House Price Predictor

Ananya Priya
UG Student, Dept. of CS Engg.
Sir M. Visvesvaraya Institute of Technology
Bengaluru, India

Anusha Gupta
UG Student, Dept. of CS Engg.
Sir M. Visvesvaraya Institute of Technology
Bengaluru, India

Divya Drishti
UG Student, Dept. of CS Engg.
Sir M. Visvesvaraya Institute of Technology
Bengaluru, India

Ms. Rekha B N
Associate Professor,Dept. of CS Engg.
Sir M. Visvesvaraya Institute of Technology
Bengaluru, India

ABSTRACT

House Prices increase every year and having a house is a basic necessity of any human being. So for searching and finding about the house and its price people usually need to go to the broker who gives them an approximation of the price of the house which includes his profit also. That doesn't provide an accurate result about the price which gives rise to a need for the system which will predict the price of the house with less error. Machine Learning is a technology which is growing rapidly nowadays. Various algorithms are there which are useful for different kinds of applications. Prediction of houses using a machine learning algorithm gives more accurate results. Here we compared different models of machine learning for predicting house prices like but Gradient Boosting Algorithm surprasses additional models of machine learning in predicting the house prices.

***Keywords*– House price prediction, machine learning, linear regression,support vector regression, decision tree regression, Gradient Boosting Regression,XGBoost Regression, Kneighbors**

I INTRODUCTION

One of the most crucial elements of modern life and economy is the property market. Due to the significant inflation or the changes in house prices raise the need to accurately predict future prices.However, for the majority of people, housing is one of their biggest outlays. Therefore, making the right choice when buying a home can help consumers save money and even occasionally turn a return on their investment.

House is a basic necessity of an individual.As development of civilization is happening at a very high rate the demand for the houses are also increasing day by day.Whenever people go to a different place they need a house to live.They usually go to the brokers regarding its price. But generally brokers tell them the price which is not so accurate.They see their own profit and tell the price to them.But in Today's world Time and money both are valuable to the people.Spending extra money for buying the house or extra time to manually computing the price of the house is very difficult hence House price predictor is not only useful for the customers who are buying the house rather it plays a vital role for the sellers too in determining the selling price of the house.

Machine learning is generally used for predicting house prices. There are various algorithms present which can determine the price of the house with more accuracy. Some of them are KNeighbors algorithm, Gradient Boosting Algorithm, Decision Tree Regression etc. These algorithms basically work on the datasets which are used to train these models and then test other datasets.There are different factors which play a key role in determining the house price which

includes the number of bathrooms, neighborhood of the house,lot area, street, year in which it is built etc. Machine learning uses all these features to determine the house price more accurately.

The house price prediction model makes use of machine learning techniques, which enable computers to learn from data, to anticipate new data. The method of predictive analytics that is most frequently used is regression.Regression methods are frequently used to create models based on many criteria to forecast prices. In this research, we attempted to use a gradient boosting regression model that could forecast property prices after comparing different models.

II LITERATURE SURVEY

Ayush Varma,Abhijit Sarma,Sagar Doshi & Rohini Nair,et al.,2020[1] have proposed various regression techniques like forest regression,boosted regression ,linear regression and neural network and they have taken the mean of all the techniques to get the better and efficient result.This idea provides more accurate result and less error in comparison to the individual algorithms used.Google Map is also used to get real time neighborhood details for more accuracy.Data mining was the main component of the system proposed.

Dr. M. Thamarai &Dr. S P. Malarvizhi ,et al.,2020[2]have suggested utilizing the Scikit-Learn Machine Learning Tool to anticipate home prices, the decision tree classification, regression, and multiple linear regression algorithms. This model was made for Godavari district of Andhra Pradesh .Also the comparison has been done between multiple linear regression and decision tree regression out of which multiple linear regression was found better.

Manasa J,Radha Gupta & Narahari N S ,et al.,2020[3] have proposed different regression technique like Support vector regression, Lasso and Ridge regression models, multiple linear regression (Least Squares), and boosting methods like Extreme Gradient Boost Regression are examples of regression techniques (XGBoost)etc are compared.In this paper the data set has nine features and is open to the public. Since the data was collected for these models in 2016 so now these are not so relevant as the data was collected for Bengaluru City which is growing rapidly. Also the methodology and an SVR method were run utilizing the retrieved housing data. The experimental findings at the end demonstrate that the suggested strategy predicts individual property prices better than the SVR method. Additionally, when making short-term predictions, the trend in house price predictions is essentially in line with the actual data.

