



House Price Prediction using KNN Algorithm

Vasanti Sakhare, Prof. Anup Gade, Prof. Abhay Rewatkar

Department of Information Technology, RTMNU
(Artificial Intelligence & Machine Learning)

Tulsiramji Gaikwad-Patil College of Engineering & Technology, Nagpur, India
sakharevasanti10@gmail.com

Abstract:

The real estate industry is one that makes frequent use of data mining. Data mining's use to forecasting real estate prices, vital housing features, and other elements is only the tip of the iceberg. Researchers have shown that uncertainty about real estate prices affects both buyers and sellers. To better understand which factors and models may predict future real estate values, a literature review is done. This study's results corroborated the superiority of artificial neural networks, SVR, and XGBoost over more conventional modeling techniques. Furthermore, our results indicate that both geographical and architectural factors have a significant impact on real estate values. Developers and academics alike may benefit greatly from this study by understanding which machine learning models provide the most reliable estimates of future property values.

Keywords- House Price Prediction, Machine Learning Model, Support Vector Regression, Artificial Neural Network

1. INTRODUCTION:

A safe and secure place to call home is very important, right up there with food and water. Housing demand increased as people's incomes increased. Some homeowners may buy a home as an investment, but the vast majority of buyers are motivated by more sentimental factors, either as a place to rest or to earn money. The housing market is a leading indicator of a country's economic health, as stated in [1]. The increased housing stock will have a multiplier impact on the economy as homeowners buy home furnishings and appliances and construction firms stock up on building supplies to meet surging housing demand. There is a correlation between the number of properties on the market and the health of the construction industry and the economy as a whole. According to reports, governments and human rights groups from all over the world have stressed the importance of having a place to call home [2].

The domestic market is heavily influenced by a country's economic, monetary, and political structures. However, as [3]

pointed out, home price fluctuations have always been a challenge for homeowners, structures, and the real estate industry, and as [4] said, the cost of owning a home has increased significantly due to the housing market's rapid internationalisation. Rising home prices might have far-reaching effects on the economy and people's standard of living. Those who want to use their home as an investment will feel the effects of this problem over time. Housing prices tend to increase in a cyclical manner in line with demand, as more people want to buy homes. To help investors make decisions and builders set prices, most stakeholders, including buyers and developers, homebuilders, and the real estate industry, would benefit from knowing the precise attributes or accurate factors influencing the house price when there are numerous variables, such as location and property demand, that may affect the house price [6]. There are several methods of forecasting (Machine Learning Model) future home prices. The models in question include the likes of support vector regression and artificial neural networks. Buyers, sellers, and developers are all in line to benefit from the house-price model. Homebuyers, property investors, and builders may all utilise the model's outputs to make educated decisions, such as the current market value of house prices. But this method has the potential to aid buyers in prioritising features that are most important to them within their budget [7]. The processes of identifying the factors that influence property values and using that knowledge to build a machine learning model that can then make predictions have been decomposed in previous research. However, this article brings together the concepts of property price forecasts and qualities into a coherent whole.

2. MACHINE LEARNING MODEL:

According to [8], there are two primary schools of thinking when it comes to anticipating the demand in the housing market: the conventional technique and the more modern quantitative model. Hedonic pricing, artificial neural networks (ANNs), and spatial regression are only a few alternatives to the conventional approach to value.

Conducting further research and using a more complex method of assessment. There are a variety of models available, so it's important to settle on the best one for making accurate predictions about future property values. Multiple studies have used Regression Analysis, making it one of the most popular models in this area of inquiry [9]. The Support Vector Regression (SVR) model is another popular approach to estimating future property prices. There are a number of key ways in which the housing market differs from the typical consumer market. [10] points out that unlike other markets, the housing market is resilient, adaptable, and not bound by geographical rigidity. For this reason, the hedonic method is favoured for determining market differentiation. The hedonic model was developed by [24] in 1939, but their work wasn't publicised until the 1960s.

Court developed this model in the early 1930s to examine the price and dependability of autos. According to [11], the pricing and precise amounts of items' features are hedonistic markers for economic actors. After perfecting the method for other real estate applications, Rosen used it to analyse house values [3]. It's possible to split Rosen's suggested paradigm or theory into two halves. An item's overall price may be calculated by first running a price regression on its individual components. The first step involves the quantification of a good's price, but does not include the generation of the inverse demand function. Therefore, the implicit pricing function from the first-stage estimate is necessary for the calculation of the inverse demand function. Taking an average, using a hedonic model, and utilising a matching methodology are the three most prevalent approaches to assessing property values, and they have all been compared in previous research. Both the basic average and the matching approaches were demonstrated to be biased when applied to the real estate market. When compared to the two most common substitutes, the hedonic model consistently produces superior outcomes [3]. According to the hedonic market theory, which forms the basis for hedonic pricing, a property's worth is directly related to the aggregate of the advantages it offers.

