House Prices Advanced Regression Techniques

MD. Suleman Khan*, Dr. Uma Rani Vanamala**

*Student, M.Tech, Department of Information Technology, Jawaharlal Nehru Technological University Hyderabad

** Professor, Department of Information Technology, Jawaharlal Nehru Technological University Hyderabad

Abstract: Since data mining can be utilized to remove important information from crude information, it is turning out to be increasingly more typical in the land business. This is on the grounds that data mining is an exceptionally viable device for estimating home estimations and tracking down significant lodging characteristics. The housing business sector and property holders are especially likewise underscores how pivotal underlying and locational factors are in deciding home estimations. The most valuable machine learning models for next concentrate on in this space will be distinguished, and lodging designers and scholastics will get a more noteworthy handle of the urgent components that influence home estimating. With the assistance of this review, housing market members will actually want to make more educated decisions and property costs will be anticipated all the more precisely.

Index Terms: House prices, Regression, Price prediction, Lasso regression

1. INTRODUCTION

One of the most important things in life is a house, along with basic needs like food, water, and other essentials. The increased demand for housing is a result of rising living standards. The majority of people buy houses simply for their shelter and means of subsistence, however other people see them as investments or assets. A nation's economy is greatly impacted by the housing market, which also positively correlates with the strength of the currency used in that nation. While contractors and homebuilders get raw materials to satisfy housing demand, homeowners buy things like furniture and appliances for their homes. This cycle serves as an example of how the supply of new homes has an impact on the economy [1].

Human rights activists and international organizations have emphasized the significance of housing, pointing out that it is deeply ingrained in each nation's political, economic, and financial systems. But price swings have always been a problem for builders, homeowners, and the real estate industry. Homes have become unaffordable in a number of nations due to significant price increase in the housing industry, which has an effect on both the national economy and the standard of living for citizens [2]. Those who consider housing as an investment are finding these price hikes especially difficult. Prices are indirectly raised by the yearly increase in demand for homes, which is impacted by factors including location and property demand. As a result, in order to make wise judgments and establish fair pricing, stakeholders—including purchasers, developers, builders, and the real estate sector—seek to identify the particular characteristics or elements that affect home prices [3]. worried about cost variances, which has prompted a lot of study into the relevant qualities and successful determining models. To decide the main qualities, this study overviews the writing and evaluates the viability of many AI models. Our outcomes approve that the best models for guaging home estimations are XGBoost and Random Forest. The concentrate

Support vector regression and artificial neural networks are two examples of machine learning models that may be used to predict house prices. For those who are building or purchasing a home, these models have several advantages. For example, they help determine pricing by offering insightful information about the current market appraisal of housing prices. These models may be used by prospective purchasers to choose homes that fit their needs and preferences in terms of features and budget [4]. Prior research has frequently concentrated on evaluating the characteristics that influence home values or independently forecasting home values using machine learning models. This piece, however, attempts to integrate the two by forecasting home values and identifying the factors that influence them at the same time [5].

Because machine learning algorithms can handle large quantities of data and produce precise predictions, they are being used more and more in the real estate sector. To forecast home values, these models can examine a number of variables, including market trends, property size, and location. In doing so, they assist interested parties in developing strategies to successfully traverse the housing market and in making well-informed judgments [6]. Moreover, these forecasting tools have the potential to improve market transparency by giving interested parties a better grasp of price dynamics and reducing the risks brought on by price volatility [7].

In summary, rising living standards and economic expansion are the main drivers of the ongoing increase in housing demand. It is vital for anyone involved in the real estate sector to comprehend the aspects that impact property prices. An effective approach for forecasting home values and examining the factors influencing them is the use of machine learning models. Homeowners, real estate speculators, and builders may all make better judgments by utilizing these models, which will ultimately help to create a more transparent and stable housing market.

2. LITERATURE SURVEY

Researchers, economists, and politicians investigate housing prices because of their influence on individuals and the economy. Understanding how various factors affect housing prices may help homebuyers, investors, and builders. Recent research and conclusions are used in this literature analysis to examine home price variables and forecasting models. Housing prices are complicated, impacted by economic conditions, geography, and market movements. Choong Wei Cheng's Petaling housing pricing research used statistical analysis to uncover major drivers. Cheng observed that location, amenities, and the economy affect house prices. His investigation showed that house prices rely on economic considerations, neighborhood value, and property characteristics [1].

Recent years have seen data-driven house price prediction gain popularity. Febrita et al. used fuzzy rule extraction to estimate home prices in Malang, East Java. Their research showed that data-driven techniques can capture housing market complexity. After reviewing historical data, scientists created a model that reliably projected house values depending on location, size, and market circumstances. This method enhanced forecast accuracy and revealed home price drivers [2].

