

Docker Image Instructions

Documents from Github

Download all the needed documents from my Github repo:

https://github.com/XiaohanSun01/Open_Science_Project

Run the R file

Download the `mushroom_classification.R` from the GitHub repo and run the file.

If there is no problem after the running, you should have the classification result saved to a file called `mushroom_classification_accuracy.txt`. This is also the file we want to get later again when we successfully run the Docker image.

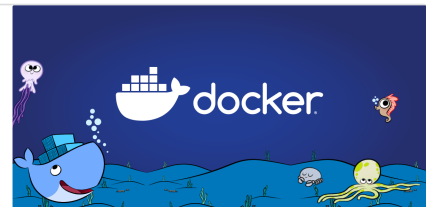
Download the Docker Desktop

1. Use the link below to download Docker Desktop and choose the one that fits your operating system.

Download Docker Desktop | Docker

Install Docker Desktop - the fastest way to containerize applications. The Docker Subscription Service Agreement has been updated. Our Docker Subscription Service Agreement includes a change to the terms for Docker Desktop. It remains free for

<https://www.docker.com/products/docker-desktop/>



2. Follow the instructions to finish the download and open the application.

Docker Image Creation

Docker Images vs. Docker Containers

Before we use Dockerfile to create a docker image, here are two definitions that I want to clarify: Docker images and Docker containers. So, images are the definition of the OS, while the containers are the actual running instances of the images. You will need to **install the image once**, but **multiple containers of the same images can be run simultaneously**.

Dockerfile

From my GitHub repo, here is one document called `Dockerfile`, which is the one to build a docker image. Below are the step-by-step explanations about this document, and you can create your own Dockerfile based

on the instructions.

1. Describe what image we are building our image from. Here we use `rocker`, a collection of Docker images for R. `4.2.0` is my R version, and you can change it here to your current R version.

```
FROM rocker/r-ver:4.2.0
```

2. Create a directory to receive the analysis. Here is the **directory for the container** and you can change it to your preferred working directory.

```
RUN mkdir /home/analysis
```

3. Use these commands below to install the packages needed from the `mushroom_classification.R`.

```
RUN R -e "install.packages('stringr')"  
RUN R -e "install.packages('purrr')"  
RUN R -e "install.packages('caret')"  
RUN R -e "install.packages('ggplot2')"  
RUN R -e "install.packages('randomForest')"
```

4. Now, we need to get the script for the analysis from the machine (host) to the container. For that, we use

```
COPY localfile pathinthecontainer
```

Note: `mushroom_classification.R` and `secondary_data_no_miss.csv` have to be in the same folder as the `Dockerfile` on your computer.

```
COPY mushroom_classification.R /home/analysis/mushroom_classification.R  
COPY secondary_data_no_miss.csv /home/analysis/secondary_data_no_miss.csv
```

5. `CMD` is the command to be run every time you will launch the docker. What we want is `mushroom_classification.R` to be sourced.

Also, we move the `mushroom_classification_accuracy.txt` from `/home/analysis` to `/home/results` for later export this file from the container to the host.

```
CMD cd /home/analysis \  
  && R -e "source('mushroom_classification.R')\" \  
  && mv /home/analysis/mushroom_classification_accuracy.txt /home/results/mushroom_classification_accuracy.txt
```

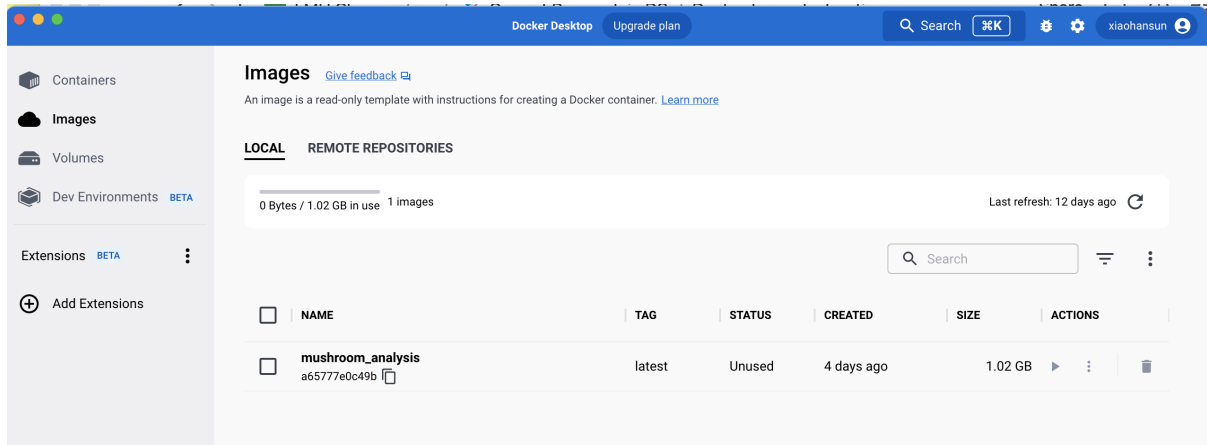
Build and Run

1. Open a terminal and direct to the directory where `Dockerfile` is located.
2. Run the command under the directory above:

```
docker build --build-arg WHEN=2022-01-05 -t mushroom_analysis .
```

Note: `WHEN` specifies the date when you build this image, `-t name` is the name of the image (here `mushroom_analysis`), and `.` means it will build the `Dockerfile` in the current working directory.

3. If the image is created successfully, it should now appear in the Docker Desktop:



4. Now, let's create a folder `docker_results` under a directory where you want to export the container content. I created the folder under the same folder in which the `Dockerfile` is saved.

```
mkdir your_prefered_directory_here
cd ~/your_prefered_directory_here
mkdir ~/your_prefered_directory_here/docker_results
```

5. Move to the directory of the created folder:

```
cd ~/your_prefered_directory_here/docker_results
```

6. Finally, we run the docker container using `docker run`.

And to export the container content, we use `-v` flag when running the container, with `/path/from/host:/path/in/container`.

Last is the image name (Here is `mushroom_classification`).

```
docker run -v ~/your_prefered_directory_here/docker_results:home/results mushroom_classification
```

7. DONE! Your analysis will be run 🎉. You should get the `mushroom_classification_accuracy.txt` again under the folder `docker_results`.