CLASSIFIED

Classified is Jordi Beernink, Roel Bouman, Jeffrey Luppes, Gerdriaan Mulder and Thijs Werrij

Sberbank Russian Housing Market Competition

https://www.kaggle.com/c/sberbank-russian-housing-market

Introduction - The competition

Predict prices of realty in Moscow area

Data set consists of:

- train.csv, 30473 entries with information about houses/apartments
- macro.csv, data on Russia's macroeconomy and financial sector

Data is from 2011 to 2015, while the test set is from 2015 to 2016

Introduction - Data exploration

Approach - Pre-processing

- Transform the data using one-hot encoding for categories, but preserving numerical data

- [missing data?]

Approach - Testing different methods

Split the training data between entries before 2015 and in 2015, for internal validation

Tested several methods:

- LinearRegression
- SGDRegression
- KNNeighbours
- Decision Tree
- RandomForests
- Gradientboosting Regressor
- Deep Learning (Keras)
- XGBoost

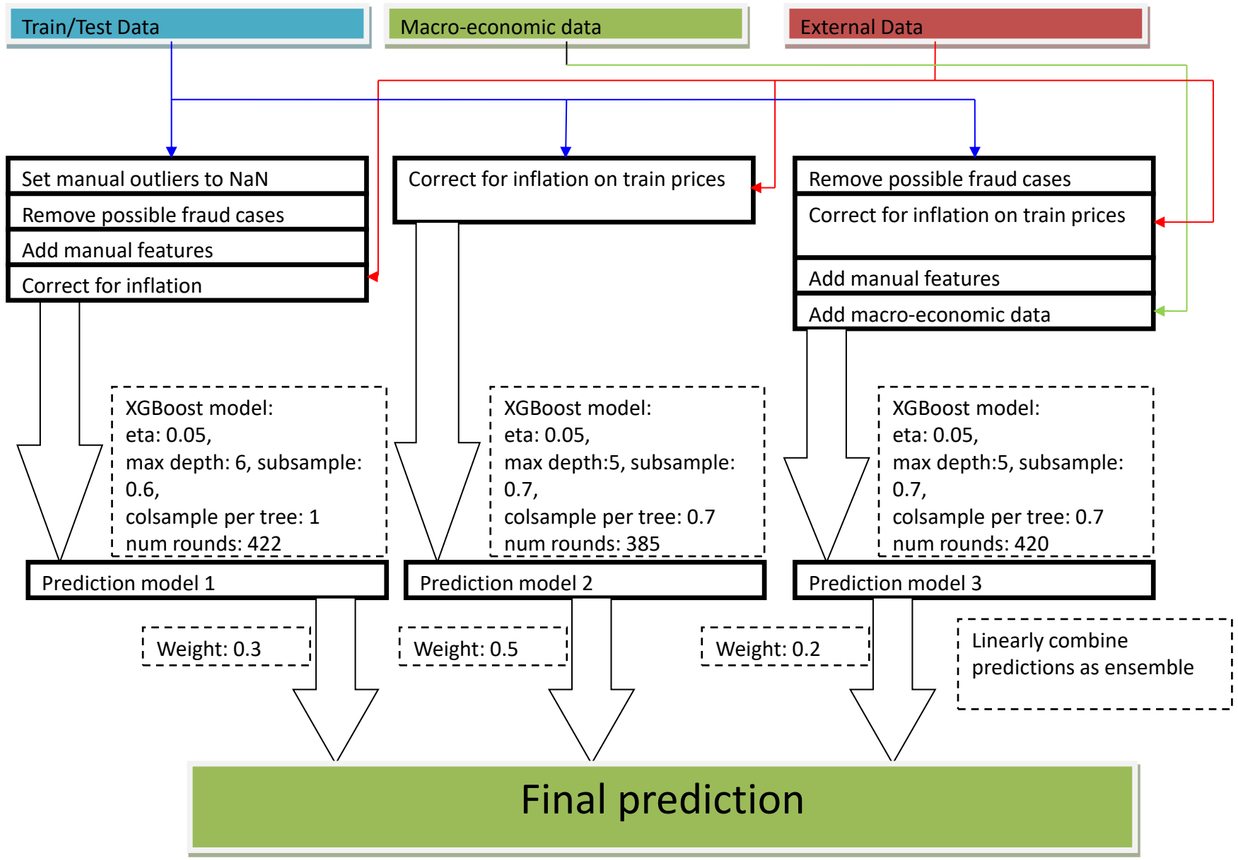
XGBoost outperformed all other methods, which later became even more notable…

Approach – XGBoost ensemble

Kernel published by Andy Harless

Widely used in the competition

Improved and outperformed the original



Results

Classifier	Method	RMSLE	Kaggle Rank (3077)
Benchmark Submission	Naïve XGBoost	0.67333	3034
XGBoost	Single	0.32575	1856
XGBoost	Ensemble with 3 models	0.31062	266
XGBoost	Ensemble with 3 models Readjustment of submission	0.31051	168
Deep Learning	Dense and Dropout	0.46745	2870
Random Forest	Naïve	6.12138	3072
SGD Regressor	Naïve	0.59560	3021
Linear Regression	Complete dataset Only 2015	0.49689	2897
KNNeighbors	Amount of neighbors (6) Removing outliers Only 2015	0.93122	3050
Decision Tree	Complete dataset Only 2015	0.58460	3020
Random Forest	Removing outliers Only 2015	0.75239	3040
Gradientboosting Regressor	Complete dataset Only 2015	0.41384	2767

- (discuss other things worth mentioning here)

Further work

Conclusion

Team CLASSIFIED

Sberbank Russian Housing Market Competition

Maybe a short

Summary of stuff

Discussed

Thank you for your attention! Any questions?