Name: - Muhammad Huzaifa Waseem (2303-KHI-DEG-021)

Pair Partner 1: - Muhammad Faizan Rafique (2303.005.KHI.DEG)

Pair Partner 2: - Syed Muhammad Hammad Irshad(2303.KHI.DEG.032)

UNIT 3.4:

Assignment

Random Forest is a popular machine learning algorithm used for analyzing data. To optimize our Random Forest model, we use several tools and processes to streamline our workflow and ensure consistent results.

Mlflow_env_vars.sh is one such tool, which allows us to provide environment variables for our project. This ensures that we are using the same environment across all stages of our workflow, from data preprocessing to model training and evaluation.

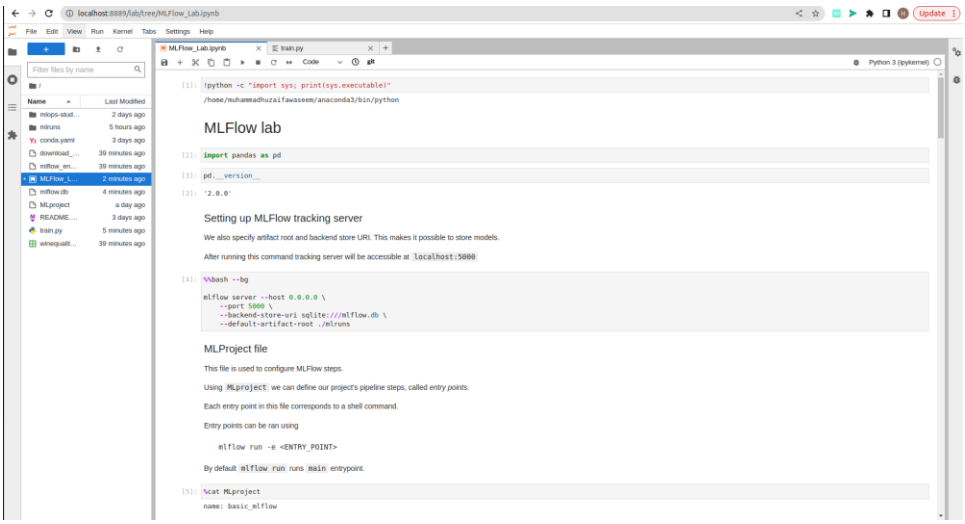
We also utilize a **conda.yaml** file to define dependencies for our project, which makes it easier to manage the libraries and packages required for our analysis.

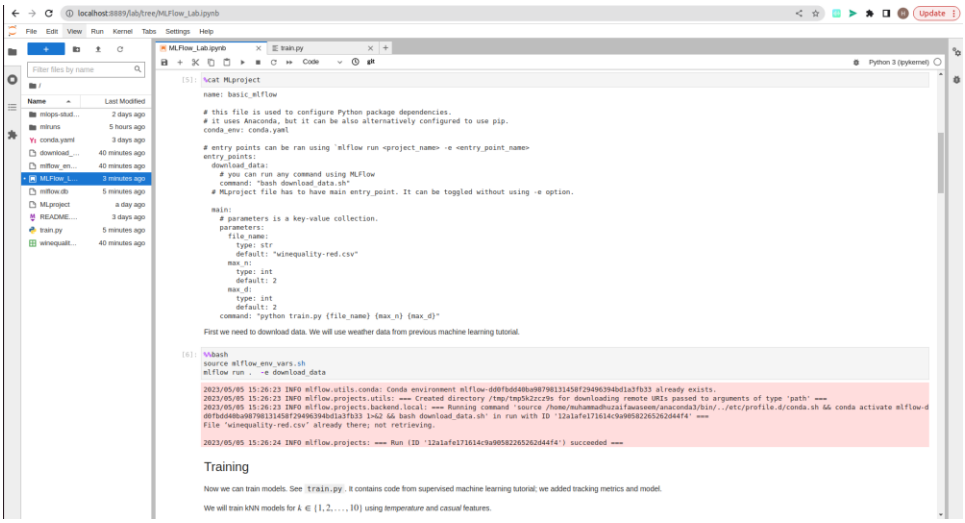
To train our model, we use **Train.py**, define the necessary functions and parameters for our Random Forest model, including the parameters max_n and max_d. These parameters are used to define the maximum number of trees in the forest and the maximum depth of the trees .This ensures that our model is optimized for our specific data and use case.

