airbnb

Group 3 - Yizhen Luo, Xiao Bai

How much are you expecting to get from your Airbnb Listing?

Key Problems

PROBLEM SCOPE

Airbnb listing revenue prediction

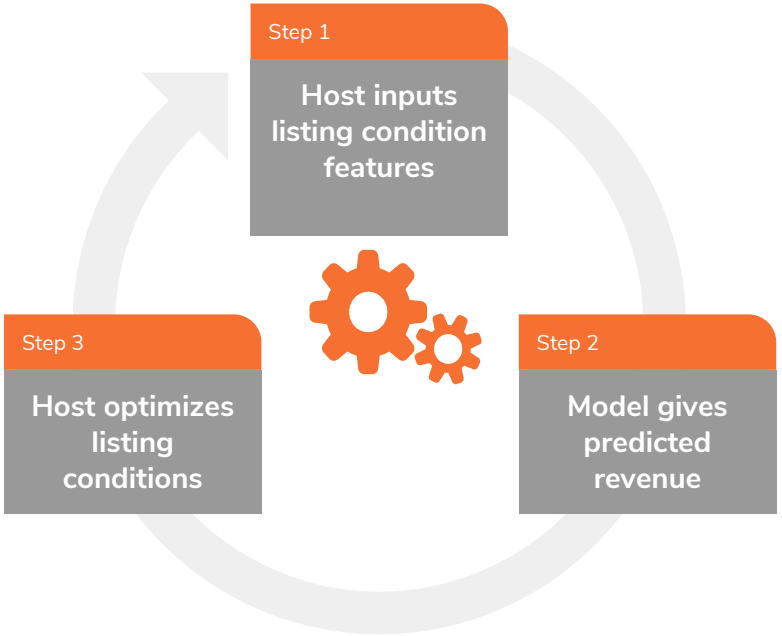
SIGNIFICANCE

FOR HOSTS

Practical tool to maximize revenue by tuning listing features

FOR MARKET OBSERVERS

Insights on consumer preference and behavior, better efficiency and allocation



Dataset:
Beijing

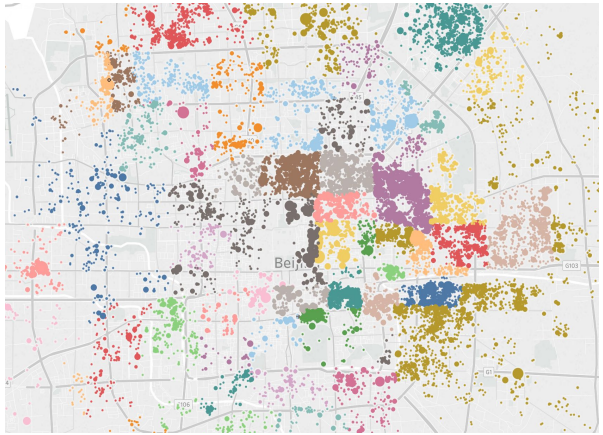
21706 listings * 89 features

Updated: Oct. 16, 2018

insideAirbnb.com

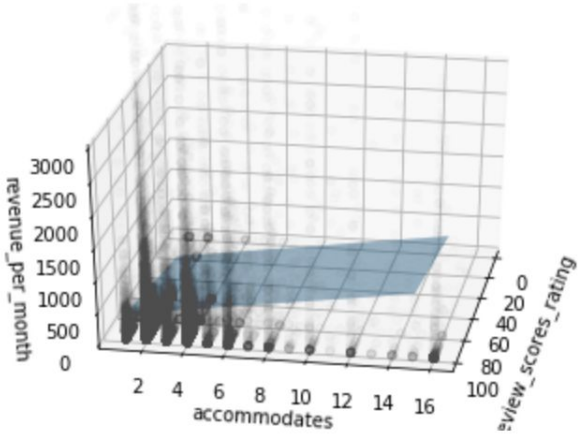
Our Hypothesis through dataset

Initial observation in the dataset: Determinant factors include Neighborhoods, Accommodate, Review Scores, etc.



NEIGHBORHOOD

Larger bubbles (representing monthly revenue) in the city center than suburban areas.



