

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](#)

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](#)

GitLab Repo Link:

[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 need to be tracked.

Stage the files

```
git add .
```

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

Committing 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 repository named dsts-assignment-1 (new repo for this assignment purpose [link here](#)). 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.

```
• ajult dsts-assignment-1 → (master) ▼ 23:41 git init
Initialized empty Git repository in D:/workspace/dsts-assignment-1/.git/
• ajult dsts-assignment-1 → (master) ▼ 23:41 git status
On branch master

No commits yet

Untracked files:
  (use "git add <file>..." to include in what will be committed)
  Dockerfile
  README.md
  Sydney Restaurants.twb
  __pycache__/
  data/
  data_exploration.ipynb
  git_workflow.md
  main.py
  predictive_modelling.ipynb
  predictive_modelling_classification.ipynb
  predictive_modelling_classification.py
  predictive_modelling_regression.py

nothing added to commit but untracked files present (use "git add" to track)
• ajult dsts-assignment-1 → (master) ▼ 23:42 git branch -m main
• ajult dsts-assignment-1 → (main) ▼ 23:43
• ajult dsts-assignment-1 → (main) ▼ 23:50 git add .
warning: in the working copy of 'data/zomato_df_final_data.csv', LF will be replaced by CRLF the next time Git touches it
warning: in the working copy of 'data_exploration.ipynb', LF will be replaced by CRLF the next time Git touches it
warning: in the working copy of 'predictive_modelling.ipynb', LF will be replaced by CRLF the next time Git touches it
warning: in the working copy of 'predictive_modelling_classification.ipynb', LF will be replaced by CRLF the next time Git touches it
• ajult dsts-assignment-1 → (main) ▼ 23:50
• ajult dsts-assignment-1 → (main) ▼ 23:50 git commit -m "initial commit"
[main (root-commit) 863d8d8] initial commit
13 files changed, 23357 insertions(+)
create mode 100644 .gitignore
create mode 100644 Dockerfile
create mode 100644 README.md
create mode 100644 Sydney Restaurants.twb
create mode 100644 data/sydney.geojson
create mode 100644 data/zomato_df_final_data.csv
create mode 100644 data_exploration.ipynb
create mode 100644 git_workflow.md
create mode 100644 main.py
create mode 100644 predictive_modelling.ipynb
create mode 100644 predictive_modelling_classification.ipynb
create mode 100644 predictive_modelling_classification.py
create mode 100644 predictive_modelling_regression.py
• ajult dsts-assignment-1 → (main) ▼ 23:51
```

Committing changes to local repo

```
• ajult dsts-assignment-1 → (main) ▼ 23:50 git commit -m "initial commit"
[main (root-commit) 863d8d8] initial commit
13 files changed, 23357 insertions(+)
create mode 100644 .gitignore
create mode 100644 Dockerfile
create mode 100644 README.md
create mode 100644 Sydney Restaurants.twb
create mode 100644 data/sydney.geojson
create mode 100644 data/zomato_df_final_data.csv
create mode 100644 data_exploration.ipynb
create mode 100644 git_workflow.md
create mode 100644 main.py
create mode 100644 predictive_modelling.ipynb
create mode 100644 predictive_modelling_classification.ipynb
create mode 100644 predictive_modelling_classification.py
create mode 100644 predictive_modelling_regression.py
• ajult dsts-assignment-1 → (main) ▼ 23:51 git remote add origin git@gitlab.com:data-science-technology-systems/dsts-assignment-1.git
• ajult dsts-assignment-1 → (main) ▼ 00:00
• ajult dsts-assignment-1 → (main) ▼ 00:02 git push -u origin main
Enumerating objects: 15, done.
Counting objects: 100% (15/15), done.
Delta compression using up to 16 threads
Compressing objects: 100% (13/13), done.
Writing objects: 100% (15/15), 2.83 MiB | 2.29 MiB/s, done.
Total 15 (delta 1), reused 0 (delta 0), pack-reused 0
To gitlab.com:data-science-technology-systems/dsts-assignment-1.git
 * [new branch] main -> main
branch 'main' set up to track 'origin/main'.
• ajult dsts-assignment-1 → (main) ▼ 00:02
• ajult dsts-assignment-1 → (main) ▼ 00:02
```

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 denoted 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](#)

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](#)

Dockerize the Models:

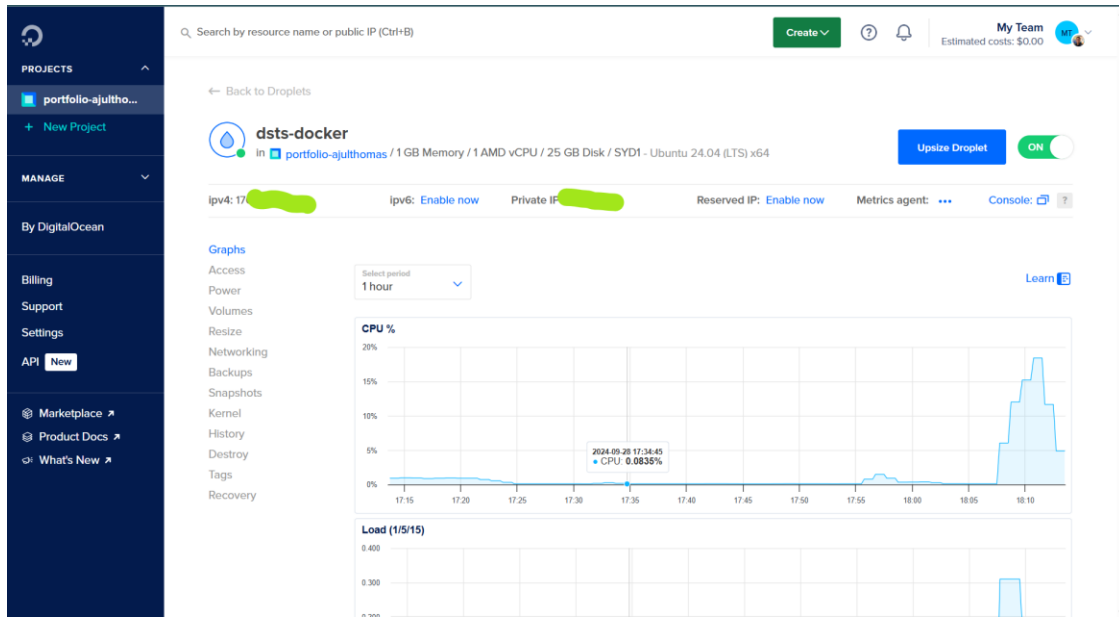
Kindly use the link below to go to the markdown file in GitLab, where I have explained in detail the various commands that I have used to dockerize the model and upload it to Docker Hub.

[docker_commands.md](#)

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

```
root@dsts-docker:~/apps#
root@dsts-docker:~/apps#
root@dsts-docker:~/apps# neofetch

      .-/+oossssoo+/-.
      `:+ssssssssssss++`
      -+ssssssssssssssyyssss+-
      .ossssssssssssssdMMMMyssso.
      /ssssssssshdmmNmmymMMhsssss/
      +ssssssshmyssssssssssssssss+
      /ssssssshMMMyhhyryhmmNMhsssss/
      .ssssssshNMhssssssssshNMhsssss.
      +ssssshhhyNMMyssssssssyNMMysssss+
      ossyNMMyNMhssssssssssshmmhssssso
      ossyNMMyNMhssssssssssshmmhssssso
      +ssssshhhyNMMyssssssssyNMMysssss+
      .ssssssshNMhssssssssshNMhsssss.
      /ssssssshNMMyhhyryhmmNMhsssss/
      +ssssssshNMMyhhyryhmmNMhsssss/
      /ssssssshdmmNmmymMMhsssss/
      .ossssssssssssssdMMMMyssso.
      -+ssssssssssssssyyssss+-
      `:+ssssssssssss++`
      .-/+oossssoo+/-.

root@dsts-docker:~/apps#
root@dsts-docker:~/apps#
root@dsts-docker:~/apps# ls
dsts-assignment-1
root@dsts-docker:~/apps# cd dsts-assignment-1/
root@dsts-docker:~/apps/dsts-assignment-1#
root@dsts-docker:~/apps/dsts-assignment-1# git pull
remote: Enumerating objects: 11, done.
remote: Counting objects: 100% (11/11), done.
remote: Compressing objects: 100% (6/6), done.
remote: Total 6 (delta 4), reused 0 (delta 0), pack-reused 0 (from 0)
Unpacking objects: 100% (6/6), 83.16 KiB | 1.08 MiB/s, done.
From https://gitlab.com/data-science-technology-systems/dsts-assignment-1
  9c86c7d..6fe2297  main    -> origin/main
Updating 9c86c7d..6fe2297
Fast-forward
   main.py                  | 1 +
   predictive_modelling_classification.ipynb | 870 ++++++
   predictive_modelling_classification.py    | 16 +-
   u3253992_assignment_1.docx               | Bin 228096 -> 228336 bytes
  4 files changed, 567 insertions(+), 320 deletions(-)
root@dsts-docker:~/apps/dsts-assignment-1#
```

