Machine Learning Final Project

House Price Estimation In Piraeus Using Machine Learning

Presentation by
Dimitrios Patiniotis Spyropoulos

Introduction

The nature of the Housing Market

A house/shelter is widely accepted to be an essential human need.

Demand for housing is either due to the fact that houses are a great investment, as they have a proven intrinsic value, or simply because they provide shelter.

The role of locality

Locality plays a critical role not only in the property price, but is also correlated with other housing attributes.

Data Used

Data Crawling

In the context of this project, data is been crawled daily by www.xe.gr

Evaluating the data

Since data does not come from an API, or at least an official organization, it must be carefully evaluated – there is a high risk of being noisy.

Preparing Data

Categorical Encoding – Label vs One Hot
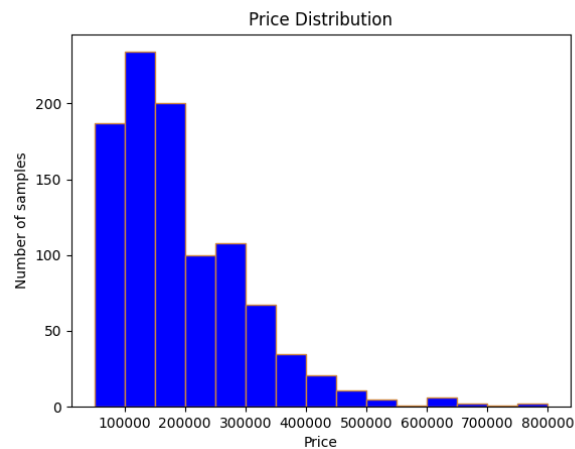
Label Encoding

MACHINE LEARNING ALGORITHMS	MSE	R SQUARED	RMSE
Bayesian Ridge Regressor	0.008	0.613	0.089
Support Vector Regressor	0.011	0.475	0.106
Random Forest Regressor	0.007	0.667	0.084
K-Neighbors Regressor	0.011	0.493	0.104
Linear Regressor	0.009	0.587	0.089

One Hot Encoding

MACHINE LEARNING ALGORITHMS	MSE	R SQUARED	RMSE
Bayesian Ridge Regressor	0.008	0.598	0.089
Support Vector Regressor	0.010	0.506	0.098
Random Forest Regressor	0.006	0.718	0.074
K-Neighbors Regressor	0.010	0.497	0.099
Linear Regressor	0.008	0.592	0.089

# Preparing Data

The case for Standard Scaling

Price Distribution

Standard Scaling

MACHINE LEARNING ALGORITHMS	MSE	R SQUARED	RMSE
Bayesian Ridge Regressor	0.303	0.608	0.550
Support Vector Regressor	0.461	0.413	0.679
Random Forest Regressor	0.363	0.680	0.673
K-Neighbors Regressor	0.550	0.288	0.742
Linear Regressor	0.305	0.612	0.553

Algorithms and Results

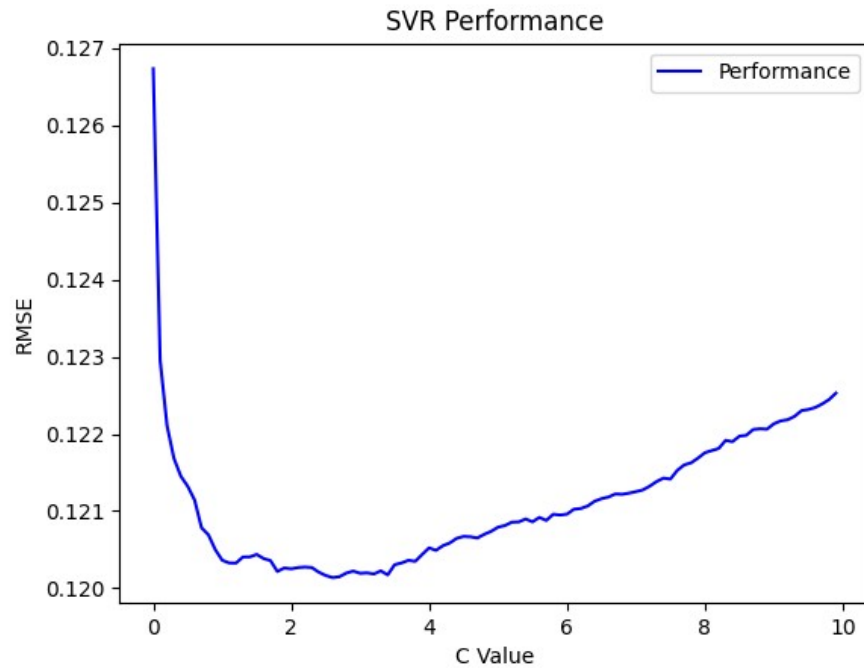| MACHINE LEARNING ALGORITHMS	MSE	R SQUARED	RMSE
Bayesian Ridge Regressor	0.008	0.598	0.089
Support Vector Regressor	0.010	0.506	0.098
Random Forest Regressor	0.006	0.718	0.074
K-Neighbors Regressor	0.010	0.497	0.099
Linear Regressor	0.008	0.592	0.089

