**Assignment - 1** **Predictive Modelling of Eating-Out Problem**

**Unit Name** : Data Science Technology and Systems PG (11523)  
**Name** : Ajul Thomas  
**Email** : u3253992@uni.canberra.edu.au  
**Date** : 29 September 2024

### My Tableau Dashboard:

**Link**:  
[Sydney Restaurant Analytics - Tableau Dashboard](https://public.tableau.com/views/SydneyRestaurants/SydneyRestaurantAnalytics?:language=en-US&:sid=&:redirect=auth&:display_count=n&:origin=viz_share_link)

### Model Evaluation:

#### **Regression Models:**

|  |  |
| --- | --- |
| Model | Test MSE |
| model\_regression\_1 | 0.1252 |
| model\_regression\_2 | 0.1254 |

#### **Classification Models:**

|  |  |
| --- | --- |
| Model | Test Accuracy |
| model\_classification\_3 (Logistic Regression) | 0.8295 |
| model\_classification\_4 (SVM) | 0.8295 |
| model\_classification\_5 (RFC) | 0.8559 |

**Docker Image Link**:  
[Docker Image - ajulthomas/dsts-assignment-1](https://hub.docker.com/r/ajulthomas/dsts-assignment-1)

**GitLab Repo Link**:  
[Data Science Technology Systems - DSTS Assignment 1](https://gitlab.com/data-science-technology-systems/dsts-assignment-1)

# Git Workflow

This document explains the commands used to set up git version control for this project.

## Setting up Local Repo

**Initialize the local Git repository**

# initialises the local repository  
git init  
  
# rename the default branch name to main  
git branch -m main

**Add .gitignore file**

Add a .gitignore file mentioning the file names or extensions that doesn’t needs to be tracked.

**Stage the files**

git add .

The . adds all files to the staging area of the local repository.

**Commiting the changes to local repository**

git commit -m "Initial commit"

## Setting up the remote or cloud repository

**Create online repository**

I have created an empty Gitlab respository named dsts-assignment-1 (new repo for this assignment purpose [link here](https://gitlab.com/data-science-technology-systems/dsts-assignment-1.git)). This code is also uploaded to my github account under the repository dsts-lab (already existing repo).

**Connect the local repo to the remote repository**

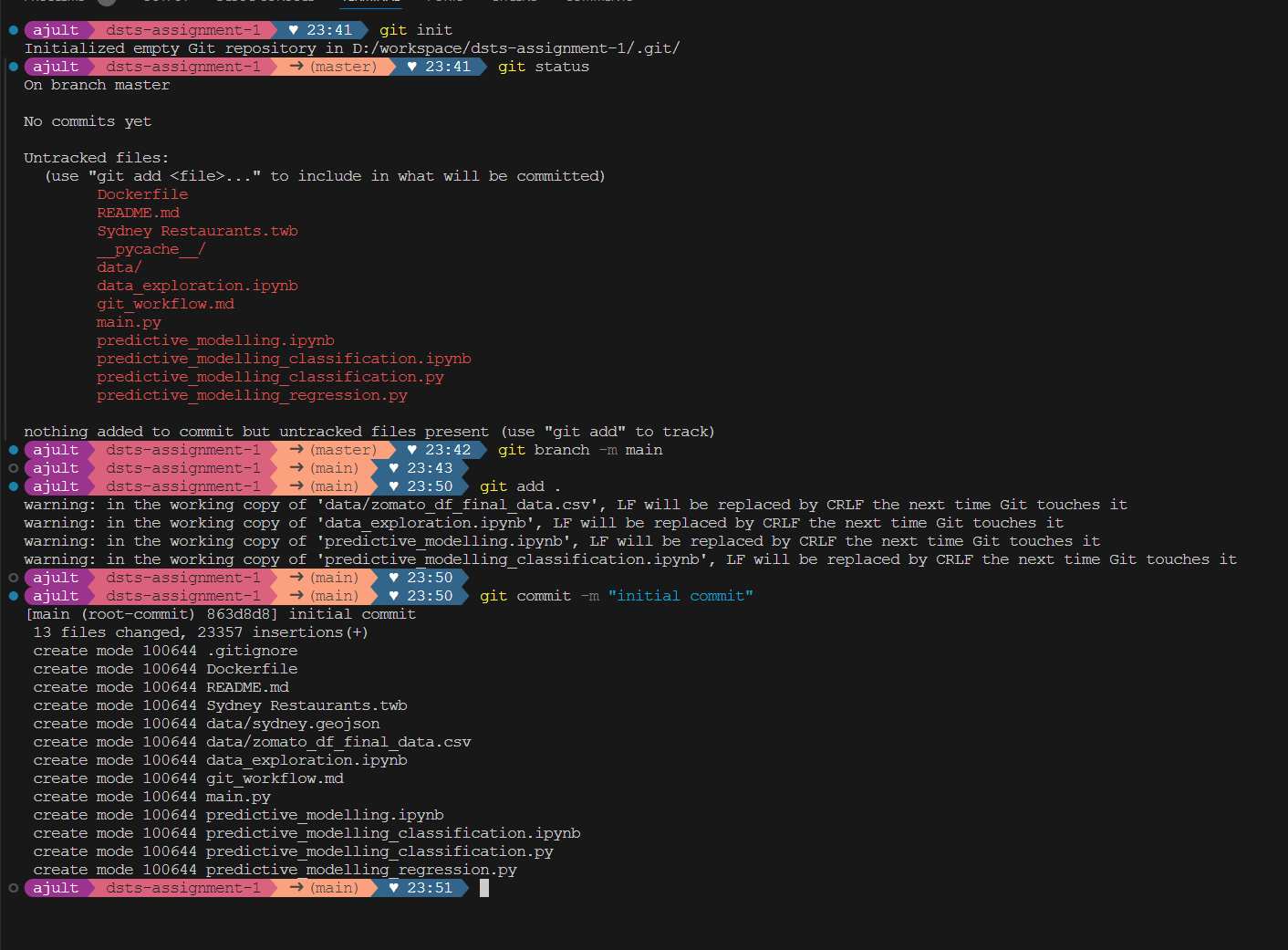
git remote add origin git@gitlab.com:data-science-technology-systems/dsts-assignment-1.git

