MLOps Assignment 5
Docker and Github Actions
Nitish Bhardwaj (B21AI056)

1. **Docker**
    a. **DockerFile creation**

MLops_assignment5 > Dockerfile > ...

```
1   FROM python:3.9-slim
2   WORKDIR /workspace
3   COPY . .
4   RUN apt-get update && apt-get install -y \
5   build-essential \
6   libssl-dev \
7   libffi-dev \
8   python3-dev
9   RUN pip install --no-cache-dir -r requirements.txt
10  EXPOSE 80
11  ENV NAME MLOpsLab
12  CMD ["python", "train.py"]
```

    b. **Docker Image build**

```
(mlflow_venv) nitish_linux@DESKTOP-CA3RQPS:~/Mlops_assignment/MLops_assignment5$ docker build -t mlops-lab5-image .
[+] Building 123.0s (10/10) FINISHED                                                                    docker:default
 => [internal] load build definition from Dockerfile                                                             0.0s
 => => transferring dockerfile: 292B                                                                             0.0s
 => [internal] load metadata for docker.io/library/python:3.9-slim                                              23.6s
 => [internal] load .dockerignore                                                                                0.0s
 => => transferring context: 2B                                                                                  0.0s
 => [1/5] FROM docker.io/library/python:3.9-slim@sha256:49f94609e5a997dc16086a66ac9664591854031d48e375945a9dbf   0.0s
 => [internal] load build context                                                                                0.0s
 => => transferring context: 7.31MB                                                                              0.0s
 => CACHED [2/5] WORKDIR /workspace                                                                              0.0s
 => [3/5] COPY . .                                                                                               0.1s
 => [4/5] RUN apt-get update && apt-get install -y build-essential libssl-dev libffi-dev python3-dev           28.9s
 => [5/5] RUN pip install --no-cache-dir -r requirements.txt                                                    65.8s
 => exporting to image                                                                                           4.5s
 => => exporting layers                                                                                          4.5s
 => => writing image sha256:7b510a384e95c3087e6de5b3e126d6a58920b98f96938373342abd308f2dd106                     0.0s
 => => naming to docker.io/library/mlops-lab5-image                                                              0.0s
```

c. Run Docker Container

```
nitish_linux@DESKTOP-CA3RQPS:~/Mlops_assignment/MLops_assignments$ docker run mlops-lab5-image
2024/10/06 00:10:48 WARNING mlflow.utils.git_utils: Failed to import Git (the Git executable is probably not on your PATH), so Git SHA is not available. Error: Failed to initialize: Bad git executable.
The git executable must be specified in one of the following ways:
    - be included in your $PATH
    - be set via $GIT_PYTHON_GIT_EXECUTABLE
    - explicitly set via git.refresh(<full-path-to-git-executable>)

All git commands will error until this is rectified.

This initial message can be silenced or aggravated in the future by setting the
$GIT_PYTHON_REFRESH environment variable. Use one of the following values:
    - quiet|q|silence|s|silent|none|n|0: for no message or exception
    - warn|w|warning|log|l|1: for a warning message (logging level CRITICAL, displayed by default)
    - error|e|exception|raise|r|2: for a raised exception

Example:
    export GIT_PYTHON_REFRESH=quiet

2024/10/06 00:10:50 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.
2024/10/06 00:10:52 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.
2024/10/06 00:10:54 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.
First 5 rows of dataset:
      crim    zn  indus  chas    nox  ...  tax  ptratio       b  lstat  medv
0  0.00632  18.0   2.31     0  0.538  ...  296     15.3  396.90   4.98  24.0
1  0.02731   0.0   7.07     0  0.469  ...  242     17.8  396.90   9.14  21.6
2  0.02729   0.0   7.07     0  0.469  ...  242     17.8  392.83   4.03  34.7
3  0.03237   0.0   2.18     0  0.458  ...  222     18.7  394.63   2.94  33.4
4  0.06905   0.0   2.18     0  0.458  ...  222     18.7  396.90   5.33  36.2

[5 rows x 14 columns]
Summary statistics of dataset:
             crim          zn       indus  ...           b       lstat        medv
count  506.000000  506.000000  506.000000  ...  506.000000  506.000000  506.000000
mean     3.613524   11.363636   11.136779  ...  356.674032   12.653063   22.532806
std      8.601545   23.322453    6.860353  ...   91.294864    7.141062    9.197104
min      0.006320    0.000000    0.460000  ...    0.320000    1.730000    5.000000
25%      0.082045    0.000000    5.190000  ...  375.377500    6.950000   17.025000
50%      0.256510    0.000000    9.690000  ...  391.440000   11.360000   21.200000
75%      3.677083   12.500000   18.100000  ...  396.225000   16.955000   25.000000
max     88.976200  100.000000   27.740000  ...  396.900000   37.970000   50.000000

[8 rows x 14 columns]
Missing values check:
crim       0
zn         0
indus      0
chas       0
nox        0
rm         0
age        0
dis        0
rad        0
tax        0
ptratio    0
b          0
lstat      0
medv       0
dtype: int64
Linear Regression MSE: 24.291119474973385
Random Forest MSE: 7.901513892156864

Comparison of Models:
Linear Regression MSE: 24.291119474973385
Random Forest MSE: 7.901513892156864
Best model: Random Forest
```

