ISSN: 2278-4632
Vol-11 Issue-01 2021

House Price Prediction Using Machine Learning Techniques

G.L.Sravanthi[1] B.Hemalatha[2] , Syed.Yasmin3 A.Bhavana[4] , J.Pooja Reddy[5], Department of Computer Science and Engineering, Vignan's Nirula Institute of Technology and Science for Women.
Pedapalakaluru, AP, India. Mail id:glsravanthi88@gmail.com

Abstract

House prices increases every year. So there is a need for a system to predict house prices in the future . Hence price prediction can help the developer determine the selling price of a house and can help the customer to arrange the right time to purchase a house. House price index represents the summarized price changes of residential housing. While for a single-family house price prediction, it needs more accurate method based on location, house type, size, build year, local amenities, and some other factors which could affect house demand and supply. We use random forest regression model to predict individual house price. More over on other hand housing value indices, the advancement of a housing cost prediction that tend to the advancement of real estate policies schemes. We utilize machine learning algorithms as a research method that develops housing price prediction models. We create a housing cost prediction model in view of machine learning algorithm models for example, random forest regression on look at their order precision execution. The housing cost prediction model is used to support a house vender or a real estate agent for better information based on the valuation of house. A recent report indicates that house sellers and buyers are increasingly turning to online research in order to estimate house price before contacting real estate agents.

1.Introduction

Buying a house is a stressful thing. Buyers are generally not aware of factors that influence the house prices. The real estate agents are trusted with the communication between buyers and sellers as well as laying down a legal contract for the transfer. This just increases the cost of houses. The cost of the house depends on many factors like rooms in the house, area outside the house, location, and build year. According to the US Census Bureau, 560,000 houses were sold in the United States in 2016. In addition, 65% of all-American families owned houses in 2016. For the Americans who sold and bought these houses, a good housing price prediction would better prepare them for what to expect before they make one of the most important financial decisions in their lives.

A recent report from the Zillow Group, a popular housing database website, indicates that house sellers and buyers are increasingly turning to online research in order to estimate house price before contacting real estate agents. Researching how much the house you are interested in is worth on your own can be difficult for multiple reasons. One particular reason is that there many factors that influence the potential price of a house, making it more complicated for an individual to decide how much a house is worth on their own without external help.

This can lead to people making poorly informed decisions about whether to buy or sell their houses and which prices are reasonable. Because houses are long term investments, it is imperative that people make their decisions with the most accurate information possible. Therefore, housing websites such as Zillow, Trulia and Redfin 1, exist to provide estimations of housing valuations based on the houses' characteristics, at no cost. Using the datasets and machine learning algorithms as the prediction model, we compute score and use them as the baseline for predicting house price. In the dataset of 1,457 sold houses I collected, the ratio of overestimated houses to underestimated houses is 3 to 2. The question of this project is what the most important factors affecting housing prices are. In order to answer the questions listed above, this project uses various machine learning algorithms.

ISSN: 2278-4632
Vol-11 Issue-01 2021

2. Literature survey

Understanding recent trends in house prices and home ownership

This paper looks at a broad array of evidence concerning the recent boom in home prices and considers what this means for future home prices and the economy. It does not appear possible to explain the boom in terms of fundamentals such as rents or construction costs. A psychological theory, that represents the boom as taking place because of a feedback mechanism. or social epidemic that encourages a view of housing as an important investment opportunity, fits the evidence better. Three case studies of past booms are considered for comparison: the US housing boom of 1950, the US farmland boom of the 1970s, and the temporary interruption 2004-5 of the UK housing boom. The paper concludes that while it is possible that prices will continue to go up as is commonly expected, there is a high probability of steady and substantial real home price declines extending over years to come.[1]

Predicting House Prices with Spatial Dependence: A Comparison of Alternative Methods

This paper compares alternative methods for taking spatial dependence into account in house price prediction. We select hedonic methods that have been reported in the literature to perform relatively well in terms of ex-sample prediction accuracy. Because differences in performance may be due to differences in data, we compare the methods using a single data set. The estimation methods include simple OLS, a two-stage process incorporating nearest neighbours' residuals in the second stage, geostatistical, and trend surface models. These models take into account submarkets by adding dummy variables or by estimating separate equations for each submarket. Based on data for approximately 13,000 transactions from Louisville, Kentucky, we conclude that a geostatistical model with disaggregated submarket variables performs best.[2]

patial dependence, housing submarkets and house price prediction

This paper compares alternative methods of controlling for the spatial dependence of house prices in a mass appraisal context. Explicit modelling of the error structure is characterized as a relatively fluid approach to defining housing submarkets. This approach allows the relevant submarket to vary from house to house and for transactions involving other dwellings in each submarket to have varying impacts depending on distance. We conclude that - for our Auckland, New Zealand, data - the gains in accuracy from including submarket variables in an ordinary least squares specification are greater than any benefits from using geostatistical or lattice methods. This conclusion is of practical importance, as a hedonic model with submarket dummy variables is substantially easier to implement than spatial statistical methods.[3]

House price prediction: hedonic price model vs. artificial neural network.