This command tells the local Git version control system that for the dsts-assignment-1 local repo, the remote repository url is as provided and tells it to store the url in variable origin.

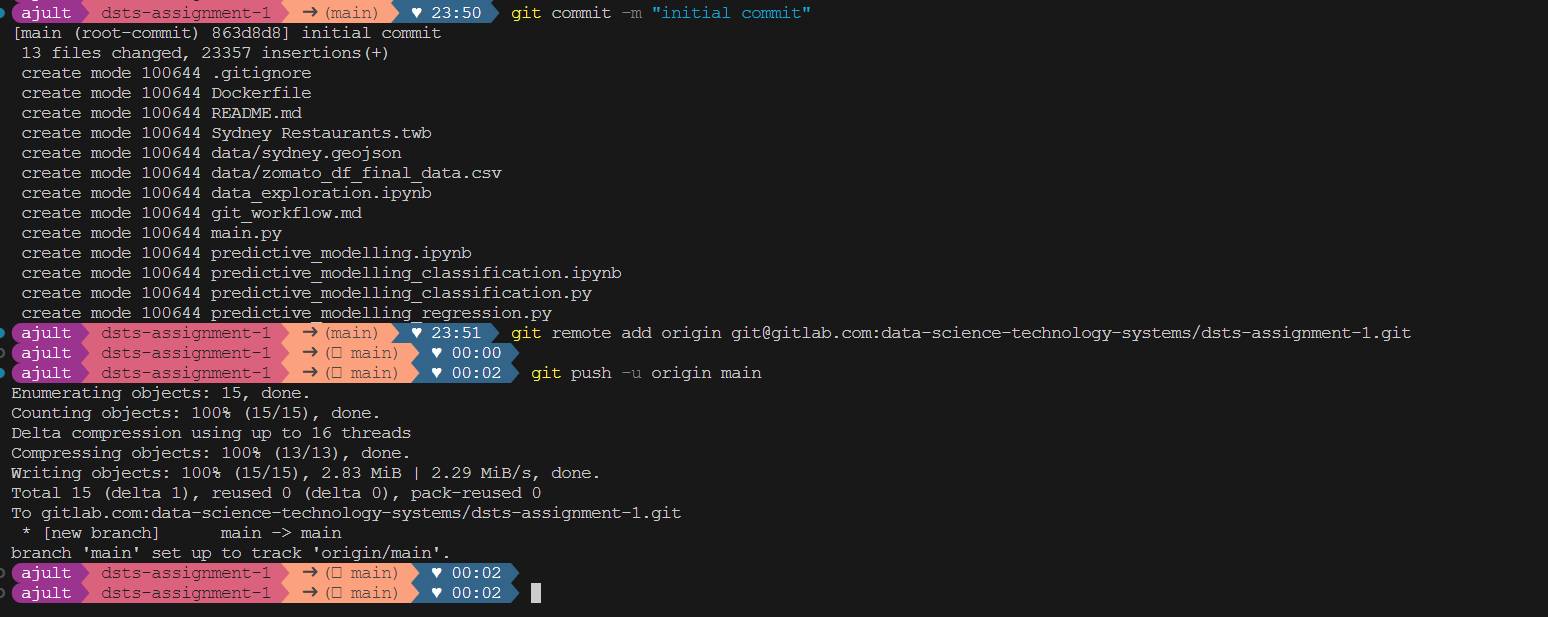
**Upload the local commits to remote repository**