3. ATTRIBUTE:

The two main elements that may be used to describe the home price forecast are the house's attributes and the model utilised to create the house price prediction. Many academics have collaborated on a model to predict future home prices, and it takes into account. The pricing of pre-existing homes in Jakarta, Indonesia, is studied using a conceptual model and questionnaires in this study. Since the primary goal of this study is to rank the characteristics or qualities that impact the price of housing, the findings confirm the validity of this analysis by showing that the attributes or elements that affect

housing price vary between housing developments in Jakarta. A home's price is determined by a number of factors. Location, structural condition, and neighbourhood are the three main elements that determine house costs.

Location: In the real estate market, location is the single most essential factor. In addition, he found that characteristics of the surrounding area greatly influenced home prices in his research. It has been determined that the property's position is permanent. All of these studies show that house costs vary greatly depending on factors like proximity to a shopping mall or whether or not the home has a view of the hills or the seashore. Structural structure, or physical qualities as defined by certain studies, is another factor that affects home prices.

Structural Characteristic: The structural feature is an easily recognisable aspect of a home, and may include things like the square footage, number of rooms, garage, and outdoor space. These physical features, provided by builders and developers to entice consumers, are exactly what they're looking for. According to their earlier research, homebuyers would put the most weight on the structural features that contribute to the property's market worth. All of these characteristics, they said in their prior research, are associated with increasing home values.

Neighborhood: The neighborhood's amenities might be included into the asking price of the home. A property's value may increase if it is in a desirable location, has good public schools, and is close to restaurants and stores. As could be anticipated, home costs skyrocket when moving from a middle-class suburb to a wealthy one. The research discovered, however, that cultural differences make these characteristics less universally applicable.

4. PROPOSED METHOD:

This dataset focuses mostly on active real estate listings that are presently available for purchase. To compile this dataset of over 1635 items and 8 attributes, researchers used a technique often known as "manual web scraping." The information was compiled from a wide variety of real estate listing websites, such as 99acres, magic bricks, and housing.com. The first seven columns describe the area type, location, Society name, number of BHK rooms, number of bathrooms, total square feet area, price, and availability. The variable of interest here is price. In order to increase precision and stop model over fitting, it is usual procedure during data cleaning to get rid of irrelevant data columns. The model is protected from the curse

of dimensionality and can make more precise predictions as a result. Cleansing has a chilling effect on geography, culture, and accessibility. After the data was cleaned up, the model had five inputs and one output: a combination of four inputs and one independent variable. The value of price, however, is debatable.

Regression Techniques: The machine learning toolset includes both supervised and unsupervised learning strategies. Supervised Learning works best when there are a few examples of the desired result in the data used to train the system. This is the gold standard for improving one's odds of obtaining reliable outcomes. We employed a supervised learning technique to do this. Regression and classification are two types of supervised learning. Different from classification, regression uses a continuous value for the dependent attribute or outcome (predicted) variable. It is possible to use predictive modelling techniques like regression analysis to look at how different factors in a dataset affect the target variable. The goal is to find the line that best fits the data points by minimising the gap between them. Several Regression methods are being evaluated on the present problem statement to see which one yields the most reliable results.

Deployment of model:

The data was split into training and testing sets for the classification of the best fitting machine learning model. The standard 80-20 split ratio is used, a typical ratio for this purpose; 80% of the data is considered as a training set and 20% as a testing set. To allow the implementation of the model, Scikit-Learn have to be imported. It is a Python Library which provides machine learning algorithms for implementation and many more features for modeling. We are performing Supervised Learning and to find out best model we will be implementing some regression algorithms which are likely to do a precise estimation of prices. The model which gives least error and most nearer value prediction will be our final model.

5. METHODS of ANALYSIS:

Many parties have a vested interest in arriving at a reliable estimate of a property's worth, including sellers, purchasers, agents, lenders, and investors. It's a challenging one, too. It's no secret that characteristics like square footage, number of rooms, and proximity to popular tourist destinations all play a role in setting a hotel's rate. When a property must be sold quickly, for example, the price will likely be lower than usual because of the urgency of the sale. There are several methods for estimating a property's selling price, but regression analysis is a common one. All regression methods take in one

or more predictor variables and output a single outcome value. In this study, we evaluate the accuracy of several machine learning techniques for estimating home prices using data including square footage, number of bedrooms and and location.

