**PREDICTING HOUSE PRICES USING MACHINE LEARNING**   
  
**PROBLEM DEFINITION:**The problem is to develop a machine learning model that can accurately predict house prices based on various features and attributes of the houses. This is a regression problem where we aim to predict a continuous numerical value (the price) rather than a categorical outcome. The accuracy of the predictions will be crucial as it can help both buyers and sellers in making informed decisions in the real estate market.   
  
**DESIGN THINKING:**  
**1.Data Source:**A good data source for house price prediction using machine learning should be Accurate, Complete, Covering the geographic area of interest, Accessible.

**2.Data Pre-processing:**  
Data pre-processing is the critical first step in any machine learning project. It involves cleaning the data, removing outliers, and handling missing values to prepare the dataset for model training. In the context of the house price prediction project, let's elaborate on the specific steps:  
 a)Duplicate Removal  
 b)Handling Missing Values  
 c)Categorical Variable Encoding  
 d)Data Normalization

**3.Feature selection :**Feature Selection is the process of identifying and selecting the most relevant features from a dataset for a given machine learning task. The goal of feature selection is to improve the performance of the machine learning model by reducing the number of features and eliminating irrelevant or redundant features. There are a variety of feature selection techniques. Some of the most common techniques include:

• Correlation-based feature selection  
• Information gain-based feature selection

**4. Model Selection**:  
Choose machine learning algorithms suitable for regression tasks. Common models for predicting house prices include linear regression, decision trees, random forests, gradient boosting, and neural networks  
(e.g., deep learning models).  
Experiment with multiple algorithms to determine which one provides the best performance for your specific dataset. You can also consider ensemble methods.

**Linear regression**: Linear regression is a simple but effective algorithm for house price Prediction. Linear regression models the relationship between the house price and the Features using a linear function. Choose the Linear Regression algorithm as your predictive model, given that you want to predict house prices, a continuous target variable.

**5.Model Training** :  
The task involves training a selected machine learning model using pre-processed data and subsequently evaluating the model's performance using key metrics such as Mean Ab solute Error (MAE), Root Mean Squared Error (RMSE), and R-squared.

**6.Evaluation Metrics:**

* Mean Absolute Error (MAE):  
  MAE measures the average absolute difference between the predicted value sand the actual target values. It provides insight into the average magnitude of errors made by the model.
* Root Mean Squared Error (RMSE):  
  RMSE is another common metric that calculates the square root of the average of squared differences between predicted and actual values. It provides information about the typical magnitude of errors and gives higher penalties to larger errors.
* S-squared (R2):  
  T-R-squared quantifies the proportion of the variance in the target variable that is explained by the model. It ranges from 0 to 1, where a higher value indicates a better fit. It is often used to assess how well the model captures the variation in the data.

**Feature Engineering:**Feature engineering is the process of creating new features or transforming existing ones to provide more meaningful information to the machine learning model. For the house price prediction project, we can consider the following feature engineering techniques:

1. .Age of the House: Create a new feature that represents the age of each house by subtracting the year built from the current year. This feature can be informative as older houses may have different pricing dynamics compared to newer ones.
2. Square Footage per Bedroom: Calculate a new feature by dividing the total square footage of a property by the number of bedrooms. This metric can provide insights into the spaciousness of bedrooms, which can be a key factor in house pricing.
3. Bathrooms per Square Foot: Compute a new feature by dividing the number of bathrooms by the total square footage. This can capture the luxury level of bathrooms relative to the property's size.
4. School District Quality: Integrate external data from a third-party API to assess the quality of the school district in which each house is located. Properties situated in neighbourhood with better school districts often command higher price.

**CONCLUSION**:  
we have established a clear understanding of our goal: to predict house prices using machine learning. We outlined a structured approach that includes data source selection, data pre-processing, feature selection, model selection, model training, and evaluation.