Applied Research on House Prediction Using Diverse Machine Learning Techniques

Maharshi Modi, Ayush Sharma, Dr P Madhavan

Department of Computer Science, SRM Institute of Science and Technology, Kattankulathur, Chennai

Abstract

With the booming civilization and ever-changing market requirements, it is essential to know the market drifts. Today prediction of house prices according to the trends is the principal essence of the study. It is imperative for an individual to understand the business trends so that he can prepare his budgetary needs according to his requirements. Real Estate is an ever-growing enterprise with an expanding society. For an investor, it is essential to comprehend the business drifts which can assist him to underwrite in the right way and augment his business throughput. Sometimes clients get dupe by the hoax market rate set up the agent due to which the real estate industry is less translucent these days. With an uptick in convince of the dataset, it's viable for a researcher to develop a model with high accuracy. The foremost intention of this design is to develop a comprehensive design that is advantageous for a business society as well as an individual which is the main nub of this investigation. This design is intended to assist a client by diminishing his fieldwork moreover extricate his time and money. Models are enlightened in diverse machine learning algorithm and they are coupled by implementing the stacking technique.

Introduction

The market is evolving day by day, today a lot of software giants are shifting towards Artificial intelligence for better decision-making and resolving some complicated difficulties in real-world using the data accessible in ample quantity. Today machine learning is applied in various sectors hence integrating it with the Real Estate sector can help an investor to intensify his business throughput also support an individual to shape his budgetary needs according to his requirements. For making it convenient to the majority of the people a user-friendly Graphical User Interface (GUI) is deployed. Through this design market trends can be analyzed and it can be a useful asset for an investor to bankroll in correct direction. Many Real Estate companies are facing breakdown in the market due to lack of such analysis and this could also lead to disruption in countries economy as Real Estate sector contributes significantly.
Through improvised analytics and higher data with accurate prediction this project aims to develop a complete model which is useful for an individual as well as an organization and solve real time problem which could save time and money.

Previous System

Due to lack of research and data available previous system didn’t have such technology and approach to predict the house prices with more enhanced efficiency. Traditionally, clients had to do the standard fieldwork, in-order to check the house price and it was considerably inconvenient for them to

investigate diverse areas which could match their stipulation. It would restrict one's need to traverse more areas due to lack of infrastructure additionally it would cost them a lot of time and money.

Even, as a business perspective, it was quite difficult for the shareholder to further analyze the market trends as a result, they end up losing money due to lack of business drifts and thus this would start to threaten the resources of the company. The market stability of real estate was quite vulnerable.

Literature Survey

[1]"On the Relation between Local Amenities and House Price Dynamics" Eli Beracha, Ben T Gilbert, Tyler Kjorstad, Kaplan Womack (2016). This article has investigated the association between house price volatility, resell value and amenities, and proves that areas with high services experience greater price volatility. It additionally asserts that there is further room for development in major cities and principal regions. For regional house price foresight, they have practiced several regression algorithms such as random forest and gradient boosting.

[2]"Defining Street-based Local Area and measuring its effect on house price using a hedonic price approach: The case study of Metropolitan London", Stephen Law (2017). In this paper, Stephen law has studied the well-built relationship among avenue-based local regions with house prices. And it reveals that using avenue-based local area is better than a regional-primary based nearby vicinity. Stephen law has made his judgment based on the dataset available from the united states housing branch.

[3] "Property Renovations and Their Impact on House Price Index Construction" Alexander N. Bogin, William M. Doerner Alexander and Williams studied the price improvements in wide-scale U.S. areas. They have considered various factors which include monetary boom, the distance for nearest public transport, close by airports and numerous such others. The consequences exhibit that the cost can be escalated by 15% in popular areas of major cities, while less distortionary effects outside of central areas and in smaller cities.

[4] ”Geographically weighted regression with a non-Euclidean distance metric: a case study using hedonic house price data", Binbin Lu, Martin Charlton, Paul Harris & A. Stewart Fotheringham(2014) , Binbin designed a geographically weighted regression design in this article to compose a study of residence prices in London, using the study's Euclidean distance (ED), road network length, and time metrics. This model aims to provide a faster and faster response to higher accuracy. The results show that people with higher incomes, however, sell houses cheaply, buy expensive homes.

[5] "Explaining house price dynamics: Isolating the role of non-fundamentals", David C. Ling, Joseph T.L. Ooi and Thao T.T. Le (2015). David and Joeseph discovered that the views and beliefs of buyers, builders and lenders are associated with the value of the house over the following two quarters at some point of booms and busts. People with a religious belief generally tend to buy homes inside their community or close to it. They have practiced numerous machine learning algorithms to accomplish their

undertakings such as regression, random forest, and support vector machine.

