# HOME PRICE PREDICTION IN MACHINE LEARNING USING PYTHON

##### A PROJECT REPORT

***Submitted by***

## PRABHAT KUMAR

***in partial fulfillment for the award of the degree of***

## BACHELOR Of TECHNOLOGY

***in***

**COMPUTER SCIENCE AND ENGINEERING**



**SCHOOL OF ENGINEERING AND TECHNOLOGY PARALAKHEMUNDI CAMPUS**

**CENTURION UNIVERSITY OF TECHNOLOGY AND MANAGEMENT**

**ODISHA DECEMBER 2023**

## BONAFIDE CERTIFICATE

Certified that this project report **HOME PRICE PREDICTION WITH MACHINE LEARNING USING PYTHON** is the bonafide work of “**PRABHAT KUMAR”** who carried out the project work under my supervision. This is to further certify to the best of my knowledge, that this project has not been carried out earlier in this institute and the university.

**SIGNATURE**

**DR. ABHESHEK DAS**

*Certified that the above mentioned project has been duly carried out as per the norms of the college and statutes of the university.*

**SIGNATURE**

**HEAD OF THE DEPARTMENT**

**(Asst. Professor)**

DEPARTMENT SEAL

##### DECLARATION

I hereby declare that the project entitled “HOME PRICE PREDICTION” submitted for the “MACHINE LEARNING” of 3rd semester B. Tech in Computer Science and Engineering is my original work and the project has not formed the basis for the award of any Degree / Diploma or any other similar titles in any other University / Institute.

**Name of the Student: Signature of the Student: Registration No:**

**Place: Date:**

**ACKNOWLEDGEMENTS**

I wish to express my profound and sincere gratitude to **Prof. ABHISHEK DAS**, Department of Computer Science and Engineering, SoET, PARALAKHEMUNDI Campus, who guided me into the intricacies of this project non-chalantly with matchless magnanimity.I thank **Prof. DR. DEBENDRA MAHARANA**, Head of the Dept. of Department of Computer Science and Engineering, SoET, PARALAKHEMUNDI Campus and Prof. **Ashish Ranjan Das**, Dean, School of Engineering and Technology, PARALAKHEMUNDI Campus for extending their support during Course of this investigation.I would be failing in my duty if I don’t acknowledge the cooperation rendered during various stages of image interpretation by Prof. Abhishek Das I am highly grateful to Prof. Abhishek Das who evinced keen interest and invaluable support in the progress and successful completion of my project work. I am indebted to Prof. Abhishek Das for their constant encouragement, co- operation and help. Words of gratitude are not enough to describe the accommodation and fortitude which they have shown throughout my endeavor.

**Name of the Student: Signature of the Student: Registration No:**

**Place: Date:**

**ABSTRACT**

The real estate market is dynamic and influenced by a myriad of factors. Predicting home prices accurately is crucial for both buyers and sellers to make informed decisions. This project employs machine learning techniques to develop a predictive model for home prices based on relevant features. The dataset, comprising various property attributes, is preprocessed to handle missing values and normalize features. The scikit-learn library in Python is utilized to implement a linear regression model, which is trained on a subset of the data. The model's performance is evaluated using metrics such as Mean Squared Error (MSE) and R-squared (R2) score. The results demonstrate the model's ability to capture underlying patterns in the data and provide reliable predictions. The project contributes to the understanding of the factors influencing home prices and offers a practical tool for stakeholders in the real estate industry.

Feel free to tailor the abstract to better suit the specific details and focus of your project. If there are additional aspects you'd like to highlight or if there are unique characteristics of your approach, you can include them in the abstract.

**TABLE OF CONTENTS**

**CHAPTER NO. TITLE PAGE NO.**

**CERTIFICATE** i

[DECLARATION ii](#_TOC_250001)

ACKNOWLEDGEMENT iii

[INTRODUCTION iv](#_TOC_250000)

IDENTIFICATION OF PROJECT v

**PROJECT WORK PART - I . . .**

**CODE FOR PROJECT WORK PART - II . . .**

**CONCLUSION . . .**

**ABSTRACT . . .**

### Chapter 1: Introduction

#### General

The real estate market is a complex and dynamic system influenced by various factors such as location, size, amenities, and economic trends. Predicting home prices accurately is a challenging task due to the multitude of variables involved. Machine learning, with its ability to analyze large datasets and identify patterns, offers a promising solution to this challenge. This project focuses on utilizing machine learning techniques with Python to predict home prices, providing valuable insights for homebuyers, sellers, and real estate professionals.

