JOURNAL OF BASIC SCIENCE AND ENGINEERING

HOUSE PRICE ESTIMATION USING MACHINE LEARNING

Sonali P.Bhoite[1], Vishal Nayakwadi[2], Mandar K Mokashi[3], Yogeshwari Mahajan[4], Tejali Katkar[5],

[1]Assistant Professor, Department of Information Technology, Vishwakarma Institute of Information Technology, Pune (India)

[2]Assistant Professor, Department of AI&DS, Zeal College of Engineering and Research,Pune (India)

[3]Assistant Professor, Department of Information Technology, Vishwakarma Institute of Information Technology, Pune (India)

[4]Assistant Professor, Department of Information Technology, Pimpri Chinchwad College of Engineering and Research, Pune

[5]Assistant Professor, Department of Artifiial Intlligence and Data science, D.Y.Patil College of Engineering , Pune (India)

Abstract—**The House Price Estimator system aims to compare an efficiency of the two popular machine learning algorithms, linear regression, and XGBoost, in predicting the selling price of a house. The proposed system will involve collecting a dataset of historical real estate transactions and selecting relevant parameters such as location, square footage, number of rooms and bathrooms, and other amenities. The data will be pre-processed, cleaned, and transformed to prepare it for modeling. The linear regression and XGBoost algorithms will be implemented and trained on the same dataset, and their performance will be evaluated using a range of measures, including R-squared, mean squared error, and root mean squared error.**

The ultimate goal of the proposed system is to determine which algorithm produces more accurate and reliable predictions and can be used to build an effective house price estimator. The results of this proposed system can help the buyers, the sellers, and the real estate personnel to derive informed decisions about pricing and selling houses.

Keywords— Machine learning, model, linear Regression, XG Boost

Introduction

The construction industry is a dynamic and complex system, where various parameters like location, size, amenities, and many more features can influence the selling price of the property. Accurately predicting the price of a house is a critical task for buyers, sellers, and real estate professionals, as it can help them make informed decisions about buying, selling, or renting a property. To develop a ML model to predict the selling/buying price of a property based on various features. The performance of two popular algorithms, linear regression, and XGBoost, in predicting the price of the house. Simple and popular, linear regression is an algorithm. for regression tasks, while XGBoost is a more complex and powerful algorithm that can handle complex relationships between features. The dataset for this proposed system will be collected from historical real estate transactions and will contain various features such as location, square footage, number of bedrooms and bathrooms, and other amenities. The data will be pre-processed, cleaned, and transformed to prepare it for modelling. Here both linear regression and XGBoost algorithms are implemented and train them on the same dataset. The performance of the models will be evaluated using several metrics like as mean squared error, root mean squared error and R-squared. The results of this proposed system can be used to build an accurate and reliable house price estimator that can assist buyers, sellers, and real estate professionals in making informed decisions. By comparing the performance of linear regression and XGBoost algorithms, we can determine which algorithm produces more accurate and reliable predictions and can be used to build an effective house price estimator.

LITERATURE REVIEW

[1] The study emphasises how crucial it is to have a precise model for predicting property prices in order to promote socioeconomic growth and public welfare. Using accessible datasets, the use of machine learning methods including Random Forest, Decision Trees, and Linear Regression is investigated for the purpose of predicting home values. The study made use of housing datasets from Carnegie Mellon University's StatLib library, which had 13 feature variables and 506 sample data. The study recognises that a variety of characteristics, including location, area, and the number of rooms, have a significant impact on housing costs; thus, all of this information is necessary to anticipate the price of a particular home. In order to examine the diverse effects of characteristics on prediction techniques, the study compares and contrasts a number of sophisticated models.

[2]The paper highlights the significance of precise house cost forecasts for potential buyers who are cautious approximately their budgets and advertise techniques. The objective of the paper is to estimate coherent house costs for non- householders based on their money related arrangements and yearnings. The paper employments relapse procedures such as Numerous straight, Edge, Rope, Flexible Net, Slope boosting, and Ad Boost Relapse to foresee house costs on a dataset. The think about analyzes the past stock and passage ranges, and figures improvements to appraise conjectured costs. The paper points to help dealers in assessing the offering fetched of a house precisely and to offer assistance individuals foresee the correct time slap to amass a house. Related variables that affect the fetched such as physical conditions, concept, and area, are too taken into thought. Generally, the paper gives important bits of knowledge into the utilize of relapse procedures for house cost forecast and highlights the significance of exact expectation models for making well-informed choices in the genuine domain industry. The consider is advantageous for both dealers and buyers who are cautious

Vol. 21, No. 1, (2024)
ISSN: 1005-0930

approximately their budgets and advertise methodologies.

