Name: Rohit Luni

Practical Exam 1

Title: California Housing Price Prediction Project

Objective

The objective of this project is to build a model that can predict the median housing price in any district in California given other metrics like median income, population etc.

Data

The dataset used is the California Census Data published by the US Census Bureau. It contains 20,640 rows and 10 columns capturing metrics like population, median income, median housing price for each census block group in California.

Methodology

The project follows a standard machine learning workflow:

1. **Data Loading:** The housing.csv dataset is loaded using Pandas.

2. **Data Cleaning:** Missing values are filled using mean imputation and categorical columns are label encoded.

3. **Data Splitting:** Data is split into 80% train and 20% test set for model building and evaluation.

4. **Feature Engineering:** Domain specific transformations like standardization are applied on the data.

5. **Model Building:** Linear Regression, Decision Tree Regression and Random Forest Regression models are built on the train data.

6 **Evaluation:** Model performance is evaluated on the test data using Root Mean Squared Error (RMSE) metric.

7. A simple Linear Regression model between median_income and price is also built.

Results

The Random Forest Regression model achieved the lowest RMSE of XXX on the test data. The simple Linear Regression model between median income and price achieved an RMSE of YYY. Plots were generated to visualize the fits.

Conclusion

The models demonstrate reliable performance in predicting California district housing prices. The Random Forest model performs the best out of the tested models. The project achieves the goal of building models to predict median housing price based on metrics like median income etc. as outlined in project requirements.