Gao et al. used multi-task learning to forecast home prices by location. Their study focuses on location as a key house pricing factor. They created a prediction model that surpassed standard approaches by merging location-based data with other criteria. This research highlighted the importance of location in home price projections and showed how sophisticated machine learning may improve prediction accuracy [3].

Phan's Melbourne home price forecast study expanded the use of machine learning methods. Phan showed that machine learning algorithms might capture housing market dynamics. His studies showed that support vector regression and artificial neural networks outperformed standard statistical approaches in prediction. This study showed that machine learning can improve home price projections and help stakeholders make decisions [4].

Song et al. predicted Chinese property prices using a dendritic neuron model. Their study showed how sophisticated neural network models can capture nonlinear housing price correlations. The dendritic neuron model could handle complicated information and accurately predict the home price index. Advanced computer methods were stressed to improve house price projections [5].

Nur et al. predicted Malang, East Java house prices using regression and particle swarm optimization. Both standard statistical approaches and optimization strategies improved predicted accuracy in their investigation. Using particle swarm optimization, they improved the regression model and prediction outcomes. This study showed that hybrid techniques may manage housing market complexity and improve prediction models [6].

Yusof and Ismail used multiple regression analysis to discover home price drivers. Their study examined economic statistics, property attributes, and market circumstances. Multiple regression analysis helped them quantify each aspect and create a reliable home pricing model. This study stressed the need for a diversified approach to housing market dynamics and prediction models [7].

Recent work has focused on house price prediction using machine learning techniques. These models can manage massive datasets, capture complicated connections, and make reliable predictions. Statistics show that machine learning models like support vector regression, artificial neural networks, and dendritic neuron models are more accurate and reliable than standard statistical approaches. These new methods allow researchers to create models that reveal home price drivers and empower stakeholders to make educated decisions.

Finally, house prices depend on economic conditions, geography, and market trends. Recent study shows that data-driven and machine learning algorithms can anticipate house prices. These models provide house purchasers, investors, and builders useful information because to their accuracy and reliability. Advanced computational methods will help us understand house price drivers and improve forecasting models as housing markets develop. These insights help stakeholders understand housing market intricacies and make educated decisions that boost economic growth.

3. METHODOLOGY

a) Proposed Work:

Random forest and Xgboost are two instances of different prediction models (otherwise called ML models) that might be utilized to expect house prices. The house-price model offers a few benefits for home developers, financial backers, and buyers. Homebuyers, financial backers, and developers will track down an abundance of data and skill in this model, remembering the valuation of house costs for the ongoing business sector, which will support setting house evaluating. In the in the mean time, this model can help imminent property holders in choosing the elements of a home that best suits their necessities and financial plan. Earlier exploration focused on looking at the elements that impact home estimations and making autonomous forecasts about home estimations utilizing ML models. Be that as it may, this post incorporates both trait and property price prediction together.

b) System Architecture:

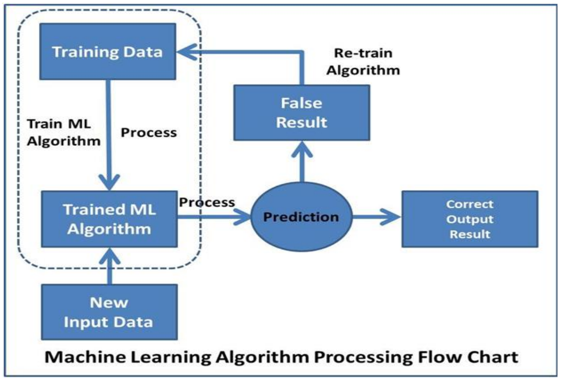

Fig 1 Proposed Architecture

In order to anticipate housing prices, the suggested system design first feeds fresh input data into a machine learning (ML) algorithm that has been trained. To forecast home prices, this algorithm analyzes the data. The findings are then divided into true and false categories. Accurate outcomes go on to give stakeholders insightful information or useful choices. Nevertheless, retraining is triggered in cases of false positives. Re-examining the training data, which has been selected and improved in light of both new and historical data insights, is part of the re-training process. The ML algorithm is retrained using this improved training set, which improves the algorithm's accuracy and predictive power. After undergoing retraining, the machine learning system is prepared to handle fresh input data and endeavors to consistently enhance the precision of subsequent forecasts. By going through this iterative cycle, the machine learning model is able to adjust to evolving data patterns and maintain its accuracy in predicting home values in the real estate market.

