

MLOPs Final Project Report

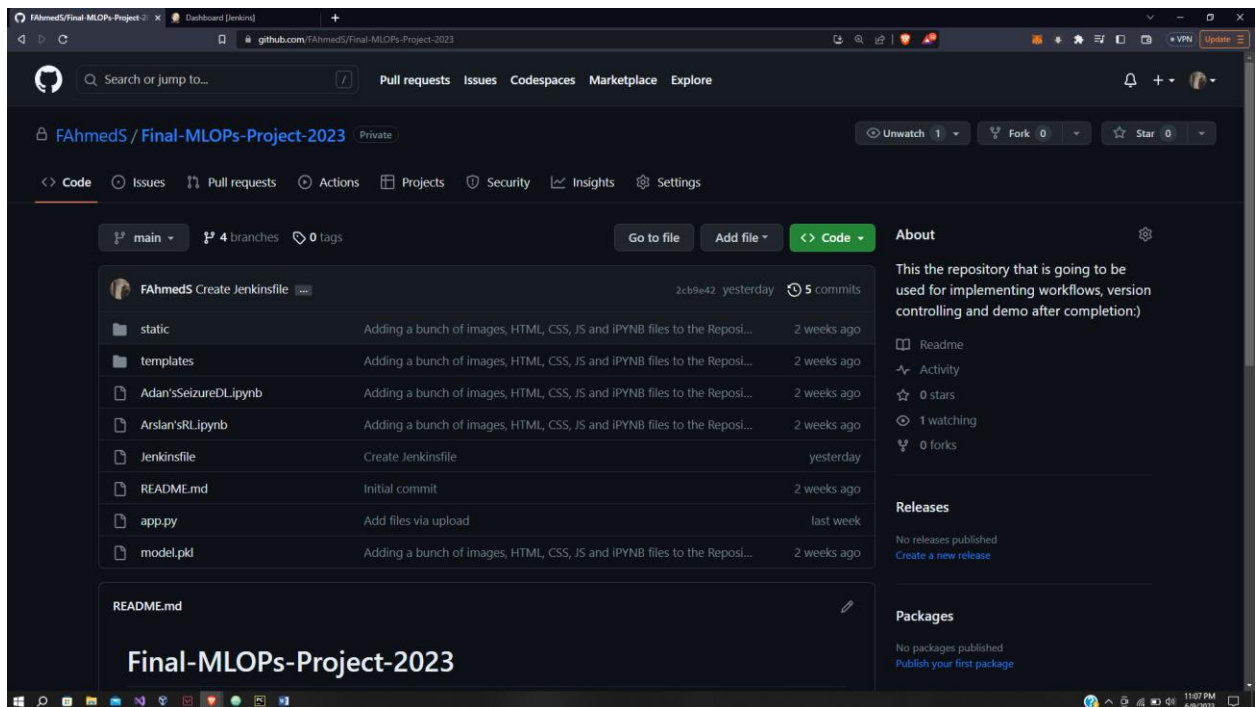
Group members:

- Faisal Ahmed Siddiqui
- Adan Nazir
- Zaid Khan

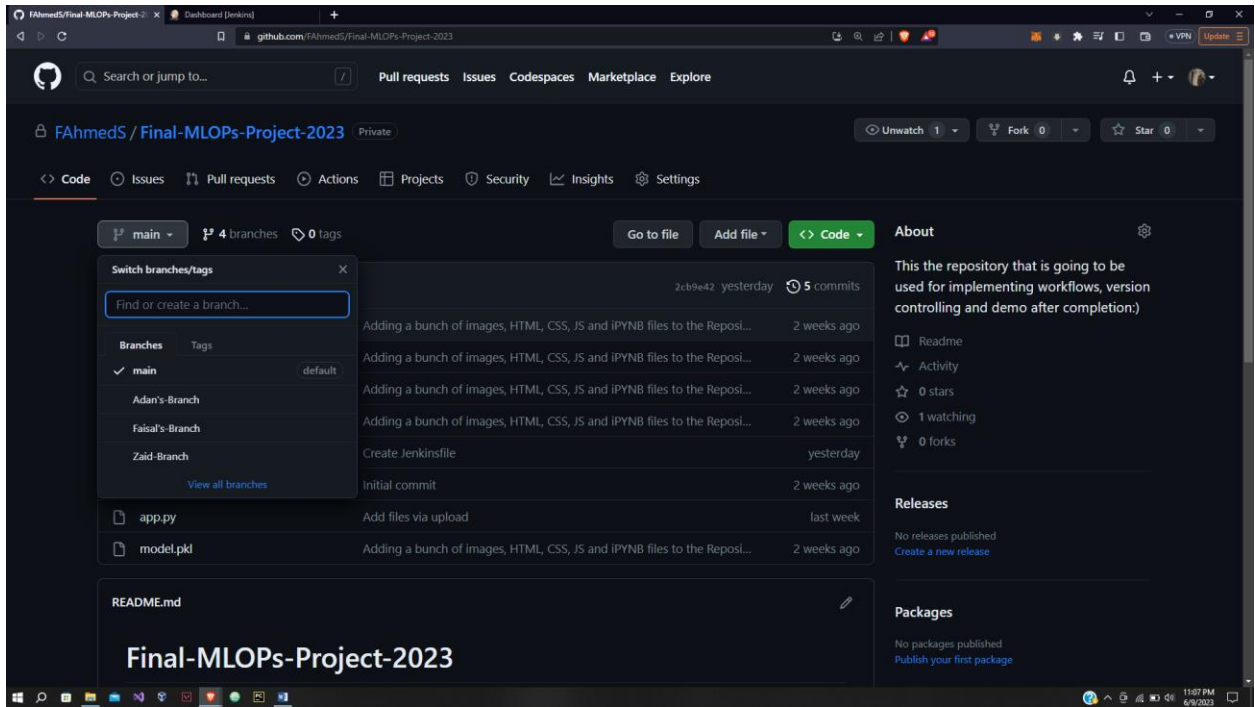
Section: AI-J

Steps: Just followed the Evaluation criteria for implementation.

