**Project Title: Bangalore House Price Prediction using KNN**

**Project Duration:** 1 Days  
**Date Completed:** April 20, 2025  
**Author:** *sagar khese*

**1. project overview:**

This project objective is to develop predictive model using k-nearest neighbours(knn) to estimate the house prices in Bangalore on various feature like size, location, number of bedrooms

**2. Dataset Overview**

* **Source file:** Bangalore house prediction .csv
* **Shape:** 5 rows × 108 columns
* **Features:** size, location, number of bedrooms, no of bath
* **Target Variable:**  price

**3. Solution Architecture**

* **Raw data -- >**
* **data processing -- >**
* **Feature encoding (:** size, location, number of bedrooms, no of bath) **-- >**
* Exploratory Data -- >
* **Model Training (KNN) -- >**
* **Model Evaluation (R2 Score, Error) -- >**
* **Model saving**

**4. Methodology**

| **Step** | **Reason** |
| --- | --- |
| **Data Collection** | Loading csv file from local library |
| **Data processing** | Checking and handling the null values |
| **Exploratory Data Analysis (EDA)** | Visualized relationships using seaborn line plots and Checked feature distributions and correlations. |
| **Feature Engineering** | Separated features x and target y and standard scaler also use |
| **Model selection** | By using KneighborsRegressor from sklearn.neighbors. |
| **Model Training**  **Evaluation** | Spilt data to x and y variable in ratio of (80 :20) and trained model with varying k values to identify optimal neighbours  R2score and viz training and validation error |
| **Deployment Prep** | Saved the model using Joblib for future use |

**5. Time Taken**

| **Task** | **Time Spent** |
| --- | --- |
| Data collection, Cleaning & EDA | 2 hours |
| Model Building | 1.30 minutes |
| Visualization | 30 minutes |
| Documentation | 1 hour |
| **Total Time** | **5 hours** |

**6. Challenges Faced**

* Understanding the distribution of real estate features across Bangalore
* Finding error in data
* Differentiating between overfitting and underfitting in high-dimensional space.

**7. Complexity**

* **Complexity:** hard
* **This project are hard to handle because of multivariate feature and the need for careful tuning of knn parameter to ensure predictive power**
* **Skills Required:** Python, Pandas, Seaborn, Scikit-learn, Data Visualization