Predicting House Prices using Machine

Learning

**Phase 1: Problem Definition and Design**

**Thinking**

**Devasree P**

**AU711021205006**

**Predicting House Prices using Machine Learning**

**Phase 1: Problem Definition and Design Thinking**

# Problem Definition:

* The problem definition for predicting house prices using machine learning involves creating a predictive model that can estimate the selling price of a house based on a set of input features. Here's a more detailed problem definition:
* Problem Statement: Develop a machine learning model that accurately predicts the selling price of residential houses based on various property features.
* The goal is to provide a valuable tool for various stakeholders in the real estate market, including buyers, sellers, and real estate professionals, to make more informed decisions about property transactions.
* Machine learning is a computational approach that allows computer systems to analyze historical data on residential properties, learn patterns and relationships within the data, and use that knowledge to make accurate predictions about the selling prices of houses.
* It enables the creation of predictive models that can estimate house prices based on various property features, helping buyers, sellers, and real estate professionals make informed decisions in the housing market.

# Design Thinking:

* Design thinking methodology for developing an Predicting House Price using Machine Learning is impressive. It covers each of the essential steps, such as:

**Data gathering:**

* Gather a comprehensive dataset with historical housing data, including various features and sale prices. Ensure the data is clean, relevant, and representative of the target market.
* If relevant, gather geospatial data, including latitude and longitude coordinates, to incorporate location-based features.
* Decide whether you want to collect data for all available properties or a specific subset, such as properties within a certain price range or geographical area.

**Data preprocessing:**

* Perform data cleaning, handle missing values, and preprocess the features (e.g., feature scaling, encoding categorical variables) as needed.

**Feature selection:**

* Create new features or transformations of existing features if it improves model performance.
* In the case of house price prediction, we can use historical data on various features of a house, such as its location, size, and amenities, to train a machine-learning model.
* Once the model is trained, it can analyze new data on a given house and make a prediction of its market value.

**Model selection:**

* Choose an appropriate machine learning algorithm or a combination of algorithms for the regression task. Experiment with different models to determine which one performs best.
* To find an algorithm that performs well on the provided dataset, it is important to experiment with a variety of algorithms.

**Evaluation:**

* Apply the chosen metrics to assess the model's performance on the testing set. Calculate and report the MAE, MSE, RMSE, and R2 score.
* Create visualizations, such as scatter plots of predicted vs. actual prices or residual plots, to visually inspect how well the model's predictions align with the actual data.

**Iterative improvement:**

* Continuously collect and update your housing dataset to include new sales data and relevant features.
* Improve data quality by addressing missing values, outliers, and inconsistencies as new data becomes available.
* Continue to preprocess the data as needed, adapting to changes in the dataset or evolving requirements.

2