PREDICTING HOUSE PRICES USING MACHINE  LEARNING

**ABSTRACT:**

House price prediction is a challenging task, as it is influenced by a variety of factors, both quantitative and qualitative. Machine learning algorithms can be used to develop predictive models that take into account these factors and produce accurate estimates of house prices. This paper presents a modular approach to house price prediction using machine learning. The approach consists of many modules.

**OBJECTIVE:**

The objectice of this project is to develop a machine learning model that accurately predicts the prices of houses based on a set of features such as location, square footage, number of bedrooms and bathrooms and other relevant factors.

**DATA SOURCE:**

Dataset Link:   [**https://www.kaggle.com/datasets/vedavyasv/usa-housing**](https://www.kaggle.com/datasets/vedavyasv/usa-housing)

A good data source for house price prediction using machine learning should be accurate, complete, covering the geographic area of interest, accessible.

**DATA PREPROCESSING:**

This module involves cleaning and pre-processing the data to ensure that it is in a suitable format for machine learning algorithms.

* Duplicate Removal
* Handling Missing Values
  + Mean Imputation
  + Median Imputation
* Categorical Variable Encoding
  + One - Hot Encoding
  + Label Encoding
* Data Normalization
  + Standardisation
  + Min max Scaling

**FEATURE SELECTION:**

This module involves creating new features from the existing data that may be more predictive of house prices.

* Correlation based feature selection
* Infromation gain based feature selection
* Recursive feature elimination
* Linear regression
* Random forest regressor
* Gradient boosting regressor

**MODEL SELECTION:**

This module involves selecting a machine learning algorithm that is well-suited to the task of house price prediction. Example: Deep learning models.

**MODEL TRAINING:**

The task involves training a selected machine learning model using preprocessed data and subsequently evaluating the model’s performance using key metrics such as Mean Absolurte Error (MAE), Root Mean Squared Error (RMSE) and R-squared.

**EVALUATION METRICS:**

* + Mean Absolute Error (MAE)
  + Root Mean Squared Error (RMSE)
  + S-Squared (R2)

**Modules for house price prediction using machine learning:**

                     The following modules can be used to develop a house price prediction model using machine learning:

1. **Load the data:**

Load the data into a machine learning library, such as NumPy or pandas.

1. **Clean the data:**

Remove any missing values, outliers, or inconsistencies from the data.

1. **Pre-process the data:**

Convert categorical features to numerical features and scale the numerical features to a common range.

1. **Feature engineering:**

Create new features from the existing data that may be more predictive of house prices. For example, you could create features such as the number of bedrooms and bathrooms per square foot, or the distance to the nearest school and park.

1. **Select the most important features:**

Use a feature selection algorithm to select the most important features for predicting house prices.

1. **Choose a machine learning algorithm:**

There are many different machine learning algorithms that can be used for house price prediction. Some popular options include linear regression, logistic regression, decision trees, and random forests.

1. **Train the model:**

Train the selected model on the prepared data.

1. **Evaluate the model:**

Evaluate the performance of the trained model on a held-out test set.

1. **Deploy the model:**

Once you have developed a house price prediction model, you can use it to make predictions on new data by passing the features of the new data to the model. The model will then output a prediction of the house price. This modular approach to house price prediction using machine learning provides a flexible and extensible framework for developing accurate and reliable predictive models.  Deploy the model to production so that it can be used to make predictions on new data.