Learning Lab 40: Docker for Data Science

# Overview

|  |
| --- |
|  |
|  |
|  |
|  |
|  |
|  |
| Errors when running the code in someone else’s machine. |
| If you put it in a docker container, they can use the same libraries as you are using for development. |
| This works with Applications and AWS as well (not just markdown) |
| File are managed by git, the software environment (OS, R, libraries) are managed by Docker. |
|  |
| **Step 1**: You create a docker image (text file with instructions). Here starting from roker image and adding on to it. It takes a loing time to build the image, but loading it takes only a few minutes. |
| **Step 2**: Then you can share this on Docker Hub. |
| Can pull images from docker hub |
| Run the image. I get the IDE with the libraries. |
| Docker only manages images, not files. You can use GitHub for this. |
| We can do this in AWS as shown above. |

# Demo / Code

|  |  |
| --- | --- |
|  | |
| Performing Grid Search using H2O and saving metrics in ml\_flow | |
| Move best model to production folder. | |
| Now the model has been built, so now we are creating a report for someone (on local machine). This should run corectly as everything is setup correctly on local machine. | |
| Now we can create a Docker Image to share – Go to docker website and install docker. | |
| Then create the docker file. |  |
|  | |
| Then push to Docker Hub and pull  See instructions in Image:   1. Step 1: Spin up an empty R Studio Sesssion. Have a username and Password associated with it as shown in the instructions. 2. Step 2: Linked Volumes: Link a folder on the machine to the RStudio session. 3. Step 3: Load GitHub repo and files. 4. Step 4: Run in detach mode with –d instead of –rm. This wil allow others to run your code and analysis. | |
| Step 5: List and stop the container. | |