International Journal of Scientific Research in Science, Engineering and Technology

Print ISSN - 2395-1990
Online ISSN : 2394-4099

OPEN ACCESS

Available Online at : www.ijsrset.com
doi : https://doi.org/10.32628/IJSRSET

Enhancing House Price Prediction and Accuracy using Particle Swarm Optimization

[1] M. Madhusudhana Rao, [2] B. Bhanu Hanumantha Rao, [3] M. Yamuna, [4] G. Revanth Mehatha, [5] G. Bhavya Nagasri

[1]Assistant Professor, [2,3,4,5] UG Student

Department of CSE, Sri Vasavi Institute of Engineering & Technology, Nandamuru, Andhra Pradesh, India

ARTICLEINFO

Article History:

Accepted: 15 April 2024
Published: 25 April 2024

Publication Issue :
Volume 11, Issue 2
March-April-2024
Page Number :
434-442

ABSTRACT

Predictive models for determining the sale price of houses in cities like Bengaluru is still remaining as more challenging and tricky task. The sale price of properties in cities like Bengaluru depends on a number of interdependent factors. Key factors that might affect the price include area of the property, location of the property and its amenities. In this research work, an analytical study has been carried out by considering the data set that remains open to the public by illustrating the available housing properties in machine hackathon platform. The data set has nine features. In this study, an attempt has been made to construct a predictive model for evaluating the price based on the factors that affect the price. Modeling explorations apply some regression techniques such as multiple linear regression (Least Squares), Lasso and Ridge regression models, support vector regression, and boosting algorithms such as Extreme Gradient Boost Regression (XG Boost). Such models are used to build a predictive model, and to pick the best performing model by performing a comparative analysis on the predictive errors obtained between these models. Here, the attempt is to construct a predictive model for evaluating the price based on factors that affects the price.

Keywords : House Price, Lasso Regression, Ridge Regression, Regression Methods

I. INTRODUCTION

The housing market is one of the most aggressive as far as estimating and same will in general shift essentially dependent on various elements; determining property cost is a significant module in decision making for both the purchasers and financial backers in supporting financial plan allotment, observing property finding tricks and deciding reasonable approaches subsequently it becomes one of the great fields to apply the ideas of AI to advance and foresee the costs with high precision. Along these lines, in this paper, we present different significant highlights to utilize while anticipating lodging costs with great exactness. We can

Copyright © 2024 The Author(s): This is an open-access article distributed under the terms of the Creative Commons Attribution **4.0 International License (CC BY-NC 4.0)**

utilize relapse models, utilizing different elements to have lower Residual Sum of Squares. While utilizing highlights in a relapse model some element designing is needed for better expectation. In a study by (Durganjali and Vani Pujitha 2019) introduced a model which has accuracy of 70%. The goal of the paper by (Bhagat, Mohokar, and Mane 2016) is to predict the efficient house pricing for real estate customers with respect to their budgets and priorities. By analyzing previous market trends and price ranges, and also upcoming developments future prices will be predicted. Advanced machine learning algorithms are demonstrated by (B and Swathi 2019) can achieve very accurate prediction of property prices, as evaluated by the performance metrics. In an article by (Azimlu, Rahnamayan, and Makrehchi 2021) A House price Valuation based on Random Forest Approach. The Mass appraisal of residential property south korea Jengei HONG. Predicting house prices is expected to help people who plan to buy a house so they can know the price range in the future, then they can plan their finances well. In addition, house price predictions are also beneficial for property investors to know the trend of housing prices in a certain location. Predicting house prices is expected to help people who plan to buy a house so they can know the price range in the future, then they can plan their finances well. In addition, house price predictions are also beneficial for property investors to know the trend of housing prices in a certain location. Application of Predicting House Prices will help people to know the price range of the house in prior based on location, area type, square feet and other factors.

