# Predicting House Prices Using Machine Learning

**Problem Statement:**

The objective of this machine learning project is to develop a predictive model that can estimate the market value or selling price of residential properties based on various input features. This predictive model will be valuable for real estate agents, homeowners, and prospective buyers in making informed decisions in the real estate market.

**Data:**

The dataset used for this project includes historical data on residential properties, comprising the following features:

**1. House Features:**

- Size of the house (in square feet)

- Number of bedrooms

- Number of bathrooms

- Lot size (in acres or square feet)

- Presence of additional amenities (e.g., pool, garage, garden)

- Year of construction or renovation

- Architectural style

- Neighborhood or location

- Other relevant architectural and structural attributes

**2. Location Features:**

- Neighborhood or postal code

- Proximity to schools, hospitals, parks, shopping centers, public transport, and other facilities

- Historical property price trends in the area

**3. Economic and Market Indicators:**

- Local or regional economic data (e.g., unemployment rate, income levels)

- Real estate market trends and indicators (e.g., interest rates, supply and demand)

**4. Target Variable:**

- Selling price of the house

**Objectives:**

**1. Data Collection and Preprocessing:** Gather a comprehensive dataset that includes the features mentioned above and preprocess the data to handle missing values, outliers, and ensure data quality.

**2. Exploratory Data Analysis (EDA):** Analyze and visualize the dataset to identify relationships between features and the target variable. Understand the distribution of house prices, detect outliers, and gain insights into the data.

**3. Feature Engineering:** Select relevant features and transform them as needed. This might involve encoding categorical variables, scaling numerical features, and creating new meaningful features.

**4. Model Selection:** Choose appropriate machine learning algorithms for regression tasks, such as linear regression, decision trees, random forests, gradient boosting, or neural networks.

**5. Model Training:** Split the dataset into training and testing sets, then train the selected machine learning models on the training data.

**6. Model Evaluation:** Evaluate the models' performance using appropriate regression metrics (e.g., Mean Absolute Error, Mean Squared Error, Root Mean Squared Error, R-squared) on the testing dataset.

**7. Hyperparameter Tuning:** Fine-tune the hyperparameters of the selected model(s) to optimize their performance.

**8. Deployment:** Deploy the trained machine learning model as an application or service that allows users to input house features and receive estimated house prices.

**Success Criteria:**

The success of this machine learning project can be determined by several factors, including:

1. Achieving a low prediction error on the testing dataset, as measured by chosen evaluation metrics.

2. The model's ability to generalize well to unseen data, indicating its reliability in predicting house prices in the real world.

3. Deployment of the model as a user-friendly application or service that can assist real estate professionals and property buyers/sellers.

**Deliverables:**

The deliverables for this project should include:

1. A well-documented Jupyter notebook or codebase that covers data collection, preprocessing, feature engineering, model development, evaluation, and hyperparameter tuning.

2. A trained machine learning model capable of predicting house prices.

3. Documentation on how to use the model for predictions.

4. Visualizations and insights from EDA.

5. A report summarizing the project, including findings, challenges, and recommendations.