

Analytical Study of House Price Prediction Models

Gauri Mhatre, Guide: Asst. Prof. Gauri Ansurkar

Keraleeya Samajam's Model College, Dombivli East, Mumbai, Maharashtra, India

gaurimhatre2001@gmail.com

Abstract:-

House Price Prediction is commonly used to evaluate the changes in price of houses. Since housing price is strongly corresponding to other factors such as location, area, population, it requires other information apart from HPP to predict individual housing price. Researches has stated that changes in house prices are due to concern of for house owners and the real estate market. Large numbers of paper adopting traditional machine learning approaches to predict housing prices accurately has been consider, but they are rarely concerned about the performance of individual models and neglect the less popular but complex models. A survey of literature is carried out to analyze the relevant attributes, factors and the most efficient models to predict the house prices. Machine learning has played major role in image detection, spam reorganization, normal speech command, product recommendation and medical diagnosis.

Keywords: -

House Price Prediction, Real estate market, Traditional machine learning.

Introduction: -

House is most important aspect of human life. Development of civilization is the beginning of the increase in demand for houses day by day. Predicting and analyzing the real estate market can evaluate the stability of the real estate market and can facilitate the government to carry out control on house prices and maintain the healthy

and stable development. Many people invest in real

estate, because sometimes, it can bring a lot of capital income. Investment in real estate is also a reflection of a local real estate development situation to a certain extent. The purpose of house price prediction and analyses is to provide basis for pricing between buyers and sellers. Different groups of buyers may focus on different house attributes so if a house's features and attributes identified through an attention mechanism match the needs of a buyer, then the house can be recommended to the buyer. Many researchers have already worked to understand and break the mysteries of the prediction of house prices.

Literature Review: -

1. Prediction of House Price Using Machine Learning

Authors: Mr. Piyush Chordia, Mr. Pratik Konde, Ms. Supriya Jadhav, Hrutik Pandhare, Prof. Shikha Pachouly

This paper studies the SVM algorithm in machine learning for house price prediction. It takes data from the user and analyze the pre-available data and uses various algorithm and classifies data and forecast the accurate price of the property. Then it confirms the accurate prediction and result also depends on the population and the quality of the training dataset. Results obtained before through SVM vs optimized SVM were then be evaluated and from

comparative analysis done in next section, SVM shows comparable value over the cryptocurrencies for this period. SVM algorithm is not acceptable for large data sets and does not execute very well when the data set has more data specimens. But when you have a data set where number of features is more than the number of rows of data, SVM can perform in that case as well and SVM can work well with image data as well. The model will be enhanced on the accuracy rate of the predicted price in the future.

2. House Price Prediction Using Machine Learning

Authors: G. Naga Satish, Ch. V. Raghavendran, M.D.Sugnana Rao, Ch.Srinivasulu

In this paper, the Decision tree machine learning algorithm is used to construct a prediction model to forecast the potential selling prices for any real estate property. Additional features like air quality and crime rate were included in dataset to help predict the prices accurately. These features are not mostly included in the datasets of other prediction systems, which makes this system different and influence people's decision while purchasing a property. The disadvantages of decision tree algorithm are that small change in the data can cause a large change in the structure of the decision tree causing instability and tree sometimes calculation can go far more complex compared to other algorithms and often involves higher time to train the model. Decision tree training is relatively expensive as the complexity and time has taken are more and inadequate for applying regression and predicting continuous values. But compared to other algorithms decision trees requires less effort for data preparation during pre-processing and does not require normalization of data, scaling of data as well. Missing values in the data also do not affect the process of building a decision tree to any

considerable extent and is very intuitive and easy to explain to technical teams as well as stakeholders. The trained model is integrated with the User Interface using the Flask Framework and system provides 89% accuracy while predicting the prices for the real estate prices.

