

House Price Prediction Using Linear Regression:

Abstract:

House price prediction is one of the most important and widely used applications of machine learning in the real estate industry. Accurate prediction of house prices helps buyers to make informed purchasing decisions, assists sellers in setting reasonable prices, and supports real estate companies in understanding market trends. The value of a house does not depend on a single factor; instead, it is influenced by multiple attributes such as the size of the house, number of rooms, location, and other related features. Because of this complexity, traditional manual methods of price estimation are often unreliable and prone to error.

With the increasing availability of historical housing data, machine learning provides an effective solution for predicting house prices by learning patterns from past records. In this project, a supervised machine learning approach is used, where the model is trained using labeled data containing house features and their corresponding prices. Since the output variable is a continuous numerical value, a regression technique is required. Linear Regression is selected as the machine learning algorithm for this project because it is simple, easy to interpret, and well suited for understanding the relationship between independent variables and the target variable.

The project follows the complete machine learning life cycle, starting from data collection and loading, followed by Exploratory Data Analysis to understand the structure and characteristics of the dataset. Data preprocessing steps are applied to handle missing values and improve data quality. The dataset is then divided into training and testing sets, and the Linear Regression model is trained using the training data. Predictions are generated for the test data, and the model performance is evaluated using standard regression metrics such as Mean Squared Error, Mean Absolute Error, and R² score. The results of this project demonstrate that Linear Regression can effectively predict house prices for a basic dataset, thereby validating the practical application of supervised learning and regression concepts.

Problem Statement:

Predicting the correct price of a house is a difficult and complex task because property prices are influenced by multiple factors that interact with each other. Factors such as the area of the house, number of bedrooms, number of bathrooms, and overall structural features play an important role in determining the final price. Along with these physical attributes, market demand and historical pricing trends also affect house values. Due to this complexity, estimating house prices using traditional or manual methods often results in inaccurate or inconsistent outcomes.

In many cases, house price estimation is done based on human judgment, personal experience, or rough market assumptions. Such approaches are subjective in nature and may vary from person to person, leading to biased pricing decisions. Without proper analysis of historical data, it becomes difficult to understand the exact relationship between house features and their corresponding prices. This lack of analytical insight can cause overpricing or underpricing of properties, which is not desirable for either buyers or sellers.

With the increasing availability of real-world housing datasets, it has become possible to apply data-driven techniques to solve this problem more effectively. Machine learning provides methods that can automatically learn patterns from past data and use those patterns to make future predictions. Since house price is a continuous numerical value, a regression-based supervised learning approach is suitable for this problem. Therefore, the main problem addressed in this project is to design and implement a machine learning model that can accurately predict house prices based on given input features using Linear Regression.

Objectives of the Project:

The primary objective of this project is to develop a machine learning model that can accurately predict house prices based on given input features. The project aims to study and understand a real-world house price dataset and analyze how different attributes of a house influence its final price. Another important objective is to perform Exploratory Data Analysis in order to gain insights into the dataset, understand data distribution, and identify any patterns or relationships between variables.

The project also focuses on preprocessing the dataset to improve data quality by handling missing values and preparing the data for model training. An additional objective is to apply a supervised learning regression technique, specifically Linear Regression, to model the relationship between independent variables and the dependent variable. Through this project, the objective is also to evaluate the performance of the developed model using standard regression metrics such as Mean Squared Error, Mean Absolute Error, and R² score. Finally, the project aims to provide practical exposure to machine learning concepts and demonstrate their real-world application in solving house price prediction problems.