Random Forest-Based House Price Prediction

B.Usha Sri
Computer Science Engineering
B V Raju Institute of Technology
Narsapur, Medak-502313
ushasri.b@bvrit.ac.in

Chaganti Santhosh Kumar Reddy
Computer Science Engineering
B V Raju Institute of Technology
Narsapur, Medak-502313
santhoshchaganti897@gmail.com

Chiluveri Rithish Kumar
Computer Science Engineering
B V Raju Institute of Technology
Narsapur, Medak-502313
rithish25062002@gmail.com

Akula Vyshnavi
Computer Science Engineering
B V Raju Institute of Technology
Narsapur, Medak-502313
vyshnaviakula81@gmail.com

Buttagalla Vinod
Computer Science Engineering
B V Raju Institute of Technology
Narsapur, Medak-502313
vinodbuttagalla@gmail.com

***Abstract*—The precise prediction of house prices is highly significant for professionals in the real estate industry, homeowners, and potential buyers. This study primarily focuses on utilizing the random forest algorithm to accurately predict house prices. A comprehensive dataset containing various features such as location, size, rooms, amenities, and historical transaction data was collected. The investigation included steps of data conditioning, designing features, instruction of models, and oversight. To assess the outcome of the model along with avoiding the situation a random-forest stance was employed, and validated cross-valid techniques were employed. Lasso regression is employed to identify the key features that significantly influence house prices. Results showed promising predictive accuracy, surpassing other baseline algorithms. Feature importance analysis highlighted the significant influence of location, size, and the number of rooms on house prices, aligning with domain knowledge. This study illustrates the precision of random forests and Lasso-regression-based ML assertions, assisting housing specialists, homeowners, and potential purchasers in making accurate choices. Future research can explore integrating additional data sources and advanced techniques to enhance prediction accuracy and address housing market challenges.**

***Index Terms*—Machine Learning, Random Forest Algorithm, Lasso Regression, Mean Squared Error (MSE), Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and R-squared Error.**

I. Introduction

Housing is an important measure of a country's economic performance. As cities' populations grow due to urbanization, so does the need for housing. Predicting House prices are an important responsibility in the real estate sector since they allow property buyers and sellers to make informed decisions. However, calculating the exact worth of a house is a difficult task that is influenced by a variety of criteria such as public facilities, location, size, age, and condition. To facilitate informed decision-making, computer-based prediction systems can be utilized. The practice of ML, a component of technological innovation (AI), employs strategies and equipment to extract conclusions from information.

The use of machine learning algorithms, particularly Random Forest, has gained significant popularity in accurately forecasting house prices by leveraging a wide range of property features. Random Forest, which is based on decision trees, demonstrates exceptional performance in both regression and classification tasks. There are multiple phases associated with creating a Random Forest model for predicting home prices, including data preparation, development of the model, and model evaluation. By utilizing several property attributes, machine learning algorithms, particularly Random Forest, provide a reliable way for evaluating home prices. In the real estate sector, accurate house price forecasting is essential, along with machine learning algorithms, particularly Random Forest, which offer a trustworthy method to accomplish this. Proper feature selection, thorough data preparation, efficient model training, and comprehensive assessment are essential for improving prediction model precision as well as dependability. The subsequent sections will go into these in further detail, presenting methodologies and results for predicting house prices using the Random Forest algorithm.

II. Literature Survey

Rinabi Tanamal et al. [1] provided a novel method for using the random forest algorithm to forecast home values in Surabaya City, Indonesia. The study focuses on examining the effects of several variables, including location, size, age, and accessibility, on the model's ability to predict outcomes accurately. To test their suggested strategy, they used a dataset of 1,000 house listings from the website Rumah123.com.They observed better outcomes when comparing their strategy with several current models, such as multiple linear regression and support vector regression. According to the study's findings, the approach that was suggested can be utilized to accurately anticipate property prices in Surabaya City and can help buyers, investors, and real estate agents make wise choices. The necessity for larger datasets, the integration of more variables like crime rates and environmental conditions, and the investigation of deep learning techniques are just a few of the constraints and prospects the authors have covered for their suggested strategy.