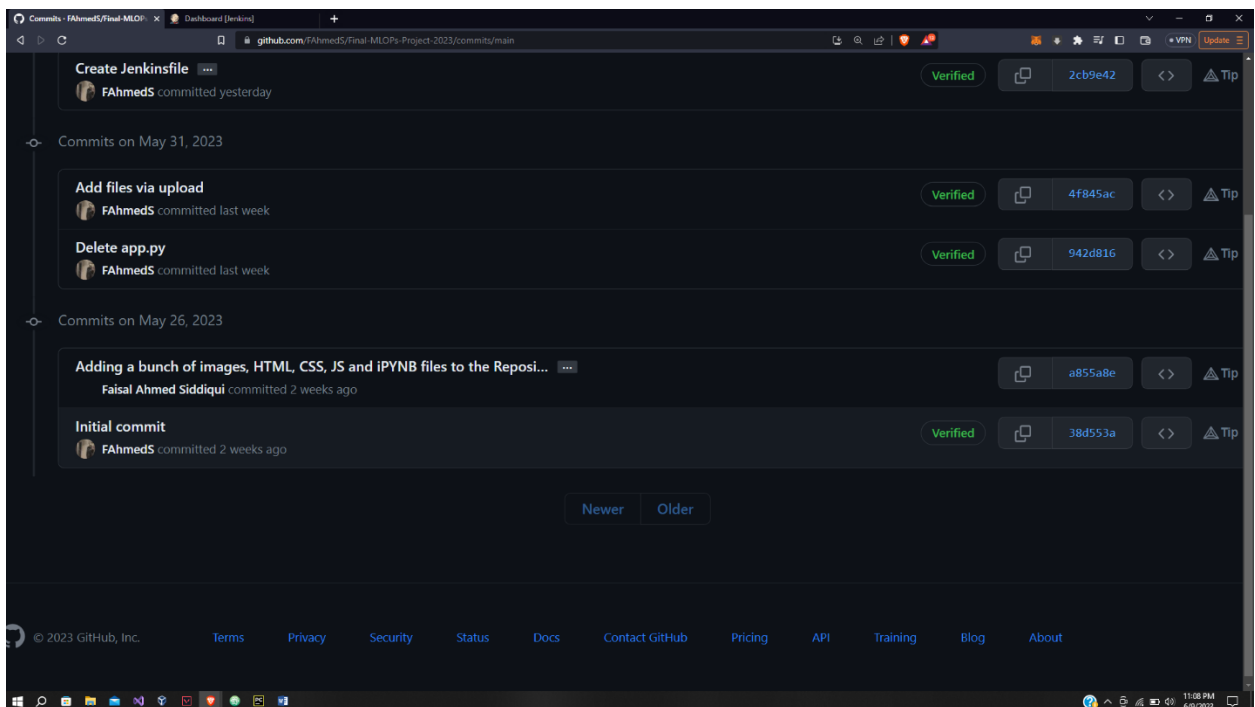
1. Github & Git used for Version controlling:



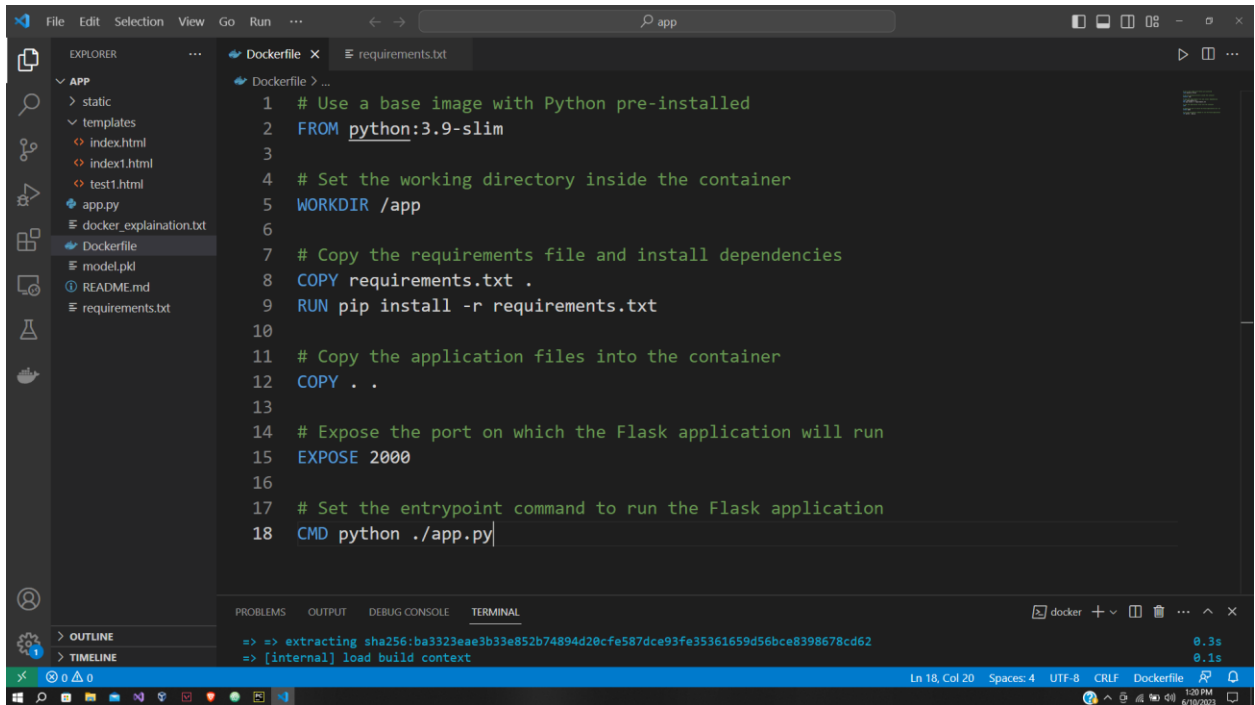
2. Branches created.



