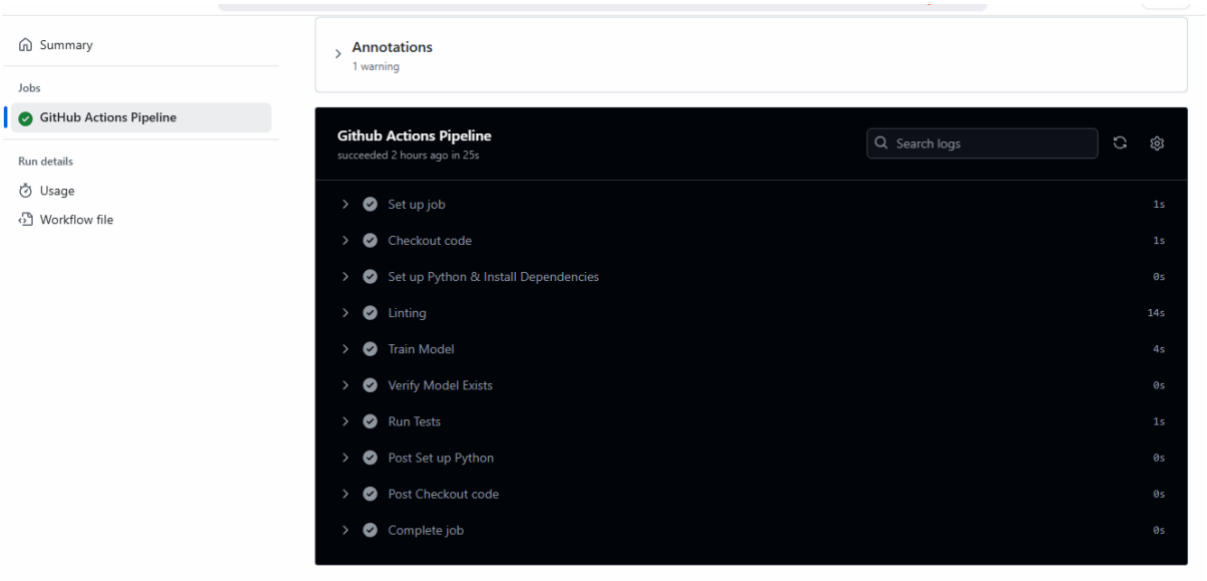
M1 – CI/CD Pipeline and Version Control

Screenshot:



Repo Link: https://github.com/deepakchirammeltw/MLOps_Assignment

Description:

Set Up Job:

Default Github process for preparing files and actions for running the workflow.

Checkout Code:

This step checks out the code from the repository, so the pipeline can access and work with the latest version of your code during the run.

Setup Python & Install Dependencies:

This step sets up the Python environment for your pipeline. It installs the specified Python version (in this case, Python 3.9) on the GitHub Actions runner, allowing subsequent steps to use Python for tasks like installing dependencies and running scripts.

This step also installs all the dependencies required to run the project. It upgrades pip and installs the dependencies listed in the requirements.txt file. Plus it installs flake8, the linter, to check for code quality and style issues.

Linting:

This step runs the flake8 linter to check for style and syntax issues in the Python code. It checks the entire repository and uses the default line length limit of 79 characters. If any linting errors or warnings are found, it will output them in the GitHub Actions log, and the pipeline will fail.

Train Model:

This step runs the training2.py script, which contains the logic for training the machine learning model. It loads the dataset, performs hyperparameter tuning using Optuna, trains the RandomForestClassifier, and saves the trained model as model2.joblib.

Verify Model File Exists:

This step checks if the model2.joblib file exists in the current working directory after training. It lists the file details to confirm that the model file was created and saved properly. If the file is not found, it will fail the pipeline and help identify any issues during model saving.

Run Tests:

This step runs a Python script (test_model.py) to test the trained model (model2.joblib). It will load the model, make predictions on a test set, and compute accuracy or other relevant metrics. If there are any issues with the model or predictions, this step can help identify them.

Post Setup Python, Post Checkout Code & Complete Job:

Default GitHub action for post job cleanup