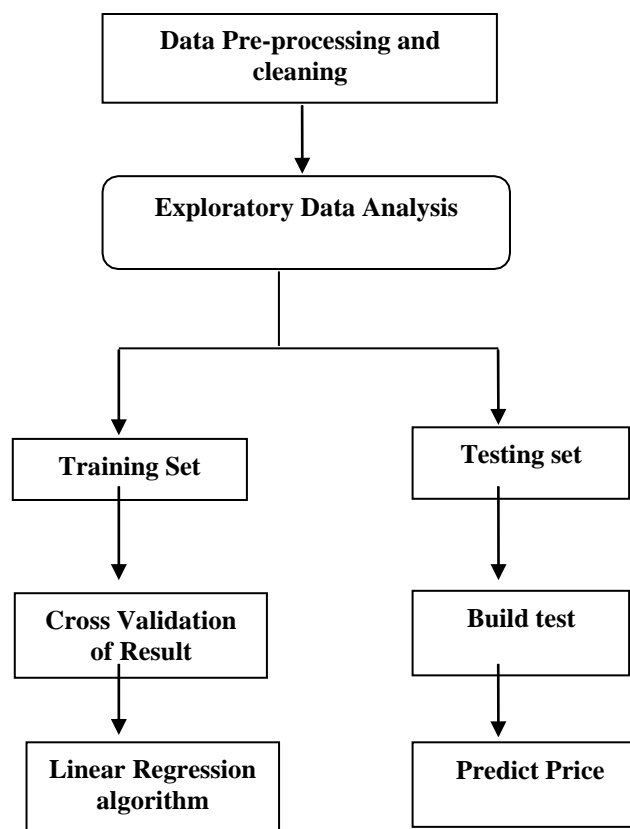


Fig.1: Block diagram showing Overall testing

MATLAB The findings are achieved by the use of neural network-based approaches. A wide range of factors that influence housing prices are considered and refined. Machine learning has been explored as a potential method of accomplishing the goal. First, we begin collecting data. The next step is data cleaning, which removes any lingering errors from the information. Preprocessing the data is the next stage. The goal of data visualisation is to graphically represent the data's distribution through the construction of a variety of charts. The home businesses' final expenses were determined with pinpoint accuracy. We use regression analysis on our house pricing dataset, but we also consider and apply several

classification algorithms, such as the SVM algorithm, decision tree algorithm, Random Forest classifier, etc., to help people buy homes at prices that are within their means. The prices of the houses are calculated using a number of different techniques. Better data and analysis have led to more accurate pricing. This would be very helpful to the general population. These results are the result of using Python's built-in data mining tools. Numerous factors affect property prices, so it's vital to consider and deal with them all. Machine learning assisted us in completing the task. It all starts with the data collection. The next step is data cleaning, which involves correcting any mistakes and making the data error-free. The data is then preprocessed. The information is then visualised in the form of various graphs. Multiple patterns of data distribution have been revealed. The model has also been tested and is ready for deployment. We found that certain classification methods had been applied to our dataset while others had not.

Therefore, we eliminated the algorithms that were not being used on our home price dataset and worked to improve the accuracy and precision of the remaining algorithms. To improve the accuracy of our categorization algorithms, we propose a novel stacking strategy. Improving the algorithms' accuracy and precision is essential for achieving better results. Incorrect findings would make it impossible for the public to make reliable predictions about future home sales prices. Accuracy and results were also enhanced via the use of data visualisation. The prices of the houses are calculated using a number of different techniques. Better data and analysis have led to more accurate pricing. This would be very helpful to the general population.

