

Project Home-a-loan

...

July 19, 2017

The Challenge



Mission:

- Predict whether a lead will convert to “lock”
- Supervised learning (Classification)

Business Values:

- Know the potential of each customer - know the ones to focus on
- Best Customers profiling - targeted marketing for new customer acquisition

Mission:

- Predict “locked-to-funded” time (efficiency)
- Supervised learning (Regression)

Business Values:

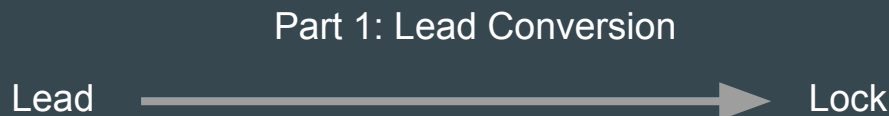
- Improve customer experience by providing an expected waiting time
- Know areas of improvement for efficiency

The data

- 9752 cases
- 417 features
- Cleaning / pre-processing:

Data Type	Examples	Processing
Numerical	Borrower income, loan amount	None
Categorical	Type of home, Education level, City of property, Gender	Dumification
Text	Goal of refinancing, Unqualified reason note	Tfidf Vectorization (limiting stop words)
Datetime	Created time, Last modified	Categorize (year, quarter, month, dow), Calculate Period (difference of dates/times), Calculate Cohort (quarter/month since initiation)

Part 1 - Leads Conversion



Mission:

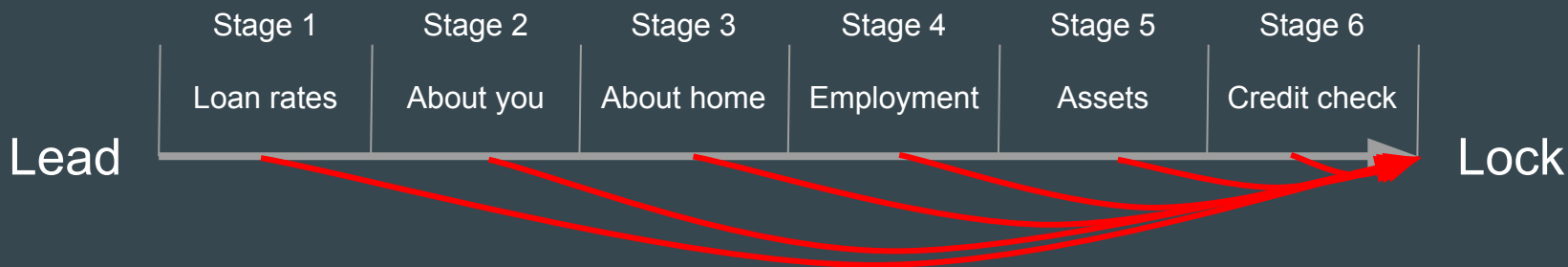
- Predict whether a lead will convert to “lock”
- Supervised learning (Classification)

Business Values:

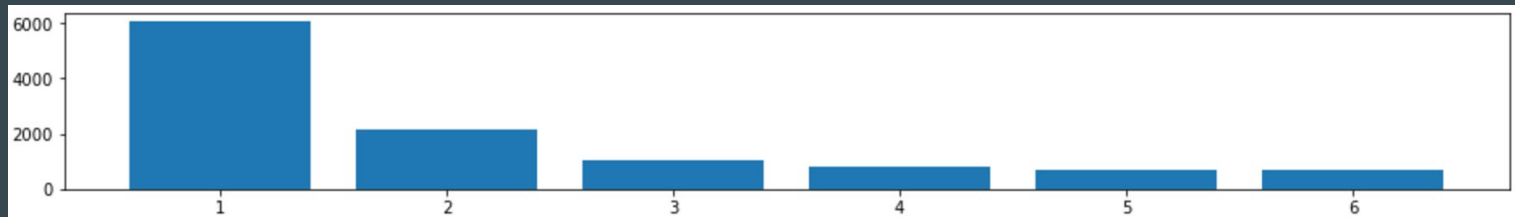
- Know the potential of each customer - know the ones to focus on
- Best Customers profiling - targeted marketing for new customer acquisition