There are about 25 articles in IEEE xplore and in 30 Scopus related to this study. In a study by (Sangani, Erickson, and Al Hasan 2017) From investment to buying a house for residence, a person investing in the housing market is interested in the potential gain. This paper presents machine learning algorithms to develop intelligent regressions models for House price prediction. The main focus of the project by (Kadu and Bamnote 2021) is to forecast house prices using real factors intended to base our assessment on each of the basic criteria i.e. which is taken into account when setting prices. The goal of this project is to learn Python and gain experience in Data Analytics,

Machine Learning, and AI. The aim of the study by (Andrle and Plašil 2019) Using the borrowingcapacity and net-present-value techniques, it evaluates housing prices in 11 Canadian Census Metropolitan Areas (CMAs). The purpose of the paper by (Priya and Gayathri Priya 2021) is to assist the seller in accurately estimating the selling price of a house. Physical circumstances, and location, among other things, were all taken into account while determining the cost. This paper by (C. Zhou 2021) House price prediction can be done by using multiple prediction models (Machine Learning Model) such as support vector regression, artificial neural network, and more.

Modeling uses machine learning algorithms, where machine learns from the data and uses them to predict a new data. The most frequently used model for predictive analysis is regression. As we know, the proposed model for accurately predicting future outcomes has applications in economics, business, banking sector, healthcare industry, e-commerce, entertainment, sports etc. One such method used to forecast house prices are based on multiple factors. In metropolitan cities like Bengaluru, the prospective home buyer considers several factors such as location, size of the land, proximity to parks, schools, hospitals, power generation facilities, and most importantly the house price. Multiple linear regression is one of the statistical techniques for assessing the relationship between the (dependent) target variable and several independent variables. Regression techniques are widely used to build a model based on several factors to predict price. In this study, we have made an attempt to build house price prediction regression model for data set that remains accessible to the public in Machine hackathon platform. We have considered five prediction models, they are ordinary least squares model, Lasso and Ridge regression models, SVR model, and XGBoost regression model. A comparative study

was carried out with evaluation metrics as well. Once we get a good fit, we can use the model to forecast monetary value of that particular housing property in Bengaluru.

In metropolitan cities like Bengaluru, the prospective home buyer considers several factors such as location, size of the land, proximity to parks, schools, hospitals, power generation facilities, and most importantly the house price. Multiple linear regression is one of the statistical techniques for assessing the relationship between the (dependent) target variable and several independent variables. Regression techniques are widely used to build a model based on several factors to predict price. Hence it will be useful for buyer to predict the price of the house so buyer can search the houses according to his budget.

Predictive models for determining the sale price of houses in cities like Bengaluru is still remaining as more challenging and tricky task. The sale price of properties in cities like Bengaluru depends on a number of interdependent factors. Key factors that might affect the price include area of the property, location of the property and its amenities. The data set has nine features. In this study, an attempt has been made to construct a predictive model for evaluating the price based on the factors that affect the price.

II.RELATED WORK

House Price Index (HPI) is commonly used to estimate the changes in housing price. Since housing price is strongly correlated to other factors such as location, area, population, it requires other information apart from HPI to predict individual housing price. There has been a considerably large number of papers adopting traditional machine learning approaches to predict housing prices accurately, but they rarely concern about the performance of individual models and neglect the less popular yet complex models. As a result, to explore various impacts of features on prediction methods, this paper will apply both traditional and advanced machine learning approaches to investigate the difference among several advanced models. This paper will also comprehensively validate multiple techniques in model implementation on regression and provide an optimistic result for housing price prediction.

The real estate market is one of the fields where machine learning can be applied to optimize and predict the price with high accuracy. Determining housing price is vital model for decision making for customers in which number of parameters can be considered to predict price of desired house. The participants that are involved in the process are not aware of the various analytical techniques available to guess the property price considering various features relating to surroundings, environment and other amenities etc. The design will help users to invest in a property without approaching an agent. It also decreases the risk involved in the transaction. Use of lasso regression is done as model because of its convertible and probability methodology on model selection. The result displays that the approach of the issue needs to be successful, and has the ability to operate predictions that would be comparative with other house price prediction models.