Dhileep Kumar et al. [2] centered on applying machine learning methods like Linear Regression, Decision trees, k-Means, and Random Forest to estimate home prices. They can train models using different features, such as ZN, INDUS, CHAS, and RAD, in the suggested system. Of the previously acquired statistics, 80% of it serves for training, and the rest, or 20%, is used for testing. The raw data is kept in the '.csv' file used here. The '.csv' file was loaded into the Jupiter notebook using pandas and Numpy, which were also used to clean and manipulate the data. Another was sci-kit-learn, which was used for actual analysis and offers several built-in functions that aid with issue-solving.MISE and cross-validation were the two strategies employed for assessing the models' performance. These techniques were employed following the collection, cleaning, examination, training, and operational examination of the model.

Suryawanshi et al. [3] proposed a machine-learning approach for predicting hotel, flat, and vacation rental prices. They examined various factors such as location, amenities, and seasonality to enhance prediction accuracy. They analyzed publicly available datasets from platforms like Expedia, Booking.com, and Airbnb. They compared different ML techniques, including regression analysis, decision trees, and random forests, to identify the most effective approach for price prediction. The study demonstrates that their suggested method reliably forecasts hotel rates, benefiting travelers, tour operators, and lodging providers in making informed decisions. They highlighted the need for larger datasets, the inclusion of additional factors like user reviews and ratings, and the exploration of deep learning techniques as future directions for their proposed strategy.

Chee Kin et al. [4] suggested a home price forecast technique based on machine learning. The Random Forest algorithm was utilized by the researchers to calculate house values, and the outcome of the suggested forecasting algorithm was evaluated employing the Boston Housing dataset 1.

Gupta et al. [5] suggested a methodology to forecast inflation in housing based on machine learning regression approaches. They projected the property values employing the Random Forest machine learning strategy on assessing the validity of the suggested forecasting approach employing the UCI Machine learning repository Boston housing dataset of 506 entries and 14 attributes.

Amena Begum et al. [6] proposed utilizing location, area, count of bedrooms, and bathrooms as inputs for machine learning algorithms to estimate house prices. They compared various techniques such as linear regression, decision trees, and neural networks, using a dataset of 21,613 house listings from the Kaggle competition "House Prices: Advanced Regression Techniques." Their strategy outperformed existing models like gradient boosting and random forests, accurately estimating house prices and aiding buyers, investors, and real estate agents. The authors highlighted the need for larger datasets, the inclusion of variables like crime rates and environmental conditions, and the exploration of deep learning techniques as prospects for their suggested strategy.

Adetunji et al. [7] proposed using the random forest to estimate property prices, showing its effectiveness in regression tasks. They investigated the effect of characteristics such as location, size, age, and amenities on forecast accuracy. A comparison analysis was carried out using linear regression and support vector regression. According to the study, random forest effectively forecasted house prices, which benefited real estate brokers, investors, and home buyers. Limitations, future research areas, and prospective applications were highlighted, offering useful information for the real estate domain.

Akilendra Pratap Singh et al. [8] proposed a machine learning-based housing price forecasting method. They employed the Linear Regression, Random Forest, and XGBoost algorithms to calculate the value of homes and tested their suggested forecasting algorithm on the California Housing dataset.

Chen et al. [9] suggested an approach that can forecast home values that employ machine learning along with deep computing methodologies. To estimate property prices and evaluate the algorithms, the authors utilized five common learning methods and deep learning strategies: Linear Regression (LR), Bayesian, Backpropagation neural network (BP neural network), Support Vector machine learning (SVM), and Deep Neural Network (DNN). The research's dataset was gathered via the Kaggle platform.

Anand G. Rawool et al. [10] proposed a revolutionary method for accurately predicting housing values, utilizing regression models, decision trees, and artificial neural networks. The study examined the effects of variables such as location, size, age, and amenities on forecasting outcomes. Publicly accessible datasets, including the Boston Housing, California Housing, and Ames Housing datasets, were utilized for data collection, preprocessing, model training, and testing. The authors compared their approach with support vector machines, random forests, and linear regression, yielding superior results. The suggested method enables accurate anticipation of home prices, aiding buyers, investors, and real estate agents in making informed decisions. Limitations and future directions were also discussed, highlighting the need for more extensive datasets, the inclusion of additional features (e.g., crime rates), and the exploration of deep learning techniques.