Part 2
(Later)

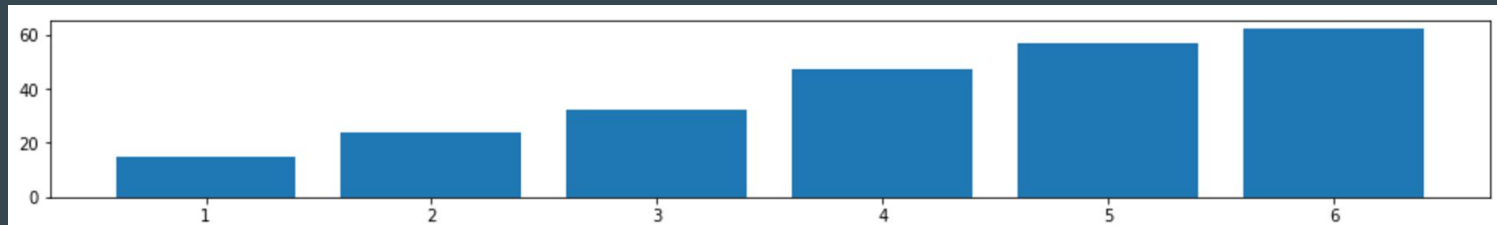
Part 1 - Leads Conversion



Sample
Sizes
(Rows)

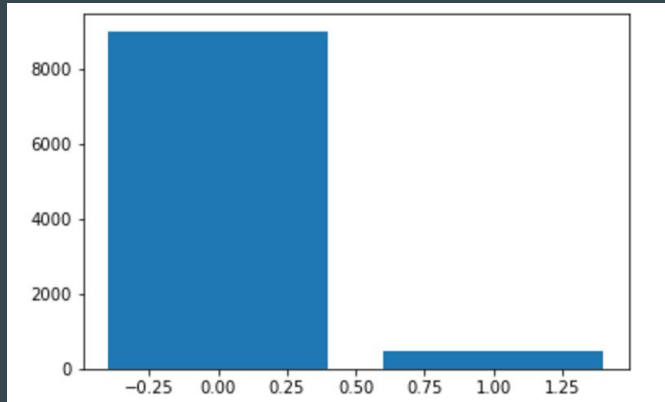


#Features
(Columns)

