# House Price Prediction

This project builds a Random Forest model to predict housing prices using data from the Kaggle competition House Prices: Advanced Regression Techniques. It demonstrates a full machine learning pipeline: data retrieval, preprocessing, model training, evaluation, and result visualization.

## Project Files

| File/Folder | Description |
|---|---|
| `data_utils.py` | Downloads and unzips data from Kaggle API |
| `model.py` | Trains a Random Forest model and evaluates it |
| `images/` | Stores output plots (prediction vs. actual, importance) |
| `report.md` | Final report for academic submission |
| `README.md` | This file |

## How to Run

1. **Place your Kaggle API key** in: ~/.kaggle/kaggle.json
2. **Download the dataset**:

```
python data_utils.py
```

3. **Train the model and generate plots**:

```
python model.py
```

4. **Check output in** `images/` **folder**:

- `pred_vs_actual.png`

- `feature_importance.png`

# Full Report

Read the full analysis and conclusions here: report.md

# Requirements

Install required packages with:

```
pip install pandas numpy matplotlib scikit-learn kaggle
```