07-03-2024 **PROJECT DISCUSSION -2**

Discussion was mainly about the dataset and existing framework such as **flower, pysyft, Fate, tensorflow federated** and how this framework works and the advantages and disadvantages of this framework.

Key Points Discussed:

Advantages and Disadvantages of the Frameworks

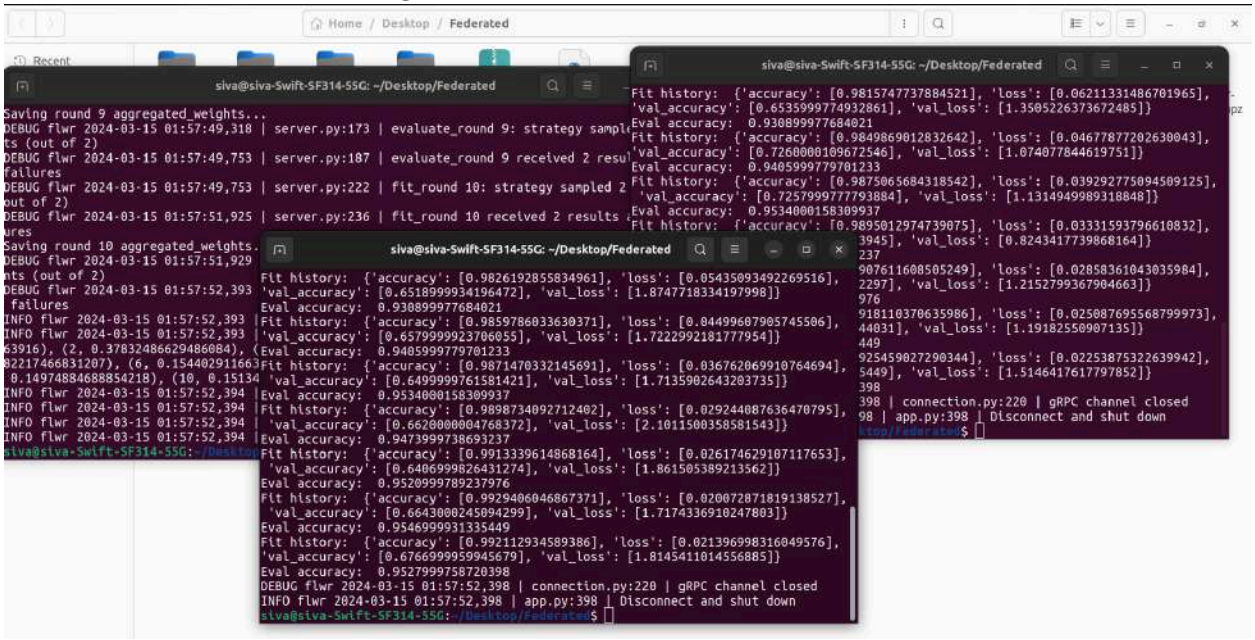
Out of all this frameworks Flower Frameworks have more advantage compared to other frameworks

Deciding The Dataset:

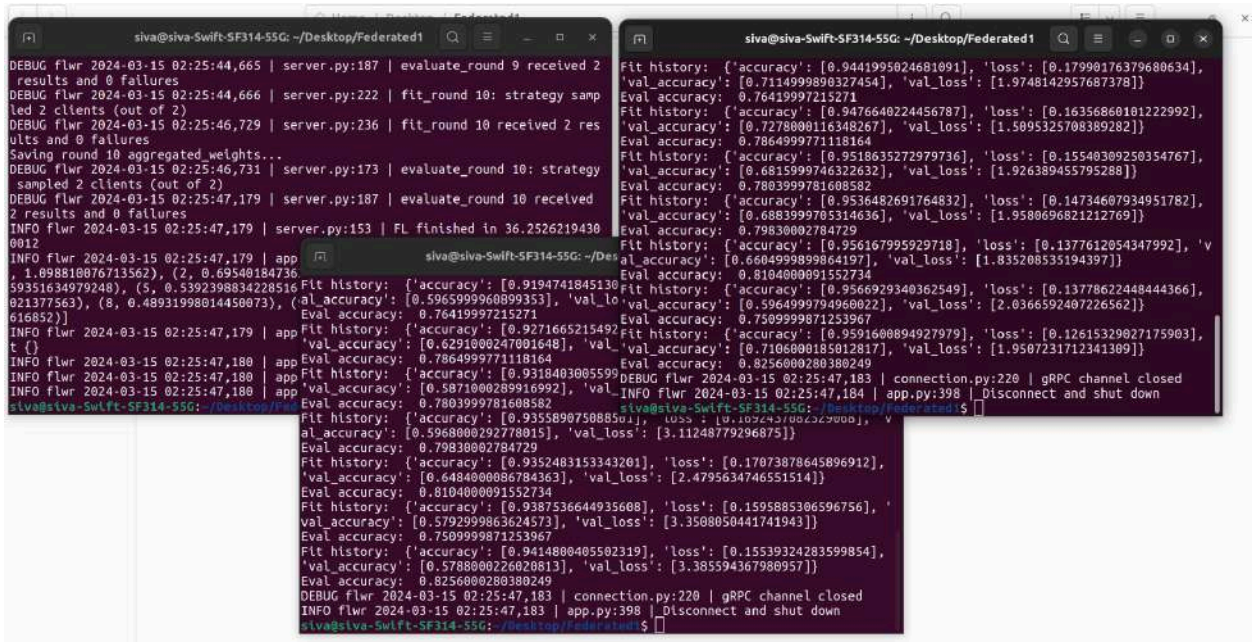
- Titanic Dataset(size: 994)
- Cancer Dataset(size:596)
- Mnist Dataset
- California House Prediction(Recommended by sir)

Showed Some Existing Model(based on class imbalance and horizontal federated learning)

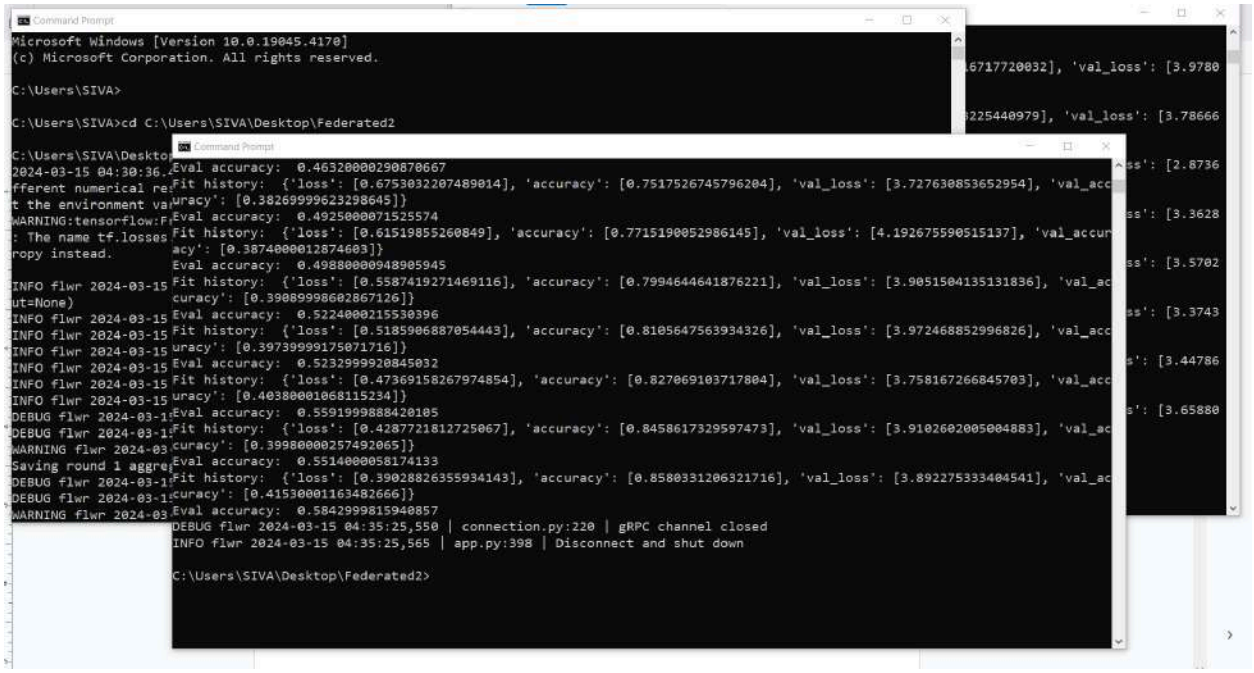
Horizontal Federated Learning - MNIST dataset(with different class distributions to each client)



