# MLOps: Assignment-1

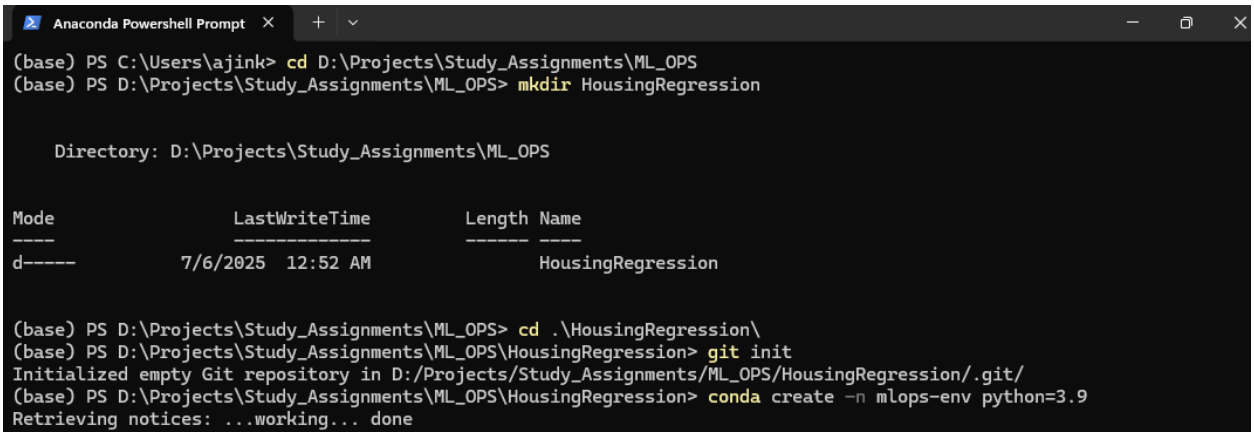
Ajinkya Ghodake (G24AI1046)

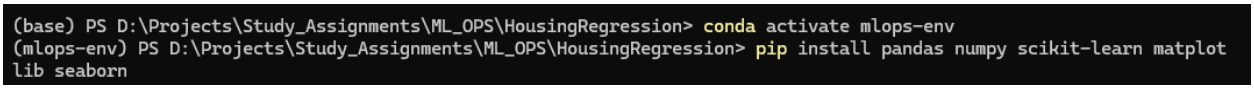
GitHub Repo: https://github.com/git-commit-acc/HousingRegression

