IJFANS INTERNATIONAL JOURNAL OF FOOD AND NUTRITIONAL SCIENCES

ISSN PRINT 2319 1775 Online 2320 7876

Research paper Volume 12, Iss 1, 2023

House Price Estimation using Machine Learning Algorithms

K.Chandrasekhar, C.Ramakrishnaiah, K.Bala Chandra Reddy

Assistant Professor(Adhoc)CSE Department,JNTUACEP,chandra507shiva@gmail.com
Assistant Professor(Adhoc)CSE Department,JNTUA, c.ramakrishnaiah08@gmail.com
Assistant Professor(Adhoc)CSE Department,JNTUACEP,kakarla505@gmail.com

Abstract—*Theproject'sgoalistoassesshowwell,twowell-known machine learning algorithms, multiple linear regression, and XGBoost Regressor, estimate the price of a house. To implement this concept, a dataset of past real estate transactions was gathered, and key features such as the property's location, size, number of bedrooms and bathrooms, and additional amenities wereidentifiedforanalysis. Thedatawas pre-processed,cleaned, and transformed to prepare it for modeling. Using the same dataset, the Multiple Linear Regression and XGBoost Regressor models were developed and trained, and their performance was assessedonthebasis ofaccuracy,cross-validation,meansquared error, and root mean squared error. The ultimate goal of the project is to determine which algorithm predicts more accurate and reliable results.*

Keywords—Machine learning, model, linear Regression, XG Boost,Onehot-encoding,Flask,Python,HTML,CSS,Bootstrap, pickling.

I. INTRODUCTION

The real estate market is a dynamic and complex system, where various factors such as location, size, amenities, and otherfeaturesinfluencethesellingcostofahome.Accurately predicting the cost of a house is a critical task for buyers, sellers,andreal estateprofessionals, asit canhelp them take informed decisions about buying, selling, or renting a property. The performance of two popular algorithms, Multiple Linear Regression, and XGBoost Regressor is evaluatedinpredictingthepriceofthehouse.MultipleLinear Regression is a simple and widely used algorithm for regression tasks, while XGBoost Regressor is a more complex and powerful algorithm that can handle complex relationships between features. The dataset used for this project was compiled from records of past real estate deals and includes information on location, area, number of bedrooms, bathrooms, and other factors. The results helped to build an accurate and reliable system that assists buyers, sellers, and real estate professionals in making informed decisions.BycomparingtheperformancesofMultipleLinear Regression and XGBoost Regressor algorithms, made determine which algorithm produces more accurate and reliable predictions.

II. LITERATUREREVIEW

[1] The study emphasizes the significance of creating a reliablehomepriceforecastmodelforsocioeconomicgrowth and public welfare. The use of machine learning algorithms including linear regression, decision trees, and random forests to anticipate home prices is examined using the provided datasets. The StatLiblibrary, which ismanaged by Carnegie Mellon University, provided the 506 sample datasets and the 13 feature variables that were used in the study. The study indicates that a number of variables, including location, area, and the number of rooms, have a significant impact on housing costs. As a result, all of this data isnecessarytoforecast individualhomeprices. In order toexplore thediverse effects ofcharacteristics on prediction approaches,theresearchcomparesanumberofsophisticated models. By thoroughly testing a variety of methods used in model execution on regression, the study also offers a positive outcome for house price prediction. Overall, the paper provides valuable insights into the use of machine learningalgorithmsforhousepricepredictionandhighlights the importance of accurate prediction models for making well-informed decisions in the real estate industry.

