**House Price Prediction**

**Table of Contents**

1. Introduction
2. Project Overview
3. Dataset Description
4. Steps Followed
   * Importing Libraries
   * Loading the Dataset
   * Handling Missing Values
   * Data Cleaning
   * Outlier Detection and Removal
   * Feature Engineering
   * Model Selection and Training
   * Model Serialization
   * Deployment
5. Technologies Used
6. Conclusion

**1. Introduction**

This document provides an overview of a Machine Learning project that predicts house prices in Bangalore. The project involves data preprocessing, feature engineering, model selection, training, and deployment using FastAPI.

**2. Project Overview**

This project aims to develop an end-to-end Machine Learning model that predicts house prices in Bangalore based on features like location, size, and total square feet. The model was trained using Ridge Regression and deployed using FastAPI.

**3. Dataset Description**

The dataset was collected from **Kaggle** and includes the following key features:

* **Location** – The area where the house is located.
* **Size** – Number of bedrooms (BHK).
* **Total Square Feet** – The total area of the house.
* **Price** – The target variable to be predicted.

**4. Steps Followed**

**4.1 Importing Libraries**

Essential libraries used:

* **pandas, numpy** – For data manipulation.
* **matplotlib** – For data visualization.
* **sklearn** – For machine learning modeling.

**4.2 Loading the Dataset**

* Imported the dataset into a **Pandas DataFrame**.
* Checked for **null values and missing data**.

**4.3 Handling Missing Values**

* Filled missing values in the **bathroom column** with the **mean value**.
* Dropped rows with significantly missing or incorrect data.

**4.4 Data Cleaning**

* Dropped **irrelevant columns**.
* Converted the **size column** into **BHK format** to ensure consistency.
* Standardized **total\_sqft values** to ensure realistic area calculations.

**4.5 Outlier Detection and Removal**

* Identified **outliers** where **a 3 BHK apartment was priced lower than a 2 BHK** in the same area and removed them.

**4.6 Feature Engineering**

* Converted the **location column** from categorical to numerical using **one-hot encoding**.

**4.7 Model Selection and Training**

* Selected **Ridge Regression** as the best machine learning model for price prediction.
* Trained the model successfully and achieved accurate predictions.

**4.8 Model Serialization**

* Saved the trained model as a **pickle file (model.pkl)**.
* Created a **JSON file** for future reference.

**4.9 Deployment**

* Created an **app.py** file for the **FastAPI** application.
* Successfully deployed the model on **FastAPI**.

**5. Technologies Used**

The following technologies were used in this project:

* **Programming Language:** Python
* **Libraries:** pandas, numpy, matplotlib, sklearn
* **Machine Learning Algorithm:** Ridge Regression
* **Deployment Framework:** FastAPI

**6. Conclusion**

This project successfully developed and deployed a **house price prediction model for Bangalore**. The model was trained using **Ridge Regression**, and it was deployed using **FastAPI** for real-time predictions.

This document provides a **comprehensive overview of the project**, covering **data preprocessing, model training, and deployment**.