Horizontal Federated Learning - MNIST fashion dataset(with different class distributions to each client)



Horizontal Federated Learning - CIFAR10 dataset (with different class distributions to each client)



BASIC MODEL OF FEDERATED LEARNING

Comparing Performance Of Centralized And Decentralized Model

In this model,there is no server,client method instead the dataset is partitioned into three datasets each with different features and the datasets are modeled using a logistic regression for each dataset and the predicted outputs are sent to the FNN model which finally predicts the label for the dataset and this predicted output are compared with the actual output for calculating the accuracy for each dataset and for each model(centralized model,decentralized model)

FOR TITANIC DATASET:

```
Running centralised training...
Train accuracy: 84.248%
Test accuracy: 82.022%
Running decentralised training...
Iteration 1, loss = 0.64399825
Iteration 2, loss = 0.56226789
Iteration 3, loss = 0.49493030
Iteration 4, loss = 0.45459877
Iteration 5, loss = 0.43634872
Iteration 6, loss = 0.43703215
Iteration 7, loss = 0.43900626
Iteration 8, loss = 0.42954290
Iteration 9, loss = 0.42684795
Iteration 10, loss = 0.42327625
Iteration 11, loss = 0.42247988
Iteration 12, loss = 0.42048271
Iteration 13, loss = 0.41931750
Iteration 14, loss = 0.41798792
Iteration 15, loss = 0.41693441
```

```
Client 0 test accuracy: 73.034%
Client 1 test accuracy: 81.461%
Client 2 test accuracy: 71.348%
Combined test accuracy: 80.337%
```

For Cancer Dataset :

(since size of the dataset is very small we getting error related to not sufficient data in train and test set)

CENTRALIZED MODEL:

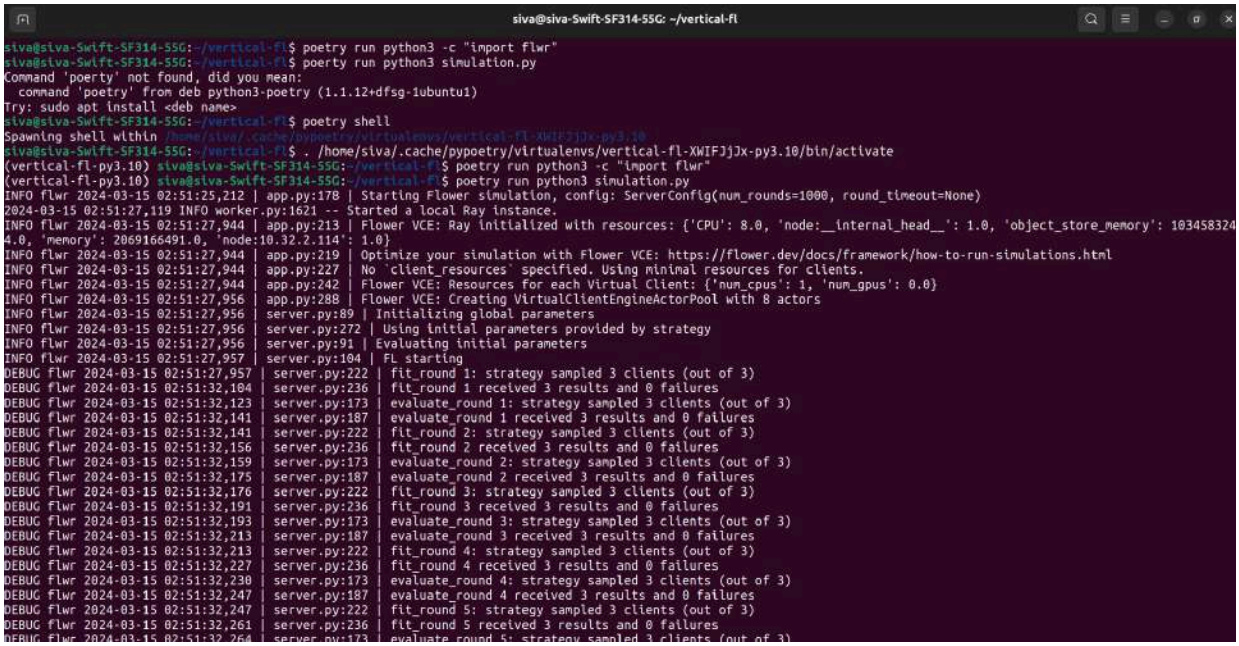
```
Running centralised training...
Train accuracy: 98.758%
Test accuracy: 95.349%
```