[3] The paper highlights the importance of lodging cost forecast models in setting up genuine domain approaches and anticipating future lodging costs in the US lodging advertise. The ponder utilizes machine learning calculations such as C4.5, RIPPER, Naïve Bayesian, and AdaBoost to create a lodging cost forecast demonstrate based on the lodging information of 5359 townhouses in Fairfax Province, Virginia, accumulated by the Numerous Posting Benefit (MLS) of the Metropolitan Territorial Data Frameworks (MRIS).The inquire about compares the classification precision execution of the distinctive machine learning calculations and proposes an made strides lodging cost expectation show to help a house dealer or a genuine domain specialist in making better-informed choices based on house cost valuation. The RIPPER calculation reliably beats the other models in the execution of lodging cost forecast, based on precision. Generally, the paper gives important bits of knowledge into the utilize of machine learning calculations for lodging cost forecast and highlights the significance of exact expectation models in helping house dealers or genuine estate agents in making better-informed decisions

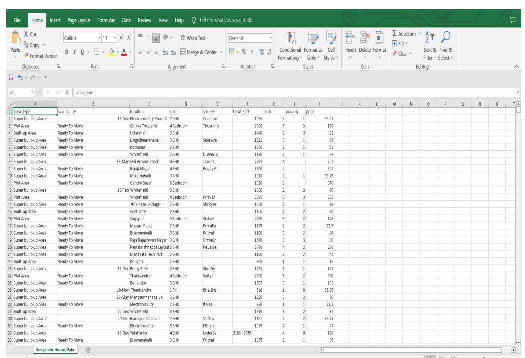


Fig 1: Dataset

The study is useful for individuals and organizations involved in the US housing market, particularly in the development of real estate policies and strategies for future housing price predictions.

[4] Talks about the application of machine learning in foreseeing house costs in a little town in the West Godavari locale of Andhra Pradesh. The proposed work employments principal machine learning calculations such as choice tree classification, choice tree relapse, and numerous straight relapse, which are executed utilizing the Scikit-Learn machine learning apparatus. The input highlights utilized in the demonstrate incorporate the number of rooms accessible in the house, the age of the house, transportation accessibility, adjacent schools, and adjacent shopping centres. The show makes a difference clients to foresee the accessibility of houses in the city and too to anticipate the costs of the houses. The article reports that the execution of different direct relapse is way better than that of choice tree relapse in foreseeing house costs. This may be due to the truth that different straight relapse is a more progressed calculation that can handle complex connections between the input highlights and the yield variable. The creators recommend that future work seem include planning a dataset with more

Vol. 21, No. 1, (2024)
ISSN: 1005-0930

highlights and utilizing more progressed machine-learning methods to build the house cost forecast demonstrate. This is a great proposal, as ore highlights may possibly make strides the exactness of the show, and progressed strategies like neural systems or slope boosting could move forward the execution indeed assist. By and large, this article highlights the value of machine learning in anticipating house costs and gives a great beginning point for assist investigate in this area.

[5] The inquire about paper points to foresee lodging costs based on different relapse strategies utilizing each essential parameter that is considered whereas deciding the cost. The proposed framework proposes a framework that makes ideal utilize of Direct Relapse, Woodland relapse, Boosted relapse, and Neural systems to give an precise forecast of lodging costs. The framework points to fulfil clients by giving exact yield and avoiding the hazard of contributing in the off-base house. Real-time neighbourhood points of interest utilizing Google maps can be utilized to get correct genuine- world valuations. The proposed framework can be overhauled with extra highlights for the customer's advantage without exasperating its centre usefulness. A future upgrade may be the expansion of bigger cities to the database to permit clients to investigate more houses, get more precision, and come to a legitimate choice. The proposed framework addresses the need of straightforwardness in the genuine domain industry and gives an arrangement for clients to make educated decisions.

