

Assignment

Build, Track, Package, Deploy and Monitor an ML Model using MLOps
Best Practices with California Housing Dataset

Subject

MLOps – AIMLCZG523

Group

Group-16

Contributors

Anup Jindal - 2023ac05472 – 100%

Yogesh Chaturvedi - 2023ac05167 – 100%

Debasish Acharya - 2023ac05417 - 100%

Anit Nair - 2023ac05503 – 100%

Project Content Details

Github Repo

<https://github.com/nairanit25/MTechAssignments>

DockerHub Repo

<https://hub.docker.com/repository/docker/nairanit25/housing-mlops/general>

<https://hub.docker.com/r/nairanit25/housing-mlops>

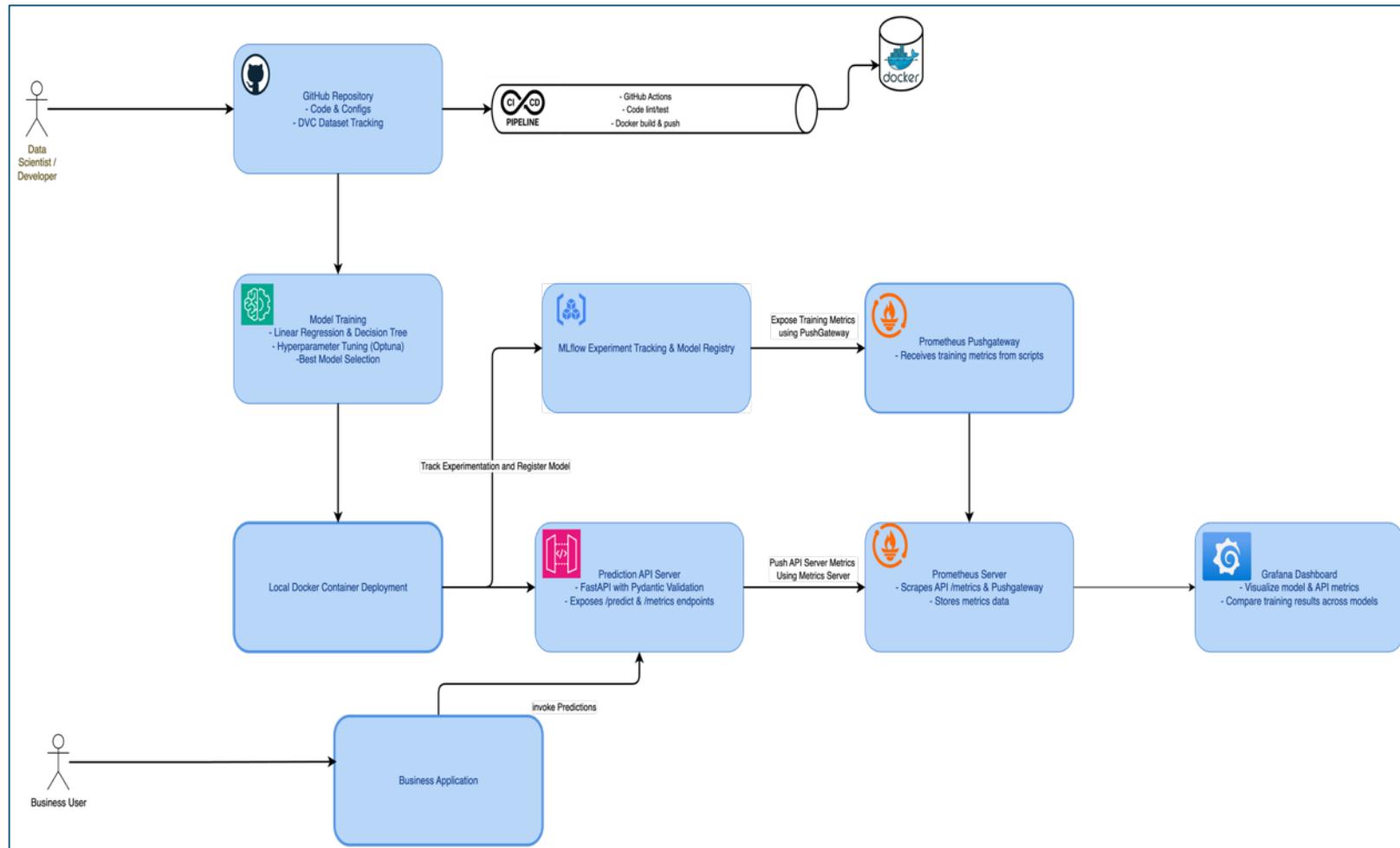
End-To-End Project Demo Video Link

<https://drive.google.com/drive/folders/18w9lkrFTBUEahCCMlFVG1FHTRQaePcn8?usp=sharing>

Note: The video captures complete end-to-end, hence spans 20 minutes. This summary documents captures each activity screenshot demonstrated in the video.

Architecture Overview (Part-6)

The below diagram represents the MLOps design for housing price prediction assignment. This assignment implements a comprehensive ML lifecycle for housing price prediction, incorporating data versioning, model training, experiment tracking, model registration through MLFlow, deployment, and monitoring.



Architecture Summary (Part-6)

Data and Code Versioning

Source code and configurations are managed in a **GitHub repository** to track all changes. The California Housing dataset is versioned using **DVC**, ensuring consistent reproducibility of data and experiments. The project maintains a clear directory structure to separate code, data, and models.

Model Development and Tracking

Two models — **Linear Regression** and **Decision Tree** — are trained using Python scripts. Hyperparameter tuning is automated with **Optuna**. **ydata_profiling** was used for basic data exploration and validation. All experiments, parameters, and evaluation metrics are logged to **MLflow Tracking Server**. The best model is selected and registered in the **MLflow Model Registry** for deployment.

Model Retraining on dataset change is retriggered using a background watcher.

API and Containerization

A prediction API is built using **FastAPI** that accepts JSON inputs and returns predictions. Input validation is handled via **Pydantic**. The API exposes Prometheus-compatible `/metrics` endpoints. The entire service is containerized using **Docker**, simplifying deployment across environments.

CI/CD Pipeline

GitHub Actions automate code testing, Docker image building, and pushing to Docker Hub. Deployment scripts enable running containers locally or on cloud infrastructure. This pipeline guarantees automated, repeatable, and consistent deployments.

Logging and Monitoring

Incoming prediction requests and responses are logged for audit and debugging. Custom metrics (training RMSE, R²; API request rates, latencies) are captured via the **Prometheus client**. Training metrics are pushed to a **Prometheus Pushgateway** from local

training scripts. Prometheus scrapes both the API and Pushgateway endpoints. Visualization and alerting are set up using **Grafana dashboards** that compare model metrics across training runs.

Grafana Dashboard captures following metrics

- **Prediction Count**
 - **Prediction Request Latency (p95)**
 - **CPU Usage**
 - **Memory Usage**
 - **Request Rate**
 - **Model RMSE Comparison**
 - **Model R2 Comparison**
-

This architecture ensures an end-to-end robust ML system incorporating version control, experiment tracking, containerized deployment, continuous integration, and real-time monitoring—enabling reproducibility, maintainability, and scalability.

Project Execution: End-To-End Steps

Git Repo Setup/Repository Structure (Part-1)

The screenshot shows a GitHub repository page for 'nairanit25/MTechAssignments'. The repository is private. The main branch is 'main' with 1 branch and 0 tags. There are 98 commits from user 'anitnair'. The commits are listed below:

Commit	Message	Time Ago
.dvc	stop tracking data/Housing.csv	last week
.github/workflows	Update ci_pipeline.yaml	2 days ago
Docker	Add test cases	2 days ago
configs	Registered best model back into registry	3 days ago
data	stop tracking data/Housing.csv	16 hours ago
monitoring	Add R2 and RMSE to Prometheus	2 days ago
src	Add R2 and RMSE to Prometheus	2 days ago
test	Add test cases	2 days ago
.dockerignore	Fix logging	3 days ago
.dvcignore	stop tracking data/Housing.csv	last week
.gitignore	docker-compose changes	5 days ago
README.md	Read Me file	3 days ago
docker-compose.yml	Add R2 and RMSE to Prometheus	2 days ago

About
No description, website, or topics provided.

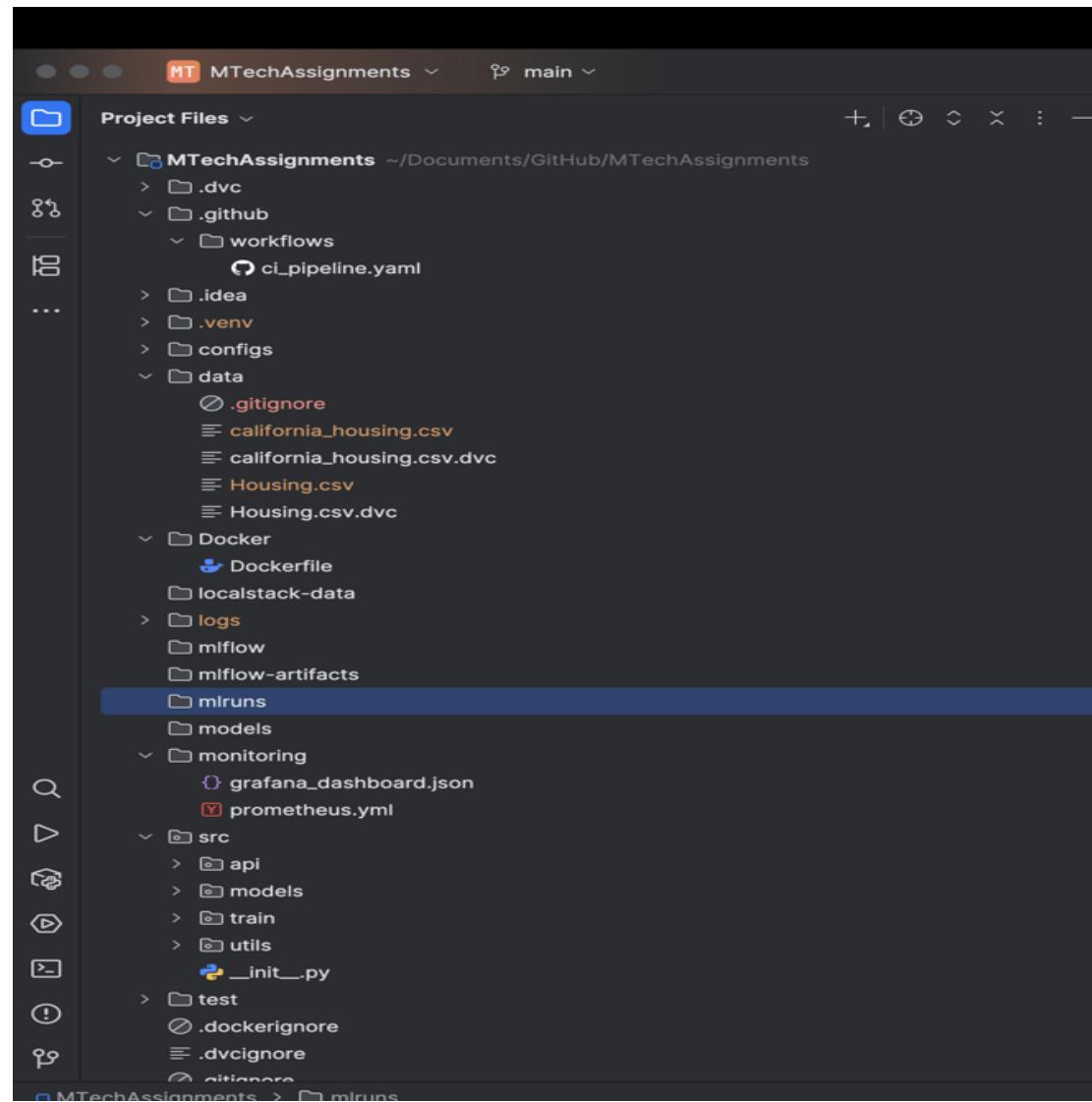
Readme
Activity
0 stars
0 watching
0 forks

Releases
No releases published
[Create a new release](#)

Packages
No packages published
[Publish your first package](#)

Languages
Python 97.7% Dockerfile 2.3%

Code pulled from Git and Data files pulled from DVC (Part-1)