Algorithms and Results
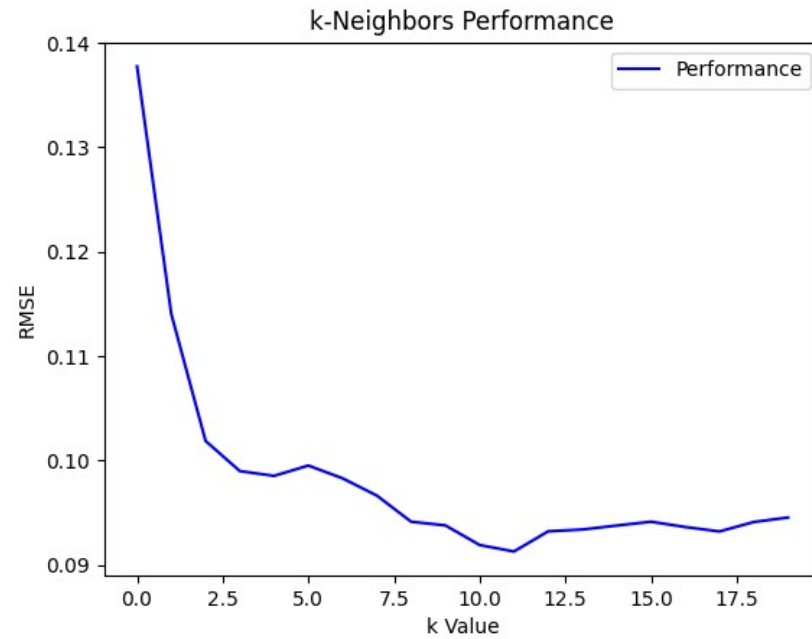
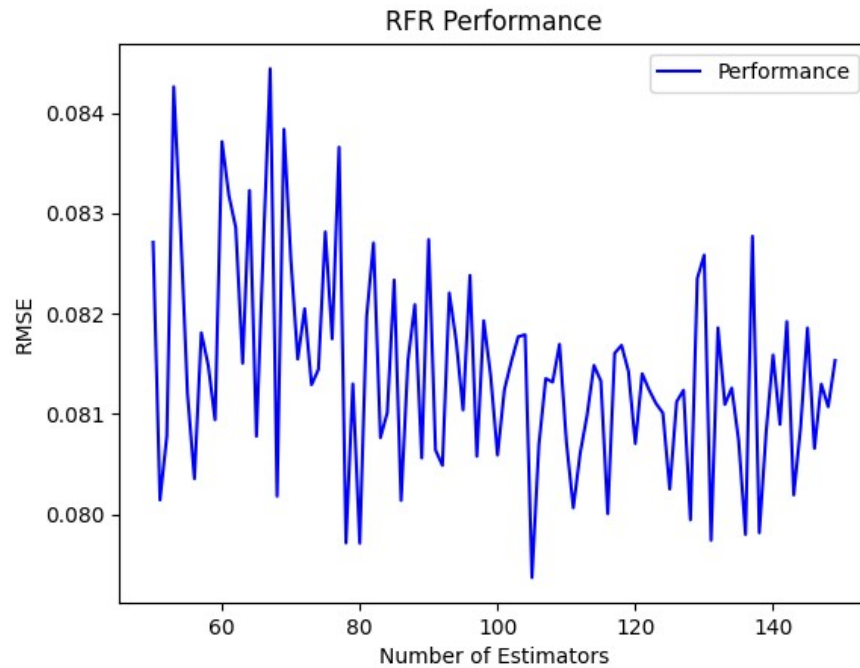# Algorithms and Results
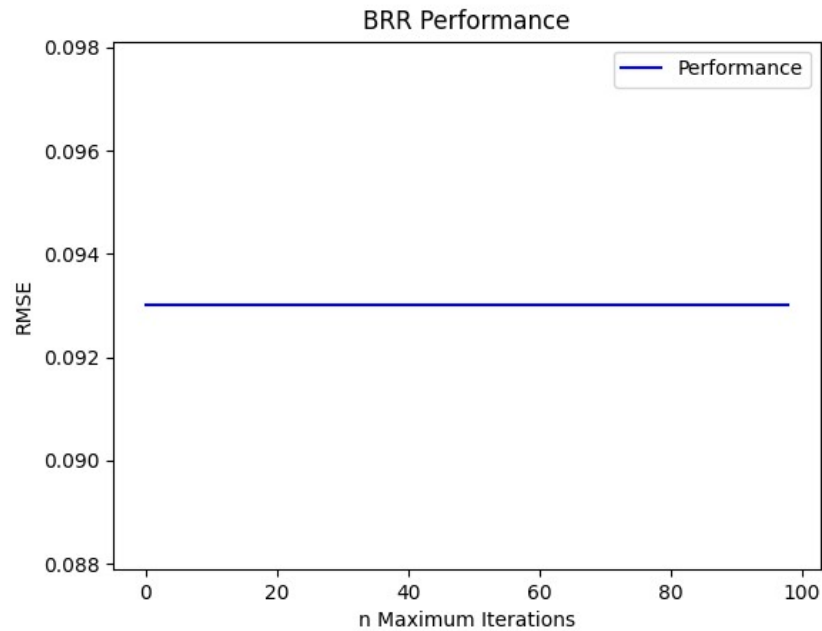
Algorithms and Results

Algorithms and Results

Further Thoughts & Conclusion

Limitations of data crawling

Hard to get much data, let alone good quality data.

Would need a 5000% increase of requests just to add 1 more feature.

Model Blindspots

The end model does not fully take into account trends and momentum.

It also does not take into account the actual selling price (just the asking price).

Source Code

You can find the code used at:

https://github.com/DimitrisPatiniotis/house_valuation