IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 11, Issue 7, July 2023

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.379

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 7, July 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1107037 |

House Price Prediction Using Random Forest Algorithm

Shreyanka K V[1], Manjunatha H T[2]

PG Student, Dept. of Master of Computer Applications, Jawaharlal Nehru New College of Engineering, Shivamogga, India

Assistant professor, Dept. of Master of Computer Applications, Jawaharlal Nehru New College of Engineering, Shivamogga, India

ABSTRACT: The parties involved in real estate, such as homeowners, purchasers, agents, creditors, and investors, an accurate estimate of its worth is essential. Prices are influenced by elements like size, number of rooms, and location, and they are subject to factors like market demand and special circumstances. In order to forecast home sales and create an accurate housing pricing model, this research employs random forest algorithms. The dataset consists of 13321 samples and 9 feature variables. The research explores the influence of features on prediction methods using conventional and cutting-edge machine learning methodologies and offers upbeat results by testing a number of regression strategies. Every year, prices rise as a result of rising demand for homes. Buyers, developers, and the real estate sector are interested in correct elements impacting home pricing.

KEYWORDS: House price prediction, Machine Learning, Random Forest Algorithm.

I. INTRODUCTION

Homeowners, appraisers, developers, tax assessors, investors, and mortgage lenders all depend on house price predictions, which are made possible by machine learning. Models' forecasting ability has increased by examining property value factors such dwelling site, quality, location, and environment. Due to rising interest rates, which undermined asset values and precipitated a dramatic decrease in home prices, the US real estate industry went through a worldwide crisis in 2005. However, the market saw an upward trend in 2012, necessitating the development of more accurate prediction models in order to safeguard the economy against foreseeable dangers. Machine learning techniques are becoming employed more frequently in prediction, which is advantageous to many different economic sectors. This study use a variety of machines to investigate the housing pricing issue.

Support vector regression and artificial neural networks are two examples of machine learning models that may be used to estimate home prices. With the use of this model, buyers, investors, and builders may estimate market valuations and choose their ideal homes depending on their budgets. Prior research has concentrated on examining factors influencing home values. In order to estimate house sales and develop an accurate housing price model, this study employs random forest algorithms.

II. LITERATURE SURVEY

Manasa and Gupta [5] have taken Bengaluru as city for case study. The property size in square feet, location, and its facilities are all key aspects affecting cost 9 different attributes are used For experimental work.

In [3], Luo suggests that to explain the factors that determine residential asset prices, most studies have concentrated on macroeconomic aspects it looks at some micro characteristics such as lot size and pool size that can be utilized as features to estimate house price in this research. random forest and support vector machine are two machine learning methods which are used to predict asset pricing.

Wang et al. [2] state that studies that do not take into account all of the factors influencing property values, provides inaccurate forecast results. As a consequence, for house prediction, the authors propose a full circle joint self- attention model. Authors employ satellite imagery to assess the environment around the residential area.

Lim et al. [1] compared the prediction performance of the ANN model, the multilayer perceptron with that of the ARIMA model in predicting the Singapore housing market.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 7, July 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1107037 |

Piao et al. [7] find that the aspects governing residential real estate prices are complex and the evaluation of useful features is unclear, leading to reduced accuracy in many traditional home price prediction system.

Varma et al. [6] worked upon regression techniques they used weighted average of the multiple techniques the selling price of a house is forecasted using fuzzy ANN and KNN.

Mukhlishin et al. [8] assess the predictive performance to the real world selling price of a house using MAPE the fuzzy technique beats neural networks and k-nearest neighbour for house price prediction in data training using part of dataset as per the findings of the experiments.

Madhuri et al. [4] focused to anticipate house prices based upon their financial capacity and objectives in continuous manner for people looking for their first potential house.

Manjunatha HT and AjitDanti. "A Novel Approach for Detection and Recognition of Traffic Signs for Automatic Driver Assistance System Under Cluttered Background" -Recent Trends on Image Processing and Pattern Recognition, Springer Nature Singapore,Pte Ltd. 2019, **RTIP2R 2018, CCIS 1035, pp. 1–8, 2019, ISBN 978-981-13-9181-1DOI-** https://link.springer.com/chapter/10.1007/978-981-13-9181-1_36.