Data versioning using DVC (Part-1)

```
~/Documents/GitHub/MTechAssignments -- zsh
[.venv] I060587@M2P4LW36YN MTechAssignments % git checkout
Your branch is up to date with 'origin/main'.
[.venv] I060587@M2P4LW36YN MTechAssignments % dvc checkout
Building workspace index
Comparing indexes
Applying changes
A     data/Housing.csv
A     data/california_housing.csv
(.venv) I060587@M2P4LW36YN MTechAssignments % ] + ] |1.00 [00:00, 115entry/s]
|4.00 [00:00, 2.61kentry/s]
|2.00 [00:00, 420file/s]
```

Basic EDA using YData Profiling (Part-1)

YData Profiling Report

Overview Alerts 8 Reproduction

Dataset statistics

Number of variables	9
Number of observations	20640
Missing cells	0
Missing cells (%)	0.0%
Duplicate rows	0
Duplicate rows (%)	0.0%
Total size in memory	1.4 MiB
Average record size in memory	72.0 B

Variable types

Numeric	9
---------	---

Variables

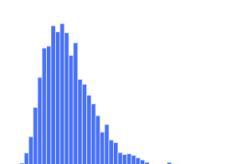
Select Columns ▾

MedInc

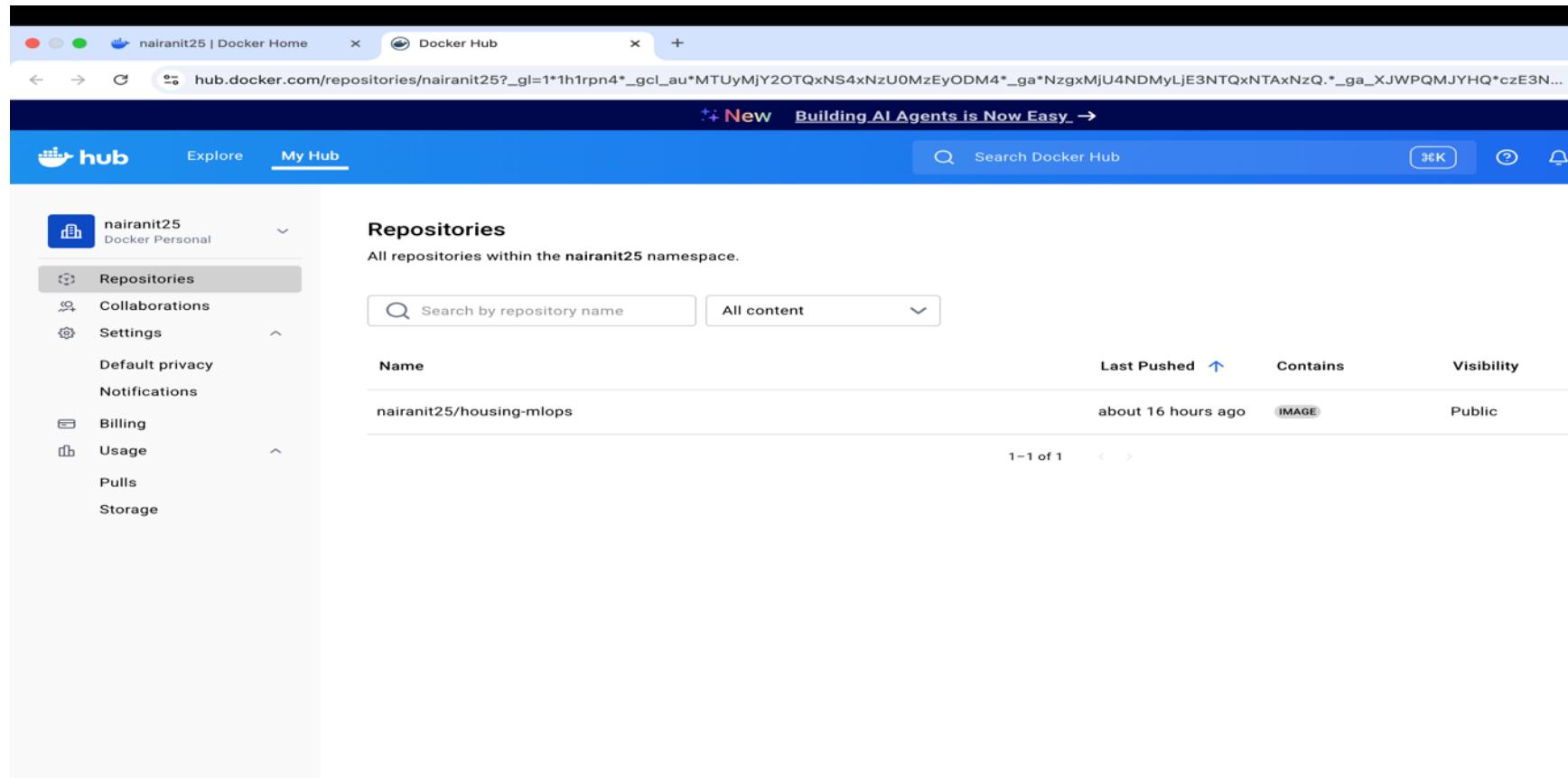
Real number (ℝ)

High correlation

Distinct	12928	Minimum	0.4999
Distinct (%)	62.6%	Maximum	15.0001
Missing	0	Zeros	0
Missing (%)	0.0%	Zeros (%)	0.0%
NaNs	0	NaNs (%)	0



DockerHub Repository (Part-4 CI/CD)



The screenshot shows the DockerHub interface for the user 'nairanit25'. The left sidebar contains navigation links for 'Repositories' (selected), 'Collaborations', 'Settings', 'Default privacy', 'Notifications', 'Billing', 'Usage', 'Pulls', and 'Storage'. The main content area is titled 'Repositories' and displays a single repository: 'nairanit25/housing-mlops'. This repository was last pushed 'about 16 hours ago' and is marked as 'IMAGE'. The visibility is set to 'Public'. A search bar at the top right allows users to search Docker Hub.

Name	Last Pushed	Contains	Visibility
nairanit25/housing-mlops	about 16 hours ago	IMAGE	Public

Commit to trigger Github Actions (Part-4 CI/CD)

The screenshot shows a GitHub commit interface for a repository named "MTechAssignments".

Top Bar: Current Repository: MTechAssignments, Current Branch: main, Fetch origin: Last fetched 7 minutes ago.

Changes Tab: Shows 11 changed files. The list includes: .DS_Store, .idea/.gitignore, .idea/libraries/conf.xml, .idea/libraries/my_test_package.xml, .idea/misc.xml, .idea/modules.xml, .idea/MTechAssignments.iml, .idea/vcs.xml, data/.DS_Store, data/.gitignore, and README.md (which is checked).

Stashed Changes: A stash named "ReadMe update" is shown. It contains a single file, README.md, with the following content:

```
@@ -0,0 +1,2 @@
1 + /california_housing.csv
2 + /Housing.csv
```

Commit Buttons: Commit 1 file to main.

Github Actions Triggered for commit (Part-4 CI/CD)

The screenshot shows the GitHub Actions interface for the repository `nairanit25/MTechAssignments/actions`. The left sidebar contains navigation links for Code, Issues, Pull requests, Actions, Projects, Security, Insights, and Settings. The main content area is titled "All workflows" and displays "31 workflow runs". A search bar at the top right allows filtering of workflow runs.

Workflow Name	Event	Status	Branch	Actor
ReadMe update	now	In progress	main	nairanit25
stop tracking data/Housing.csv	16 hours ago	6m 52s	main	nairanit25
Data files	16 hours ago	7m 17s	main	nairanit25
stop tracking data/california_housing.csv	16 hours ago	6m 56s	main	nairanit25
Add R2 and RMSE to Prometheus	2 days ago	7m 3s	main	nairanit25
Update ci_pipeline.yaml	2 days ago	7m 4s	main	nairanit25
Add test cases	2 days ago	1m 21s	main	nairanit25
Update ci_pipeline.yaml	2 days ago	-	main	nairanit25

Continuous Build and Test (Part-4 CI/CD)

The screenshot shows a GitHub Actions pipeline run details page. The URL in the browser is github.com/nairanit25/MTechAssignments/actions/runs/16856892743/job/47751218765. The pipeline name is **ReadMe update #31**.

The pipeline consists of two jobs:

- build-and-test** (Succeeded): This job includes steps for Set up job, Checkout repository, Set up Python, Install dependencies, Run tests, Check DVC data tracking, Verify MLflow Tracking URI, Post Set up Python, Post Checkout repository, and Complete job.
- docker-build-and-push** (In Progress): This job is currently running.

On the left sidebar, there are sections for Summary, Jobs (selected), Run details, Usage, and Workflow file.

At the top right, there are buttons for Cancel workflow and three dots for more options.

Continuous Test (Part-4 CI/CD)

build-and-test

succeeded 20 hours ago in 1m 22s

Search logs  

Step	Description	Time
>  Set up Python		0s
>  Install dependencies		1m 7s
▼  Run tests		5s
<pre>1 ► Run pip install pytest 12 Requirement already satisfied: pytest in /opt/hostedtoolcache/Python/3.12.11/x64/lib/python3.12/site-packages (8.3.2) 13 Requirement already satisfied: iniconfig in /opt/hostedtoolcache/Python/3.12.11/x64/lib/python3.12/site-packages (from pytest) (2.1.0) 14 Requirement already satisfied: packaging in /opt/hostedtoolcache/Python/3.12.11/x64/lib/python3.12/site-packages (from pytest) (25.0) 15 Requirement already satisfied: pluggy<2,>=1.5 in /opt/hostedtoolcache/Python/3.12.11/x64/lib/python3.12/site-packages (from pytest) (1.6.0) 16 ===== test session starts ===== 17 platform linux -- Python 3.12.11, pytest-8.3.2, pluggy-1.6.0 18 rootdir: /home/runner/work/MTechAssignments/MTechAssignments 19 plugins: hydra-core-1.3.2, typeguard-4.4.4, anyio-4.10.0, mock-3.14.0 20 collected 9 items 21 22 test/test_model_registry.py [100%] 23 24 ===== warnings summary ===== 25 ../../../../../../opt/hostedtoolcache/Python/3.12.11/x64/lib/python3.12/site-packages/pydantic/_internal/_config.py:323 26 /opt/hostedtoolcache/Python/3.12.11/x64/lib/python3.12/site-packages/pydantic/_internal/_config.py:323: PydanticDeprecatedSince20: Support for class-based `config` is deprecated, use ConfigDict instead. Deprecated in Pydantic V2.0 to be removed in V3.0. See Pydantic V2 Migration Guide at https://errors.pydantic.dev/2.11/migration/ 27 warnings.warn(DEPRECATION_MESSAGE, DeprecationWarning) 28 29 -- Docs: https://docs.pytest.org/en/stable/how-to/capture-warnings.html 30 ===== 9 passed, 1 warning in 2.23s =====</pre>		

Continuous Deployment: Docker-Build and Push (Part-4 CI/CD)

The screenshot shows a GitHub Actions workflow named "docker-build-and-push" that has just succeeded. The workflow consists of several steps:

- Set up job (1s)
- Checkout repository (1s)
- Set up Docker Buildx (7s)
- Log in to DockerHub (0s)
- Build and push Docker image (2m 28s)
- Post Build and push Docker image (0s)
- Post Log in to DockerHub (0s)
- Post Set up Docker Buildx (4s)
- Post Checkout repository (0s)
- Complete job (0s)

The "build-and-test" job is also listed as completed.

Continuous Deployment: Deploy (Part-4 CI/CD)

The screenshot shows a GitHub Actions pipeline run details page. The URL in the browser is github.com/nairanit25/MTechAssignments/actions/runs/16856892743/job/47751320262. The pipeline name is "ReadMe update #31".

The left sidebar shows the pipeline structure:

- Summary
- Jobs
 - build-and-test (green checkmark)
 - docker-build-and-push (green checkmark)
 - deploy** (highlighted with a blue bar, green checkmark)
- Run details
- Usage
- Workflow file

The main content area displays the "deploy" job details:

deploy
succeeded now in 2m 33s

Search logs

Log steps:

- > Set up job
- > Checkout repo
- > Install Docker Compose
- > Deploy using Docker Compose
- > Post Checkout repo
- > Complete job

Github Actions: Successfully triggered the pipeline and completed for the commit (Part-4 CI/CD)

The screenshot shows a GitHub Actions pipeline run summary for a commit. The pipeline consists of three jobs: build-and-test, docker-build-and-push, and deploy. All jobs completed successfully with a total duration of 6m 58s.