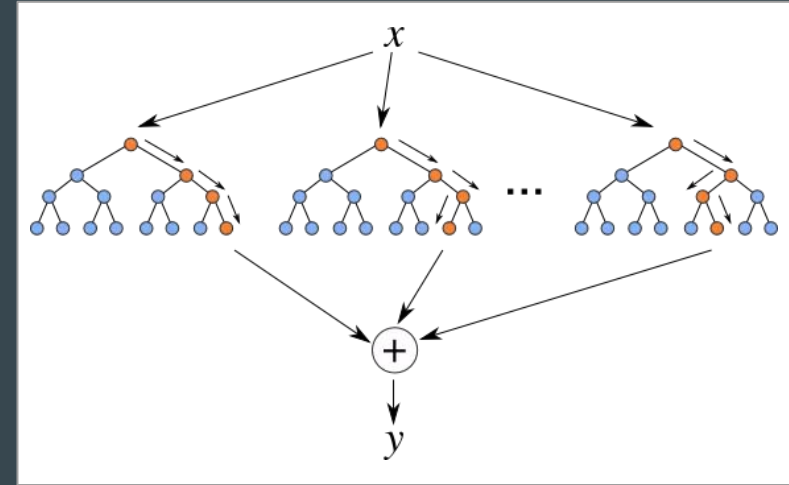


Part 1 - Leads Conversion - Methodology

Random Oversampling for Imbalance Classes

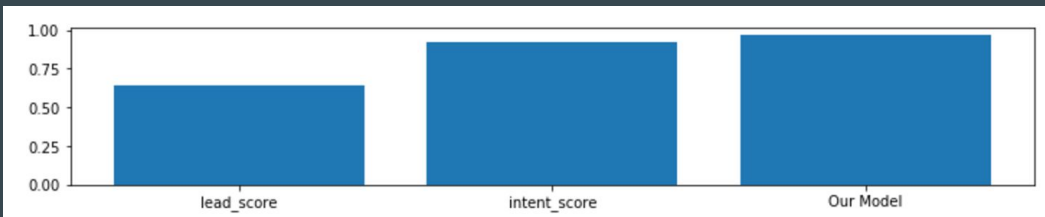


Random Forest Classifier



Part 1 - Leads Conversion - Results

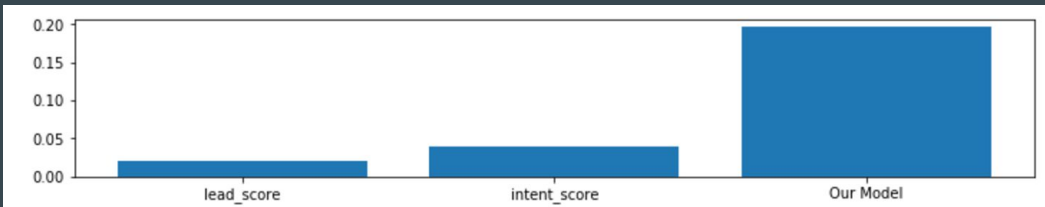
Accuracy



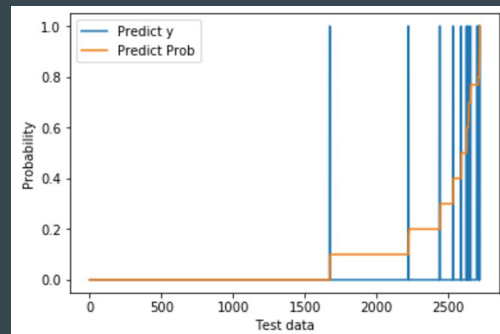
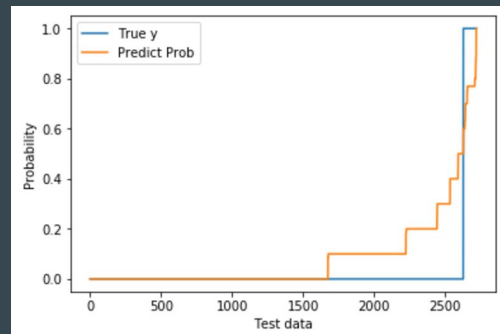
F1-score



