HOUSE PRICE PREDICTION USING MACHINE LEARNING ALGORITHM

[1]Mrs.N.Vishnu Devi, [2]Mr.B.Sundaresan, [3]Mr.T. J.Shiyam Sundar, [4]Mr.S.Naveen

[1]Assistant Professor, [2,3,4]UG Student
Department Of Information Technology
K.L.N College Of Engineering
Pottapalayam, Sivaganga Distrtict, Tamil Nadu, India.

***Abstract*- This research delves into the domain of house price prediction, employing the Random Forest algorithm within the framework of machine learning. Emphasizing crucial factors like location, amenities, and house type, the primary objective is to construct a robust model that effectively assists users in making informed decisions while selecting homes within their budget. The study systematically evaluates the Random Forest algorithm's performance in providing accurate predictions for house prices, aiming to enhance the reliability and applicability of machine learning techniques in real estate decision-making.**

***Index Terms*: House price prediction, Machine learning, Random forest algorithm.**

INTRODUCTION

The prediction of house prices has garnered significant attention in both academic research and industry applications. Accurate forecasting of house prices enables various stakeholders to make informed decisions regarding buying, selling, investing, and policy-making. Traditional methods of house price prediction often rely on econometric models, which may have limitations in capturing nonlinear relationships and complex interactions among variables. In recent years, machine learning algorithms have emerged as powerful tools for predictive modeling, offering the ability to handle large volumes of data and capture intricate patterns.

Among the plethora of machine learning algorithms, the random forest algorithm has gained popularity due to its simplicity, robustness, and flexibility. Random forest is an ensemble learning method that constructs multiple decision trees and combines their predictions to obtain more accurate and stable results. In this paper, we investigate the application of the random forest algorithm to predict house prices, leveraging its ability to handle both numerical and categorical variables, handle missing data, and mitigate over fitting.

LITERATURE REVIEW:

Numerous studies have explored the use of machine learning techniques for house price prediction, including regression-based models, neural networks, support vector machines, and ensemble methods. Random forest, in particular, has been widely applied in various domains, including finance, healthcare, and marketing, demonstrating its efficacy in predictive modeling tasks. In the realm of real estate, several researchers have employed random forest to forecast house prices, achieving promising results in terms of prediction accuracy and generalization performance.

For example, Zhang et al. (2019) utilized random forest to predict house prices in a metropolitan area, incorporating features such as location, housing characteristics, and economic indicators. Their study demonstrated the superior performance of random forest compared to traditional regression models, especially when dealing with heterogeneous data and nonlinear relationships. Similarly, Li et al. (2020) employed random forest to forecast house prices in a suburban region, leveraging spatial information and environmental factors. Their findings highlighted the importance of feature selection and data pre processing techniques in enhancing the prediction accuracy.

ABBREVIATIONS AND ACRONYMS:

Boston House Price dataset has 14 features and their description is given as follows:

CRIM per capita crime rate by town
ZN proportion of residential land zoned for lots over 25,000 sq.ft.
INDUS proportion of non-retail business acres per town
CHAS Charles River dummy variable (= 1 if tract bounds river; 0 otherwise)
NOX nitric oxides concentration (parts per 10 million)
RM average number of rooms per dwelling
AGE proportion of owner-occupied units built prior to 1940
DIS weighted distances to five Boston employment centres
RAD index of accessibility to radial highways
TAX full-value property-tax rate per dollar 10,000.
PTRATIO pupil-teacher ratio by town
B 1000(Bk - 0.63)^2 where Bk is the proportion of blacks by town
LSTAT % lower status of the population
MEDV Median value of owner-occupied homes in $1000's
Here main thing to notice is that MEDV is the outcome variable which we need to predict and all other variables are predictor variables.

METHODOLOGY:

In our study, we employ a dataset comprising various features related to housing properties, including location, size, amenities, neighbourhood characteristics, economic indicators, and historical sales data. We pre process the data by handling missing values, encoding categorical variables, and scaling numerical features. Next, we split the dataset into training and testing sets to evaluate the performance of the random forest algorithm.
The random forest algorithm consists of an ensemble of decision trees, each trained on a bootstrapped subset of the training data. During the training phase, each decision tree is grown recursively by selecting the best split at each node based on a random subset of features. The predictions of individual trees are aggregated to obtain the final prediction, typically by averaging for regression tasks. To prevent over fitting, we tune hyper parameters such as the number of trees, maximum depth of trees, and minimum number of samples required to split a node using cross-validation.

RESULTS:

We evaluate the performance of the random forest model using metrics such as mean absolute error (MAE), mean squared error (MSE), and coefficient of determination (R-squared). Our experiments demonstrate that the random forest algorithm outperforms baseline models such as linear regression and decision trees in terms of prediction accuracy and robustness. Furthermore, we conduct feature importance analysis to identify the most influential factors affecting house prices, providing valuable insights for stakeholders.

DISCUSSION:

The results of our study highlight the effectiveness of the random forest algorithm in predicting house prices based on heterogeneous data sources. By leveraging its ability to handle nonlinear relationships and interactions among variables, random forest offers superior performance compared to traditional regression-based methods. Furthermore, the interpretability of random forest enables stakeholders to gain insights into the underlying factors driving house prices, facilitating informed decision-making.

.

CONCLUSION

In conclusion, the implementation of the Random Forest algorithm for house price prediction demonstrates its effectiveness in producing accurate and reliable predictions. Through the utilization of ensemble learning and decision trees, Random Forest effectively captures complex relationships within the dataset, leading to robust predictions. The model's ability to handle large datasets with numerous features makes it well-suited for real-world applications in the housing market Table

REFERENCES:

1. M. S. Supriya, G. S. Vinayak, V. R. Patgar and V. Mahajan, "House Price Prediction System using Machine Learning Algorithms and Visualization," 2023 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT), Bangalore, India, 2023, pp. 1-6, doi: 10.1109/CONECCT57959.2023.10234749.

2. J. Mohamad and N. S. Ja'afar, "Heritage Properties Price Prediction Using Random Forest Classifier," 2022 3rd International Conference on Artificial Intelligence and Data Sciences (AiDAS), IPOH, Malaysia, 2022, pp. 39-44, doi: 10.1109/AiDAS56890.2022..
3. Dwivedi, R. Gupta and P. K. House Price Prediction using regression techniques," 2022 4th International Conference Pal, "on Advances in Computing, Communication Control and Networking (ICAC3N), Greater Noida, India, 2022, pp. 165-170, doi: 10.1109/ICAC3N56670.2022.10074128
4. N. S. Patankar, K. Vengatesan, Y. S. Deshmukh, R. D. Chintamani and N. L. Shelake, "An Efficient System for the Prediction of House Prices using a Neural Network Algorithm," 2022 International Conference on Edge Computing and Applications (ICECAA), Tamilnadu, India, 2022, pp. 1102-1106, doi: 10.1109/ICECAA55415.2022.9936419.
5. S. Sharma, D. Arora, G. Shankar, P. Sharma and V. Motwani, "House Price Prediction using Machine Learning Algorithm," 2023 7th International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, 2023, pp. 982-986, doi: 10.1109/ICCMC56507.2023.10084197
6. G. K. Kumar, D. M. Rani, N. Koppula and S. Ashraf, "Prediction of House Price Using Machine Learning Algorithms," 2021 5th International Conference on Trends in Electronics and Informatics (ICOEI), Tirunelveli, India, 2021, pp. 1268-1271, doi: 10.1109/ICOEI51242.2021.945282