iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** V **Month of publication:** May 2022

DOI: https://doi.org/10.22214/ijraset.2022.43190

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue V May 2022- Available at www.ijraset.com

House Price Prediction Using Machine Learning

Kumari Sandhya[1], Sarwar Siddiqui[2]
[1, 2]CSE Dept, SRMCEM Lucknow

Abstract: Real estate is the least transparent industry in our ecosystem. Housing prices keep changing day in and day out and sometimes are hyped rather than being based on valuation. Predicting housing prices with real factors is the main crux of our research project. This paper outlines how to predict housing costs using various regression techniques using the Python library. The proposed method takes into account the sophisticated aspects used in the house price calculation and provides a more accurate forecast. This paper uses machine learning to explain how the house price model works and which datasets are used in the proposed model. Predictive models to determine the selling prices of homes in cities like Bangalore are maintained because of more than difficult and sophisticated tasks. The price of real estate for sale in cities like Bangalore depends on several factors involved.

Keywords: Housing Price Prediction , linear regression, XGBoost,, Machine Learning, Lasso Regression, Ridge Regression, Random Forest Regression.

I. INTRODUCTION

House is one of the basic needs for a person and their prices vary from place to place depending on available amenities like parking space, locality etc. Buying a home is one of the biggest and most important choices for a family as they put all of their funds into investments and cover them over time with loans. Modeling using machine learning algorithms, when machines learn data and use it provide new data. Like all of us namely the proposed model for accurate forecasting future results have economic applications, banking, business, e-commerce, healthcare industry, entertainment etc. In metropolitan cities like Bangalore, potential buyers consider a number of factors such as location, land size, distance from parks, schools, hospitals, power generation facilities and especially the price of the house. As we see suitable, we can use the model to predict monetary value of this particular residence property in Bangalore.

A. Objective

This project is proposed to predict house prices and get more accurate and better results. This number will be of great help to everyone because house price is a topic that many people care about, whether they are rich or middle class, because one can never judge price or estimate the price of a home based on local or properties available. To perform this task python programming language is used.

B. Machine Learning

Machine learning is an area of artificial intelligence that allows PC frameworks to learn and improve their execution using information. Machine learning is an area of artificial intelligence that allows PC frameworks to learn and improve their execution using information.

C. Python

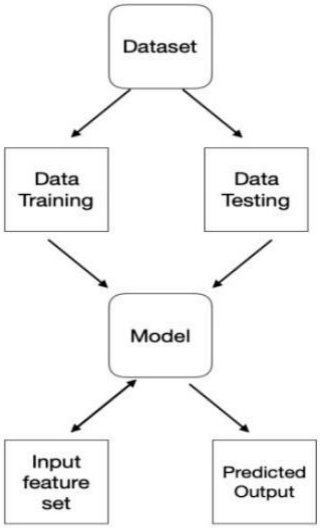
Python is a high-level programming language for general purpose programming. It allows explicit programming on large and small scales. It is an easy to read language. Python supports various libraries such as Pandas, NumPy, SciPy, Matplotlib etc. Python is a particularly useful language for improving the web and advancing programming.

II. LITERATURE REVIEW

There are a number of things that affect home prices. In this exploration, divide these components into the three essential sets, which have state of being, thought , and territory. States of being are homes confined through a residence that may be visible through human recognizes, which includes the variety of the residence, the quantity of rooms, the availability of kitchen and parking space, the openness of the backyard nursery, the sector of land and structures, and the age of the residence, whilst the concept is an concept supplied through architects who can pull in capacity, for instance, the opportunity of a mild home, robust and inexperienced condition, and global magnificence condition.

©IJRASET: All Rights are Reserved

III. METHODOLOGY



A. Data Description

Each record in the database describes a Bangalore suburb or town. The data was drawn from the Bangalore Standard Metropolitan Statistical Area (SMSA) in 2017. The attributes are defined as follows :-

1) Area_type : Area of the Property in which they exist
2) Availability property's status as in 'ready to move' or still under construction.
3) Location : name of locality
4) Size : No. of Bedrooms along with 1 Hall and 1 kitchen.
5) Society : Name of the society
6) total_sqft : Area of the Property in square feet
7) Bath : No. of bathrooms
8) Price : price in Inr

B. Data Collection

Data collection is the systematic process of gathering information about variables. It helps to find answers to questions, hypothesizes too much and evaluates results. Collecting data as the pathway to social events and estimating data on targeted factors within the built-in framework, at this stage allowing related queries to be answered and evaluate the results.

C. Data Visualization

Data Visualization is the pictorial or graphical representation of information. It enables to grasp difficult concepts or identify new patterns. This includes creating and investigating visual representations of information.