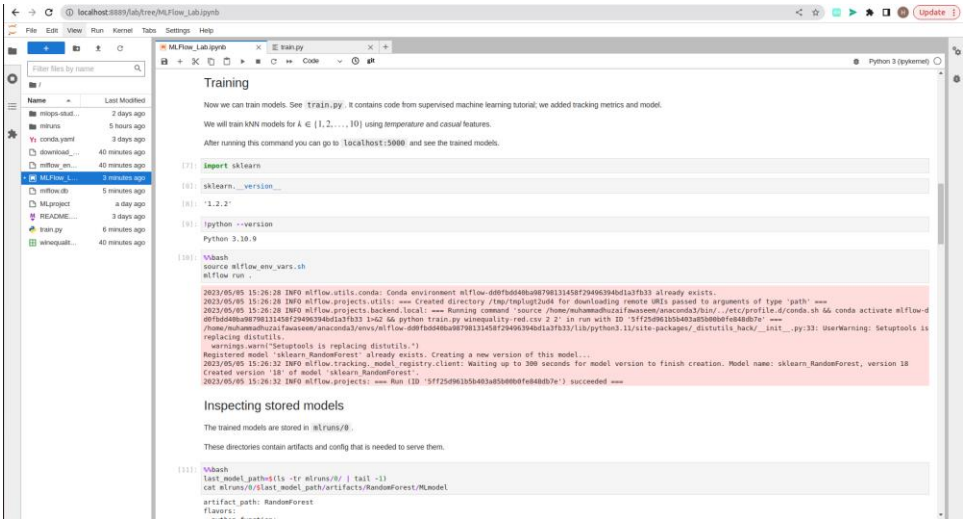
Additionally, **Download_data.sh** is called in our code to import or download data in csv format. This script ensures that we have access to the necessary data required for our analysis.

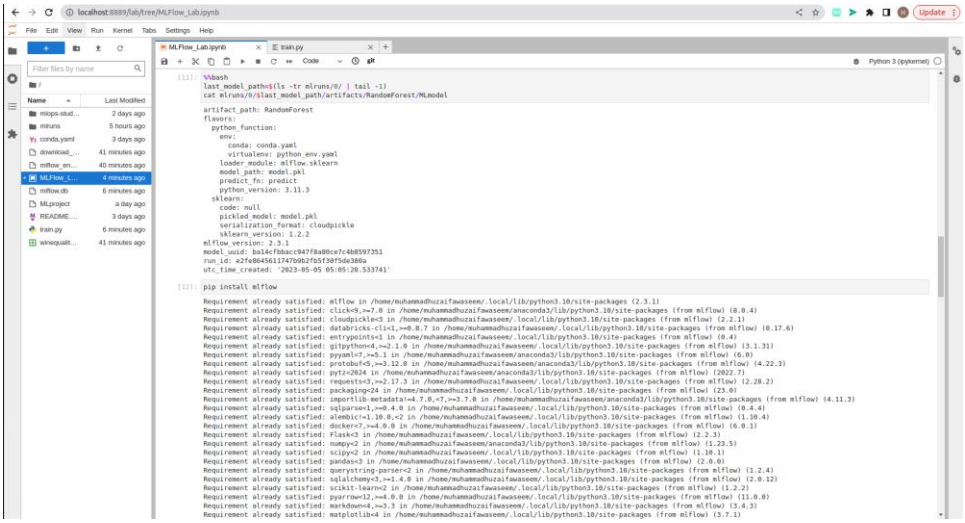
Finally, we log all our work to ensure transparency and reproducibility. By utilizing these tools and processes, we are able to efficiently and effectively analyze our data using Random Forest.