3. Created meaningful commit



4. **Docker/Containerization:** For containerization used VS code with docker plugin, then made Dockerfile and requirements.txt file for image making:

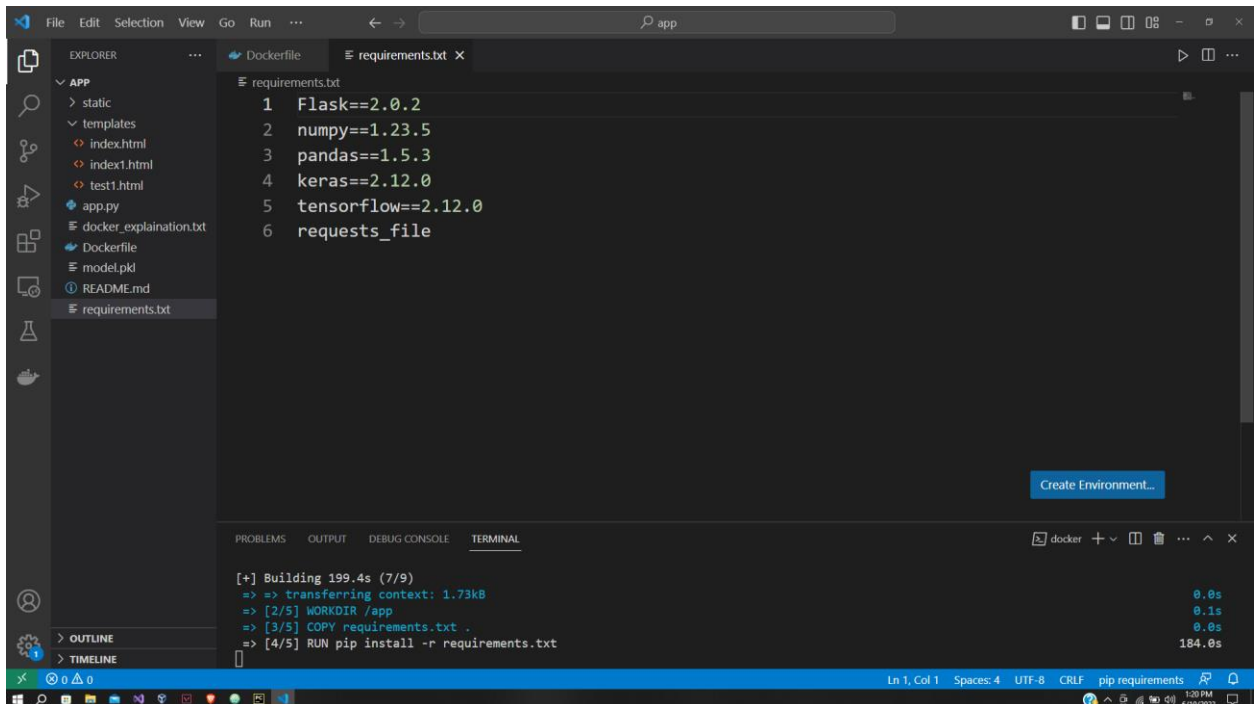


```
1 # Use a base image with Python pre-installed
2 FROM python:3.9-slim
3
4 # Set the working directory inside the container
5 WORKDIR /app
6
7 # Copy the requirements file and install dependencies
8 COPY requirements.txt .
9 RUN pip install -r requirements.txt
10
11 # Copy the application files into the container
12 COPY . .
13
14 # Expose the port on which the Flask application will run
15 EXPOSE 2000
16
17 # Set the entrypoint command to run the Flask application
18 CMD python ./app.py
```

PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL

=> extracting sha256:ba3323eae3b33e852b74894d20cfe587dce93fe35361659d56bce8398678cd62 0.3s
=> [internal] load build context 0.1s

Ln 18, Col 20 Spaces: 4 UTF-8 CRLF Dockerfile 6/10/2023



```
1 Flask==2.0.2
2 numpy==1.23.5
3 pandas==1.5.3
4 keras==2.12.0
5 tensorflow==2.12.0
6 requests_file
```

PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL

[+] Building 199.4s (7/9)
=> transferring context: 1.73kB 0.0s
=> [2/5] WORKDIR /app 0.1s
=> [3/5] COPY requirements.txt . 0.0s
=> [4/5] RUN pip install -r requirements.txt 184.0s

Ln 1, Col 1 Spaces: 4 UTF-8 CRLF pip requirements 6/10/2023

```
7 # Copy the requirements file and install dependencies
8 COPY requirements.txt .
9 RUN pip install -r requirements.txt
10
11 # Copy the application files into the container
12 COPY . .
13
14 # Expose the port on which the Flask application will run
15 EXPOSE 2000
16
17 # Set the entrypoint command to run the Flask application
18 CMD python ./app.py
```

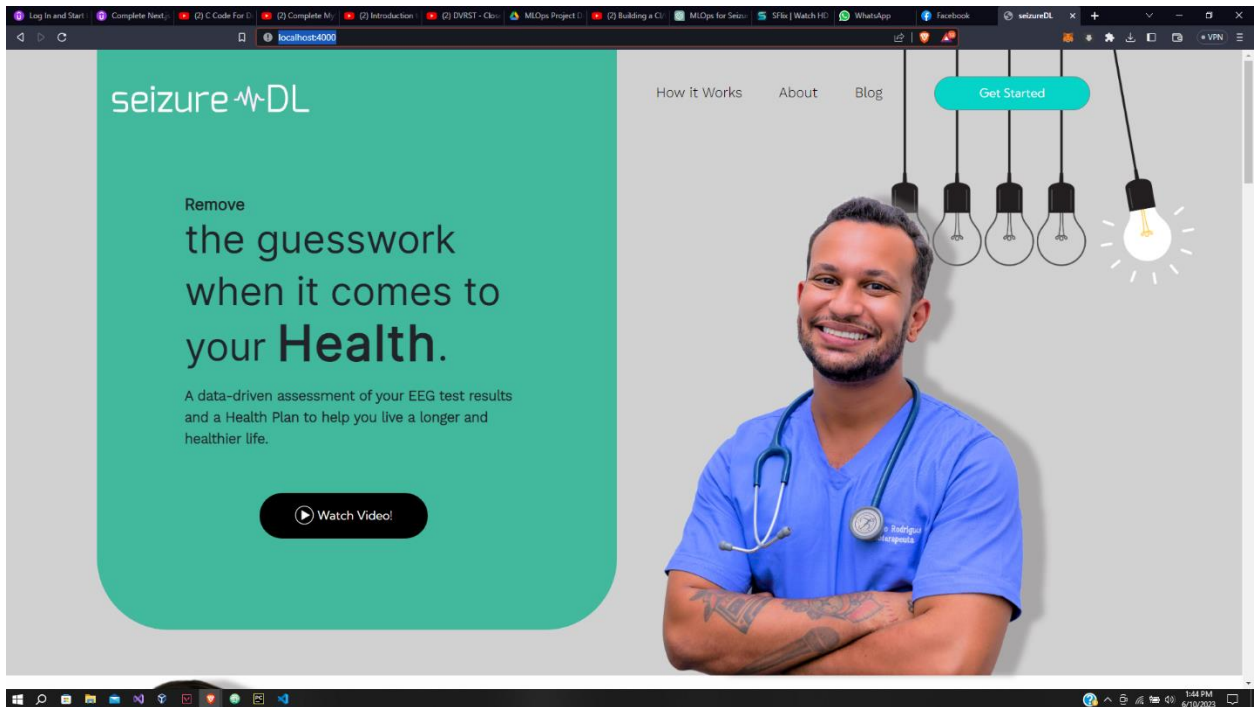
```
PS C:\Users\HP\Desktop\FYP Stuff\FYP2_MID_EVAL\app> docker build -t mlops-docker-final .
[+] Building 35.6s (7/9)
=> extracting sha256:ba3323eae3b33e852b74894d20cfe587dce93fe35361659d56bce8398678cd62 0.3s
=> [internal] load build context 0.1s
=> transferring context: 1.73kB 0.0s
=> [2/5] WORKDIR /app 0.1s
=> [3/5] COPY requirements.txt . 0.0s
=> [4/5] RUN pip install -r requirements.txt 20.2s
=> # 95.2/95.2 kB 1.4 MB/s eta 0:00:00
=> # Collecting numpy==1.23.5
=> # Downloading numpy-1.23.5-cp39-cp39-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (17.1 MB)
=> # 17.1/17.1 MB 1.8 MB/s eta 0:00:00
=> # Collecting pandas==1.5.3
=> # Downloading pandas-1.5.3-cp39-cp39-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (12.2 MB)
```

