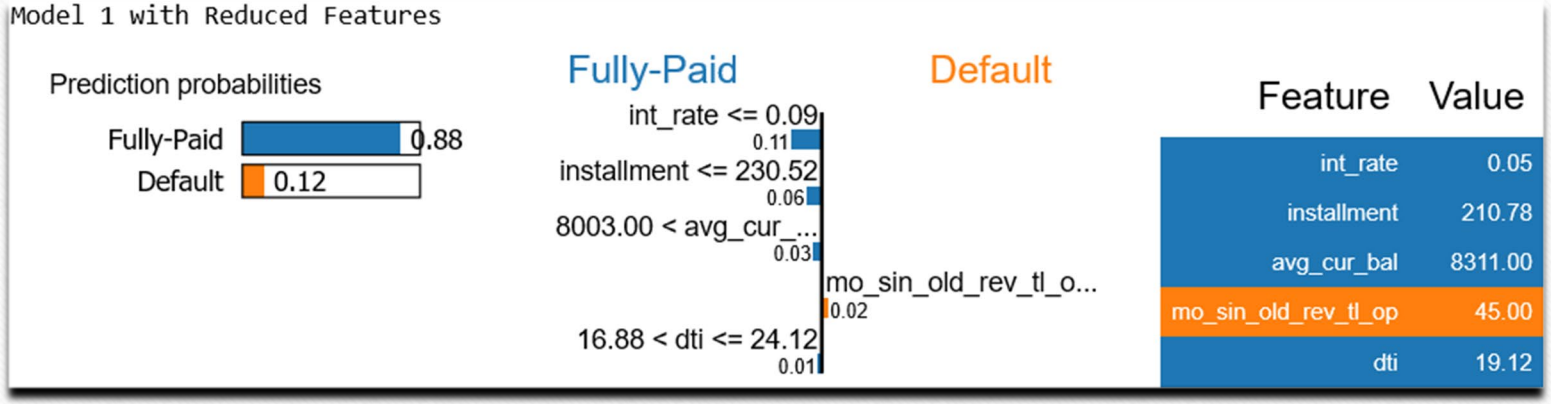
model_name	true_neg	true_pos	false_neg	false_pos	accuracy	precision	recall	f1_score	training_time
base_model	22692	241	4845	187	0.820061	0.563084	0.047385	0.087414	68
feature_reduction	22684	224	4862	195	0.819167	0.534606	0.044042	0.081381	41
grid_search	22777	145	4941	102	0.819667	0.587045	0.028510	0.054378	1179

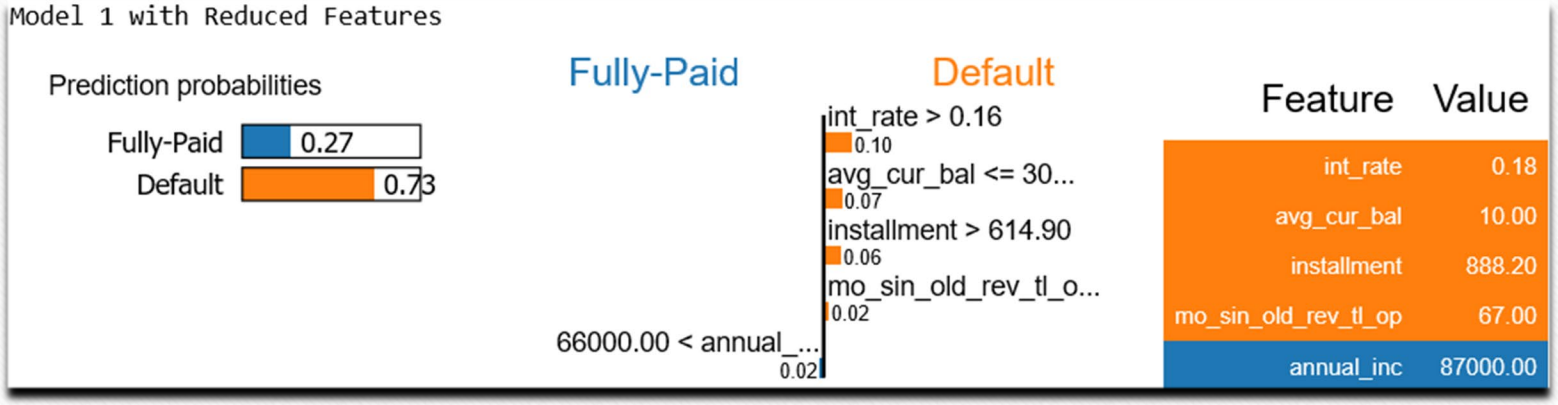
Loan Classifier: Model 1 (Feature Reduction)