Pipeline Summary:

- Triggered via push 7 minutes ago
- Status: Success
- Total duration: 6m 58s
- Artifacts: -

Workflow File: ci_pipeline.yaml

On: push

```
graph LR; A[build-and-test] -- "1m 28s" --> B[docker-build-and-push]; B -- "2m 44s" --> C[deploy];
```

Jobs:

- build-and-test
- docker-build-and-push
- deploy

Run details:

- Usage
- Workflow file

Docker Image: Latest Image pushed to Dockerhub (Part-4 CI/CD)

The screenshot shows a web browser window displaying the Docker Hub interface. The URL in the address bar is `hub.docker.com/repositories/nairanit25`. The page title is "hub". The main content area is titled "Repositories" and shows a single repository entry:

Name	Last Pushed	Contains	Visibility
nairanit25/housing-mlops	2 minutes ago	IMAGE	Public

On the left side, there is a sidebar with the following menu items:

- Repositories (selected)
- Collaborations
- Settings
- Default privacy
- Notifications
- Billing
- Usage
- Pulls
- Storage

How likely are you to recommend Docker Hub to another developer?

Not at all likely 0 1 2 3 4 5 6 7 8 9 10 Extremely likely

powered by InMoment

Continuous Deployment: Trigger Deploy to Local Docker (Part-4 CI/CD)

```
...GitHub/MTechAssignments — docker-compose up --build --force-recreate  ↵
[(.venv) I0e05870M2P4LW36YN MTechAssignments % git checkout
Your branch is up to date with 'origin/main'.
[(.venv) I0e05870M2P4LW36YN MTechAssignments % dvc checkout
Building workspace index
Comparing indexes
Applying changes
A     data/Housing.csv
A     data/california_housing.csv
[(.venv) I0e05870M2P4LW36YN MTechAssignments % docker-compose up --build --force-recreate
[+] Running 1/27
  grafana [           ] Pulling
    2c9f7a78edd5 Pulling fs layer
    1b3ff0832fcf Pulling fs layer
    8a0da51253d0 Pulling fs layer
    85130e653e20 Pulling fs layer
    70c28aee5b48 Pulling fs layer
    47662a07db22 Pulling fs layer
    261b7bddf7e3 Pulling fs layer
    d2fcc48b8f71 Pulling fs layer
    6e174226ea69 Pulling fs layer
  localstack Pulling
  prometheus [           ] Pulling
    197a814c6a2c Pulling fs layer
    d932d57a546b Pulling fs layer
    7fb55c4e976b Pulling fs layer
    feadeb862692 Pulling fs layer
    76619c1988eb Pulling fs layer
    2dfc78ad941 Pulling fs layer
    fd1d3a5a5f79 Pulling fs layer
    95d87e5ba19a Pulling fs layer
  pushgateway [           ] Pulling
    6ce8bb87a9754 Pulling fs layer
    4f4fb700ef54 Pulling fs layer
    d2f8aae8d888 Pulling fs layer
    1d0712eaef887 Pulling fs layer
    f255518b7291 Pulling fs layer
  mlflow Pulled

```

Deployment Status: All required services and the API server deployed to Local Docker (Part-4 CI/CD)

The screenshot shows the Docker Desktop interface. The left sidebar has 'Containers' selected. The main area displays container statistics and a list of running containers.

Container CPU usage: 0.57% / 1200% (12 CPUs available)

Container memory usage: 956.61MB / 7.47GB

Actions: Show charts

Name	Container ID	Image	Port(s)	CPU (%)	Last started	Actions
mtechassignments	-	-	-	0.57%	3 minutes ago	[Edit]
mlflow-server	7ede9b63cb2b	python:3.12-slim	5000:5000	0.03%	3 minutes ago	[Edit]
pushgateway	1ba11096077b	prom/pushgateway:latest	9091:9091	0%	3 minutes ago	[Edit]
prometheus	c080c8e425c0	prom/prometheus:latest	9090:9090	0%	3 minutes ago	[Edit]
localstack	85011fe572d6	localstack/localstack	4566:4566	0.07%	3 minutes ago	[Edit]
grafana-1	f257865b4829	grafana/grafana:latest	3000:3000	0.47%	3 minutes ago	[Edit]
api-1	b258a10c81ba	mtechassignments-api	8000:8000	0%	3 minutes ago	[Edit]

Showing 7 items

Walkthroughs:

- Multi-container applications
- Containerize your application

Application Status: API Server up with no model registered yet! (part-4)

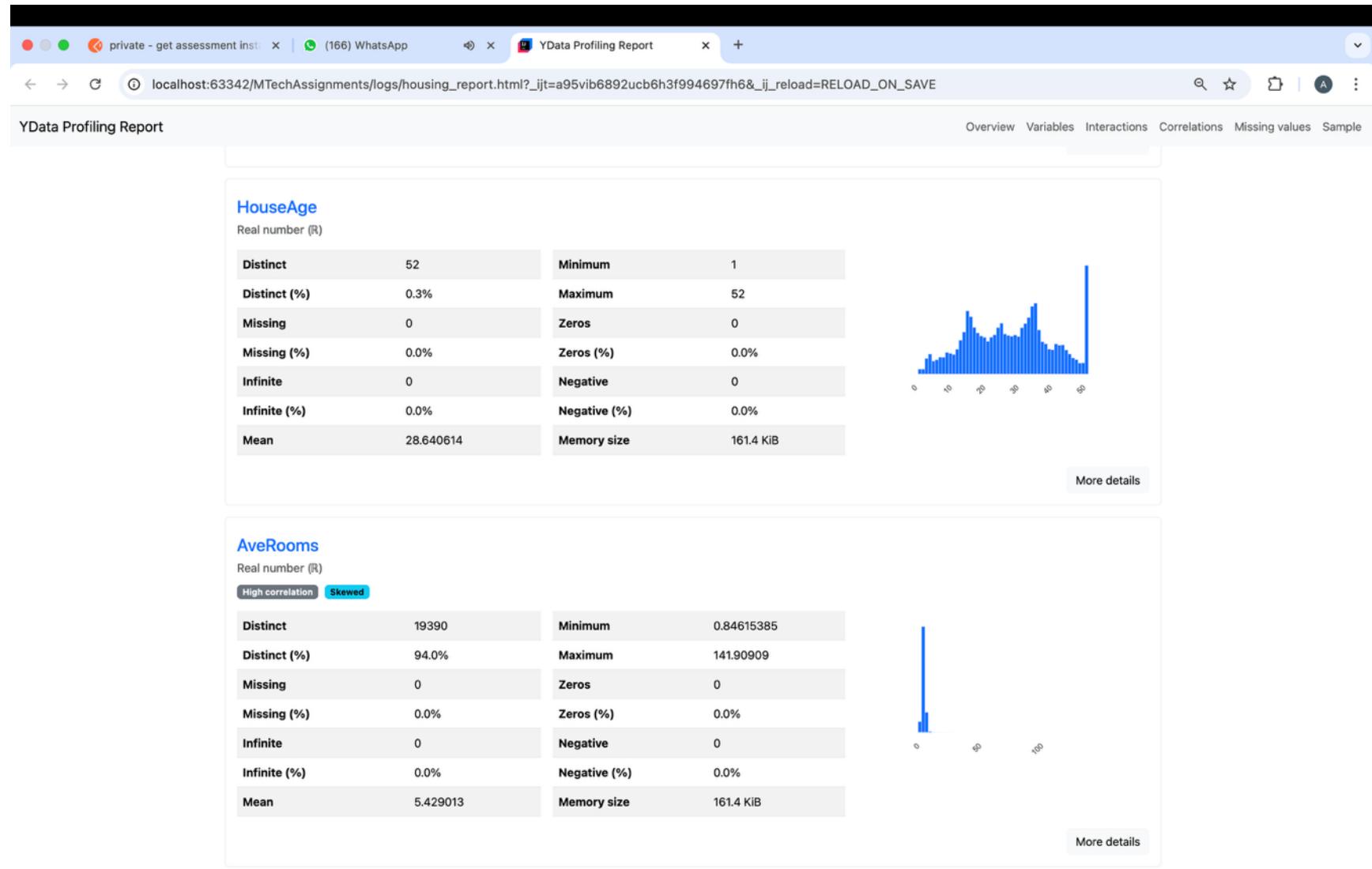
The screenshot shows the Docker Desktop interface with the following details:

- Left Sidebar:** Contains links for Ask Gordon (BETA), Containers (selected), Images, Volumes, Builds, Models (BETA), MCP Toolkit (BETA), Docker Hub, Docker Scout, and Extensions.
- Top Bar:** Shows "docker.desktop PERSONAL", a search bar, and various icons for settings and notifications.
- Container Details:** The container name is "mtechassignments-api-1". It has an ID "f9d1b364a9ed" and is running on port "8000:8000".
- Status:** The status is "Running (5 minutes ago)" with a green "OK" icon.
- Logs Tab:** Active tab showing the application logs. The logs output the following text:

```
INFO: Started server process [1]
INFO: Waiting for application startup.
2025-08-10 04:11:24,021 - src.api.app - INFO - app.py:32 - Starting MLOps Housing Price Prediction Inference App
2025-08-10 04:11:24,022 - src.api.app - INFO - app.py:34 - Starting to load model from registry
2025-08-10 04:15:30,296 - src.models.model_registry - INFO - model_registry.py:42 - No registered models found.
2025-08-10 04:15:30,297 - src.api.app - INFO - app.py:49 - No Available models found in the registry: []
INFO: Application startup complete.
INFO: Uvicorn running on http://0.0.0.0:8000 (Press CTRL+C to quit)
INFO: 172.18.0.5:44698 - "GET /metrics HTTP/1.1" 307 Temporary Redirect
INFO: 172.18.0.5:44698 - "GET /metrics/ HTTP/1.1" 200 OK
INFO: 172.18.0.5:39658 - "GET /metrics HTTP/1.1" 307 Temporary Redirect
INFO: 172.18.0.5:39658 - "GET /metrics/ HTTP/1.1" 200 OK
INFO: 172.18.0.5:58336 - "GET /metrics HTTP/1.1" 307 Temporary Redirect
INFO: 172.18.0.5:58336 - "GET /metrics/ HTTP/1.1" 200 OK
INFO: 172.18.0.5:38916 - "GET /metrics HTTP/1.1" 307 Temporary Redirect
INFO: 172.18.0.5:38916 - "GET /metrics/ HTTP/1.1" 200 OK
INFO: 127.0.0.1:43226 - "GET /health HTTP/1.1" 200 OK
INFO: 172.18.0.5:33472 - "GET /metrics HTTP/1.1" 307 Temporary Redirect
INFO: 172.18.0.5:33472 - "GET /metrics/ HTTP/1.1" 200 OK
```

On the right side, there are several blue icons for actions like copy, share, and refresh.

Optuna Hyperparameter Tuning Report (Part 2)



MLFlow Experiments Recording (Part-2)

The screenshot shows the MLflow UI interface. At the top, there's a toolbar with various icons. Below it is a header bar with the 'mlflow' logo, version '3.0.0', and tabs for 'Experiments', 'Models', and 'Prompts'. On the right of the header are buttons for 'GitHub' and 'Docs'. The main area is titled 'Experiments' and shows a table of runs for the experiment 'housing_price_prediction_training'. The table has columns for 'Run Name', 'Created', 'Dataset', 'Duration', 'Source', and 'Models'. There are two rows in the table, both labeled 'model_comparison'. Both rows show a creation time of '2 hours ago' and a duration of '4.7s'. The 'Source' column indicates they used 'train_m...' datasets. The 'Models' column is empty. At the bottom of the table, it says '4 matching runs'.