5. Created image successfully then ran container on the localhost 4000, flask application runs on the calhost:4000 AND then stopped container as well:

```
1 Flask==2.0.2
2 numpy==1.23.5
3 pandas==1.5.3
4 keras==2.12.0
5 tensorflow==2.12.0
6 requests_file
```

```
PS C:\Users\HP\Desktop\FYP Stuff\FYP2_MID_EVAL\app> docker images
REPOSITORY TAG IMAGE ID CREATED SIZE
mlops-docker-final latest 80584a4c8530 10 minutes ago 2.76GB
mlops-docker-tested1 latest 89534610445b 11 days ago 325MB
gcr.io/k8s-minikube/kicbase v0.0.37 01c0ce65fff7 4 months ago 1.15GB
PS C:\Users\HP\Desktop\FYP Stuff\FYP2_MID_EVAL\app> docker container run -d -p 4000:2000 mlops-docker-final
b0f8e71488f07ed29936dca2e8c1f770778cf8bb3315927ae09d156b08de87e6
PS C:\Users\HP\Desktop\FYP Stuff\FYP2_MID_EVAL\app> docker container ls
CONTAINER ID IMAGE COMMAND CREATED STATUS PORTS NAMES
b0f8e71488f0 mlops-docker-final "/bin/sh -c 'python ..." 16 seconds ago Up 14 seconds 0.0.0.0:4000->2000/tcp reverent_jackson
PS C:\Users\HP\Desktop\FYP Stuff\FYP2_MID_EVAL\app> docker container stop b0f
b0f
PS C:\Users\HP\Desktop\FYP Stuff\FYP2_MID_EVAL\app> docker container ls
CONTAINER ID IMAGE COMMAND CREATED STATUS PORTS NAMES
PS C:\Users\HP\Desktop\FYP Stuff\FYP2_MID_EVAL\app>
```

6. Container running on localhost:4000:



7. CI/CD pipeline using Jenkins:

- i. For this first I integrated the GitHub to Jenkins via webhook successfully.
- ii. As the repo was private so I had to make/add credentials for SSH authorization.
- iii. For now, just created random Jenkins file to check whether build is success or not.