Danh Phan ,et al.,2018[4] have proposed a combination of Stepwise and Support Vector Machine which is based on mean squared error measurement in this model. Regression Trees ,Polynomial Regression and Neural Network were used in the model and results were compared.So it was found that regression trees give good prediction results just like linear regression while polynomial regression gives less errors.Also Neural network does not give satisfactory results.While principle component analysis and SVM gave the better result in terms of accuracy.

Maida Ahtesham, Narmeen Zakaria Bawany & Kiran Fatima ,et al.,2020[5] have forecast housing prices, we suggested the Gradient Boosting Model XGBoost. The Open Data Pakistan website is where the dataset was obtained. XGBoost was primarily chosen because it supports parallel processing, making it faster and more flexible. Different features like location,property type etc were taken into consideration for determining the price of the house.Accuracy was good which shows that using this model(XGBoost) compared to other models was proven better in terms of accuracy.

Debanjan Banerjee & Suchibrota Dutta,et al., 2017[6] have proposed a model using various classifiers. All of these classifiers are supervised models. Because supervised classifiers draw on historical data and house price prediction necessitates historical data, their use is acceptable. The effectiveness of various classifiers has been assessed based on their accuracy, precision, sensitivity, and specificity. Although the experiment results show that Random Forest provides more accuracy, this particular classifier is also more prone to overfitting.

Feng Wang, Yang Zou, Haoyu Zhang and Haodong Shi et al.,2019[7]have put out a model that employs deep learning and an ARIMA model, with the home price prediction model being built using the Tensor Flow framework. 13 explanatory variables are initially selected as the key, in addition. We compare the outcomes of deep learning and machine learning models first. Second, we demonstrate how focus mechanisms can improve prediction accuracy using a variety of attention models. Third, to confirm the performance of our suggested

model. After the raw housing data is acquired from the Internet using Scrapy, the data preparation stage comes next.

Quang Truong, Minh Nguyen, Hy Dang, Bo Mei,et al.,2019[8] have proposed Hybrid Regression technique which examines various housing price forecast methods.Three separate machine learning approaches—Random Forest, XGBoost, and LightGBM—as well as two machine learning techniques should be used for the best outcomes. Stacked Generalization and hybrid regression The strategy used in this is stacked generalization. When accuracy is the top priority, the regression technique is the best choice.

Pei-ying wang, Chiao-ting chen, Jain-wun su, Ting-yun wang & Szu-hao huang,et al.,2021[9] have put up a reliable model for accurate house prediction. It is used to examine the housing markets in Kaohsiung and Taichung. In order to complete the information about the house in this study, heterogeneous data is added into the model.

Mansi Jain, Himani Rajput, Neha Garg & Pronika Chawla ,et al.,2020[10] have proposed different regression techniques and use of different python libraries to predict the price of the houses.It also briefs various mathematical and numerical techniques used for predicting the price. Cross validation,classification algorithm like SVM Algorithm, Decision tree algorithm etc, data mining techniques & machine learning was used in this model. Firstly data collection then data cleaning and then data preprocessing has been used.

Table 1. SUMMARY OF LITERATURE SURVEY

Ref No.	Year	Authors	Research Problem	Data Source	Algorithm	Results
[11]	2019	G. Naga Satish, Ch. V. Raghavendran, M.D.Sugnana Rao, Ch.Srinivasulu	House Price Prediction Using Machine Learning	The dataset is taken from kaggle.	Machine learning algorithm, lasso regression process and neural system,Gradient Boosting algorithm.	The accuracy of the gradient algorithm, which has an accuracy of 91%, is higher than the accuracy of the other two algorithms, which are 76.15 and 76.14%
[12]	2020	Choujun Zhan, Yonglin Liu, Wangling Chen, Zeqiong Wu, Zefeng Xie	Housing prices prediction with deep learning: an application for the real estate market in Taiwan	The available real estate transaction data from the Taiwan Ministry of Housing were used in this investigation.	Deep learning, housing prices prediction, time series	CNN is regarded as the best model because its R2 value is greater than 0.945.
[13]	2019	CH.Raga Madhuri, Anuradha G, M.Vani Pujitha	House Price Prediction Using Regression Techniques: A Comparative Study	Information on home sales that is based on actual data from Kingcounty. It is a publicly available dataset of that	Multiple linear, Ridge, LASSO, Elastic Net, Gradient boosting and AdaBoost Regression.	All algorithms are compared and it was found that gradient boosting has higher accuracy than all other algorithms used.