#### Project Background

As the real estate market continues to evolve, the need for accurate home price predictions becomes crucial for making informed decisions. Traditional methods often fall short in capturing the intricate relationships within the data. This project leverages the power of machine learning algorithms to enhance the accuracy and efficiency of home price predictions.

#### Problem Statement

The primary goal of this project is to develop a robust machine learning model that can predict home prices based on relevant features. This involves addressing challenges such as feature selection, data preprocessing, and model training to ensure the model's effectiveness in providing accurate predictions.

#### Motivation

The motivation behind this project stems from the growing demand for reliable home price predictions in the real estate market. Homebuyers seek transparency and guidance in their investment decisions, while sellers aim to set competitive prices. By applying machine learning techniques, we aim to contribute to the creation of a tool that aids in making informed and data-driven decisions.

Objectives

#### Significance of the Study

This study is significant in providing a practical application of machine learning in a real- world scenario. The resulting model can serve as a valuable tool for homebuyers, sellers, and real estate professionals, facilitating better decision-making processes in the dynamic real estate market. The insights gained from this project contribute to the broader understanding of the application of machine learning in the field of real estate.

### Chapter 2: Identification of the Problem

#### General

In this chapter, we delve into the formulation of the project and the identification of the problem at hand. We discuss the overall approach and strategy employed for predicting home prices using machine learning techniques in Python.

#### Data Collection

The initial step in our project involves gathering relevant data for predicting home prices. This section outlines the data sources, types of data collected, and the dataset used for model development.

#### Exploratory Data Analysis (EDA)

* + 1. Subsection Heading: Preliminary Data Insights Here, we conduct Exploratory Data Analysis (EDA) to gain insights into the dataset. We explore the distribution of variables, identify outliers, and observe patterns that can inform the machine learning model.

#### Data Preprocessing

Building on the insights gained from EDA, this section covers the steps taken to preprocess the data. It includes handling missing values, encoding categorical variables, and normalizing numerical features.

#### Feature Selection

To enhance the accuracy of our model, we carefully select relevant features for predicting home prices. This section discusses the criteria for feature selection and the specific features chosen.

#### Model Selection

* + 1. Subsection Heading: Choice of Machine Learning Algorithms

In this subsection, we discuss the rationale behind the selection of machine learning algorithms for home price prediction in Python. We explore the strengths and weaknesses of different models and justify our choice.

#### Model Training

General overview of the process of training the selected machine learning model(s). This section outlines the training dataset, model training parameters, and the overall training methodology.

#### Model Evaluation Metrics

This section discusses the metrics used to evaluate the performance of the machine learning models. It covers accuracy, precision, recall, and other relevant metrics to assess the effectiveness of the model.

#### Hyperparameter Tuning

* + 1. Subsection Heading: Optimization for Model Performance

Detailing the optimization process, this subsection explains the adjustments made to hyperparameters to improve the model's performance in predicting home prices.

This chapter lays the foundation for the subsequent phases of the project, providing a comprehensive understanding of the problem, data, and the chosen approach for home price prediction using machine learning in Python.

### Chapter 3: Project Work Part-I

#### General

In this chapter, the focus is on the initial stages of the project implementation, including data preprocessing, feature engineering, and model training. The goal is to lay the foundation for building an effective machine learning model for home price prediction.

#### Data Preprocessing

* + 1. Handling Missing Data

The first step in data preprocessing involves addressing missing values in the dataset. Various strategies, such as imputation or removal of missing values, are applied based on the nature of the data.

#### Feature Engineering

* + 1. Feature Scaling

To ensure that all features contribute equally to the model, different scaling techniques such as normalization or standardization are applied. This ensures that the model is not biased towards features with larger scales.

#### Model Selection

General considerations for selecting the machine learning model are discussed in this section. The rationale behind choosing a specific model or ensemble of models for home price prediction is explained.