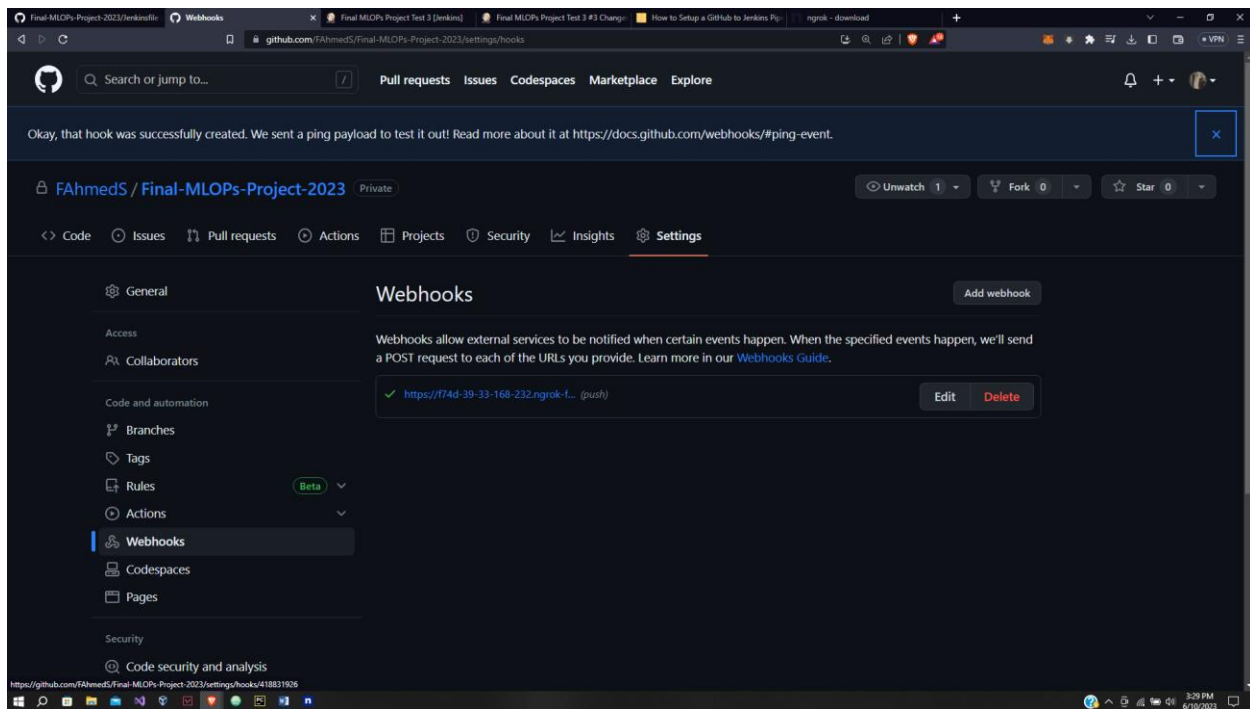
```

Command Prompt
(c) Microsoft Corporation. All rights reserved.

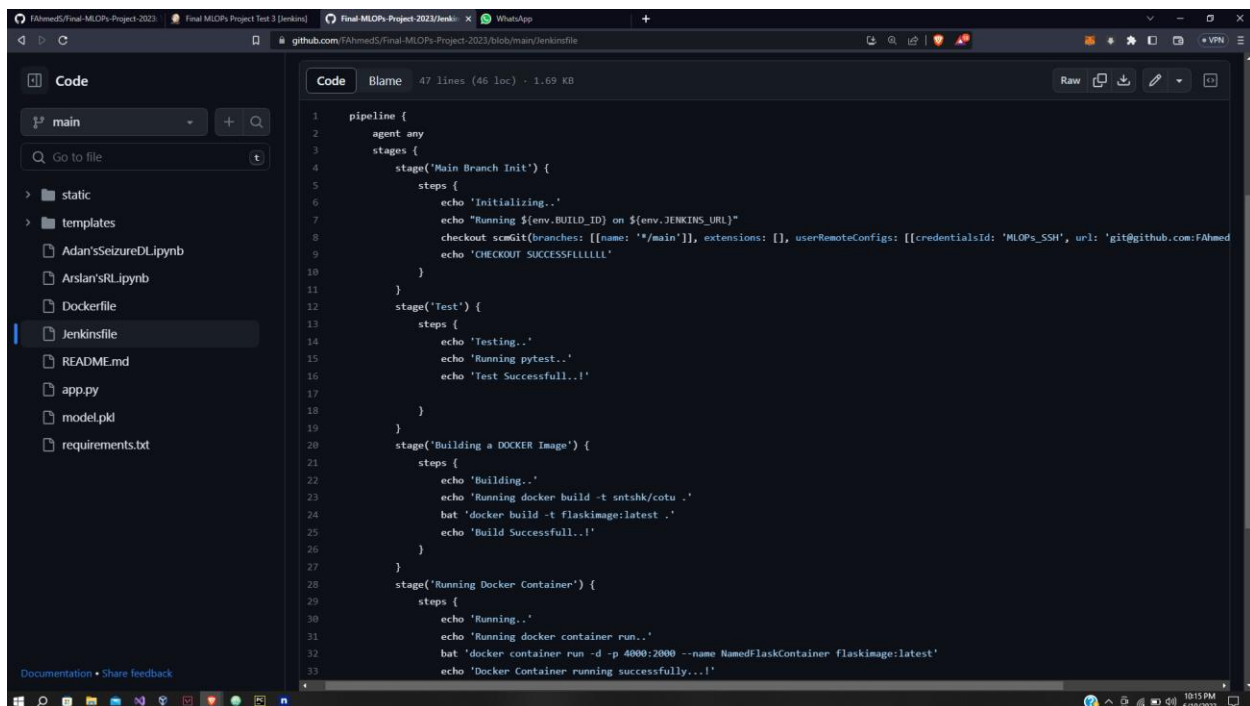
C:\Users\VP>ngrok http 8080
'ngrok' is not recognized as an internal or external command,
operable program or batch file.

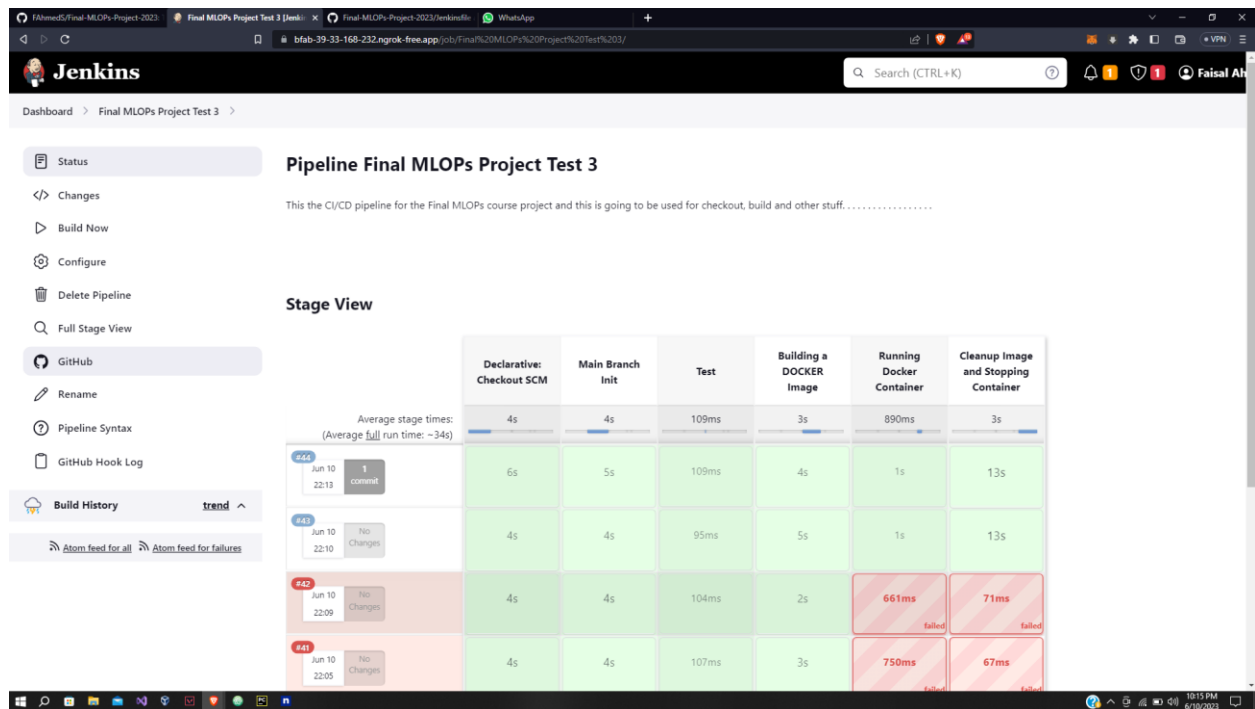
C:\Users\VP>ssh-keygen -t ed25519 -C 1191674@mu.edu.pk
Generating public/private ed25519 key pair.
Enter file in which to save the key (C:\Users\VP/.ssh/id_ed25519):
Created directory 'C:\Users\VP/.ssh'.
Enter passphrase (empty for no passphrase):
Enter same passphrase again:
Your identification has been saved in C:\Users\VP/.ssh/id_ed25519.
Your public key has been saved in C:\Users\VP/.ssh/id_ed25519.pub.
The key fingerprint is:
SHA256:13+FR2J9WncRDT3806u3N6L0D5a/yc89Kt9mhuUA 1191674@mu.edu.pk
The key's randomart image is:
+--[ED25519 256]--+
|      .+..+..+      |
|      o  +  o  +    |
|      +o*  .+*+     |
|      o+  .o+       |
|      S .o  +..+    |
|      o .o  +..+    |
|      .o+..+       |
|      .o+..+       |
|      +o+..+       |
+-----[SHA256]-----+
C:\Users\VP>