The objective of this study is to empirically compare the predictive power of the hedonic model with an artificial neural network model on house price prediction. A sample of 200 houses in Christchurch, New Zealand is randomly selected from the Harcourt website. Factors including house size, house age, house type, number of bedrooms, number of bathrooms, number of garages, amenities around the house and geographical location are considered. Empirical results support the potential of artificial neural network on house price prediction, although previous studies have commented on its black box nature and achieved different conclusions.[4]

3. Proposed System:-

Simple linear regression is useful for finding relationship between two continuous variables. One is predictor or independent variable and other is response or dependent variable. It looks for statistical relationship but not deterministic relationship. Relationship between two variables is said to be deterministic if one variable can be accurately expressed by the other. For example, using temperature in degree Celsius it is possible to accurately predict Fahrenheit. Statistical relationship is not accurate in determining relationship between two variables. For example, relationship between height and weight.

ISSN: 2278-4632
Vol-11 Issue-01 2021

We have a dataset which contains information about relationship between 'number of hours studied' and 'marks obtained'. Many students have been observed and their hours of study and grade are recorded. This will be our training data. Goal is to design a model that can predict marks if given the number of hours studied. Using the training data, a regression line is obtained which */will give minimum error. This linear equation is then used for any new data. That is, if we give number of hours studied by a student as an input, our model should predict their mark with minimum error.

Y(pred) = b0 + b1*x

The values b0 and b1 must be chosen so that they minimize the error. If sum of squared error is taken as a metric to evaluate the model, then goal to obtain a line that best reduces the error.

$$\text{Error} = \sum_{i=1}^{n}(actual\_output - predicted\_output) ** 2$$

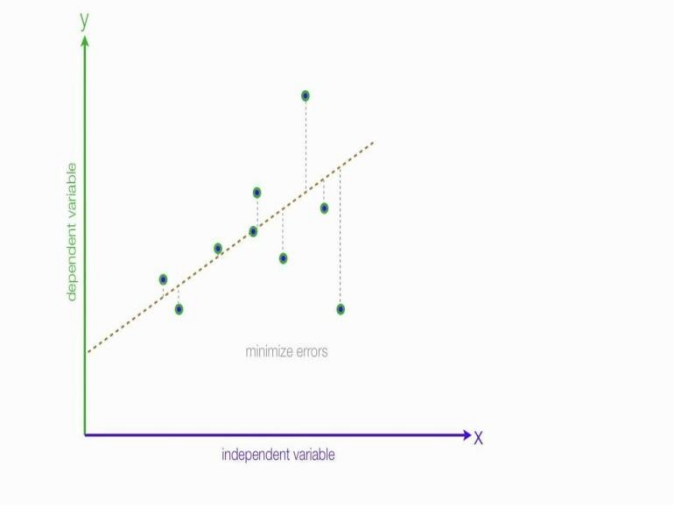
If we don't square the error, then positive and negative point will cancel out each other.For model with one predictor.

$$b_0 = \bar{y} - b_1\bar{x}$$

$$b_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

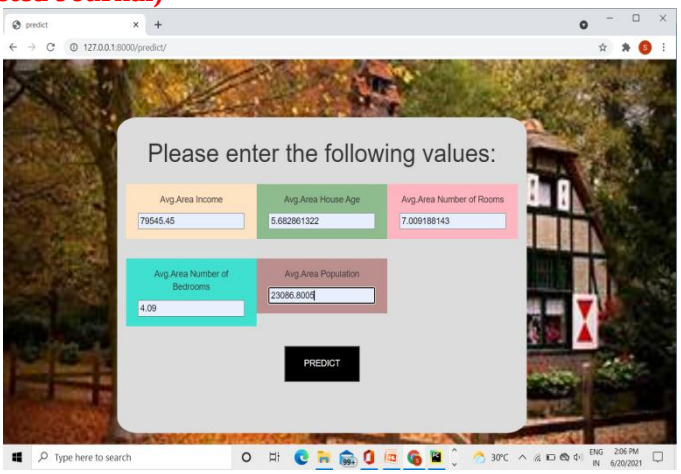
Apart from above equation co-efficient of the model can also be calculated from normal equation.

$$\text{Theta} = (X^T X)^{-1} X^T Y$$



Graph 1: Predicting price using Linear Regression.