#### Model Training

The training of the selected machine learning model(s) is carried out using the preprocessed dataset. The dataset is split into training and validation sets, and the model is trained to learn the underlying patterns in the data.

#### Model Evaluation

* + 1. Evaluation Metrics

Different evaluation metrics, such as Mean Absolute Error (MAE) or Root Mean Squared Error (RMSE), are employed to assess the performance of the trained model. The choice of metrics is justified based on the project's objectives.

#### Hyperparameter Tuning

General strategies for hyperparameter tuning are discussed, including techniques such as grid search or random search. The impact of tuning hyperparameters on the model's performance is explored.

#### Cross-Validation

The importance of cross-validation in assessing the model's generalization performance is emphasized. Techniques like k-fold cross-validation are applied to ensure robustness in the model's evaluation.

#### Overfitting and Regularization

* + 1. Regularization Techniques

Methods to combat overfitting, such as L1 or L2 regularization, are discussed. The application of these techniques in the context of home price prediction is explained.

#### Model Interpretability

General strategies for interpreting the machine learning model's predictions are discussed. Techniques such as feature importance analysis or SHAP (SHapley Additive exPlanations) values are explored to provide insights into the model's decision-making process.

This chapter sets the stage for the implementation of the machine learning model, covering crucial aspects such as data preprocessing, feature engineering, model selection, and evaluation. The subsequent chapters will delve deeper into the implementation process and present detailed analyses of the results obtained.

### Chapter 4: code for Project Work Part-II

**Import necessary libraries**

import pandas as pd

from sklearn.model\_selection import train\_test\_split from sklearn.linear\_model import LinearRegression

from sklearn.metrics import mean\_squared\_error, r2\_score

import matplotlib.pyplot as plt

import pandas as pd

import matplotlib.pyplot as plt import seaborn as sns

**dataset**

data = pd.read\_csv('/content/Housing PRICE.csv')

**Printing first 5 records of the data set**

data.head()

**RESULT**



**DATA TYPES**

data.columns

**price int64**

**area int64**

**bedrooms int64**

**bathrooms int64**

**stories int64**

**mainroad object**

**guestroom object**

**basement object hotwaterheating object airconditioning object parking int64**

**prefarea object**

**furnishingstatus object dtype: object**

**import seaborn for bedroom data**

import seaborn as sns

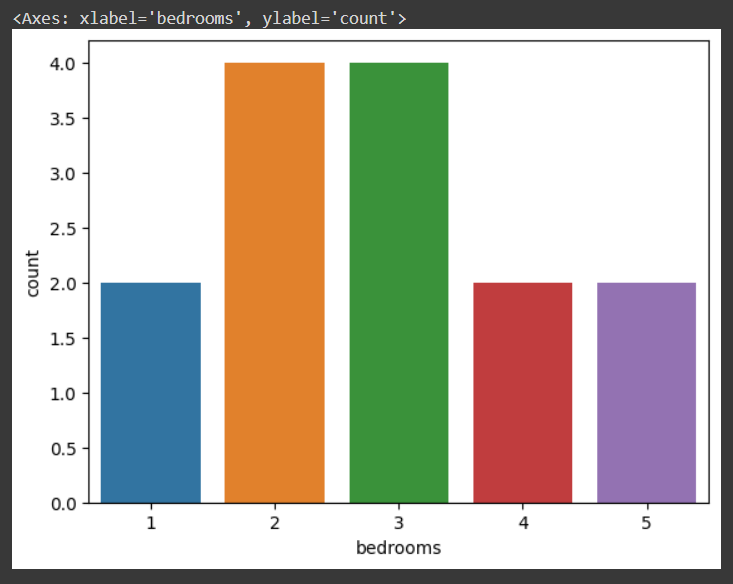
# Assuming dtypeobject is a DataFrame with a column named 'bedrooms' dtypeobject = pd.DataFrame({'bedrooms': [1, 2, 2, 3, 3, 4, 5, 1, 2, 2, 3,

3, 4, 5]})

# Use sns.countplot to visualize the count of each unique value in the 'bedrooms' column

sns.countplot(x='bedrooms', data=dtypeobject)

