# **ONT | Task-6**

## **Federated Learning**

## **Introduction**

Federated Learning (FL) is a machine learning paradigm where models are trained across multiple decentralized devices or servers holding local data samples, without exchanging them. It is designed to protect user privacy by keeping data localized and aggregating model updates centrally. Key techniques include:

* **Federated Averaging:** Aggregates model weights from multiple clients to update a global model.
* **Client-Server Architecture:** Clients train models on local data and send updates to a central server, which aggregates and updates the global model.
* **Privacy Preservation:** Techniques such as differential privacy and secure multi-party computation are used to ensure data privacy.

## **Objectives**

The goal of this task was to implement Federated Learning to train models on datasets from the EU and USA networks separately. The effectiveness of Federated Learning was assessed by comparing it with centralized learning approaches.

## **Data Preparation**

**Dataset Description**

* **EU Network Dataset:** Data representing network traffic from EU-based sources.
* **USA Network Dataset:** Data representing network traffic from USA-based sources.

**Data Preprocessing**

* **Handling Missing Values:** Imputation or removal of missing values.
* **Normalization:** Scaling features to a standard range to improve model performance.

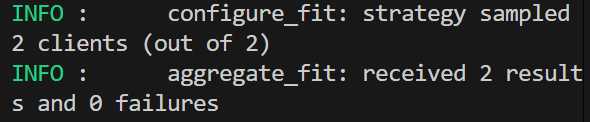
## **Federated Learning Implementation**

**Framework Selection**

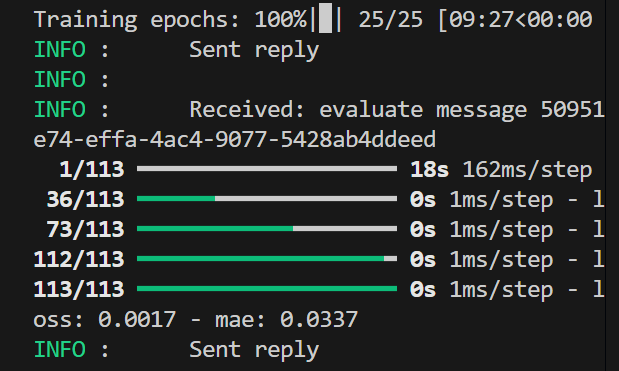
* **Framework:** Flower (FLWR) was selected for its ease of use and flexibility in Federated Learning setups.

**Server and Client Configuration**

* **Server:** Aggregated model updates and managed communication with clients.



* **Clients:** Trained models on local datasets and communicated updates to the server.



**EU Network Dataset**

* **Local Subsets:** The EU dataset was split into subsets for different clients.
* **Model Training:** Federated Learning models were trained using these subsets.
* **Evaluation:** The performance of the aggregated model was assessed using a validation set.

**USA Network Dataset**

* **Local Subsets:** Similar to the EU dataset, the USA dataset was split into local subsets.
* **Model Training:** Federated Learning models were trained on these subsets.
* **Evaluation:** The aggregated model’s performance was evaluated on a validation set.

## **Model Training and Evaluation**

**Federated Learning Model Training**

* **Training Process:** Models were trained across multiple rounds of Federated Learning with updates aggregated each round.
* **Metrics:** Models were evaluated using accuracy, precision, recall, and F1-score.

**Centralized Model Training**

* **Comparison:** A centralized model was trained on the entire dataset for comparison with the Federated Learning models.

## **Comparison and Analysis**

**Performance Comparison**

* **Federated Learning vs. Centralized Learning:** Performance metrics such as accuracy, precision, recall, and F1-score were compared between Federated Learning and centralized approaches.

**Benefits and Challenges**

* **Benefits:** Data privacy, decentralized training, and reduced data transfer.
* **Challenges:** Communication overhead, potential heterogeneity in client data, and slower convergence.

**Visualizations**

* **Training and Validation Loss Curves:** Plots showing the loss over training epochs.
* **Accuracy Plots:** Visual representation of model accuracy over time.

**Results Summary**

* **Server Logs:** The server successfully initiated Federated Learning rounds, aggregated model updates, and evaluated the global model. Performance metrics showed improvement over the rounds.
* **Client Logs:** Clients trained models locally and sent updates to the server. The models were trained and evaluated successfully on both EU and USA datasets.

## **Conclusion**

The Federated Learning setup demonstrated the ability to collaboratively train models across different datasets while preserving data privacy. The comparison with centralized models highlighted the trade-offs between privacy and performance.