Khan Sohil Liyaqatullah et al. [11] suggested a housing price forecasting framework that makes use of algorithms for Machine Learning including linear regression, Decision Tree, K-Means, and also Random Forest Regression. Their study employed 80% of the data gathered from the known collection to train models and the rest, or 20%, for testing.

Anirudh Kaushal et al. [12] utilized multiple linear regression to predict house prices. They looked examined the association between house prices and factors like location, size, rooms, amenities, and neighborhood features. For training and evaluation, publically available real estate datasets were used. Preprocessing techniques like missing value handling and normalization were used, as well as assessment measures such as mean squared error, mean absolute error, and r-squared value. The study indicated that multiple linear regression was a more reliable and interpretable method of predicting property prices than other regression-based models. The study shed light on methodology preprocessing approaches and prospective uses in real estate decision-making. They admit that further research is required to validate their conclusions and address shortcomings.

Smith et al. [13] used real estate data with features like location, size, and amenities. They employed machine learning algorithms (e.g., linear regression, decision trees, k-nearest neighbors) to predict prices. Feature selection was used to identify important features. Their support vector regression model trained on 50,000 US real estate transactions achieved high accuracy, with XGBoost Regression performing best. They suggested improving the model's usability by developing a user interface for accessing real estate price predictions for various locations.

J. J. Wang et al. [14] in the system that forecasts prices for houses, the pertinent data marked as such is used for online training of the ANN. The ANN gains the ability to forecast home prices and then uses the unlabeled data to make the forecast. It consists of up of a total of three layers of information: a data entry module, a covered layer, and a layer that provides output. In the final layer, there's merely one neuron that collects currents via inputs. Chau predicted the memristor based on the symmetry of circuit theory. By altering the measured voltage that crosses the memristor, the memristance, or staunch opposition of the memristor, may be personalized to an appropriate magnitude. The ann can automatically alter the weight online in training mode depending on the bp algorithm. The precision of prediction is indicated by the square of the difference between the intended and anticipated results.

Sifei Lu et al. [15] looked at the engineering of innovative features and suggested a hybrid Lasso and Gradient boosting regression model that ensures superior prediction. They chose features using Lasso regression. To determine the ideal number of characteristics that can enhance the prediction, they performed numerous iterations of feature engineering. The Kaggle score evaluation was improved when more features were introduced. As a result, they increased the 79 provided features by 400. Additionally, they tested using Ridge, Lasso, and gradient boosting and discovered that 230 features provided the greatest result after using Lasso to remove the unnecessary features. In 2017 it is simple to use but not that much efficient for prediction. Finally, they recommended creating a separate algorithm to detect and projected aberrant transaction costs.

The literature investigation highlights the significant improvements made in the use of machine learning methodologies along with strategies for home price estimation. Regression modeling, artificial neural networks, decision tree models, support vector regression, as well as random forest regression, constitute a few of the frequently employed machine learning algorithms to anticipate the value of homes. The choosing of attributes, simplifying the model, and inclusion of geographical and temporal variables can all significantly improve the accuracy of forecasting home price models. The latest study additionally indicates that hybrid approaches that combine deep neural networks, sentiment analysis, and machine learning are capable of boosting the reliability of systems to anticipate the cost of real estate.

Existing Work

The latest research on utilizing machine learning algorithms for predicting home prices has yielded promising results in terms of accuracy and effectiveness. Several studies that examined the effectiveness of various algorithms discovered that ensemble approaches, such as random forest and gradient boosting, typically outperformed other algorithms, like linear regression and SVM. The size of the dataset, the count of features included, and the complexity of the model are just a few variables that may affect how effective these models are. Yet, current research on house price prediction models has shown that machine learning algorithms may produce precise and effective predictions and can aid in improving real estate market decision-making.

Recent studies on home price prediction provide evidence for the rationale behind the random forest method we chose for our study. These studies have repeatedly demonstrated that ensemble methods—like the random forest—perform better than other algorithms, such as linear regression and SVM. The choice can be made because random forests can manage big datasets with intricate properties. While deep learning models demand more time and money for learning, the random forest method balances accuracy and efficiency. The random forest method was chosen for our study because of its accuracy and efficiency in predicting home prices, which have been demonstrated.

