Price Prediction of Major Rental Platforms in Shanghai via Machine Learning Methods

Machine Learning with
Professor Enric Junqué de Fortuny

Eden Wu | Xinyao Han | Yile Xu

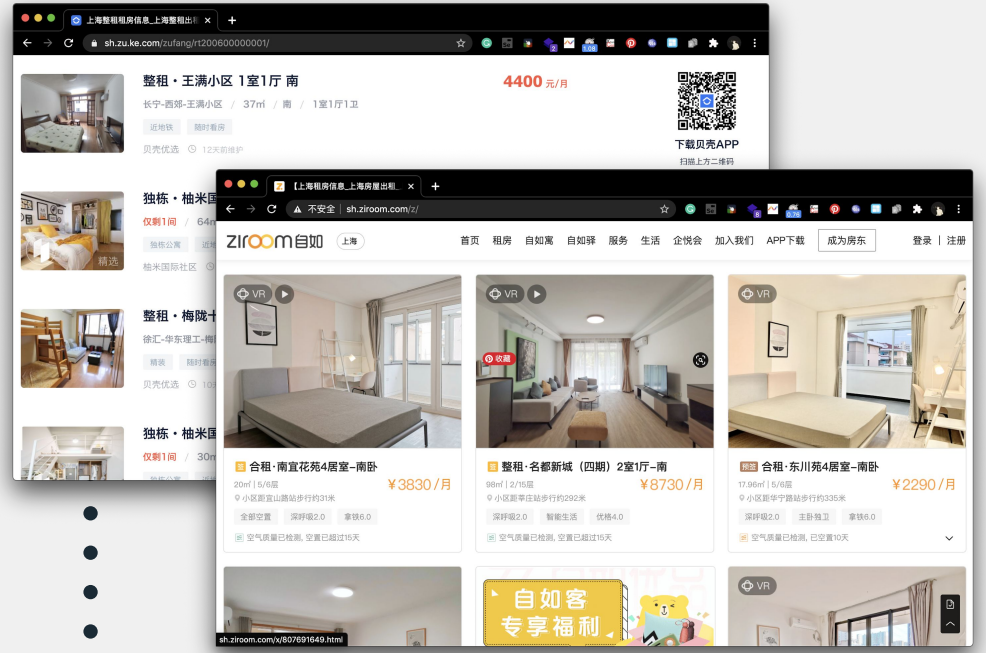
TABLE OF CONTENTS

01.

Problem Statement

Indicate Fake Rental Listings with Rental House Price Predictions

PROBLEM STATEMENT



PRICE ESTIMATION

The average price with respect to community, floor, decoration, area etc.

PRICE PREDICTION

02.

DESCRIPTIVE ANALYSIS

Dataset and important features analysis

Descriptive Analytics I

Crawled Data
- Encryptors
- OCR

Baidu Map API
- Lng and Lat
- Subway info

H3 location Index

Dataset Retrieving Process

Y variable

	count	mean	std	min	25%	50%	75%	max
Price	19137	10824.97	9109.47	1020.00	4300.00	670.00	16000.00	40000.00
Area	19137	102.97	71.58	6.00	51.00	74.00	140.00	902.00
Longitude	17676	121.457	0.115	120.458	121.391	121.455	121.531	121.926
Latitude	17676	31.214	0.122	30.720	31.179	31.218	31.238	31.825
Bedrooms	19137	2.34	1.41	0.0	1.0	2.0	3.0	9.0
Livingrooms	19137	1.43	0.62	0.0	1.0	1.0	2.0	6.0
Bathrooms	362	1.64	0.62	0.0	1.0	1.0	2.0	7.0
Floor	16236	16.35	11.11	1.0	6.0	17.0	27.0	121.0
NextToSubway	19137	0.56	0.50	0.0	0.0	1.0	1.0	1.0
Exquisite	11763	0.69	0.46	0.0	0.0	1.0	1.0	1.0
OpenForVisits	11763	0.19	0.39	0.0	0.0	0.0	0.0	1.0