Run Name	Created	Dataset	Duration	Source	Models
model_comparison	2 hours ago	-	4.7s	train_m...	
model_comparison	2 hours ago	-	6.1s	train_m...	
					(+)

4 matching runs

MLFlow Best Model in Registry (Part-2)

The screenshot shows the MLflow UI interface on a web browser. The URL in the address bar is `localhost`. The top navigation bar includes links for Experiments, Models, and Prompts, along with GitHub and Docs links. The main content area is titled "MLOps_Housing_Price_Prediction_Best_Model_Registration_Experiment". On the left, there is a sidebar titled "Experiments" with a search bar and two entries: "MLOps_Housing_Price_Prediction_Best_Model_Registration_Experiment" (selected) and "housing_price_prediction_training". The main panel displays a table of "Runs" under the "Experimental" tab. The table has columns: Run Name, Created, Dataset, Duration, Source, and Models. One row is visible, showing "Registering_housing_p..." created "2 hours ago" with a duration of "2.5s", sourced from "best_m..." and associated with "model". A footer at the bottom indicates "1 matching run".

Run Name	Created	Dataset	Duration	Source	Models
Registering_housing_p...	2 hours ago	-	2.5s	best_m...	model

Train Linear Regression Model – California Housing (Part-2)

```
(.venv) I060587@M2P4LW36YN MTechAssignments % python -m src.train.train_models --data-path data/california_housing.csv --algorithms 'decision_tree'
/Users/I060587/Documents/GitHub/MTechAssignments/.venv/lib/python3.9/site-packages/urllib3/_init__.py:35: NotOpenSSLWarning: urllib3 v2 only supports OpenSSL 1.1.1+, currently the 'ssl' module is compiled with 'Li
bressl 2.8.3'. See: https://github.com/urllib3/urllib3/issues/3020
    warnings.warn(
2025-08-10 09:50:824 - __main__ - INFO - train_models.py:230 - Starting Training parameters data-path: data/california_housing.csv, trails: 5, algorithms: ['decision_tree']
Upgrade to ydata-sdk
Improve your data and profiling with ydata-sdk, featuring data quality scoring, redundancy detection, outlier identification, text validation, and synthetic data generation.
Register at https://ydata.ai/register
100%|██████████| 9/9 [00:00<00:00, 276.68it/s]
Summarize dataset: 100%|██████████| 99/99 [00:03<00:00, 28.06it/s, Completed]
Generate report structure: 100%|██████████| 1/1 [00:01<00:00, 1.07s/it]
Render HTML: 100%|██████████| 1/1 [00:00<00:00, 1.18it/s]
Export report to file: 100%|██████████| 1/1 [00:00<00:00, 13.13it/s]
2025-08-10 09:33:57,810 - __main__ - INFO - train_models.py:247 - Train: 14448, Val: 3096, Test: 3096
2025-08-10 09:33:57,952 - __main__ - INFO - train_models.py:254 - Starting Training, Algo to train: decision_tree
/Users/I060587/Documents/GitHub/MTechAssignments/.venv/lib/python3.9/site-packages/sklearn/base.py:493: UserWarning: X does not have valid feature names, but DecisionTreeRegressor was fitted with feature names
    warnings.warn(
/Users/I060587/Documents/GitHub/MTechAssignments/.venv/lib/python3.9/site-packages/sklearn/base.py:493: UserWarning: X does not have valid feature names, but DecisionTreeRegressor was fitted with feature names
    warnings.warn(
2025/08/10 09:33:59 WARNING mlflow.models.model: 'artifact_path' is deprecated. Please use 'name' instead.
2025/08/10 09:34:02 WARNING mlflow.models.model: Model logged without a signature and input example. Please set 'input_example' parameter when logging the model to auto infer the model signature.
Registered model 'decision_tree_housing_price_predictor' already exists. Creating a new version of this model...
2025/08/10 09:34:02 INFO mlflow.store.model_registry.abstract_store: Waiting up to 300 seconds for model version to finish creation. Model name: decision_tree_housing_price_predictor, version 2
Created version '2' of model 'decision_tree_housing_price_predictor'.
2025-08-10 09:34:02,468 - __main__ - INFO - train_models.py:198 - Decision Tree - Val R2: 0.6831, Val RMSE: 0.64
    ↗ View run whimsical-donkey-894 at: http://localhost:5000/#/experiments/967784594619596523/runs/bc7e479d846c4c63b41f20239afc3662
    ✓ View experiment at: http://localhost:5000/#/experiments/967784594619596523
    ↗ View run model_comparison at: http://localhost:5000/#/experiments/967784594619596523/runs/a5a34f4be27443918b14b39eefa37a25
    ✓ View experiment at: http://localhost:5000/#/experiments/967784594619596523
2025-08-10 09:34:02,497 - __main__ - INFO - train_models.py:301 -
=====
2025-08-10 09:34:02,497 - __main__ - INFO - train_models.py:302 - TRAINING RESULTS SUMMARY
2025-08-10 09:34:02,497 - __main__ - INFO - train_models.py:303 - =====
2025-08-10 09:34:02,497 - __main__ - INFO - train_models.py:310 - decision_tree: R2 = 0.6831
2025-08-10 09:34:02,497 - __main__ - INFO - train_models.py:312 - Training completed!
(.venv) I060587@M2P4LW36YN MTechAssignments %
```

Train Decision Tree Model – California Housing (Part-2)

```
...GitHub/MTechAssignments — docker-compose up --build --force-recreate          -/Documents/GitHub/MTechAssignments — zsh          -/Documents/GitHub/MTechAssignments — zsh +  
(.venv) I060587@M2P4LW36YN MTechAssignments % python -m src.train.train_models --data-path data/california_housing.csv --algorithms 'linear_regression'  
/Users/I060587/Documents/GitHub/MTechAssignments/.venv/lib/python3.9/site-packages/urllib3/_init__.py:35: NotOpenSSLSWarning: urllib3 v2 only supports OpenSSL 1.1.1+, currently the 'ssl' module is compiled with 'Li  
bressl 2.8.3'. See: https://github.com/urllib3/urllib3/issues/3020  
warnings.warn(  
2025-08-10 09:33:12,842 - __main__ - INFO - train_models.py:230 - Starting Training parameters data-path: data/california_housing.csv, trails: 5, algorithms: ['linear_regression']  
Upgrade to ydata-sdks  
Improve your data and profiling with ydata-sdks, featuring data quality scoring, redundancy detection, outlier identification, text validation, and synthetic data generation.  
Register at https://ydata.ai/register  
100%|██████████| 9/9 [00:00<00:00, 257.07it/s]  
Summarize dataset: 100%|██████████| 99/99 [00:03<00:00, 27.92it/s, Completed]  
Generate report structure: 100%|██████████| 1/1 [00:01<00:00, 1.13it/s]  
Render HTML: 100%|██████████| 1/1 [00:00<00:00, 1.12it/s]  
Export report to file: 100%|██████████| 1/1 [00:00<00:00, 8.13it/s]  
2025-08-10 09:33:23,911 - __main__ - INFO - train_models.py:247 - Train: 14448, Val: 3096, Test: 3096  
2025-08-10 09:33:23,911 - __main__ - INFO - train_models.py:247 - Starting Training. Algo: linear_regression  
2025-08-10 09:33:26 WARNING mlflow.models.model: 'artifact_dir' is deprecated. Please use 'name' instead.  
2025-08-10 09:33:29 WARNING mlflow.models.model: Model logged without a signature and input example. Please set 'input_example' parameter when logging the model to auto infer the model signature.  
Registered model 'linear_regression_housing_price_predictor' already exists. Creating a new version of this model...  
2025-08-10 09:33:29 INFO mlflow.store.model_registry.abstract_store: Waiting up to 300 seconds for model version to finish creation. Model name: linear_regression_housing_price_predictor, version 2  
Created version '2' of model 'linear_regression_housing_price_predictor'.  
2025-08-10 09:33:30,008 - __main__ - INFO - train_models.py:73 - Type of model_registered: <class 'mlflow.models.model.ModelInfo'>  
2025-08-10 09:33:30,008 - __main__ - INFO - train_models.py:76 - Registered Model Name: linear_regression_housing_price_predictor, Experiment Run Info (model-run-id): 083929daac5244d882ca5a2539785e34  
2025-08-10 09:33:30,008 - __main__ - INFO - train_models.py:77 - Registered Model Version: 2, model uri: /models:/m-cf844dc251c4e3db3b017fe279c25ad  
2025-08-10 09:33:30,008 - __main__ - INFO - train_models.py:78 - Flow Experiment run_id: 083929daac5244d882ca5a2539785e34  
2025-08-10 09:33:30,008 - __main__ - INFO - train_models.py:79 - Registered Model Metadata: None, Registered Model Param: {'alpha': '1.0', 'algorithm': 'linear_regression', 'regularization': 'ridge'}  
2025-08-10 09:33:30,249 - __main__ - INFO - train_models.py:116 - Linear Regression - Val R2: 0.5848, Val RMSE: 0.74  
➤ View run clean-gnu-509 at: http://localhost:5000/#/experiments/967784594619596523/runs/083929daac5244d882ca5a2539785e34  
✓ View experiment at: http://localhost:5000/#/experiments/967784594619596523  
➤ View run model_comparison at: http://localhost:5000/#/experiments/967784594619596523/runs/380d11b01d994a4890502523793f6704  
✓ View experiment at: http://localhost:5000/#/experiments/967784594619596523  
2025-08-10 09:33:30,283 - __main__ - INFO - train_models.py:301 - ======  
=====  
2025-08-10 09:33:30,283 - __main__ - INFO - train_models.py:302 - TRAINING RESULTS SUMMARY  
2025-08-10 09:33:30,283 - __main__ - INFO - train_models.py:303 - ======  
2025-08-10 09:33:30,283 - __main__ - INFO - train_models.py:310 - linear_regression: R2 = 0.5848  
2025-08-10 09:33:30,283 - __main__ - INFO - train_models.py:312 - Training completed!  
(.venv) I060587@M2P4LW36YN MTechAssignments %
```

Register best performing Model to MLFlow Model Registry (part-2)

```
...GitHub/MTechAssignments -- docker-compose up --build --force-recreate          ~/Documents/GitHub/MTechAssignments -- zsh                                     ~/Documents/GitHub/MTechAssignments -- zsh +  
[.venv] I0605870M2P4LW36YN MTechAssignments % python -m src.models.best_model_selector  
/Users/I060587/Documents/GitHub/MTechAssignments/.venv/lib/python3.9/site-packages/urllib3/_init__.py:35: NotOpenSSLWarning: urllib3 v2 only supports OpenSSL 1.1.1+, currently the 'ssl' module is compiled with 'Li  
bressl 2.8.3'. See: https://github.com/urllib3/urllib3/issues/3020  
    warnings.warn(  
2025-08-10 09:34:42,260 - src.models.model_registry - INFO - model_registry.py:159 -  
Successfully retrieved 2 most recent versions for 'linear_regression_housing_price_predictor'.  
2025-08-10 09:34:42,260 - __main__ - INFO - best_model_selector.py:51 -  
Best model found:  
2025-08-10 09:34:42,261 - __main__ - INFO - best_model_selector.py:52 - Model Name: linear_regression_housing_price_predictor, Version: 2  
2025-08-10 09:34:42,261 - __main__ - INFO - best_model_selector.py:53 - R2: 0.5848267710408837, RMSE: 0.7354415694664408  
2025-08-10 09:34:42,261 - __main__ - INFO - best_model_selector.py:51 -  
Best model found:  
2025-08-10 09:34:42,261 - __main__ - INFO - best_model_selector.py:52 - Model Name: linear_regression_housing_price_predictor, Version: 2  
2025-08-10 09:34:42,261 - __main__ - INFO - best_model_selector.py:53 - R2: 0.5848267710408837, RMSE: 0.7354415694664408  
2025-08-10 09:34:42,261 - __main__ - INFO - best_model_selector.py:88 - Best model to be registered in registry: {best_model_data}  
2025-08-10 09:34:42,318 - src.models.model_registry - INFO - model_registry.py:72 - Successfully retrieved details for model 'linear_regression_housing_price_predictor' version 2.  
2025-08-10 09:34:42,318 - src.models.model_registry - INFO - model_registry.py:115 - model_uri : models:/linear_regression_housing_price_predictor/2  
Downloading artifacts: 100% [██████████] 5 / 5 [00:00<00:00, 13851.73it/s]  
2025-08-10 09:34:42,382 - __main__ - INFO - best_model_selector.py:98 - loading model: linear_regression_housing_price_predictor from registry: Pipeline(steps=[('scaler', StandardScaler()), ('regressor', Ridge())])  
2025-08-10 09:34:42,382 - __main__ - INFO - best_model_selector.py:100 - Registering best model as target: housing_price_predictor  
2025-08-10 09:34:42,509 - src.models.model_registry - INFO - model_registry.py:173 - Logging and registering model under new name 'housing_price_predictor'...  
2025/08/10 09:34:42 WARNING milflow.models.model: 'artifact_path' is deprecated. Please use 'name' instead.  
2025/08/10 09:34:44 WARNING milflow.models.model: Model logged without a signature and input example. Please set 'input_example' parameter when logging the model to auto infer the model signature.  
Registered model 'housing_price_predictor' already exists. Creating a new version of this model...  
2025/08/10 09:34:44 INFO milflow.store.model_registry.abstract_store: Waiting up to 300 seconds for model version to finish creation. Model name: housing_price_predictor, version 2  
Created version '2' of model 'housing_price_predictor'.  
2025-08-10 09:34:44,912 - src.models.model_registry - INFO - model_registry.py:183 - Registered Model Name: housing_price_predictor, Experiment Run Info (model-run-id): b739b298d7dc460abe0b3c85dfc2cc0c  
2025-08-10 09:34:44,912 - src.models.model_registry - INFO - model_registry.py:184 - Registered Model version: 2, model_uri Info: models:/m-956660e5a8234fe9fb72fd4a3ffb024e  
2025-08-10 09:34:44,912 - src.models.model_registry - INFO - model_registry.py:185 - Registered Model Param: {}  
▲ View run Registering_housing_price_predictor at: http://localhost:5000/#/experiments/954136587637183516/runs/b739b298d7dc460abe0b3c85dfc2cc0c  
✓ View experiment at: http://localhost:5000/#/experiments/954136587637183516  
2025-08-10 09:34:45,076 - __main__ - INFO - best_model_selector.py:102 - Registered best model with target: housing_price_predictor  
[.venv] I0605870M2P4LW36YN MTechAssignments %
```