Manjunatha HT and AjitDanti. "Detection and Classification of Potholes in Indian Roads using Wavelet Based Energy Modules"IEEE**- 978-1-5386-9319-3/19© 2019 ,SCOUPS Nature**

Manjunatha HT, AjitDanti, Arunkumar KL, Rohith D" Indian Road Lanes Detection Based on Regression and clustering using Video processing Techniques", 3rd International Conference on Recent Trends in Image Processing and Pattern Recognition (RTIP2R,2020). Dr. Babasaheb Ambedkar Marathwada University, Aurangabad. Springer, Scopus Indexed 3rd and 4th January 2020.**Springer Nature Singapore Pte Ltd. 2021 K. C. Santosh and B. Gawali (Eds.): RTIP2R 2020, CCIS 1380, pp. 1–14, 2021.** https://doi.org/10.1007/978-981-16-0507-9_17

Manjunatha HT andAjitDanti," Indian traffic sign board recognition using Normalized Correlation Method", International Journal of Computer Engineering and Applications **(IJCEA), Volume XII, Issue III, March 18, ISSN 2321-3469**

Manjunatha HT and AjitDanti, "Segmentation of Traffic Sign Board in a cluttered background using Using Digital Image Processing", **National Conference on Network Security, Image Processing and Information Technology, March 2017**.

III.PROBLEM STATEMENT

The objective of this project is to develop a house price prediction model using the random forest algorithm. The model should accurately predict the prices of residential properties based on various input features, such as the size of the property, number of bedrooms and bathrooms, location, proximity to amenities, and other relevant factors.

The housing market is complex and influenced by numerous variables, making it challenging to determine accurate property values. By employing the random forest algorithm, we aim to create a robust and reliable prediction model that takes into account the interactions between different features and their impact on house prices.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 7, July 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1107037 |

IV. METHODOLOGY

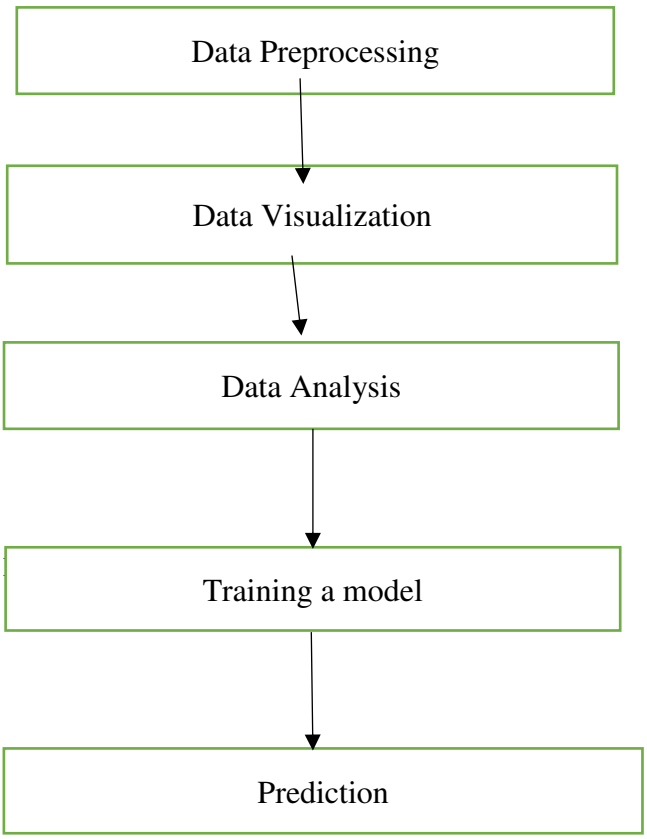

Fig 1: Block diagram of house price prediction

Data gathering, data preprocessing, feature selection, model training, model assessment, and prediction are some of the processes in the Random Forest algorithm's technique for predicting home prices. Data preparation include cleaning, handling missing values, and translating the data into a format appropriate for the Random Forest algorithm whereas data collection entails gathering historical data on home prices and other pertinent parameters. Feature selection is the process of choosing the dataset's most pertinent characteristics by applying methods including correlation analysis, principal component analysis, and feature significance ranking. When developing a prediction model, the Random Forest technique is trained on pre-processed data before several decision trees are built using a variety of randomly chosen subsets of data. The performance and interpretability of the model are enhanced by strategies like bagging and feature significance.

An technique for supervised machine learning called Random Forest is used to solve classification and regression issues. Based on ensemble learning, which integrates several classifiers to address complicated issues and enhance model performance, it can tackle complex problems. To increase predicted accuracy, the method employs many decision trees on subsets of a dataset and takes the average. Higher accuracy and overfitting are prevented by the larger number of trees in the forest.

