IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

MASTERING HOUSE PROPERTY VALUATION USING MACHINE LEARNING

[1]K. Harsha Vardhan,[2]Dr K Ramesh Babu, [3]SK.Asif, [4]K.Indu,[5]S.Bharani,[6]G.Mahesh babu

[1]Assistant professor, [2]Professor ,[3,4,5,6]Research scholar,

[1,2,3,4,5,6]CSE Department "GVR&S College of Engineering and Technology, Guntur, Andhra Pradesh, India

ABSTRACT: FORESEEING HOME ESTIMATIONS PRECISELY IS FUNDAMENTAL FOR VARIOUS LAND RELATED EXERCISES, LIKE FINANCIAL PLANNING, BUYING, AND SELLING. IN THIS FIELD, MACHINE LEARNING CALCULATIONS HAVE BECOME BASIC INSTRUMENTS FOR MAJOR AREAS OF STRENGTH FOR BUILDING MODELS. THIS EXPLORATION UTILIZES MACHINE LEARNING STRATEGIES TO GIVE A TOTAL WAY TO DEAL WITH LODGING COST FORECAST. THERE ARE A FEW SIGNIFICANT STAGES IN THE RECOMMENDED SYSTEM. TO BEGIN WITH, DATA IS ASSEMBLED FROM VARIOUS SOURCES, LIKE LAND POSTINGS, SEGMENT INFORMATION, FINANCIAL POINTERS, AND PAST DEALS INFORMATION. THEN, INCLUDE DESIGNING IS FINISHED TO EXTRICATE APPROPRIATE INFORMATION AND FURTHER DEVELOP THE MODEL'S ANTICIPATING SKILL. FOR MODEL PREPARATION, AN ASSORTMENT OF MACHINE LEARNING TECHNIQUES ADA BOOST REGRESSION, CAT BOOST, XG BOOST, LIGHT GBM GRADIENT BOOSTING ARE UTILIZED, INCLUDING INCLINATION SUPPORTING OF LAND, CHOICE TREES, IRREGULAR TIMBERLANDS, AND STRAIGHT RELAPSE. R2 VALUE OF GRADIENT BOOSTING REGRESSOR IS 0.780372 AND THAT IS THE HIGHEST R2 VALUE WHEN COMPARED TO OTHERS.

INDEX TERMS - MACHINE LEARNING, HOUSE PRICE PREDICTION, GRADIENT BOOSTING REGRESSOR ETC.,

1.INTRODUCTION

Over quite a while in the past, there is physically choose the cost of any property. Individuals who don't have a clue about the genuine cost of that specific house and they endure deficiency of cash. The house value expectation of the house is finished utilizing different ML calculations like Less fatty Relapse, Choice Tree Relapse, K-Means Relapse and Arbitrary Woods Relapse. 80% of information structure kwon dataset is utilized for preparing reason and staying 20% of information utilized for testing reason. This work applies different methods, for example, highlights, names, decrease strategies and change procedures like property blends, set missing traits as well as searching for new connections. This all demonstrates that house value expectation is an arising research region and it requires the information on ML. Lately, the housing market has become progressively unique and intricate, driven by different financial, social, and ecological variables. Anticipating house costs precisely is significant for various partners, including purchasers, dealers, realtors, and financial backers. Conventional techniques for valuation frequently depend on manual appraisal In and authentic information examination, which can be tedious and inclined to human blunder. ML offers a promising answer for this test by utilizing progressed calculations to dissect immense measures of information and concentrate significant examples. By using ML procedures, we can foster models that foresee house costs with more noteworthy precision as well as adjust to changing economic situations in genuine time.here means to investigate the utilization of ML in foreseeing house costs, using a dataset containing different elements like area, size, conveniences, and neighbourhood qualities. Via preparing and calibrating ML models on verifiable lodging information, we look to foster a prescient model that can gauge the selling cost of a house in light of its credits.