Successful Training Retrigger on dataset Change (Bonus Assignment)

```
...itHub/MTechAssignments — docker-compose up --build --force-recreate ... ...ents/GitHub/MTechAssignments — python -m src.train.watch_and_train .../Documents/GitHub/MTechAssignments — zsh +
```

sion_tree']
Upgrade to ydata-sdk
Improve your data and profiling with ydata-sdk, featuring data quality scoring, redundancy detection, outlier identification, text validation, and synthetic data generation.
Register at <https://ydata.ai/register>
100%|██████████| 9/9 [00:00<00:00, 239.80it/s]
Summarize dataset: 100%|██████████| 99/99 [00:03<00:00, 25.78it/s, Completed]
Generate report structure: 100%|██████████| 1/1 [00:01<00:00, 1.10s/it]
Render HTML: 100%|██████████| 1/1 [00:00<00:00, 1.16it/s]
Export report to file: 100%|██████████| 1/1 [00:00<00:00, 11.96it/s]
2025-08-10 18:31:11,768 - __main__ - INFO - train_models.py:247 - Train: 14446, Val: 3096, Test: 3096
2025-08-10 18:31:11,905 - __main__ - INFO - train_models.py:254 - Starting Training, Algo to train: decision_tree
/Users/I060587/Documents/GitHub/MTechAssignments/.venv/lib/python3.9/site-packages/sklearn/base.py:493: UserWarning: X does not have valid feature names, but DecisionTreeRegressor was fitted with feature names
warnings.warn
/Users/I060587/Documents/GitHub/MTechAssignments/.venv/lib/python3.9/site-packages/sklearn/base.py:493: UserWarning: X does not have valid feature names, but DecisionTreeRegressor was fitted with feature names
warnings.warn
2025-08/10 18:31:13 WARNING miflow.models.model: 'artifact_path' is deprecated. Please use 'name' instead.
2025-08/10 18:31:16 WARNING miflow.models.model: Model logged without signature and input example. Please set 'input_example' parameter when logging the model to auto infer the model signature.
Registered model 'decision_tree_housing_price_predictor' already exists. Creating a new version of this model...
2025-08/10 18:31:16 INFO miflow.store.model_registry.abstract_store: Waiting up to 300 seconds for model version to finish creation. Model name: decision_tree_housing_price_predictor, version 2
Created version '2' of model 'decision_tree_housing_price_predictor'.
2025-08-10 18:31:16,513 - __main__ - INFO - train_models.py:198 - Decision Tree - Val R²: 0.6849, Val RMSE: 0.65
View run secretive-pug-807 at: http://localhost:5000/#/experiments/492744333528392905/runs/b19acb3b40c44917bfdbe5400390d796
View experiment at: http://localhost:5000/#/experiments/492744333528392905
View run model_comparison at: http://localhost:5000/#/experiments/492744333528392905/runs/4cd8d110da4509a3ab50b3b62f0a7b
View experiment at: http://localhost:5000/#/experiments/492744333528392905
2025-08-10 18:31:16,548 - __main__ - INFO - train_models.py:381 -
=====
2025-08-10 18:31:16,548 - __main__ - INFO - train_models.py:382 - TRAINING RESULTS SUMMARY
2025-08-10 18:31:16,548 - __main__ - INFO - train_models.py:383 - =====
2025-08-10 18:31:16,548 - __main__ - INFO - train_models.py:310 - decision_tree: R² = 0.6849
2025-08-10 18:31:16,548 - __main__ - INFO - train_models.py:312 - Training completed!
/Users/I060587/Documents/GitHub/MTechAssignments/.venv/lib/python3.9/site-packages/urllib3/_init__.py:35: NotOpenSSLWarning: urllib3 v2 only supports OpenSSL 1.1.1+, currently the 'ssl' module is compiled with 'Libressl 2.8.3'. See: <https://github.com/urllib3/urllib3/issues/3020>
warnings.warn
2025-08-10 18:31:18,710 - src.models.model_registry - INFO - model_registry.py:159 -
Successfully retrieved 2 most recent versions for 'linear_regression_housing_price_predictor'.
2025-08-10 18:31:18,710 - __main__ - INFO - best_model_selector.py:51 -
Best model found:
2025-08-10 18:31:18,710 - __main__ - INFO - best_model_selector.py:52 - Model Name: linear_regression_housing_price_predictor, Version: 2
2025-08-10 18:31:18,710 - __main__ - INFO - best_model_selector.py:53 - R2: 0.5826636004227627, RMSE: 0.7470256149575868
2025-08-10 18:31:18,710 - __main__ - INFO - best_model_selector.py:51 -
Best model found:
2025-08-10 18:31:18,710 - __main__ - INFO - best_model_selector.py:52 - Model Name: linear_regression_housing_price_predictor, Version: 1
2025-08-10 18:31:18,710 - __main__ - INFO - best_model_selector.py:53 - R2: 0.5848267718408837, RMSE: 0.7354415694664408
2025-08-10 18:31:18,710 - __main__ - INFO - best_model_selector.py:88 - Best model to be registered in registry: {best_model_data}
2025-08-10 18:31:18,754 - src.models.model_registry - INFO - model_registry.py:72 - Successfully retrieved details for model 'linear_regression_housing_price_predictor' version 1.
2025-08-10 18:31:18,754 - src.models.model_registry - INFO - model_registry.py:115 - model_uri : models:/linear_regression_housing_price_predictor/1
Downloading artifacts: 100%|██████████| 5/5 [00:00<00:00, 7428.81it/s]
2025-08-10 18:31:18,824 - __main__ - INFO - best_model_selector.py:98 - loading model: linear_regression_housing_price_predictor from registry: Pipeline(steps=[('scaler', StandardScaler()), ('regressor', Ridge())])
2025-08-10 18:31:18,825 - __main__ - INFO - best_model_selector.py:100 - Registering best model as target: housing_price_predictor
2025-08-10 18:31:18,965 - src.models.model_registry - INFO - model_registry.py:173 - Logging and registering model under new name 'housing_price_predictor'...
2025/08/10 18:31:18 WARNING miflow.models.model: 'artifact_path' is deprecated. Please use 'name' instead.
2025/08/10 18:31:21 WARNING miflow.models.model: Model logged without a signature and input example. Please set 'input_example' parameter when logging the model to auto infer the model signature.
Registered model 'housing_price_predictor' already exists. Creating a new version of this model...
2025/08/10 18:31:21 INFO miflow.store.model_registry.abstract_store: Waiting up to 300 seconds for model version to finish creation. Model name: housing_price_predictor, version 2
Created version '2' of model 'housing_price_predictor'.
2025-08-10 18:31:21,466 - src.models.model_registry - INFO - model_registry.py:183 - Registered Model Name: housing_price_predictor, Experiment Run Info (model-run-id): 5b9a254ca9f44209bd1ff4df78e8ac71
2025-08-10 18:31:21,466 - src.models.model_registry - INFO - model_registry.py:184 - Registered Model version: 2, model_uri Info: models:/m=70b98ea94980478a9fb447365bbc1d3
2025-08-10 18:31:21,455 - src.models.model_registry - INFO - model_registry.py:185 - Registered Model Param: {}
View run Register_housing_price_predictor at: http://localhost:5000/#/experiments/252478190093217419/runs/5b9a254ca9f44209bd1ff4df78e8ac71
View experiment at: http://localhost:5000/#/experiments/252478190093217419
2025-08-10 18:31:21,650 - __main__ - INFO - best_model_selector.py:102 - Registered best model with target: housing_price_predictor
[Watcher] ✅ Retraining completed successfully

Containerized API Service on best performing registered model for predictions (Part-3)

Ask Gordon BETA

Containers / mtechassignments-api-1

mtechassignments-api-1

f9d1b364a9ed ⏺ mtechassignments-api:latest
8000:8000 ↗

STATUS
Running (6 seconds ago)

Logs Inspect Bind mounts Exec Files Stats

```
2025-08-10 04:18:53,818 - src.models.model_registry - INFO - model_registry.py:88 - Found latest_version name = housing_price_predictor : latest version: 1 for model 'housing_price_predictor'.
2025-08-10 04:18:53,818 - src.api.app - INFO - app.py:55 - latest model version available in registry: <ModelVersion: aliases=[], creation_timestamp=1754799525450, current_stage='None', deployment_job_state=<ModelVersionDeploymentJobState: current_task_name='', job_id='', job_state='DEPLOYMENT_JOB_CONNECTION_STATE_UNSPECIFIED', run_id='', run_state='DEPLOYMENT_JOB_RUN_STATE_UNSPECIFIED', description='', last_updated_timestamp=1754799525450, metrics=None, model_id=None, name='housing_price_predictor', params=None, run_id='b453187df96a4a85b389872524c768ce', run_link='', source='models:/m-5c0d217393104ff9b6d9750a00fd1912', status='READY', status_message=None, tags={'best_model_type': 'linear_regression_housing_price_predictor', 'dataset': 'california-housing', 'optimization_framework': 'optuna', 'registration_date': '2025-08-10T09:48:45.480810', 'run_id': 'b453187df96a4a85b389872524c768ce'}, user_id='', version='1'
2025-08-10 04:18:53,846 - src.models.model_registry - INFO - model_registry.py:72 - Successfully retrieved details for model 'housing_price_predictor' version 1.
2025-08-10 04:18:53,847 - src.models.model_registry - INFO - model_registry.py:115 - model_uri : models:/housing_price_predictor/1

Downloading artifacts:  0% | 0/5 [00:00<?, ?it/s]
Downloading artifacts: 20% [██████████] 1/5 [00:00<00:00, 35848.75it/s]
Downloading artifacts: 40% [██████████] 2/5 [00:00<00:00, 22671.91it/s]
Downloading artifacts: 60% [██████████] 3/5 [00:00<00:00, 20593.96it/s]
Downloading artifacts: 80% [██████████] 4/5 [00:00<00:00, 7692.44it/s]
Downloading artifacts: 100% [██████████] 5/5 [00:00<00:00, 2678.01it/s]
Downloading artifacts: 100% [██████████] 5/5 [00:00<00:00, 2600.31it/s]
2025-08-10 04:18:53,883 - src.api.app - INFO - app.py:58 - loading model: housing_price_predictor from registry: Pipeline(steps=[('scaler', StandardScaler()), ('regressor', Ridge())])
2025-08-10 04:18:53,883 - src.api.app - INFO - app.py:66 - Model housing_price_predictor loaded successfully
2025-08-10 04:18:53,884 - src.api.app - INFO - app.py:67 - loaded model info : {'name': 'housing_price_predictor', 'version': '1', 'uri': 'model_uri'}
INFO: Application startup complete.
INFO: Uvicorn running on http://0.0.0.0:8000 (Press CTRL+C to quit)
INFO: 127.0.0.1:52602 - "GET /health HTTP/1.1" 200 OK
INFO: 172.18.0.5:38414 - "GET /metrics HTTP/1.1" 307 Temporary Redirect
INFO: 172.18.0.5:38414 - "GET /metrics/ HTTP/1.1" 200 OK
```

Engine running RAM 7.36 GB CPU 0.08% Disk: 20.83 GB used (limit 452.13 GB)

Terminal Update

Swagger APIs Published (Part-3)

The screenshot shows a web browser window with the address bar set to 'localhost'. The main content area displays the Swagger UI for a 'Housing Price Prediction API Server'.

