IJNRD.ORG ISSN : 2456-4184
INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG
An International Open Access, Peer-reviewed, Refereed Journal

HOUSE PRICE PREDICTION USING MACHINE LEARNING

UJJWAL KUMAR, RISHU KUNWAR, DR. NEHA GARG.
[1,2] *STUDENT, DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING*
FACULTY OF ENGINEERING AND TECHNOLOGY
MANAV RACHNA INTERNATIONAL INSTITUTE OF RESEARCH AND STUDIES, FARIDABAD, HARYANA, INDIA
[3]*ASSISTANT PROFESSOR, DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING*
FACULTY OF ENGINEERING AND TECHNOLOGY
MANAV RACHNA INTERNATIONAL INSTITUTE OF RESEARCH AND STUDIES, FARIDABAD, HARYANA, INDIA

Abstract

This paper presents a comprehensive overview of a website designed to predict house prices based on the past market trends, also aims to address the challenges faced by both buyers and sellers in the real estate market. Buyers focus on finding a suitable home/flat within their budget, and consider their investment on house to increases over a period of time. On the other hand, Sellers aim to sell their homes at the best possible price. Since house prices are subject to fluctuations, customers often face difficulties in purchasing a house at the right time before prices change in the near future. To address this major issue in the real estate market, we are designing a machine learning model for predicting house prices. Machine learning techniques play a vital role in this project by providing more precise house price estimations based on user preferences such as location, number of rooms, and air quality, among others.

The project will employ various machine learning algorithms, including Linear Regression, Decision Tree, Random Forest, and Support Vector Regression. Ultimately, this solution will enable both buyers and sellers to negotiate their priorities more efficiently, minimizing financial and time losses.

Keywords : ***Machine Learning, Random Forest, Linear Regression , Decision Tree, Support Vector Regression (S.V.R).***

I. INTRODUCTION

The House Price Index (H.P.I) serves as a metric for measuring the average price changes of houses in numerous countries. such as the U.S. Federal Housing Finance Agency H.P.I and the U.K. Right Move H.P.I[1]. However, a challenge arises when attempting to predict the price of a specific house while considering variables like its age, location, and number of rooms, rather than solely relying on repeat sales from previous years. To address this significant issue, machine learning techniques are employed to provide more accurate price predictions by incorporating attributes like rooms, location, and age. Additionally, previous year data is utilized to train the model and generate precise predicted values[2][3][4][5].

In this project, the machine learning process involves the collection of data from various sources, and a step-by-step approach utilizing machine learning algorithms, from data collection to generating the predicted output.

II. LITERATURE REVIEW

We are conducting an analysis of various Machine Learning algorithms in this project to enhance the training of our Machine Learning model. The study focuses on housing cost trends, which serve as indicators of the current economic situation and have direct implications

for buyers and sellers. The actual cost of a house depends on numerous factors, including the number of bedrooms, bathrooms, and location. In rural areas, the cost tends to be lower compared to cities. Additionally, factors such as proximity to highways, malls, supermarkets, job opportunities, and good educational facilities greatly influence house prices.

To address this issue, our research paper presents a survey on predicting house prices by analyzing given features. We employed different Machine Learning models, including Linear Regression, Decision Tree, and Random Forest, to construct a predictive model with their working accuracy. Our approach involved a step-by-step process, encompassing Data Collection, Pre-Processing Data, Data Analysis to Model Building

III. METHODOLOGY / WORKFLOW

The workflow of process has been shown in figure 1 and discussed as follow-

III.A. Data collection

The process of data collection entails gathering a dataset that contains important variables such as the number of bedrooms, location, bathrooms, etc., along with their corresponding sale prices [9].

III.B Data preprocessing

Data preprocessing involves clean and preprocessing the data to handle missing values and categorical variables which involves techniques like imputation, normalization and feature scaling.

III.C. Feature selection

Feature selection basically involves analyzing the dataset and select the most relevant features that are likely to have significant impact on house prices(e.g. location).

III.D. Splitting the dataset

It split the dataset into training and testing sets, in which TrainingDataset is used to train the machine learning model, while Testing Dataset is used to evaluate the performance.

III.E. Model training

It will choose a suitable learning algorithm for regression (linear regression or random forest) and train it using the training dataset.

III.F. Model evaluation

The evaluation of the trained model will be conducted by employing suitable metrics like Mean Squared Error (MSE) and Root Mean Squared Error (RMSE). These metrics will assess the performance and accuracy of the model.

III.G.. Model tunning

In this step, the parameters of the model are adjusted in order to discover the optimal configuration that minimizes the discrepancy between the predicted prices and the actual prices.

