Machine Learning Engineer Nanodegree – capstone proposal

Predicting apartment prices in Bucharest

Adrian Ciubotaru
December 6th 2021

Domain Background

This project is about the world's largest asset class, estimated to be worth $277 trillion at the end of 2020. I am talking about real estate. Naturally, such a vast and lucrative domain has attracted problem solvers from within the machine learning community, who attempted to answer one of the two main questions for the industry: what's the right value for the price and what's the right value for the rent?

Problem Statement

It is difficult to figure out how to appropriately price an apartment in Bucharest. The largest website only offers the range for a neighborhood and consulting an appraiser is both expensive and time consuming.

Datasets and Inputs

Around 2500 listings scraped from www.imobiliare.ro. All apartments are from Bucharest, from a subset of all existing neighborhoods. The listings include information such as rooms, floor, building age, as well as sections of free text, where other amenities are mentioned. The dataset needs to be cleaned and go through feature selection.

Solution Statement

I will create a regression model that, based on inputted characteristics of the apartment, will return a reasonable price.

Benchmark Model

Currently the website offers a range for all properties within a neighborhood. The range goes from the minimum listing price of any property in the area all the way to the max listing price of any property in the area. A reasonable benchmark model would be *(Min Price + Max Price)/2*.

Evaluation Metrics

For this exercise, I will use RMSE (root-mean-square error) as an evaluation metric.

Project Design

- Clean data & create features
- Use AutoML to determine the best algorithm
- Expose final model endpoint
- Create a simple web page with a form where any user can input apartment features in order to receive a price prediction