Predicting House Prices using Machine Learning

This project provides us an overview on how to predict house prices using various machine learning models. This proposed model considers as the most accurate model used for calculating the house prices and provides a most accurate prediction.

Scope:

Predicting the price of a house helps for determining the selling price of the house in a particular region and it helps people to find the correct time to buy a home. The objective is to develop a 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.

Applications:

- To eliminate the need of real estate agents to gain information regarding house prices.
- To provide best prices to users without getting cheated.
- To enable users to search home as per their budget.

Dataset:

The dataset that we use in our project is given below:
https://www.kaggle.com/datasets/vedavyasv/usa-housing

Steps involved in making our project:

1.Data Source:

In order to proceed further into this project, we should collect the data, then we should pre-process the data and then we perform exploratory data analysis on the given data set.

2. Data Preprocessing:

Data preprocessing is the process of cleaning our data set. Clean and preprocess the data, handle missing values, and convert categorical features into numerical representations and outlier removal.

- Dealing with Missing Values:

    Dealing with the problem of missing values because some machine learning models don't accept data with missing values. For that we can see the number of missing values in our dataset.

- Deleting Some Unimportant Columns:

  We will delete columns that are not useful in our analysis.

- Outlier Removal:

  Handling outliers, it consists of identifying the outliers and hence removing those outliers.

3.Feature Selection:

Feature selection is a common machine learning technique used to build a simplified model for understanding and to enhance generalization by removing irrelevant or redundant information. Machine learning models accept only numbers as input, and since our dataset contains categorical features, we encode them in order for our dataset to be suitable for modeling. In order to make all algorithms work properly with our data, we need to scale the features in our dataset. For that, we will use a helpful function named StandardScaler() from the popular Scikit-Learn Python package.

4. Model Selection:

The important step is choosing a machine learning algorithm. In this project , XGBoost algorithm is used for predicting the price of the house using a machine learning approach. XGBoost stands for extreme gradient boosting, where gradient boosting is implemented with several additional features focusing on performance and speed. With careful parameter tuning, it is capable of training highly accurate models.

5. Model training:

We train our model for house price prediction. The data is broken down into two modules: a Training set and Test set. The training set includes the target variable. We must initially train the model. Then, test the model on the test dataset and get the results.

6.Evaluation:

Evaluate the model's performance using metrics like Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared. Here, we evaluate the model performance by comparing its predictions with the actual true values of tested data using the MAE metric.

7.Comparison:

Based on mean absolute error, the XGBoost algorithm is expected to perform superior to the other models to predict the housing price.

Architectural diagram:
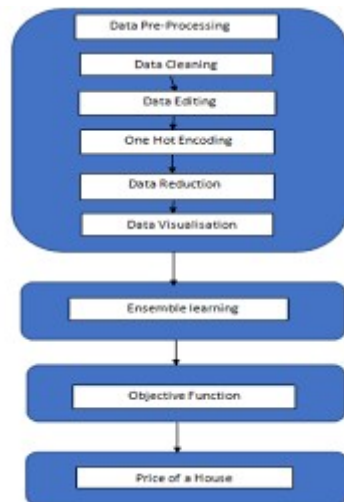
Conclusion:

This phase aims to predict the house prices accurately to the next level using a machine learning model. As we move forward, this project will help people to predict house prices.