```



8. Now Building Docker image and run container using Jenkinsfile workflow, for this made some changes to Jenkinsfile from Github and Builds were successful as shown below:





9. Now for the Apache Airflow, converted my whole flask application code to DAG compatible.

```

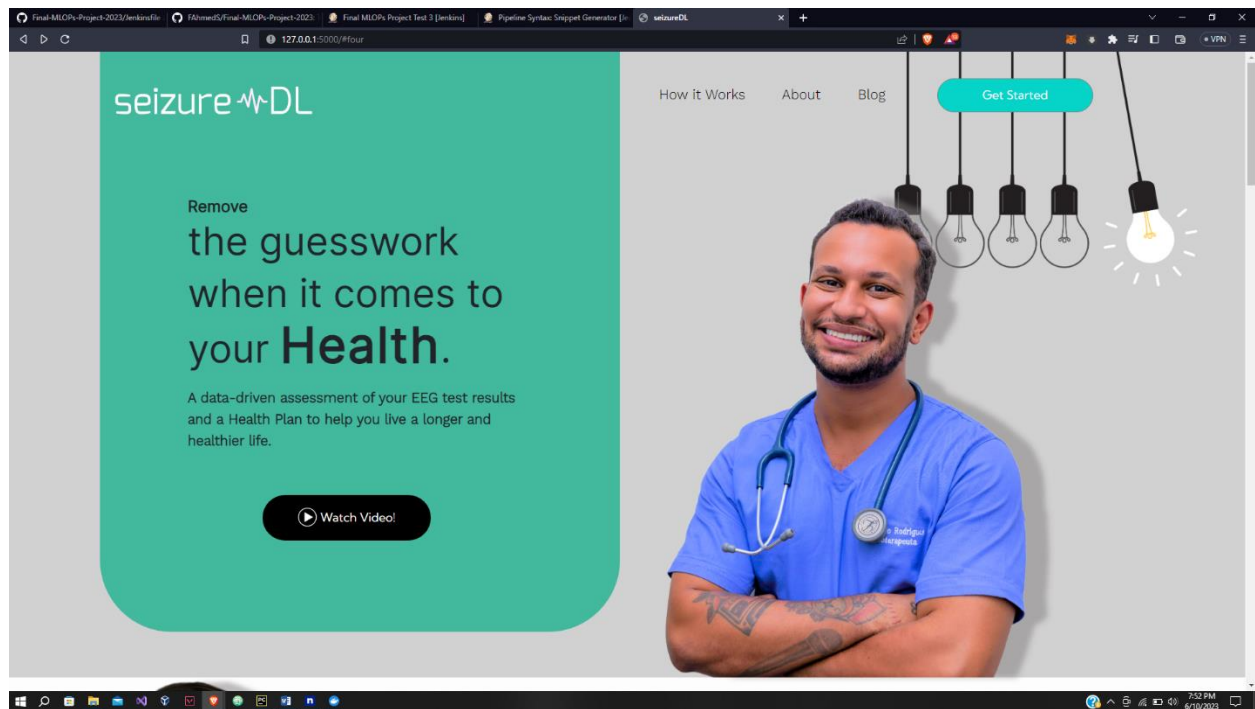
pythonProject-airflow  Version control
Project
  pythonProject-airflow C:\Users\HP\PylamProjects
    static
    templates
    app.py
    model.pkl
  External Libraries
  Scratches and Consoles

app.py
37     return render_template('index1.html', results="Predicted Value is {:.3f}, Patient is less likely to get Seizure.".format(
38     else:
39     return render_template('index1.html', results="Predicted Value is {:.3f}, Patient is likely to get Seizure.".format(val))
40
41 dag = DAG(
42     'flask_airflow_dag',
43     description='DAG for Flask and Airflow integration',
44     schedule_interval=None,
45     start_date=datetime(2023, 5, 25),
46 )
47
48 load_model_task = PythonOperator(
49     task_id='load_model_task',
50     python_callable=load_model,
51     dag=dag,
52 )
53
54 run_flask_app_task = PythonOperator(
55     task_id='run_flask_app_task',
56     python_callable=run_flask_app,
57     dag=dag,
58 )
59
60 @app.route('/')
run_flask_app()

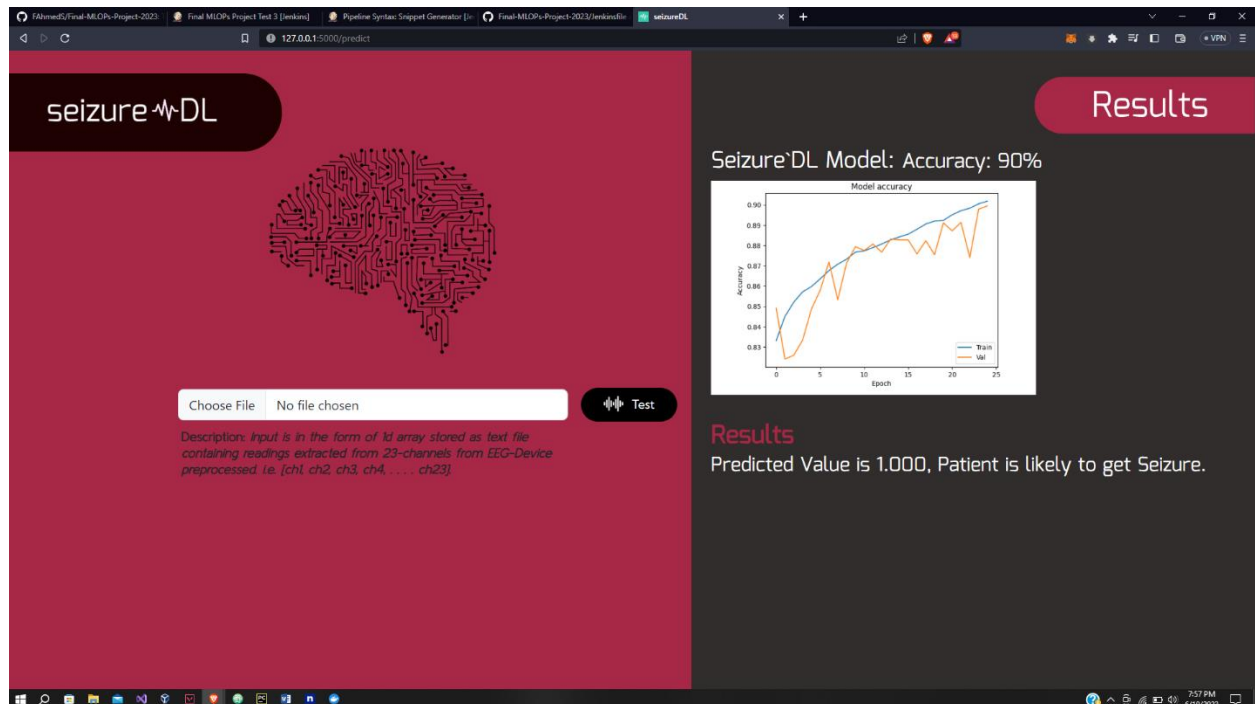
Run app
[2023-06-10T19:51:02.979+0500] {internal.py:224} INFO - Press CTRL+C to quit
[2023-06-10T19:51:03.008+0500] {internal.py:224} INFO - * Restarting with watchdog (windowsapi)
WARNING:root:OSError while attempting to symlink the latest log directory
2.6.1
Starting the Flask Application!
[2023-06-10T19:51:06.056+0500] {internal.py:224} WARNING - * Debugger is active!
[2023-06-10T19:51:06.710+0500] {internal.py:224} INFO - * Debugger PIN: 871-982-517
pythonProject-airflow  app.py
22:33 CNLF UTF-8 4 spaces Python 3.10
7:51 PM 6/10/2023

