EstateHunter

Team

Vasil Todorov, FN 25459

Problem Description

The estate market in Sofia is going crazy and there are many overpriced properties. It would be good to have some kind of automatic filtering for good bargains.

Methodology/Algorithm

I am going to train a NN over most of the estates on the market which are sold online (imoti.bg etc.). I will experiment with different architectures / normalizations and other methods (probably try something “cutting edge” from a new paper if there is time, but can’t promise) and try out different ways to grade the accuracy. I might run some classic ML algorithm like SVM to compare the results. I can also try training the NN with a well known set from Kaggle and then adjust the NN by training it with the data from the properties in Sofia. I might also try some feature extraction from the descriptions like is it close to some Metro station, does it have a view to Vitosha etc.

Once I got a well designed NN, I will evaluate all the estates on the market by dividing the dataset to 10 batches and training 10 times the NN each time with 1 batch that would be evaluated and 9 batches for training.

Related Work

The house prediction problem is well known, there are some existing datasets like the *Boston* one and this one in Kaggle - https://www.kaggle.com/c/house-prices-advanced-regression-techniques

I will check the the best solutions and read the information and experience the authors provide.

Evaluation Plan

I will try different ways to evaluate the accuracy. One possible is just taking the net difference in % of the real value. I will try to get accuracy at least 93-95%, otherwise it wouldn’t be that useful.

Minimal Requirements

The dataset will be at least 10000 examples and will be collected with a web scraper.

There is prior work done to the problem and a human who keeps an eye on the market can evaluate the result fairly good.

The result can be evaluated automatically on a dataset and comparing the predicted price with the real one.