

# House Price Prediction

## Micro Project for Practical Machine Learning

*by*

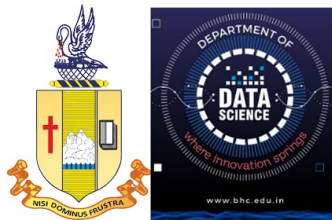
**Charumathi M**

**225229106**

*Submitted To*

**Dr. K. RAJKUMAR**

**Course Instructor**



**DEPARTMENT OF DATA SCIENCE  
BISHOP HEBER COLLEGE (AUTONOMOUS)  
TIRUCHIRAPPALLI 620017**

**MARCH 2023**

# CERTIFICATE

I hereby acknowledge that this project is the original work done by me for the requirements for Micro Project in Practical Machine Learning Course. This micro project is not copied from internet or whatsoever.

Tiruchirappalli

20 March 2023

Your Name and signature

# Table of Contents

Chapter	Title
1	Abstract
2	Background and Motivation
3	Problem Statement and Dataset Description
4	Existing Methodology
5	Proposed Methodology and Solution
6	Model Implementation
7	Testing and Evaluation
8	Model Deployment
9	Model Archival in github and Demo in Youtube
10	Conclusion and Future Work
11	References

## **Chapter 1. ABSTRACT**

House price prediction is an important problem in the field of real estate. With the increase in availability of data, machine learning algorithms have become increasingly popular for addressing this problem. In this study, we investigate the use of machine learning algorithms to predict house prices . The dataset consists of various features such as the number of bedrooms, location, and square footage of the house. We use supervised regression techniques such as linear regression and support vector regression to train our models on this dataset. We compare the performance of these models and provide insights on their strengths and weaknesses. Our results indicate that machine learning algorithms can accurately predict house prices and can be a valuable tool for real estate professionals and home buyers.

## **Chapter 2. Background and Motivation**

House price prediction is an important topic for real estate investors, agents, and other stakeholders. The motivation for predicting house prices is mainly driven by the desire to have accurate and reliable estimates of housing prices in a particular location, which can help people make informed decisions. Accurate predictions of house prices can also help investors and homebuyers identify potential areas for profitable investment or guide them in making well-informed decisions when dealing with real estate.

In machine learning, house price prediction can be approached as a regression problem, where the model is trained with historical data on housing prices and features such as location, property size, number of rooms, and other relevant factors. The goal is to train an accurate model that can make predictions on new data with high precision and reliability. Many different machine learning algorithms and techniques can be applied to house price prediction, including regression models, decision trees, random forests, SVMs, neural networks, and others.

Overall, the background and motivation for house price prediction using machine learning lies in the need for accurate and reliable estimates of housing prices, which can help people make informed decisions, guide investments and real estate transactions. Additionally, advancements in machine learning and the availability of large datasets have made it possible to create more sophisticated models for real estate markets.

## **Chapter3. Problem Statement and Dataset Description**

prices of real estate properties are sophisticatedly linked with our economy.

Despite this, we do not have accurate measures of house prices based on the vast amount of data available.

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 for most consumers

### **Dataset description:**

House price datasets typically include a variety of features that can impact the price of a house, such as the number of bedrooms and bathrooms, the square footage of the house, the location, and the age of the house. The dataset may also include information on the sale date or listing date, the sale price or listing price, and any additional features or amenities that the house may have.

The data may be collected from a variety of sources, such as real estate websites or public records, and may be preprocessed and cleaned to remove any missing or erroneous data. The dataset description may provide more information on the source of the data, the preprocessing steps that were taken, and any additional information needed to interpret the dataset.

Overall, the description should provide a detailed overview of the dataset and its features, which is essential for understanding and working with the data.

## **Chapter 4: Existing Methodology**

There are currently existing methodologies in house price prediction, which includes traditional statistical models such as linear regression model .

```
from sklearn.linear_model import LinearRegression  
from sklearn.metrics import r2_score
```

## **Chapter 5: Proposed Methodology and Solution**

There are many proposed methodologies in house price prediction, including traditional statistical models such as Random Forest Classifier, Gradient Boosting Classifier various models.

```
from sklearn.ensemble import RandomForestRegressor  
from sklearn.metrics import mean_squared_error  
from sklearn.ensemble import GradientBoostingRegressor
```

**Solution:**

RMSE for Random Forest Classifier : **130802.05987876646**  
RMSE for Gradient Boosting Classifier : **134690.590701**

## **Chapter 6: Model Implementation**

### **Random Forest Classifier:**

Random forests or random decision forests is an ensemble learning method for classification, regression and other tasks that operates by constructing a multitude of decision trees at training time. For classification tasks, the output of the random forest is the class selected by most trees. For regression tasks, the mean or average prediction of the individual trees is returned.