2.LITERATURE SURVEY

In the [1] authorized by Maida Ahtesham, Narmeen Zakaria Bawany, Kiran Fatima(2020). A number of variables, such as a home's location, size, and number of bedrooms, affect its price. This study uses the gradient boosting model XG Boost to forecast house prices .The 38.961 records in the publicly accessible dataset about Karachi City were obtained from an open real estate platform in Pakistan. In the [2] authorized by The Danh Phan(2018). Past real estate transactions are analyzed machine learning techniques to find practical models for buyers and sellers of real estate. The wide disparity in home prices between the priciest and most affordable districts of Melbourne is made clear. In the [3] authorized by Aman Chaurasia , Inam Ui Haq(2023). In order to improve the accuracy of house price predictions, they have developed a machine learning-based algorithm in this research article. The suggested model concurrently makes use of machine learning algorithms and data pre-processing methods. Actual home price data is used to assess the suggested model's efficacy. In the [4] authorized by Chen Chee Kin, Zailan Arabee Bin Abdual Salam, Kadhar Batcha Nowshath (2022). Numerous digital technologies have been created to assess the purchasers' budgetary restrictions and property marketing strategies. This article aims to forecast house prices for those who do not own a property, taking into account their financial capabilities and desired lifestyle.A variety of methods, including chatbots, artificial neural networks, and machine

learning, will be used to generate the anticipated prices. In the [5] authorized by Ze Li(2021), a house price index, or HPI, is important for providing reliable information to people who need it, including bank finance departments, real estate investment firms, and homeowners. We investigate the correlations between the frequency, HPI, and flavor components in 99326 samples.place name, place id, level, period, year of sale or rental, and hpi type and index.An improved model, such as xgboost, might be selected to enhance the prediction outcome. In the [6] authorized by H. Zhang et al, Four regression algorithms are trained using only textual data, non-textual data, or both. Our results show that by using exclusively the description data with Word2Vec and a Deep Learning model, we can achieve good performance. However, the best overall performance is obtained when using both textual and non-textual features. Overall, we observe that combining the textual and non-textual features improves the learned model and provides performance benefits when compared against using only one of the feature types. And uses final developed model with Word2Vec and Deep Learning to predict the house price. In the [7] authorized by Dr.M.Thamarai, Proposed work makes use of the attributes or features of the houses such as number of bedrooms available in the house, age of the house, travelling facility from the location, school facility available nearby the houses and Shopping malls available nearby the house location. House availability based on desired features of the house and house price prediction are modeled in the proposed work and the model is constructed for a small town in West Godavari district of Andhra pradesh. The work involves decision tree classification, decision tree regression and multiple linear regression and is implemented using Scikit-Learn Machine Learning Tool. In the [8] authorized by T. Swetha Chowdary, The price of a property is affected by a number of variables, including its location, size, age, condition, and recent market trends. Homes in upscale neighbourhoods or in places with high cost of living are typically more costly. The size of the house and the amount of land it sits on both significantly affect the cost. Regression methods like linear regression, decision tree regression,and lasso regression are some examples of machine learning techniques that are increasingly utilised to forecast home values. However, it's crucial to make sure that the data utilised for prediction is reliable and indicative of the market in order to produce more accurate forecasts and assist buyers and sellers in making smarter decisions. In the [9] authorized by Amey Thakur, they propose to implement a house price prediction model of Bangalore, India. It's a Machine Learning model which integrates Data Science and Web Development. We have deployed the app on the Heroku Cloud Application Platform. Housing prices fluctuate on a daily basis and are sometimes exaggerated rather than based on worth. The major focus of this project is on predicting home prices using genuine factors. Here, we intend to base an evaluation on every basic criterion that is taken into account when establishing the pricing. The goal of this project is to learn Python and get experience in Data Analytics, Machine Learning, and AI. In the [10] authorized by ALAN IHRE, The algorithms were selected from an assessment of previous research and the intent was to compare their relative performance at this task. Software implementations for the experiment were selected from the scikit learn Python library and executed to calculate the error between the actual and predicted sales price using four different metrics. Hyperparameters for the algorithms used were optimally selected and the cleaned data set was split using five-fold cross-validation to reduce the risk of bias. An optimal subset of hyperparameters for the two algorithms was selected through the grid search algorithm for the best prediction. The Random Forest was found to consistently perform better than the kNN algorithm in terms of smaller errors and be better suited as a prediction model for the house price problem. With a mean absolute error of about 9 % from the mean price in the best case, the practical usefulness of the prediction is rather limited to making basic valuations.

