Predicting Loan Defaults

Nikesh Patel

11.12.13

Lending Club is a P2P Loan Originator

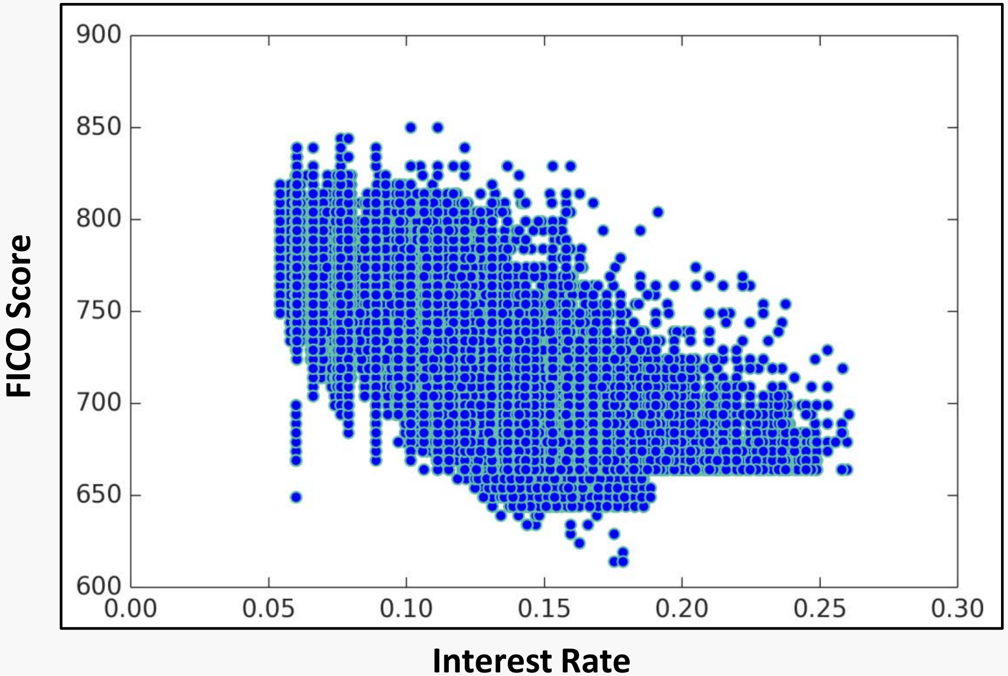
1. Borrower applies for loans on LC website
2. Approved loans are assigned Risk Grade and Interest Rate
3. Investors can view loan info/stats and purchase "shares"
4. LC charges loan application and share transaction fees

Problem and Approach

Can we Predict which Loans will Default?

- LC publishes a rich dataset for all historical and “in funding” loans
- LC’s Risk Grading algorithm is a “black box”

FICO Score vs. Interest Rate



Data: Loans from 2007-Present

Features:

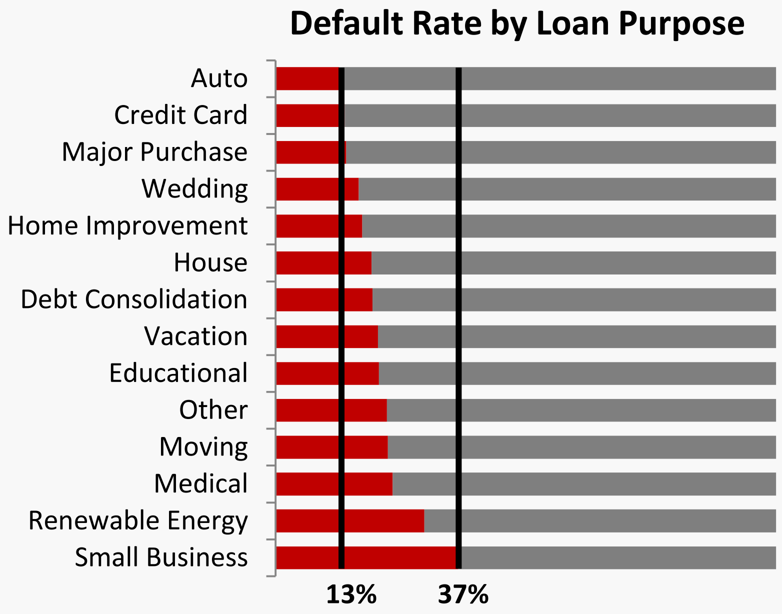
- Loan details (Amount, Length, Purpose)
- Income, Employment, Location, Mortgage/Rent
- Credit History (FICO, Delinquencies, Balances, # Account/Types)
- Current Status (Fully Paid, Current, Late, Default/Charged Off)