PROPOSED SYSTEM

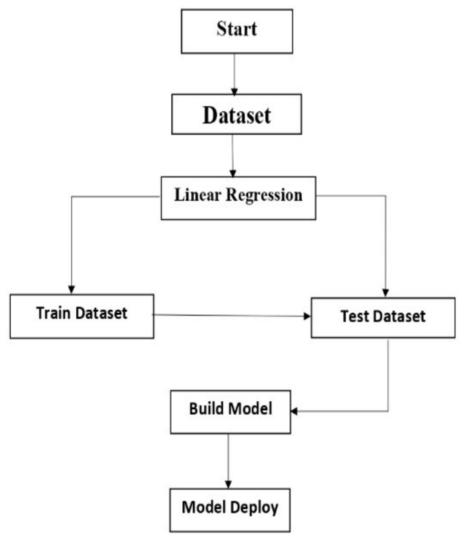
Predicting house prices using linear regression involves using historical data on house prices and related features such as the number of bedrooms, the size of the house, location, and age of the house, to create a predictive model. Here is a basic methodology for house price prediction using linear regression:

1. Data collection: Collect data on house prices and related parameters like as the place, size, number of bedrooms, age, etc. This data can be obtained from public databases, real estate websites, or other sources.
2. Data preparation: Clean and prepare the data for analysis. This may involve removing missing values, outliers, or irrelevant data.
3. Feature selection: Select relevant features that are most likely to impact the house price. For example, location, size, and age of the house are common features used in house price prediction.
4. Data splitting: seperate the data into training and testing sets. The training set is used to build the predictive model, and the testing set is used to evaluate the accuracy of the model. Model building: Build a linear regression model using the training data. The model will involve selecting appropriate variables and tuning the model hyperparameters.

Vol. 21, No. 1, (2024)
ISSN: 1005-0930

5. Model evaluation: Evaluate the performance of the model using the testing data. This may involve calculating metrics such as mean squared error, mean absolute error, and R-squared.
6. Model improvement: If the model's efficiency is not satisfactory, refine the model by adjusting the model hyperparameters or selecting additional features.
7. Deployment: Once the model is deemed satisfactory, it can be deployed to make predictions on new data.

Note that this is just a basic methodology, and there are many other techniques that can be used to improve the accuracy of the model, such as regularization, feature engineering, and ensemble methods.



Flow Chart

Fig 2: Flowchart

Dataset

The dataset used in the House Price Estimator proposed system would typically contain various features related to the real estates transactions, such as location, square footage, number of bedrooms and bathrooms, and other amenities. The dataset could be collected from various sources such as real estate websites, property listing services, and public real estate databases. In addition to being sufficiently vast and diverse to guarantee that the model can produce correct forecasts for a range of locations, property kinds, sizes, and amenities, the dataset should also be sufficiently large to give the model numerous instances from which to learn.

Vol. 21, No. 1, (2024)
ISSN: 1005-0930

The dataset should also be preprocessed and cleaned to remove any missing values, outliers, or irrelevant data.

Tools and Techniques

Techniques used for the House Price Estimator proposed system are:

1. Programming Languages: Python is used for libraries and frameworks.
2. Libraries: libraries in Python used for machine learning tasks, such as Scikit-learn, XGBoost, Pandas, NumPy, Matplotlib, and Seaborn.
3. Data Preprocessing Techniques: Data preprocessing techniques such as data cleaning, missing value imputation, feature scaling, and one- hot encoding are used to prepare the data for modeling.
4. Evaluation Metrics: Evaluation metrics such as mean squared error, and root mean squared error is used to assess the performance of the model.
5. Deployment Techniques: Techniques for
6. web development frameworks can be used to deploy the model in a real- world scenario.

PROPOSED METHODOLOGY

1. Defining a problem: The problem statement is to develop a machine learning model to predict the selling price of a house based on various features.
2. Collect and preprocess the data: The data for the proposed system will be collected from historical real estate transactions and pre-processed to clean the data, handle missing values, and perform feature engineering to extract and transform relevant features.
3. Divide the data into training and test sets: To assess the model's performance on unobserved data, the data will be divided into training and test sets.
4. Choose and implement the model: The two machine learning algorithms chosen for the proposed system are linear regression and XGBoost. Both algorithms will be implemented using the sci-kit- learn library in Python.
5. Train the model: The linear regression and XGBoost models will be trained on the training data using suitable training algorithms and hyperparameters.
6. Evaluate the model: The performance of the models will be evaluated on the test data using various metrics such as mean squared error, and mean absolute error.
7. Deploy the model: The final model can be deployed in a real-world scenario by creating a user interface and integrating the model with other systems.
8. Monitor and update the model: The model should be monitored regularly to ensure that it is performing well and updated periodically to incorporate new data or features.