In this machine learning paper, we analyzed the real estate property prices in Montreal. The information on the real ´ estate listings was extracted from Centris.ca and duProprio.com. We predicted both asking and sold prices of real estate properties based on features such as geographical location, living area, and number of rooms, etc. Additional geographical features such as the nearest police station and fire station were extracted from the Montreal Open Data Portal. We used and compared ´ regression methods such as linear regression, Support Vector Regression (SVR), k-Nearest Neighbours (kNN), and Regression Tree/Random Forest Regression. We predicted the asking price with an error of 0.0985 using an ensemble of kNN and Random Forest algorithms. In addition, where applicable, the final price sold was also predicted with an error of 0.023 using the Random Forest Regression. We will present the details of the

prediction questions, the analysis of the real estate listings, and the testing and validation results for the different algorithms in this paper. In addition, we will also discuss the significances of our approach and methodology.

III.PROPOSED SYSTEM

We are going to construct a predictive model for evaluating the price based on the factors that affect the price. Modeling explorations apply some regression techniques such as multiple linear regression (Least Squares), Lasso and Ridge regression models, support vector regression, and boosting algorithms such as Extreme Gradient Boost Regression (XG Boost). Such models are used to build a predictive model, and to pick the best performing model by performing a comparative analysis on the predictive errors obtained between these models. Here, the attempt is to construct a predictive model for evaluating the price based on factors that affects the price.

Advantages:

- Proposed system is totally focused on predicting house in Bangalore city.
- Multiple machine learning algorithms is used for predicting the house prices in different locations.

3.1 MODULES DESCRIPTION

Supervised Classification (Training Dataset):

The data has been divided into two parts i.e., training and testing data in the 70:30 ratios. Learning algorithms have been applied on the training data and based on the learning, predictions are made on the test data set.

Supervised Classification (Test Dataset):

The test dataset is 30% of the total data. Supervised learning algorithms have been applied on the test data and the output obtained is compared with the actual output.

Pandas: Pandas is an open source, BSD-licensed library providing high-performance, easy-to-use data structures and data analysis tools for the Python programming language.

Numpy: NumPy is a general-purpose array-processing package. It provides a high-performance multidimensional array object, and tools for working with these arrays. It is the fundamental package for scientific computing with Python.

MatPlotLib: matplotlib.Pyplot is a plotting library used for 2D graphics in python programming language. It can be used in python scripts, shell, web application servers and other graphical user interface toolkits

Scikit-learn: Scikit-learn is a free machine learning library for Python. It features various algorithms like support vector machine, random forests, and k-neighbors, and it also supports Python numerical and scientific libraries like NumPy and SciPy.

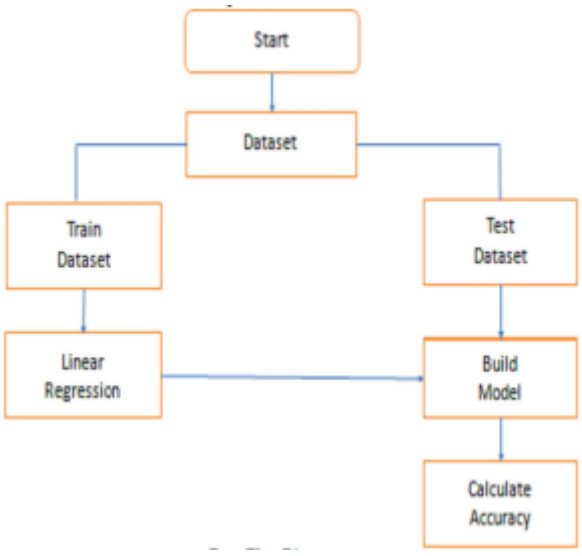

Fig 1: Block Diagram of the proposed system

III. RESULTS AND DISCUSSION

In [1]:
```
import numpy as np
import pandas as pd
from matplotlib import pyplot as plt
%matplotlib inline
import matplotlib
matplotlib.rcParams["figure.figsize"] = (20,10)
import seaborn as sns
from sklearn import preprocessing
from sklearn import model_selection
import sklearn
import xgboost
```

```
Matplotlib is building the font cache; this may take a moment.
```

