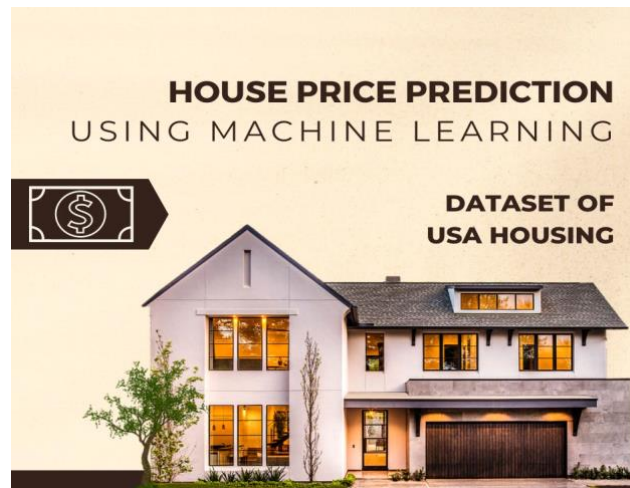# HOUSE PRICE PREDICTION
USING MACHINE LEARNING

REG NO: 950621104089

NAME : SAVARI JEEVA JEBASTIN P

OBJECTIVE :-

The objective 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, squarefootage, number of bedrooms and bathrooms, and other relevant factors.

PHASE 1: PROBLEM DEFINITON & DESIGN THINKING

Date	28 September 2023
Team ID	GROUP 4
Project Name	HOUSE PRICE PREDICITON USING MACHINE LEARNING

ABSTRACT :

Predicting house prices is a crucial task in the real estate industry, as it aids buyers, sellers, and investors in making informed decisions. Machine learning techniques have gained prominence in recent years for their ability to analyze vast amounts of data and provide accurate predictions. This project aims to develop a house price prediction system using machine learning algorithms. The system will utilize various features such as property size, location, number of bedrooms, bathrooms, and other relevant factors to predict the selling price of houses. The project will involve data preprocessing, feature engineering, model selection, and evaluation to build a robust and accurate prediction model. The outcome of this project will be a valuable tool for homeowners, real estate agents, and property investors to estimate house prices effectively.

PROBLEM STATEMENT:-

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 :-

A good data source for house price prediction using machine learning should be Accurate, Complete, Covering the geographic area of interest, Accessible.

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

DATA PREPROCESSING:

Data preprocessing is the critical first step in any machine learning project. It involves cleaning the data, removing outliers, and handling missing values to prepare the dataset for model training. In the context of the house price prediction project.
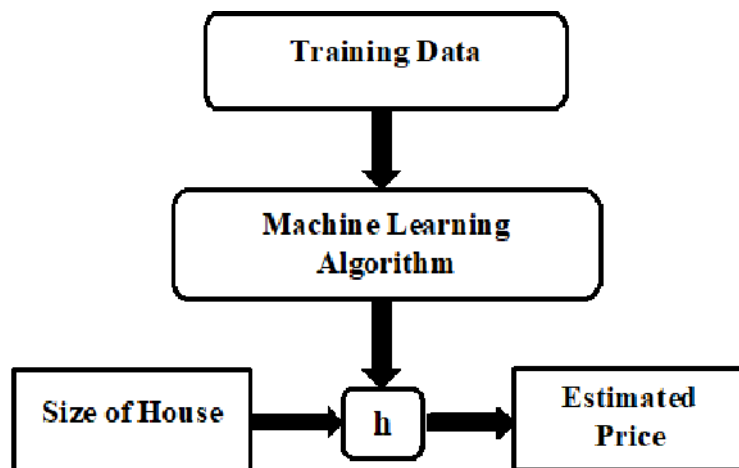
FEATURE SELECTION:

Feature Selection is the process of identifying and selecting the most relevant features from adataset for a given machine learning task. The goal of feature selection is to improve the performance of the machine learning model by reducing the number of features and eliminating irrelevant or redundant features.

MODEL SELECTION:

Choose machine learning algorithms suitable for regression tasks. Common models for predicting house prices include linear regression, decision trees, random forests, gradientboosting, and neural networks.

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 Absolute Error (MAE), Root Mean Squared Error (RMSE), and RM-square.

LINEAR REGRESSION:

Linear regression is a simple but effective algorithm for house priceprediction. Linear regression models the relationship between the house price and the features using a linear function.

RANDOM FOREST REGRESSION:

Random forest regression is a more complex algorithm that builds a multitude of decision trees to predict the house price. Random forest regression are typically more accurate than linear regression models, but they can be more computationally expensive to train.