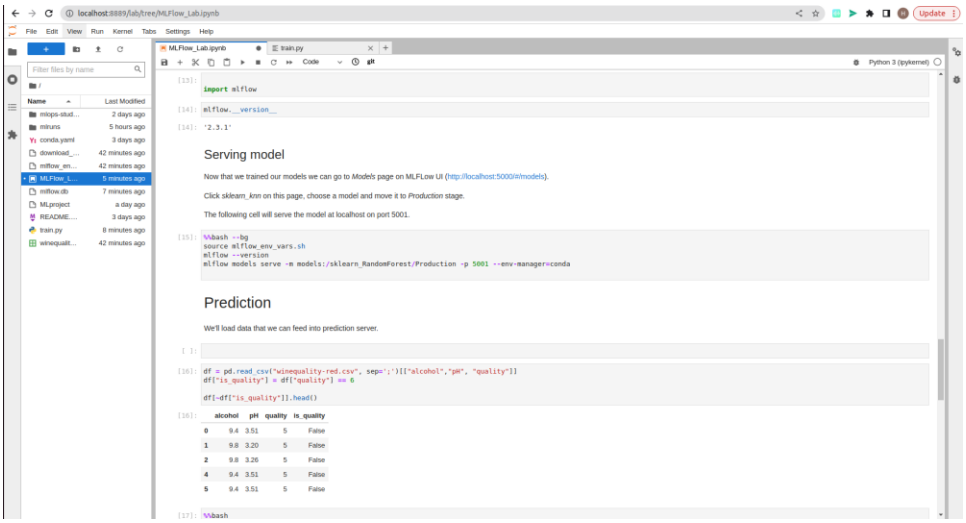
MLFlow_lab.ipynb(code):