III. Research Methodology

We utilized the Random Forest method in this project to predict house prices since, as we saw, it produced results with the highest accuracy and fewest errors. Fig.1 demonstrates the workflow of our research paper, which focuses on forecasting house prices using the Random Forest method. We focused on the most crucial characteristics that had a significant impact on the price of homes, such as the distance from a main road, the accessibility of amenities for the general public, the availability of parking, the number of bedrooms in the home, etc.

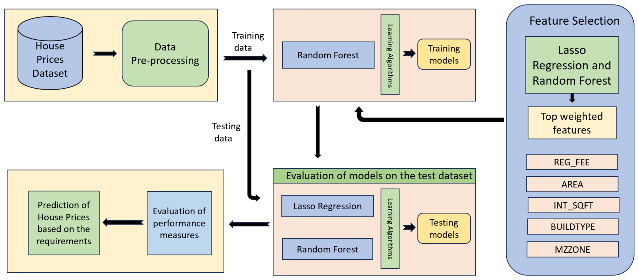

Fig.1. Research Workflow Diagram

A. Dataset Description

Among the datasets available in the UCI machine learning repository, you can find the Chennai Housing Sales Price dataset, which encompasses a diverse range of information. This particular dataset consists of 22 attributes in Fig.2, comprising 10 categorical attributes and 12 numerical attributes. It encompasses a total of 7,109 recorded house details. To streamline the analysis, a subset of 12 features is selected from the original 22, with these 12 features serving as inputs for the model. Among these 12 features, 11 are considered independent variables, while 1 is the dependent variable used for prediction.

	PRT_ID	AREA	INT_SQFT	DATE_SALE	DIST_MAINROAD	N_BEDROOM	\
0	P03210	Karapakkam	1004	04-05-2011	131	1.0	
1	P09411	Anna Nagar	1986	19-12-2006	26	2.0	
2	P01812	Adyar	909	04-02-2012	70	1.0	
3	P05346	Velachery	1855	13-03-2010	14	3.0	
4	P06210	Karapakkam	1226	05-10-2009	84	1.0	

	N_BATHROOM	N_ROOM	SALE_COND	PARK_FACIL	...	UTILITY_AVAIL	STREET	MZZONE	\
0	1.0	3	AbNormal	Yes	...	AllPub	Paved	A	
1	1.0	5	AbNormal	No	...	AllPub	Gravel	RH	
2	1.0	3	AbNormal	Yes	...	ELO	Gravel	RL	
3	2.0	5	Family	No	...	NoSewr	Paved	I	
4	1.0	3	AbNormal	Yes	...	AllPub	Gravel	C	

	QS_ROOMS	QS_BATHROOM	QS_BEDROOM	QS_OVERALL	REG_FEE	COMMIS	SALES_PRICE
0	4.0	3.9	4.9	4.330	380000	144400	7600000
1	4.9	4.2	2.5	3.765	760122	304049	21717770
2	4.1	3.8	2.2	3.090	421094	92114	13159200
3	4.7	3.9	3.6	4.010	356321	77042	9630290
4	3.0	2.5	4.1	3.290	237000	74063	7406250

Fig.2. A Snippet of Raw Dataset

The data set has been divided into two parts: 20% to be tested and 80% to be trained to ensure that the system is properly analyzed. The first section offers a comprehensive evaluation of its efficacy on unobserved data in addition to a sizable amount of knowledge for learning and developing behaviors.

B. Data Classification

The strategy of random forests is employed for projecting house prices by utilizing a variety of variables connected to the place of residence, which include property characteristics, structural aspects, amenities, and socioeconomic trends. The goal is to develop a predictive model that accurately estimates the sale price of a property.

Location-based Features: Location-based features capture geographic factors that influence house prices. These include the neighborhood's desirability, proximity to amenities such as schools, hospitals, shopping centers, and transportation accessibility. These features help capture the local market dynamics and demand for properties in specific areas.

Property Characteristics: Property characteristics encompass important attributes of a house, such as its size, number of bedrooms and bathrooms, lot area, and floor area. These features provide insights into the property's capacity to accommodate residents, its overall space, and its layout.

Structural Attributes: Structural attributes consider aspects such as the age of the property, construction quality, and architectural style. The age of a property can impact its condition and potential maintenance costs. Construction quality and architectural style can contribute to its overall appeal and market value.