III.H. Model prediction

Now after getting the model satisfaction data, we can use it to givenew predicted final data based on new input involving the training data[6].

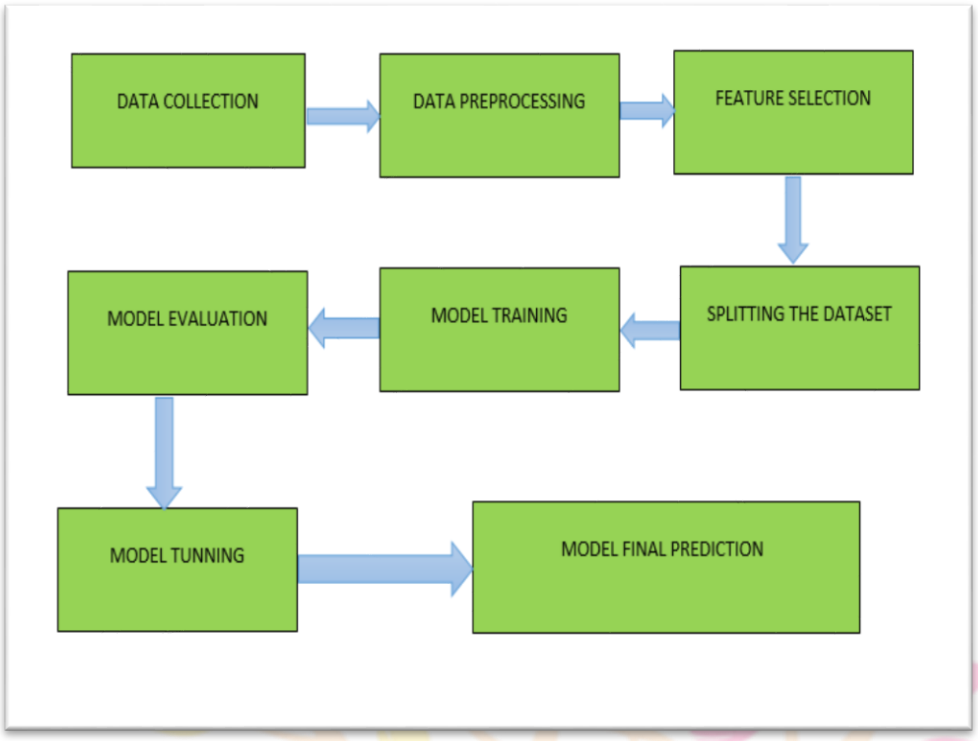

Fig 1 Methodology.

Proposed Methodology-

The method utilized the housing dataset available on Kaggle and provides a comparison of the performances of four machine learning algorithms [7]. The dataflow diagram is shown in figure 2. There are four machine learning algorithms linear regression, random forest, Decision tree and support vector regression. One with best accuracy will be used for predicting house price[8].

