

Architecture Overview:

Dataset – California Housing Dataset (pulled from sklearn datasets) - used load_data.py

Processed data – used python to pull and preprocess the data - used preprocess.py and version the changes in dataset using DVC. Used S3 storage to save versions of datasets.

Train a model – Logistic regression, Decision tree. - used train_model.py to train a model using MLflows we experimented the model and derived best model out of two models.

API – used FlaskAPI to create api request that interact with model and predict the value based on model outcome.

Package building and containerisation – Used Docker and required library setup to create image and run the image in Docker hub, Docker file is used to configure the required environment and model setup.

GitHub Actions – used automate the CI/CD process to deploy the model into production after lint test on the code base.

Prometheus & Grafana – used to Build the dashboard based on predicted values.

Environment setup – used requirement.txt file in the project to install all necessary libraries

Demo video: [click here](#)

Docker Hub Image link: [click here](#)

GitHub repo: [click here](#)

Instructions of execution - you can find [README.md](#) file in GitHub repo detailed explanation about the execution are noted there and listed group members info too.

Sample input to the model:

```
curl -X POST http://localhost:8000/predict -H "Content-Type: application/json" \
-d '{"MedInc": 8.3252, "HouseAge": 41.0, "AveRooms": 6.9841, "AveBedrms": 1.0238, "Population": 322.0, "AveOccup": 2.5556, "Latitude": 37.88, "Longitude": -122.23, "MedHouseVal": 69.3}'
```

Sample output from the model:

```
{"prediction":4.214499999999999}
```