

House Price Prediction using Machine Learning

A Project Report

Submitted in partial fulfillment of the requirements

Of

Techshaksham Certificate

By

Magesh S

Gunalakshmi N S

Pavithra P

Joselin Sophia Rani J

Under the Esteemed Guidance of

Prof. Raja P

ACKNOWLEDGMENT

We have taken efforts in this project “House Price Prediction”. We with immense pleasure and commitment would like to present the project assignment.

We are highly indebted to Prof. RAJA P and for their guidelines and constant supervision as well as for providing necessary information regarding the project and also their support in completing the project.

We are sincerely thankful to the teammates who helped us complete the project.

We would like to thanks Government College of engineering Tirunelveli for adding the course Artificial intelligence using cloud computing.

ABSTRACT

House Price Prediction (HPP) is commonly used to estimate the changes in housing price. Since housing price strongly correlated to other factors such as location, area, population, it requires other information apart from HPP to predict individual housing price. House Price increase every year, so there is a need for a system to predict House prices in the future. We aim to make evaluations based on every basic parameter that is considered while determining the price.

CHAPTER 1

INTRODUCTION

Problem Statement:

- Problem faced during buying a house
 - Buying a house is a stressful thing.
 - Buyers are generally not aware of factors that influence the house prices.
 - Many problems are faced during buying a house.
- Hence real estate agents are trusted with the communication between buyers and sellers as well as laying down a legal contract for the transfer. This just create a middle man and increases the cost of houses.

Problem Definition:

- In India, there are multiple real estate classified websites where properties are listed for sell/buy/rent and more. However, in each of these websites we can see lot of inconsistencies in terms of pricing of an apartment and there are some cases when similar apartments are priced differently and thus there is lot of in-transparency.
- Sometimes the consumers may feel the pricing is not justified for a particular listed apartment but there no way to confirm that either. Proper and justified prices of properties can bring in a lot of transparency and trust back to the real estate industry, which is very important as for most consumers especially in India the transaction prices are quite high and addressing this issue will help both the customers and the real estate

industry in the long run. We propose to use machine learning and artificial intelligence techniques to develop an algorithm that can predict housing prices based on certain input features.

Expected Outcomes:

- The proposed solution for the Bangalore House Price Prediction involves implementing a supervised machine learning algorithm, such as regression model. The data set will be pre-processed to handle missing values outlier and categorical variables. The dataset will then be split into training and testing sets to train and evaluate the model's performance. Features engineering may be applied to extract relevant information from the available features. Various regression algorithms such as linear regression, decision tree or ensemble methods like random forest will be tested to determine the most accurate model. Hyper parameter tuning and cross-validation technique will be employed to enhance the model's robustness. The selected model will be deployed to predict house price based on input features.

CHAPTER 2

LITERATURE SURVEY

House Price Prediction using Machine learning algorithms:

The demand for renting and owning homes has increased as a result of increased urbanization. As a result, figuring out a better technique to computer property prices that truly represent market prices has become a popular issue. The research focuses on applying machine learning methods such as simple linear regression (SLR), multiple linear regression (MLR), and neural networks to properly determine the house price (NN). The algorithm with the lowest mean square error (MSE) is picked as the best for estimating the price of a property. This will assist both sellers and buyers in determining the optimal price for House.

CHAPTER 3

PROPOSED METHODOLOGY

System Design

- DATA INGESTION AND PREPROCESSING
- TRAINING PIPELINE
- MODEL PERSISTENCE
- DEPLOYMENT
- USER INTERFACE

Modules Used:

- Pandas - for data manipulation
- Scikit-learn for machine learning algorithm
- Linear regression for the linear regression model.

PROCESS:

DATA COLLECTION:

- Gather a comprehensive dataset with relevant features such as property size, location, amenities, and historical pricing.

DATA PREPROCESSING:

- Handle missing values, outliers, and encode categorical variables.
- Explore feature engineering techniques to extract valuable information.

SPLIT THE DATA:

- Divide the dataset into training and testing sets.
- Typically 80-20 or 70-30 splits are common.

TRAIN RIDGE REGRESSION MODEL:

- Implement Ridge Regression using a machine learning library like scikit-learn in Python.
- Tune hyper parameters, especially the regularization parameter (alpha) using cross-validation.