				particular US region.		
[14]	2017	Li Li and Kai-Hsuan Chu	Prediction of Real Estate Price Variation Based on Economic Parameters	For analysis and forecasting, the variance in real estate in Taipei, Taiwan between 2005 and 2015 that was relevant to the general public was taken into account.	Back propagation neural network (BPN) and radial basis function neural network (RBF)	The prediction of house price variation was not as accurate using the NN model as compared to the other two indexes, Cathy and Sinyi house price index
[15]	2020	Syafiqah Jamil, Thuraiya Mohd, Suraya Masrom, Norbaya Ab Rahim	Machine Learning Price Prediction on Green Building Prices	The GB dataset was taken into consideration which covers Kuala Lumpur District, Malaysia.	Linear Regression, Decision Tree, Random Forest, Ridge and Lasso Regression	Out of all the algorithm techniques applied to build the model for house price prediction ,Decision Tree outperformed others.
[16]	2018	Marco De Nadai and Bruno Lepri	The economic value of neighborhoods: Predicting real estate prices from the urban environment	The information for the eight Italian cities was gathered from various sources. Real estate properties made up a sizable dataset.	Gradient Boosted Tree and XGBoost	The algorithm was applied to cities and it was found that the features of the neighborhood seem to drive more than 20% of the home price.Also the prediction error drops by 60%.
[17]	2017	Sifei Lu, Zengxiang Li, Zheng Qin, Xulei Yang, Rick Siow Mong Goh	A Hybrid Regression Technique for House Prices Prediction	The dataset, which included 79 explanatory variables, was provided by Kaggle for certain Home Equity Transactions in Ames, Jowa.	Ridge Regression methodology, hybrid Lasso and Gradient boosting regression, and creative feature engineering approach	The hybrid regression model, which used 65% Lasso and 35% Gradient Boost Algorithm, produced the best result of 0.11260.
[18]	2019	Yiyang Luo	Residential Asset Pricing Prediction using Machine Learning	The data comes from the Ames. The Ames housing dataset served as the foundation for the Kaggle house price competition.	Random forest,support vector machine,OLS Regression	The outcome demonstrated that it obtained R squared > 0.9, indicating that the feature chosen can explain factors influencing home price.
[19]	2020	Bandar Almaslukh	A Gradient Boosting Method for Effective	Using real estate market data from the past that was	Gradient Boosting Method, Randomized	It is clear that the OGB model yields the best R-squared

			Prediction of Housing Prices in Complex Real Estate Systems	gathered from Sindian District in New Taipei City, Taiwan and stored in the UCI repository	Search Algorithm	result and the worst results for the MAE, RMSE, and MAPE metrics.
[20]	2018	J. J. WANG , S. G. HU1 , X. T. ZHAN , Q. LUO , Q. YU , ZHEN LIU , T. P. CHEN , Y. YIN , SUMIO HOSAKA , AND Y. LIU	Predicting House Price With a Memristor-Based Artificial Neural Network	Boston suburb data is utilized	BP algorithm	After learning to do so while operating in the training mode, the ANN can accurately predict the price of a house in the predicting mode.

III IMPLEMENTATION

We will be using the Gradient Boosting Model to predict the house price after comparing other models in terms of accuracy. Gradient Boosting is an Algorithm which basically uses the weak learners and combines them to form a strong learner.Each time new models are trained to basically remove or reduce the loss function of the previous model by using Gradient Descent Algorithm.

The various steps consists of:
1.Importing various Libraries
2.Loading the dataset and get brief information about the dataset like statistical information
3.Preprocessing the dataset
4.Splitting the dataset into training and testing data
5.Feature Scaling
6.Selection of model among various models
7.Training of Model and Validating the model

Data Collection:- Dataset is collected for house price prediction.It should contain various features which will be helpful for determining the price of the house accurately like lot area, roofstyle, year in which house is built, basement quality of the house, neighborhood of the house etc. This dataset should be loaded for training and testing the model but it can be in any format like sql so it should be changed to csv before loading.

Data Preprocessing:- This step consists of various substeps:
Data Understanding:- Knowing the data and its various features so that impact of different features on the target variable can be determined. It is done by getting statistical information of the dataset or brief information of the dataset.

Data Cleaning:-It is the process to detect the irrelevant data or to detect incomplete data. Various rows of the dataset can contain missing values of one or more features.But machine learning models generally give error if it has to deal with the missing values. So missing values should be handled before training that dataset, either the whole feature should be dropped if that feature doesn't impact the price of the house or else some other values should be replaced in that place. Various methods are there to compute those values which can be using statistics like mean,median mode or using Machine Learning Algorithms etc.

