Bengaluru House Price Prediction

Objective: The aim of this project is to predict house prices in Bengaluru using various machine learning regression models. This will help potential homebuyers and real estate investors make informed decisions.

Dataset: The dataset used for this project is sourced from Kaggle and includes features such as area type, location, size, total square feet, number of bathrooms, balcony count, and the price of the house.

Data Preprocessing:

- Handling missing values by dropping columns with a high percentage of missing data and filling missing values in columns with low percentages.
- Converting categorical features to numeric using techniques like one-hot encoding.
- Scaling features using Min-Max scaling.
- Detecting and removing outliers using methods like Z-score.

Feature Engineering:

- Converting the 'total_sqft' feature to a numeric format by handling ranges and non-numeric entries.
- Extracting the number of bedrooms (BHK) from the 'size' feature.
- Creating new features to enhance the predictive power of the models.

Models Used:

- **Linear Regression:** Assumes a linear relationship between the input features and the target variable.
- **Lasso Regression:** Uses L1 regularization to prevent overfitting by penalizing large coefficients.
- **Ridge Regression:** Similar to Lasso, but uses L2 regularization.
- **Random Forest Regression:** An ensemble model that builds multiple decision trees and averages their predictions.
- **Gradient Boosting Regression:** Builds trees sequentially, with each tree correcting errors made by the previous ones.
- **Support Vector Machine:** Finds the hyperplane that best separates the data into different classes, adapted here for regression.
- **XGBoost:** An optimized version of gradient boosting that is efficient and effective for large datasets.

Evaluation Metrics:

- **Root Mean Squared Error (RMSE):** Measures the square root of the average squared differences between predicted and actual values.
- **Mean Absolute Error (MAE):** Measures the average magnitude of errors in a set of predictions, without considering their direction.
- **R-squared (R^2):** Indicates the proportion of the variance in the dependent variable that is predictable from the independent variables.

Conclusion: The project aims to identify the most effective model for predicting house prices in Bengaluru, providing accurate price estimates to assist buyers and sellers in making informed decisions.