[2] The article emphasizes the significance of precise house price forecasts for prospective buyers who are wary of their budgets and market approaches. Based on their financial plans and objectives, the paper's goal is to estimate coherent housing prices for non-homeowners. Using a dataset and regression techniques such multiple linear, ridge, LASSO, elastic net, gradient boosting, and ada boost regression, the study forecasts home price values. To predict speculated pricing, the study forecasts developments and evaluates historical product and fare ranges. The purpose of the document is to aid sellers in precisely evaluating a home's selling price as well as to assist readers in determining the precise amount of time needed to accumulate a home. Physicalconditions,concepts,andlocationarealsotakeninto account as related elements that affect cost. The study underlines the significance of precise prediction models for making knowledgeable judgments in the real estate market, andoverallitoffersinsightful informationontheapplication ofregressiontechniquesforhousepriceprediction.Both

ISSN PRINT 2319 1775 Online 2320 7876
Research paper Volume 12, Iss 1, 2023

buyers and sellers that are careful with their finances and market strategies will get benefit from the study.

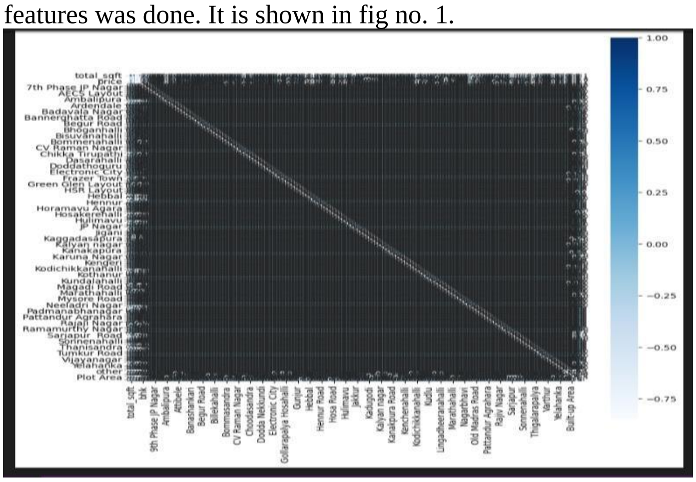
[3] In formulating real estate regulations and forecasting future house prices in the US housing market, the study emphasizes the importance of housing price prediction models. The study creates a housing price prediction model based on housing data for 5359 townhouses in Fairfax County, Virginia, collected by the Multiple Listing Service (MLS) of the Metropolitan Regional Information Systems. These machine learning algorithms include C4.5, RIPPER, Nave Bayesian, and AdaBoost (MRIS). The research evaluates the classification accuracy performance of several machinelearningalgorithmsandoffersanenhancedhousing pricepredictionmodelin ordertoassistahomeseller orreal estate agent in making judgements based on home price valuation. Based on the accuracy, the RIPPER algorithm routinelysurpassesthe other models when it comes tohouse price forecasts. Overall, the study highlights the importance of exact prediction models in assisting home sellers or real estate agents in making more informed judgements and provides valuable information about the application of machine learning algorithms for housing price prediction. PeopleandorganizationsactiveintheUShousingmarketcan benefitfromthestudy,especiallywhendevelopingrealestate regulations and methods for forecasting house prices in the future.[4] The paper investigates the application of machine learning to forecast house price values at a tiny town in Andhra Pradesh's West Godavari area. The suggested study makes use of basic machine learning methods implemented with the Scikit-Learn machine learning framework, such as decision tree classification, decision tree regression, and multiplelinearregression.Themodel'sinputfeaturesinclude the home's age, the number of bedrooms it has, the accessibility of transportation, the proximity of shopping centers and educational institutions, and the number of bedrooms it has. Users can forecast the cost and availability of homes in the city with the help of the model. This might be the more sophisticated multiple linear regression techniquethathandleintricateconnectionsbetweentheinput features and the output variable. The authors speculate that future studies may create a dataset with more variables and develop the house price prediction model using more advanced machine-learning techniques. This is a smart idea because adding more features could perhaps increase the model'saccuracy,andusingcutting-edgemethodslikeneural networksorgradientboostingcouldboostperformanceeven more. Overall, thisarticleemphasizeshowmachinelearning isusedtoanticipatehousingprices, anditoffersanexcellent place tostart for future studyin this field.[5] Thegoal ofthe project is to forecast house prices utilizing a variety of regression approaches and every fundamental factor that affects pricing The research recommends a method that utilizes linear regression, forest regression, boosted regression, andneural networks to offer a precise prediction of housing prices. The system seeks to please users by producingaccurateresultsandreducingthechanceofmaking the wrong investment in real estate. To obtain precise real-world assessments, Google Maps' real-time neighborhood information might be employed. Without affecting its fundamental functionality, the suggested system can be updated with new features for the advantage of the client. Morecitiescouldbeaddedtothedatabaseinalaterupdate toenableuserstolookatmorehomes,gaingreateraccuracy, andmaketherightchoice.Theinitiativeseeks toaddressthe lackofopennessintherealestatesectorandofferscustomers a way to make educated choices.