git push -u origin main

This command pushes command from local main branch to remote(Gitlab repo) main branch.



Committing changes to local repo



Pushing commits to remote repo

## Building a Docker Image

**Build the docker image**

docker build -t dsts-assignment:1 .

The above command asks docker engine to build a docker image with tag dsts-assignment:1 as per the commands provided in the current directory donoted by ‘.’

**Run the docker image**

docker run dsts-assignment:1

The above command runs the docker image named dsts-assignment:1.

## Uploading the docker image to Dockerhub

**Create a docker account**

I have created a docker account under the username ajulthomas

**Tagging my docker image in accordance with the dockerhub requirements**

docker tag dsts-assignment:1 ajulthomas/dsts-assignment-1:v1.0

**Pushing the image to Docker hub**

docker push ajulthomas/dsts-assignment-1:v1.0

**Docker hub Repo Link**

[get the image from dockerhub here](https://hub.docker.com/r/ajulthomas/dsts-assignment-1)

### Git Version Control:

Kindly use the link below to access the markdown file on GitLab, where I have explained in detail, with screenshots, various Git commands used for setting up a local repository and pushing it to the GitLab repository.

[git\_workflows.md](https://gitlab.com/data-science-technology-systems/dsts-assignment-1/-/blob/ae1a5907fd57432974d862cba84f1080f24f3594/git_workflow.md)

