

House Price Prediction Using Machine Learning

Dev Mulchandani¹

UG Student

Department of Information Technology
G H Rasoni College of Engineering Nagpur

Swati Shamkuwar²

Professor

Department of Information Technology
G H Rasoni College of Engineering Nagpur

Abstract:- This project provides us an overview on how to predict house prices using various machine learning models with the help of different python libraries. This proposed model considers as the most accurate model used for calculating the house price and provides a most accurate prediction. This provides a brief introduction which will be needed to predict the house price. This project consists of what and how the house price model works with the assistance of machine learning technique using scikit-learn and which datasets we will be using in our proposed model. Predicting the price of a house helps for determine the selling price of the house in a particular region and it help people to find the correct time to buy a home. In this task on House Price Prediction using machine learning, our task is to use data to create a machine learning model to predict house prices in the given region. We will implement a linear regression algorithm on our dataset. By using real world data entities, we are going to predict the price of the house in that area. For better results we require data pre-processing units to improve the efficiency of the model. for this project we are using supervised learning, which is a part of machine learning. We have to go through different attributes of the dataset

I. INTRODUCTION

In recent years, the real estate market has witnessed unprecedented fluctuations, driving both excitement and apprehension among buyers, sellers, and investors alike. Amidst this dynamic landscape, the ability to accurately predict house prices has emerged as a pivotal tool for informed decision-making and strategic planning within the housing sector. Recognizing the complexities inherent in price determination, researchers and practitioners have turned to advanced computational methods, particularly machine learning, to unravel the intricate patterns underlying housing market dynamics.

This research paper endeavors to contribute to the burgeoning literature on house price prediction by offering a comprehensive analysis of state-of-the-art machine learning techniques in this domain. By leveraging rich datasets comprising diverse features such as location, property characteristics, economic indicators, and demographic information, we aim to develop robust models capable of forecasting house prices with a high degree of accuracy.

II. LITERATURE SURVEY

Home prices reflect current market conditions and are a concern for buyers and sellers. Location also plays a role for many reasons. Manually estimating home prices is a difficult task and is often inaccurate, so many machines have been developed to predict home prices. Sifei Lu, Zengxiang Li, Zheng Qing, Xulei Yang, Rick Siow Mong Goh proposed a high-level real estate estimate using horizontal lines. can give us a good estimate of home prices based on other variables. They used linear regression on Ames dataset, so its accuracy is high. Project estimation house price has two modules, admin and user. Admins can add locations and view locations. Admins have the right to increase speed according to the area of the room. Users can search for locations and see estimated house prices for that location. Home prices reflect current market conditions and are a concern for buyers and sellers. There are many factors that affect the price of a home, such as the number of bedrooms and bathrooms, and the price of a home also varies depending on its location. Manually predicting home prices is difficult and often not very good.

Therefore, many systems have been developed to estimate housing prices. In this research, these elements are divided into three main categories: existing status, idea and original. Our property is limited by the human dimensions of the house, such as the size of the house, the number of rooms, the kitchen and parking space, the openness of the children's playground, the area of the land and buildings. The age of the building and the idea are the ideas that architects put forward to attract buyers, such as the possibility of a small house, sustainable and green products and good conditions. The location of the house has a great impact on its value. This is because the park chooses standard land prices. The district also selected important streets where businesses such as schools, parks, crisis centers, and welfare centers are located, as well as family businesses such as shopping malls, gourmet tours, or nicer places. Accuracy of home price predictions. Fairfax County urban housing data was analyzed and the classification accuracy of various algorithms was compared. Then, a better prediction model was developed to help real estate agents by supporting decision making based on home value assessment.

➤ Proposed System

- The proposed approach involves a comprehensive methodology, including:
- Data collection and preprocessing for creating a clean and standardized dataset.
- Feature engineering to enhance the model's predictive capabilities.
- Evaluation and selection of regression algorithms through rigorous testing.
- Incorporation of interpretability features for user trust and understanding.
- Development of a user-friendly interface for model integration.
- Continuous monitoring and updating mechanisms for model adaptation.