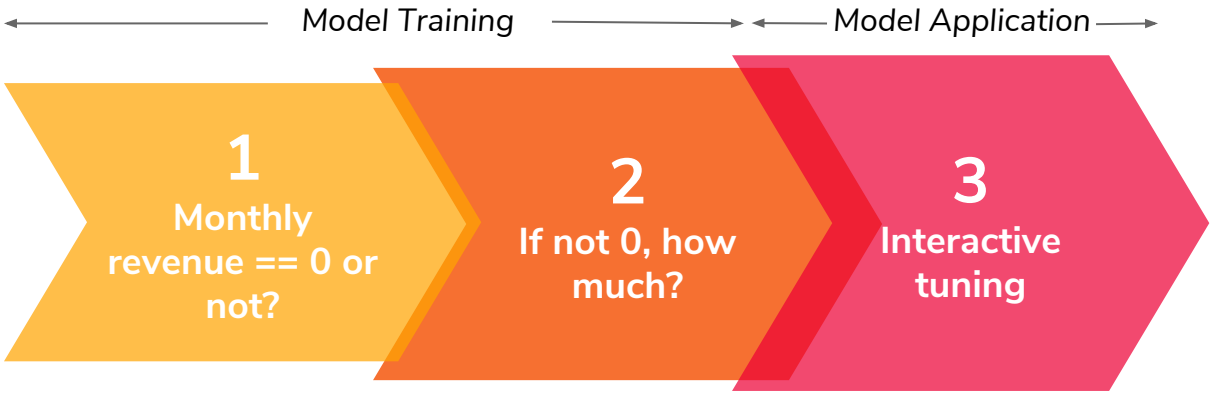
REVIEW SCORES & ACCOMMODATES

Higher review scores, higher accommodates, higher revenue.

Data Preprocessing

Our Approach

40% of the listings in Beijing are not making revenue at all



Classification

- Logistic Reg
- Naive Bayes
- Neural Network
- Random Forest

Accuracy = 70%

Revenue Prediction

- Linear Regression
- Decision Tree
- Bagging
- Boosting
- Stacking

Interactive Mode

- Revenue prediction through real-time parameters input
- Real-time tuning

Model Evaluation Metric

*STEP 1: Classification - **Accuracy***

*STEP 2: Revenue Prediction - **MSPE: Mean Squared Percentage Error***

- Accuracy - work especially well when data set is close to balanced

- MSPE - Invariant to scale: No need to normalize!

STEP 1
Classification

Model Training Results *with CV exp. NN*

1 Monthly revenue == 0 or not?

Logistic Regression

- Training score: 0.6912233748439262
- Testing score: 0.6919996717755733

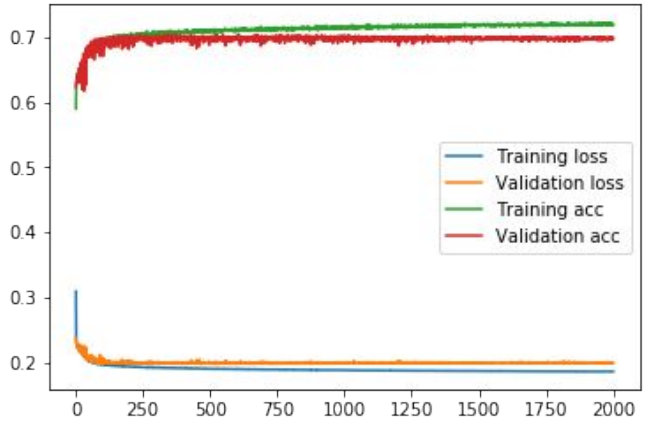
Naive Bayes

- Training score: 0.6650383529343664
- Testing score: 0.662564597603657

Neural Network

- Training score: 0.7288800118537075
- Testing score: 0.7014566751764529

Neural Network *produces the best result so far…*



STEP 1
Classification

Model Training Results *with CV*

1 Monthly revenue == 0 or not?

Random Forest

- Training score: 0.7035061360625914
- Testing score: 0.7029635801891778

AdaBoost

- Training score: 0.7083140468765632
- Testing score: 0.7034106444711669

***AdaBoost** produces the best result, with 70.34% accuracy*

STEP 1
Classification

Model Training Results

1 Monthly revenue == 0 or not?

Step 1 Model Accuracy	Training	Testing
Naïve Bayes	66.50%	66.26%
Logistic Regression	69.12%	69.20%
Neural Network	72.89%	70.15%
Random Forest	70.35%	70.30%
AdaBoost	70.83%	70.34%
Constant Model (c = 0)	39.85%	39.17%
Constant Model (c = 1)	60.15%	60.83%
Random Guess	49.96%	49.86%