```

10. Which works fine as can be seen:



11. Prediction Value is shown:



12. Initializing DVC


```
!dvc init
```

Initialized DVC repository.

You can now commit the changes to git.

```
+-----+
|               |
|  DVC has enabled anonymous aggregate usage analytics.  |
|  Read the analytics documentation (and how to opt-out) here:  |
|  <https://dvc.org/doc/user-guide/analytics>                |
|               |
+-----+
```

What's next?

- Check out the documentation: <<https://dvc.org/doc>>
- Get help and share ideas: <<https://dvc.org/chat>>
- Star us on GitHub: <<https://github.com/iterative/dvc>>

13. Add data to DVC

```
In [15]: !dvc add data/data.csv
```

'' Checking graph

Adding...

```
!
 0% Checking cache in '/home/user/.dvc/cache'| 0/? [00:00<?, ?files/s]
```

```
!
 0%|          |Transferring          0/? [00:00<?, ?file/s]
 0%|          |Transferring          0/1 [00:00<?, ?file/s]
```

```
!
 0%|          |Checking out data/data.csv 0/? [00:00<?, ?files/s]
 0%|          |Checking out data/data.csv 0/1 [00:00<?, ?files/s]
100%|██████████|Checking out data/data.csv 1/1 [00:01<00:00, 1.33s/files]
100% Adding...|████████████████████████████████████████|1/1 [00:05, 5.98s/file]
```

To track the changes with git, run:

```
git add data/.gitignore data/data.csv.dvc
```

To enable auto staging, run:

```
dvc config core.autostage true
```

14. Setting up remote storage

```
In [27]: !dvc remote add -d storage gdrive://10D5lejHosPJtL6NJtxAoh03s7PdD4Yi
```

Setting 'storage' as a default remote.

```
In [28]: !git commit .dvc/config -m "Configure remote storage"
```

```
[master 36cb8e8] Configure remote storage
Committer: Parrot Live user <user@localhost.localdomain>
Your name and email address were configured automatically based
on your username and hostname. Please check that they are accurate.
You can suppress this message by setting them explicitly. Run the
following command and follow the instructions in your editor to edit
your configuration file:
```

```
git config --global --edit
```

After doing this, you may fix the identity used for this commit with:

```
git commit --amend --reset-author
```

15. Pushing data to remote storage

```
In [31]: !dvc push
```

```
96% |██████████| /home/user/.dvc/cache/3c/51f600M/626M [04:06<00:10, 1.32MB/s]
96% |██████████| /home/user/.dvc/cache/3c/51f604M/626M [04:07<00:10, 2.23MB/s]
97% |██████████| /home/user/.dvc/cache/3c/51f605M/626M [04:08<00:10, 2.12MB/s]
97% |██████████| /home/user/.dvc/cache/3c/51f607M/626M [04:09<00:10, 2.00MB/s]
97% |██████████| /home/user/.dvc/cache/3c/51f608M/626M [04:10<00:09, 1.99MB/s]
97% |██████████| /home/user/.dvc/cache/3c/51f610M/626M [04:11<00:11, 1.55MB/s]
98% |██████████| /home/user/.dvc/cache/3c/51f611M/626M [04:12<00:10, 1.55MB/s]
98% |██████████| /home/user/.dvc/cache/3c/51f613M/626M [04:13<00:07, 1.79MB/s]

98% |██████████| /home/user/.dvc/cache/3c/51f614M/626M [04:14<00:07, 1.69MB/s]
98% |██████████| /home/user/.dvc/cache/3c/51f616M/626M [04:15<00:06, 1.78MB/s]
99% |██████████| /home/user/.dvc/cache/3c/51f617M/626M [04:16<00:06, 1.47MB/s]
99% |██████████| /home/user/.dvc/cache/3c/51f619M/626M [04:18<00:06, 1.29MB/s]
99% |██████████| /home/user/.dvc/cache/3c/51f620M/626M [04:19<00:04, 1.45MB/s]
99% |██████████| /home/user/.dvc/cache/3c/51f622M/626M [04:19<00:02, 1.64MB/s]
99% |██████████| /home/user/.dvc/cache/3c/51f623M/626M [04:20<00:01, 1.76MB/s]
100% |██████████| /home/user/.dvc/cache/3c/51f624M/626M [04:20<00:00, 1.99MB/s]
100% Transferring|██████████| [1/1 [04:25<00:00, 265.83s/file]
1 file pushed
```

16. Pulling dataset from DVC:

```
1 import dvc
2 import git
```

```
1 !git init
2 !dvc init
```

Reinitialized existing Git repository in /home/user/.git/
Initialized DVC repository.