Amenities and Facilities: Amenities and facilities refer to the presence or absence of specific features that enhance the quality of living. This may include the availability of parking spaces, gardens, swimming pools, security systems, or other recreational amenities. These features can positively influence the perceived value of a property.

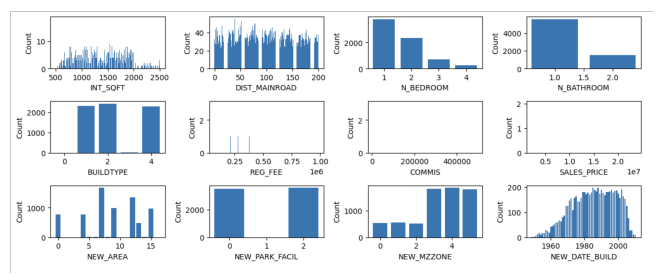

Fig.3.Count of records for each unique value in each attribute

Socioeconomic Factors: Socioeconomic factors incorporate demographic information, economic indicators, crime rates, and school district ratings. These factors provide insights into the overall desirability and livability of an area. Demographic information, such as population density and income levels, can indicate the purchasing power of potential buyers. Economic indicators reflect the health and growth of the local economy, which can impact housing demand. Crime rates and school district ratings contribute to the perception of safety and educational opportunities, influencing property values.

C. Data Preprocessing

Data preprocessing is a crucial aspect of house price prediction, involving several essential steps. The initial data, which includes property characteristics, location-based features, and socioeconomic factors, were gathered from various sources. To ensure data quality, duplicate records were eliminated, and missing values were handled appropriately. As shown in Fig.3, we generated a graph to enumerate the count and distinct values contained in every attribute, allowing for a full understanding of the dataset. Feature scaling was applied to standardize the scale of all features, while categorical variables were encoded using techniques like one-hot encoding. Lastly, the dataset was divided into separate training and testing sets to facilitate model development and evaluation. imputation.

	INT_SQFT	DIST_MAINROAD	N_BEDROOM	N_BATHROOM	BUILDTYPE	REG_FEE	COMMIS	SALES_PRICE	NEW_AREA	NEW_PARK_FACIL	NEW_MZZONE	NEW_DATE_BUILD
0	1004	131	1.0	1.0	1	380000	144400	7600000	12	2	0	1967
1	1986	26	2.0	1.0	1	760122	304049	21717770	4	0	3	1995
2	909	70	1.0	1.0	1	421094	92114	13159200	0	2	4	1992
3	1855	14	3.0	2.0	4	356321	77042	9630290	15	0	2	1988
4	1226	84	1.0	1.0	4	237000	74063	7406250	12	2	1	1979

, Fig.4.A Snippet of Cleaned Dataset

- Data Cleaning: We eradicate redundant entries, deal with unusual values, check and update data for precision and regularity, and maintain data reliability. These tactics boost the precision and grade of our study, producing solid data insights.
- Handling Missing Values: Resolving unfinished values by managing missing data boosts the dependability of our house price forecasts, ensuring the precision of our algorithms. We treat values that aren't present in our house price prediction dataset using techniques like regression. Upon completing the data cleaning process and handling missing values, the dataset is visualized in Fig.4, showcasing its refined form.
- Data Splitting: We partitioned the cleaned information between test and training groups for machine learning. In general, training makes use of 80% of the data while validation makes use of the remaining 20%. The exact split ratio may vary depending on the dataset's size and requirements.

D. Feature Selection

Fig.5. Features selection using Lasso Regression

- Lasso Regression: In our study, we utilize Lasso-based regression analysis, which incorporates L1 regularization to encourage sparsity in the model by penalizing less significant feature coefficients. We control the level of regularization using the lambda parameter, which affects the coefficients. By simultaneously selecting features and reducing coefficients, our Lasso regression approach enhances interpretability. Fig.5 illustrates the graph of feature selection using lasso regression, offering detailed insights into the factors that have a significant impact on house prices. In our analysis, we leverage cross-validation to optimize results, particularly for high-dimensional datasets with potentially irrelevant properties.

- Random Forest Algorithm: In our research, the parameters to forecast the cost of property were identified by employing the Random Forest algorithm. By training a group of decision trees, the approach estimated the significance of each characteristic as shown in Fig.6 by examining its effect on projected accuracy. As a result, we were able to identify the crucial traits for estimating house prices. We choose these top-ranked attributes to build a more specific model. This method significantly increased the effectiveness and interpretability of our house price forecast model.