2. Github Actions

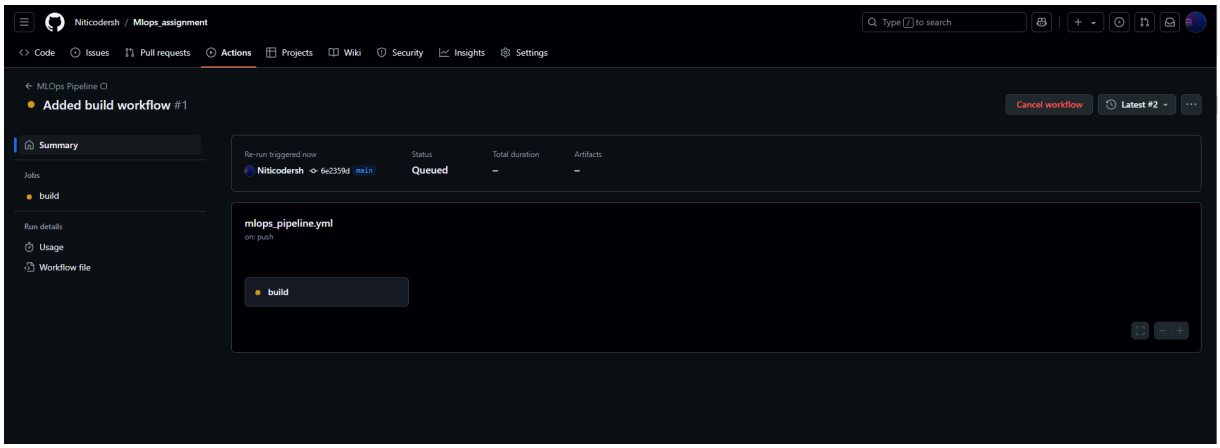
a. Creating .yml file

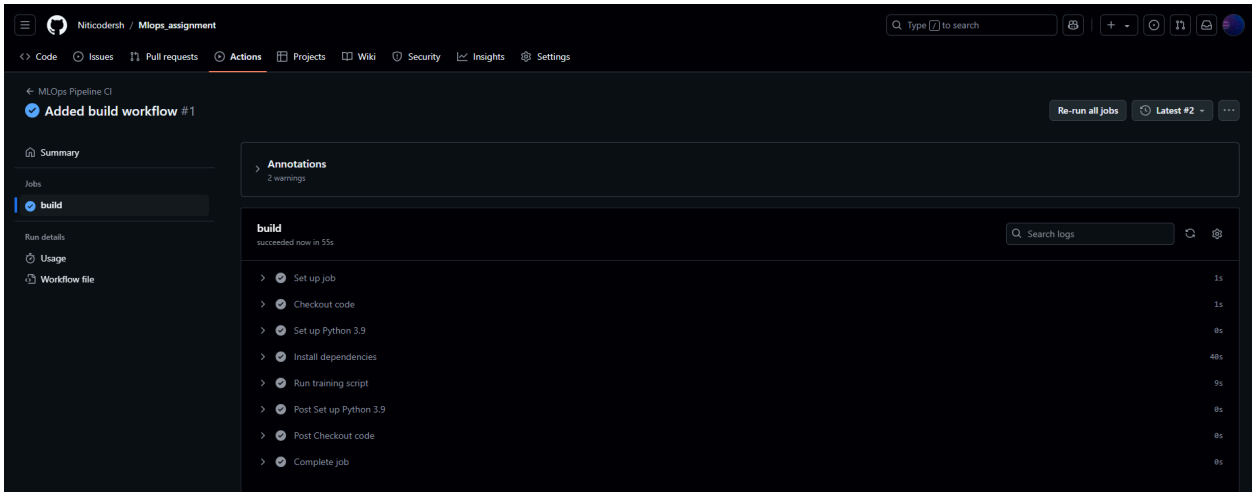
.github > workflows > mlops_pipeline.yml

```yaml
name: MLOps Pipeline CI
on:
    push:
        branches:
            - main
        paths:
            - 'MLops_assignment5/**'
    pull_request:
        branches:
            - main
        paths:
            - 'MLops_assignment5/**'
jobs:
    build:
        runs-on: ubuntu-latest
        steps:
        - name: Checkout code
          uses: actions/checkout@v2
        - name: Set up Python 3.9
          uses: actions/setup-python@v2
          with:
            python-version: '3.9'

        - name: Install dependencies
          run: |
            pip install --upgrade pip
            pip install -r MLops_assignment5/requirements.txt
        - name: Run training script
          run: |
            python MLops_assignment5/train.py
```

b. WorkFlow Results




Niticodersh / Mlops_assignment
Type / to search
Code
Issues
Pull requests
Actions
Projects
Wiki
Security
Insights
Settings
MLOps Pipeline CI
Added build workflow #1
Re-run all jobs
Latest #2
Summary
Jobs
build
Run details
Usage
Workflow file
Annotations
2 warnings
build
succeeded now in 55s
Search logs
Set up job 1s
Checkout code 1s
Set up Python 3.9 0s
Install dependencies 40s
Run training script 9s
Post Set up Python 3.9 0s
Post Checkout code 0s
Complete job 0s