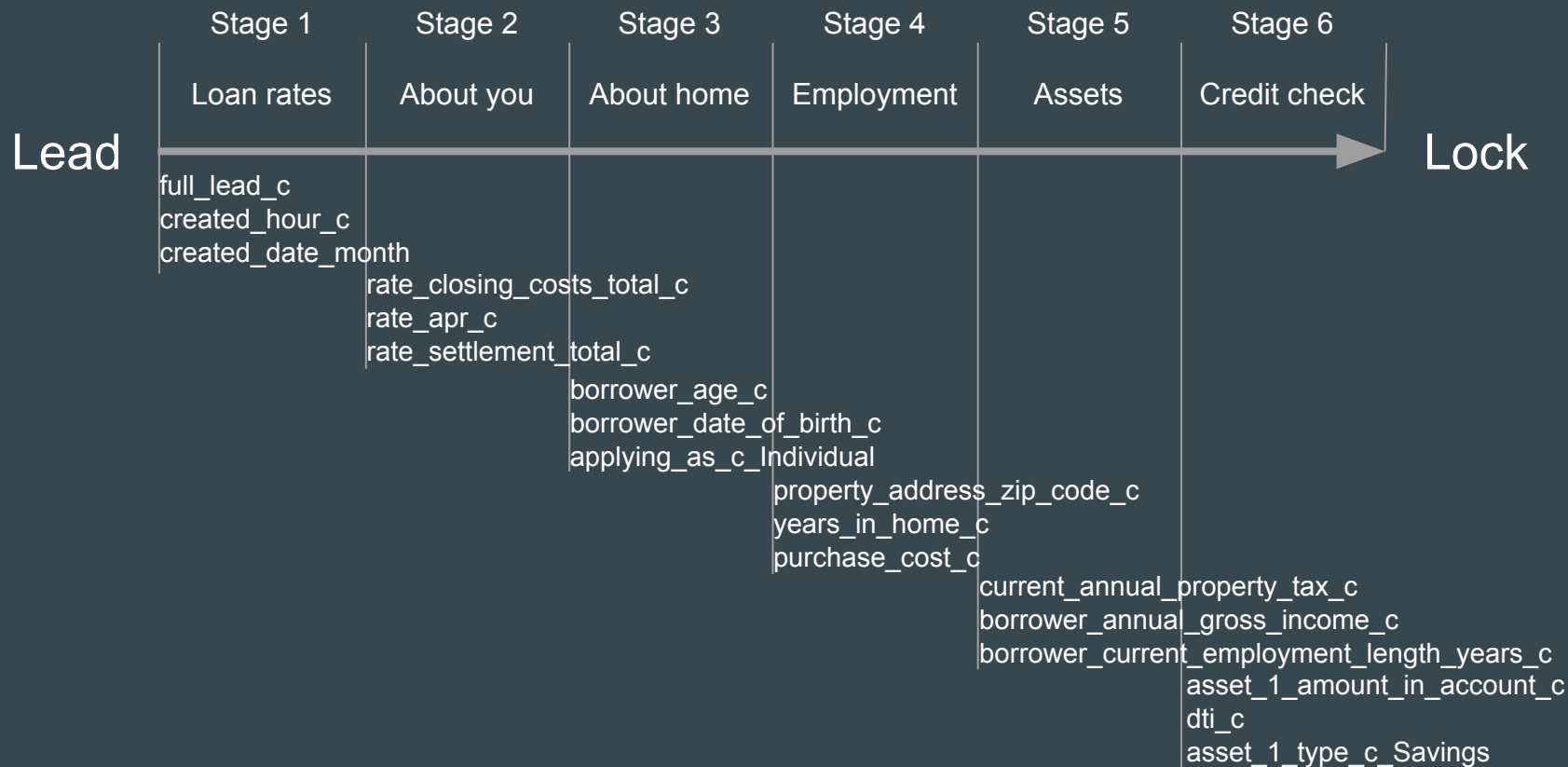
Precision



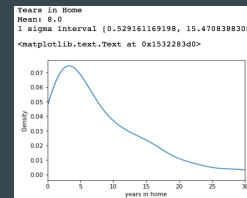
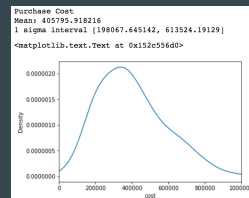
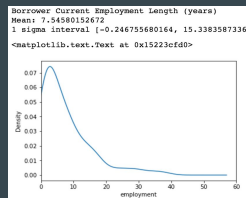
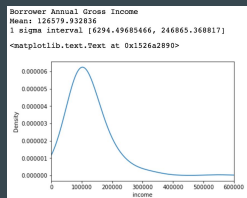
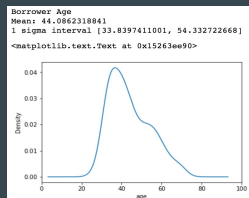
Predicted Probability v.s. True Value