c) Dataset Collection:

A thorough dataset collection is necessary for this study on data mining-based house price prediction. A broad range of parameters should be included in the dataset, such as structural features like the number of bedrooms, bathrooms, and overall square footage, as well as locational considerations like proximity to amenities, schools, and transit hubs. Economic factors including interest rates, past pricing patterns, and state of the local market should also be mentioned. To guarantee accuracy and applicability, the information should come from reputable real estate databases, public documents, and surveys. To get the dataset ready for analysis, preparation operations like filling in missing values and normalizing the data would be required. With the use of this extensive dataset, machine learning models—Random Forest and XGBoost in particular—will be able to estimate housing prices with greater accuracy and pinpoint the key variables affecting these values.

d) Data Processing:
In order to employ machine learning models like Random Forest and XGBoost to accurately estimate home prices, data processing is an essential stage in the preparation of the dataset. First, the gathered data has to be cleaned up. This includes getting rid of duplicates, filling in the gaps with imputation or deletion, and fixing any mistakes or inconsistencies. The data must next be converted into a format that is appropriate for analysis. Using methods like one-hot encoding, categorical variables—such neighborhood names or property types—should be transformed into numerical values. Normalization or standardization of numerical characteristics may be necessary to guarantee that each feature contributes equally to the performance of the model. Moreover, feature engineering may be used to generate new, pertinent qualities from the available data, improving the prediction ability of the model. Preventing biased outcomes also requires the identification and elimination of outliers. To assess the performance of the model, the data should be divided into training and testing sets once it has been cleaned and converted. Following processing, the data is prepared for entry into machine learning models, which allow precise estimation of home values and identification of significant influencing variables. The validity and dependability of the prediction models are guaranteed by this meticulous data processing.

e) Feature Extraction:
Improving the forecast accuracy of home price models requires feature extraction. From the original data, new, pertinent characteristics are identified and created throughout this process. For example, further information may be obtained by extracting attributes like the age of the property, the distance to the closest city center, or school quality ratings. Capturing the subtle impacts on home values is made easier by converting unrefined characteristics into more meaningful measurements, such as price per square foot or the proportion of bathrooms to bedrooms. Efficient feature extraction guarantees that the most relevant and informative data are supplied into machine learning models, such as Random Forest and XGBoost.

f) Training & Testing:
The dataset is split into two subsets for the training and testing phases of the suggested architecture for predicting home prices: training data and testing data. The machine learning algorithm learns patterns and correlations within the data by using the training data to teach it. The testing data is used to evaluate the algorithm's performance and predicted accuracy once it has been trained. When the model is applied in real-world circumstances, this procedure guarantees that it can accurately estimate home values and that it generalizes well to new data. The robustness and dependability of the model are improved by iterative improvement through training and testing.

g) Algorithms:
Random Forest Algorithm:
Several decision tree classifiers are used in the Random Forest algorithm, an ensemble technique that improves predicting accuracy. Using a random feature selection and a fraction of the training data, each tree in the forest is constructed separately. Because of the trees' diversity, Random Forest performs well on a variety of machine learning classification and regression problems by reducing overfitting and enhancing resilience.

Gradient Boosting Algorithm:
In order to reduce prediction errors, gradient boosting is a potent machine learning approach that generates a succession of weak learners, usually decision trees, progressively. In regions where prior models have underperformed, iteratively fitting new models produces predictions that are more accurate. Gradient Boosting is a popular technique for jobs like classification, regression, and ranking that call for a high degree of prediction accuracy. It can manage big datasets with ease while retaining interpretability.

4. CONCLUSION

In conclusion, this work collected existing research on key home price drivers and assessed data mining methods for house price prediction. Strategic placement, such as accessibility to retail malls, affects housing values, unlike rural areas without such conveniences. Investors, purchasers, and developers need accurate prediction models to set fair property prices.

The assessment of variables utilized by earlier studies showed that Random Forest and XGBoost models predict property values. These models employ numerous input factors to show substantial positive correlations with home prices across datasets. They help stakeholders navigate the difficult real estate market with their solid performance.

This research seeks to help future studies construct viable housing price prediction models. The results from this study can be used to improve existing models or generate new ones to improve forecast accuracy and real-world applicability. Researcher refinement and validation of these models can improve real estate prediction, promoting informed decision-making and sustainable market practices.

This study established the foundation for sophisticated machine learning approaches in real estate and emphasizes the necessity of precise price projections for global housing market transparency and efficiency.