IV. MACHINE LEARNI NG ALGORITHMS

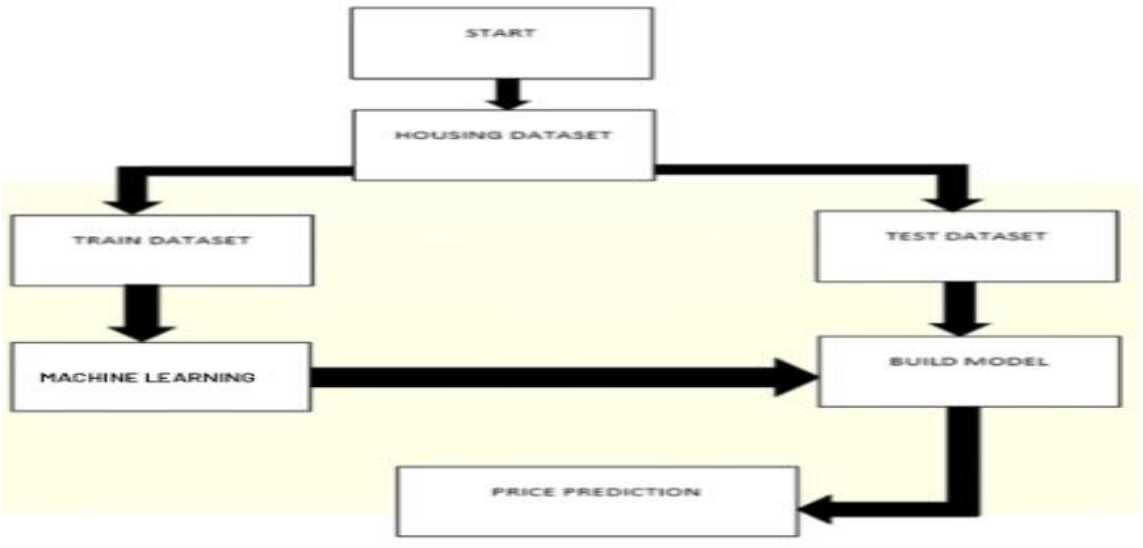

Fig. 2 Machine Learning Algorithm for Proposed Method

IV.A. Linear regression

Linear regression used to predict the value of a variable based on the value of another variable. The variable we want to predict is called the dependent variable. The variable we are using to predict the other variables value is called the independent variable.

IV.B. Decision tree

Decision tree is a non-parametric supervised learning algorithm, which is utilized for both classification and regression tasks. In this,a flowchart-like structure where each internal node represents a feature, each branch represents a decision, and each leaf node represents an outcome or prediction.

IV.C. Support Vector Regression (S.V.R)

Support Vector Regression is a supervised learning algorithm that is used to predict discrete values. The goal is to find a hyperplane that best fits the training data while also controlling the margin or deviation of the data points from the hyperplane

IV.D. Random forest

It is a Supervised Machine Learning Algorithm that is used widely in Classification and Regression problems. It builds decision trees on different samples and takes their majority vote for classification and average in case of regression.

V. PERFORMANCE

The performance of the algorithms evaluated using the R-squared value, which measures the amount of variance explained by the dependent variable [6]. A higher R-squared value indicates better algorithm performance. Upon comparing the values, it can be concluded that Random Forest outperforms the other three algorithms, achieving an R-squared value of 0.90. In contrast, the remaining algorithms, namely Support Vector Machines (SVM), 2).

Linear Regression, and Decision Tree, achieved R-squared values of 0.79, 0.86, and 0.77 respectively [6], as shown in figure 3.

CLASSIFIER	ACCURACY(%)	SENSTIVITY(%)	SPECIFICITY(%)	FINAL AUC
S.V.M	82.16	66.10	91.90	0.79
LINEAR REGRESSION	82.16	69.80	89.70	0.86
RANDOM FOREST	84.30	71.40	71.40	0.90
DECISION TREE	78.51	68.50	84.60	0.77

Figure 3 Ml Algorithms Statistics.

Therefore, Random Forest demonstrates the highest performance with an AUC of 0.90, making it the preferred choice for predicting house prices using the collected data [6].

VI. CONCLUSION

In summary, this research paper concludes that employing the Random Forest machine learning algorithm will lead to more precise pricing predictions using the collected data, which serves as the core component of machine learning. Subsequently, the trained data can be utilized to generate new predicted values. This approach offers cost savings and reduces the need for extensive physical efforts for both buyers and sellers. Notably, Random Forest predicted house prices with the highest accuracy value of 0.90.

REFERENCES

[1]House Price Index, *Federal Housing Agency .https://www.fhfa.gov/* (accessed,February,2023).

[2] Quang Truong, Minh Nguyen, Hy Dang, Bo Mei – "*House Price Prediction via Improved Machine Learning Techniques*" ,2019,United States.

[3]Fan C,Cui Z,Zohng X ,*House Prices Prediction With machine learning Algorthms* ,2018,ICMLC.

[4]Phan TD ,*Housing Price Prediction Using achine Learning Algorithms,Australia*,2018,ICMLDE.

[5]MU J, Wu F,Zhang A, *housing value Forcasting based on Machine Learning Methods,Abstract and Applied Analysis* 2014.

[6]Anand G. Rawool,Dattatray V.Rogye,Sainath.Rane,Sr.Vinayak A.Bharadi – "*House Price Prediction Using Machine Learning*",2021,Mumbai University.

[7] Mitchell, T. M. (2018). *Machine Learning* (1st ed., pp. 1-10). Mcgraw Hill Education. https://www.cin.ufpe.br/~cavmj/Machine%20-%20Learning%20-%20Tom%20Mitchell.pdf

[8] Saleh, H. (2018*). Machine Learning Fundamentals: Use Python and scikit-learn to get up and running with the hottest developments in machine learning*. *Packt Publishing Ltd*.".

[9] M. Jain, H.Rajput,N.Garg and p.Chawla,"*Prediction Of House Pricing using Machine Learning with python*,"2020 International Conference On electronics and Sustainable CommunicationSystems(ICESC),Coimbatore,India,2020,pp.570-574,doi: 10.1109/ICESC48915.2020.9155839.