Support Vector Machine (SVM) Model Report

Dataset Used

The dataset used for this analysis is the Housing dataset (https://raw.githubusercontent.com/sorif95/ML-Assignment/main/Housing.csv) , which contains various features related to house properties, including price, area, number of bedrooms, bathrooms, stories, and parking.

Preprocessing Steps

1. Missing values were handled using mean, median, mode imputation, and KNN imputation for bathrooms.
2. Categorical variables were encoded using Label Encoding.
3. Numerical features were standardized using StandardScaler.
4. A new binary classification feature 'parking_binary' was created.
5. Feature selection was performed using a Random Forest classifier.
6. Dimensionality reduction was applied using PCA to retain 5 principal components.

Model Performance Results

Classification Metrics

Test Accuracy: 0.91
Precision: 0.69
Recall: 1.00
F1-score: 0.81

Regression Metrics

MSE: 0.47
RMSE: 0.69
MAE: 0.46
R^2 score: 0.66

Observations on Performance Changes

1. The SVM classifier with RBF kernel and class weighting improved class balance handling.
2. Feature selection using Random Forest helped focus on relevant variables, reducing noise.
3. PCA helped reduce dimensionality, improving model efficiency without sacrificing performance.
4. The classification model improved recall significantly but had lower precision.
5. The regression model performed well with an R^2 score of 0.66, showing good predictive power.