purpose	title	addr_city	addr_state	msa	acc_now_delinq	acc_open_past_24mths	bc_open_to_buy	percent_bc_gt_75	bc_util	dti
Credit card refinancing	Debt Consolidation Loan	Plantation	FL		0	6	1995	100	85.2	18.24
Credit card refinancing	Out of Debt in 3!	PARAGOULD	AR		0	8	38164	16.7	26	11.28
Debt consolidation	Debt consolidation	Highlands Ranch	CO		0	5	7684	16.7	58	4.01
Debt consolidation	Peace Of Mind	MEMPHIS	TN		0	8	11874	16.7	47	14.06
Debt consolidation	Goodbye to debt	Wirtz	VA		0	3	2112	0	50.9	19.02
Credit card refinancing	Credit Card Payoff	Chicago	IL		0	0	64	100	97.1	4.62
Credit card refinancing	CreditCardsNG	Fairfield	CT		0	6	5888	33.3	45.5	26.58
Debt consolidation	Debt consolidation	Grand Marais	MN		0	0	26200	14.3	41.6	3.87
Credit card refinancing	Credit card refinancing	PRATTVILLE	AL		0	4	2279	40	74.4	8.94
Credit card refinancing	credit card payoff	SPRING HILL	FL		0	8	71555	25	20.5	8.64
Debt consolidation	Finale	RICHLAND	PA		0	0	3828	50	68.6	4.57
Credit card refinancing	Credit card refinancing	Pembroke Pines	FL		0	1	18128	66.7	68	15.27
Credit card refinancing	Credit card refinancing	HOMESTEAD	FL		0	6	32299	0	5.6	7.71
Debt consolidation	consolidation	Los Angeles	CA		0	1	4343	50	85.2	18.18
Credit card refinancing	Consolidation loan	SANTA CLARITA	CA		0	5	2119	25	69.7	4.82
Debt consolidation	last console	phx	AZ		0	5	12477	25	24.8	16.05
Debt consolidation	Consolidate	Gardena	CA		0	2	7054	50	70.2	9.51
Debt consolidation	Credit card debt consolidation	Atkinson	NH		0	1	46695	0	16.2	9.13
Credit card refinancing	cc refi	Colfax	CA		0	5	54755	42.9	37.4	18.55

Analysis – Cleaning/Processing

Challenges:

- Due to loan durations (36/60 month), few have fully matured
- LC's changed its credit policy in 2010; features are not consistent
- Collinearity (e.g. Revolving Balance vs. High Credit)



Feature Selection:

- Loan Fundamentals (Amnt, Duration)
- Loan Purpose
- "Signals" (Inquiries w/n 6 months, FICO)

Analysis – Techniques

Logistic Regression

- Struggled with classifying bad loans
- Very high false negative rate

Naive Bayes

- Better at classifying defaults than Regression
- High false positive rate

Decision Trees

- Good performance from AdaBoost
- Lower false positive rate, but not strictly better than Bayes

Results

Feature Importance

- Small Business loans are risky!
- Income & Income/Loan Amount ratio
- Available credit utilization
- Not Important: Location, Housing Status

AUC maxed out at 0.55 ☹

	precision	recall	f1-score	support
False	0.82	0.98	0.89	25498
True	0.54	0.08	0.14	6048

- False Negatives > True Positives
- Disappointing, but makes sense: if Lending Club could reliably predict defaults, they wouldn't approve them!
- Recall: Sample is composed of approved loans, not all applications

Conclusions & Next Steps

Is there still value in the Model?

- Use the low "False Negative" model to screen loans
- Identify local misclassification within Risk Grades

Opportunities for Improvement

- "Financial Stability" model for refis
- Sample expansion through new originators

Next Steps

- Reverse-engineer LC model
- De-Anonimization to incorporate new features
- Analyze recovery rate (e.g. Late →Current) for secondary market opps.