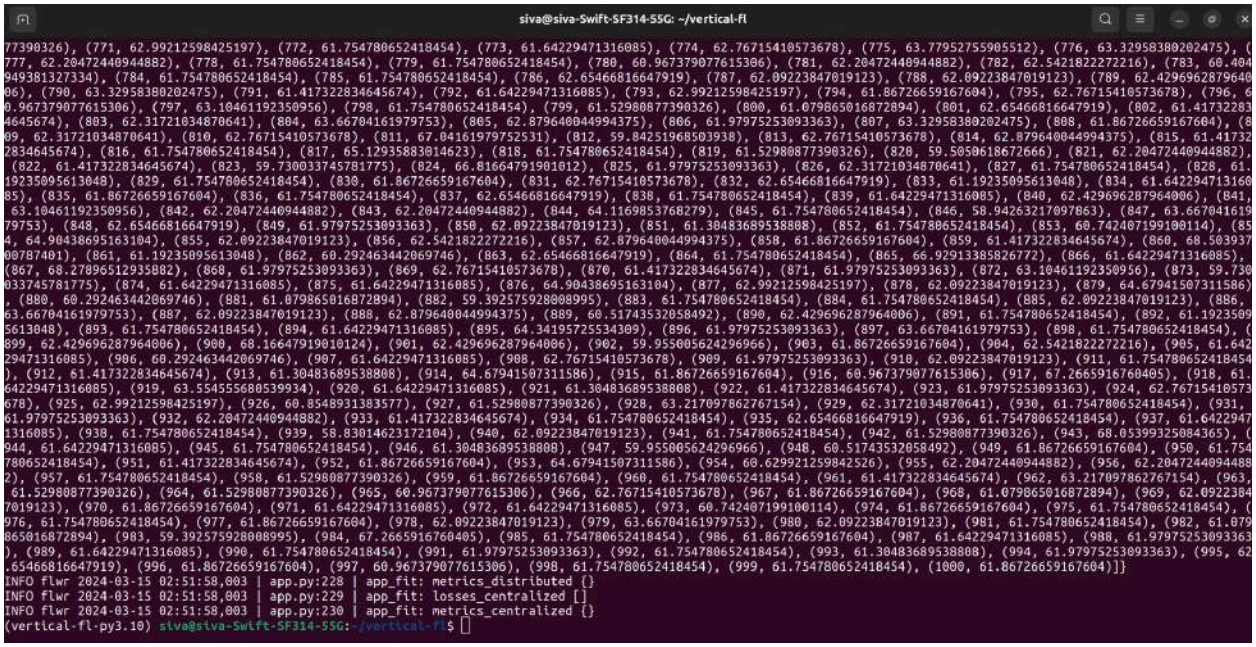
DECENTRALIZED MODEL:

```
Training loss did not improve more than tol=0.000100 for 10 consecutive epochs.
Stopping.
Client 0 test accuracy: 96.512%
Client 1 test accuracy: 94.186%
Client 2 test accuracy: 94.186%
Combined Test accuracy: 96.512%
```