Table 1 Data Description Without Dummies

Descriptive Analytics II

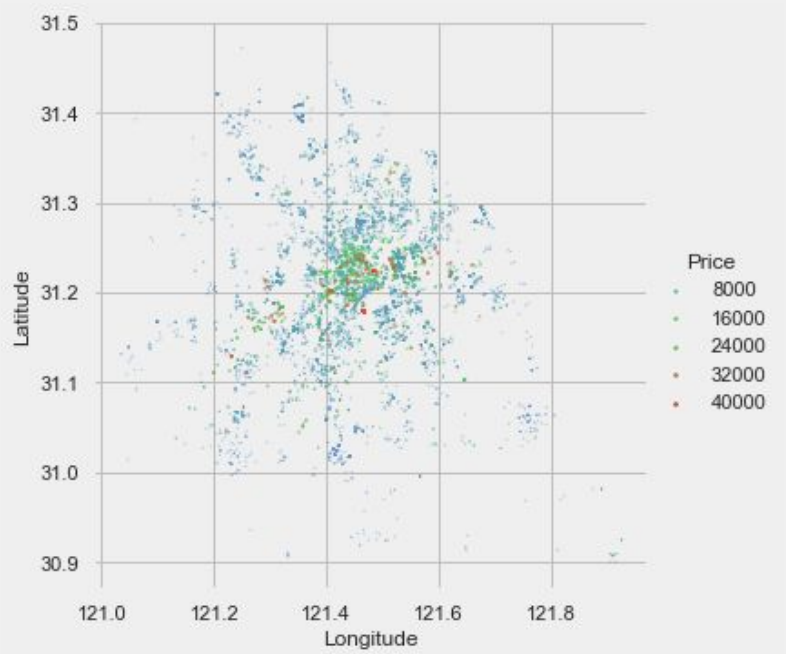


Fig.2. Y Distribution with Longitude and Latitude

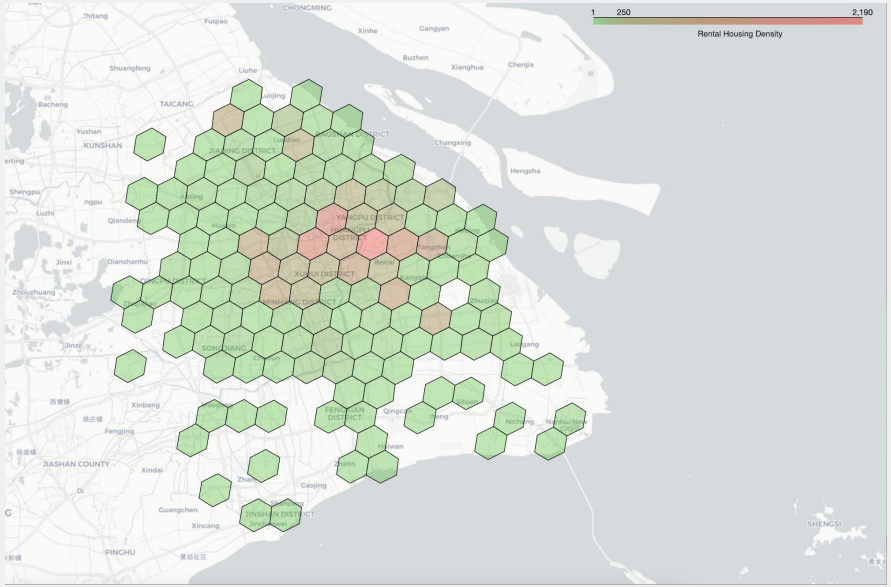


Fig.3. H3 Index and Rental Density Visualization in Shanghai

Descriptive Analytics III

