

# Assignment structure description

SIT, Data Science, W2D2  
Martinson Michael

Git repo link - [https://github.com/MartinsonMichael/SIT\\_DS\\_W2D2](https://github.com/MartinsonMichael/SIT_DS_W2D2)

I did this assignment mostly locally, soni order to show installation and setup of **git** and **dvc** I attached terminal logs (not full log, but with comments) - file `unix_console_log.txt`

Also, for the flask server and the test client I created python scripts (located in `srs/` folder, `flaskApp.py` and `client.py` accordingly) and jupyter-notebooks (located in `notebooks/`).

Note!

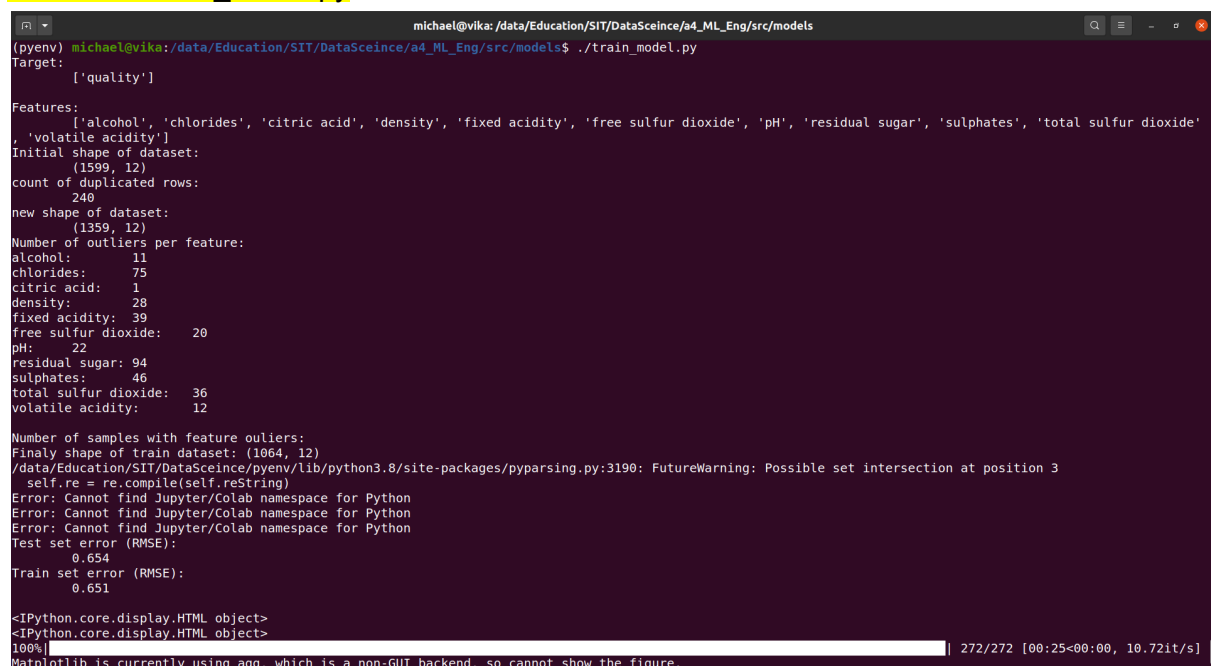
- For launching a server you need an ml model, so you pull it with dvc, or create it with `train_model.py` (located in `src/models/`).
- Scripts can not work properly with wrong file structure, so in order to launch, pull everything from git and dvc (but dvc requires my personal gdrive token...).