III. Methodology

To make predictions about house prices using Multiple Linear Regression and the XGBoost Regressor algorithm, a model was created using past data on house prices and relevant features such as location, property type, number of balconies and galleries, number of bedrooms, hall, and kitchen (BHK), and square footage. Here is a basic methodology for house price estimation:

1. Data collection: It involved the collection of different values for all the independent variables from public databases and repositories.
2. Data preparation: It involved removingoutliersand unnecessaryfeaturesfromthedatasetitinvolvedthe following steps:
   a. Removing null values from the dataset by dropping such columns from the dataset.
   b. Creatinganewcolumnforuniquelocations.
   c. Converting the values specified in the range to single-digit values.
   d. Removingoutliersforpricepersquarefeettoset a limit for the price by using standard deviation and mean:
      upper_limit = data.price_sqft.mean() + data.price_sqft.std()
      lower_limit = data.price_sqft.mean() - data.price_sqft.std()
   e. Removinga2BHKhousewhosepriceper sqftis less than the mean of 1bhk house price.
3. Feature selection: Here the selection of the following features was done which included:
   a. Location.
   b. Areatype.
   c. Numberofgalleries.
   d. Numberofbathrooms.
   e. Sizeinsquare feet.
   f. BHK.

Similarly, an understanding ofthecorrelation between these features was done. It is shown in fig no. 1.