Vol. 21, No. 1, (2024)
ISSN: 1005-0930

Overview of Algorithms used :

In the House Price Estimation, two machine learning algorithms are used : linear regression and XGBoost, to predict the selling price of a house based on various features. Here's how these algorithms work.

Linear Regression:

Linear regression, also known as logarithmic regression, is a statistical method that describes the relationship between an independent variable and a dependent variable or multiple independent variables by applying a linear regression equation to the underlying data.It tries to find the best-fit line that can predict the target variable (i.e. selling price) based on the input features (e.g. location, square footage, number of bedrooms, etc.).The linear regression algorithm assumes a linear relationship between the target variable and the input features, which means that the predicted value of the target variable is a linear relation of the input features. In this the Linear Regression class from the sci-kit-learn library in Python is used to implement linear regression. Once the model is trained, it can be used to make predictions on new data.

XGBoost:

XGBoost (Extreme Gradient Boosting) is a more complex and powerful algorithm that can handle complex relationships between functions. It is an ensemble learning algorithm that combines predictions from multiple decision trees to make a final prediction. XGBoost uses gradient boosting, which is a technique that iteratively adds decision trees to the model, each one trying to correct the errors of the previous tree. In this, the XGB Regressor class from the XGBoost library is used to implement XGBoost. Once the model is trained, it can be used to make predictions on new data.

RESULTS

The performance of the models is evaluated against experimental data using various metrics such as root mean square error, root mean square error, and R-squared. By comparing the performance of linear regression and XGBoost algorithms, we can determine which algorithm produces more accurate and reliable predictions and can be used to build an effective house price estimation model.

Fig 3: Linear Regression Graph

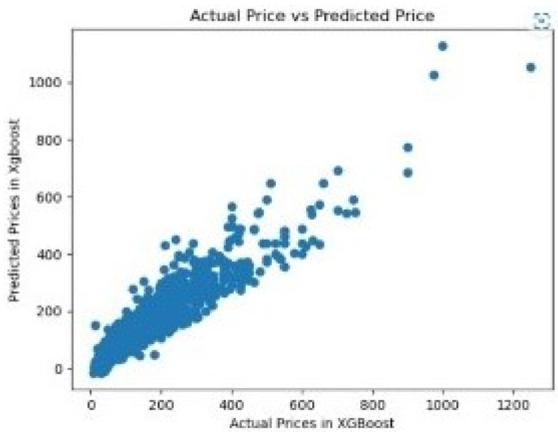


Fig 4: XGBoost Graph

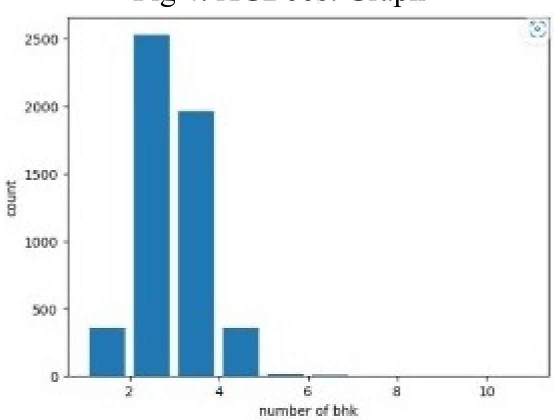


Fig 5: Room Bhk Graph

SCOPE OF THE PROPOSED SYSTEM

The House Price Estimator has a wide range of future scopes, some are - incorporating additional features like there are numerous other features, such as crime rate, school district, transportation, and economic indicators that can significantly impact the selling price of a house. Incorporating these features in the model can improve the accuracy of property price estimator. Utilizing different algorithms: In the proposed methodology, compared the performance of linear

Vol. 21, No. 1, (2024)
ISSN: 1005-0930

regression and XGBoost algorithms in predicting the selling price of a house. However, there are numerous other algorithms available for regression tasks, like decision trees, random forests, and neural networks. Building a recommendation system: The House Price Estimator can be used to build a recommendation system that can provide personalized recommendations to buyers based on their preferences and budget. The system can use machine learning algorithms to recommend houses that meet the buyer's criteria and are within their budget.