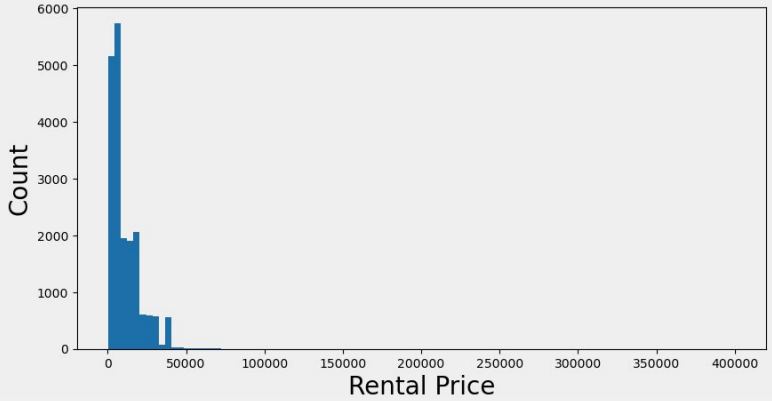


Fig.4. Y Distribution Before Clipping

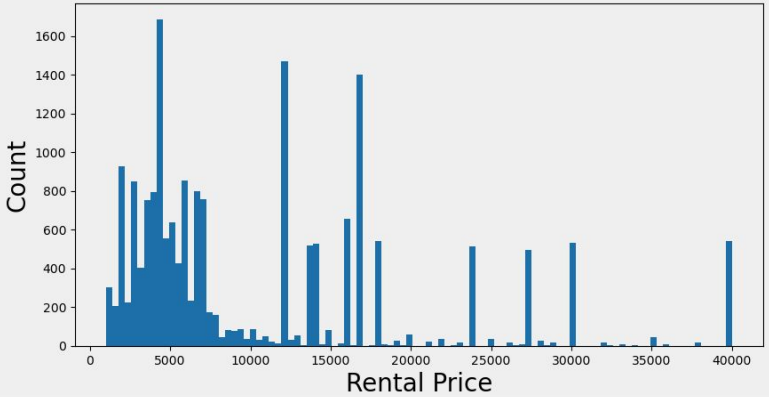


Fig.5. Y Distribution After Clipping

03.

Machine Learning

Using RF, GBR, Extra
And Ensemble Methods

Machine Learning

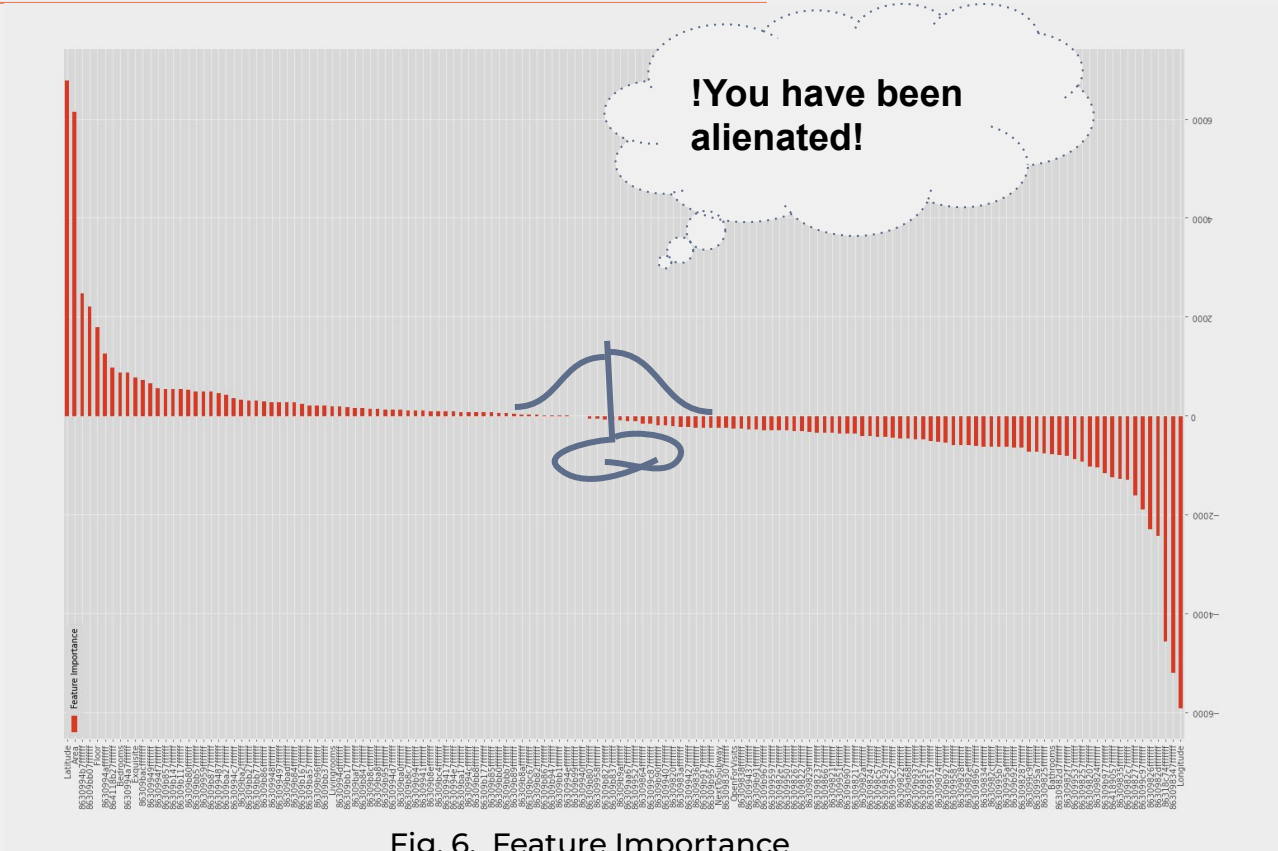
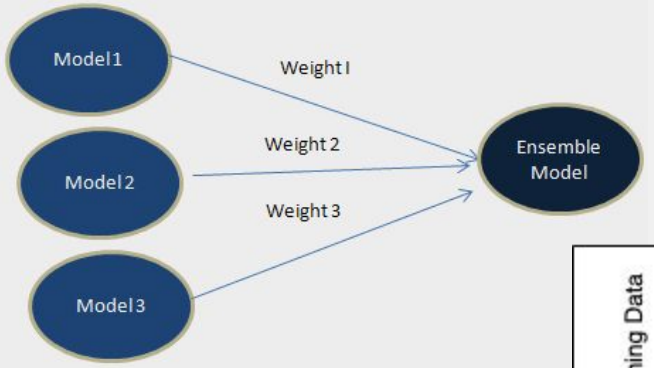