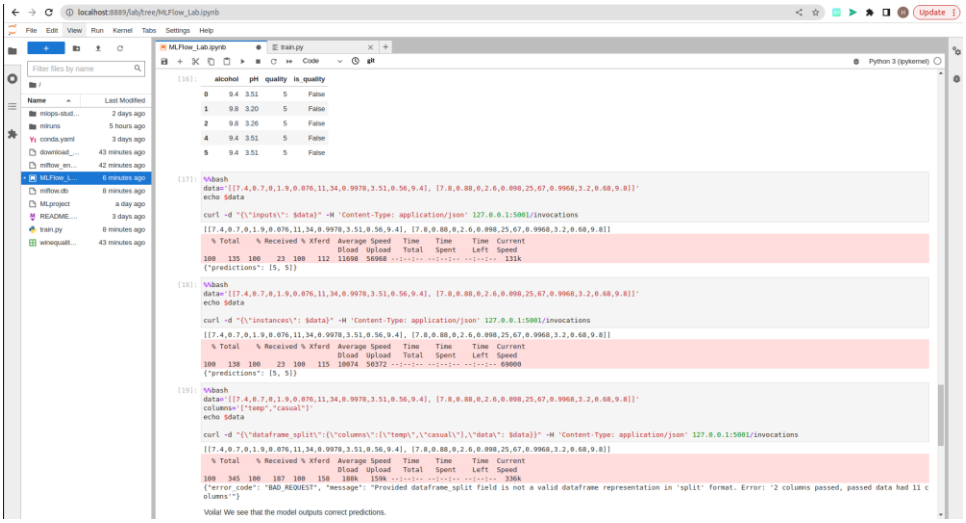



localhost:8889/lab/tree/MLFlow_Lab.ipynb
File Edit View Run Kernel Tabs Settings Help
MLFlow_Lab.ipynb
train.py
Code
Python 3 (ipykernel)
[5]: %cat MLproject
name: basic_mlflow
this file is used to configure Python package dependencies.
it uses Anaconda, but it can be also alternatively configured to use pip.
conda_env: conda.yaml
entry points can be ran using `mlflow run <project_name> -e <entry_point_name>
entry_points:
download_data:
you can run any command using MLFlow
command: "bash download_data.sh"
MLproject file has to have main entry_point. It can be toggled without using -e option.
main:
parameters is a key-value collection.
parameters:
file_name:
type: str
default: "winequality-red.csv"
max_n:
type: int
default: 2
max_d:
type: int
default: 2
command: "python train.py {file_name} {max_n} {max_d}"
First we need to download data. We will use weather data from previous machine learning tutorial.
[6]: %%bash
source mlflow_env_vars.sh
mlflow run . -e download_data
2023/05/05 15:26:23 INFO mlflow.utils.conda: Conda environment mlflow-dd0fbdd40ba987981314587294963940d1a3fb33 already exists.
2023/05/05 15:26:23 INFO mlflow.projects.utils: === Created directory /tmp/tmp5k2zcz9s for downloading remote URIs passed to arguments of type 'path' ===
2023/05/05 15:26:23 INFO mlflow.projects.backend.local: === Running command 'source /home/muhammadhuzaifawaseem/anaconda3/bin/../etc/profile.d/conda.sh && conda activate mlflow-d
d0fbdd40ba987981314587294963940d1a3fb33 1>&2 && bash download_data.sh' in run with ID '12a1afe171614c9a90582265262d44f4' ===
File 'winequality-red.csv' already there; not retrieving.
2023/05/05 15:26:24 INFO mlflow.projects: === Run (ID '12a1afe171614c9a90582265262d44f4') succeeded ===
Training
Now we can train models. See train.py . It contains code from supervised machine learning tutorial; we added tracking metrics and model.
We will train kNN models for $k \in \{1, 2, \ldots, 10\}$ using temperature and casual features.