All ML models are doing significantly better than ***constant model 0, 1, and random guess***

STEP 2
Revenue Prediction

Model Training Results

2 If not 0, how much?

Linear Regression

```
MSPE train: 8.671572715260696
MSPE test: 9.910066494046127
```

Polynomial Linear Regression (order = 2)

```
MSPE train: 4.526501221258536
MSPE test 336.1048972111548
```

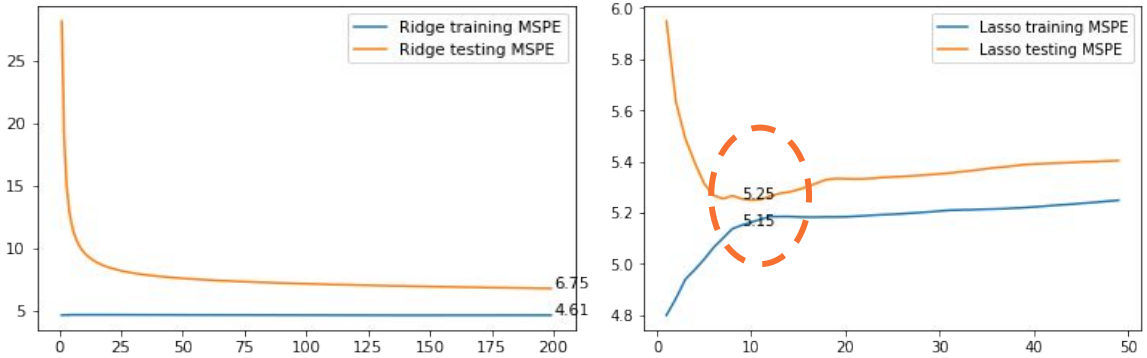
Severely overfits!

STEP 2
Revenue Prediction

Model Training Results

2 If not 0, how much?

Polynomial Linear Regression with Regularization




```
MSPE train: 5.151295614181206
MSPE test: 5.25409806583349
```

SGD LR with CV yields similar result...

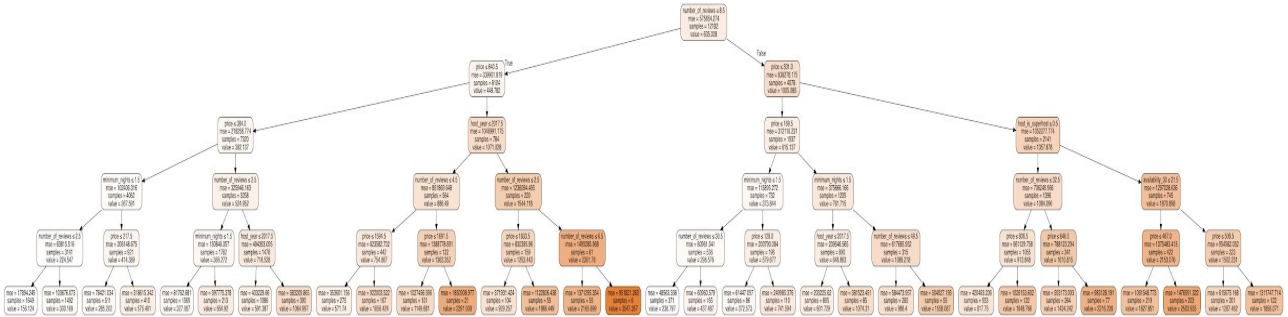
***Polynomial LR with Lasso Regularization alpha = 9** performs the best so far...*

STEP 2
Revenue Prediction

Model Training Results

Decision Tree (Regression Tree)

2 If not 0, how much?



A regression tree of depth 5

STEP 2
Revenue Prediction

Model Training Results

2 If not 0, how much?

Decision Tree (Regression Tree)

```
MSPE train: 0.00050992085843816 89
MSPE test: 6.261112703727595
```

Completely overfits!