### Dockerize the Models:

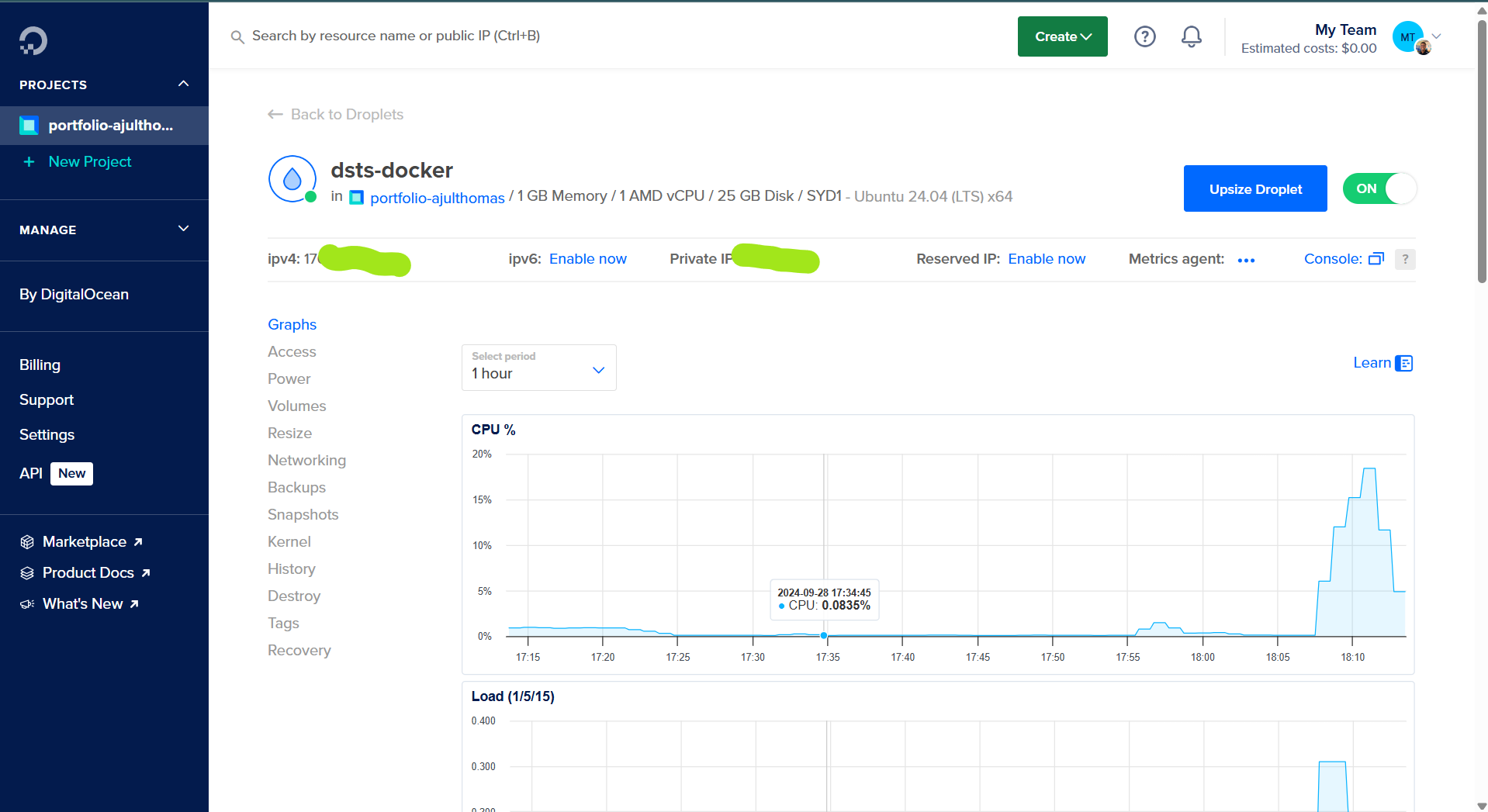
Kindly use the link below to go to the markdown file in GitLab, where I have explained in detail the various commands that I have used to dockerize the model and upload it to Docker Hub.

[docker\_commands.md](https://gitlab.com/data-science-technology-systems/dsts-assignment-1/-/blob/main/docker_commands.md?ref_type=heads)