3. PROPOSED MODEL

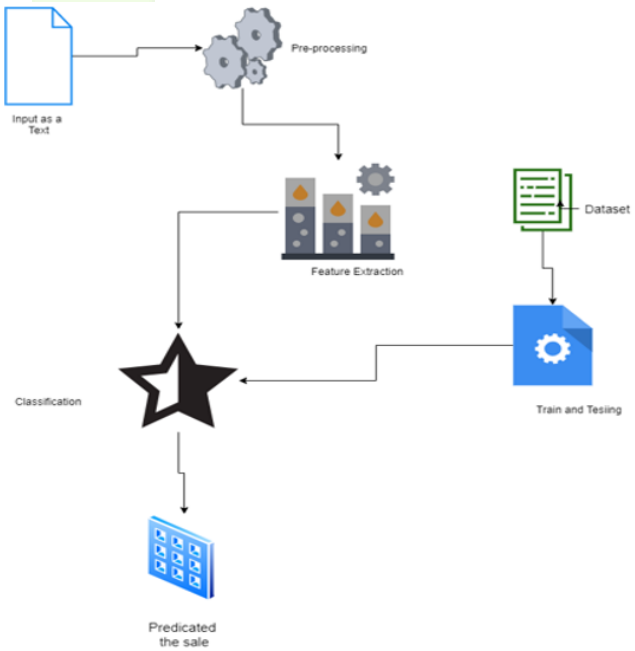


Fig 3.1 System Architecture

4.RESEARCH METHODOLOGY

1.Data Collection: Gather comprehensive datasets containing information about past real estate transactions, including property features (e.g., size, number of bedrooms and bathrooms, amenities), location details, sale prices, and transaction dates.Utilize public real estate databases, government records, online listing platforms, and third-party data providers to collect relevant data.Ensure data quality by addressing missing values, inconsistencies, and outliers through data cleaning techniques.

2. Data Preprocessing: Perform exploratory data analysis (EDA) to gain insights into the distribution and relationships between variables. Handle categorical variables by encoding them into numerical representations using techniques such as one-hot encoding or label encoding. Standardize or normalize numerical features to ensure uniformity and mitigate the scale effect. Address missing values through imputation techniques such as mean, median, or mode substitution, or employ advanced methods like predictive imputation or iterative imputation.

3. Feature Engineering: Extract meaningful features from the raw data to capture relevant information influencing house prices.Create new features or transform existing ones based on domain knowledge and intuition. Feature selection techniques like correlation analysis, feature importance scores, or recursive feature elimination may be employed to identify the most predictive features and reduce dimensionality**.**

4. Model Selection: Evaluate a variety of machine learning algorithms suitable for regression tasks, including linear regression, decision trees, random forests, support vector machines (SVM), gradient boosting machines (GBM), and neural networks. Consider the trade-offs between model complexity, interpretability, and predictive performance.Utilize techniques such as grid search or randomized search for hyperparameter tuning to optimize model performance.

5. Model Training: Split the dataset into training, validation, and test sets to evaluate model performance. Train the selected machine learning models on the training data using appropriate optimization algorithms. Employ cross-validation techniques such as k-fold cross-validation to assess model generalization and mitigate overfitting.

6. Model Evaluation: Evaluate the trained models on the validation set using appropriate regression evaluation metrics such as mean squared error (MSE), root mean squared error (RMSE), mean absolute error (MAE), and R-squared. Compare the performance of different models to select the best-performing one for deployment.

