Feature-Based Airbnb Price Prediction

GeekHub
Brian Mike Tony Logan

Inspiration

Since ... Short-time leasing becomes a trendy lifestyle for nowadays' generation.

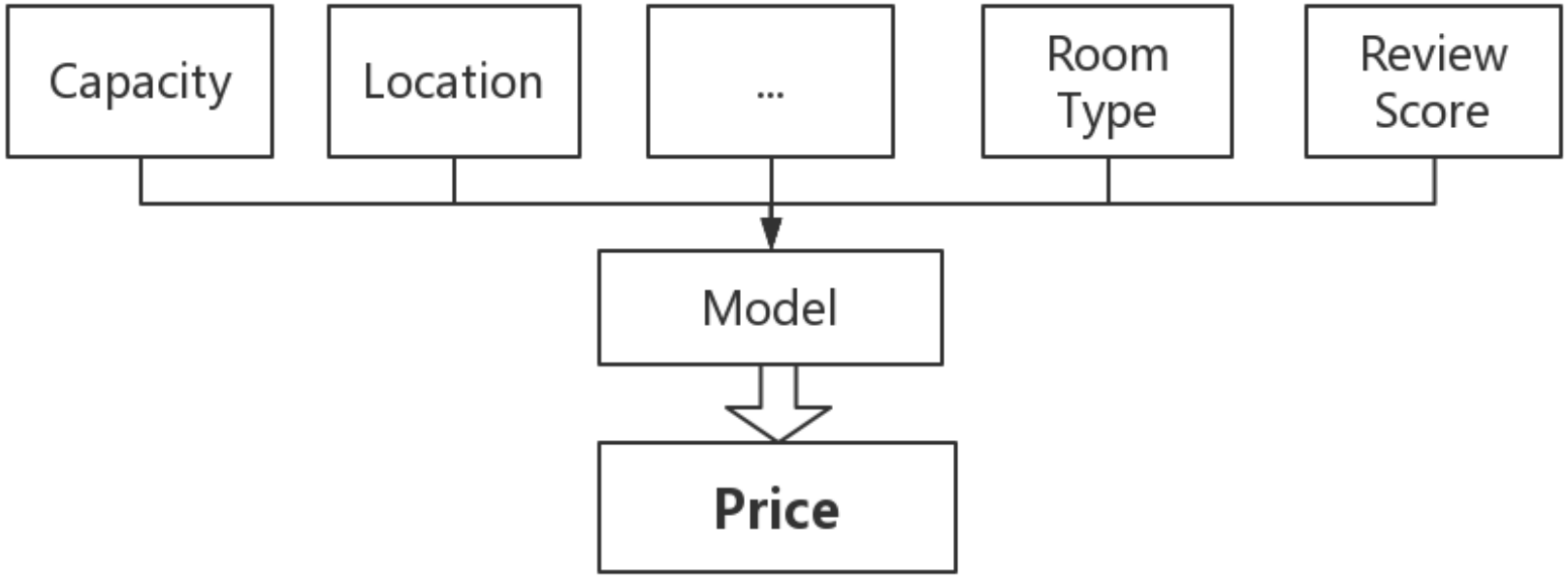
As a result … Airbnb has a booming market in Beijing with increasing joiners.

But … New hosts often have a hard time on properly pricing their property

Problem Statement

Given a series of data describes the property's features

Output the reasonable/best price point for the host.



Data Pre-Processing

- Drop Properties with no reviews (i.e. No people rent it/pay the price yet)
- Drop Properties that lacks to many features (less than 100 rows of them)
- Drop outliers handpicked via inspection (i.e. Check the Actual id on www.airbnb.com)
- Drop Properties with unreasonable prices (i.e. the host didn't actually want to Airbnb it)