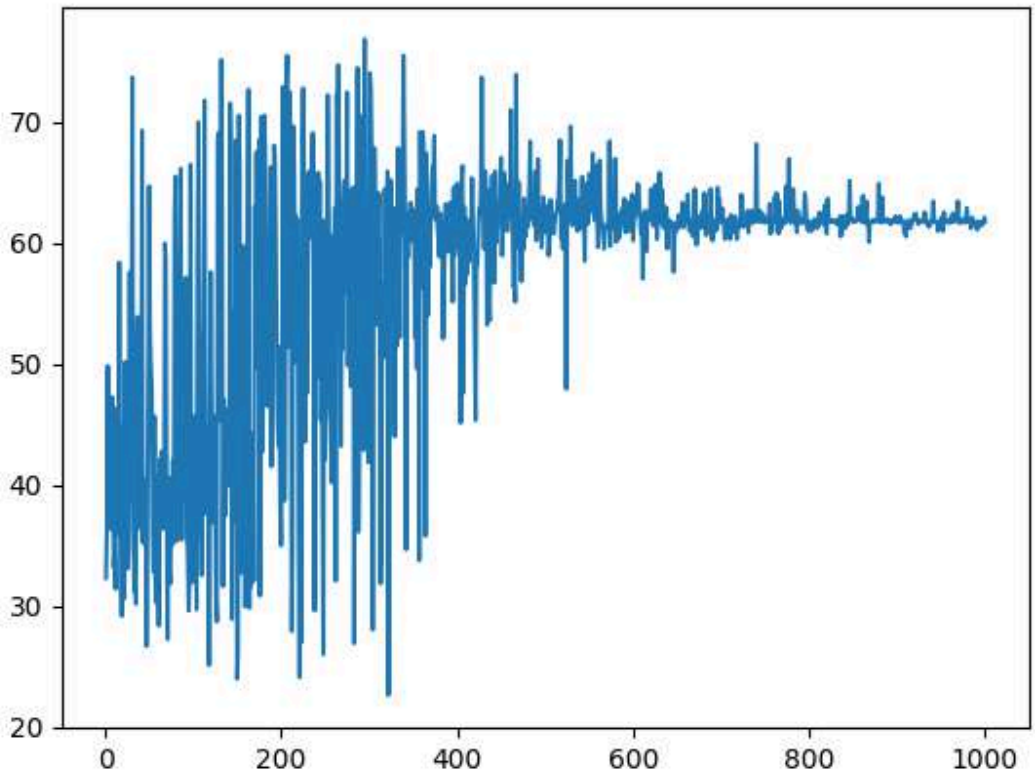
FLOWER FRAMEWORK

Vertical Federated Learning - Titanic data set (Features of the data set has been distributed to three clients)


