House Price Prediction System

Sahil Datir[1] Siddhant Pande[2] Shreyash Jawalkar[3] Prof.Sonali Guhe[4]
[4]Assistant Professor
[1,2,3,4]Department of Information Technology
[1,2,3,4]G.H.Raisoni College of Engineering, Nagpur, India

Abstract — This paper proposes a model that considers all the required factors so that the family should get a perfect house according to their needs. Weights are automatically assigned to transaction data and important features of all of the house buyers are identified by an attention mechanism. The model collects data from public facilities such as schools, colleges, and metro stations, it uses satellite maps to analyze the environment around the house. To improve prediction accuracy the model considers the interaction between two different features. To make the model more user-friendly we have also added some new features for house pricing.

Keywords: Multiple Linear Regression, Public Facilities, Hypothesis

I. INTRODUCTION

Predicting house prices is a challenging task due to the different methods used by different people and the varied parameters that can influence the results. The demand for houses that meet people's needs and showcase their high standards has been on the rise. In India, people do a lot of research before finalizing a particular house, and making a house pricing system is an essential tool for making informed decisions.

This paper proposes a house pricing system that uses multiple linear regression to predict house prices. The system considers two groups of factors that affect house prices: house conditions and surrounding environmental conditions. The well developed and premium surrounding that offer everything near the house costs more. As all of these factors are very essential and greatly affect house pricing we must have all the data on all these things so that when we search for a house with such factors it would not take long to find the houses. The authors collect all the different types of data in their database so that all the features are properly working and we can summarize the important features for buyers. They use a spatial transformer network to fetch the features of the image. By using satellite, house transaction data, metro stations, railway stations, bus stations, parks, schools, and other public places authors test their hypothesis. All house buyers have their own needs and requirements if they want nearby grocery shops, laundry, hardware shops, shopping malls, schools, and colleges for their families. If most of the features that are required by the buyer match only then we can suggest the house to the buyer and he will be interested in buying the house.

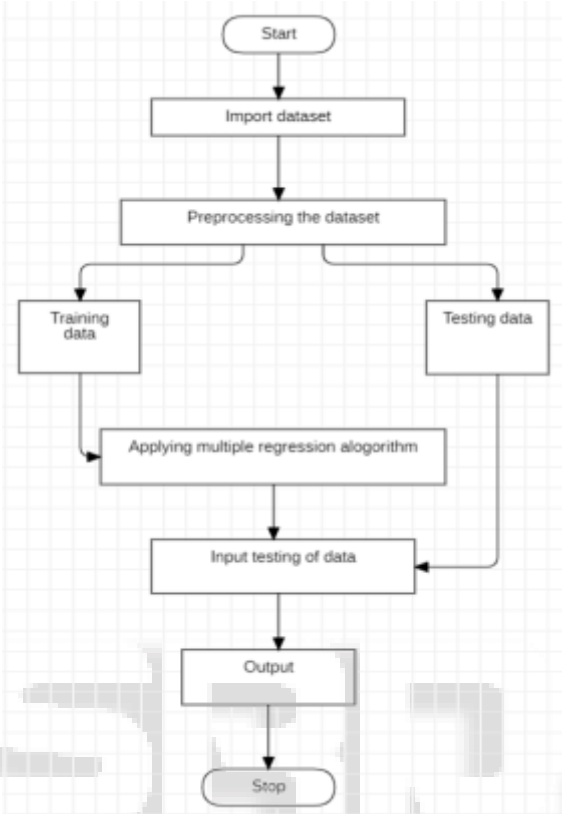

Fig. 1: Model Work Flow

We compare the performance of these models with other deep learning models and machine learning models. The model can learn the features and also sort and make a summary for the house buyers so that the accuracy of prediction can be increased and investigate the factors that influence the house pricing. According to the results, the model outperforms all the other models and has higher accuracy than any other model.

This paper has a great approach to house price prediction by comparing both environmental conditions and the surrounding environment of the house. This system helps the buyers and proposes suggestions that match their requirements the most. By using the public surrounding environment and satellite maps and other related factors as input, the model can learn the features also sort and make a summary for the house buyers so that the accuracy of prediction can be increased and investigate the factors that influence the house pricing.

II. METHODOLOGY:

Multiple regression is a mathematical process used to identify the relationship between one dependent variable and one or more independent variables. In Predicting house prices model dependent variable Y is the price and the independent variable X is the features of the house. The major aim of Multiple regression is to predict the value of the dependent

variable (price) according to the independent variables (features). The value of independent variables is already present in the dataset.

Multiple regression is mathematically represented as an equation

$$Y = p + q1X1 + q2X2 + \ldots.. + qnXn$$

Here Y is the dependent variable, which is unknown, whose value is not present in the dataset, and X1,..., Xn is n number of independent variables, which is known, whose value is present in the dataset. When calculating the values of, q1, ..., qn, regression analysis makes sure that the dependent variable (price) is maximally predicted from the dataset of independent variables. Data Collection: Collect data on house prices, demographic information of the region, and other relevant variables that can impact the price of a house. Ensure that the data collected is sufficient and representative of the population being studied.