A common machine learning approach for predictive analysis is linear regression, which may forecast continuous or numerical variables like sales, salary and product price. A slanted straight line illustrating the linear connection between the variables is shown between the dependent variable (y) and one or more independent variables.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 7, July 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1107037 |

i. **Data Collection:** For home price identification, the Boston housing dataset was utilized, which was originally part of the UCI Machine Learning Repository. It has 9 feature variables and 13321 samples. The objective is to predict home values using the available features.

ii. **Data Preprocessing:** Cleaning, addressing missing values, and converting the data into a format appropriate for analysis using the Random Forest method are all parts of data preparation.

iii. **Data Analysis:** The pre-processed data used in the Random Forest technique is used to train numerous decision trees, which are then combined to produce a final prediction model. The performance and interpretability are improved by strategies like bagging and feature significance.

iv. **Training a model:** By applying methods like correlation analysis and relevance ranking, significant features in a dataset that affects home prices are found through the feature selection process.

v. **Prediction:** After model assessment, the trained Random Forest algorithm makes precise forecasts about future property values using pertinent features.

V. RESULTS AND DISCUSSION

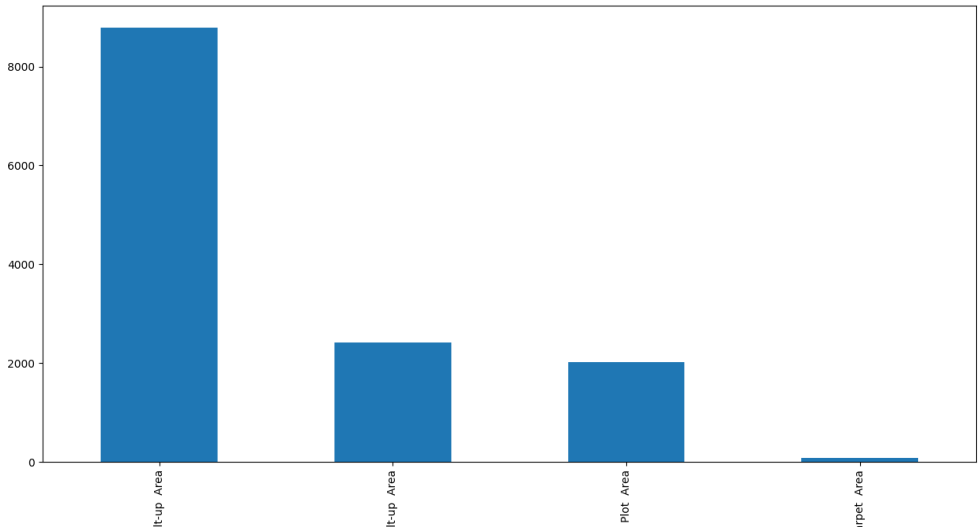

Fig 2: graph of the area

The fig 2 shows the graphical representation of types of area depend upon their height from the seal level.
The graph plotted area versus amount will be prescribed according to the area and their dimensions in the house price.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 7, July 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1107037 |

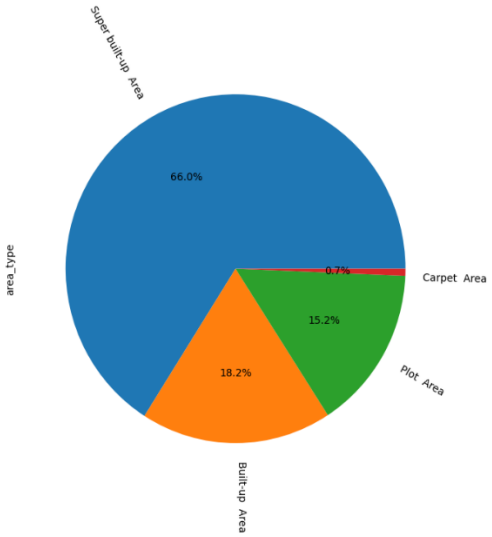

Fig 3: pie chart of types of area

The fig 3 shows the graphical representation of percentage of area built-up and carpet area and plotted area.
The chart shows the percentage of area development in the location built-up areas carpet area plot area and type of area.

	area_type	availability	location	size	society	total_sqft	bath	balcony	price
0	False	False	False	False	False	False	False	False	False
1	False	False	False	False	False	False	False	False	False
2	False	False	False	False	True	False	False	False	False
3	False	False	False	False	False	False	False	False	False
4	False	False	False	False	True	False	False	False	False
...	...	...	...	...	...	...	...	...	...
13315	False	False	False	False	False	False	False	False	False
13316	False	False	False	False	True	False	False	True	False
13317	False	False	False	False	False	False	False	False	False
13318	False	False	False	False	False	False	False	False	False
13319	False	False	False	False	True	False	False	False	False

[13320 rows x 9 columns]

Fig 4: Table of rows and columns involved in the price prediction

The fig 4 shows the graphical representation of rows and columns involved in the testing data.
The above table shows number of rows and columns in the price prediction.