5. FUTURE SCOPE

Looking ahead, there are a number of interesting directions that future research on housing price prediction might go. First off, improving the accuracy of prediction models might involve using sophisticated data sources like sentiment analysis from social media and geolocation data. Furthermore, the integration of dynamic elements such as policy modifications and economic indicators might yield more comprehensive understanding of market oscillations. Furthermore, building interpretable AI models will improve stakeholders' confidence and comprehension of forecasts. Finally, concentrating on scalable solutions that take into account big data difficulties will guarantee that models continue to work well as datasets get larger. Future research in these areas can further the field and help make more informed judgments about development, market regulation, and real estate investment.

REFERENCE

[1] A. S. Temür, M. Akgün, and G. Temür, "Predicting Housing Sales in Turkey Using Arima, Lstm and Hybrid Models," J. Bus. Econ. Manag., vol. 20, no. 5, pp. 920–938, 2019, doi: 10.3846/jbem.2019.10190.

[2] A. Ebekozien, A. R. Abdul-Aziz, and M. Jaafar, "Housing finance inaccessibility for low-income earners in Malaysia: Factors and solutions," Habitat Int., vol. 87, no. April, pp. 27–35, 2019, doi: 10.1016/j.habitatint.2019.03.009.

[3] A. Jafari and R. Akhavian, "Driving forces for the US residential housing price: a predictive analysis," Built Environ. Proj. Asset Manag., vol. 9, no. 4, pp. 515–529, 2019, doi: 10.1108/BEPAM-07-2018-0100.

[4] Choong Wei Cheng, "Statistical Analysis of Housing Prices in Petaling," Universiti Tunku Abdul Rahman, 2018.

[5] R. E. Febrita, A. N. Alfiyatin, H. Taufiq, and W. F. Mahmudy, "Data-driven fuzzy rule extraction for housing price prediction in Malang, East Java," 2017 Int. Conf. Adv. Comput. Sci. Inf. Syst. ICACSIS 2017, vol. 2018-Janua, pp. 351–358, 2018, doi: 10.1109/ICACSIS.2017.8355058.

[6] G. Gao et al., "Location-Centered House Price Prediction: A Multi-Task Learning Approach," pp. 1–14, 2019, [Online]. Available: http://arxiv.org/abs/1901.01774.

[7] T. D. Phan, "Housing price prediction using machine learning algorithms: The case of Melbourne city, Australia," Proc. - Int. Conf. Mach. Learn. Data Eng. iCMLDE 2018, pp. 8–13, 2019, doi: 10.1109/iCMLDE.2018.00017.

[8] Y. Y. S. Song, T. Zhou, H. Yachi, and S. Gao, "Forecasting house price index of China using dendritic neuron model," PIC 2016 - Proc. 2016 IEEE Int. Conf. Prog. Informatics Comput., pp. 37–41, 2017, doi: 10.1109/PIC.2016.7949463.

[9] R. Aswin Rahadi, S. K. Wiryono, D. P. Koesrindartoto, and I. B. Syamwil, "Factors Affecting Housing Products Price in Jakarta Metropolitan Region," Int. J. Prop. Sci., vol. 6, no. 1, pp. 1–21, 2016, doi: 10.22452/ijps.vol6no1.2.

[10] A. Nur, R. Ema, H. Taufiq, and W. Firdaus, "Modeling House Price Prediction using Regression Analysis and Particle Swarm Optimization Case Study : Malang, East Java, Indonesia," Int. J. Adv. Comput. Sci. Appl., vol. 8, no. 10, pp. 323–326, 2017, doi: 10.14569/ijacsa.2017.081042.

[11] A. Yusof and S. Ismail, "Multiple Regressions in Analysing House Price Variations," Commun. IBIMA, vol. 2012, pp. 1–9, 2012, doi: 10.5171/2012.383101.

[12] A. Osmadi, E. M. Kamal, H. Hassan, and H. A. Fattah, “Exploring the elements of housing price in Malaysia,” Asian Soc. Sci., vol. 11, no. 24, pp. 26–38, 2015, doi: 10.5539/ass.v11n24p26.

[13] T. L. Chin and K. W. Chau, “A critical review of literature on the hedonic price model,” Int. J. Hous. Sci. Its Appl., vol. 27, no. 2, pp. 145–165, 2003.

[14] M. J. Ball, “Recent Empirical Work on the Determinants of Relative House Prices,” Urban Stud., vol. 10, no. 2, pp. 213–233, 1973, doi: 10.1080/00420987320080311.

[15] M. Rodriguez, “Managing Corporate Real Estate: Evidence from the Capital Markets.” Journal of Real Estate Literature, 1996.