Decision Tree through GridSearchCV tuning maximum depth and minimum sample per leaf

```
MSPE train score: 2.334169333770033
MSPE test score: 4.321409383536813
```

Better…

STEP 2
Revenue Prediction

Model Training Results - Ensemble

2 If not 0, how much?

Bagging - Random Forest through GridSearchCV

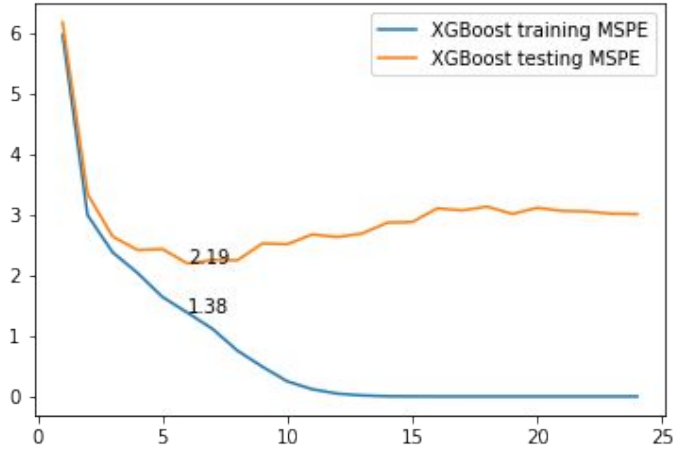
```
{'max_depth': 10, 'min_samples_leaf': 10}

MSPE train score: 2.3465543959061312
MSPE test score: 3.4341842442484145
```

STEP 2
Revenue Prediction

Model Training Results - Ensemble

Boosting - XGBoost tuning max_depth

2 If not 0, how much?




```
MSPE train: 1.380951715588963
MSPE test: 2.193337226402705
```

XGBoost Regressor with Max_depth = 6 *gives the best performance so far…*

STEP 2
Revenue Prediction

Model Training Results - Ensemble

2 If not 0, how much?

Stacking


```
MSPE train: 1.3367183901824902
MSPE test: 2.158744966789093
```

*Stacking **Random Forest** with **XGBoost** brings improvement*

STEP 2
Revenue Prediction

Model Training Results

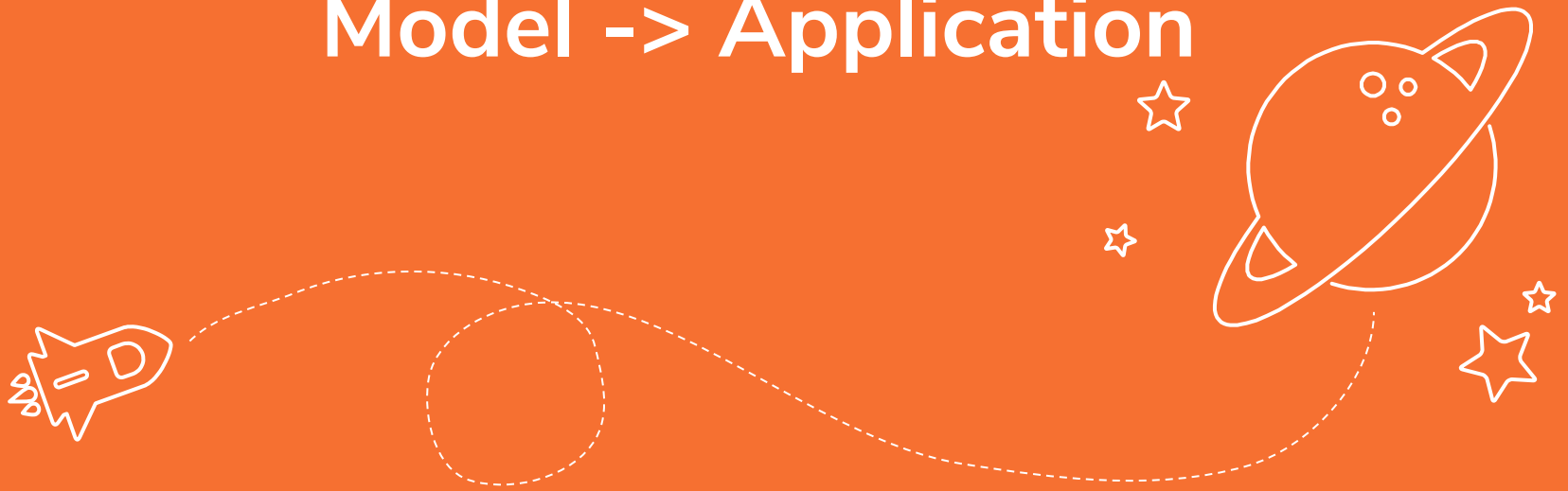
2 If not 0, how much?