cloning repo to digital ocean

```

root@dsts-docker:~/apps/dsts-assignment-1#
root@dsts-docker:~/apps/dsts-assignment-1#
root@dsts-docker:~/apps/dsts-assignment-1#
root@dsts-docker:~/apps/dsts-assignment-1# docker build -t ajulthomas/dsts-assignment-1:latest .
[*] Building 34.1s (10/10) FINISHED
docker:default
-> [internal] load build definition from Dockerfile
-> [internal] load metadata for docker.io/library/python:3.9-slim
-> [internal] load metadata for registry-4.docker.io
-> [internal] load build context
-> [internal] transferring context: 57B
-> [1/4] FROM docker.io/library/python:3.9-slim@sha256:1b123790982ba55e4b02b10341e3d309dd13b4ee4527b080705b3336635179d0
-> RUN python3 docker.io/library/python:3.9-slim@sha256:1b123790982ba55e4b02b10341e3d309dd13b4ee4527b080705b3336635179d0
-> RUN sha256:1b123790982ba55e4b02b10341e3d309dd13b4ee4527b080705b3336635179d0 / 10.41MB / 10.41MB
-> RUN sha256:1b123790982ba55e4b02b10341e3d309dd13b4ee4527b080705b3336635179d0 / 1.75MB / 1.75MB
-> RUN sha256:1b123790982ba55e4b02b10341e3d309dd13b4ee4527b080705b3336635179d0 / 3.00MB / 3.00MB
-> RUN sha256:1b123790982ba55e4b02b10341e3d309dd13b4ee4527b080705b3336635179d0 / 28.13MB / 28.13MB
-> RUN sha256:1b123790982ba55e4b02b10341e3d309dd13b4ee4527b080705b3336635179d0 / 1.81MB / 1.81MB
-> RUN sha256:1b123790982ba55e4b02b10341e3d309dd13b4ee4527b080705b3336635179d0 / 14.14MB / 14.14MB
-> RUN sha256:1b123790982ba55e4b02b10341e3d309dd13b4ee4527b080705b3336635179d0 / 24MB / 24MB
-> EXTRACTING sha256:1b123790982ba55e4b02b10341e3d309dd13b4ee4527b080705b3336635179d0 / 1.81MB
-> EXTRACTING sha256:1b123790982ba55e4b02b10341e3d309dd13b4ee4527b080705b3336635179d0 / 3.00MB
-> EXTRACTING sha256:1b123790982ba55e4b02b10341e3d309dd13b4ee4527b080705b3336635179d0 / 28.13MB
-> EXTRACTING sha256:1b123790982ba55e4b02b10341e3d309dd13b4ee4527b080705b3336635179d0 / 1.81MB
-> [internal] load build context
-> [internal] transferring context: 14.13MB
-> [3/4] WORKDIR /app
-> [3/4] COPY . /app
-> [3/4] RUN pip install --no-cache-dir -r requirements.txt
-> EXTRACTING to image
-> EXTRACTING image
-> writing image sha256:1b123790982ba55e4b02b10341e3d309dd13b4ee4527b080705b3336635179d0 to docker.io/ajulthomas/dsts-assignment-1:latest
root@dsts-docker:~/apps/dsts-assignment-1#
root@dsts-docker:~/apps/dsts-assignment-1#

```

building docker image

```

root@dsts-docker:~/apps/dsts-assignment-1#
root@dsts-docker:~/apps/dsts-assignment-1#
root@dsts-docker:~/apps/dsts-assignment-1#
root@dsts-docker:~/apps/dsts-assignment-1# docker run ajulthomas/dsts-assignment-1:latest

```

```

-----
Model Results
-----

Test MSE model_regression_1 : 0.12541003492532116

Test MSE model_regression_2 : 0.12528647851532118

Test accuracy model_classification_3 : 0.8295059151009047

-----

root@dsts-docker:~/apps/dsts-assignment-1#
root@dsts-docker:~/apps/dsts-assignment-1#
root@dsts-docker:~/apps/dsts-assignment-1#

```

docker image output

```

root@dsts-docker:~/apps/dsts-assignment-1#
root@dsts-docker:~/apps/dsts-assignment-1#
root@dsts-docker:~/apps/dsts-assignment-1# docker push ajulthomas/dsts-assignment-1:latest
The push refers to repository [docker.io/ajulthomas/dsts-assignment-1]
d21249fb3b6d: Pushed
c4a1da9c6a38: Pushed
9864b41e34f0: Pushed
9e599118e168: Mounted from library/python
e228adf1886f: Mounted from library/python
fb5ccd0db472: Mounted from library/python
8d853c8add5d: Mounted from library/python
latest: digest: sha256:02769a46a02ac3d4f20b2e23169b26881b415633f5678b011b114e473a3c3271 size: 1789
root@dsts-docker:~/apps/dsts-assignment-1#
root@dsts-docker:~/apps/dsts-assignment-1#
root@dsts-docker:~/apps/dsts-assignment-1#

```

Pushing the docker image to DockerHub

```

ajult dsts-assignment-1 → (Main) 18:49 docker run ajulthomas/dsts-assignment-1:latest
Unable to find image 'ajulthomas/dsts-assignment-1:latest' locally
latest: Pulling from ajulthomas/dsts-assignment-1
302e3ee49805: Pull complete
4c0965d39195: Pull complete
fdeec85abba: Pull complete
62a08b8dd4f5: Pull complete
0658bc6dea35: Pull complete
20e148127380: Pull complete
b161ab917077: Pull complete
Digest: sha256:02769a46a02ac3d4f20b2e23169b26881b415633f5678b011b114e473a3c3271
Status: Downloaded newer image for ajulthomas/dsts-assignment-1:latest

-----
Model Results
-----

Test MSE model_regression_1 : 0.12541003492532107

Test MSE model_regression_2 : 0.12522450428030196

Test accuracy model_classification_3 : 0.8295059151009047

-----

```

Running the docker image on my personal laptop.