In [2]:
```
home = pd.read_csv("Bengaluru_House_Data.csv")
home.head()
```

Out[2]:

	area_type	availability	location	size	society	total_sqft	bath	balcony	price
0	Super built-up Area	19-Dec	Electronic City Phase II	2 BHK	Coomee	1056	2.0	1.0	39.07
1	Plot Area	Ready To Move	Chikka Tirupathi	4 Bedroom	Theanmp	2600	5.0	3.0	120.00
2	Built-up Area	Ready To Move	Uttarahalli	3 BHK	NaN	1440	2.0	3.0	62.00
3	Super built-up Area	Ready To Move	Lingadheeranahalli	3 BHK	Soiewre	1521	3.0	1.0	95.00
4	Super built-up Area	Ready To Move	Kothanur	2 BHK	NaN	1200	2.0	1.0	51.00

Fig 2. Results Screenshot

Data Preprocessing!

In [4]:
```
## finding null values in %form

round(100*(home.isnull().sum()/len(home.index)),2)
```

Out[4]:
```
area_type        0.00
availability     0.00
location         0.01
size             0.12
society         41.31
total_sqft       0.00
bath             0.55
balcony          4.57
price            0.00
dtype: float64
```

In [5]:
```
#removing NaN values from the dataset
home.dropna(inplace =True)
```

In [6]:
```
home = home.drop(columns='society')
```

In [7]:
```
home.reset_index(drop= True, inplace =True)
```

In [8]:
```
home['bhk'] = home['size'].str.split().str[0]
home['bhk'].dropna(inplace = True)
home['bhk'] = home['bhk'].astype('int')
```

Fig 3. Results Screenshot

In [13]:
```
## displaying only the continous variables from the dataset
## to determine the variables which have outliers and those which needs to be removed
fig = plt.figure(figsize = (10,8))
for index,col in enumerate(cont_):
    plt.subplot(3,2,index+1)
    sns.boxplot(y = cont_.loc[:,col])
fig.tight_layout(pad = 1.0)
```

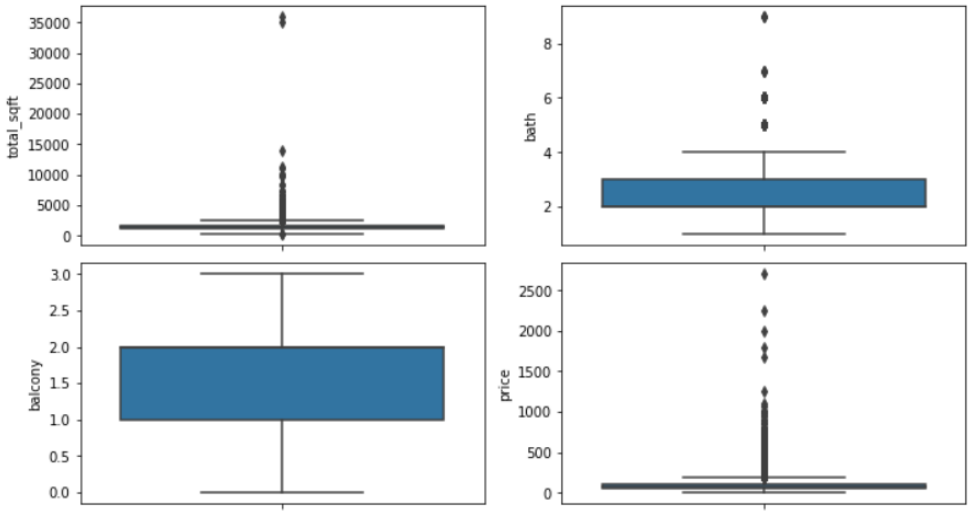

Fig 4. Results Screenshot

In [24]:
```
## representing Numerical Data and Visualizing the same usin Distplot to gain further info

num_ = home.select_dtypes(exclude = 'object')
fig = plt.figure(figsize =(10,8))
for index, col in enumerate(num_):
    plt.subplot(3,2,index+1)
    sns.distplot(num_.loc[:,col],kde = False)
fig.tight_layout(pad = 1.0)
```

```
c:\users\user\appdata\local\programs\python\python36\lib\site-packages\seaborn\distributions.py:2551: FutureWarning: `distplot` is a deprecated function and will be removed in a future version. Please adapt your code to use either `displot` (a figure-level function with similar flexibility) or `histplot` (an axes-level function for histograms).
  warnings.warn(msg, FutureWarning)
```

