Tab 1

CALIFORNIA_HOUSING_LINEAR_REGRESSION_PROBLEM

Project_Title:-

California Housing Price Prediction using Linear Regression

Shaik Khasif Ahmad | Internship Task 1|Day-1|

1. Objective:-

- Build a Linear Regression model to predict median house values
- Understand feature impact on housing prices

2. Dataset Description:-

- California Housing dataset from Scikit-learn
- 8 numerical features
- Target: Median House Value
- Train-test split: 80–20

3. Exploratory Data Analysis (EDA):-

- Dataset inspected for shape, columns, and missing values
- No missing values found
- Target variable shows continuous distribution

Tab 2

4. Model & Methodology:-

- Linear Regression model implemented using Scikit-learn
- Model trained on training data
- Predictions generated on test data

5. Evaluation Metrics:-

- MAE: 0.533
- RMSE: 0.746
- R^2 Score: 0.576

6. Results & Observations:-

- Model shows reasonable predictive performance
- Predictions follow a linear trend with actual values

7. Improvement Ideas:-

- Apply feature scaling
- Use Ridge or Lasso regression
- Try non-linear models like Random Forest or XGBoost

8. Conclusion:-

- Linear Regression provides a strong baseline model
- Further tuning can improve accuracy