Feature Transformation:-It is basically the process of transforming the feature representation from one to another.As our dataset contains numerical feature as well as categorical feature so it is needed to convert the categorical feature into numerical one so that statistics can be applied to them.The various methods used are ordinal Encoding, Label Encoding, One Hot Encoding etc depends upon the data on which it is applied.

Splitting the Dataset Into Training And Testing data:- The dataset is then splitted into training and testing data in which training data is basically used to train the model while testing data is used to predict the value of the target variable.It can be done in the ratio of 80-20% or else different dataset can be used for that.

Feature Scaling:- It is basically the process which transform the value of the features on a similar scale.This is basically done so that contribution of each and every features towards

Determination of house price should be equal.There will be no dominating feature.This process depends on the application on which it is applying.

Model Selection:- There are various models which can be used to determine or predict the house price but not all the models are equally accurate.So a comparison has been done to see which is more accurate and then model is selected according to that.Also here we have used Regression as we have to determine continuous values.We have used linear regression, Support vector regression,Kneighbors, Decision Tree, Gradient Boosting,Random Forest,XGBoost Regression.

Training of Model And Validating:- We trained the model using above algorithms,validated the model using K-Fold Cross Validation Technique and selected Gradient Boosting Algorithm.

```
Training model with its Score : LinearRegression   [-1.7285974116877867e+25]
Training model with its Score : Support Vector Regression   [-0.052133638285474 2]
Training model with its Score : KNeighborsRegressor   [0.559462077907711]
Training model with its Score : DecisionTreeRegressor   [0.6572375726092475]
Training model with its Score : GradientBoosting Regression   [0.8700799181660634]
Training model with its Score : RandomForest Regression   [0.8443453596140967]
Training model with its Score : XGBoost Regression   [0.8420479088628255]
```

IV CONCLUSION

In this paper, we used an optimized Gradient Boosting Regression Model to predict the house price which provides 87% accuracy. This model is trained with a number of features and impact of each feature is taken into consideration while determining the price.For validating the model K-Fold Cross Validation is used which is again an advanced technique which is used so that validation can be done properly.Metrics such as mean squared error, R-squared score, and root mean squared error are used to assess the algorithm's efficiency.House price prediction has several benefits for purchasers and clients, including improved choice-making, which allows buyers and consumers to make more informed decisions about buying or selling a house then cost savings, which helps to save money by negotiating better deals with avoiding costly mistakes, and transparency in the real estate market which gives transparency by making property prices more accessible to buyers and consumers. This can aid in the reduction of data disparities and the creation of a more competitive and efficient market.

To summarize, house price prediction is a complicated task that involves an in-depth knowledge of the information set, suitable preprocessing, feature selection, model selection, and validation.

IV REFERENCES

[1] A. Varma, A. Sarma, S. Doshi and R. Nair, "House Price Prediction Using Machine Learning and Neural Networks," 2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT), Coimbatore, India, 2018, pp. 1936-1939, doi: 10.1109/ICICCT.2018.8473231.

[2] Dr. M. Thamarai & Dr. S P. Malarvizhi ,2020 "House Price Prediction Modeling using Machine Learning",I.J. Information Engineering and Electronic Business, 2020, 2, 15-20 IEEE Xplore Part Number: CFP20K58-ART; ISBN: 978-1-7281-4167-1

[3] J. Manasa, R. Gupta and N. S. Narahari, "Machine Learning based Predicting House Prices using Regression Techniques," 2020 2nd International Conference on Innovative Mechanisms for Industry Applications (ICIMIA), Bangalore, India, 2020, pp. 624-630, doi: 10.1109/ICIMIA48430.2020.9074952.

[4] T. D. Phan, "Housing Price Prediction Using Machine Learning Algorithms: The Case of Melbourne City, Australia," 2018 International Conference on Machine Learning and Data Engineering (iCMLDE), Sydney, NSW, Australia, 2018, pp. 35-42, doi: 10.1109/iCMLDE.2018.00017.

[5] M. Ahtesham, N. Z. Bawany and K. Fatima, "House Price Prediction using Machine Learning Algorithm - The Case of Karachi City, Pakistan," 2020 21st International Arab Conference on Information Technology (ACIT), Giza, Egypt, 2020, pp. 1-5, doi: 10.1109/ACIT50332.2020.9300074.

