Predicting house Prices Using Machine Learning.

**Problem Statement:**

The goal of this machine learning project is to create a predictive model that can estimate the selling price of residential houses based on a set of input features. This prediction will be valuable for both homebuyers and sellers, as well as real estate agents, to make informed decisions in the housing market.

**Data:**

The dataset used for this project contains historical data on residential properties, including information such as:

**1. Features:**

- Size of the house (in square feet)

- Number of bedrooms

- Number of bathrooms

- Neighbourhood or location

- Year of construction

- Lot size

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

- Distance to key facilities (e.g., schools, hospitals, public transport)

- Other relevant attributes

**2. Target Variable:**

- Selling price of the house

**Objectives:**

**1.** **Data Collection**: Gather and pre process a comprehensive dataset that includes features and target variable (selling prices) for a diverse set of residential properties.

**2. Exploratory Data Analysis (EDA):** Analyse and visualize the dataset to gain insights into the relationships between different features and the target variable. Identify any missing data, outliers, or data quality issues.

**3**. **Feature Engineering:** Select relevant features and preprocess the data to ensure it is suitable for machine learning algorithms. This may include handling missing values, encoding categorical variables, scaling numerical features, and addressing outliers.

4. **Model Selection**: Choose appropriate machine learning algorithms for regression, such as linear regression, decision trees, random forests, or gradient boosting. Evaluate various models based on performance metrics like Mean Absolute Error (MAE), Mean Squared Error (MSE), and R-squared.

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

**6. Model Evaluation:** Evaluate the models' performance using the testing dataset and choose the best-performing model based on the evaluation metrics.

**7. Hyper parameter Tuning:** Fine-tune the hyper parameters of the selected model(s) to optimize their performance.

**8. Deployment:** Deploy the trained model as an application or service that can take input data (house features) and provide predicted house prices.

**Success Criteria:**

The success of this machine learning project will be determined by the following criteria:

1. Achieving a low prediction error (measured by metrics like MAE, MSE, and R-squared) on the testing dataset.

2. The model's ability to generalize to new, unseen data.

3. Deployment of the model for real-world use, allowing users to obtain accurate house price estimates.

**Deliverables:**

The deliverables for this project will include:

1. A well-documented Jupyter notebook or codebase that covers data preprocessing, feature engineering, model selection, training, 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.

.