Project Title: *Predicting House Prices using Machine Learning*

Dataset Link: https://www.kaggle.com/datasets/vedavyasv/usa-housing

Project Steps

Phase 1: Problem Definition and Design Thinking

In this part you will need to understand the problem statement and create a document on what have you understood and how will you proceed ahead with solving the problem. Please think on a design and present in form of a document.

Problem Definition: The problem is to predict house prices using machine learning techniques. 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. This project involves data preprocessing, feature engineering, model selection, training, and evaluation.

Design Thinking:

- **Data Source:** Choose a dataset containing information about houses, including features like location, square footage, bedrooms, bathrooms, and price.
- **Data Preprocessing:** Clean and preprocess the data, handle missing values, and convert categorical features into numerical representations.
- **Feature Selection:** Select the most relevant features for predicting house prices.
- **Model Selection:** Choose a suitable regression algorithm (e.g., Linear Regression, Random Forest Regressor) for predicting house prices.
- **Model Training:** Train the selected model using the preprocessed data.
- **Evaluation:** Evaluate the model's performance using metrics like Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared.

Phase 2: Innovation

The packages need to be imported are as follows:

```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.metrics import r2_score, mean_absolute_error,mean_squared_error
from sklearn.linear_model import LinearRegression
from sklearn.linear_model import Lasso
from sklearn.ensemble import RandomForestRegressor
from sklearn.svm import SVR
import xgboost as xg
```

Ensemble Technique:

Ensemble methods in machine learning usually produce more accurate solutions than a single model would. The ensemble methods in machine learning combine the insights obtained from multiple learning models to facilitate accurate and improved decisions.

Deep learning technique:

Deep learning drives many artificial intelligence (AI) applications and services that improve automation, performing analytical and physical tasks without human intervention. Deep learning neural networks, or artificial neural networks, attempts to mimic the human brain through a combination of data inputs, weights, and bias. These elements work together to accurately recognize, classify, and describe objects within the data.

Gradient Boosting:

Gradient Boosting is a powerful boosting algorithm that combines several weak learners into strong learners, in which each new model is trained to minimize the loss function such as mean squared error or cross-entropy of the previous model using gradient descent. In each iteration, the algorithm computes the gradient of the loss function with respect to the predictions of the current ensemble and then trains a new weak model to minimize this gradient. The predictions of the new model are then added to the ensemble, and the process is repeated until a stopping criterion is met.

Phase 3: Development part 1

In this part you will begin building your project by loading and preprocessing the dataset. Start building the house price prediction model by loading and preprocessing the dataset. Load the housing dataset and preprocess the data.

The code for predicting the house price using machine learning is given in the below link**:**

https://github.com/juliaan03/NM/blob/main/AI_Phase3.pdf

Phase 4: Development part 2

In this part you will continue building your project.
Continue building the house price prediction model by

- Feature selection
- Model training
- Evaluation.

The complete code is given in the below link:

https://github.com/juliaan03/NM/blob/main/AI_Phase4.ipynb

The readme file gives the detailed information about the execution of the code:

https://github.com/juliaan03/NM/blob/main/README.md