7. Deployment and Monitoring

Deploy the chosen model into production to make real-time predictions on new or unseen data. Implement monitoring mechanisms to track model performance and drift over time, retraining the model periodically to maintain its accuracy and relevance. In summary, the methodology of house price prediction using machine learning involves a systematic approach encompassing data collection, preprocessing, feature engineering, model selection, training, evaluation In, deployment, and monitoring. By following these steps, stakeholders can develop accurate and robust prediction models to inform decision-making in the real estate domain.

5.DATASET INFORMATION

The dataset for house price prediction using machine learning typically contains various attributes related to properties, locations, and sale prices. Here's the typical information you would find in such a dataset:

i). Property Features:

- Size of the property (e.g., square footage, number of rooms)
- Number of bedrooms and bathrooms
- Type of property (e.g., single-family home, condominium, apartment)
- Amenities (e.g., swimming pool, garage, backyard)
- Age of the property
- Condition of the property (e.g., renovated, new construction)

ii). Location Details:

- Address or geographical coordinates (latitude and longitude)
- Neighbourhood information (e.g., crime rates, school district, amenities)
- Proximity to key locations (e.g., city center, public transportation, parks)

iii).Sale Price Information:

- Sale price of the property
- Date of sale or listing
- Currency of the sale price (e.g., USD, EUR)
- Any additional costs or fees associated with the sale

iv). Market Indicators:

- Economic indicators (e.g., GDP growth rate, inflation rate)
- Real estate market trends (e.g., housing market index, mortgage rates)
- Demographic data (e.g., population density, household income)

v). Additional Features:

- Any other relevant features that may influence property prices, such as:
- Accessibility to amenities (e.g., shopping centers, restaurants)
- Environmental factors (e.g., air quality, noise pollution)
- Historical data on property prices in the area

The dataset should ideally be comprehensive, covering a diverse range of properties across different locations and time periods. It should also be clean and well-structured, with minimal missing values or inconsistencies.

Datasets for house price prediction are available from various sources, including government agencies, real estate listings websites, and research organizations. Some popular datasets used in research and practice include the "House Prices: Advanced Regression Techniques" dataset from Kaggle, which contains detailed information on residential properties in Ames, Iowa, and the "California Housing Prices" dataset from the Start Lib repository, which includes housing data from the 1990 California census.When working

with a dataset for house price prediction, it's essential to thoroughly understand the data, perform exploratory data analysis (EDA), and preprocess the data appropriately before training machine learning models. This ensures the accuracy and reliability of the prediction models built on the dataset.

6.ALGORITHM USED

i)Ada Boost Regressor: Ada Boost Regressor is a machine learning algorithm used for regression tasks. It works by iteratively training weak learners, typically decision trees, and focusing on instances that are harder to predict correctly.

ii)XG Boost:XG Boost (Extreme Gradient Boosting) is a popular and powerful machine learning algorithm used for both regression and classification tasks. It is known for its efficiency, speed, and accuracy, and it often outperforms other gradient boosting algorithms.

iii)Cat Boost : Cat Boost is another popular gradient boosting library like XG Boost and Light GBM. It's specifically designed to handle categorical features efficiently, hence the name "Cat Boost." It offers high performance, support for large datasets, and robustness against overfitting.

iv) Light GBM: Light GBM (Light Gradient Boosting Machine) is a gradient boosting framework developed by Microsoft. It's known for its speed, efficiency, and accuracy, particularly on large datasets. Light GBM is designed to handle large-scale and high-dimensional data efficiently and is widely used in both industry and academia.

v)Gradient Boosting:A potent machine learning method for both regression and classification issues is gradient boosting.

1. Boosting Technique: Gradient Boosting trains a sequence of weak learners (usually decision trees) one after the other, fixing mistakes produced by the earlier trees.
2. Objective Function Optimization: Gradient Boosting adds trees while optimizing a loss function.
3. Gradient Descent: The term "gradient" in Gradient Boosting describes the process of minimizing the loss function by means of gradient descent.
4. Shrinkage: Gradient Boosting frequently uses a shrinkage parameter to reduce overfitting and enhance generalization.
5. Tree Structure and Depth: Every tree that is introduced to the ensemble is usually shallow; these trees are also called basic learners or weak learners.
6. Clearing Requirement: Gradient Boosting keeps adding trees until a certain quantity is attained.