Proposed System

(1) : Data Analysis And Validation - In this section, we will load the dataset, check for data duplication and redundancy after which clean and trim the dataset to reduce data inconsistency and enhance the analysis. The dataset might incorporate missing values and null values which can cause inconsistency. To achieve better results data needs to thoroughly validated and pre-processed. Supervised machine learning classification algorithms will be used to extract patterns and give the dataset. Data validation is a vital undertaking to be executed to attain results with better accuracy. (2) : Data Collection - The dataset acquired for predicting the house prices is classified into two divisions: Training set and Test set. The data model that was built using various machine learning techniques is applied to the training set. The Test set is used to check the accuracy of the model. (3) : Pre-processing - The data obtained might include missing values that may result in inconsistency. Records need to be pre-processed to strengthen the effectiveness of the set of rules to accomplish better performance. It is important to exclude the outliers and also to achieve the variable conversion. Based on the correlation amongst attributes it was observed that attributes that are significant individually include property area, sqr-feet, a distance which is the most powerful among all. Some variables such as applicant income and co-applicant earnings are not notable alone. Data unification, data compression, and data transformation are also to regard real estate records. For elementary analysis, the data is diminished to a few minimum quantities of records. Initially, the Attributes which are significant to forming a real estate credibility prediction are recognized with information gain as the attribute-evaluator and Ranker as the search-method.

(4) : Construction and Training the model - The huge dataset will then be divided into different sub-datasets. These instances will then be trained separately for each training dataset to form a classifier. Such datasets are trained using knowledge of the ensemble. Ensemble learning uses various techniques of machine learning to achieve performance variability and produce accurate results. Test data can be processed through each classifier and check for the results they provide. In this model, a voting classifier will be used which used multiple models to make predictions for each data point. The foresight by the various model is recognized as a “vote”. The foresight which we get from the majority of models is chosen as the final prediction.

Conclusion

The analysis of the survey on the House Price Prediction based on the Ensemble approach led to various conclusions, a proper system that has better accuracy and better decision-making approach models are needed in the market which can be useful for business as well as an individual day to day use. Another inference drawn is developing countries that are facing real estate disruption due to lack of such technology can be a valuable asset to intensify their economy as the real estate sector appends a significant amount to a countries economy.

References

[1] **Eli Beracha, Ben T Gilbert, Tyler Kjorstad, Kaplan Womack,** "On the Relation between Local Amenities and House Price Dynamics", Journal of Real estate Economics, Aug. 2016.

[2] **Stephen Law**, "Defining Street-based Local Area and measuring its effect on house price using a hedonic price approach: The case study of Metropolitan

[3] **Alexander N. Bogin, William M. Doerner**, "Property Renovations and Their Impact on House Price Index Construction",https://www.fhfa.gov/PolicyProgramsResearch/Research/PaperDocuments/wp1702.pdf

[4] **Binbin Lu, Martin Charlton, Paul Harris & A. Stewart Fotheringham,** "Geographically weighted regression with a non-Euclidean distance metric: a case study using hedonic house price data", International Journal of Geographical Information Science, pp. 660-681, Jan 2014.

[5] **David C. Ling, Joseph T.L. Ooi and Thao T.T. Le**, "Explaining house price dynamics: Isolating the role of non-fundamentals", Journal of Money, Credit and Banking, vol. 47, Issue S1, pp. 87-125, March/April 2015.

[6] **Joep Steegmans, Wolter Hassink**, "Financial position and house price determination: An empirical study of income and wealth effects", Journal of Housing Economics, vol. 36, pp. 8-24, June 2017.

[7] **Sean Holly, M. Hashem Pesarana, Takashi Yamagata**, "A Spatio-temporal model of house prices in the USA", Journal of Econometrics, vol. 158, Issue 1, pp. 160–173, Sep. 2010.

[8] **Marco Helbich, Wolfgang Brunauer, Eric Vaz, Peter Nijkamp,** "Spatial Heterogeneity in Hedonic House Price Models: The Case of Austria", Urban Studies, vol. 51, Issue 2, Feb. 2014

[9] **Ayush Varma ; Abhijit Sarma ; Sagar Doshi ; Rohini Nair** "House Price Prediction Using Machine Learning and Neural Networks", 2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT) ,10.1109/ICICCT.2018.8473231 , 18116205 , April. 2018

[10] **Rushab Sawant, Saurabh Jain , Tushar Tiwari, Yashwant Jangid, Ms.Ankita Gupta** "A Multi Feature Based Housing Price Prediction for Indian Market Using Machine Learning" , International Journal of Computer & Mathematical Sciences IJCMS ISSN 2347 – 8527 Volume 6, Issue 12 December 2017