3. Housing Price Prediction via Improved Machine Learning Techniques

Authors: - Quang Truong, Minh Nguyen, Hy Dang

In this paper, different models were investigating for house price prediction. Three different types of Machine Learning methods including Random Forest, eXtreme Gradient Boosting (XGBoost), and Light Gradient Boosted Machine (LightGBM) and two techniques in machine learning including Hybrid Regression and Stacked Generalization Regression are compared and analyzed for best solutions. Even though all of those methods achieved desirable results, different models have their own advantage and disadvantage. The Random Forest method has the lowest error on the training set but is liable to be overfitting. Its time complexity is high since the dataset has to be fit multiple times in random forest method. Random Forest method's disadvantage is that it requires much computational power as well as resources as it builds numerous trees to combine their outputs and it also requires much time for training as it combines a lot of decision trees to determine the class and due to the ensemble of decision trees, it also suffers interpretability and fails to determine the significance of each variable. It has advantages that it creates as many trees on the subset of the data and combines the output of all the trees and reduces the overfitting problem in decision trees and also reduces the differences and therefore improves the accuracy. Random Forest works well with both categorical and continuous variables and can be used to solve

both classification as well as regression problems and Random Forest can automatically handle missing values.

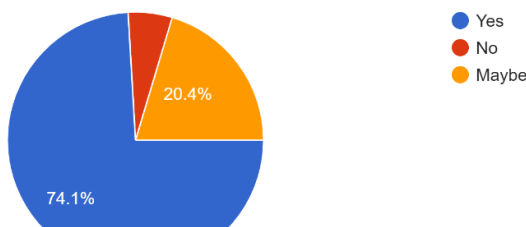
The XGBoost and LightGBM are decent methods when comparing accuracy, but their time complications are the best, especially LightGBM. Indeed though LightGBM and XGBoost are both asymmetric trees, LightGBM grows leaf-wise (horizontally) while XGBoost grows level-wise (vertically). LightGBM grows the tree widely, resulting in smaller and faster models compared to XGBoost. The Hybrid Regression method is simple but performs a lot better than the three former methods due to the generalization. Finally, the Stacked Generalization Regression method has a complicated architecture, but it is the best choice when perfection is the top priority. Indeed, though Hybrid Regression and Stacked Generalization Regression deliver satisfactory results, time complexity must be taken into consideration since both of them contain Random Forest, a high time complexity model.

Questionnaire: -

- 1. House prices are expected to rise further?
- 2. Do you think foreign investors are important in global housing market?
- 3. Does housing market have impact on economic and financial stability?
- 4. What is your level of knowledge in house price market?
- 5. Is it worth investing in real estate market?
- 6. Rate the factors on your level which affect the housing prices?
- 7. Do you think house price analysis is important?

Results: -

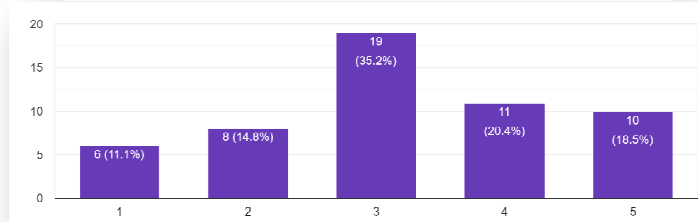
1.



When people were asked if they think that prices are expected to rise further, 74.1% people think rise in prices of houses is expected and 20.4% thinks maybe and 5.6% thinks no.

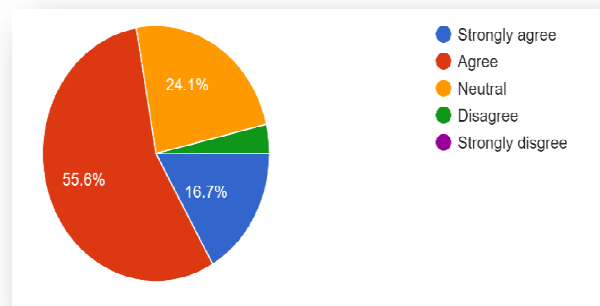
2.

When asked about foreign investors are important in global housing market, 35.2% people have neutral



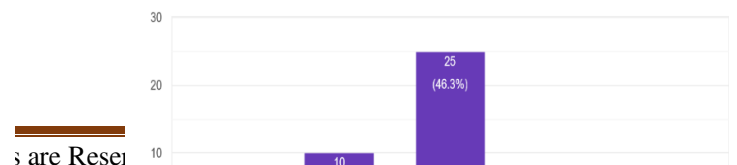
views and 11.1% people strongly disagree with the foreign investors being important in global estate market and 18.5% people are strongly agree.

3.