GRADIENT BOOSTING REGRESSION:

Gradient boosting regression is another complex algorithm that builds a sequence of decision trees to predict the house price. Gradient boosting regression are typically more accurate than random forest regression, but they can be even more computationally expensive to train.

CATBOOST CLASSIFIER:

CatBoost is a machine learning algorithm implemented byYandex and is open-source. It is simple to interface with deep learning framework such as Apple's Core ML and Google's TensorFlow. Performance, ease-of-use, and robustness are the main advantages of the CatBoost library.

EVALUATION METRICES:-

MEAN ABSOLUTE ERROR (MAE):

MAE measures the average absolute difference between the predicted values and the actual target values. It provides insight into the average magnitude of errors made bythe model.

ROOT MEAN SQUARE ERROR (RMSE):

RMSE is another common metric that calculates the square root of the average of squared differences between predicted and actual values. It provides information about thetypical magnitude of errors and gives higher penalties to larger errors.

S-SQUARED (R2):

T-R-squared quantifies the proportion of the variance in the target

variable that is explained by the model. It ranges from 0 to 1, wherea higher valueindicates a better fit. It is often used to assess how well the model captures the variation in the data.

AGE OF HOUSE:

Create a new feature that represents the age of each house by subtracting the year built from the current year. This feature can be informative as older houses may have different pricing dynamics compared to newer ones.

.

- **Square Footage per Bedroom:** Calculate a new feature by dividing the total square footage of a property by the number of bedrooms. This metric can provide insights into thespaciousness of bedrooms, which can be a key factor in house pricing.

- **Bathrooms per Square Foot:** Compute a new feature by dividing the number of bathrooms by the total square footage. the luxury level .

- **School District Quality:** Integrate external data from a third-party API to assess thequality of the school district in which each house is located. Properties situated in neighborhoods with better school

APPLICATIONS:

House price prediction using machine learning has a wide range of practical applications in the real estate industry and beyond. Here are some notable applications:

- **Real Estate Investment:** Investors can use house price prediction models to assess the potential return on investment for properties. This helps them make informed decisions about buying, selling, or holding real estate assets.

- **Home Buying Assistance:** Prospective homebuyers can use these predictions to estimate the affordability of properties in different neighborhoods and make informed decisions about purchasing a home.

- **Selling Price Estimation:** Sellers can use machine learning models to estimate the selling price of their homes.

- **Property Valuation:** Real estate professionals, including appraisers and agents, can benefit from automated property valuation tools to provide more accurate and data-driven estimates to clients.

- **Market Analysis:** Machine learning models can be used to analyze real estate market trends and forecast future price changes, which is valuable for both buyers and sellers.

- **Portfolio Management:** Real estate investment companies can optimize their property portfolios by predicting future property values and making data-driven decisions about acquisitions and sales.

- **Risk Assessment:** Lenders and financial institutions use house price predictions to assess the risk associated with mortgage loans and determine loan-to-value ratios.

- **Property Tax Assessment**: Local governments can use machine learning models to assess property values for tax purposes, ensuring a fair and accurate tax system.

- **Home Renovation Planning**: Homeowners planning renovations can use property value predictions to estimate the potential return on investment for specific upgrades.

- **Housing Assistance Programs:** Government agencies can use predictions to allocate resources for housing assistance programs based on areas with the greatest need.

CONCLUSION:

In conclusion, house price prediction using machine learning has emerged as a powerful tool with widespread applications in the real estate industry and beyond. Through the application of advanced algorithms and data analysis techniques, machine learning models can provide accurate and data-driven estimates of property values. This technology has the potential to transform the way we buy, sell, invest in, and manage real estate.

By leveraging historical data and a variety of property-related features, these models enable real estate professionals, investors, homebuyers, and sellers to make more informed decisions.

They can assess property values, identify market trends, and optimize investment strategies. Additionally, machine learning-based predictions enhance the efficiency of property valuation, risk assessment, and portfolio management processes.

However, it's important to recognize that house price prediction using machine learning is not without its challenges. Ensuring data quality, addressing bias, and maintaining model accuracy over time are ongoing concerns. Moreover, the interpretation of complex models remains a topic of interest, particularly in providing users with transparency and insights into how predictions are made.

In the ever-evolving landscape of real estate and data science, house price prediction using machine learning continues to advance, offering valuable insights and opportunities for those involved in the housing market. As technology and methodologies evolve, this field holds the potential to reshape the way we approach real estate transactions and investment decisions, making them more data-informed and, ultimately, more efficient and equitable.