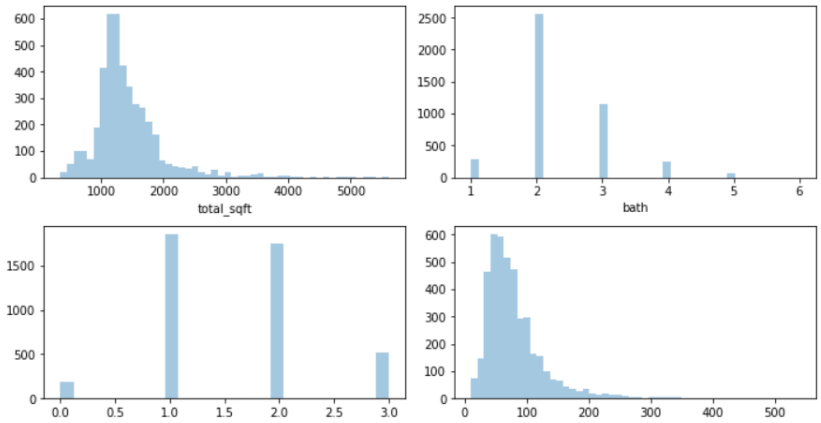

Fig 5. Results Screenshot

Machine Learning Part

In [36]:
```
# from the model selection module import train_test_split for the ML training and testing.
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X1,y,test_size=0.3,random_state=10)
```

In [37]:
```
## importing the required libraries for Machine Learning

from sklearn.model_selection import cross_val_score,cross_val_predict
from sklearn.linear_model import LinearRegression
lr = LinearRegression()
```

In [39]:
```
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error,r2_score
lr.fit(X_train,y_train)
y_pred = lr.predict(X_test)
acc = mean_squared_error(y_pred,y_test)
rscore = r2_score(y_pred,y_test)
print(f'Accuracy with Linear regression : {rscore}')
```

```
Accuracy with Linear regression : 0.8995320160983539
```

In [40]:
```
from sklearn.linear_model import Lasso
lasso_reg = Lasso()
lasso_reg.fit(X_train,y_train)
y_pred = lasso_reg.predict(X_test)
acc = mean_squared_error(y_pred,y_test)
rscore = r2_score(y_pred,y_test)
print(f'Accuracy with Lasso regression : {rscore}')
```

```
Accuracy with Lasso regression : 0.8608167697809426
```

Fig 6. Results Screenshot

In [42]:
```
from sklearn.linear_model import Ridge
Ridge_reg = Ridge()
Ridge_reg.fit(X_train,y_train)
y_pred = Ridge_reg.predict(X_test)
acc = mean_squared_error(y_pred,y_test)
rscore = r2_score(y_pred,y_test)
print(f'Accuracy with Ridge regression : {rscore}')
```

```
Accuracy with Ridge regression : 0.8994713868000848
```

In [44]:
```
from xgboost import XGBRegressor
xgb_reg = XGBRegressor()
xgb_reg.fit(X_train,y_train)
y_pred = xgb_reg.predict(X_test)
acc = mean_squared_error(y_pred,y_test)
rscore = r2_score(y_pred,y_test)
print(f'Accuracy with Extreme Gradient Boosting regression : {rscore}')
```

```
Accuracy with Extreme Gradient Boosting regression : 0.9070356585958991
```

Fig 7. Results Screenshot

V. CONCLUSION

An optimal model does not necessarily represent a robust model. A model that frequently use a learning algorithm that is not suitable for the given data structure. Sometimes the data itself might be too noisy or it could contain too few samples to enable a model to accurately capture the target variable which implies that the model remains fit. When we observe the evaluation metrics obtained for advanced regression models, we can say both behave in a similar manner. We can choose either one for house price prediction

compared to basic model. With the help of box plots, we can check for outliers. If present, we can remove outliers and check the model's performance for improvement.