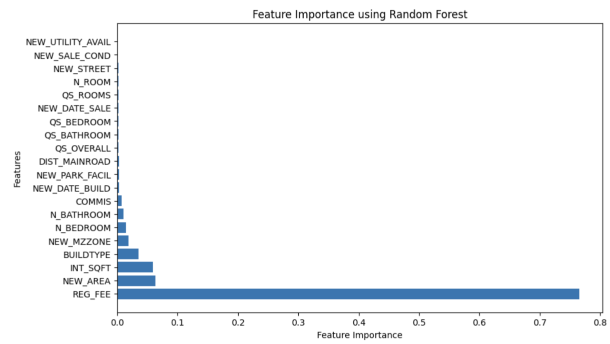

Fig.6. Features selection using Random Forest Regression

IV. Results

A. Lasso Regression

Lasso Regression research we conducted produced encouraging results for feature selection. L1 regularisation was employed to efficiently shrink parameters and several important characteristics. The Lasso-based algorithm can be capable to decrease forecasting mistakes, and boosting the accuracy of house price projections, according to the computed MSE of 2156326591156.32. Additionally, the calculated R-squared value of 0.854 shows that the chosen attributes may explain about 85.4% of the variation in house prices. These outcomes show how effectively the Lasso Regression technique handles complex data sets with a wide range of possible parameters, offering insightful information for predicting home prices.

B. Random Forest Algorithm

In our research, we utilized the Random Forest algorithm to predict house prices, achieving an impressive accuracy rate of 97%. This signifies the algorithm's effectiveness in capturing the intricate relationships between the features and the target variable. We evaluated the profitability of our hypothesis using metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), and R-squared. The obtained MAE value of 496797.40, MSE value of 411560850852.61, and R-squared value of 0.972 highlight the strong performance of our model.

```python
# Print the predicted prices
# Create a dictionary with the input values
input_values = {
    'INT_SQFT': 1986,
    'DIST_MAINROAD': 26,
    'N_BEDROOM': 2.0,
    'N_BATHROOM': 1.0,
    'BUILDTYPE_Commercial': 1,
    'REG_FEE': 760122,
    'COMMIS': 304049,
    'NEW_AREA_Anna Nagar': 1,
    'NEW_PARK_FACIL_No': 1,
    'NEW_MZZONE_RH': 1,
    'NEW_DATE_BUILD_1995': 1
}
print("Predicted House Prices:")
for price in predicted_prices:
    print(price)
```

✓ 0.0s

```
Predicted House Prices:
21111783.9
```

Fig.7. An example test case for house price prediction

V. Conclusions

By examining the coefficients of the selected features, it was observed that variables such as property size, number of bathrooms, and location proximity were influential in predicting house prices. Other features with near-zero coefficients were deemed less significant. This information can be visualized in Table.1, which presents the RMSE, MAE, MSE, and R-squared values for both Random Forest and Lasso Regression algorithms. Lasso Regression successfully predicted house prices by leveraging L1 regularization and feature selection. It achieved accurate estimates and effectively handled high-dimensional datasets. Based on the observation depicted in Fig.7, we can observe that the test case incorporates input values from Fig.2, representing the unprocessed dataset. Our algorithm has generated house price predictions that exhibit a high level of similarity to the actual house prices.

Table. 1. RMSE, MAE, MSE, and R Squared Values for Random Forest and Lasso Regression Algorithms

Values → / Algorithms ↓	RMSE	MSE	MAE	R Squared
Random Forest	641530.08	411560850852.61	496797.40	0.97
Lasso Regression	1468443.59	2156326591156.32	1157120.82	0.85

Random Forest algorithm exhibits high accuracy in predicting house prices by effectively capturing complex feature-target relationships.

VI. Future Enhancement

To improve the research on house price prediction using the Random Forest algorithm, several areas for future enhancement can be explored. These include advancing feature engineering techniques, investigating ensemble learning methods, considering temporal analysis, integrating external data sources, and enhancing model interpretability. These enhancements aim to derive more informative features, enhance predictive performance, capture temporal dynamics, incorporate broader data perspectives, and gain a better understanding of feature importance.