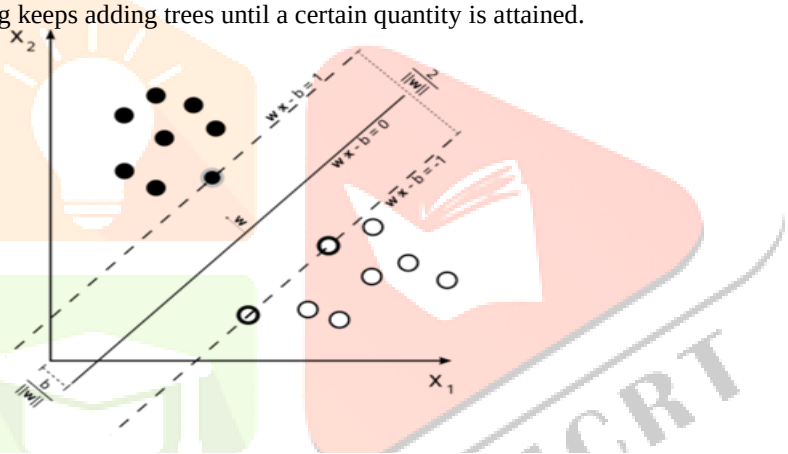


Fig 3.2. Supervised Learning

A support-vector machine is a supervised learning model that divides the data into regions separated by a linear boundary. Here, the linear boundary divides the black circles from the white

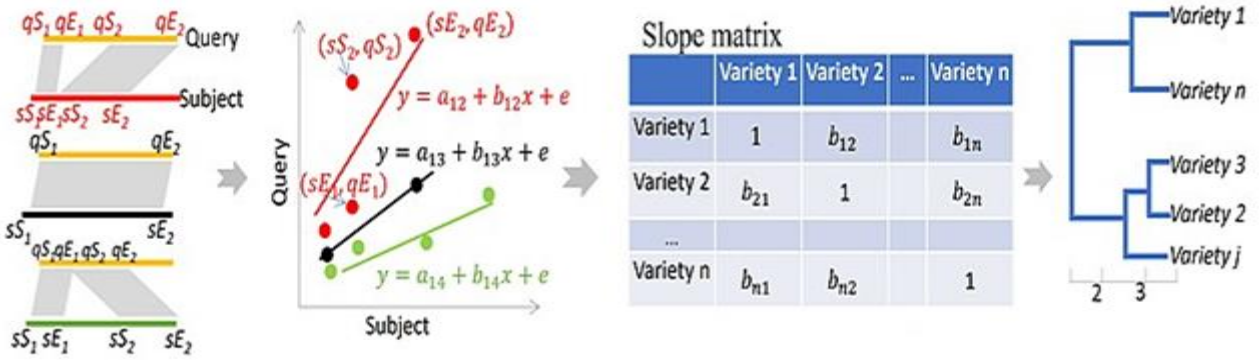


Fig 3.3 Unsupervised Learning

Clustering via Large Indel Permuted Slopes, CLIPS, turns the alignment image into a learning regression problem. The varied slope (b) estimates between each pair of DNA segments enables to identify segments sharing the same set of indels.