Figno.1:Heatmapshowingacorrelationbetweenfeatures

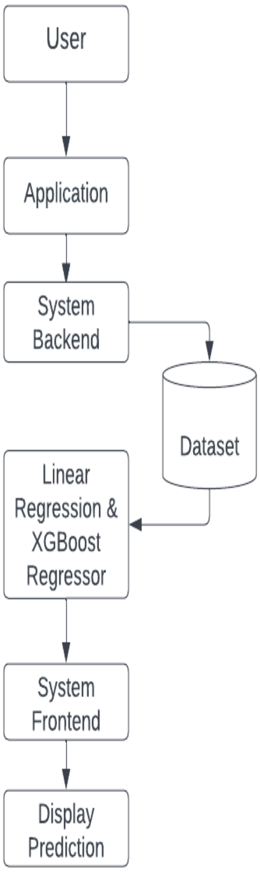
IJFANS INTERNATIONAL JOURNAL OF FOOD AND NUTRITIONAL SCIENCES

ISSN PRINT 2319 1775 Online 2320 7876

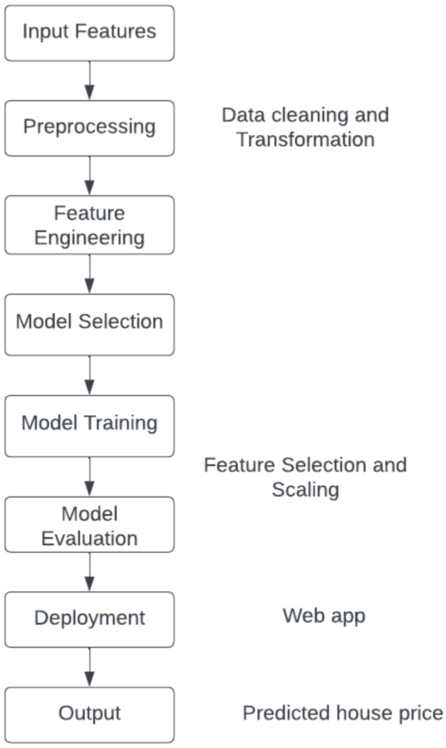
 Volume 12, Iss 1, 2023

4. Datasplitting:Thedatasetwasdividedintotwosets: 80% for training the model and 20% for testing purposes.
5. Model building: Herethebuiltmodel was stored in a .pkl file using the pickling technique byusing the dump()functiontoincreasetheefficiencyandspeed of predictions as training the model for each prediction is not a feasible process.
6. Model evaluation: To evaluate the model, metrics such as mean squared error, mean absolute error, accuracy, and cross-validation were utilized.
7. Model improvement: Here the data augmentation was carried out by adding additional rows to the dataset. Similarly, the one-hot encoding technique wasappliedtotheareatypeandlocationcolumnsin the dataset for model's performance improvement, by converting the categorical data into numerical data. This was done by using the getDummies().
8. Deployment: The model was deployed using the flask framework. Here the graphical user interface was created using HTML and the bootstrap framework of CSS.

i) Flowchartofmodels:



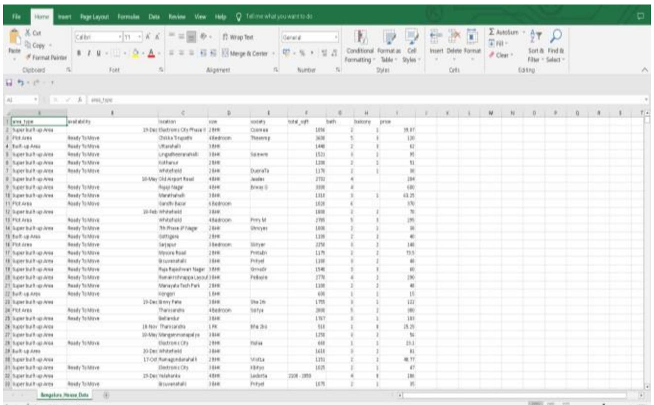
Figno2:Flowchartofsystem



Figno3:Flowchartofmodel

ii) Dataset

The dataset used for building the system comprised features related to real estate transactions, such as the property'slocation,squarefootage,numberofbedrooms andbathrooms,andadditionalamenities.Thetotal datasetincluded13321rows.Figno. 4showsthedataset used.



Figno4:Dataset

iii) AlgorithmsusedintheProject:

In this project, two machine learning algorithms, Multiple linearregression,andXGBoostRegressor wereemployedto forecastthesellingpriceofapropertyusingvariousfeatures. Thealgorithmsutilizedinthisproject to thebuilt project are as follows:

1) MultipleLinearRegression:
   - Multiple Linear regression is a simple and widely used algorithm for regression tasks.

- By utilizing the input features, the algorithms determined the best-fit line that could predict the targetvariable,whichinthiscaseisthesellingprice of the property.
- To implement Multiple Linear Regression in this project, the Regression class from the sci-kit-learn along with sklearn python libraries were used.
- Formulatocalculatetheprediction:

$$y = \beta0 + \beta1x1 + \beta2x2 + \beta3x3+..+ \beta nxn$$

2) XGBoostRegressor:

- A more sophisticated and potent algorithm that can handle intricate interactions between features is XGBoost (Extreme Gradient Boosting).
- It is an ensemble learning algorithm that combines severaldecisiontrees'predictionstoproduceasingle outcome.
- Gradient boosting, a technique used by XGBoost, it iteratively adds decision trees to the model while each tree attempts to fix the flaws of the prior tree.
- To implement XGBoost Regressor in project, XGBRegressor class from the XGBoost library was used.
- FormulaofXGBoostRegressorusedforcalculations is as follows:

$$\hat{y} = \sum_{k=1}^{K} f_k(x)$$

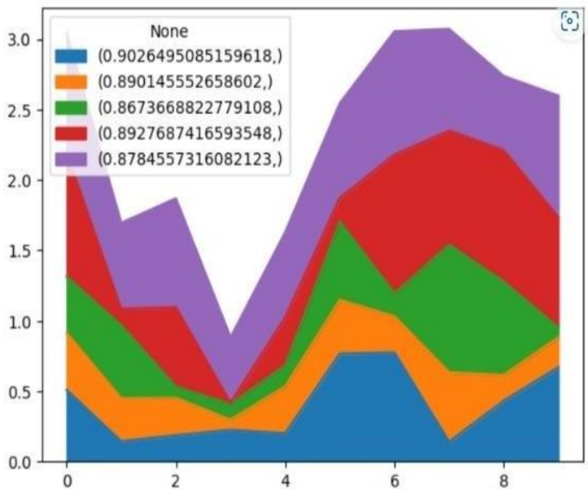
iv) MultipleLinearRegressionvsXGBoostRegressor

	MultipleLinear Regression	XGBoost Regressor
1.Accuracy	88.62%	89.76%
2.RSME	0.8	0.5
3.MAE	0.5	0.3

TableI.ComparisonOfAlgorithms

After evaluation of the accuracy, mean squared error, and mean absolute error, it was determined that the XGBoost Regressor model performed better than the Multiple Linear Regression model.Thefigno. 9showsthatineverycasethe XGBoost Regressor worked better.

IV. Results

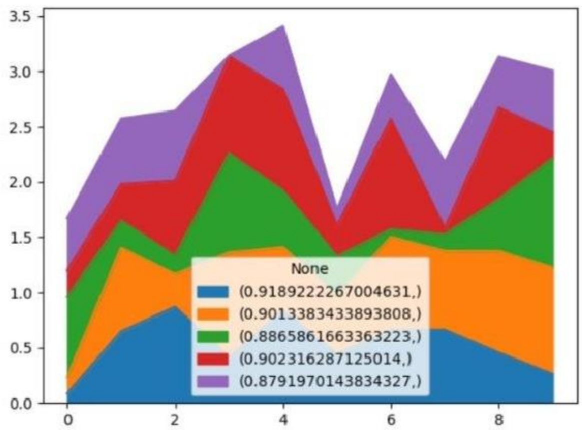


Figno.5:PerformanceofLinearRegression

Theresultsofmodelevaluationusingcrossvalidationfor Linear Regression are shown in figure no. 5.

Figno.6Actual pricesvspredictedpricesin multiplelinear regression

Thefigureno.6showstheactualpricevs thepredictedprice in Multiple Linear Regression.



Figno.7PerformanceofXGBoostRegressor

TheresultsoftheXGBoostRegressormodelevaluationusing cross-validation is shown in fig no.7.

Figno.8ActualpricesvspredictedpricesinXGBoost regressor

Theactualpricevsthepredictedpriceisshown infigure no.8

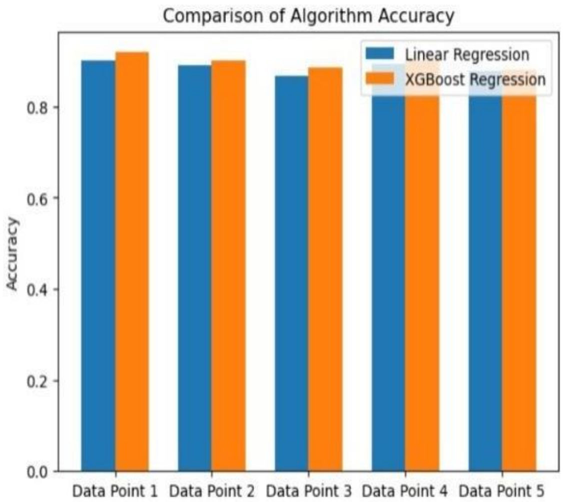


Fig9:AccuracyofMultipleLinearRegressionvsXGBoost Regressor

ThecomparisonbetweentheaccuracyofMultipleLinear Regression and XGBoost Regressor is shown in fig no.9.