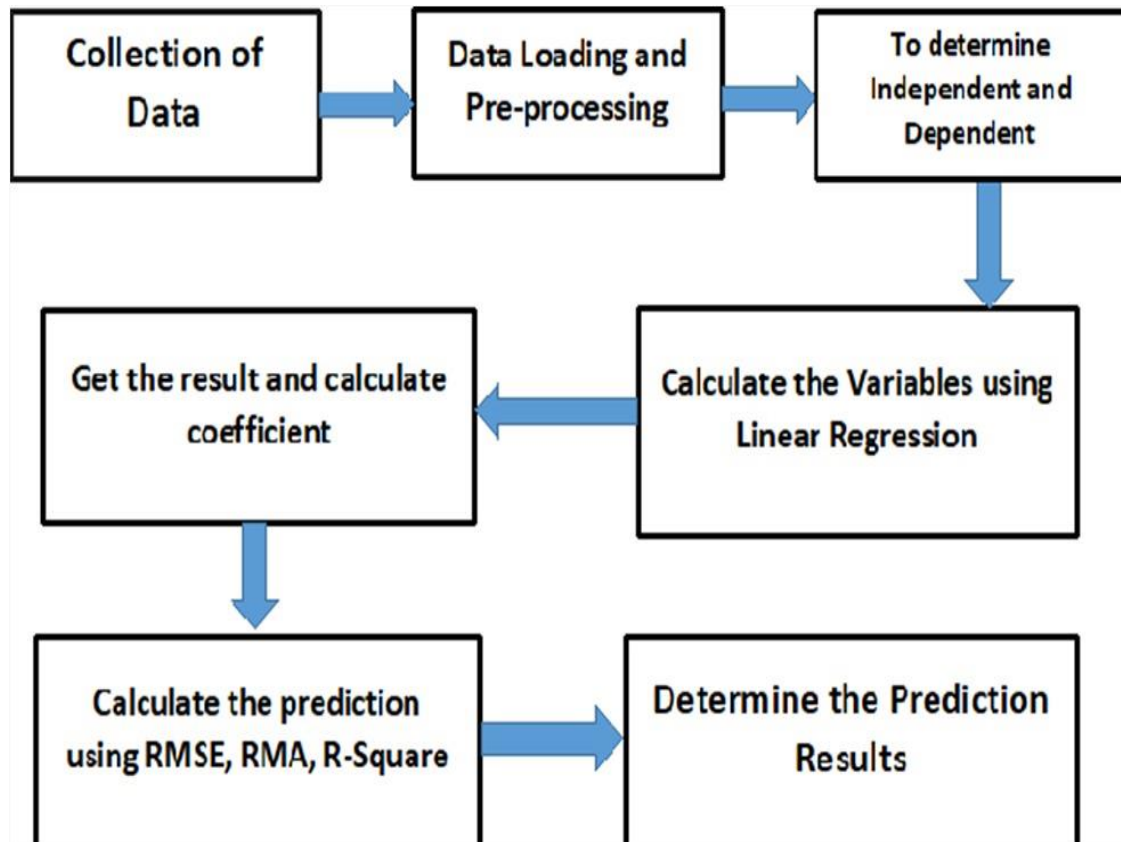
MODEL EVALUATION:

- Assess the model's performance on the testing set using metrics like Mean Squared Error or R-squared.

PERSIST THE MODEL:

- Save the trained Ridge Regression model to a file using serialization (e.g., Jollibee or pickle) for later deployment.

Data Flow Diagram



Advantages

House Price Prediction can help the developer determine the selling price of a house and can help the customer to arrange the right time to purchase a house. There are three factors that influence the price of a house which include physical conditions, concept and location.

Requirement Specification

Software Requirements:

- **Flask** used to create online apps. It offers a straightforward and adaptable method for developing Python-based web applications and APIs (Application Programming Interfaces).
- **Pandas** used for working with data sets. It has functions for analysing, cleaning, exploring, and manipulating data.
- **Numpy** used to perform a wide variety of mathematical operations on arrays.
- **Scikit-learn** used to implement machine learning models and statistical modelling.
- **Pickle** used to serializes the object first before writing it to a file.
- **Requests** is a library for making HTTP requests. It provides an easy-to-use interface that makes working with HTTP very simple.

CHAPTER 4

IMPLEMENTATION and RESULT

Collection of data:

Data processing techniques and processes are numerous. We collected data for USA/Mumbai real estate properties from various real estate websites. The data would be having attributes such as Location, carpet area, built- up area, age of the property, zip code, price, no of bedrooms etc. We must collect the quantitative data which is structured and categorized. Data collection is needed before any kind of machine learning research is carried out. Dataset validity is a must otherwise there is no point in analyzing the data.

Data preprocessing:

Data preprocessing is the process of cleaning our data set. There might be missing values or outliers in the dataset. These can be handled by data cleaning. If there are many missing values in a variable we will drop those values or substitute it with the average value.

Training the model:

Since the data is broken down into two modules: a Training set and test set, we must initially train the model. The training set includes the target variable. The decision tree regression algorithm is applied to the training data set. The decision tree builds a regression model in the form of a tree structure.

Testing and integration with UI:

The trained model is applied to test dataset and house price are predicted. The trained model is then integrated with the front end using Flask in python.

CHAPTER 5

CONCLUSION

In conclusion, using machine learning in python is a powerful tool for predicting house prices. By gathering and cleaning data, visualizing patterns, and training and evaluating our models, we can make informed decisions in the dynamic world of real estate.

By leveraging advanced algorithms and data analysis, we can make accurate predictions and inform decision-making processes. This approach empowers buyers, sellers, and investors to make informed choices in a dynamic and competitive market, ultimately maximizing their opportunities and outcomes.

ADVANTAGES:

House Price Prediction can help the developer determine the selling price of a house and can help the customer to arrange the right time to purchase a house. There are three factors that influence price of a house which include physical conditions, concept and location.

SCOPE:

In the future, the GUI can be made more attractive and interactive. It can also be turned into any real estate sale website where sellers can give the details and house for sale and buyers can contact according to the details given on the website. To simplify it for the user, there can also be a recommending system to recommend real estate properties to the user based on the predicted price. The current dataset only

includes a few locations of Bangalore city, expanding it to other cities and states of India is the future goal.

REFERENCES

1. Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems, Third Edition, October, 2022.
2. Maharshi Modi, Ayush Sharma, Dr. P. Madhavan Applied Research on House Price Prediction Using Diverse Machine Learning Techniques, International Journal of Scientific & Technology Research Volume 9, Issue 04, April 2020.
3. Dr. M. Thamarai, Dr. S P. Malarvizhi House Price Prediction Modeling Using Machine Learning, I.J. Information Engineering and Electronic Business, 2020, 2, 15-20.

GITHUB LINKS USED FOR REFERENCE:

1. <https://github.com/Shreyas3108/house-price-prediction>
2. <https://github.com/JangirSumit/BostonHousePricePrediction>

Project Source Code

<https://github.com/Magesh20103/HousePricePredictor>