

Task 6: Linear Regression – House Price Prediction

Tools:

- Python, Pandas, NumPy
- Scikit-learn
- Matplotlib

Dataset:

- House Prices

Hints / Mini Guide:

1. Load the California housing dataset and convert it into a Pandas DataFrame to view columns and target values clearly.
2. Perform basic inspection using `.head()`, `.info()`, and `.describe()` to understand numerical ranges and data types.
3. Separate input features (`X`) and target variable (`y`) and confirm that `y` is continuous numeric.
4. Split the dataset into training and testing sets using `train_test_split` with a fixed `random_state`.
5. Train a Linear Regression model using `LinearRegression()` and fit it using training data.
6. Predict on test data and create a small comparison table showing actual vs predicted results.
7. Evaluate the model using MAE and RMSE to measure prediction error properly.
8. Plot predicted vs actual values using scatter plot to visually check performance quality.
9. Interpret model coefficients to identify the most impactful features on house prices.

Deliverables:

- Notebook (.ipynb)
- MAE + RMSE report
- Predicted vs Actual plot

Final Outcome:

Intern builds a complete Regression model and understands evaluation and interpretation.

Interview Questions Related To Above Task:

- What is the use of Linear Regression?
- What is RMSE and why is it important?
- Why do we split train and test data?
- What does coefficient represent?
- What is underfitting?

📌 Task Submission Guidelines

- ⌚ **Time Window:**

You can complete the task anytime between 10:00 AM to 10:00 PM on the given day. Submission link closes at 10:00 PM.

- 🔍 **Self-Research Allowed:**

You are free to explore, Google, or refer to tutorials to understand concepts and complete the task effectively.

- 🛠️ **Debug Yourself:**

Try to resolve all errors by yourself. This helps you learn problem-solving and ensures you don't face the same issues in future tasks.

- 💸 **No Paid Tools:**

If the task involves any paid software/tools, do not purchase anything. Just learn the process or find free alternatives.

- 📁 **GitHub Submission:**

Create a new GitHub repository for each task.

Add everything you used for the task — code, datasets, screenshots (if any), and a short README.md explaining what you did.

⬆️ Submit Here:

After completing the task, paste your GitHub repo link and submit it using the link below:

- 👉 [\[Submission Link\]](#)

Best
of
Luck