References

[1] R. Tanamal, N. Minoque, T. Wiradinata, Y. Soekamto, and T. Ratih, "House price prediction model using random forest in surabaya city," TEM Journal, pp. 126-132, Feb. 27, 2023, doi:DOI: 10.18421/TEM121-17

[2] D. Kumar, S. Sakthivel, and K. Kalachelvi, "House price prediction using random forest and CNN algorithm," International Research Journal of Modernization in Engineering Technology and Science, vol. 4, issue 8, pp. 5, Aug. 2022.

[3] R. Suryawanshi, N. Thakur, K. Waghmare, M. Kshirsagar, R. Gaurkar, and M. Wagh, "Accommodation price prediction using machine learning," in International Journal of Progressive Research in Science and Engineering, vol. 3, no. 04, pp. 1-6, Apr. 2022.

[4] C. Chee Kin, Z. Arabee Bin Abdul Salam, and K. Batcha Nowshath, "Machine learning based house price prediction model," 2022 International Conference on Edge Computing and Applications (ICECAA), Tamilnadu, India, 2022, pp. 1423-1426, doi: 10.1109/ICECAA55415.2022.9936336.

[5] A. Gupta, S. K. Dargar and A. Dargar, "House prices prediction using machine learning regression models," 2022 IEEE 2nd International Conference on Mobile Networks and Wireless Communications (ICMNWC), Tumkur, Karnataka, India, 2022, pp. 1-5, doi: 10.1109/ICMNWC56175.2022.10031728.

[6] A. Begum, N. J. Kheya, and Md. Z. Rahman, "Housing price prediction with machine learning," International Journal of Innovative Technology and Exploring Engineering (IJITEE), vol. 11, issue 3, Jan. 2022, ISSN: 2278-3075.

[7] A. B. Adetunji, O. N. Akande, F. A. Ajala, O. Oyewo, Y. F. Akande, and G. Oluwadara, "House price prediction using random forest machine learning Technique," in The 8th International Conference on Information Technology and Quantitative Management, 2021-2022, pp. 8, doi:0.1016/j.procs.2022.01.100

[8] A. P. Singh, K. Rastogi, and S. Rajpoot, "House price prediction using machine learning," 2021 3rd International Conference on Advances in Computing, Communication Control and Networking (ICAC3N), Greater Noida, India, 2021, pp. 203-206, doi: 10.1109/ICAC3N53548.2021.9725552.

[9] Y. Chen, R. Xue, and Y. Zhang, "House price prediction based on machine learning and deep learning methods," 2021 International Conference on Electronic Information Engineering and Computer Science (EIECS), Changchun, China, 2021, pp. 699-702, doi: 10.1109/EIECS53707.2021.9587907.

[10] A. G. Rawool, D. V. Rogye, S. G. Rane, and Dr. V. A. Bharadi, "House price prediction using machine learning," IRE Journals, vol. 4, no. 11, pp. 5, May 2021.

[11] K. S. Liyaqatullah1, Usman Malik, Q. M. Basheeruddin, "House price prediction using machine learning" International Journal of Advanced Research in Computer and Communication Engineering(IJARCCE), Vol. 10, Issue 4, April 2021, pp.385-393 doi: 10.17148/IJARCCE.2021.10468

[12] Kaushal, Anirudh, & Shankar, Achyut (2021), "House price prediction using multiple linear regression," SSRN Electronic Journal. https://doi.org/10.2139/ssrn.3833734

[13] D. Smith, S. Rodrigues, V. Rodrigues, and P. Shah, "Real estate price prediction," in International Journal of Engineering Research and Technology (IJERT),2021.

[14] J. J. Wang, S. G. Hu, X. T. Zhan, Q. Luo, Q. Yu, Zhen Liu, T. P. Chen, Y. Yin, Sumio Hosaka, and Y. Liu, "Predicting house price with a memristor-based artificial neural network," in IEEE Access, vol. 6, pp. 16523-16528, 2018, doi: 10.1109/ACCESS.2018.2814065.

[15] S. Lu, Z. Li, Z. Qin, X. Yang, and R. S. M. Goh, "A hybrid regression technique for house prices prediction," 2017 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM), Singapore, 2017, pp. 319-323, doi:10.1109/IEEM.2017.8289904.