V. SCOPEOFTHEPROJECT

Theprojecthasawiderangeoffuturescopebyincorporating additional features like crime rate, school district, transportation,andeconomicindicatorsthatcansignificantly impact the selling price of a house. Incorporating these featurescanimprovetheaccuracyofthemodels.Thissystem comparedtheperformanceofMultipleLinearregressionand XGBoostRegressoralgorithmsinpredictingthesellingprice of a house. However, there are numerous other algorithms availablefor regressiontasks,suchasdecisiontrees,random forests, and neural networks which can be used. Lastly integration with real estate agents for building a complete systemwhereuserscancheckthepriceofthehouseandalso contacttheagentfor furthertransactionsbyusingthesystem can also be implemented.

VI. CONCLUSION

This project serves as a significant example of applying machinelearningtechniquesintherealestatesector.Through the use of algorithms like Multiple Linear Regression and XGBoost Regressor, we can make precise predictions about the selling price of a property based on a range of relevant features.Thispredictionisusefulfor buyers,sellers,andreal estateprofessionalstomakeinformeddecisionsaboutpricing and selling houses. The results of this project are useful for various stakeholders in the real estate industry. Buyers can usethisProjecttomakeinformeddecisionsaboutpurchasing a property, while sellers can use it to price their property correctlyandsellitquickly.Realestateprofessionalscanuse the estimator to advise their clients on pricing and selling strategies. In conclusion, this project is a worthyapplication of machine learning in the real estate.

REFERENCES

[1] Amena Begum, Nishad Jahan Kheya, and Md. Zahidur Rahman,"Housing Price Prediction using Machine Learning," InternationalJournal of Innovative Technology and Exploring Engineering(IJITEE), Volume 11 Issue 3, January 2022, doi:10.35940/ijitee.C9741.0111322.

[2] CH.Raga Madhuri, Anuradha G, and M.Vani Pujitha,"House PricePrediction Using Regression Techniques: A Comparative Analysis,"IEEE 6th International Conference on Smart Structures and Systems,2019 ICSSS, doi:10.1109/ICSSS.2019.8882834

[3] Jae Kwon Bae and I. BoglaevByeonghwa Park, "Using machinelearning algorithms for home price prediction: The case of FairfaxCounty,Virginiahousingdata,"FundamentaInformaticae1(1):11 –23,July 2021, doi:10.22995/scmi.2021.1.1.03.

[4] Dr.M.Thamarai,Dr. S P. Malarvizhi,"HousePrice Prediction ModelsUsingMachineLearning",I.J.InformationEngineeringandElectronicBusiness, 2020, 2, 15-20, published online April 2020 in MECS(http://www.mecs-press.org/), doi: 10.5815/ijieeb.2020.02.03

[5] AyushVarma, AbhijitSarma,SagarDoshi,RohiniNair,"HousePricePredictionUsingMachineLearningAndNeuralNetworks",Publishedin: 2018 Second International Conference on InventiveCommunicationandComputationalTechnologies(ICICCT), DOI:10.1109/ICICCT.2018.8473231

[6] Quang Truong, Minh Nguyen, Hy Dang, Bo Mei."Housing PricePrediction via Improved Machine Learning Techniques", 2019International Conference on Identification, Information andKnowledgein 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 ExploringEngineering (IJITEE) ISSN: 2278-3075, Volume-8 Issue-9, July2019,doi:10.17577/IJERTV10IS040322

[8] Nor Hamizah Zulkifley,Shah Alam, Shuzlina Abdul,Rahman IsmailIbrahim,"HousePricePredictionusinga MachineLearningModel: ASurvey 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,"PredictionofHousePricingUsingMachine 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 PricePrediction Using Linear Regression", 2021 JETIR October 2021,Volume8,Issue10