Content of solution zip folder:

- `0_MMD_SIT_W2D2_WineQuality.ipynb` - original jupyter-notebook with data exploration, model creating, etc
- `flaskApp.py` and `client.py` - python scripts with flask server and test client (should be placed in `src/`)
- `10_MMD_flask_server.ipynb` and `11_MMD_client.ipynb` - the same as scripts but in a format of jupyter-notebooks (should be placed in `notebooks/`)
- `train_model.py` - script to train model from scratch (should be placed in `src/models/`)
- `unix_console_log.txt` - just a description of unix commands
- `SIT_DS_Martinson_W2D2_ml_eng.pdf` - this document with descriptions

Example of work:

`srs/models/train_model.py`



```
michael@vika: /data/Education/SIT/DataScience/a4_ML_Eng/src/models
(pynv) michael@vika:/data/Education/SIT/DataScience/a4_ML_Eng/src/models$ ./train_model.py
Target:
['quality']

Features:
['alcohol', 'chlorides', 'citric acid', 'density', 'fixed acidity', 'free sulfur dioxide', 'pH', 'residual sugar', 'sulphates', 'total sulfur dioxide', 'volatile acidity']
Initial shape of dataset:
(1599, 12)
count of duplicated rows:
240
new shape of dataset:
(1359, 12)
Number of outliers per feature:
alcohol: 11
chlorides: 75
citric acid: 1
density: 28
fixed acidity: 39
free sulfur dioxide: 20
pH: 22
residual sugar: 94
sulphates: 46
total sulfur dioxide: 36
volatile acidity: 12

Number of samples with feature outliers:
Finally shape of train dataset: (1064, 12)
/data/Education/SIT/DataScience/pynv/lib/python3.8/site-packages/pyparsing.py:3190: FutureWarning: Possible set intersection at position 3
  self.re = re.compile(self.reString)
Error: Cannot find Jupyter/Colab namespace for Python
Error: Cannot find Jupyter/Colab namespace for Python
Error: Cannot find Jupyter/Colab namespace for Python
Test set error (RMSE):
0.654
Train set error (RMSE):
0.651

<IPython.core.display.HTML object>
<IPython.core.display.HTML object>
100% | 272/272 [00:25<00:00, 10.72it/s]
Matplotlib is currently using agg, which is a non-GUI backend, so cannot show the figure.
```

## src/client.py

```
michael@vika:/data/Education/SIT/DataSceince/a4_ML_Eng/src$ ./client.py
Now client is working in 'localhost' mode
Got response from Wine Prediction server: <Response [200]>
{"predictions": [5.015809882480472]}
```

## src/flaskApp.py

```
(pyenv) michael@vika:/data/Education/SIT/DataSceince/a4_ML_Eng/src$ ./flaskApp.py
* Serving Flask app "flaskApp" (lazy loading)
* Environment: production
  WARNING: This is a development server. Do not use it in a production deployment.
  Use a production WSGI server instead.
* Debug mode: off
* Running on http://127.0.0.1:5000/ (Press CTRL+C to quit)
Exception in thread Thread-1:
Traceback (most recent call last):
  File "/usr/lib/python3.8/threading.py", line 932, in _bootstrap_inner
    self.run()
  File "/usr/lib/python3.8/threading.py", line 1254, in run
    self.function(*self.args, **self.kwargs)
  File "/data/Education/SIT/DataSceince/pyenv/lib/python3.8/site-packages/flask_ngrok.py", line 70, in start_ngrok
    ngrok_address = _run_ngrok()
  File "/data/Education/SIT/DataSceince/pyenv/lib/python3.8/site-packages/flask_ngrok.py", line 31, in _run_ngrok
    ngrok = subprocess.Popen([executable, 'http', '5000'])
  File "/usr/lib/python3.8/subprocess.py", line 854, in __init__
    self._execute_child(args, executable, preexec_fn, close_fds,
  File "/usr/lib/python3.8/subprocess.py", line 1702, in _execute_child
    raise child_exception_type(errno_num, err_msg, err_filename)
PermissionError: [Errno 13] Permission denied: '/tmp/ngrok/ngrok'
127.0.0.1 - - [22/Apr/2021 10:15:07] "POST / HTTP/1.1" 200 -
```