Step 2 Model MSPE	Model Parameters	Traing	Testing
Linear Regression	-	8.67	9.91
Polynomial LR	*Lasso, alpha = 9*	5.15	5.25
Decision Tree	*max_depth = 10, min_samples_leaf = 5*	2.33	4.32
Random Forest	*max_depth = 10, min_samples_leaf = 10*	2.35	3.43
XGBoost	*max_depth = 6*	1.38	2.19
Stacking	*Random Forest + XGBoost*	1.34	2.16
Constant Model	*c = mean*	31.94	32.76

*We choose **stacking model** for Step 2.*

STEP 3

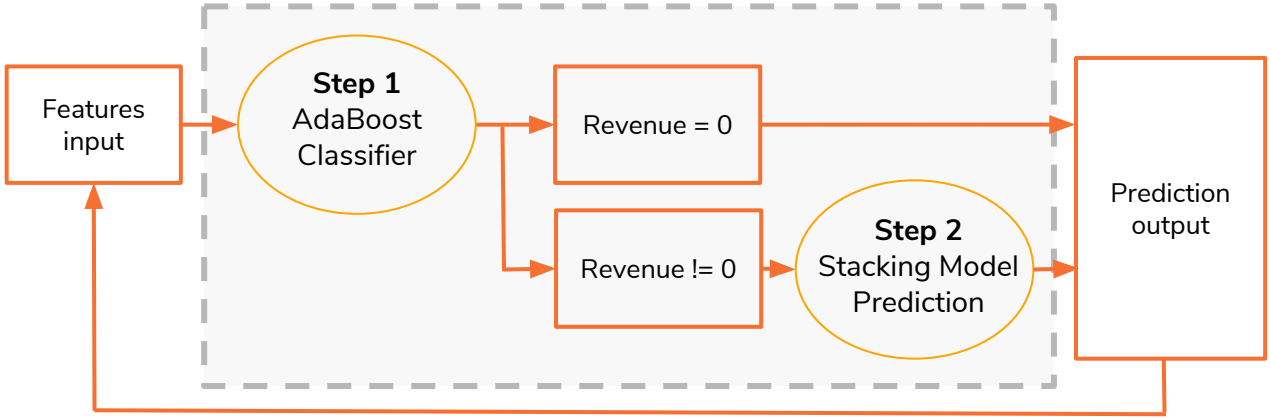
Model -> Application



STEP 3
Interactive Tuning

Compiling two models

Interactive tuning



STEP 3
Interactive Tuning

Interactive Tuning Interface

Interactive tuning

The expected monthly revenue of my listing is: ¥ 419.6476799855554

About My Listing Initial Setup:

My listing neighbourhood is... Dongzhimen

My listing property type is... Apartment

My listing room type is... Entire home/apt

My listing can accommodate # people... 2

My listing has # bathrooms... 2

My listing has # bedrooms... 2

My listing has # beds.. 2

My listing bed type is... Real Bed

The expected monthly revenue of my listing is: ¥ [498.87791317]

About Me:

I have been an Airbnb host since ... 2017

I respond to my guests... within an hour

I am Airbnb Super Host ...

I have # listings... 3

The expected monthly revenue of my listing is: ¥ [498.87791317]

About My Listing Initial Setup:

My listing neighbourhood is Dongchimen

Future Work

If we had more time...

Better model

Listing description

Review ***sentiments***

Listing/host ***picture***

...

Tuning Suggestion

Recommend which feature to tune and how to tune to improve the expected revenue most easily

Suggested pricing

2 private rooms instead of 1 entire apartment

...

Real-time data updates

Update the dataset with Airbnb server in real time

Questions?

airbnb

REFERENCE

P. Choudhary, A. Jain, and R. Baijal, “Unravelling Airbnb Predicting Price for New Listing,” arXiv:1805.12101 [q-fin], May 2018. [Online serial]. Available: {http://arxiv.org/abs/1805.12101}. [Accessed Dec. 5, 2018].

K. Calixte, J. Curran, L. Y. S. Lai, and E. Ram, “Airbnb: Predicting Loyalty,” Semantic Scholar, 2016. [Online]. Available: {https://bit.ly/2UjJsEB}. [Accessed Dec. 5, 2018].

E. Tang and K. Sangani, “Neighborhood and Price Prediction for San Francisco Airbnb Listings,” Stanford School of Engineering, Stanford University, Stanford, CA, 2015. [Online serial]. Available: {http://cs229.stanford.edu/proj2015/236_report.pdf