---

## GitHub Repository Setup

```
mkdir HousingRegression
cd HousingRegression
git init
git remote add origin https://github.com/git-commit-acc/HousingRegression.git
git branch -M main
git push -u origin main
```



```
(mlops-env) PS D:\Projects\Study_Assignments\ML_OPS\HousingRegression> pip freeze > requirements.txt
(mlops-env) PS D:\Projects\Study_Assignments\ML_OPS\HousingRegression> mkdir .github


    Directory: D:\Projects\Study_Assignments\ML_OPS\HousingRegression


Mode                 LastWriteTime         Length Name
----                 -------------         ------ ----
d-----          7/6/2025  12:57 AM                .github


(mlops-env) PS D:\Projects\Study_Assignments\ML_OPS\HousingRegression> mkdir .github/workflows


    Directory: D:\Projects\Study_Assignments\ML_OPS\HousingRegression\.github


Mode                 LastWriteTime         Length Name
----                 -------------         ------ ----
d-----          7/6/2025  12:57 AM                workflows


(mlops-env) PS D:\Projects\Study_Assignments\ML_OPS\HousingRegression> New-Item utils.py -type file


    Directory: D:\Projects\Study_Assignments\ML_OPS\HousingRegression


Mode                 LastWriteTime         Length Name
----                 -------------         ------ ----
-a----          7/6/2025  12:57 AM              0 utils.py


(mlops-env) PS D:\Projects\Study_Assignments\ML_OPS\HousingRegression> New-Item regression.py -type file


    Directory: D:\Projects\Study_Assignments\ML_OPS\HousingRegression


Mode                 LastWriteTime         Length Name
----                 -------------         ------ ----
-a----          7/6/2025  12:57 AM              0 regression.py


(mlops-env) PS D:\Projects\Study_Assignments\ML_OPS\HousingRegression> New-Item README.md -type file


    Directory: D:\Projects\Study_Assignments\ML_OPS\HousingRegression


Mode                 LastWriteTime         Length Name
----                 -------------         ------ ----
-a----          7/6/2025  12:57 AM              0 README.md


(mlops-env) PS D:\Projects\Study_Assignments\ML_OPS\HousingRegression> New-Item .\.github\workflows\ci.yml -type fil
e


    Directory: D:\Projects\Study_Assignments\ML_OPS\HousingRegression\.github\workflows


Mode                 LastWriteTime         Length Name
----                 -------------         ------ ----
-a----          7/6/2025  12:57 AM              0 ci.yml


(mlops-env) PS D:\Projects\Study_Assignments\ML_OPS\HousingRegression> echo "# Housing Regression ML Pipeline" > REA
DME.md
(mlops-env) PS D:\Projects\Study_Assignments\ML_OPS\HousingRegression> echo "" >> README.md
(mlops-env) PS D:\Projects\Study_Assignments\ML_OPS\HousingRegression> echo "This repository contains a complete ML
pipeline for predicting house
>> prices using the Boston Housing dataset." >> README.md
(mlops-env) PS D:\Projects\Study_Assignments\ML_OPS\HousingRegression> git add .
```

```
(mlops-env) PS D:\Projects\Study_Assignments\ML_OPS\HousingRegression>  git config --global user.email "g24ai1046@iitj.ac.in"
(mlops-env) PS D:\Projects\Study_Assignments\ML_OPS\HousingRegression>  git config --global user.name "git-commit-acc"
(mlops-env) PS D:\Projects\Study_Assignments\ML_OPS\HousingRegression> git add .
(mlops-env) PS D:\Projects\Study_Assignments\ML_OPS\HousingRegression> git commit -m "Initial commit: Setup repository structure"
[master (root-commit) 07b5982] Initial commit: Setup repository structure
 5 files changed, 0 insertions(+), 0 deletions(-)
 create mode 100644 .github/workflows/ci.yml
 create mode 100644 README.md
 create mode 100644 regression.py
 create mode 100644 requirements.txt
 create mode 100644 utils.py
(mlops-env) PS D:\Projects\Study_Assignments\ML_OPS\HousingRegression> git remote add origin https://github.com/git-commit-acc/HousingRegression.git
(mlops-env) PS D:\Projects\Study_Assignments\ML_OPS\HousingRegression> git branch -M main
(mlops-env) PS D:\Projects\Study_Assignments\ML_OPS\HousingRegression> git push -u origin main
info: please complete authentication in your browser...
To https://github.com/git-commit-acc/HousingRegression.git
 ! [rejected]        main -> main (fetch first)
error: failed to push some refs to 'https://github.com/git-commit-acc/HousingRegression.git'
```

```
(mlops-env) PS D:\Projects\Study_Assignments\ML_OPS\HousingRegression> git push origin main --force
Enumerating objects: 7, done.
Counting objects: 100% (7/7), done.
Delta compression using up to 20 threads
Compressing objects: 100% (4/4), done.
Writing objects: 100% (7/7), 921 bytes | 921.00 KiB/s, done.
Total 7 (delta 0), reused 0 (delta 0), pack-reused 0 (from 0)
remote: This repository moved. Please use the new location:
remote:   https://github.com/git-commit-acc/HousingRegression_practice.git
To https://github.com/git-commit-acc/HousingRegression.git
 + 71988f9...07b5982 main -> main (forced update)
(mlops-env) PS D:\Projects\Study_Assignments\ML_OPS\HousingRegression> git checkout -b reg_branch
Switched to a new branch 'reg_branch'
(mlops-env) PS D:\Projects\Study_Assignments\ML_OPS\HousingRegression> git checkout reg_branch
Already on 'reg_branch'
(mlops-env) PS D:\Projects\Study_Assignments\ML_OPS\HousingRegression> python regression.py
Linear Model: MSE = 24.29, R2 = 0.67
Decision Tree: MSE = 11.18, R2 = 0.85
Random Forest: MSE = 8.73, R2 = 0.88
(mlops-env) PS D:\Projects\Study_Assignments\ML_OPS\HousingRegression> git checkout hyper_branch
error: pathspec 'hyper_branch' did not match any file(s) known to git
(mlops-env) PS D:\Projects\Study_Assignments\ML_OPS\HousingRegression> git checkout -b hyper_branch
Switched to a new branch 'hyper_branch'
(mlops-env) PS D:\Projects\Study_Assignments\ML_OPS\HousingRegression> New-Item hyper.py -type file


    Directory: D:\Projects\Study_Assignments\ML_OPS\HousingRegression


Mode                 LastWriteTime         Length Name
----                 -------------         ------ ----
-a----          7/6/2025   1:09 AM              0 hyper.py


(mlops-env) PS D:\Projects\Study_Assignments\ML_OPS\HousingRegression> python hyper.py

Tuning Ridge...
Best Params for Ridge: {'alpha': 0.1, 'fit_intercept': True, 'solver': 'svd'}
Ridge - MSE: 24.30, R²: 0.67

Tuning Decision Tree...
Best Params for Decision Tree: {'criterion': 'squared_error', 'max_depth': 20, 'min_samples_split': 2}
Decision Tree - MSE: 22.23, R²: 0.70

Tuning Random Forest...
Best Params for Random Forest: {'max_depth': 10, 'min_samples_split': 2, 'n_estimators': 50}
Random Forest - MSE: 9.44, R²: 0.87
(mlops-env) PS D:\Projects\Study_Assignments\ML_OPS\HousingRegression>
```