III. METHODOLOGY

➤ Data Collection:

Gather comprehensive information that includes a variety of characteristics relevant to your home value estimate, including property characteristics (For example size , age , number of bedrooms/bathrooms) , location (For example , neighborhood , proximity to amenities) , economic measures (e.g., interest rates, unemployment rate), demographics (e.g., census rate, median income), and historical sales data. complete and consistent.

➤ Data Preprocessing:

Cleanse data using techniques such as imputation, outlier detection, and data validation to resolve missing values, outliers, and inconsistencies. Input categorical variables using methods such as single-bit encoding or label encoding to convert them into numerical representations suitable for modeling.

➤ Feature Engineering:

Conduct exploratory data analysis (EDA) to gain insight into the distribution and relationship between different features.

Create new features by combining, transforming, or extracting data from existing variables to capture relationships and patterns in the data.

Includes empirical information and expert opinions to develop factors that are good predictors of home prices, such as neighborhood characteristics, housing market conditions, and employment.

➤ Model Development:

Experiment with various machine learning algorithms, including transformation models (for example , linear regression, ridge regression, lasso regression), tree methods (for example , decision tree, random forest, gradient boosting) , support vector machines (SVM), and neural networks. To train and evaluate the performance of different models, split the dataset into training set, validation set, and test set as appropriate

➤ Model Evaluation:

The performance of each model is assessed by evaluating appropriate parameters such as mean error (MAE), mean square error (MSE), root mean square error (RMSE) and coefficient of determination (R-squared). Or the process of cross-checking different models to determine the best way to predict house prices.

➤ Interpretability and Validation:

Interpret the results of training models using techniques such as factor analysis, partial dependency plots, and nonlinear interpretation models such as SHAPley Contribution Explanations to understand that it does not affect the price of the house. Non-standard data or real-world testing to ensure it is reliable and usable in real-world applications. Evaluation of results to promote innovation and transparency.