~~[price < (50 * accomondates) OR (50 * beds)]~~

~~[Price > 20,000 OR price = 9999]~~

~~[beds=50 OR bathrooms=101 OR min_nights > 3]~~

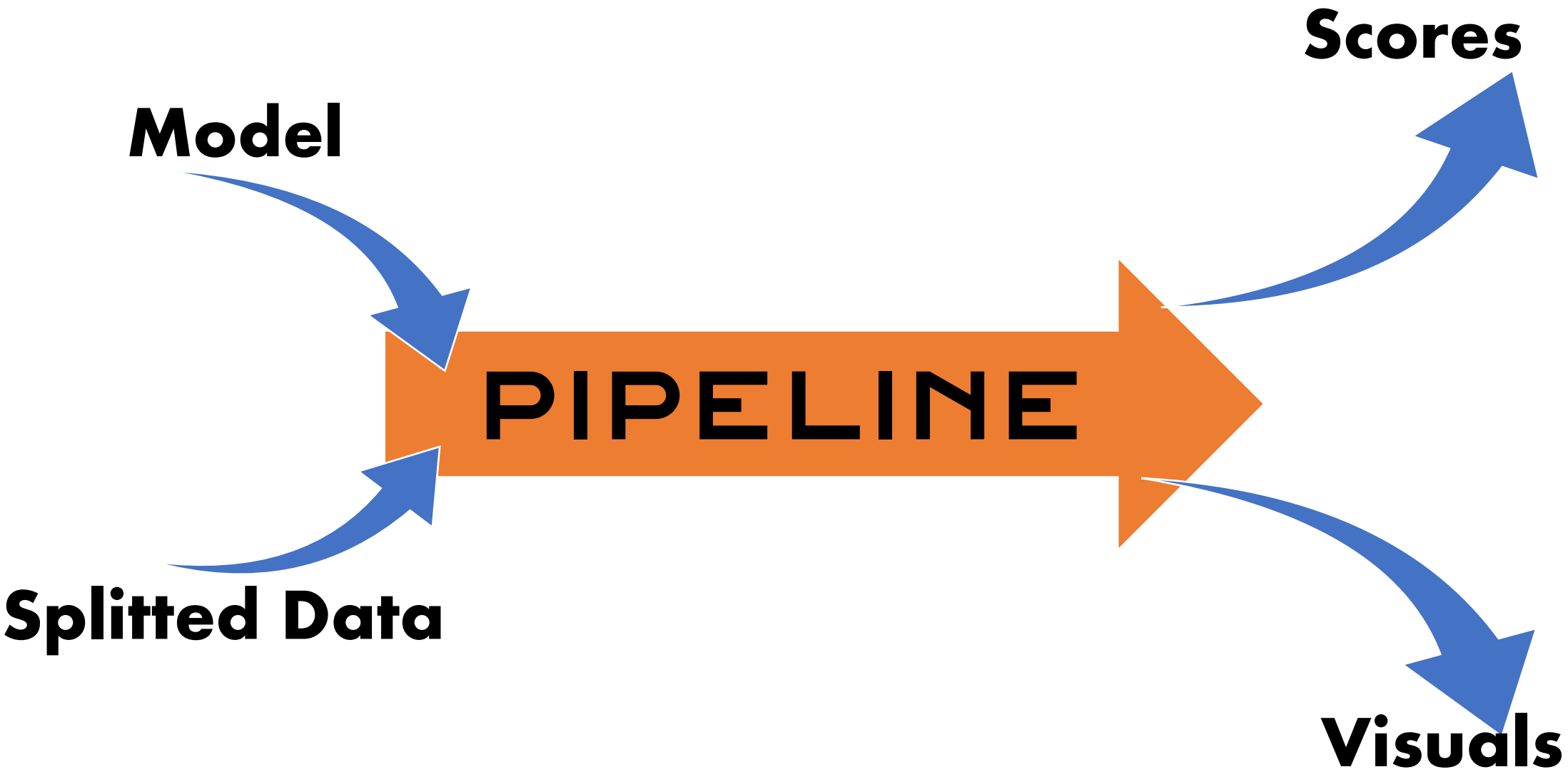
- Fill up a few fixable nulls.