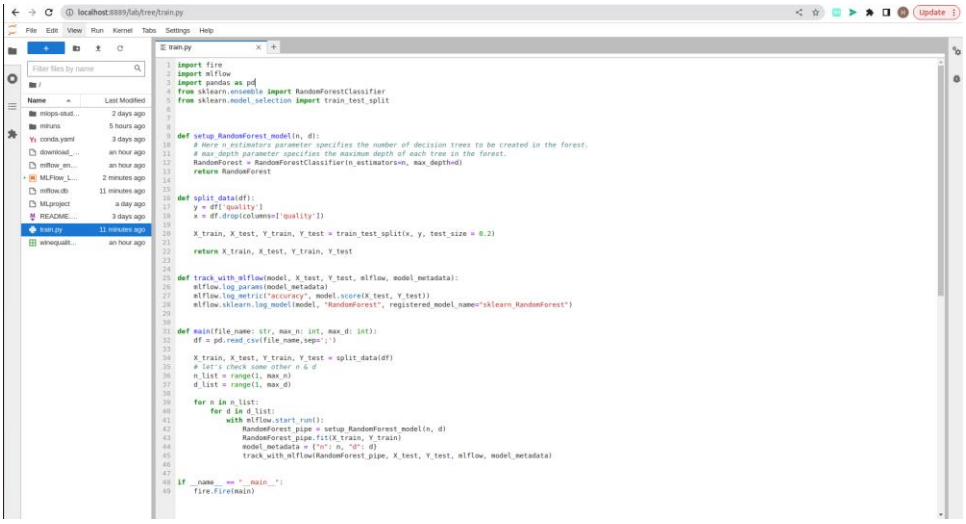

localhost:8889/lab/tree/MLFlow_Lab.ipynb
File Edit View Run Kernel Tabs Settings Help
MLFlow_Lab.ipynb
train.py
Code
Python 3 (ipykernel)
Training
Now we can train models. See train.py . It contains code from supervised machine learning tutorial; we added tracking metrics and model.
We will train kNN models for $k \in \{1, 2, \ldots, 10\}$ using temperature and casual features.
After running this command you can go to localhost:5000 and see the trained models.
[7]: import sklearn
[8]: sklearn.__version__
[8]: '1.2.2'
[9]: !python --version
Python 3.10.9
[10]: %%bash
source mlflow_env_vars.sh
mlflow run .
2023/05/05 15:26:28 INFO mlflow.utils.conda: Conda environment mlflow-dd0fbdd40ba987981314587294963940d1a3fb33 already exists.
2023/05/05 15:26:28 INFO mlflow.projects.utils: === Created directory /tmp/tmplugt2ud4 for downloading remote URIs passed to arguments of type 'path' ===
2023/05/05 15:26:28 INFO mlflow.projects.backend.local: === Running command 'source /home/muhammadhuzaifawaseem/anaconda3/bin/../etc/profile.d/conda.sh && conda activate mlflow-d
d0fbdd40ba987981314587294963940d1a3fb33 1>&2 && python train.py winequality-red.csv 2 2' in run with ID '5ff25d961b5b403a85b00b0fe848db7e' ===
/home/muhammadhuzaifawaseem/anaconda3/envs/mlflow-dd0fbdd40ba987981314587294963940d1a3fb33/lib/python3.11/site-packages/_distutils_hack/__init__.py:33: UserWarning: Setuptools is
replacing distutils.
warnings.warn("Setuptools is replacing distutils.")
Registered model 'sklearn_RandomForest' already exists. Creating a new version of this model...
2023/05/05 15:26:32 INFO mlflow.tracking._model_registry.client: Waiting up to 300 seconds for model version to finish creation. Model name: sklearn_RandomForest, version 18
Created version '18' of model 'sklearn_RandomForest'.
2023/05/05 15:26:32 INFO mlflow.projects: === Run (ID '5ff25d961b5b403a85b00b0fe848db7e') succeeded ===
Inspecting stored models
The trained models are stored in mlruns/0 .
These directories contain artifacts and config that is needed to serve them.
[11]: %%bash
last_model_path=$(ls -tr mlruns/0/ | tail -1)
cat mlruns/0/$last_model_path/artifacts/RandomForest/MLmodel
artifact_path: RandomForest
flavors:


localhost:8889/lab/tree/MLFlow_Lab.ipynb
File Edit View Run Kernel Tabs Settings Help
MLFlow_Lab.ipynb
train.py
Code
Python 3 (ipykernel)
[11]: %%bash
last_model_path=$(ls -tr mlruns/0/ | tail -1)
cat mlruns/0/$last_model_path/artifacts/RandomForest/MLmodel
artifact_path: RandomForest
flavors:
python_function:
env:
conda: conda.yaml
virtualenv: python_env.yaml
loader_module: mlflow.sklearn
model_path: model.pkl
predict_fn: predict
python_version: 3.11.3
sklearn:
code: null
pickled_model: model.pkl
serialization_format: cloudpickle
sklearn_version: 1.2.2
mlflow_version: 2.3.1
model_uuid: ba14cfbbacc947f8a80ce7c4b8597351
run_id: e2fe864561174f9b2fb5f30f5de380a
utc_time_created: '2023-05-05 05:05:20.533741'
[12]: pip install mlflow
Requirement already satisfied: mlflow in /home/muhammadhuzaifawaseem/.local/lib/python3.10/site-packages (2.3.1)
Requirement already satisfied: click<9,>=7.0 in /home/muhammadhuzaifawaseem/anaconda3/lib/python3.10/site-packages (from mlflow) (8.0.4)
Requirement already satisfied: cloudpickle<3 in /home/muhammadhuzaifawaseem/.local/lib/python3.10/site-packages (from mlflow) (2.2.1)
Requirement already satisfied: databricks-cli<1,>=0.8.7 in /home/muhammadhuzaifawaseem/.local/lib/python3.10/site-packages (from mlflow) (0.17.6)
Requirement already satisfied: entrypoints<1 in /home/muhammadhuzaifawaseem/.local/lib/python3.10/site-packages (from mlflow) (0.4)
Requirement already satisfied: gitpython<4,>=2.1.0 in /home/muhammadhuzaifawaseem/.local/lib/python3.10/site-packages (from mlflow) (3.1.31)
Requirement already satisfied: pyyaml<7,>=5.1 in /home/muhammadhuzaifawaseem/anaconda3/lib/python3.10/site-packages (from mlflow) (6.0)
Requirement already satisfied: protobuf<5,>=3.12.0 in /home/muhammadhuzaifawaseem/anaconda3/lib/python3.10/site-packages (from mlflow) (4.22.3)
Requirement already satisfied: pytz<2024 in /home/muhammadhuzaifawaseem/anaconda3/lib/python3.10/site-packages (from mlflow) (2022.7)
Requirement already satisfied: requests<3,>=2.17.3 in /home/muhammadhuzaifawaseem/.local/lib/python3.10/site-packages (from mlflow) (2.28.2)
Requirement already satisfied: packaging<24 in /home/muhammadhuzaifawaseem/.local/lib/python3.10/site-packages (from mlflow) (23.0)
Requirement already satisfied: importlib-metadata!=4.7.0,<7,>=3.7.0 in /home/muhammadhuzaifawaseem/anaconda3/lib/python3.10/site-packages (from mlflow) (4.11.3)
Requirement already satisfied: sqlparse<1,>=0.4.0 in /home/muhammadhuzaifawaseem/.local/lib/python3.10/site-packages (from mlflow) (0.4.4)
Requirement already satisfied: alembic!=1.10.0,<2 in /home/muhammadhuzaifawaseem/.local/lib/python3.10/site-packages (from mlflow) (1.10.4)
Requirement already satisfied: docker<7,>=4.0.0 in /home/muhammadhuzaifawaseem/.local/lib/python3.10/site-packages (from mlflow) (6.0.1)
Requirement already satisfied: Flask<3 in /home/muhammadhuzaifawaseem/.local/lib/python3.10/site-packages (from mlflow) (2.2.3)
Requirement already satisfied: numpy<2 in /home/muhammadhuzaifawaseem/anaconda3/lib/python3.10/site-packages (from mlflow) (1.23.5)
Requirement already satisfied: scipy<2 in /home/muhammadhuzaifawaseem/.local/lib/python3.10/site-packages (from mlflow) (1.10.1)
Requirement already satisfied: pandas<3 in /home/muhammadhuzaifawaseem/.local/lib/python3.10/site-packages (from mlflow) (2.0.0)
Requirement already satisfied: querystring-parser<2 in /home/muhammadhuzaifawaseem/.local/lib/python3.10/site-packages (from mlflow) (1.2.4)
Requirement already satisfied: sqlalchemy<3,>=1.4.0 in /home/muhammadhuzaifawaseem/.local/lib/python3.10/site-packages (from mlflow) (2.0.12)
Requirement already satisfied: scikit-learn<2 in /home/muhammadhuzaifawaseem/.local/lib/python3.10/site-packages (from mlflow) (1.2.2)
Requirement already satisfied: pyarrow<12,>=4.0.0 in /home/muhammadhuzaifawaseem/.local/lib/python3.10/site-packages (from mlflow) (11.0.0)
Requirement already satisfied: markdown<4,>=3.3 in /home/muhammadhuzaifawaseem/.local/lib/python3.10/site-packages (from mlflow) (3.4.3)
Requirement already satisfied: matplotlib<4 in /home/muhammadhuzaifawaseem/.local/lib/python3.10/site-packages (from mlflow) (3.7.1)