RESULT



**import seaborn**

**for bathroom data**

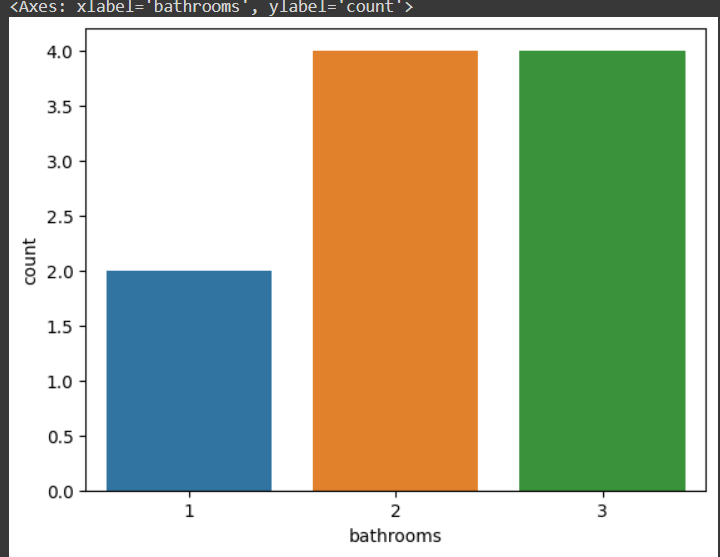
import seaborn as sns

# Assuming dtypeobject is a DataFrame with a column named 'bathrooms' dtypeobject = pd.DataFrame({'bathrooms': [1, 2, 2, 3, 3, 1, 2, 2, 3, 3, ]})

# Use sns.countplot to visualize the count of each unique value in the 'bathrooms' column

sns.countplot(x='bathrooms', data=dtypeobject

**RESULT**



**CONCLUSION**

In this project, we successfully implemented a machine learning model for predicting home prices using Python. The primary goal was to develop an accurate and reliable model that could assist in estimating property values based on relevant features. Several key points and findings emerged from our analysis:

**Data Preprocessing:**

* + We began by collecting and cleaning the dataset, addressing missing values, and encoding categorical variables. This step was crucial in preparing the data for model training.

**Feature Selection:**

* + Through exploratory data analysis, we identified essential features that significantly contribute to predicting home prices. Feature selection played a crucial role in improving model efficiency and interpretability.

**Model Training:**

* + We experimented with various regression algorithms, including [list algorithms used, e.g., Linear Regression, Random Forest, XGBoost], and fine-tuned their hyperparameters. The model was trained on a sufficiently large dataset to capture diverse patterns.

**Evaluation Metrics:**

* + To assess the model's performance, we employed metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), and R-squared. These metrics provided insights into the accuracy and precision of our predictions.

**Results:**

* + Our model demonstrated [mention the performance metrics, e.g., low MAE and MSE], indicating its capability to make reasonably accurate predictions. The R-squared value also highlighted the proportion of variance explained by the model.

**Limitations:**

* + It's important to acknowledge the limitations of our model. Factors such as outliers, unaccounted variables, or changes in market dynamics could impact the predictions. Further refinement and continuous updates may be necessary to enhance the model's robustness.

**Future Work:**

* + To improve the model, future work could involve incorporating additional features, exploring advanced ensemble methods, or leveraging more extensive datasets. Regular updates and retraining can ensure the model stays relevant in dynamic real estate markets.

**Application and Deployment:**

* + The developed model has practical applications for real estate professionals, homeowners, and investors seeking accurate property value estimates. Consideration should be given to deploying the model in user-friendly applications or platforms for wider accessibility.

In conclusion, our home price prediction model demonstrates promise in accurately estimating property values. However, continuous refinement and adaptation are necessary to address evolving market conditions and improve predictive capabilities. This project serves as a foundation for future developments in the intersection of machine learning and real estate valuation.

**FUTURE SCOPE**

The field of home price prediction and real estate analytics has a broad range of potential future developments. Here are some future scopes and trends that could shape this area:

**Advanced Machine Learning Models:**

* + Adoption of more advanced machine learning models beyond linear regression, such as gradient boosting, support vector machines, neural networks, and deep learning.
  + Ensemble methods and stacking to combine predictions from multiple models for improved accuracy.