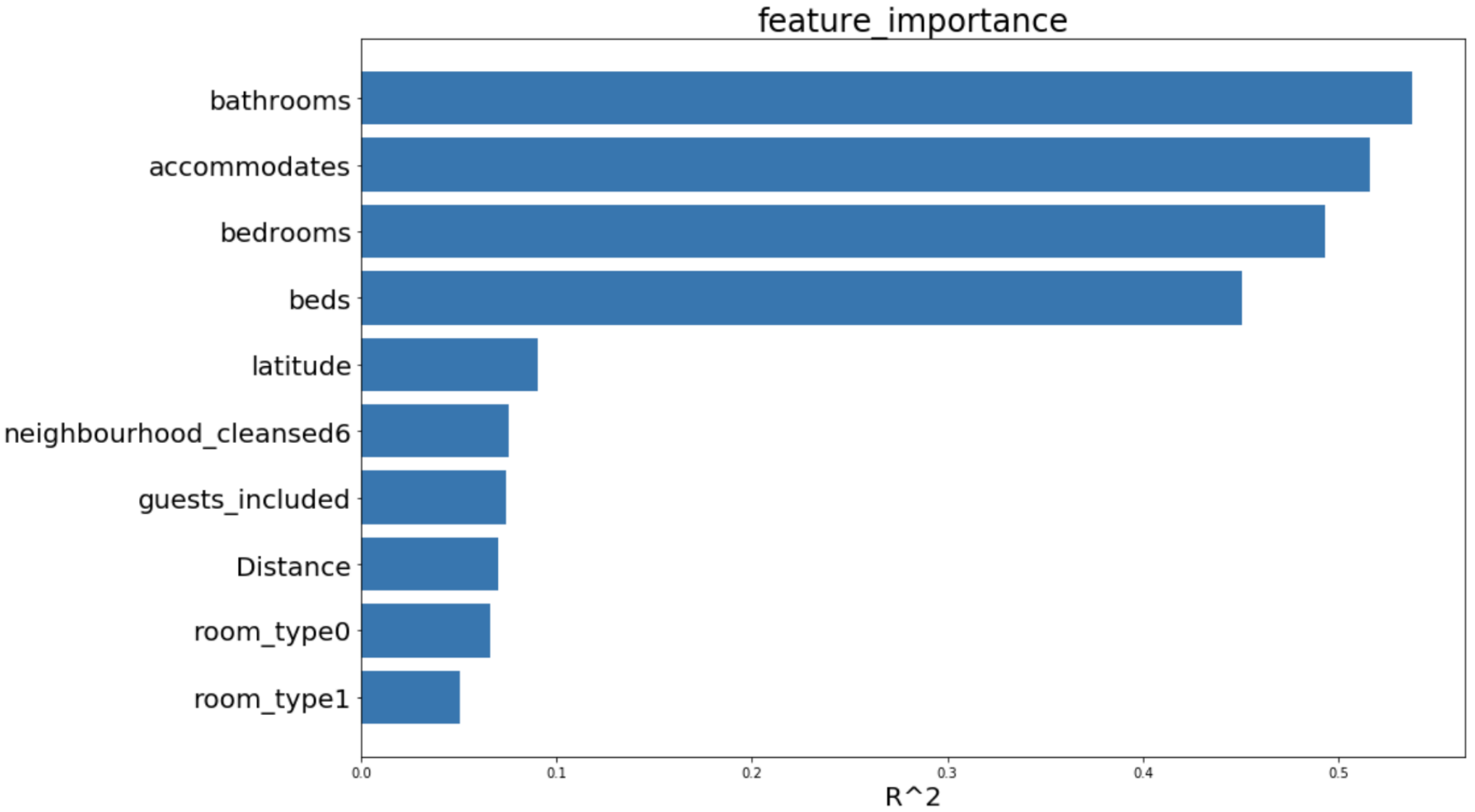
Feature Engineering

- Feature Selection - Drop features we're uncapable to deal with or irrelevant

All words	Irrelevants	Missing values
Summary Abstract	The landlord ID Host since	Weekly price Square feet
e.g. The neighborhood has a …	e.g. 192875 e.g. 527062	All none