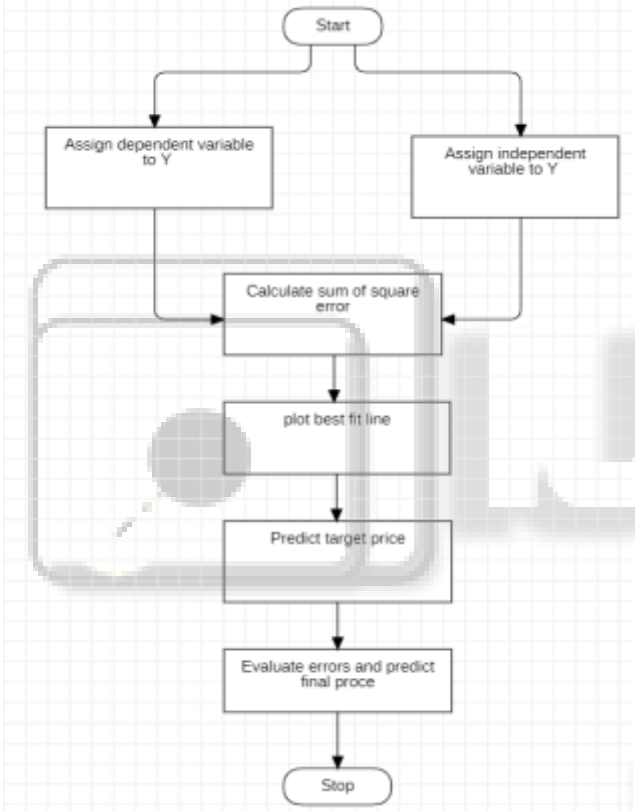

Fig. 2: Algorithm

A. *Data Preparation:*

Clean and preprocesses the data. This step involves checking for missing values, handling outliers, and normalizing the data into a format that is suitable for analysis. Identify independent variables: Identify the independent variables that can impact the price of a house. These variables can include the location of the house, the age of the house, the size of the house, the number of rooms, the crime rate in the area, and other relevant factors.

B. *Select the dependent variable:*

Select the dependent variable, which is the house price, and ensure that it is continuous and normally distributed. Model Selection: Choose the appropriate regression model based on the nature of the data and the research question. In this case, multiple regression would be suitable as various independent variables impact dependent variables. Build the model: Evaluate the model coefficient of the model using methods such as least squares estimation. Evaluate the model: Evaluate the model using measures such as root mean square error(RMSE), Root-squared, adjusted Root-squared, and many more. These measures indicate how better the model will fit the data and how accurately it can predict house prices. Interpret the results: Interpret the results of the regression analysis, including the coefficients, their significance, and their interpretation. This information can provide insights into the factors that impact house prices. Test the model: Test the model using a holdout dataset or cross-validation to ensure that it can generalize well to new data.

III. Conclusion:

In this study, we have developed a house pricing system using some different ideas that are not available in the market and we believe that they will be more helpful in this system. We also tried to do something different than the normal linear regression technique. Instead, we used multiple linear regression so that the selected factors give effective results.

IV. System Framework:

1) Import Dataset and process
2) Fill the requirements and all the needed features.
3) Calculate the sum of error
4) Calculate price of plot according to requirements

Reference:

[1] R. Manjula et al., "Real estate value prediction using multivariate regression models", *Materials Science and Engineering Conference Series*, vol. 263, no. 4, 2017.
[2] N. Jain, P. Kalra and D. Mehrotra, "Analysis of Factors Affecting Infant Mortality Rate Using Decision Tree in R Language" in Soft Computing: Theories and Applications, Singapore:Springer, pp. 639-646, 2019.
[3] House Price Prediction using a Machine Learning Model: A Survey of Literature Nor Hamizah Zulkifley, Shuzlina Abdul Rahman, Ubaidullah Nor Hasbiah, Ismail Ibrahim https://www.researchgate.net/publication/347584803_House_Price_Prediction_using_a_Machine_Learning_Model_A_Survey_of_Literature
[4] Employing Machine Learning for House Price Prediction priyansh varma, vidhu gaba, Nikita Malik https://www.researchgate.net/publication/344099603_Employing_Machine_Learning_for_House_Price_Prediction
[5] House Price Prediction Using Regression Techniques: A Comparative Study https://ieeexplore.ieee.org/document/8882834
[6] House Price Prediction Using Regression Techniques: A Comparative Study https://ieeexplore.ieee.org/document/8882834
[7] Induced stress and particle behavior of the electro-rheological suspensions at higher shear rates probed by electrical and mechanical responses https://ieeexplore.ieee.org/document/972552
[8] Machine Learning based House Price Prediction Model https://ieeexplore.ieee.org/document/9936336

