

Setup, Training, and Evaluation Guide

This guide walks you through how to set up the environment, train the model, and evaluate its performance on the California Housing dataset.

1. Environment Setup

1.1 Clone the repository:

```
git clone https://github.com/yourusername/housing-model.git
```

```
cd housing-model
```

1.2 Create and activate a virtual environment (recommended):

```
python -m venv .venv
```

```
source .venv/bin/activate # Windows: .venv\Scripts\activate
```

1.3 Install dependencies:

```
pip install -r requirements.txt
```

Alternatively, you can use 'poetry install' or 'conda env create -f environment.yml' if provided.

2. Train the Model

2.1 Launch the Jupyter notebook:

```
jupyter lab
```

2.2 Open the notebook:

```
notebooks/california_housing_model.ipynb
```

2.3 Run the cells in order to:

- Load and explore the dataset
- Engineer new features (e.g., rooms_per_household)
- Standardize features
- Train baseline and advanced models (Linear Regression, Random Forest)
- Tune hyperparameters using GridSearchCV

3. Evaluate the Model

The notebook will automatically:

- Calculate MAE, RMSE, and R^2
- Visualize actual vs. predicted results
- Show feature importances
- (Optional) Use SHAP for deeper interpretability

The final model achieves $R^2 \geq 0.80$ on the test set, exceeding the project success threshold.