Part 1 - Leads Conversion - Interpretation



Part 1 - Leads Conversion - Interpretation



Feature	Borrower Age	Borrower's Annual Income	Borrower current employment	Home purchase cost	Years in home	Property zip code
Mean	44	127,000	7.5	406,000	8	98103
Common range	33 - 54	6,300 - 247,000	0 - 15	190,000 - 614,000	0.5 - 15	98103, 92691, 98125, 93003, 94110

Part 2 - Loan Processing

Part 1
(Done)

Part 2: Loan Processing

Lock



Fund

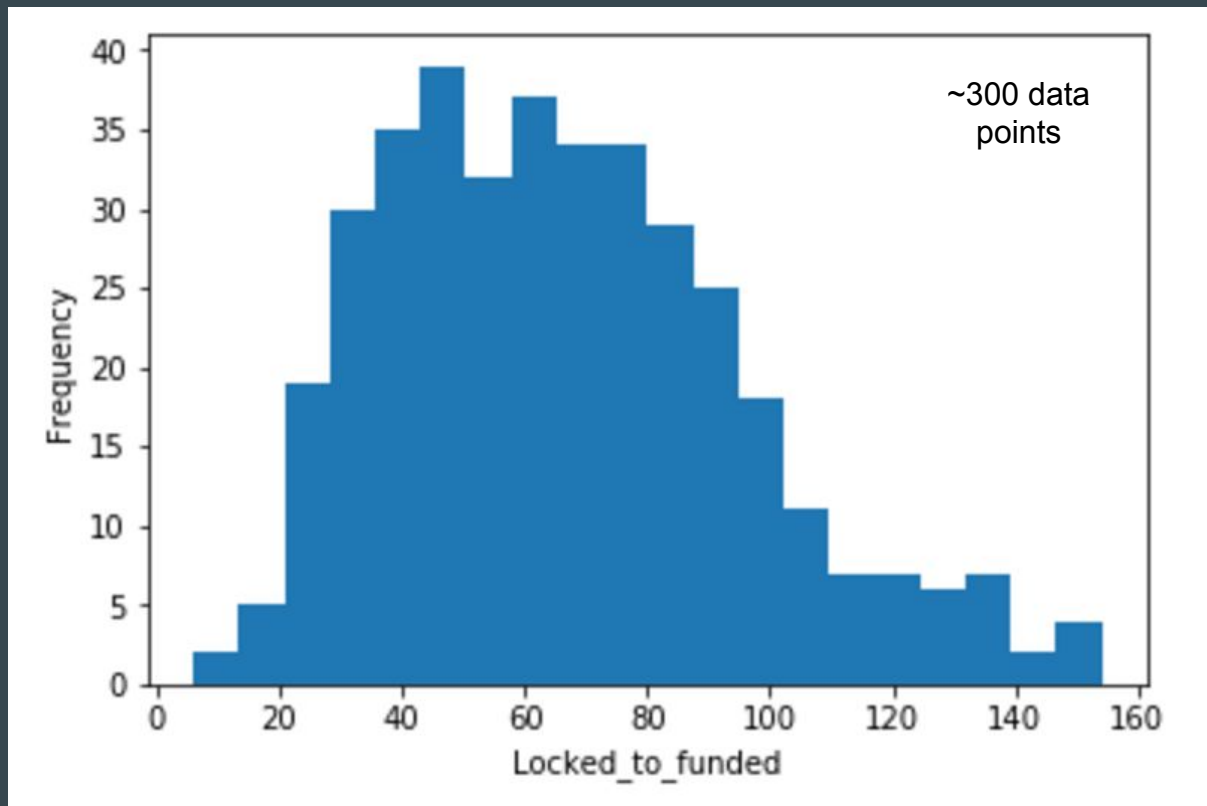
Mission:

- Predict “locked-to-funded” time (efficiency)
- Supervised learning (Regression)