7. RESULTS

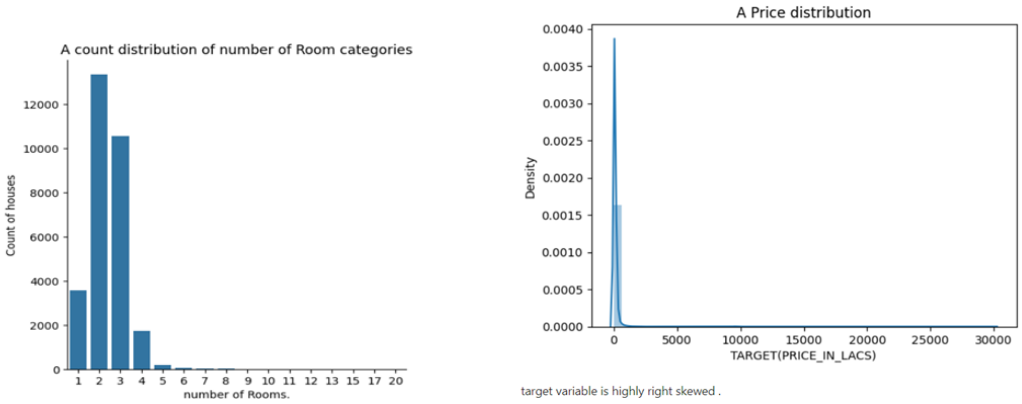


Fig 4.2 Price Distribution

The graph describes the ratio between density and Target (price in lakhs). The target variable is highly right skewed.