CONCLUSION

The House Price Estimation is a valuable application of machine learning in the real estate industry. By using algorithms such as linear regression and XGBoost, accurate prediction of selling price of a property based on several parameters becomes easy. This prediction can be useful for buyers, sellers, and real estate professionals to make informed decisions about pricing and selling houses. The results of the proposed system can be useful for various stakeholders in the real estate industry. Buyers can use the house price estimator to make informed decisions about purchasing a property, while sellers can use it to price their property correctly and sell it quickly. Real estate professionals can use the estimator to advise their clients on pricing and selling strategies. In conclusion, the House Price Estimation technique is a valuable application of machine learning in the real estate industry. By accurately predicting the selling price of a house, buyers, sellers, and real estate professionals can make informed decisions and improve their outcomes.

References

[1] Amena Begum, Nishad Jahan Kheya, Md. Zahidur Rahman, “Housing Price Prediction with Machine Learning,” International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075 (Online), Volume-11 Issue-3, January 2022,doi: http://dx.doi.org/10.35940/ijitee.C9741.0111322.

[2] CH.Raga Madhuri, Anuradha G,M.Vani Pujitha,”House Price Prediction Using Regression Techniques: A Comparative Study”, IEEE 6th International Conference on smart structures and systems ICSSS 2019,doi:10.1109/ICSSS.2019.8882834

[3] I. BoglaevByeonghwa Park a , Jae Kwon Bae, “Using machine learning algorithms for housing price prediction: The case of Fairfax County, Virginia housing data,” July 2021 Fundamenta Informaticae 1(1):11-23, doi:10.22995/scmi.2021.1.1.03

[4] Dr. M. Thamarai ,Dr. S P. Malarvizhi,” House Price Prediction Modeling Using Machine Learning”, I.J. Information Engineering and Electronic Business, 2020, 2, 15-20 Published Online April 2020 i n MECS (h t t p : / /www.mecs-press.org/),doi: 10 . 5815 / ijieeb.2020.02.03

[5] Ayush Varma, Abhijit Sarma, Sagar Doshi, Rohini Nair,” House Price Prediction Using Machine Learning And Neural Networks”,Published in: 2018 Second International

Conference on Inventive Communication and Computational Technologies (ICICCT), DOI: 10.1109/ICICCT.2018.8473231

[6] Quang Truong, Minh Nguyen, Hy Dang, Bo Mei.”Housing Price Prediction via Improved Machine Learning Techniques”, 2019 International Conference on Identification, Information and Knowledge in the Internet of Things (IIKI2019),doi: https:// doi.org/10.1016/j.procs.2020.06.111

[7] G. Naga Satish, Ch. V. Raghavendran, M.D.Sugnana Rao, Ch.Srinivasulu “House Price Prediction Using Machine Learning”, International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-8 Issue-9, July 2019,doi:10.17577/IJERTV10IS040322

[8] Nor Hamizah Zulkifley,Shah Alam, Shuzlina Abdul,Rahman Ismail Ibrahim,” House Price Prediction using a Machine Learning Model: A Survey of Literature”, I.J. Modern Education and Computer Science, 2020, 6, 46-54 Published Online December 2020 in MECS,doi: 10.5815/ijmecs.2020.06.04

[9] Satya Mohan Chowdary, A. Jaswanthi , P. Jaya Sri, S. Rajendranadh,N.Snehith,” Prediction of House Pricing Using Machine Learning with Python” , September 2021| IJIRT | Volume 8 Issue 4 | ISSN: 2349-6002,DOI: 10.1109/ICESC48915.2020.9155839

[10] Siddhant Burse , Dhriti Anjaria, Hrishikesh Balaji,” Housing Price Prediction Using Linear Regression”, 2021 JETIR October 2021, Volume8,Issue10

[11] J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[12] I. S. Jacobs and C. P. Bean, “Fine particles, thin films and exchange anisotropy,” in Magnetism, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.

[13] K. Elissa, “Title of paper if known,” unpublished.

[14] R. Nicole, “Title of paper with only first word capitalized,” J. Name Stand. Abbrev., in press.

[15] Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, “Electron spectroscopy studies on magneto-optical media and plastic substrate interface,” IEEE Transl. J. Magn. Japan, vol. 2, pp. 740–741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].

[16] M. Young, The Technical Writer’s Handbook. Mill Valley, CA: University Science, 1989.