## Branching Strategy and Usage

- **main**: contains README initially
- **reg_branch**: basic regression models (Linear, Decision Tree, Random Forest)
- **hyper_branch**: GridSearchCV hyperparameter tuning

### Run:

```
git checkout reg_branch
python regression.py

git checkout hyper_branch
python hyper.py
```

---

## Dataset Overview

- Dataset Source: http://lib.stat.cmu.edu/datasets/boston
- Records: 506
- Features: 13
- Target: MEDV (Median house value)
- Train-Test Split: 80%-20%

**Note:** No feature scaling required for tree models. For linear models, optional scaling can be done.

---

## Models Evaluated

### 1. Ridge Regression (L2 Regularization)

- Hyperparameters: `alpha`, `solver`, `fit_intercept`

### 2. Decision Tree Regressor

- Hyperparameters: `max_depth`, `min_samples_split`, `criterion`

### 3. Random Forest Regressor

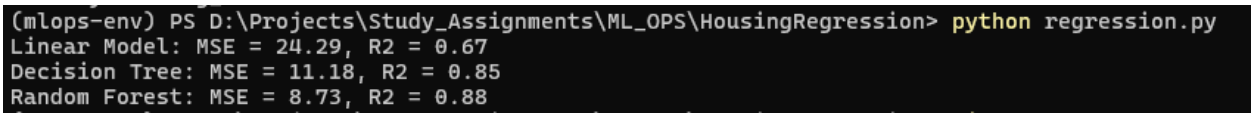
- Hyperparameters: `n_estimators`, `max_depth`, `min_samples_split`

---

## Results Summary

### Regression Baseline

- **Linear Regression**: MSE: 24.29, $R^2$: 0.67
- **Decision Tree**: MSE: 11.18, $R^2$: 0.85
- **Random Forest**: MSE: 8.73, $R^2$: 0.88



### After Hyperparameter Tuning

- **Ridge**
  Best Params: `{alpha: 0.1, fit_intercept: True, solver: 'auto'}`
  MSE: 24.30, $R^2$: 0.67

- **Decision Tree**
  Best Params: `{criterion: 'squared_error', max_depth: 20, min_samples_split: 10}`
  MSE: 22.23, $R^2$: 0.70

- **Random Forest**
  Best Params: `{n_estimators: 50, max_depth: 10, min_samples_split: 2}`
  MSE: 9.44, $R^2$: 0.87

```
(mlops-env) PS D:\Projects\Study_Assignments\ML_OPS\HousingRegression> python hyper.py

Tuning Ridge...
Best Params for Ridge: {'alpha': 0.1, 'fit_intercept': True, 'solver': 'svd'}
Ridge - MSE: 24.30, R²: 0.67

Tuning Decision Tree...
Best Params for Decision Tree: {'criterion': 'squared_error', 'max_depth': 20, 'min_samples_split': 2}
Decision Tree - MSE: 22.23, R²: 0.70

Tuning Random Forest...
Best Params for Random Forest: {'max_depth': 10, 'min_samples_split': 2, 'n_estimators': 50}
Random Forest - MSE: 9.44, R²: 0.87
(mlops-env) PS D:\Projects\Study_Assignments\ML_OPS\HousingRegression>
```

---

## CI/CD Workflow

- GitHub Actions executes `regression.py` or `hyper.py` on push to corresponding branches.
- Checks Python setup, dependency installation, dataset load, and script success.
- Defined in `.github/workflows/ci.yml`

---

## Key Insights

1. Random Forest consistently performs best in both baseline and tuned forms.
2. Hyperparameter tuning notably improves decision trees and Ridge performance.
3. Tree-based models better capture complex patterns than linear models.
4. Ridge remains a fast, interpretable baseline option.

---

## Conclusion

This assignment demonstrates the importance of structured pipelines in MLOps. Automating model training and evaluation with GitHub Actions ensures repeatability and

reliability. Among models, tuned Random Forest achieved the highest accuracy, while Ridge offered simplicity with moderate performance. The end-to-end workflow reflects the best practices in deploying reproducible machine learning solutions.

---

**Submitted by:** Ajinkya Ghodake (G24AI1046)