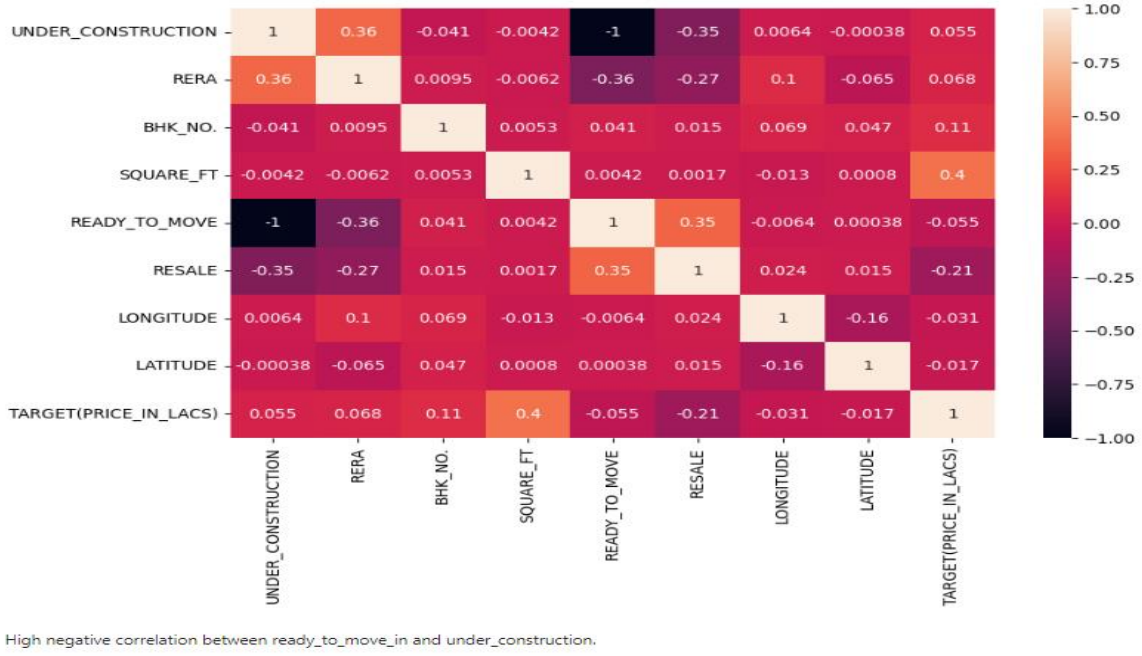


Fig 4.3 HeatMap

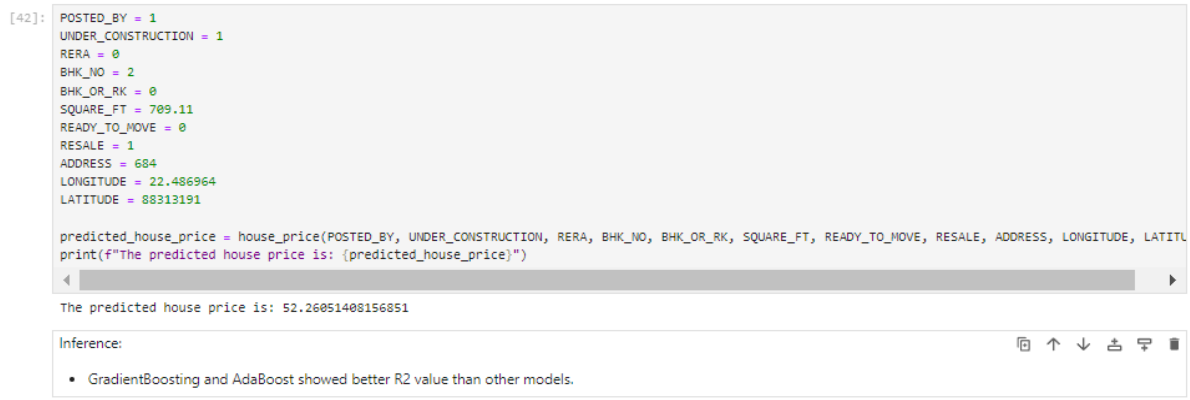
Heat Map is a visual representation of data where value are depicted by colour.
Heatmaps are used to show relationships between two variables, one plotted on each axis. By observing how cell colors change across each axis, you can observe if there are any patterns in value for one or both variables.

Tested Regressors:

Table 4.4 R2 Score OF House Price Prediction

	Regressor	Mean Squared Error	Mean Absolute Error	R2 Score
0	AdaBoost	137282.498055	145.078777	0.748044
1	GradientBoosting	119668.190992	46.463855	0.780372
2	XGBoost	260302.164156	40.052288	0.522265
3	CatBoost	167895.562661	43.459234	0.691860
4	LightGBM	206448.611504	45.276948	0.621103

R2 score must be in between 1 and 0. The highest R2 score Regressor will be used in the project As GradientBoosting Regressor has the highest R2 score, here considered GB Regressor to complete the work.

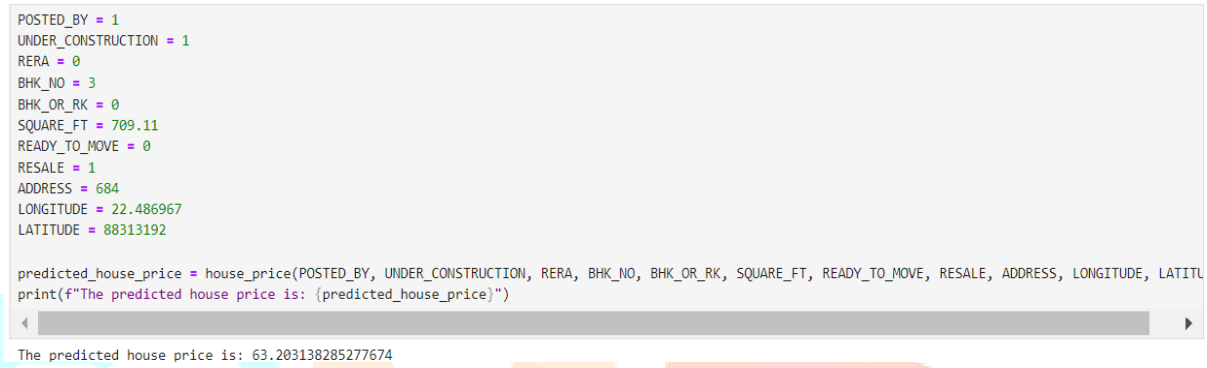


```
[42]: POSTED_BY = 1
      UNDER_CONSTRUCTION = 1
      RERA = 0
      BHK_NO = 2
      BHK_OR_RK = 0
      SQUARE_FT = 709.11
      READY_TO_MOVE = 0
      RESALE = 1
      ADDRESS = 684
      LONGITUDE = 22.486964
      LATITUDE = 88313191

      predicted_house_price = house_price(POSTED_BY, UNDER_CONSTRUCTION, RERA, BHK_NO, BHK_OR_RK, SQUARE_FT, READY_TO_MOVE, RESALE, ADDRESS, LONGITUDE, LATITU
      print(f"The predicted house price is: {predicted_house_price}")

      The predicted house price is: 52.26051408156851

      Inference:

      • GradientBoosting and AdaBoost showed better R2 value than other models.
```