### **Gradient Boosting:**

Gradient Boosting is a system of machine learning boosting, representing a decision tree for large and complex data. It relies on the presumption that the next possible model will minimize the gross prediction error if combined with the previous set of models. The decision trees are used for the best possible predictions.

### **Library used:**

- **Pandas**
- **Seaborn**
- **Matplotlib**
- **Numpy**



## Chapter 7: Testing and Evaluation

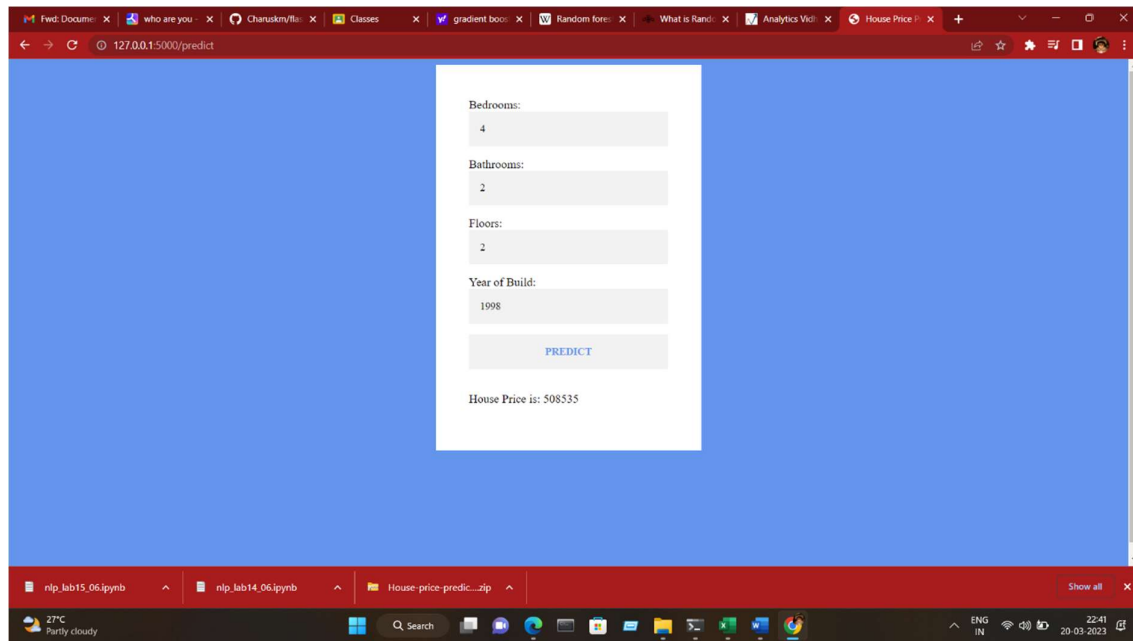
Model testing and evaluation is a crucial step in house price prediction. After training the model on a training dataset, it is necessary to evaluate its performance using a separate testing dataset. One common approach for evaluating the performance of regression models is to calculate metrics such as Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and R-squared value.

## Chapter 8: Model Deployment

Flask is a **web application framework written in python**, in simple terms it helps end users interact with your python code (in this case our ML models) directly from their web browser without needing any libraries, code files, etc.

Flask enables you to create web applications very easily, hence enabling you to focus your energy more on other important parts of a ML lifecycle like EDA, feature engineering, etc. Here in this blog I will give you a walkthrough on how to build a simple web application out of your ML Model and deploying it eventually.

Created a flask app deployment on house price prediction.



## Chapter 9: Model Archival in github and Demo in Youtube

**Github link:**

[https://github.com/Charuskm/houseprice\\_prediction](https://github.com/Charuskm/houseprice_prediction)

**Youtube Link:**

<https://youtu.be/qymEMo7o07I>

## **Chapter 10: Conclusion and Future Work**

We all have experienced a time when we have to look up for a new house to buy. But then the journey begins with a lot of frauds, negotiating deals, researching the local areas and so on.

So to deal with this kind of issues, Today we will be preparing a machine learning based model, trained on the House Price prediction Dataset.

## **Chapter 11: References**

<https://machinelearningprojects.net/house-price-prediction/>