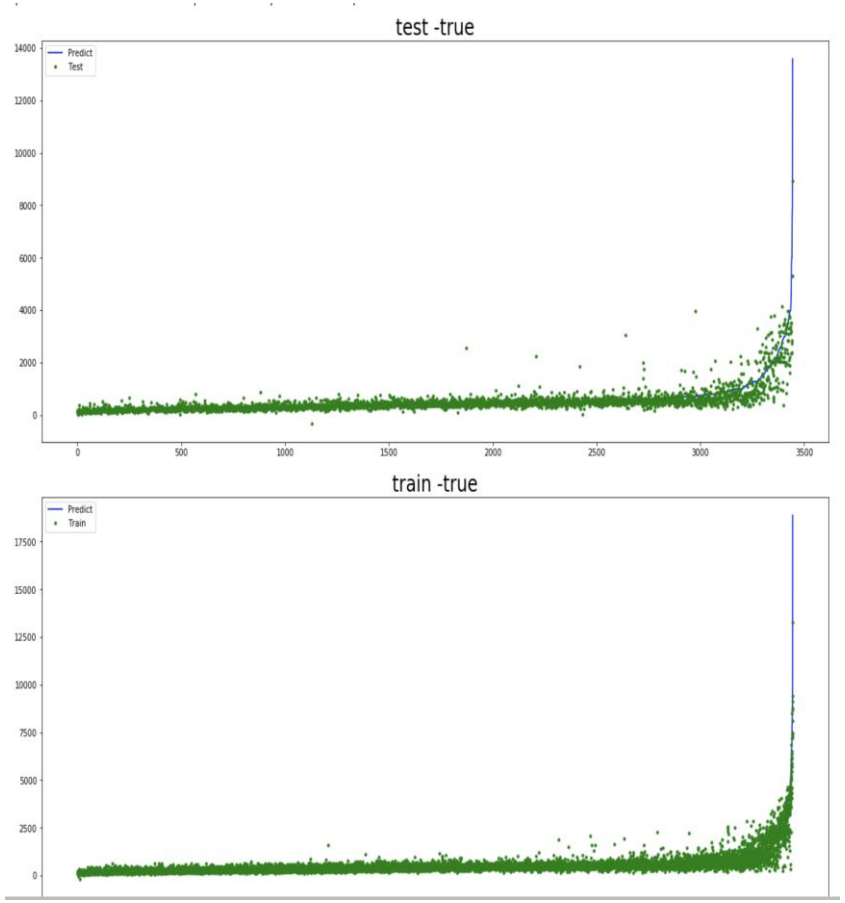
- One-hot Encoding - Transform categorical features into numericals
- Feature Generating - Adding new features (Transform coordinates to distances)