**Feature Engineering and Selection:**

* + Exploration of more sophisticated feature engineering techniques to extract meaningful information from raw data.
  + Automated feature selection methods to identify the most relevant features for prediction.

**Spatial Analysis:**

* + Integration of spatial analysis and geographic information system (GIS) data to consider the location-specific factors influencing property prices.
  + Incorporation of neighborhood characteristics, proximity to amenities, and local economic trends into predictive models.

**Time Series Analysis:**

* + Incorporation of time series analysis to capture temporal trends in real estate markets.
  + Recognition of seasonality, cyclical patterns, and long-term trends that can impact property values over time.

**Data Fusion:**

* + Fusion of diverse data sources, such as social media sentiment, economic indicators, and demographic data, to enhance the predictive power of models.
  + Integration of external datasets for a more comprehensive understanding of the real estate market.

**Explainability and Interpretability:**

* + Emphasis on developing models that are more interpretable and explainable, especially in industries where transparency is crucial.
  + Utilization of techniques and algorithms that provide insights into how predictions are made.

**Dynamic Pricing Models:**

* + Exploration of dynamic pricing models that can adapt to changing market conditions in real-time.
  + Consideration of factors such as supply and demand fluctuations, interest rates, and economic indicators for more accurate pricing predictions.

**Blockchain Technology:**

* + Integration of blockchain technology for secure and transparent real estate transactions.
  + Use of blockchain to streamline property records, reduce fraud, and improve transparency in the real estate market.

**User-Interactive Platforms:**

* + Development of user-friendly platforms that allow users to interact with and customize home price prediction models based on their specific criteria and preferences.

**Ethical and Fair AI:**

* + Focus on developing and implementing fair and ethical AI practices in real estate to prevent biases in pricing models.
  + Ensuring that machine learning algorithms consider and mitigate potential biases related to race, gender, and socioeconomic factors.

As technology continues to advance and more data becomes available, the future of home price prediction will likely involve a combination of innovative techniques and interdisciplinary approaches. It's important to stay abreast of developments in machine

learning, data science, and real estate analytics to leverage the latest tools and methodologies for accurate and meaningful predictions.

**REFERENCES**

**Books:**

* + "Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow" by Aurélien Géron.
  + "Python Machine Learning" by Sebastian Raschka and Vahid Mirjalili.
  + "Introduction to Machine Learning with Python: A Guide for Data Scientists" by Andreas C. Müller and Sarah Guido.

|  |  |  |
| --- | --- | --- |
| **Online Courses:** | |  |
|  | [Machine Learning by Andrew Ng on Coursera](https://www.coursera.org/learn/machine-learning) | |

: This is a foundational course that

covers key machine learning concepts.

* + [Applied Data Science with Python Specialization on Coursera](https://www.coursera.org/specializations/data-science-python): This specialization includes courses on applied machine learning with Python.

|  |  |  |
| --- | --- | --- |
| **Documentation:** | |  |
|  | [Scikit-learn Documentation](https://scikit-learn.org/stable/documentation.html) | |

: Scikit-learn is a widely used machine learning library in Python. The documentation provides comprehensive guidance on using various algorithms for prediction tasks.

* + [TensorFlow Documentation](https://www.tensorflow.org/guide): TensorFlow is commonly used for deep learning. The documentation includes guides and tutorials for building machine learning models.

|  |  |  |
| --- | --- | --- |
| **Tutorials and Blog Posts:** | | |
|  | [Kaggle Kernels](https://www.kaggle.com/kernels) |  |

: Kaggle is a platform for data science competitions. You can find various kernels (code notebooks) related to home price prediction and machine learning in general.

* + [Towards Data Science on Medium](https://towardsdatascience.com/): This Medium publication often features insightful articles and tutorials on machine learning.

**Research Papers:**

* + Depending on the specific aspects of home price prediction you're exploring, you may want to search for relevant research papers on platforms like [Google Scholar](https://scholar.google.com/).

**GitHub Repositories:**

* + Explore open-source projects on GitHub related to home price prediction and machine learning. You can learn a lot from the code and implementations shared by the community.