

Task 1

House Price Prediction using Linear Regression

Name: *kaku harsh*

Task: AI & ML – Task 1

1. Introduction

This project aims to build and evaluate a Linear Regression model using the California Housing dataset. The task demonstrates the complete machine learning workflow including data loading, exploration, training, and evaluation.

2. Dataset & EDA

The California Housing dataset contains information such as median income, house age, number of rooms, and location-based features.

EDA showed that:

- Dataset contains no missing values
 - 8 independent variables
 - One target variable: Median House Value
-

3. Model Building

The dataset was split into 80% training and 20% testing data.

A Linear Regression model was trained using scikit-learn.

4. Evaluation Metrics

The model performance was evaluated using:

- Mean Absolute Error (MAE)
- Root Mean Squared Error (RMSE)
- R^2 Score

These metrics indicate how close predictions are to actual house prices.

5. Results & Visualization

A scatter plot of actual vs predicted values shows a strong linear relationship, indicating good model performance.

6. Improvement Ideas

- Apply feature scaling
 - Try Ridge or Lasso Regression
 - Use advanced models like Random Forest
-

Conclusion

In this project, a Linear Regression model was successfully built and evaluated using the California Housing dataset. The complete machine learning workflow was followed, including data loading, exploratory data analysis, train-test splitting, model training, and performance evaluation.

The model achieved reasonable accuracy as indicated by the MAE, RMSE, and R^2 score, demonstrating a strong linear relationship between the input features and the target variable (median house value). The Actual vs Predicted plot further confirmed that the model predictions closely follow the true values.

Although Linear Regression provided a good baseline model, its performance can be further improved by applying feature scaling, handling outliers, and experimenting with advanced regression techniques such as Ridge, Lasso, or ensemble-based models. Overall, this project effectively demonstrates the fundamentals of machine learning and serves as a strong foundation for more complex predictive modeling tasks.