Scores
Model
PIPELINE
Splitted Data
Visuals

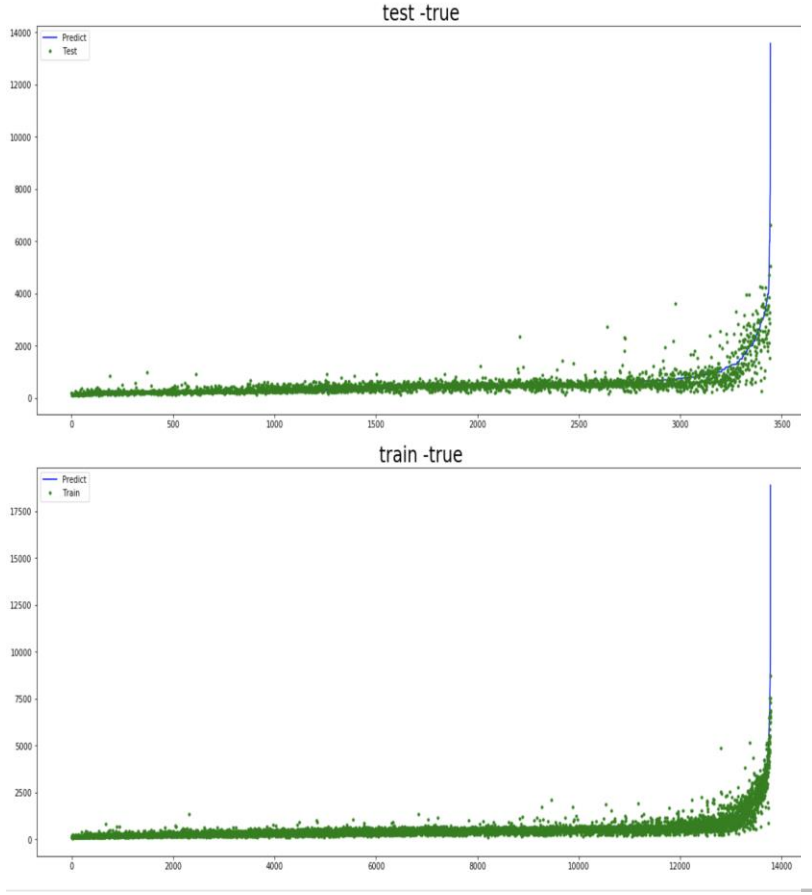

feature_importance
bathrooms
accommodates
bedrooms
beds
latitude
neighbourhood_cleansed6
guests_included
Distance
room_type0
room_type1
0.0
0.1
0.2
0.3
0.4
0.5
R^2

Model Name	R^2 Score	MSE
GradientBoostingRegressor	0.7309	93735.296
RandomForestRegressor	0.7005	104310.6571
BaggingRegressor (base on decision tree)	0.6625	117552.8415
NerualNetwork	0.6220	127463.1387

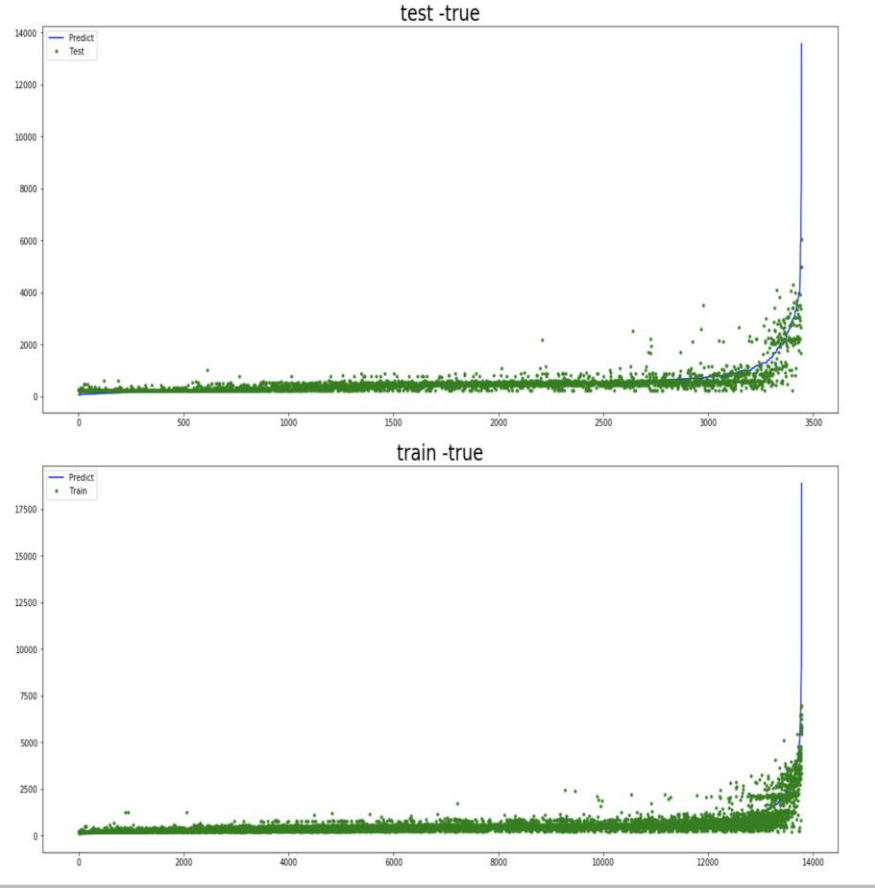
GradientBoostingRegressor



RandomForestRegressor



BaggingRegressor (base on decision tree)



Thank you!