6. RESULTS:

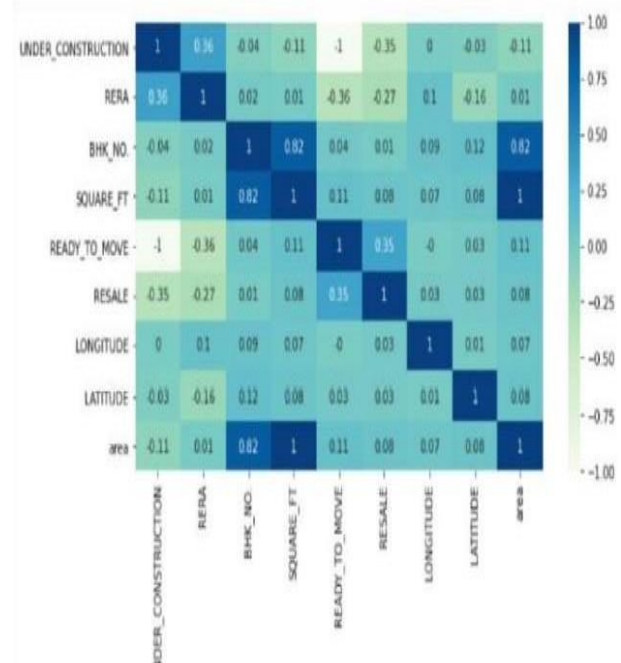


Fig.2: Heatmap for correlation between attributes

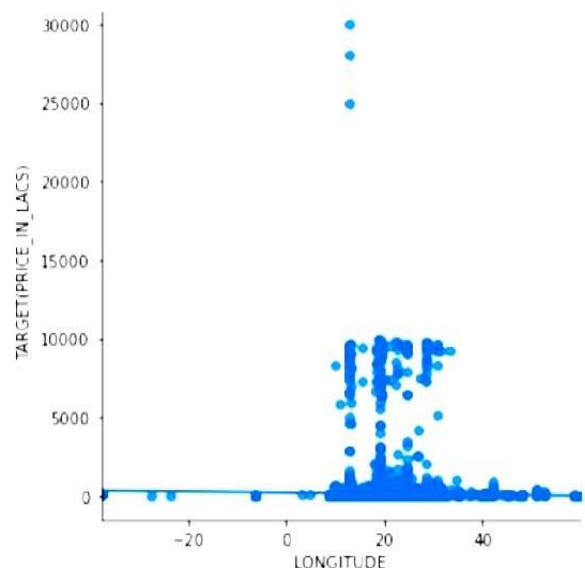


Fig.3: Line-plot (Implot) of Longitude Parameter



7. CONCLUSION:

In this work, we looked at the existing literature on forecasting property prices using data mining methods. Houses in prime locations, such as in close proximity to a shopping centre or other amenities, will cost more than similarly situated homes in more remote rural regions because of the greater availability of supporting services in the former.

A reliable model for predicting home prices would be helpful for investors and purchasers in determining a property's fair market worth. This research was conducted to provide light on the factors that have been employed by other scholars to foretell housing market movements in the future. The study indicates that SVR, ANN, and XGBoost may all be helpful in forecasting the direction of the housing market. These models were constructed with many different factors in mind, and they have shown to be rather accurate in projecting the direction of property market values. The study's primary goal was to offer data that researchers might use to create a model that reliably predicts movements in home prices in the future. Further testing of our results on a working prototype is required.

8. REFERENCES:

- [1] Feng Wang; Yang Zou; Haoyu Zhang; Haodong Sh, House Price Prediction Approach based on Deep Learning and ARIMA Model 2021 2nd Global Conference for Advancement in Technology (GCAT) IEEE Xplore 2020
- [2] D Preethi; Neelu Khare Performance Evaluation of Shallow learning techniques and Deep Neural Network for Cyber Security 2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE) IEEE Xplore: 27 April 2020.
- [3] Yihao Chen; Runtian Xue; Yu Zhang House price prediction based on machine learning and deep learning methods 2021 International Conference on Electronic Information Engineering and Computer Science (EIECS)IEEE Xplore: 09 November 2021.
- [4] Ankush Verma; Chetan Nagar; Nikita Singhi; Nupur Dongariya; Nidhi Sethi, Predicting House Price in India Using Linear Regression Machine Learning Algorithms, a review, 2022 3rd International Conference on Intelligent Engineering and Management (ICIEM) IEEE Xplore: 17 August 2022.
- [5] Sapiyah Binti Sakri; Zaiton Ali, Znet: Analysis of the Dimensionality Issues in House Price Forecasting Modeling, 2022 Fifth International Conference of Women in Data Science at Prince Sultan University (WiDS PSU) IEEE Xplore: 03 May 2022.
- [6] Jiahao Yang; Big Data Analyzing Techniques in Mathematical House Price Prediction Model 2022 IEEE International Conference on Electrical Engineering, Big Data and Algorithms (EEBDA) IEEE Xplore: 06 April 2022.
- [7] Uysal, İ., Güvenir, H. A. "An overview of regression techniques for knowledge discovery", The Knowledge Engineering Review, Vol. 14:4, 1999, 319±340 (KER 14404) Printed in the United Kingdom.
- [8] Bhuriya, Dinesh, et al. "Stock market prediction using a linear regression.", Electronics, Communication and Aerospace Technology (ICECA), 2017 International conference of. Vol. 2. IEEE 2017
- [9] Li, Li, and Kai-Hsuan Chu "Prediction of real estate price variation based on economic parameters", Applied System Innovation (ICASI), 2017 International Conference in IEEE, 2017
- [10] Pedregosa, Fabian, et al. "Scikit-learn: Machine learning in Python", Journal of machine learning research 12.Oct (2011): 2825-2830.
- [11] Maharshi Modi, Ayush Sharma, Dr. P. Madhavan "Applied Research On House Price Prediction Using Diverse Machine Learning Techniques", International Journal of Scientific & Technology Research Volume 9, Issue 04, April 2020.