Default Prediction with Loan Classifier





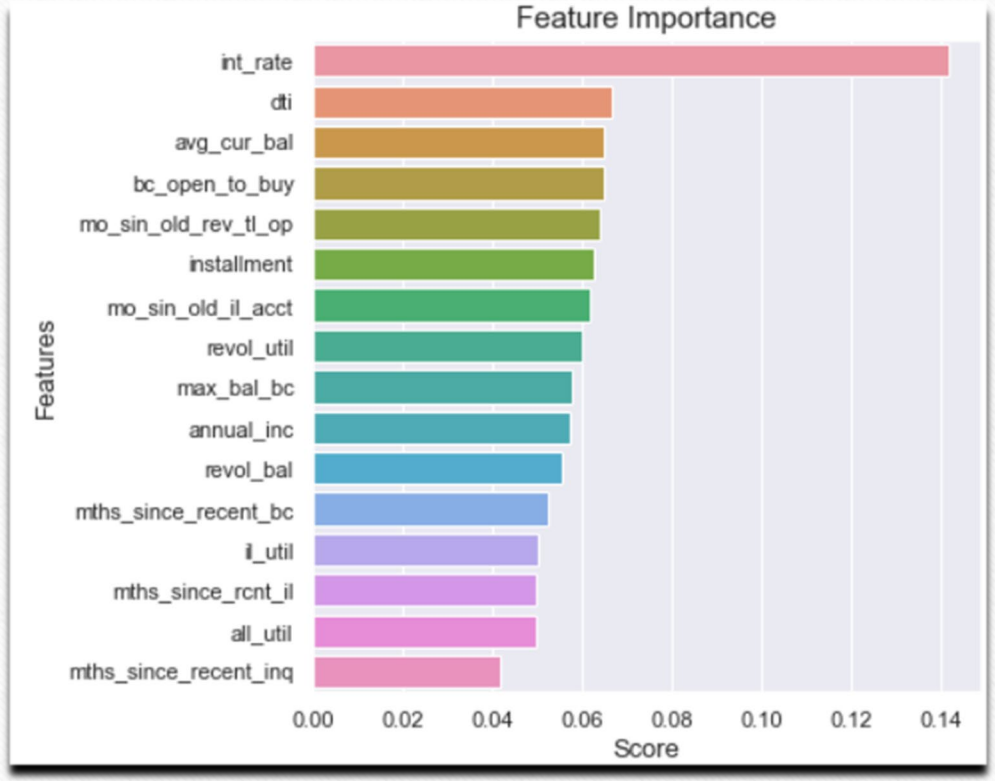
Random Forest Regressor: Estimate Months-in-Loan

Model Evaluation & Selection | Model Prediction

Months-in-Loan Estimator Model Evaluation

model_name	mse	rmse	mae	R-squared	training_time
base_model	15.385022	3.922374	3.237386	0.871849	240.061900
feature_reduction	15.510222	3.938302	3.246415	0.871484	87.549685
grid_search	16.500856	4.062125	3.373749	0.109093	338.585933

Months-in-Loan Estimator
Model 1 (Feature Reduction)



Model Prediction

loan_amnt	installment	actual_label	predict_label	prob_fully_paid	prob_default	actual_mil	predict_mil	predict_error	predict_mil_68cf	predict_mil_95cf
7000	210.78	fully-paid	fully-paid	0.88	0.12	7	9.9	2.9	(5.93, 13.81)	(1.96, 17.71)
1600	48.74	fully-paid	fully-paid	0.98	0.02	15.9	9.3	-6.6	(5.33, 13.21)	(1.36, 17.11)
20000	693.51	fully-paid	fully-paid	0.69	0.31	9	4.9	-4.1	(0.93, 8.81)	(0, 12.71)
7000	240.85	fully-paid	fully-paid	0.77	0.23	3	5.3	2.3	(1.33, 9.21)	(0, 13.11)
20000	684.33	fully-paid	fully-paid	0.84	0.16	14.9	7	-7.9	(3.03, 10.91)	(0, 14.81)
10000	275.34	default	default	0.47	0.53	9.1	7.6	-1.5	(3.63, 11.51)	(0, 15.41)
8000	251.4	default	fully-paid	0.97	0.03	1	4.2	3.2	(0.23, 8.11)	(0, 12.01)
10000	381.13	default	default	0.48	0.52	7	6.8	-0.2	(2.83, 10.71)	(0, 14.61)
6000	224.18	fully-paid	fully-paid	0.86	0.14	10	7	-3	(3.03, 10.91)	(0, 14.81)
15000	481.06	fully-paid	fully-paid	1	0	11	7.6	-3.4	(3.63, 11.51)	(0, 15.41)
8950	321.19	default	fully-paid	0.63	0.37	2	7.7	5.7	(3.73, 11.61)	(0, 15.51)
4800	154.71	fully-paid	fully-paid	0.96	0.04	0	7.7	7.7	(3.73, 11.61)	(0, 15.51)
4800	166.52	fully-paid	fully-paid	0.75	0.25	11	9	-2	(5.03, 12.91)	(1.06, 16.81)
30000	606	fully-paid	fully-paid	0.8	0.2	5	7.1	2.1	(3.13, 11.01)	(0, 14.91)
35000	745.03	default	fully-paid	0.79	0.21	3	5.2	2.2	(1.23, 9.11)	(0, 13.01)
4000	152.46	default	fully-paid	0.83	0.17	3	8.1	5.1	(4.13, 12.01)	(0.16, 15.91)

Recommendations to Investors

- Invest in a fractional part of a loan
- Invest in 36 months loan term
- Invest in loan that has
  - Low interest rate
  - Low installment
  - Low average current balance of all accounts

Thank You!