House Price Prediction using Machine Learning

-Sahil Jethva

Date: 12 – 05 – 2022

Abstract

House price forecasting is an important topic in real estate. The literature attempts to derive useful knowledge from historical data on property markets. Machine learning techniques are applied to analyze historical property transactions in India to discover useful models for house buyers and sellers. Revealed is the high discrepancy between house prices in the most expensive and most affordable suburbs in the city of Mumbai. Moreover, experiments demonstrate that the Multiple Linear Regression that is based on mean squared error measurement is a competitive approach.

Problem Statement

- In India, there are multiple real estate classified websites where properties are listed for sell/buy/rent purposes such as 99acres, housing, Common Floor, magic bricks, and more. However, on each of these websites, we can see a lot of inconsistencies in terms of pricing of an apartment and there are some cases when similar apartments are priced differently thus there is a lot of in-transparency.
- Sometimes, consumers may feel the pricing is not justified for a particular listed apartment, but there is no way to confirm that either. Proper and justified prices of properties can bring a lot of transparency and trust back to the real estate industry, which is very important as for most consumers, especially in India the transaction prices are quite high, and addressing this issue will help both the customers and the real estate industry in the long run.
- We propose to use machine learning and artificial intelligence techniques to develop an algorithm that can predict housing prices based on certain input features.

Market/Customer/Business Need Assessment

- We need a proper prediction on the real estate and the houses in the housing market we can see a mechanism that runs throughout the properties for buying and selling.
- Buying a house will be a lifetime goal for most individuals but there are a lot of people making huge mistakes buying the properties most of the people are buying properties unseen from the people they don't know by seeing the advertisements and all over the grooves coming around.
- In the housing market 2017, there is a survey that in the year 2016 the house sold in America were about 5.42 million but the starter home inventory was down up to 10.7% from 2015.
- One of the common mistakes is buying properties that are too expensive but it's not worth it. To maintain the transparency among customers and also the comparison can be made easily through this model. If the customer finds the price of the house at some given website higher than the price predicted by the model, so he can reject that house.

Target Specifications and Characterization

- Before committing to the project several groundworks should be done so there is a need for a literature review. I have analyzed many papers regarding price prediction related to the house markets and other different sectors. I have used the recent and latest technologies; our main goal is to get more accuracy than the previous works.
- The business application of this algorithm is that classified websites can directly use this algorithm to predict prices of new properties that are going to be listed by taking some input variables and predicting the correct and justified price i.e., avoid taking price inputs from customers and thus not letting any error creeping in the system.

External Search

- I have used the real estate price prediction Kaggle dataset for this project. Dataset can be found here: https://www.kaggle.com/datasets/quantbruce/real-estate-price-prediction?select=Real+estate.csv
- The set consists of 400 entries which contain entries about the data, by using the following dataset I have done an analysis using machine learning algorithms.

Sample dataset:

	No	X1 transaction date	X2 house age	X3 distance to the nearest MRT station	X4 number of convenience stores	X5 latitude	X6 longitude	Y house price of unit area
0	1	2012.917	32.0	84.87882	10	24.98298	121.54024	37.9
1	2	2012.917	19.5	306.59470	9	24.98034	121.53951	42.2
2	3	2013.583	13.3	561.98450	5	24.98746	121.54391	47.3
3	4	2013.500	13.3	561.98450	5	24.98746	121.54391	54.8
4	5	2012.833	5.0	390.56840	5	24.97937	121.54245	43.1
5	6	2012.667	7.1	2175.03000	3	24.96305	121.51254	32.1
6	7	2012.667	34.5	623.47310	7	24.97933	121.53642	40.3
7	8	2013.417	20.3	287.60250	6	24.98042	121.54228	46.7
8	9	2013.500	31.7	5512.03800	1	24.95095	121.48458	18.8
9	10	2013.417	17.9	1783.18000	3	24.96731	121.51486	22.1

Applicable Constraints

- Continuous data collection and maintenance.
- Product usage can be found not so reassuring as models have low accuracy, Optimization is required in this case.
- Marketing the product to a Training centre or a Software development company.
- The product developed should be responsive such that customers who are not educated will be able to easily use the product.

Applicable Regulations

- The above-taken Dataset is obtained for the purpose of the experiment, it cannot be used in a commercial setting. Thus, Acquiring the data to train the model is a must, for that Data protection and privacy regulations.
- Government Regulations for small businesses

Business Opportunity

House Price prediction, is important to drive Real Estate efficiency. As earlier, House prices were determined by calculating the acquiring and selling price in a locality. Therefore, the House Price prediction model is essential in filling the information gap and improving Real Estate efficiency.

Concept Generation

Based on the housing rate and area, one can predict the housing value and value of the property. You can use a Machine learning Algorithm in the Recommendation System which can help buyers to find details and unique insights on properties they're interested in.

Final Report Prototype

The product takes the following functions to perfect and provide a good result:

Back End
Model or Webapp Development: This must be done before releasing the service. A lot of manual supervised machine learning must be performed to optimize the automated tasks.
- Performing EDA to realize the dependent and independent features.
- Algorithm training and optimization must be done to minimize overfitting of the model and hyperparameter tuning.

Front End
- Different user interfaces: The user must be given many options to choose from in terms of parameters. This can only be optimized after a lot of testing and analysis of all the edge cases.
- Feedback system: A valuable feedback system must be developed to understand the customer's needs that have not been met. This will help us train the models constantly.

Code Implementation:

GitHub link for code: https://github.com/sahil-2505/Sahil-Jethva-Feynn_Labs_project-1

Conclusion

The main goal of this project is to determine the prediction for prices which we have successfully done using different machine learning algorithms like KNN regressor, OLS regression. Every system and new software technology can help in the future to predict the prices. price prediction can be improved by adding many attributes like surroundings, marketplaces, and many other related variables to the houses. The predicted data can be stored in the databases and an app can be created for the people so they would have a brief idea and they would invest the money in a safer way.