API Overview:

Housing Price Prediction API Server 1.0 OAS 3.1

[/openapi.json](#)

API Server for housing price prediction with monitoring

default

Operations:

- GET / Root**
- GET /health Health Check**
- POST /predict Predict** (highlighted with a green border)

Schemas

CaliforniaHousing > Expand all object

HTTPValidationError > Expand all object

HealthResponse > Expand all object

PredictionRequest > Expand all object

PredictionResponse > Expand all object

Invoke Prediction (Part-3)

The screenshot shows a REST API endpoint for invoking a prediction. The interface includes:

- Parameters:** No parameters.
- Request body required:** application/json
- Request body:**

```
{
  "features": {
    "median_income": 1,
    "housing_median_age": 1,
    "avg_rooms_per_household": 1,
    "avg_num_bedrooms_per_house": 1,
    "Population": 1,
    "avg_household_members": 32,
    "Latitude": 0,
    "Longitude": -124
  }
}
```
- Responses:**
 - Curl:**

```
curl -X 'POST' \
'http://localhost:8000/predict' \
-H 'accept: application/json' \
-H 'Content-Type: application/json' \
-d {
  "Features": {
    "median_income": 1,
    "housing_median_age": 1,
    "avg_rooms_per_household": 1,
    "avg_num_bedrooms_per_house": 1,
    "Population": 1,
    "avg_household_members": 32,
    "Latitude": 0,
    "Longitude": -124
  }
},
```
 - Request URL:** http://localhost:8000/predict
 - Server response:**

Code	Details
200	Response body <pre>{ "prediction": 17.70857552110135, "current_timestamp": 175479584.0715935, "processing_time": 0.0044193267822265625 }</pre> Download
	Response headers <pre>access-control-allow-credentials: true access-control-allow-origin: http://localhost:8000 content-length: 112 content-type: application/json date: Sun, 10 Aug 2025 04:19:43 GMT server: uvicorn vary: Origin x-process-time: 0.006834506988525391</pre>

File Based Logging: Model Logging (Part-5)

The screenshot shows a terminal window with several tabs open, each displaying log files. The tabs include `metrics.log`, `model.log`, `california_housing.csv`, `models.log`, `app.log`, `api.log`, `decision_tree.py`, and `grafana_dashboard`. The `model.log` tab is currently active and displays the following log entries:

```
1 2025-08-07 16:27:41,574 - __main__ - INFO - best_model_selector.py:51 - 
2 Best model found:
3 2025-08-07 16:27:41,574 - __main__ - INFO - best_model_selector.py:52 - Model Name: linear_regression_housing_price_predictor, Version: 1
4 2025-08-07 16:27:41,574 - __main__ - INFO - best_model_selector.py:53 - R2: 0.5848267710408837, RMSE: 0.7354415694664408
5 2025-08-07 16:27:41,574 - __main__ - INFO - best_model_selector.py:88 - Best model to be registered in registry: {best_model_data}
6 2025-08-07 16:27:41,652 - __main__ - INFO - best_model_selector.py:98 - loading model: linear_regression_housing_price_predictor from registry
7 2025-08-07 16:27:41,652 - __main__ - INFO - best_model_selector.py:100 - Registering best model as target: housing_price_predictor
8 2025-08-07 16:27:44,361 - __main__ - INFO - best_model_selector.py:102 - Registered best model with target: housing_price_predictor
9 2025-08-07 18:54:20,715 - __main__ - INFO - best_model_selector.py:51 - 
10 Best model found:
11 2025-08-07 18:54:20,715 - __main__ - INFO - best_model_selector.py:52 - Model Name: linear_regression_housing_price_predictor, Version: 1
12 2025-08-07 18:54:20,715 - __main__ - INFO - best_model_selector.py:53 - R2: 0.5848267710408837, RMSE: 0.7354415694664408
13 2025-08-07 18:54:20,715 - __main__ - INFO - best_model_selector.py:88 - Best model to be registered in registry: {best_model_data}
14 2025-08-07 18:54:20,800 - __main__ - INFO - best_model_selector.py:98 - loading model: linear_regression_housing_price_predictor from registry
15 2025-08-07 18:54:20,800 - __main__ - INFO - best_model_selector.py:100 - Registering best model as target: housing_price_predictor
16 2025-08-07 18:54:23,309 - __main__ - INFO - best_model_selector.py:102 - Registered best model with target: housing_price_predictor
17 2025-08-07 20:50:47,817 - __main__ - INFO - best_model_selector.py:51 - 
18 Best model found:
19 2025-08-07 20:50:47,817 - __main__ - INFO - best_model_selector.py:52 - Model Name: linear_regression_housing_price_predictor, Version: 1
20 2025-08-07 20:50:47,817 - __main__ - INFO - best_model_selector.py:53 - R2: 0.5848267710408837, RMSE: 0.7354415694664408
21 2025-08-07 20:50:47,817 - __main__ - INFO - best_model_selector.py:88 - Best model to be registered in registry: {best_model_data}
22 2025-08-07 20:50:47,890 - __main__ - INFO - best_model_selector.py:98 - loading model: linear_regression_housing_price_predictor from registry
23 2025-08-07 20:50:47,890 - __main__ - INFO - best_model_selector.py:100 - Registering best model as target: housing_price_predictor
24 2025-08-07 20:50:50,427 - __main__ - INFO - best_model_selector.py:102 - Registered best model with target: housing_price_predictor
25 2025-08-08 13:28:15,199 - __main__ - INFO - best_model_selector.py:51 - 
26 Best model found:
27 2025-08-08 13:28:15,199 - __main__ - INFO - best_model_selector.py:52 - Model Name: linear_regression_housing_price_predictor, Version: 1
28 2025-08-08 13:28:15,199 - __main__ - INFO - best_model_selector.py:53 - R2: 0.5848267710408837, RMSE: 0.7354415694664408
29 2025-08-08 13:28:15,199 - __main__ - INFO - best_model_selector.py:88 - Best model to be registered in registry: {best_model_data}
30 2025-08-08 13:28:15,272 - __main__ - INFO - best_model_selector.py:98 - loading model: linear_regression_housing_price_predictor from registry
31 2025-08-08 13:28:15,272 - __main__ - INFO - best_model_selector.py:100 - Registering best model as target: housing_price_predictor
32 2025-08-08 13:28:17,811 - __main__ - INFO - best_model_selector.py:102 - Registered best model with target: housing_price_predictor
33 2025-08-08 16:32:58,156 - __main__ - INFO - best_model_selector.py:51 - 
34 Best model found:
35 2025-08-08 16:32:58,156 - __main__ - INFO - best_model_selector.py:52 - Model Name: linear_regression_housing_price_predictor, Version: 1
36 2025-08-08 16:32:58,156 - __main__ - INFO - best_model_selector.py:53 - R2: 0.5848267710408837, RMSE: 0.7354415694664408
37 2025-08-08 16:32:58,156 - __main__ - INFO - best_model_selector.py:88 - Best model to be registered in registry: {best_model_data}
38 2025-08-08 16:32:58,249 - __main__ - INFO - best_model_selector.py:98 - loading model: linear_regression_housing_price_predictor from registry
```

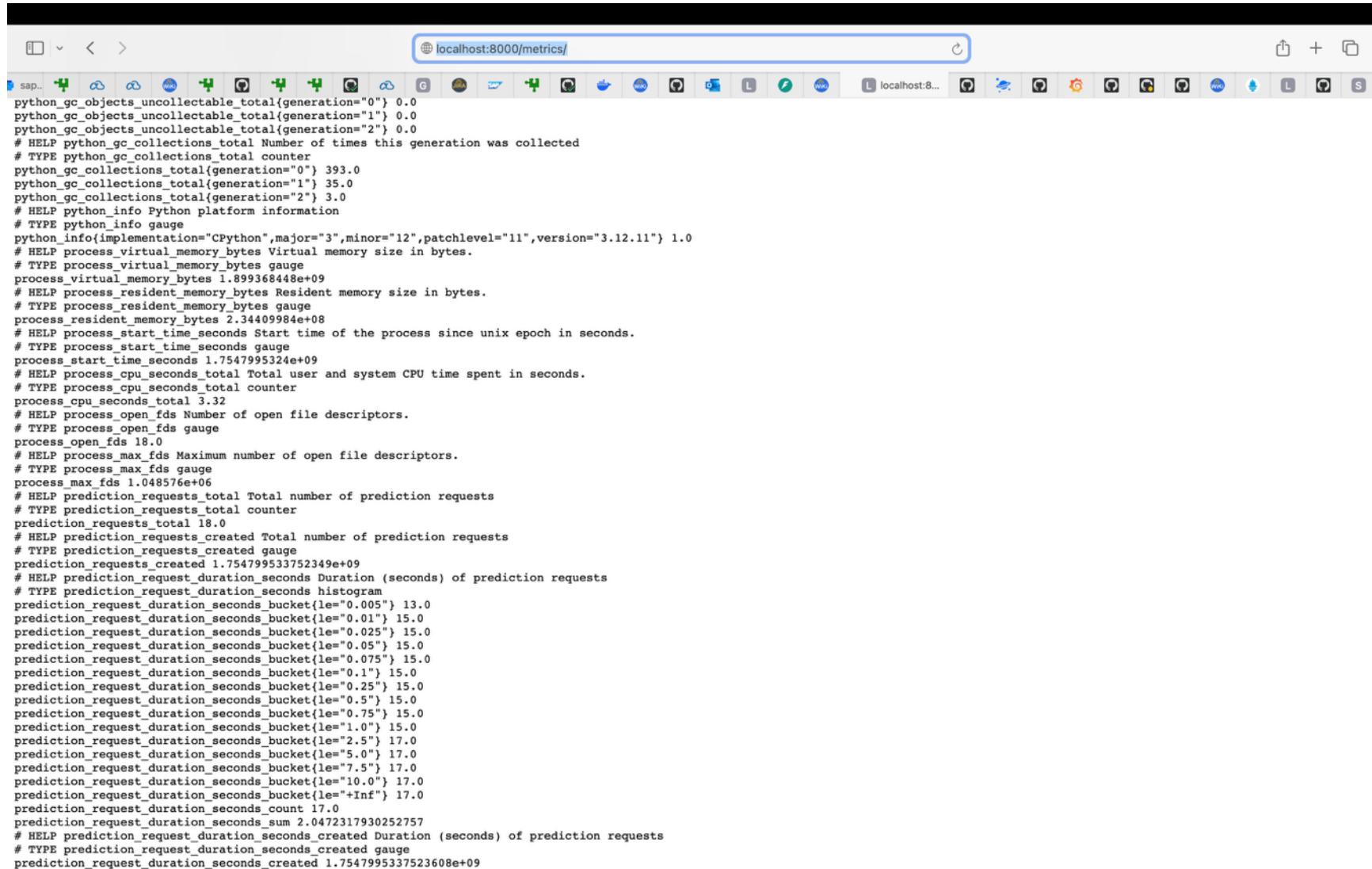
The terminal also shows the file structure of the project, including `MTechAssignments`, `logs`, and various configuration and source code files.