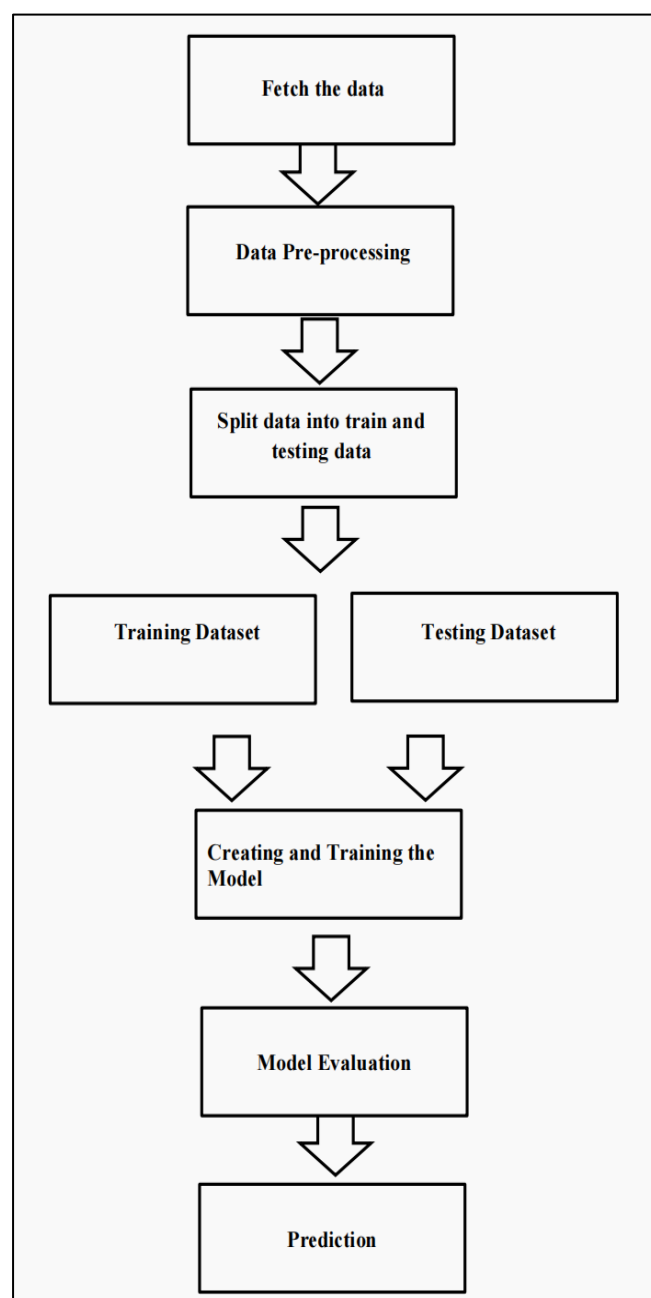


Fig 1 Interpretability and Validation

IV. CONCLUSION

Through empirical analysis and comparative studies, we have demonstrated the effectiveness of various algorithms in accurately forecasting house prices, emphasizing the importance of robust feature selection, model optimization, and interpretability. Our findings underscore the interdisciplinary nature of house price prediction, highlighting the need for collaboration between data scientists, domain experts, policymakers, and industry practitioners to inform evidence-based decision-making and drive sustainable development in the housing sector. Looking forward, future research directions may focus on refining model interpretability techniques, addressing data quality challenges, and integrating alternative data sources to further advance our understanding and predictive capabilities in the dynamic real estate market landscape.

REFERENCES

- [1]. Housing Price Prediction Using Machine Learning Algorithms: The Case of Melbourne City, Australia, The Danh Phan, 2018 International Conference on Machine Learning and Data Engineering (iCMLDE)
- [2]. Predicting Sales Prices of the Houses Using Regression Methods of Machine Learning, Parasich Andrey Viktorovich ; Parasich Viktor Aleksandrovich ; Kaftannikov Igor Leopoldovich ; Parasich Irina Vasilevna, 2018 3rd Russian-Pacific Conference on Computer Technology and Applications (RPC)
- [3]. Real Estate Value Prediction Using Linear Regression, Nehal NGhosalkar ; Sudhir N Dhage, 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA)
- [4]. Predicting Housing Market Trends Using Twitter Data, Marlon Velthorst ; Cicek Güven, 2019 6th Swiss Conference on Data Science (SDS)
- [5]. House Price Prediction Using Machine Learning and Neural Networks, Ayush Varma ; Abhijit Sarma ; Sagar Doshi ; Rohini Nair, 2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT)
- [6]. Forecasting house price index of China using dendritic neuron model, Ying Yu ; Shuangbao Song ; Tianle Zhou ; Hanaki Yachi ; Shangce Gao, 2016 International Conference on Progress in Informatics and Computing (PIC) 41
- [7]. Prediction of real estate price variation based on economic parameters, Li Li ; Kai-Hsuan Chu, 2017 International Conference on Applied System Innovation (ICASI)
- [8]. Predicting house sale price using fuzzy logic, Artificial Neural Network and K-Nearest Neighbor, Muhammad Fahmi Mukhlisin ; Ragil Saputra ; Adi Wibowo, 2017 1st International Conference on Informatics and Computational Sciences (ICICoS)
- [9]. Comprehensive Analysis of Housing Price Prediction in Pune Using Multi-Featured Random Forest Approach, Rushab Sawant ; Yashwant Jangid ; Tushar Tiwari ; Saurabh Jain ; Ankita Gupta, 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA)
- [10]. Time-Aware Latent Hierarchical Model for Predicting House Prices, Fei Tan ; Chaoran Cheng ; Zhi Wei, 2017 IEEE International Conference on Data Mining (ICDM)