Fig. 6. Feature Importance

Mean Value	MAE
Linear	5259166187082 88.19
Ridge	2242.85
Lasso	2240.55
RF	449.21
GBR	912.14
LinSVR	3606.98
Ela	2786.13
Bay	2241.18
Ker	28585.18
Extra	435.86

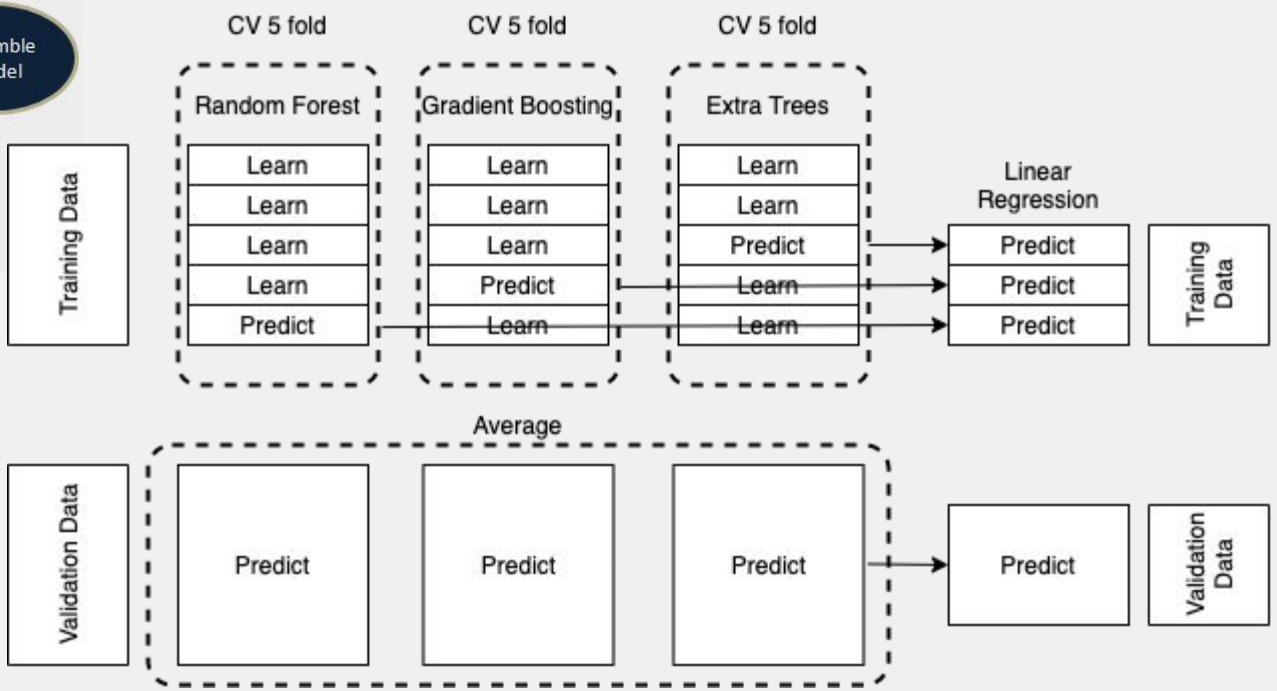
Fig. 7. MAE After Rough Estimation

Machine Learning



Weighted Average

Stacking



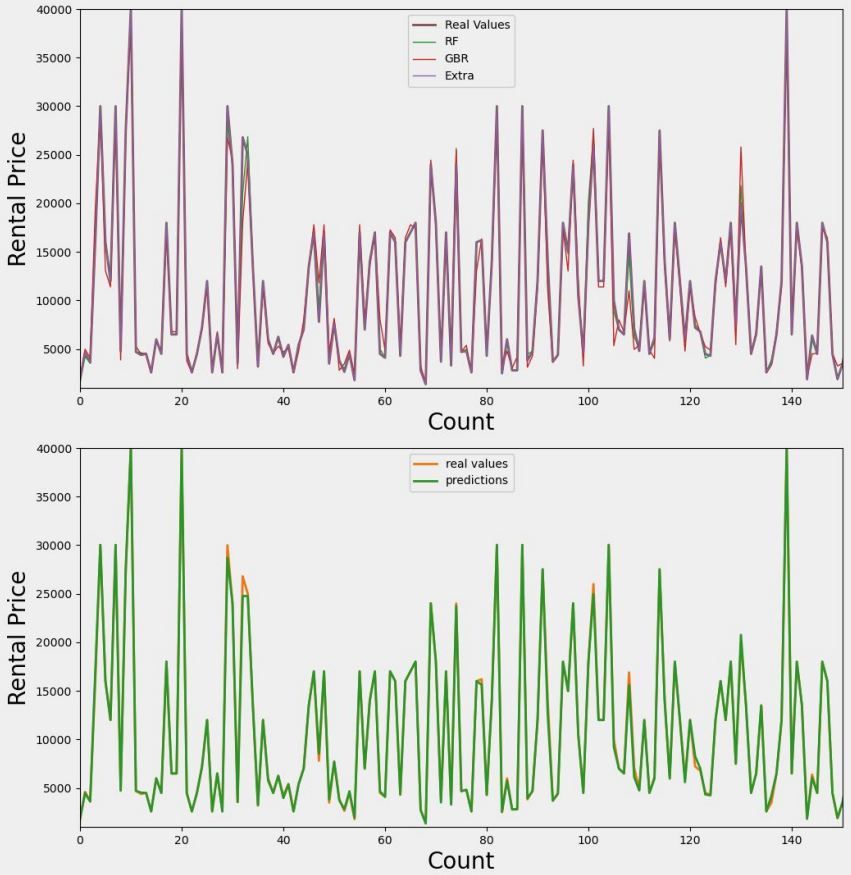
Machine Learning

PREDICTION RESULT

Model	MAE	
	Training Set	*Validation Set*
Mean Price	7271.53	7126.62
Kernel Ridge	882.22	988.19
Random Forest	449.66	613.62
Gradient Boosting	474.60	636.98
Extra Trees	436.77	594.99
Weight Average	447.68	584.92
Stacking Model	429.25	587.57

Table 2 Prediction Results

Fig. 8. Prediction Accuracy Before and After Fine Tuning and Ensemble



04.

Future Work

Optimized Price Prediction
And Fake Rental Housing Identification

Future Work

Price Prediction:

- Coupling effect of multiple regression models
- Re-learn ability of machine learning models
- Combination of Machine Learning and Deep Learning methods
- Driven factors for the good performance of tree-based models
- Faster ways to fit complex models

Fake Rental Housing Identification:

- Datasets w/ “real” and “fake” labels
- Brand new machine learning models for clustering

Future Work


IN FUTURE WORK
EVERYTHING IS POSSIBLE
Mon
Tue-Thur
Fri -Sat
Sunday
EMEFKA.SK


Adventuretime-blog
I'M SUPER PUMPED!

THANKS!

Our GitHub code:

https://github.com/EdenWuyifan/Fake-Rental-Listing-Identification

Acknowledgement:

We thank our professor Enric Junqué de Fortuny from NYU Shanghai Computer Science Department, who provided insight and expertise that greatly assisted the project.