D. Data pre-processing

This is the process of transforming the data before it is fed to the algorithm. It is used to convert raw information into a clean data set. This is a information mining strategy that involves transferring raw information into a logical organization Enter raw information in a logical organization. The result of preprocessing data is the final dataset used for preparation and reason for testing.

E. Data Cleaning

Data cleaning is the process of detecting and removing errors to increase the value of data. Data cleaning is performed using data processing tools. That's way to identify and modify off-base records from a record set, table or database. It finds the missing information and replaces the scrambled information. information is edited to ensure it is correct and accurate.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

IV. REGRESSION MODELS AND EVALUATION METRICS USED

Linear regression is one of the most popular algorithms in statistics and machine learning. The goal of a linear regression model is to find the relationship between one or more characteristic variables.

A. Basic Linear Model

The formulation for multiple regression model is :

$$Y = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \alpha_3 X_3 + \ldots\ldots\ldots + \alpha_p X_p$$

The assumptions in the model are

1) x The error terms are normally distributed.
2) The error terms have constant variance.
3) The model carries out a linear relationship between the target variable and the functions.

B. Ridge Regression

Ridge regression is a method of estimating the coefficients of multiple-regression models in scenarios where linearly independent variables are highly correlated. It has been used in many fields including econometrics, chemistry, and engineering. In mathematical form , the model can be expressed as y=xb+e. Ridge regression is a model tuning method that is used to analyze any data that suffers from multicollinearity.

C. Lasso Regression

Lasso regression is a regularization technique. It is used over regression methods for a more accurate prediction. This model uses shrinkage. Shrinkage is where data values are shrunk towards a central point as the mean. The lasso procedure encourages simple, sparse models. The word “LASSO” stands for **L**east **A**bsolute **S**hrinkage and **S**election **O**perator. It is a statistical formula for the regularisation of data models and feature selection.

D. SVR (Support Vector Regression)

SVMs or Support Vector Machines are one of the most popular and widely used algorithm for dealing with classification problems in machine learning.
In machine learning, Support Vector Machines are supervised learning models with associated learning algorithms that analyze data used for classification and regression analysis. In Support Vector Regression, the straight line that is required to fit the data is referred to as **hyperplane**.

E. XGBoost Regression Model

XGBoost stands for Extreme Gradient Boosting and is the most efficient method for regression or classification problems. This is a decision tree based algorithm using the gradient boosting framework. This provides the ability for the to have a significant impact on the performance of the model.

F. Random Forest Regression

Random Forest Regression is a supervised learning algorithm that uses ensemble learning method for regression. Ensemble learning method is a technique that combines predictions from multiple machine learning algorithms to make a more accurate prediction than a single model.

V. CONCLUSION

Home selling prices are calculated using various algorithm. Selling price calculated with better accuracy and accuracy than. This will be of great help to people. Various factors that affect home prices need to be considered and addressed.