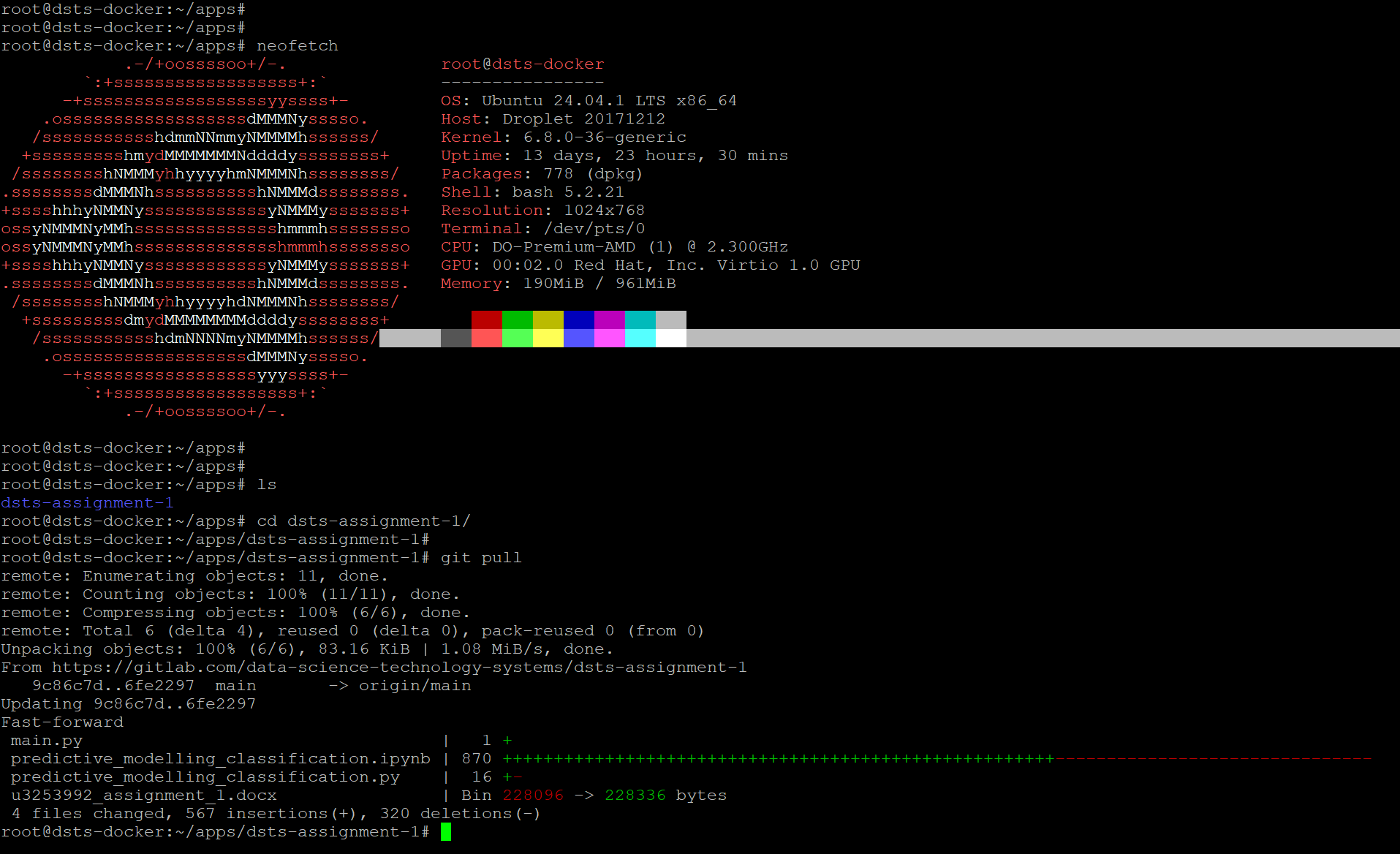
# Edge Computing

I have cloned the Git repo and run the docker containers on Digital Ocean droplets, which are virtual compute solutions offered by digital ocean. I find them very useful to test out Data Science deployment pipeline.

I used putty to connect to the remote virtual machine via ssh.



