

Real Estate Market Prediction

Team :

Tuhin Sarkar | Vatsal Rathod | Victor Okoro | Pierre-Louis Ehret | Aayusi Biswas

Overview

The Real Estate market has been an important section of the society ever since civilizations came into existence. Analysis of Real Estate Market has been an integral part of the decision making procedure. Our application works on Machine Learning Algorithms to predict the rates of a property upto 2 years in future. Currently, it works on data specific to Brooklyn.

Problem

Real Estate Consultants have been ripping off money in the name of 'expert predictions' since a long time. The lack of knowledge on the buyer's part has been fueling their confidence. Another issue is that builders and architects often end up investing in the wrong properties when left on their own. This can eventually add up to the loan losses a country has to bear.

Solution

We came up with a mobile application that can predict the price of a specified property based on the Neighbourhood, the building class, build year and area of a particular plot. Our intended interface is very simple. A user has to download our app, fill in the required fields and hit the graph button. The result will be the predictions of future prices of the pointed property as a graph.

Procedure

After web scraping and acquiring enough data about the Brooklyn Real Estate Market, we began by cleaning the data. We removed all the rows and columns with a NULL value. Then we tried to condense the data by learning about historic events like recession of 2008 and price hike of 2016-17. This helped us condense our training data to be from 2010 to 2015 and testing data of 2016-17. We were left with 74k of data rows and 26 columns as opposed to 400k data rows and 109 column that we began with. We ran a correlation matrix to find out the features that affect the sale price the most with respect to time.

We label encoded all the strings/object type features. We normalized the sale price so as to reduce the skewness of the data. This was the end of data processing.

For the training part, we made chunks of the entire training data so as to fasten the training process and increase accuracy. We then formed functions for model training and model

evaluation for different kind of training methods like Ridge, Lasso, gradient boost, xgb and Random Forest because our data was large and we needed to perform feature selection. The best testing accuracy on the basis of closeness to the actual price was found with the XGBoost model and was 66% percent for 2016 and 76% percent for 2017. The training accuracy was 99% and validation accuracy 80%.

Experience

We started by looking for data for Delhi, India and failed miserably. That's when we realised that the Indian government sites aren't updated regularly.

With that, we shifted to Melbourne which had little historical data records to offer. Finally, we hit a jackpot with Brooklyn.

Understanding data correlation, selecting features according to relevance etc were what we learned while data cleaning and processing.

And then finally, we learnt that all prediction models have specific features and structures. It's all about what data you have and how you shape your code.

Proposed UI



Resources Used:

Ng, Aron. "Machine Learning for a London Housing Price Prediction Mobile Application."

Imperial College London.

"The Home of Data Science & Machine Learning." *Countries of the World* | Kaggle,

www.kaggle.com/.

"Welcome to NYC.gov | City of New York." *Gowanus Canal – History*, www.nyc.gov/.