House Price Prediction Using Linear Regression

1. Project Overview

This project focuses on analyzing and predicting house prices using a Linear Regression model. The goal is to build a machine learning model that accurately predicts house prices based on various features such as area, number of bedrooms, and additional amenities.

2. Dataset Description

Total Records: 545

Total Features: 13

Target Variable: price (house price)

Key Features:

Numerical: area, bedrooms, bathrooms, stories, parking

Categorical: mainroad, guestroom, basement, hotwaterheating, airconditioning, prefarea, furnishingstatus

Data Exploration:

The dataset contains no missing values.

Categorical features were label encoded to convert them into numerical values.

3. Data Preprocessing

Encoding: Used LabelEncoder for categorical features.

Train-Test Split:

Training Data: 80% (436 samples)

Testing Data: 20% (109 samples)

4. Model Training

Algorithm Used: Linear Regression

Training Process:

The model was trained on the processed dataset using sklearn’s LinearRegression().

Predictions were made on the test data.

5. Model Performance

Mean Absolute Error (MAE): 979,679.69

R² Score: 0.649

This means the model explains 64.9% of the variance in house prices.

6. Hyperparameter Tuning

Technique Used: GridSearchCV (5-fold cross-validation)

Hyperparameters Tuned:

fit\_intercept: [True, False]

Best Model Configuration:

fit\_intercept = True

Best R² score on validation set: 0.647

7. Conclusion

The model provides moderate accuracy (65%) in predicting house prices.

To improve results, further techniques such as feature engineering, non-linear models, or ensemble learning can be explored.