VI. FUTURE WORK

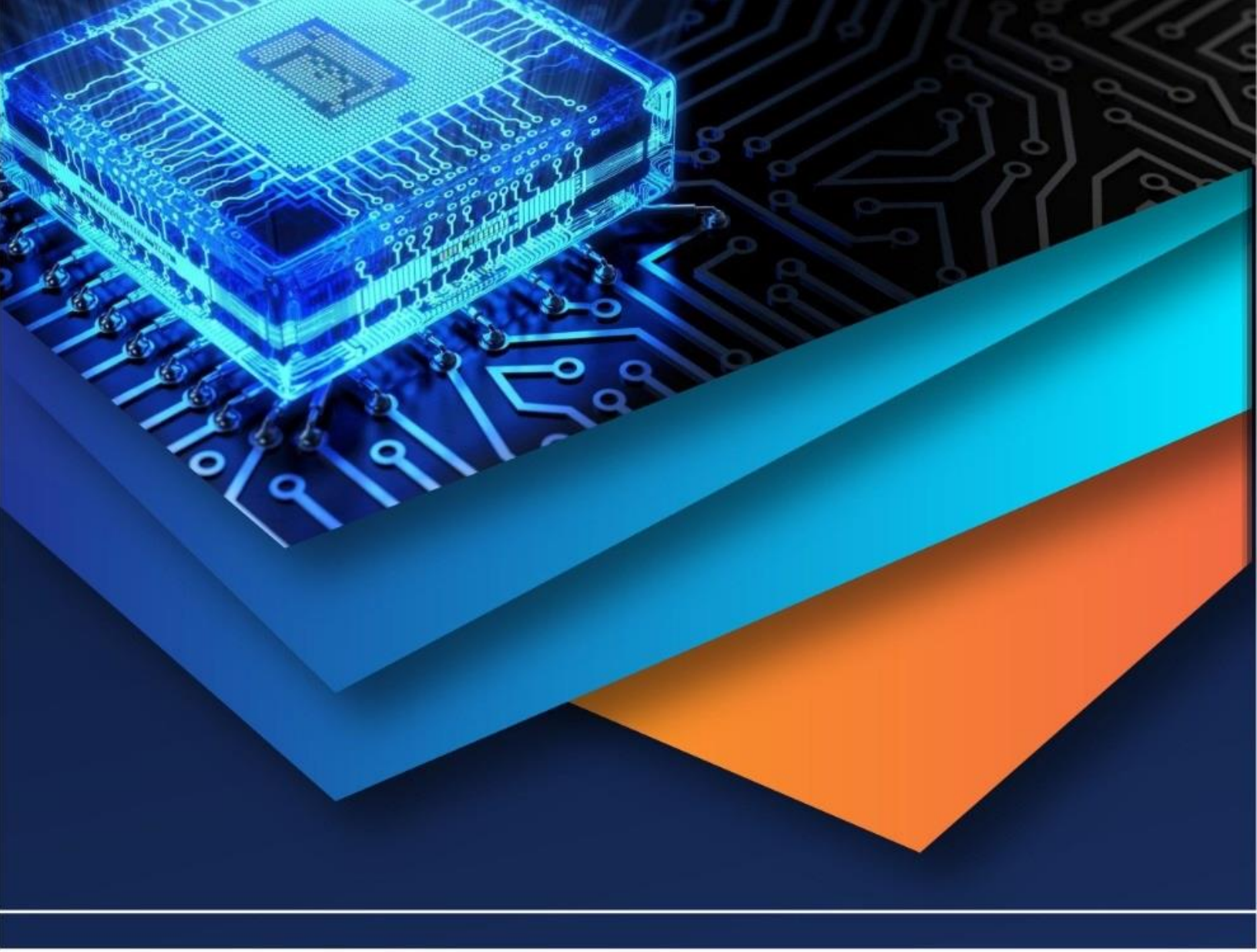
The accuracy of the gadget may be improved. Several extra cites may be protected withinside the gadget if the scale and computational strength will increase of the gadget. In addition, we can integrate different UI/UX methods to better visualize the results in in a more interactive way using Augmented Reality.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

REFERENCES

[1] H.L. Harter,Method of Least Squares and some alternatives-Part II.International Static Review.1972,43(2) ,pp. 125-190.
[2] J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68-73.
[3] Lu. Sifei et al,A hybrid regression technique for house prices prediction. In proceedings of IEEE conference on Industrial Engineering and Engineering Management: 2017
[4] R. Victor,Machine learning project:Predicting Boston house prices with regression in towards datascience.
[5] S. Neelam,G. Kiran,Valuation of house prices using predictive techniques, Internal Journal of Advances in Electronics and Computer Sciences:2018,vol 5,issue-6
[6] S. Abhishek.:Ridge regression vs Lasso,How these two popular ML Regression techniques work. Analytics India magazine,2018.
[7] S.Raheel.Choosing the right encoding method-Label vs One hot encoder. Towards datascience,2018
[8] Raj, J. S., & Ananthi, J. V. (2019). Recurrent Neural Networks and Nonlinear Prediction in Support Vector Machines. Journal of Soft Computing Paradigm (JSCP), 1(01), 33-40.
[9] Predictinghousepricesin Bengaluru(Machine Hackathon) https://www.machinehack.com/course/predicting -house-prices-inbengaluru/
[10] Raj, J. S., & Ananthi, J. V. (2019). Recurrent neural networks and nonlinear prediction in support vector machines. Journal of Soft Computing Paradigm (JSCP), 1(01), 33-40.
[11] Pow, Nissan, Emil Janulewicz, and L. Liu (2014). Applied Machine Learning Project 4 Prediction of real estate property prices in Montréal.
[12] Wu, Jiao Yang(2017). Housing Price prediction Using Support Vector Regression.
[13] Limsombunchai, Visit. 2004.House price prediction: hedonic price model vs. artificial neural network.New Zealand Agricultural and Resource Economics Society Conference.
[14] Rochard J. Cebula (2009).The Hedonic Pricing Model Applied to the Housing Market of the City of Savannah and Its Savannah Historic Landmark District; The Review of Regional Studies 39.1 (2009), pp. 9– 22.
[15] Gu Jirong, Zhu Mingcang, and Jiang Liuguangyan. (2011).Housing price based on genetic algorithm and support vector machine". In: Expert Systems with Applications 38 pp. 3383–3386.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)