**Project-3**

## Project Title:

House Price Prediction Using Machine Learning

### Introduction

This project aims to build a machine learning model to predict house prices based on various features such as location, number of rooms, and other housing characteristics. Accurate price prediction is crucial for buyers, sellers, and real estate agents to make informed decisions.

### 2. Dataset Description

We used the **California Housing Dataset**, which contains information about different houses in California. The dataset includes 8 features such as:

* MedInc: Median income in the area
* HouseAge: Median house age
* AveRooms: Average number of rooms per household
* AveBedrms: Average number of bedrooms
* Population: Population of the block
* AveOccup: Average occupancy
* Latitude and Longitude of the house location

The target variable is the **Median house value**.

### 3. Data Exploration

* The dataset contains 20,640 samples.
* There are no missing values.
* The features have different scales, so scaling is necessary.
* Some features are correlated with house price, such as Median Income.

### 4. Data Preprocessing

* Split dataset into training (80%) and testing (20%) sets.
* Features scaled using StandardScaler to normalize input data for better model performance.

### 5. Model Selection and Training

* Used **Random Forest Regressor** as the prediction model due to its robustness and ability to capture non-linear relationships.
* Model trained on the scaled training data with 100 trees (estimators).

### 6. Model Evaluation

* Root Mean Squared Error (RMSE) on test data: **0.53** (lower is better)
* R-squared (R²) score on test data: **0.81** (higher is better, max 1)
* The evaluation metrics indicate that the model predicts house prices with good accuracy.

### 7. Results Visualization

The scatter plot of actual vs predicted prices shows that the predictions closely follow the actual values with minor deviations.

### 8. Conclusion

The Random Forest Regression model effectively predicts house prices based on available features. It can be further improved by hyperparameter tuning, adding more features, or using advanced models like Gradient Boosting or XGBoost.

### 9. Future Work

* Perform hyperparameter tuning for better accuracy.
* Explore additional datasets and features such as proximity to schools or amenities.
* Deploy the model as a web application for interactive price prediction.

### 10. Appendix

* Python code and saved model files are attached separately.
* Dataset used: California Housing Dataset from scikit-learn.