Summative Assignment
Deployment

System Deployment

You will demonstrate the deployment of your Machine learning model. As a Machine Learning engineer, you have been tasked with creating an ML Pipeline and scaling and monitoring it on a cloud platform of your choice.

Your tasks will involve:

Creating a Machine Learning Classification model offline and saving it as a file.

Evaluate the model(s) using all the metrics required on a Jupyter Notebook and demonstrate how good the model(s) is.

1. Create the whole pipeline process with Python functions.
2. Load the pipeline on the cloud platform in such a way that the model can be retrained again, and create a trigger for retraining the model when the need arises.
3. Demonstrate the evaluation process of the model in production.
4. Simulate a flood of requests (using software like Locust) send them to the model and show how the model responds to these requests. Record and show the latency and response time of the requests with different numbers of docker containers.
5. Demonstrate how a user uploads values/features and the model predicts
6. A User should be able to upload new data and trigger retraining.
7. Web applications have to be dockerized.

The Web UI Should:

1. Have a page where you can make predictions on
2. Have a homepage with data visualizations on it
3. Have a page where you can upload new data
4. Have a feature where one can trigger a retraining process

Github Repo Directory Structure

Project_name/

README.md

notebook/

```
│   ├──project_name.ipynb
│
├── src/
│   ├── preprocessing.py
│   ├── model.py
│   └── prediction.py
│
├── data/
│   ├──train/
│   └── test/
│
└── models/
    ├── _model_name.pkl
    └── _model_name.tf
```

Requirements

1. A link to the GitHub repo.
    1. The README.md should have clear instructions on:
        1. URL where applicable
        2. The project description
        3. And clear steps on how to set it up
    2. Notebook
        1. Should contain all preprocessing steps as python Functions
        2. Model Training
        3. Model Test / Prediction Functions
    3. The model file
        1. Pickle (.pkl)
        2. Or tensorflow (.tf) file

1. Deployment Package. This could be a
    1. Public URL (Render, Streamlit, etc)
    2. Docker Image
    3. A mobile app
    4. Desktop app

Submission Instructions
You will have two attempts during submission Make sure to submit the following in each attempt respectively

1. The first attempt will be a Zip File of the GitHub Code Repository
2. The second attempt will be a GitHub Repository URL