[6] D. Banerjee and S. Dutta, "Predicting the housing price direction using machine learning techniques," 2017 IEEE International Conference on Power, Control, Signals and Instrumentation Engineering (ICPCSI), Chennai, India, 2017, pp. 2998-3000, doi: 10.1109/ICPCSI.2017.8392275.

[7] F. Wang, Y. Zou, H. Zhang and H. Shi, "House Price Prediction Approach based on Deep Learning and ARIMA Model," 2019 IEEE 7th International Conference on Computer Science and Network Technology (ICCSNT), Dalian, China, 2019, pp. 303-307, doi: 10.1109/ICCSNT47585.2019.8962443.

[8] Quang Truong, Minh Nguyen, Hy Dang, Bo Mei,2019 "Housing Price Prediction via Improved Machine Learning",International Conference on Identification, Information and Knowledge in the Internet of Things (IIKI2019)

[9]P. -Y. Wang, C. -T. Chen, J. -W. Su, T. -Y. Wang and S. -H. Huang, "Deep Learning Model for House Price Prediction Using Heterogeneous Data Analysis Along With Joint Self-Attention Mechanism," in IEEE Access, vol. 9, pp. 55244-55259, 2021, doi: 10.1109/ACCESS.2021.3071306.

[10] M. Jain, H. Rajput, N. Garg and P. Chawla, "Prediction of House Pricing using Machine Learning with Python," 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC), Coimbatore, India, 2020, pp. 570-574, doi: 10.1109/ICESC48915.2020.9155839.

[11] G. Naga Satish, Ch. V. Raghavendran, M.D.Sugnana Rao, Ch.Srinivasulu ,2019 "House Price Prediction using Machine Learning" International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-8 Issue-9, July 2019

[12]C. Zhan, Z. Wu, Y. Liu, Z. Xie and W. Chen, "Housing prices prediction with deep learning: an application for the real estate market in Taiwan," 2020 IEEE 18th International Conference on Industrial Informatics (INDIN), Warwick, United Kingdom, 2020, pp. 719-724, doi: 10.1109/INDIN45582.2020.9442244.

[13]C. R. Madhuri, G. Anuradha and M. V. Pujitha, "House Price Prediction Using Regression Techniques: A Comparative Study," 2019 International Conference on Smart Structures and Systems (ICSSS), Chennai, India, 2019, pp. 1-5, doi: 10.1109/ICSSS.2019.8882834.

[14]L. Li and K. -H. Chu, "Prediction of real estate price variation based on economic parameters," 2017 International Conference on Applied System Innovation (ICASI), Sapporo, Japan, 2017, pp. 87-90, doi: 10.1109/ICASI.2017.7988353.

[15] S. Jamil, T. Mohd, S. Masrom and N. Ab Rahim, "Machine Learning Price Prediction on Green Building Prices," 2020 IEEE Symposium on Industrial Electronics & Applications (ISIEA), TBD, Malaysia, 2020, pp. 1-6, doi: 10.1109/ISIEA49364.2020.9188114.

[16] M. De Nadai and B. Lepri, "The Economic Value of Neighborhoods: Predicting Real Estate Prices from the Urban Environment," 2018 IEEE 5th International Conference on Data Science and Advanced Analytics (DSAA), Turin, Italy, 2018, pp. 323-330, doi: 10.1109/DSAA.2018.00043.

[17] S. Lu, Z. Li, Z. Qin, X. Yang and R. S. M. Goh, "A hybrid regression technique for house prices prediction," 2017 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM), Singapore, 2017, pp. 319-323, doi: 10.1109/IEEM.2017.8289904.

[18] Y. Luo, "Residential Asset Pricing Prediction using Machine Learning," 2019 International Conference on Economic Management and Model Engineering (ICEMME), Malacca, Malaysia, 2019, pp. 193-198, doi: 10.1109/ICEMME49371.2019.00046.

[19] B. Almaslukh, "A Gradient Boosting Method for Effective Prediction of Housing Prices in Complex Real Estate Systems," 2020 International Conference on Technologies and Applications of Artificial Intelligence (TAAI), Taipei, Taiwan, 2020, pp. 217-222, doi: 10.1109/TAAI51410.2020.00047.

[20] J. J. Wang et al., "Predicting House Price With a Memristor-Based Artificial Neural Network," in IEEE Access, vol. 6, pp. 16523-16528, 2018, doi: 10.1109/ACCESS.2018.2814065.