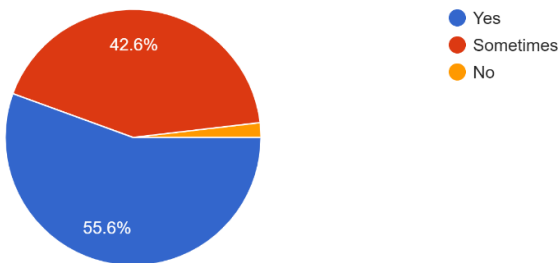
When asked about does housing market have impacted on economic and financial stability, 16.7% people strongly agrees and 55.6% people agree with the question and 24.1% people have neutral thinking and 3.7% people disagree.

4.



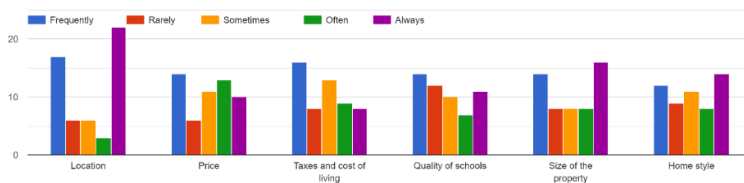
When asked about their knowledge in house price market 13% people are very familiar about house price market, 9.3% people are familiar and 46.3% have neutral knowledge and 18.5% are less familiar with house price market and 13% people are not familiar.

5.



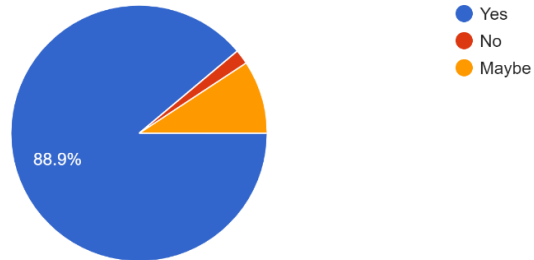
When asked about worth investing in real estate market 55.6% people thinks real estate market investment is worth and 42.6% people think sometimes it is worth investment and 1.9% people think no.

6.



When asked to rate the factors which affects the housing prices like location, price, taxes and cost of living, quality of schools, size of the property and home style.

7.



When asked about what they think house price analysis in important and 88.9% people think house price analysis should be done and 9.3% people think maybe analysis should be done and 1.9% people think no.

Descriptive Statistics:-

Descriptive statistics is a summary statistic that quantitatively describes or summarizes features from a collection of information. Here are some results about finding.

What is your level of knowledge in house price market?

References: -

1. <https://www.ijraset.com/research-paper/prediction-of-house-price-using-mi>
2. <https://deliverypdf.ssrn.com/delivery.php?ID=124083127024114093018113091024081087057013006006039043089019066023003124074085082068056000031041111006030011003096101030065081062074008052087000006083072103119016004025051013094120017021026010103119086007003080105103064069082096113121014080101064116030&EXT=pdf&INDEX=TRUE>
3. <https://reader.elsevier.com/reader/sd/pii/S1877050920316318?token=8B01E89FB8CA7B677060B14D7AC955C7B02436FEC523E95F542F8F50EB519A89106B636DEAD48B7831EB6A8E349D0FC2&originRegion=eu-west1&originCreation=20221103133342>

Conclusion: -

The accurate price prediction model would examine and analyzed and allows the investors or house buyers to determine the realistic price of a house as well as the house developers to decide the affordable house price. This paper addressed the attributes used by previous researchers to forecast a house price using various prediction models. Taken together, the results of the survey haveshown the potential of SVR the model that will be enhanced on the accuracy rate of the forecasted price. In second model, the Decision tree machine learning algorithm is used to construct a prediction model to predict potential selling prices for any real estate property and e system provides 89% accuracy. In third model, three different types of machine learning methods including Random Forest, LightGBM and XGBoost and two techniques in machine learning including Hybrid Regression and Stack Generalization Regression. Forest method has the lowest error on the training set time complexity is high. XGBoost and LightGBM are decent methods but their time complexities are the best, especially LightGBM. Hybrid Regression method is simple but performs a lot better than the three previous methods due to the generalization. Stacked Generalization Regression method has a complicated architecture, but it is the best choice when accuracy is the top priority. These models were developed based on several inputattributes and they work significantly positive with house price.

Mean	2.907407407
Standard Error	0.156990651
Median	3
Mode	3
Standard Deviation	1.153640972
Sample Variance	1.330887491
Kurtosis	-0.30510125
Skewness	0.186783893
Range	4
Minimum	1
Maximum	5
Sum	157
Count	54
Average	17.87927763