You can now commit the changes to git.

```
-----
DVC has enabled anonymous aggregate usage analytics.
Read the analytics documentation (and how to opt-out) here:
<https://dvc.org/doc/user-guide/analytics>
-----
```

What's next?

- Check out the documentation: <<https://dvc.org/doc>>
- Get help and share ideas: <<https://dvc.org/chat>>
- Star us on GitHub: <<https://github.com/iterative/dvc>>

```
1 !dvc remote add -d storage gdrive://10D51ejuHosPjtL6NJtxAoh03s7PdD4Yi
```

Setting 'storage' as a default remote.

```
1 !dvc pull
```

```
0% Transferring|                                |0/1 [00:00<?,    ?file/s]
!
0%|          |10D51ejuHosPjtL6NJtxAoh03s7PdD4Yi/30.00/? [00:00<?,    ?B/s]
0%|          |10D51ejuHosPjtL6NJtxAoh03s7PdD4Y0.00/626M [00:00<?,    ?B/s]
16%|█        |10D51ejuHosPjtL6NJtxAoh03s7P100M/626M [01:31<08:01,    1.15MB/s]
```

17. Setting up MLFlow and integrating it with Custom ML model:

```
[17] import mlflow.tensorflow
      experiment_name = "New Experiment"

      # Create a new experiment
      experiment_id = mlflow.create_experiment(experiment_name)

      print("New experiment created with ID:", experiment_id)

      mlflow.set_tracking_uri("http://localhost:5000")
      experiment = mlflow.get_experiment_by_name("New Experiment")

      # Get the experiment URI
      experiment_uri = experiment.artifact_location

      print("Experiment URI:", experiment_uri)
      mlflow.set_tracking_uri(experiment_uri)

      # Start a new MLflow run
      pkl_file="model.pkl"
      # Load your .pkl file
      with mlflow.start_run():
          optimizer = tf.keras.optimizers.Adam(learning_rate=0.011)
          model.compile(optimizer=optimizer, loss='binary_crossentropy', metrics=['accuracy'])
          hist=model.fit(X_train_norm,train_labels, validation_data=(X_val_norm,val_labels), epochs=5, batch_size=128,verbose=1)
          mlflow.log_artifact(pkl_file)
          mlflow.log_param("learning_rate",0.23)
          mlflow.log_param("batch_size",64)
          mlflow.log_param("num_epochs",5)
          mlflow.log_param("Optimizer","Adam")
          mlflow.log_metric("Accuracy",85.94)
          mlflow.log_param("Loss","Binary Cross Entropy")
          mlflow.log_metric("Precision",0.8742)
          mlflow.log_metric("Recall",0.92)
          mlflow.log_metric("F1 score",0.8916)
```

The above snippet contains the basic workflow of MLFlow, in which we log the model's parameters i.e Learning rate, Batch Size, Optimizer and the metrics i.e Accuracy, Precision, Recall and F1-score. We train the model, which is a Sequential (Custom) model for our FYP, and then we carry 2 different experiments on the same model and demonstrate our findings using MLFLOW UI.

18. Model Architecture:

```
model = tf.keras.Sequential([
    tf.keras.layers.Dense(848, activation='relu', input_shape=(23,)),
    tf.keras.layers.BatchNormalization(),

    tf.keras.layers.Dense(944, activation='relu'),
    tf.keras.layers.BatchNormalization(),

    tf.keras.layers.Dense(112, activation='relu'),
    tf.keras.layers.BatchNormalization(),
    tf.keras.layers.Dropout(0.2),

    tf.keras.layers.Dense(64, activation='relu'),
    tf.keras.layers.BatchNormalization(),

    tf.keras.layers.Dense(688, activation='relu'),
    tf.keras.layers.BatchNormalization(),
    tf.keras.layers.Dropout(0.2),

    tf.keras.layers.Dense(720, activation='relu'),
    tf.keras.layers.BatchNormalization(),
    tf.keras.layers.Dropout(0.2),

    tf.keras.layers.Dense(400, activation='relu'),
    tf.keras.layers.BatchNormalization(),

    tf.keras.layers.Dense(304, activation='relu'),
    tf.keras.layers.BatchNormalization(),
    tf.keras.layers.Dropout(0.2),

    tf.keras.layers.Dense(32, activation='relu'),
    tf.keras.layers.BatchNormalization(),

    tf.keras.layers.Dropout(0.28),
    tf.keras.layers.Dense(1, activation='sigmoid')
```

The above snippet contains a 10 layered Fully Connected Network, for our FYP. The model has a training accuracy of 90.18% on 25 epochs, and a test accuracy of 90.72%

19. ML Flow Experiments:

Run details

Run ID:	c404ffc364e4410e8fab0d506eb65e96	5edca36ca01a45459ff173bffc3c4e6
Run Name:	gentle-fox-648	efficient-elk-707
Start Time:	2023-06-11 13:07:42	2023-06-11 12:16:57
End Time:	2023-06-11 13:08:07	2023-06-11 12:17:21
Duration:	24.9s	24.8s

Parameters

Show diff only

Loss	Binary Cross Entropy	Binary Cross Entropy
Optimizer	Adam	Adam
batch_size	64	128
learning_rate	0.23	0.011
num_epochs	5	5

The above snippet contains a detailed comparison of the model's trainable parameters for 2 distinct and successful runs. He simply tweaked the model's learning rate and batch size to see if it has any co-variation with the accuracy or not.

22. Further details:

Metrics

Show diff only

Accuracy	85.94	87.5
F1 score	0.892	0.892
Precision	0.874	0.874
Recall	0.92	0.92

Tags

Show diff only

No tags to display.

As we can see, the accuracy plummets by ~2%, when the batch size and learning rate is altered, which denotes that the ideal batch size should be 128 and the learning rate should remain unchanged.