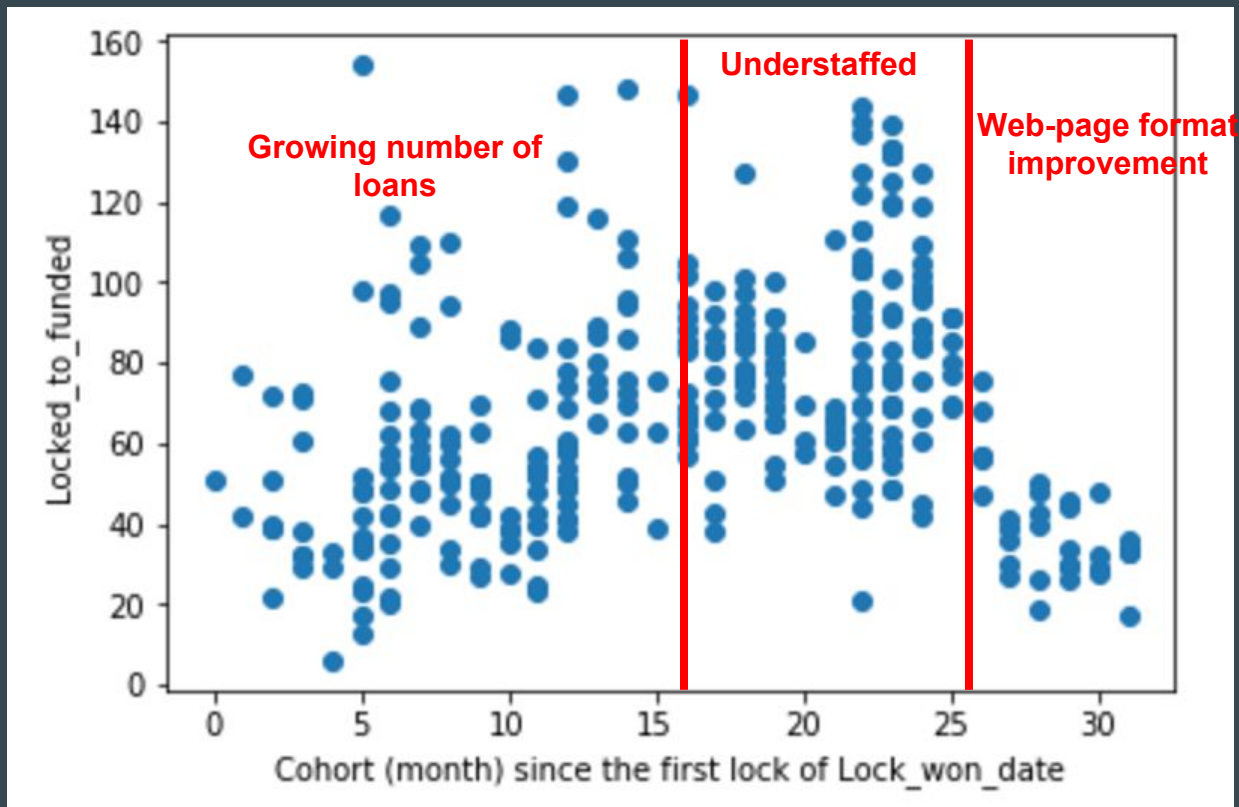
Business Values:

- Improve customer experience by providing an expected waiting time
- Know areas of improvement for efficiency

Part 2 - Loan Processing



Part 2 - Loan Processing



Conclusions

Part 1

- Multistage model construction, Oversampling, Random Forest Classification
- Outperform existing models in Accuracy, F1-score, Precision
- Import feature extracted, Best Customer Profiling

Part 2

- Too few data points (~300) to detect signal for prediction
- Time cohort analysis indicated trends in outcome

Future work

- Implement prediction pipeline into existing workflow
- Include new features (from internal and external sources) into prediction models

Thank you

Wei Cheung

cheungwz@gmail.com

github.com/weizhic