# **HOUSE PRICE PREDICTION**

* DESCRIPTION:

House price prediction is the process of using data analysis and machine learning techniques to estimate the selling or market value of residential properties. This prediction is based on various factors and features associated with the property and its surrounding area.

House price prediction is a valuable application of machine learning and data science, helping homeowners, buyers, and real estate professionals make informed decisions about property transactions. The accuracy of predictions depends on the quality and quantity of data, feature engineering, and the chosen machine learning algorithm.

* What is the need to create this model?

Creating a house price prediction model serves several important needs and purposes, both for individuals and organizations involved in the real estate industry and for homeowners or potential buyers. Here are some key reasons why there is a need to develop house price prediction models:

**Informed Decision-Making**: House price prediction models provide valuable information to homeowners, buyers, and sellers. Buyers can use these models to assess whether a property is reasonably priced, helping them make informed decisions about their investments. Sellers can use them to set competitive listing prices.

**Real Estate Investment**: For real estate investors, accurate price predictions are essential. Investors rely on these predictions to identify undervalued properties and make profitable investment decisions.

**Market Analysis**: Real estate professionals and analysts use house price prediction models to analyze local housing markets. These models can help identify trends, patterns, and areas of potential growth or decline in property values.

**Financial Planning**: Homeowners often use house price predictions as a part of their financial planning. Knowing the expected appreciation or depreciation of a property can impact long-term financial strategies, including retirement planning.

**Home Improvement Investments**: Homeowners considering renovations or improvements to their property can benefit from price predictions. Knowing how these improvements may affect the resale value of the house is valuable information.

**Competitive Advantage**: Real estate agents and agencies can gain a competitive advantage by offering clients access to accurate house price predictions. It enhances their service and helps clients make confident decisions.

* Which Algorithm is used?

House price prediction models can employ various machine learning algorithms to estimate property values accurately. The choice of algorithm depends on factors such as the dataset size, complexity, and the specific characteristics of the problem. Here are some common algorithms used in house price prediction:

**Linear Regression**: Linear regression is one of the simplest and most interpretable algorithms used for house price prediction. It models the relationship between the independent variables (features) and the target variable (house price) as a linear equation. Multiple linear regression can be used when there are multiple features.

**Decision Trees**: Decision tree-based algorithms like Decision Trees, Random Forest, and Gradient Boosting are popular for their ability to capture complex nonlinear relationships in the data. They can handle both numerical and categorical features.

**Support Vector Machines (SVM):** SVM is a powerful algorithm for regression tasks, including house price prediction. It tries to find a hyperplane that best fits the data points while maximizing the margin between data points and the hyperplane.

**K-Nearest Neighbors (K-NN):** K-NN is a simple yet effective algorithm for regression tasks. It predicts the house price based on the average of the K-nearest neighbors' prices in the training dataset.

**Neural Networks**: Deep learning techniques, particularly feedforward neural networks and convolutional neural networks (CNNs), can be used for house price prediction. They are capable of handling large and complex datasets but may require more data and computational resources.

**Lasso**:short for "Least Absolute Shrinkage and Selection Operator," is a machine learning algorithm used for regression tasks and feature selection. It's a regularization technique that is particularly useful when dealing with high-dimensional datasets or when you want to select a subset of the most important features for a predictive model. Lasso is commonly used in machine learning and statistics for various applications, including linear regression and logistic regression.

Here we used **Linear Regression** in our dataset:

Linear regression is a fundamental machine learning algorithm used extensively in house price prediction and various other regression tasks. It models the relationship between a dependent variable (in this case, house prices) and one or more independent variables

(features) in a linear fashion. Here is a detailed explanation of the linear regression algorithm used in house price prediction:

**1. Linear Model**: In simple linear regression, we have one independent variable (feature), and the relationship is modeled as a straight line:

y = mx + b

y is the dependent variable (house price).

x is the independent variable (a feature, such as square footage).

m is the slope of the line, representing the change in house price for a unit change in the feature.

b is the intercept, representing the base price of the house when the feature is zero.

**2. Multiple Linear Regression**: In house price prediction, you typically have multiple independent variables (features) influencing the house price. Multiple linear regression extends the simple linear regression model to accommodate multiple features:

y = b₀ + b₁x₁ + b₂x₂ + ... + bₖxₖ

y is still the dependent variable (house price).

x₁, x₂, ..., xₖ are the independent variables (features).

b₀ is the intercept (base price).

b₁, b₂, ..., bₖ are the coefficients associated with each feature, representing their impact on the house price.

**3. Objective Function**: In linear regression, the goal is to find the values of the coefficients (b₀, b₁, b₂, ...) that minimize the residual sum of squares (RSS) or the mean squared error (MSE). The objective function is to minimize:

MSE = Σ(yᵢ - ŷᵢ)² / n

yᵢ is the actual house price for the ith example.

ŷᵢ is the predicted house price for the ith example.

n is the number of examples in the dataset.

**4. Model Training**: The training process involves finding the optimal values of the coefficients that minimize the MSE. This can be done using various optimization techniques, including gradient descent, normal equations, or matrix factorization.

**Gradient Descent**: It iteratively adjusts the coefficients in the direction that reduces the MSE until convergence is reached.

**Normal Equations**: This approach directly computes the coefficients by solving a set of linear equations.

**Matrix Factorization**: Linear regression can also be expressed in matrix form, where the coefficients are calculated using linear algebra techniques.

**5. Model Evaluation**: After training, the model's performance is evaluated on a separate dataset, typically a test dataset. Common evaluation metrics for house price prediction include Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R-squared (R²).

**6. Assumptions**: Linear regression assumes that the relationship between the independent and dependent variables is linear, that the residuals (errors) are normally distributed and have constant variance (homoscedasticity), and that there is little or no multicollinearity among the independent variables.

* Dataset:

Source: Kaggle

For purpose of our project,dataset we used consist of house attributes such as bathroom,balcony,bhk,squarefeet area,location,crime rate etc.

* Distribution of Dataset

We can distribute the data in two parts, wiz., training and testing data. Thetraining data is a to train the machine learning model and the testing data is used to check the accuracy of the model.

Here, the model is trained using four regression algorithms which are Linear Regression, lasso, decision tree.

* Output:

After the model is programmed and implemented, the output will be displayed as it is shown below in which the graphical user interface is added using tkinter.and it can also give output without gui .



