Practical-6

Aim: Deployment of ML project in docker using Flask

Task 1: Ensure that the required libraries are installed:

- Install the Docker Command Line Interface Tool from: https://docs.docker.com/desktop/
- Install Flask library (https://flask.palletsprojects.com/en/2.3.x/installation/)
- Install gunicorn library (https://docs.gunicorn.org/en/latest/install.html)

pip install Flask

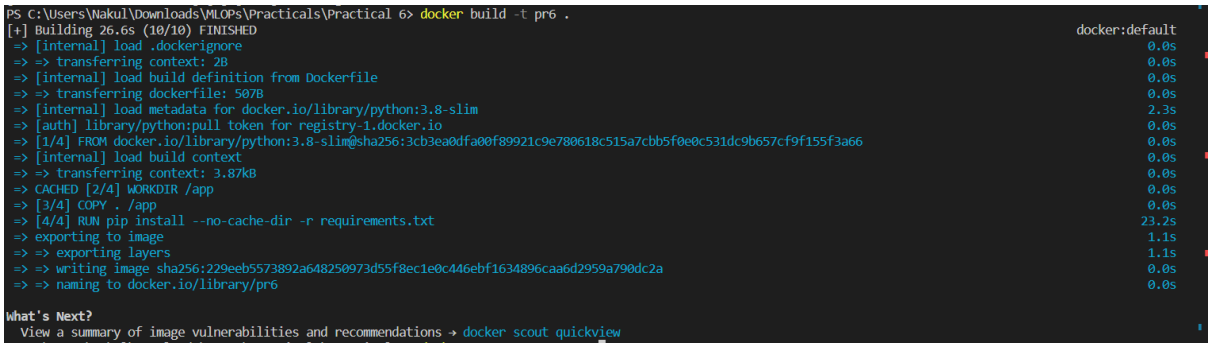
pip install gunicorn

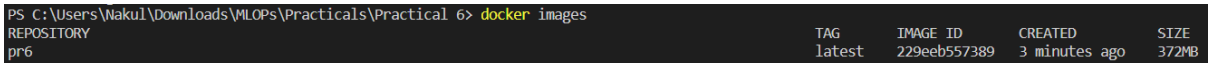
Task 2: Create the docker file using the steps described in theory material.

Practical 6 > Dockerfile > ...

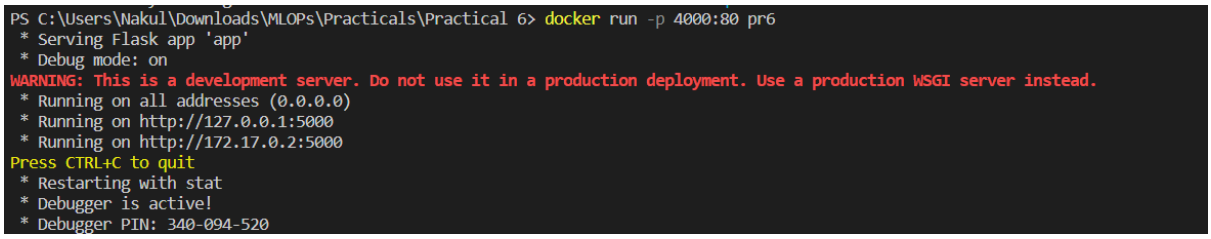
```
# Use an official Python runtime as a parent image
FROM python:3.8-slim

# Set the working directory to /app
WORKDIR /app

# Copy the current directory contents into the container at /app
COPY . /app

# Install any needed packages specified in requirements.txt
RUN pip install --no-cache-dir -r requirements.txt

# Make port 80 available to the world outside this container
EXPOSE 80

# Run app.py when the container launches
CMD ["python", "app.py"]

```

Task 3: Create the docker image using docker build command.





Task 4: Run the docker container to execute the docker image and host the machine learning model using gunicorn wsgi server.



Task 5: Compare the performance of the model in docker container and flask script deployment.

The performance of Docker compared to local hosting can vary depending on several factors, and it's not always straightforward to determine whether Docker is faster or slower.

Docker might be faster due to optimized resource allocation, efficient caching, and consistent environments, especially if the Docker container is well-tailored for the application.

Docker might be slower if the overhead introduced by containerization is significant, or if there are misconfigurations impacting resource utilization.

Conclusion:

The decision on whether Docker hosting is faster or slower depends on the specific characteristics of your application and how well it is configured for Docker. It’s recommended to perform detailed benchmarking and resource monitoring to draw accurate conclusions for your particular use case.