siva@siva-Swift-SF314-55G: ~/vertical-fl
siva@siva-Swift-SF314-55G:~/vertical-fl$ poetry run python3 -c "import flwr"
siva@siva-Swift-SF314-55G:~/vertical-fl$ poetry run python3 simulation.py
Command 'poerty' not found, did you mean:
command 'poetry' from deb python3-poetry (1.1.12+dfsg-1ubuntu1)
Try: sudo apt install <deb name>
siva@siva-Swift-SF314-55G:~/vertical-fl$ poetry shell
Spawning shell within /home/siva/.cache/pypoetry/virtualenvs/vertical-fl-XWIFJjJx-py3.10
siva@siva-Swift-SF314-55G:~/vertical-fl$. /home/siva/.cache/pypoetry/virtualenvs/vertical-fl-XWIFJjJx-py3.10/bin/activate
(vertical-fl-py3.10) siva@siva-Swift-SF314-55G:~/vertical-fl$ poetry run python3 -c "import flwr"
(vertical-fl-py3.10) siva@siva-Swift-SF314-55G:~/vertical-fl$ poetry run python3 simulation.py
INFO flwr 2024-03-15 02:51:25,212 | app.py:178 | Starting Flower simulation, config: ServerConfig(num_rounds=1000, round_timeout=None)
2024-03-15 02:51:27,119 INFO worker.py:1621 -- Started a local Ray instance.
INFO flwr 2024-03-15 02:51:27,944 | app.py:213 | Flower VCE: Ray initialized with resources: {'CPU': 8.0, 'node:__internal_head__': 1.0, 'object_store_memory': 1034583244.0, 'memory': 2069166491.0, 'node:10.32.2.114': 1.0}
INFO flwr 2024-03-15 02:51:27,944 | app.py:219 | Optimize your simulation with Flower VCE: https://flower.dev/docs/framework/how-to-run-simulations.html
INFO flwr 2024-03-15 02:51:27,944 | app.py:227 | No `client_resources` specified. Using minimal resources for clients.
INFO flwr 2024-03-15 02:51:27,944 | app.py:242 | Flower VCE: Resources for each Virtual Client: {'num_cpus': 1, 'num_gpus': 0.0}
INFO flwr 2024-03-15 02:51:27,956 | app.py:288 | Flower VCE: Creating VirtualClientEngineActorPool with 8 actors
INFO flwr 2024-03-15 02:51:27,956 | server.py:89 | Initializing global parameters
INFO flwr 2024-03-15 02:51:27,956 | server.py:272 | Using initial parameters provided by strategy
INFO flwr 2024-03-15 02:51:27,956 | server.py:91 | Evaluating initial parameters
INFO flwr 2024-03-15 02:51:27,957 | server.py:104 | FL starting
DEBUG flwr 2024-03-15 02:51:27,957 | server.py:222 | fit_round 1: strategy sampled 3 clients (out of 3)
DEBUG flwr 2024-03-15 02:51:32,104 | server.py:236 | fit_round 1 received 3 results and 0 failures
DEBUG flwr 2024-03-15 02:51:32,123 | server.py:173 | evaluate_round 1: strategy sampled 3 clients (out of 3)
DEBUG flwr 2024-03-15 02:51:32,141 | server.py:187 | evaluate_round 1 received 3 results and 0 failures
DEBUG flwr 2024-03-15 02:51:32,141 | server.py:222 | fit_round 2: strategy sampled 3 clients (out of 3)
DEBUG flwr 2024-03-15 02:51:32,156 | server.py:236 | fit_round 2 received 3 results and 0 failures
DEBUG flwr 2024-03-15 02:51:32,159 | server.py:173 | evaluate_round 2: strategy sampled 3 clients (out of 3)
DEBUG flwr 2024-03-15 02:51:32,175 | server.py:187 | evaluate_round 2 received 3 results and 0 failures
DEBUG flwr 2024-03-15 02:51:32,176 | server.py:222 | fit_round 3: strategy sampled 3 clients (out of 3)
DEBUG flwr 2024-03-15 02:51:32,191 | server.py:236 | fit_round 3 received 3 results and 0 failures
DEBUG flwr 2024-03-15 02:51:32,193 | server.py:173 | evaluate_round 3: strategy sampled 3 clients (out of 3)
DEBUG flwr 2024-03-15 02:51:32,213 | server.py:187 | evaluate_round 3 received 3 results and 0 failures
DEBUG flwr 2024-03-15 02:51:32,213 | server.py:222 | fit_round 4: strategy sampled 3 clients (out of 3)
DEBUG flwr 2024-03-15 02:51:32,227 | server.py:236 | fit_round 4 received 3 results and 0 failures
DEBUG flwr 2024-03-15 02:51:32,230 | server.py:173 | evaluate_round 4: strategy sampled 3 clients (out of 3)
DEBUG flwr 2024-03-15 02:51:32,247 | server.py:187 | evaluate_round 4 received 3 results and 0 failures
DEBUG flwr 2024-03-15 02:51:32,247 | server.py:222 | fit_round 5: strategy sampled 3 clients (out of 3)
DEBUG flwr 2024-03-15 02:51:32,261 | server.py:236 | fit_round 5 received 3 results and 0 failures


siva@siva-Swift-SF314-55G: ~/vertical-fl
INFO flwr 2024-03-15 02:51:58,003 | app.py:228 | app_fit: metrics_distributed {}
INFO flwr 2024-03-15 02:51:58,003 | app.py:229 | app_fit: losses_centralized []
INFO flwr 2024-03-15 02:51:58,003 | app.py:230 | app_fit: metrics_centralized {}
(vertical-fl-py3.10) siva@siva-Swift-SF314-55G:~/vertical-fl$

Accuracy plot vs Epochs(Vertical FL)



References

1. "Communication-Efficient Learning of Deep Networks from Decentralized Data"

   H. Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, Blaise Agüera y Arcas
   Link: https://arxiv.org/abs/1602.05629

2. "A Comparative Evaluation of FedAvg and Per-FedAvg Algorithms for Dirichlet Distributed Heterogeneous Data"

   Hamza Reguieg1 , Mohammed El Hanjri2 , Mohamed El Kamili1 , Abdellatif Kobbane2
   Link: https://arxiv.org/abs/2309.01275

3. "Heterogeneous Federated Learning: State-of-the-art and Research Challenges"
MANG YE, XIUWEN FANG, and BO DU, PONG C. YUEN, ,DACHENG TAO

   Link: https://dl.acm.org/doi/10.1145/3625558