ISSN: 2278-4632
Vol-11 Issue-01 2021



4. Results:-

After executing houseapp.py program, it results in a URL corresponding to local host. Then we need to copy the URL and paste it in any of the browser like Google, Firefox, etc and run it.The following will be the HTML page that results after running.

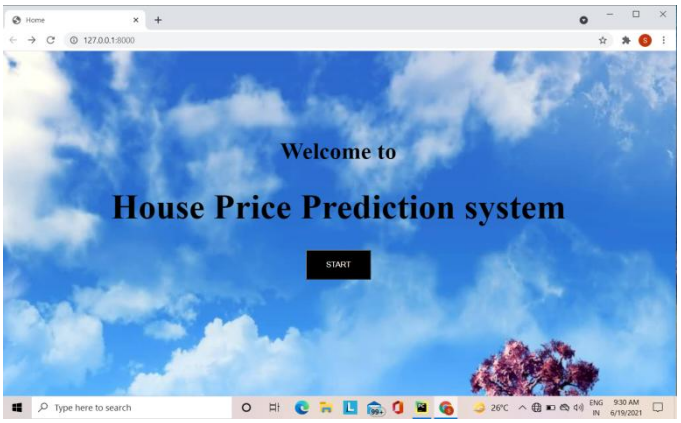


Fig:-Image representing user input

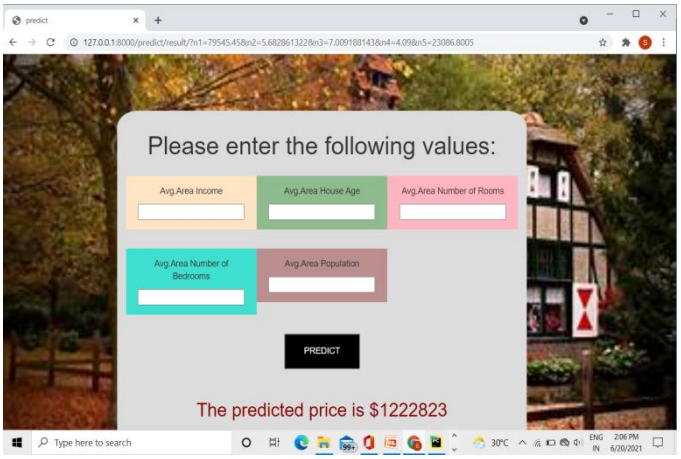


Fig:-Image representing the house price of the given input.

5. CONCLUSION:-

ISSN: 2278-4632
Vol-11 Issue-01 2021

We have mentioned the step by step procedure to analyze the dataset and find the correlation between the parameters. The dataset was then given as an input to algorithms and calculate the accuracy. We found that Random Forest Regression giving an accuracy of 94.32% respectively. For future work, we recommend that working on large dataset would yield a better and real picture about the model. We have undertaken only few Machine Learning algorithms, but we need to train many other algorithms and understand their predicting behaviour for continuous values too. By improving the error values this research work can be useful for development of applications.

REFERENCES :-

1. R. J. Shiller, "Understanding recent trends in house prices and home ownership," National Bureau of Economic Research, Working Paper 13553, Oct. 2007.
2. Torgo, Luis, and Joao Gama. "Regression using classification algorithms." Intelligent Data Analysis 1.4 (1997).
3. https://www.kaggle.com/ohmets/feature-selection-forregression/data
4. https://medium.com/@gurupratap.matharu/end-to-end-machine-learning-project-onpredicting-housing-prices-using-regression-7ab7832840ab https://www.kaggle.com/camnugent/california-housing-prices https://peltarion.com/knowledge-center/tutorials/predict-california-house-prices https://www.kaggle.com/subashdump/california-housing-price-prediction/data
5. Park, Byeonghwa, and Jae Kwon Bae. "Using machine learning algorithms for housing price prediction: The case of Fairfax County, Virginia housing data." Expert Systems with Applications 42.6 (2015): 2928-2934.
6. Limsombunchai, Visit. "House price prediction: hedonic price model vs. artificial neural network."New Zealand Agricultural and Resource Economics Society Conference. 2004.
7. S. C. Bourassa, E. Cantoni, and M. Hoesli, "Predicting house prices with spatial dependence: a comparison of alternative methods.