VI. FUTURE WORK

It is necessary to check before deciding whether the built model should or should not be used in a real-world setting .The data has been collected in 2016 and Bengaluru is growing in size and population rapidly. So, it is very much essential to look into the relevancy of data today. The characteristics present in the data set are not sufficient to describe house prices in Bengaluru. The dataset considered is quite limited and there are a lot of features, like the presence of pool or not, parking lot and others, that remain very relevant when considering a house price. The property has to be categorized either as a flat or villa or independent house. Data collected from a big urban city like Bengaluru would not be applicable in a rural city, as for equal value of feature prices, which will be comparatively higher in the urban area. We can build models through advanced techniques namely random forests, neural networks, and particle swarm optimization to improve the accuracy of predictions.

II. REFERENCES

[1]. Neelam Shinde, Kiran Gawande. "Valuation of house prices using Predictive Techniques", International Journal of Advances in Electronics and Computer Science, ISSN: 2393-2835, Volume-5, Issue-6, Jun.-2018 pp.

[2]. Sifei Lu, Zengxiang Li, Zheng Qin, Xulei Yang, Rick SiowMong Goh, "A Hybrid Regression Technique for House Price Prediction", December 2017.

[3]. Adyan Nur Alfiyatin, Hilman Taufiq, Ruth EmaFebrita, Wayan Firdaus Mahmudy, "Modeling House Price Prediction using Regression Analysis and Particle Swarm Optimization", (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 8, No. 10, 2017.

[4]. Prof. Pradnya Patil Assistant Professor. House Price Prediction Using Machine Learning and RPA. (IRJET). 2020

[5]. Alisha Kuvalekar, Shivani Manchewar, SidhikaMahadik. House Price Forecasting Using Machine Learning. Proceedings Of The 3rd International Conference On Advances In Science & Technology (Icast). 2020.

[6]. I.J. Information Engineering and Electronic Business, 2020, 2, 15-20 Published Online April 2020 in MECS (http://www.mecs-press.org/)DOI: 10.5815/ijieeb.2020.02.03

[7]. H.L. Harter,Method of Least Squares and some alternatives-Part II.International Static Review.1972,43(2) ,pp. 125-190.

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

[9]. Lu. Sifei et al,A hybrid regression technique for house prices prediction. In proceedings of IEEE conference on Industrial Engineering and Engineering Management: 2017.

[10]. R. Victor,Machine learning project:Predicting Boston house prices with regression in towards datascience.

[11]. 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

[12]. S. Abhishek.:Ridge regression vs Lasso,How these two popular ML Regression techniques work. Analytics India magazine,2018.

[13]. S.Raheel.Choosing the right encoding method-Label vs One hot encoder. Towards datascience,2018.

[14]. Raj, J. S., & Ananthi, J. V. (2019). Recurrent Neural Networks and Nonlinear Prediction in Support Vector Machines. Journal of SoftComputing Paradigm (JSCP), 1(01), 33-40.

[15]. Predicting house prices in Bengaluru(Machine Hackathon) https://www.machinehack.com/course/predicting-house-prices-inbengaluru/

[16]. 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.

[17]. Pow, Nissan, Emil Janulewicz, and L. Liu (2014). Applied MachineLearning Project 4 Prediction of real estate property prices in Montréal. [12] Wu, Jiao Yang(2017). Housing Price prediction Using Support Vector Regression.

[18]. Limsombunchai, Visit. 2004.House price prediction: hedonic price model vs. artificial neural network.New Zealand Agricultural and Resource Economics Society Conference.

[19]. 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

[20]. 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.

[21]. Danny P. H. Tay and David K. H. Ho.(1992)Artificial Intelligence and the Mass Appraisal of Residential Apartments. In: Journal of Property Valuation and Investment 10.2 pp. 525–540.