File Based Logging: Training Logs (Part-5)

The screenshot shows a terminal window with the title bar "MTechAssignments" and tab "main". The terminal displays log files from a project named "MTechAssignments". The log files include "metrics.log", "model.log", "california_housing.csv", "models.log", "app.log", "api.log", "decision_tree.py", and "grafana_dashboard". The "models.log" tab is active, showing log entries from a Python script named "train_models.py". The log entries detail the training process of a decision tree model, including the start of training, the creation of the decision tree, and the final results with an R-squared value of 0.6831 and an RMSE of 0.64. The log also includes training results summaries and completed training messages. The terminal interface includes a sidebar with project files like ".dvc", ".github", "workflows", ".idea", "venv", "configs", "data", "Docker", "localstack-data", "logs", and "src". The bottom of the terminal shows the file path "MTechAssignments > logs > models.log", the line number "178", and the timestamp "2025-08-05 19:06:03,213".

```
INFO - train_models.py:232 - Starting Training decision_tree
INFO - train_models.py:170 - Decision Tree - Val R2: 0.6831, Val RMSE: 0.64
=====
2025-08-05 16:23:38,908 - __main__ - INFO - train_models.py:279 - TRAINING RESULTS SUMMARY
2025-08-05 16:23:38,908 - __main__ - INFO - train_models.py:280 - =====
2025-08-05 16:23:38,908 - __main__ - INFO - train_models.py:287 - decision_tree: R2 = 0.6831
2025-08-05 16:23:38,908 - __main__ - INFO - train_models.py:289 - Training completed!
2025-08-05 16:26:42,566 - __main__ - INFO - train_models.py:225 - Train: 14448, Val: 3096, Test: 3096
2025-08-05 16:26:42,712 - __main__ - INFO - train_models.py:232 - Starting Training linear_regression
2025-08-05 16:26:47,968 - __main__ - INFO - train_models.py:108 - Linear Regression - Val R2: 0.5848, Val RMSE: 0.74
2025-08-05 16:26:48,022 - __main__ - INFO - train_models.py:278 -
=====
2025-08-05 16:26:48,022 - __main__ - INFO - train_models.py:279 - TRAINING RESULTS SUMMARY
2025-08-05 16:26:48,022 - __main__ - INFO - train_models.py:280 - =====
2025-08-05 16:26:48,022 - __main__ - INFO - train_models.py:287 - linear_regression: R2 = 0.5848
2025-08-05 16:26:48,022 - __main__ - INFO - train_models.py:289 - Training completed!
2025-08-05 18:52:17,952 - __main__ - INFO - train_models.py:225 - Train: 14448, Val: 3096, Test: 3096
2025-08-05 18:52:18,206 - __main__ - INFO - train_models.py:232 - Starting Training linear_regression
2025-08-05 18:52:24,243 - __main__ - INFO - train_models.py:108 - Linear Regression - Val R2: 0.5848, Val RMSE: 0.74
2025-08-05 18:52:24,288 - __main__ - INFO - train_models.py:278 -
=====
2025-08-05 18:52:24,288 - __main__ - INFO - train_models.py:279 - TRAINING RESULTS SUMMARY
2025-08-05 18:52:24,288 - __main__ - INFO - train_models.py:280 - =====
2025-08-05 18:52:24,288 - __main__ - INFO - train_models.py:287 - linear_regression: R2 = 0.5848
2025-08-05 18:52:24,288 - __main__ - INFO - train_models.py:289 - Training completed!
2025-08-05 18:53:06,908 - __main__ - INFO - train_models.py:225 - Train: 14448, Val: 3096, Test: 3096
2025-08-05 18:53:07,044 - __main__ - INFO - train_models.py:232 - Starting Training decision_tree
2025-08-05 18:53:11,878 - __main__ - INFO - train_models.py:170 - Decision Tree - Val R2: 0.6831, Val RMSE: 0.64
2025-08-05 18:53:11,918 - __main__ - INFO - train_models.py:278 -
=====
2025-08-05 18:53:11,918 - __main__ - INFO - train_models.py:279 - TRAINING RESULTS SUMMARY
2025-08-05 18:53:11,918 - __main__ - INFO - train_models.py:280 - =====
2025-08-05 18:53:11,918 - __main__ - INFO - train_models.py:287 - decision_tree: R2 = 0.6831
2025-08-05 18:53:11,918 - __main__ - INFO - train_models.py:289 - Training completed!
2025-08-05 19:03:57,884 - __main__ - INFO - train_models.py:225 - Train: 14448, Val: 3096, Test: 3096
2025-08-05 19:06:03,064 - __main__ - INFO - train_models.py:225 - Train: 14448, Val: 3096, Test: 3096
2025-08-05 19:06:03,213 - __main__ - INFO - train_models.py:232 - Starting Training linear_regression
```

API server log metrics exposed through /metrics API (part-5)



```
python_gc_objects_uncollectable_total{generation="0"} 0.0
python_gc_objects_uncollectable_total{generation="1"} 0.0
python_gc_objects_uncollectable_total{generation="2"} 0.0
# HELP python_gc_collections_total Number of times this generation was collected
# TYPE python_gc_collections_total counter
python_gc_collections_total{generation="0"} 393.0
python_gc_collections_total{generation="1"} 35.0
python_gc_collections_total{generation="2"} 3.0
# HELP python_info Python platform information
# TYPE python_info gauge
python_info{implementation="CPython",major="3",minor="12",patchlevel="11",version="3.12.11"} 1.0
# HELP process_virtual_memory_bytes Virtual memory size in bytes.
# TYPE process_virtual_memory_bytes gauge
process_virtual_memory_bytes 1.899368448e+09
# HELP process_resident_memory_bytes Resident memory size in bytes.
# TYPE process_resident_memory_bytes gauge
process_resident_memory_bytes 2.34409984e+08
# HELP process_start_time_seconds Start time of the process since unix epoch in seconds.
# TYPE process_start_time_seconds gauge
process_start_time_seconds 1.7547995324e+09
# HELP process_cpu_seconds_total Total user and system CPU time spent in seconds.
# TYPE process_cpu_seconds_total counter
process_cpu_seconds_total 3.32
# HELP process_open_fds Number of open file descriptors.
# TYPE process_open_fds gauge
process_open_fds 18.0
# HELP process_max_fds Maximum number of open file descriptors.
# TYPE process_max_fds gauge
process_max_fds 1.048576e+06
# HELP prediction_requests_total Total number of prediction requests
# TYPE prediction_requests_total counter
prediction_requests_total 18.0
# HELP prediction_requests_created Total number of prediction requests
# TYPE prediction_requests_created gauge
prediction_requests_created 1.754799533752349e+09
# HELP prediction_request_duration_seconds Duration (seconds) of prediction requests
# TYPE prediction_request_duration_seconds histogram
prediction_request_duration_seconds_bucket{le="0.005"} 13.0
prediction_request_duration_seconds_bucket{le="0.01"} 15.0
prediction_request_duration_seconds_bucket{le="0.025"} 15.0
prediction_request_duration_seconds_bucket{le="0.05"} 15.0
prediction_request_duration_seconds_bucket{le="0.075"} 15.0
prediction_request_duration_seconds_bucket{le="0.1"} 15.0
prediction_request_duration_seconds_bucket{le="0.25"} 15.0
prediction_request_duration_seconds_bucket{le="0.5"} 15.0
prediction_request_duration_seconds_bucket{le="0.75"} 15.0
prediction_request_duration_seconds_bucket{le="1.0"} 15.0
prediction_request_duration_seconds_bucket{le="2.5"} 17.0
prediction_request_duration_seconds_bucket{le="5.0"} 17.0
prediction_request_duration_seconds_bucket{le="7.5"} 17.0
prediction_request_duration_seconds_bucket{le="10.0"} 17.0
prediction_request_duration_seconds_bucket{le="+Inf"} 17.0
prediction_request_duration_seconds_count 17.0
prediction_request_duration_seconds_sum 2.0472317930252757
# HELP prediction_request_duration_seconds_created Duration (seconds) of prediction requests
# TYPE prediction_request_duration_seconds_created gauge
prediction_request_duration_seconds_created 1.7547995337523608e+09
```

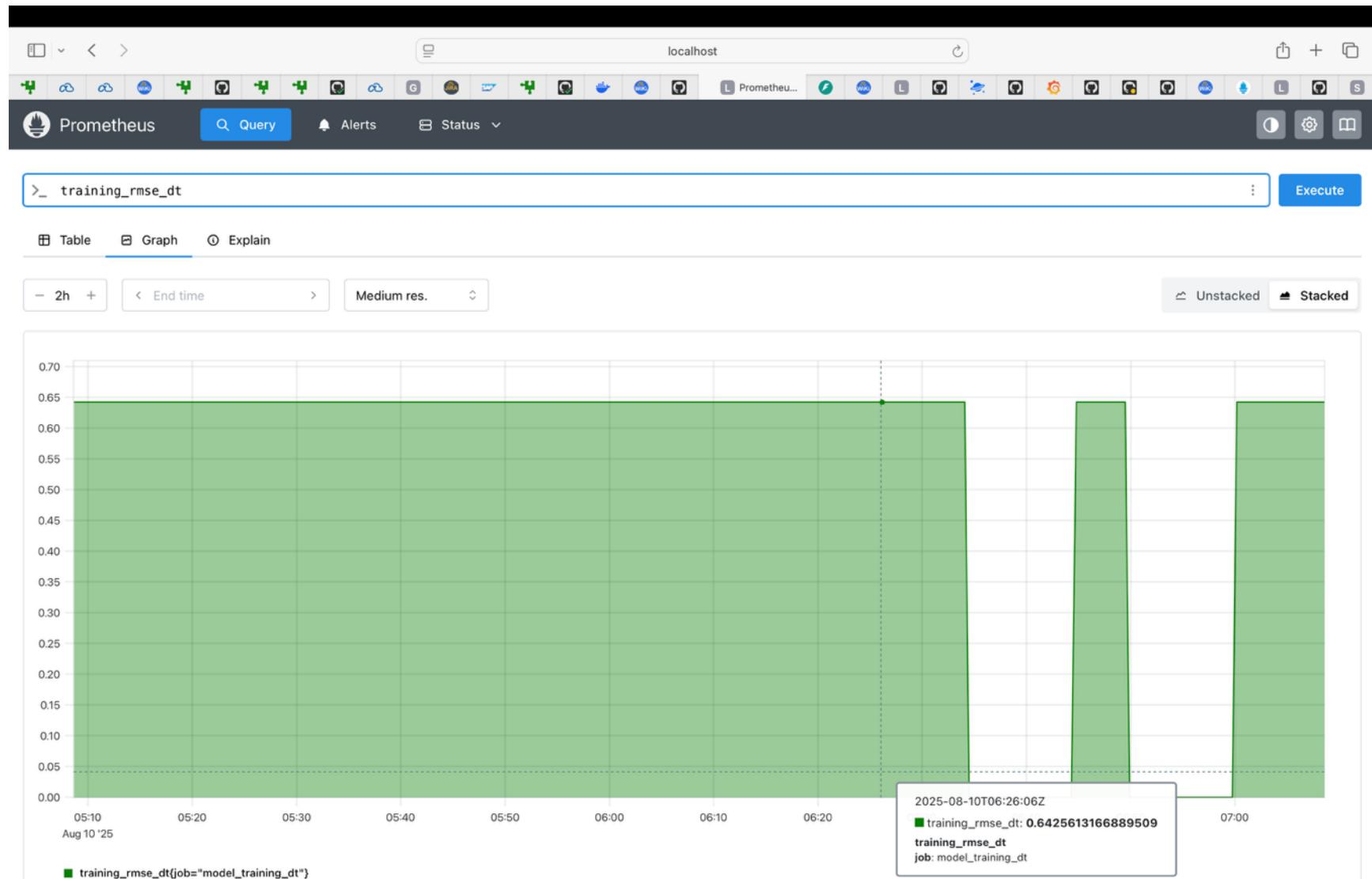
Model Training metrics exposed through Pushgateway (Bonus Assignment)

The screenshot shows the Pushgateway web interface running on localhost. It displays two groups of metrics:

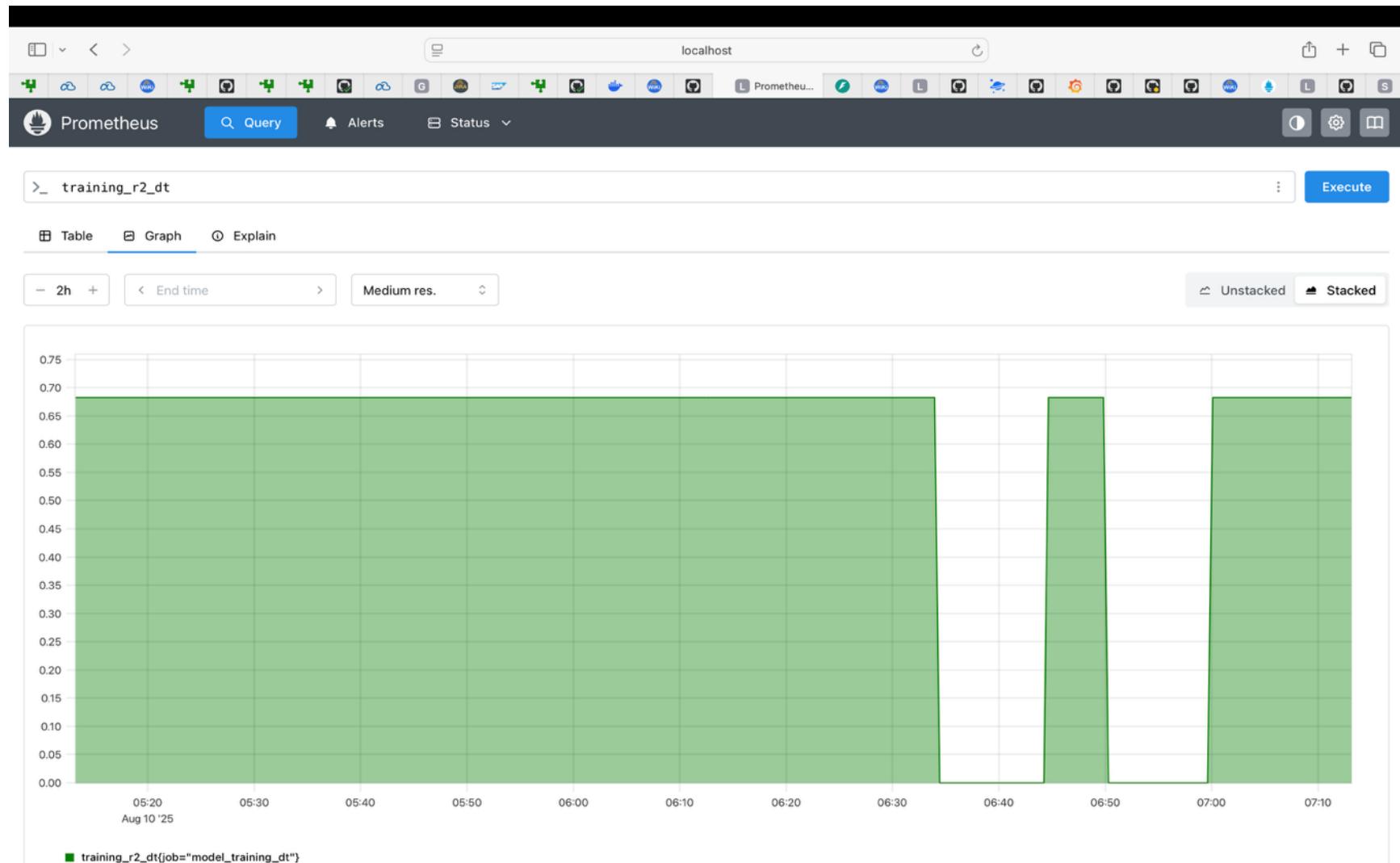
- Group 1 (job="model_training_dt")**: Contains the metric `push_failure_time_seconds`. Description: Last Unix time when changing this group in the Pushgateway failed. Type: GAUGE. Last pushed: 2025-08-09T11:39:25Z.
- Group 2 (job="model_training_lr")**: Contains the metric `push_time_seconds`. Description: Last Unix time when changing this group in the Pushgateway succeeded. Type: GAUGE. Last pushed: 2025-08-09T11:39:25Z.
- Group 3 (training_r2_lr)**: Contains the metric `R-squared score`. Description: R-squared score. Type: GAUGE. Last pushed: 2025-08-09T11:39:25Z.
- Group 4 (training_rmse_lr)**: Contains the metric `Root Mean Squared Error`. Description: Root Mean Squared Error. Type: GAUGE. Last pushed: 2025-08-09T11:39:25Z.

Each metric entry includes a "Delete Group" button on the right side of its respective card.

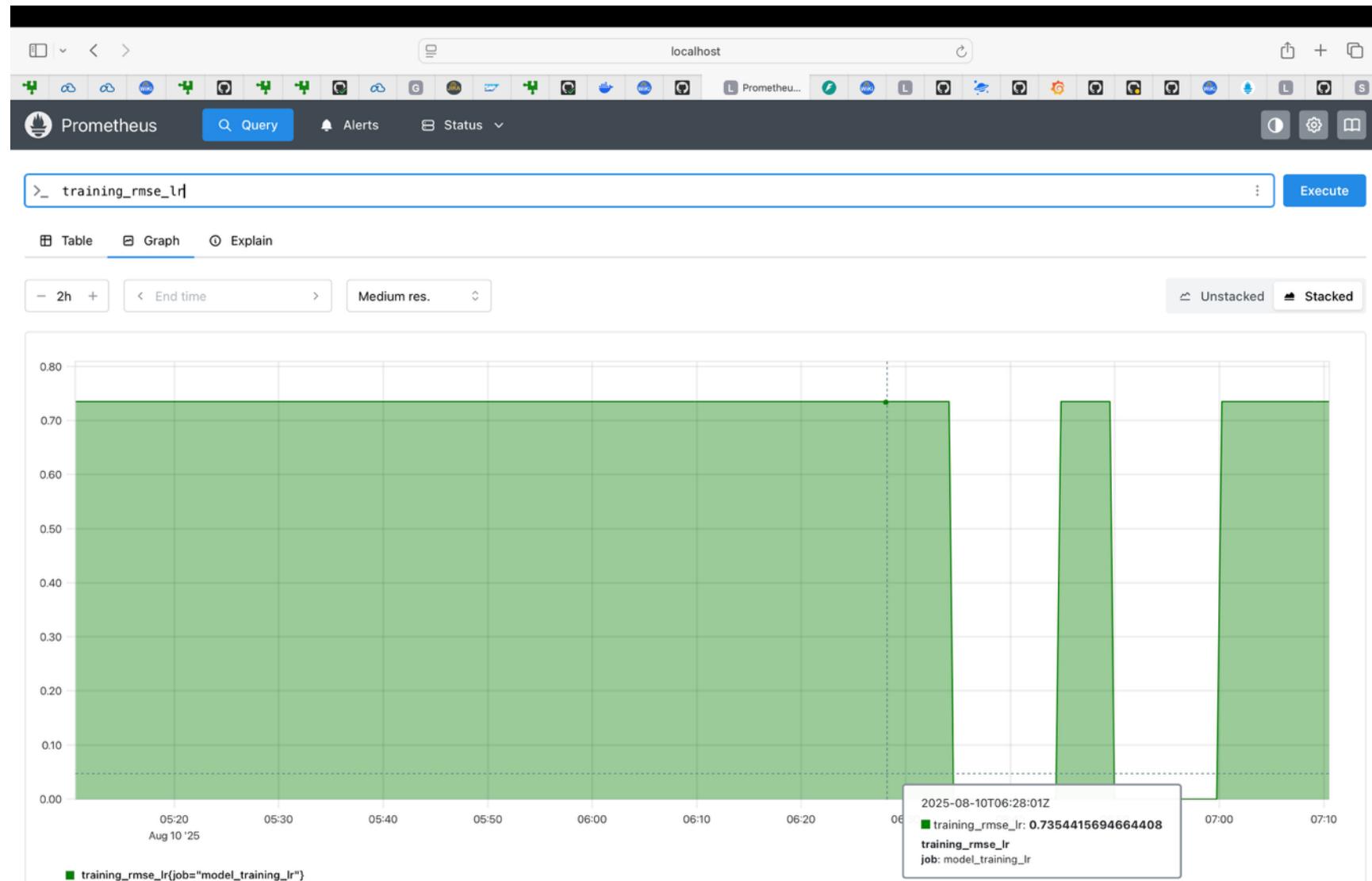
Prometheus Dashboard Query for Decision Tree – RMSE (Bonus Assignment)



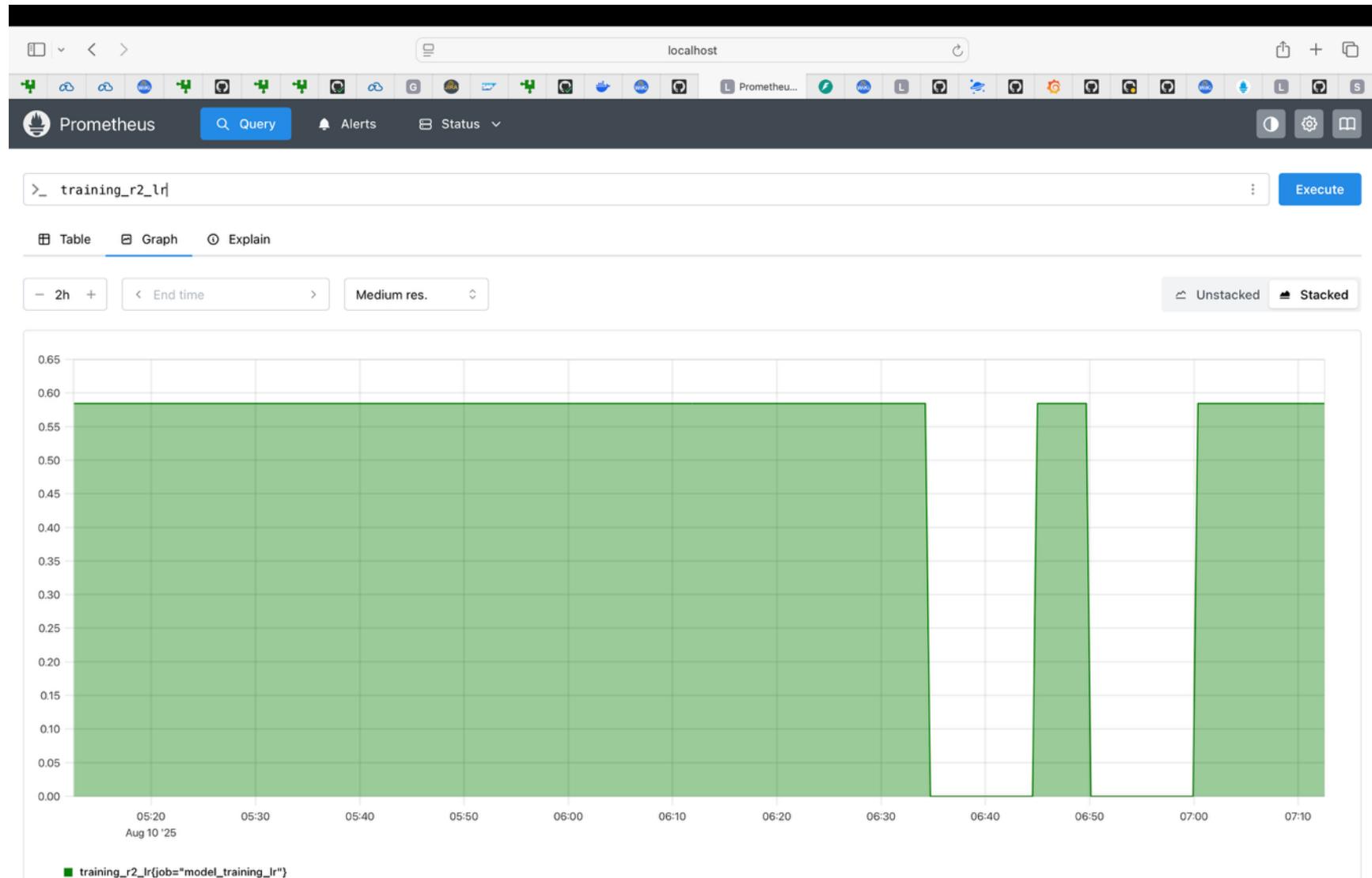
Prometheus Dashboard Query for Decision Tree – R2 (Bonus Assignment)



Prometheus Dashboard Query for Linear Regression – RMSE (Bonus Assignment)



Prometheus Dashboard Query for Linear Regression – R2 (Bonus Assignment)



Grafana Dashboard for Request and Training Metrics (Bonus Assignment)

