Housing prices

Iselin Weltzin, Martin Tunge Sterri, Markus Ludvig Gustavsen / Gruppe 17,
18.11.2022

DESCRIBE THE PROBLEM

SCOPE

One way for real estate investors to earn profits today is by finding undervalued property that they then can flip (Beattie 2022). The project's goal is to create and deploy a machine learning algorithm that is able to predict the cost of purchasing a house, given certain parameters. This is a service many real estate investors do manually today, when considering options for future transactions. Given the success of our project, it is possible to write a program which will receive input from local listings and compare it to the predicted house value, giving the user crucial information regarding the potential value of the asset.

The easiest business metric to measure by is in this case the difference between listed value and value predicted by the model, either by percentage or actual profits in a given currency. When comparing the predicted value with the listed asking price, one may notice the ones with the largest difference, suggesting a tentative value ready to be extracted. In this specific project the stakeholders are the ones developing, fueling the need for perfection by one’s own self interest. After the reachable milestone of compiling a group and framing the project, the actual development and testing of the model is the most time consuming, although our prediction of several months of working was somewhat exaggerated.

The first version of the model will just use a small number of metrics, and can be used by anyone. This is ment as a proof of concept. It is also a way of testing and interacting with the potential model, before doing something more advanced. The possibility to get some of these metrics is not great, and testing some of them to see if these can be used is information that can be taken on to further development.

Depending on where we want to take the model, using things like deep learning for image evaluation of the house and so on, it has the potential to take a lot of computer power.

METRICS

As a privately owned project, the minimum requirements for the lavished stakeholders is to turn a profit, preferably a percentage surpassing the cost of inflation, so that the investment is seen as a success to all observers. The most prominent and decisive feature is the accuracy of our prediction, where a small correction in value can be critical to profitability of the model.

One way of measuring whether or not the model can be used to find good matches, would be to do real world tests, where we can see if someone can make more profit when using

the model. This would be time consuming and difficult to do at a scale where smaller benefits would be measurable.

An alternative would be to test it on historical data. This is brings their own problems

DATA

The data used in training and testing the model is the Ames Housing dataset, a more modern and expanded version of the Boston Housing dataset. The accuracy and consistency of the data used is difficult to fully ascertain, but given the size of the dataset, and the amount of features, gives us a lot of room to assess and find the best predictors for a house price.

For further development, it is crucial to look at how we can access real life data to be used for these predictions. It could be a good idea to figure out where we should use this model. A greater region could give us more opportunities to discover undervalued property, but it could also lead us to having less information. One realistic way of accessing information could be web scraping. Imagine that we have two regions where the primary webpage for selling houses is different. These webpages could have different information. The model we would create for one webpage, could use more information, than if we had to make a general model that could be used on both.

For finding the target, we can use two tactics. One is to look at the predicted price. With the help of this, you could see if you could find some outliers. In some cases it would also be possible to find the sale prices. This information would be extra valuable. It is available in at least certain regions like in the UK where you can find sales prices back to May 2000 (rightmove, n.d.).

MODELING

The machine learning models we would like to explore are regression models. These models are good for finding a numeric value. They are good at predicting structured data (rows and columns). We expect that in real life, it would be easy to find specific values so that this type of prediction would fitt. We can note that after having looked at the data, one of the best predictors was stuff like “Overall quality”. This would probably not be easy to find in a listing, but this could indicate that some form of image processing could be useful for evaluating the price.

The model does not necessarily need to be super accurate, as long as it can be used for flagging houses that are undervalued. This could mean that the model should be fitted so that it is better to predict a higher price, then one that is too low. The model then becomes more like an upper bound. When the listed price is “much” lower than the predicted price it could be flagged so that a human can look at it.

DEPLOYMENT

Our model will be available online. A user should be able to fill in all the input fields and then get an estimated price. We used *heroku* to do this.

In practical use, it would make more sense that the model can process a lot of houses, so that we can find the outliers. We can get the data with the help of web scraping so you wouldn't have to fill it out yourself. In that model, there wouldn't be a cap on how many attributes we can use. There would most likely not be too many new houses listed, and it is not critical to find the value in seconds, so the time that the model uses would not be a big concern.

In our model, there are more concerns connected to time and the number of attributes. The dataset that we used had 79 attributes excluding the id and the price. This is too many for it to be practical for users, so we limited it to 10.

One thing to think about when it comes to house prices, it is that they change over time. Old data can become useless quite quickly, there therefore has to be a way of getting new data to train on quite quickly. It might be relevant to create a form of a predictive model that tries to predict where the price is going based on some “real time” indicators.

REFERENCES

Beattie, A. (2022, July 14). *5 Simple Ways to Invest in Real Estate*. Investopedia. Retrieved November, 2022, from https://www.investopedia.com/investing/simple-ways-invest-real-estate/

rightmove. (n.d.). *Sold House Prices - Check UK house prices paid on Rightmove*. Rightmove. Retrieved November, 2022, from https://www.rightmove.co.uk/house-prices.html