VI. CONCLUSION AND FUTURE WORK

Both the Random Forest technique and linear regression work well for predicting housing prices. Random Forests can manage intricate interactions between features and the target variables and are resistant to overfitting. They give feature importance and successfully manage missing data and outliers. With coefficients for each feature and less processing expense, linear regression is more straight forward and easy to understand. In complicated datasets with non-linear correlations, Random Forests typically beat Linear Regression in predicting accuracy. However, for interpretability and simplicity, linear regression can still be helpful, giving details about the strength and direction of the link between characteristics and home prices. The environment, dataset properties, and analytic objectives all affect which method is used.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 7, July 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1107037 |

REFERENCES

[5] J. Manasa, R. Gupta, and N. Narahari, "Machine learning based predicting house prices using regression techniques," in 2020 2nd International conference on innovative mechanisms for industry applications (ICIMIA). IEEE, 2020, pp. 624–630.

[3] Y. Luo, "Residential asset pricing prediction using machine learning," in 2019 International Conference on Economic Management and Model Engineering (ICEMME). IEEE, 2019, pp. 193–198.

[2] P.-Y. Wang, C.-T. Chen, J.-W. Su, T.-Y. Wang, and S.- H. Huang, "Deep learning model for house price prediction using heterogeneous data analysis along with joint self-attention mechanism," IEEE Access, vol. 9, pp. 55 244–55 259, 2021.

[1] W. T. Lim, L. Wang, Y. Wang, and Q. Chang, "Housing price prediction using neural networks," in 2016 12th International conference on natural computation, fuzzy systems and knowledge discovery (ICNC-FSKD). IEEE, 2016, pp. 518–522.

[7] Y. Piao, A. Chen, and Z. Shang, "Housing price prediction based on cnn," in 2019 9th international conference on information science and technology (ICIST). IEEE, 2019, pp. 491–495.

[6] A. Varma, A. Sarma, S. Doshi, and R. Nair, "House price prediction using machine learning and neural networks," in 2018 second international conference on inventive communication and computational technologies (ICICCT). IEEE, 2018, pp. 1936–1939.

[8] M. F. Mukhlishin, R. Saputra, and A. Wibowo, "Predicting house sale price using fuzzy logic, artificial neural network and k-nearest neighbor," in 2017 1st International Conference on Informatics and Computational Sciences (ICICoS). IEEE, 2017, pp. 171–176.

[4] C. R. Madhuri, G. Anuradha, and M. V. Pujitha, "House price prediction using regression techniques: a comparative study," in 2019 International conference on smart structures and systems (ICSSS). IEEE, 2019, pp. 1–5.

[9] "Classification of Vehicle Make Based on Geometric Features and Appearance-Based Attributes Under Complex Background" Aunkumar K L, Ajit Danti, Manjunatha H T, Springer 1035 (CCIS), pp 41-48

[10] "A NOVEL APPROACH FOR VEHICLE RECOGNITION BASED ON THE TAIL LIGHTS GEOMETRICAL FEATURES IN THE NIGHT VISION" Arunkumar K L, Ajit Danti, International Journal of Computer Engineering and Applications, Volume XII, Issue I, Jan. 18, www.ijcea.com ISSN 2321-3469

[11] "A Novel Approach for Detection and Recognition of Traffic Signs for Automatic Driver Assistance System Under Cluttered Background" Manjunatha H T Arunkumar K L, Ajit Danti, Springer 1035 (CCIS), pp 407-419

[12] "Estimation of vehicle distance based on feature points using monocular vision", Arunkumar K L, Ajit Danti, Manjunatha H T, IEEE 8816996 (2019), 1-5

Peruse and Recognition of Old Kannada Stone Inscription Characters, CM Nrupatunga, KL Arunkumar, Springer, Singapore, 2020

[13] "Classification of Vehicle Type on Indian Road Scene Based on Deep Learning", Arunkumar K L, Ajit Danti, Manjunatha HT , D Rohith, Springer, Singapore 1380 (2021), 1-10

[14] "Indian Road Lanes Detection Based on Regression and clustering using Video Processing Techniques", Manjunatha H T, Ajit Danti, ArunKumar K L, D Rohith, Springer, Singapore 1380 (CCIS), 193-206

[15] "Recognition of Vehicle using geometrical features of a tail light in the night vision", Arunkumar K L, Ajit Danti ,National Conference on Computation Science and Soft Computing (NCCSSC-2018)

INNO SPACE
SJIF Scientific Journal Impact Factor

Impact Factor: 8.379

doi cross ref

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details