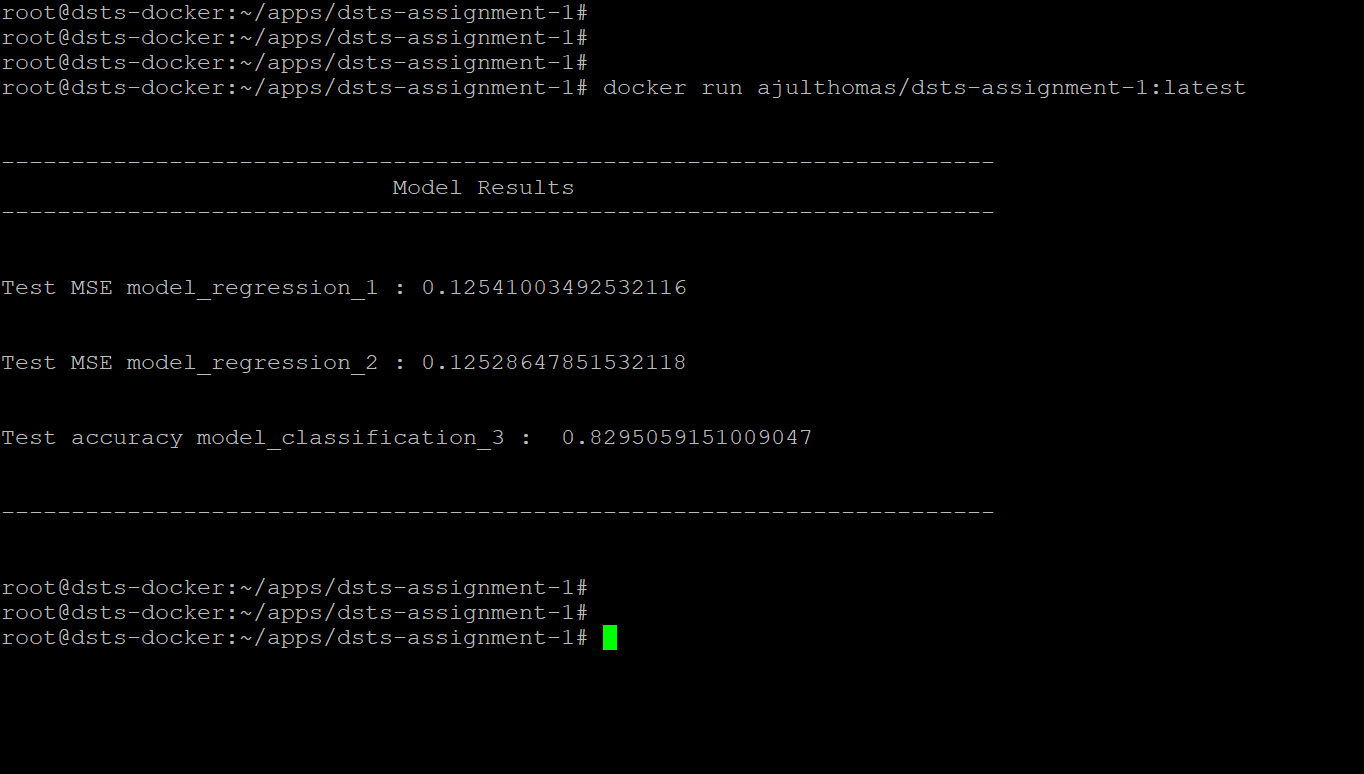
digital ocean droplet info



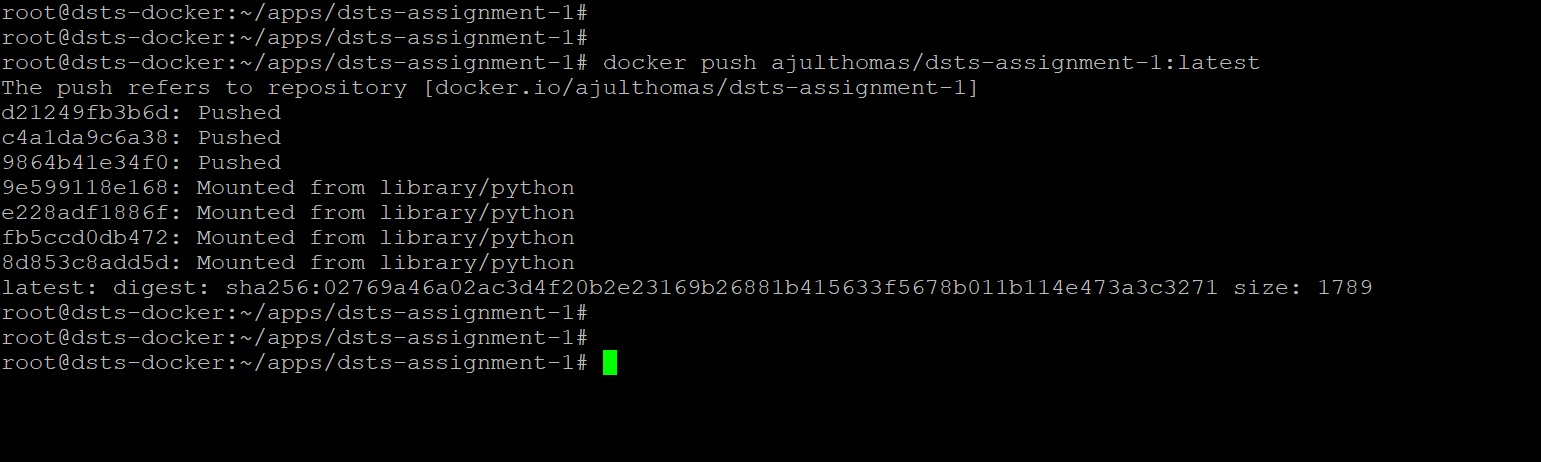
cloning repo to digital ocean



building docker image



docker image output



Pushing the docker image to DockerHub

A screenshot of a computer

Description automatically generated

Running the docker image on my personal laptop.