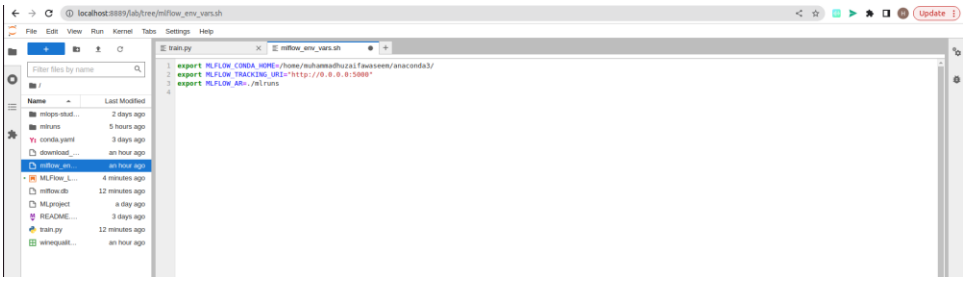




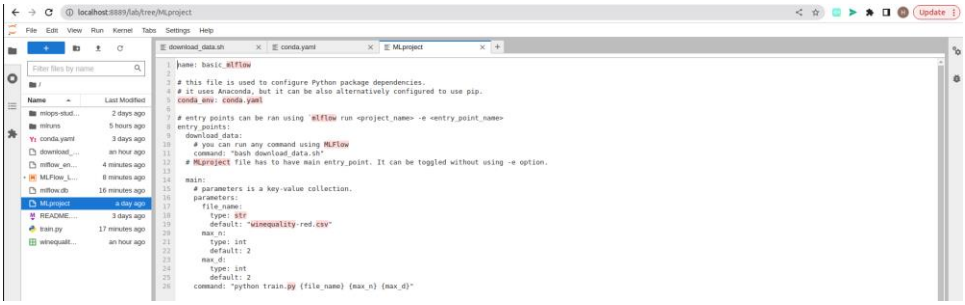
Train.py:



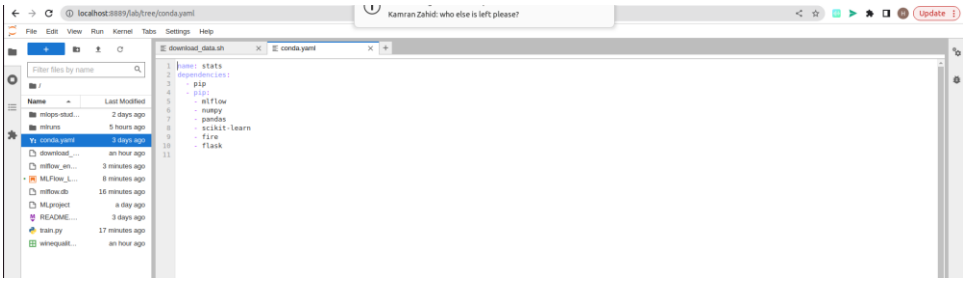
Mlflow_env_vars.sh:



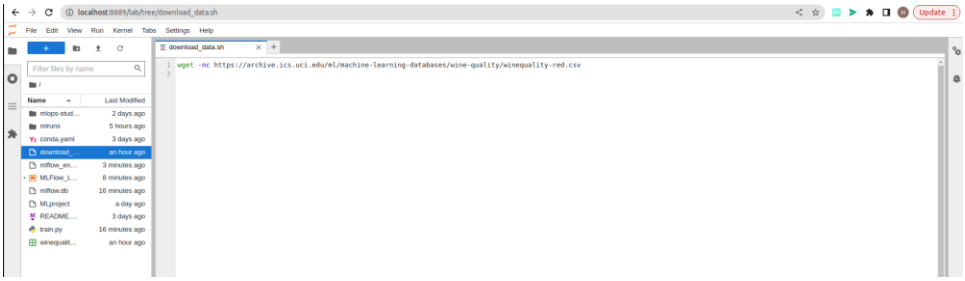
MLproject:



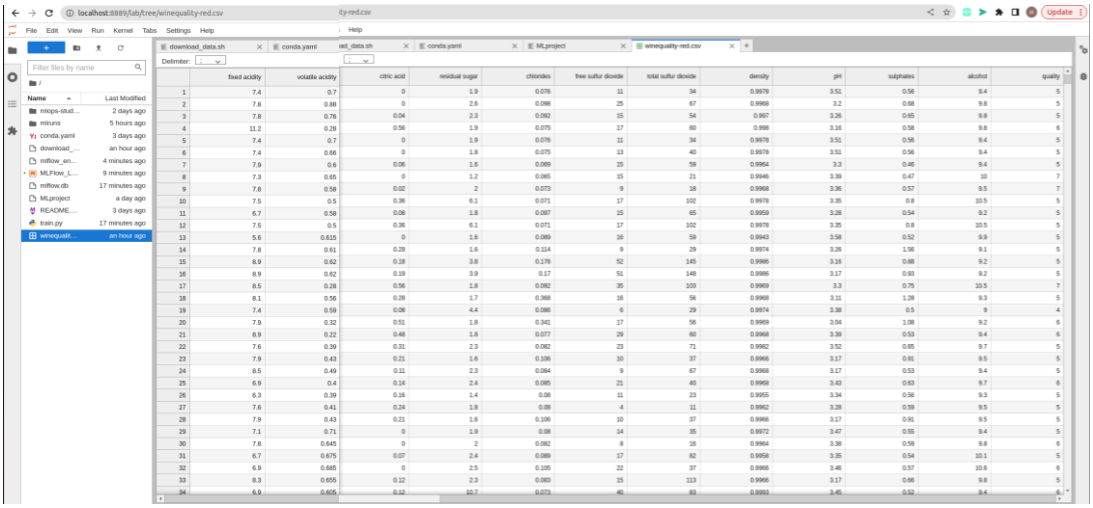
Conda.yaml:



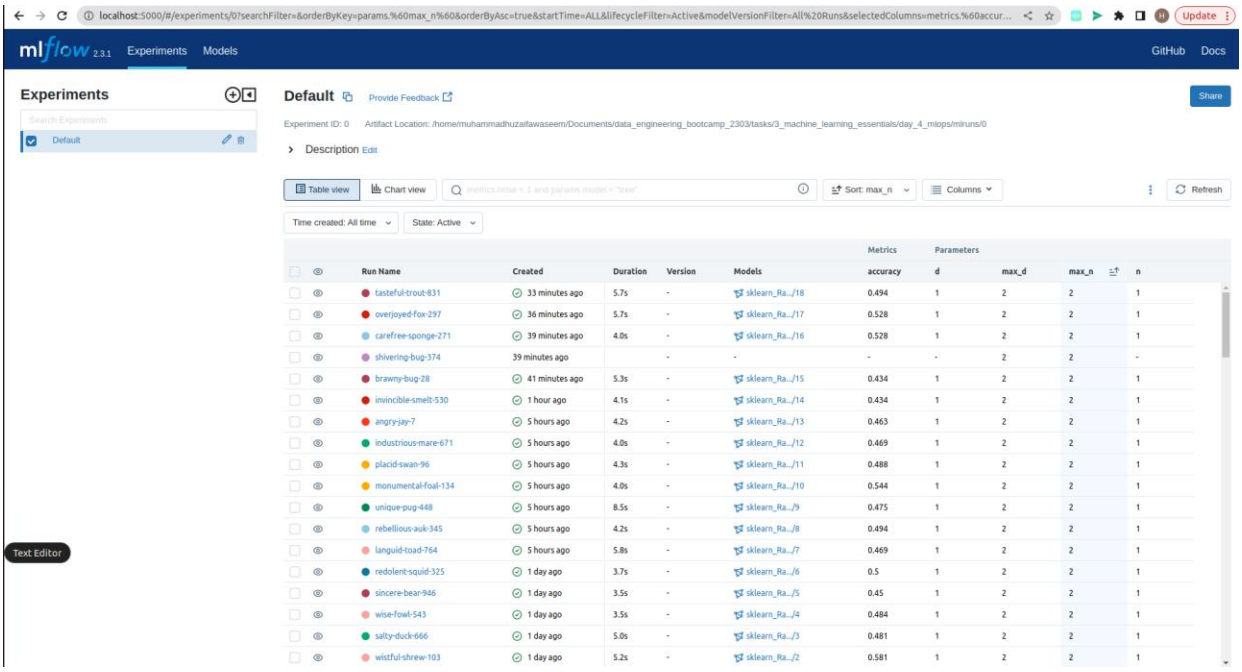
Download_data.sh:



Winequality-red.csv:



Logs:


mlflow 2.3.1
Experiments
Models
GitHub
Docs
Default
Share
Table view
Chart view
Refresh
Run Name
Created
Duration
Version
Models
accuracy
d
max_d
max_n
n
Text Editor