Fig 4.5 Result generation

When we give the inputs such as BHK number, square feet, Resale, Address, Longitude and Latitude, we get the predicted price of the house. The above screenshot shows the predicted price for a house which is located under 22.486964 and 88313191

Here's another result generation



```
POSTED_BY = 1
UNDER_CONSTRUCTION = 1
RERA = 0
BHK_NO = 3
BHK_OR_RK = 0
SQUARE_FT = 709.11
READY_TO_MOVE = 0
RESALE = 1
ADDRESS = 684
LONGITUDE = 22.486967
LATITUDE = 88313192

predicted_house_price = house_price(POSTED_BY, UNDER_CONSTRUCTION, RERA, BHK_NO, BHK_OR_RK, SQUARE_FT, READY_TO_MOVE, RESALE, ADDRESS, LONGITUDE, LATITU
print(f"The predicted house price is: {predicted_house_price}")

The predicted house price is: 63.203138285277674
```

Fig 4.6 another Result generation

8.CONCLUSION

machine learning techniques offer immense potential for House price prediction, here used AdaBoost Regressor, XG Boost Regressor, Cat Boost Regressor, Light GBM Regressor, Gradient Boosting Regressor to fulfil the requirements. When compared every Regressor one by one got to know that Gradient Boost Regressor have high R2 value. R2 value should be in between 1 and 0. The R2 value of Gradient Boosting Regressor is 0.780372 and that is the highest R2 value when compared to others. So, its concluded that Gradient Boosting Regressor is better than rest.

9.FUTURE SCOPE

10.REFERENCES

[1]Zhang, Hanxiang, Yansong Li, and Paula Branco. "Describe the house and I will tell you the price: House price prediction with textual description data." *Natural Language Engineering* (2023): 1-35.

[2]Thamarai, M., and S. P. Malarvizhi. "House Price Prediction Modeling Using Machine Learning." *International Journal of Information Engineering & Electronic Business* 12.2 (2020).

[3]Vineeth, Naalla, Maturi Ayyappa, and B. Bharathi. "House price prediction using machine learning algorithms." *Soft Computing Systems: Second International Conference, ICSCS 2018, Kollam, India, April 19–20, 2018, Revised Selected Papers 2*. Springer Singapore, 2018.

[4]Amey Thakur, Mega Satish. "BANGALORE HOUSE PRICE PREDICTION." *Department of Computer Engineering, University of Mumbai, Mumbai, MH, India, https://www. academia. edu/79097924/BANGALORE_HOUSE_PR ICE_PREDICTION* (2021).

[5]Engström, Isak, and Alan Ihre. "Predicting house prices with machine learning methods." (2019).

[6]Ahtesham, Maida, Narmeen Zakaria Bawany, and Kiran Fatima. "House price prediction using machine learning algorithm-the case of Karachi city, Pakistan." *2020 21st International Arab Conference on Information Technology (ACIT)*. IEEE, 2020.

[7]Soltani, Ali, et al. "Housing price prediction incorporating spatio-temporal dependency into machine learning algorithms." *Cities* 131 (2022): 103941.

[8]Thamarai, M., and S. P. Malarvizhi. "House Price Prediction Modeling Using Machine Learning." *International Journal of Information Engineering & Electronic Business* 12.2 (2020).

[9]Kin, Chen Chee, Zailan Arabee Bin Abdul Salam, and Kadhar Batcha Nowshath. "Machine learning based house price prediction model." *2022 International Conference on Edge Computing and Applications (ICECAA)*. IEEE, 2022.

[10]Li, Ze. "Prediction of house price index based on machine learning methods." *2021 2nd International Conferenceon Computing and Data Science (CDS)*. IEEE, 2021.