Politecnico di Torino

FEDERATED LEARNING: EMPIRICAL STUDY WITH *CIFAR-100* AND *SHAKESPEARE*

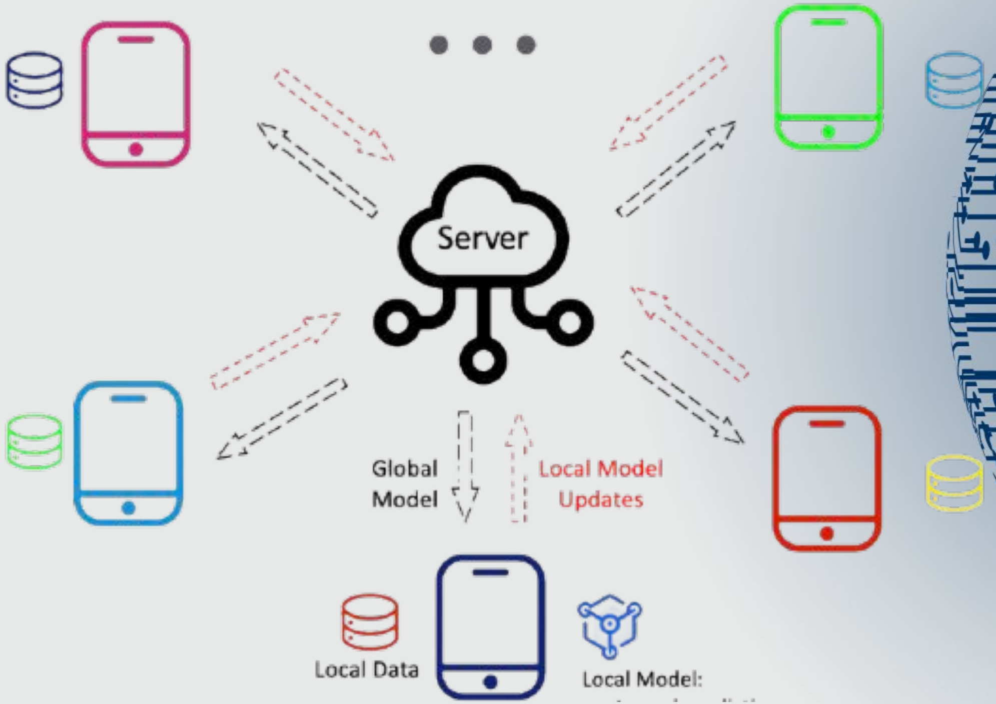
Analyzing Communication, Data Heterogeneity and their effect on Training

Fiata Rosamaria: s329502
Galtieri Chiara: s331345
Taormina Nicolò: s331853

Advanced Machine Learning - February 10, 2025

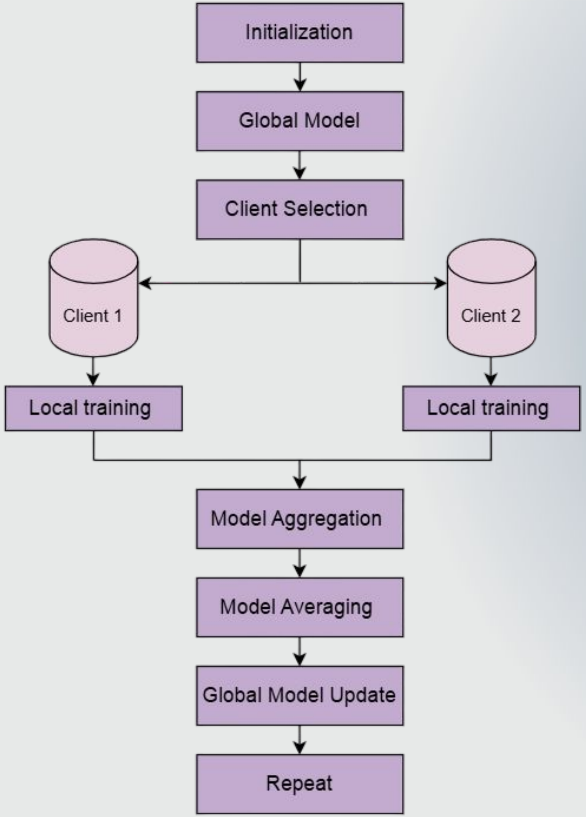
INTRODUCTION

Federated Learning (FL) enables multiple clients to train a model keeping their data local



BACKGROUND

FEDERATED AVERAGING ALGORITHM (FedAvg)



https://www.educative.io/answers/what-is-federated-averaging-fedavg

CHALLENGES IN FL

- ***STATISTICAL HETEROGENEITY***

  Non-IID data leads to biased models.

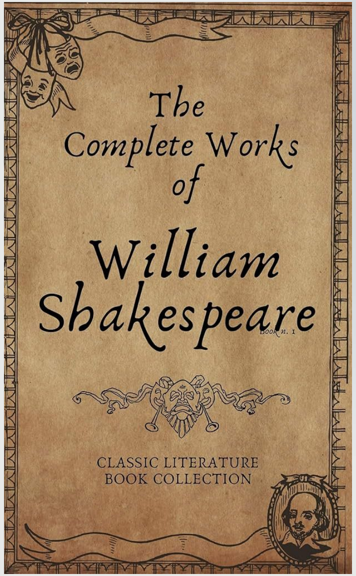
- ***CLIENT PARTICIPATION***

  Some clients may have more data than others.

OUR EXPERIMENTS DATASET

CIFAR-100

SHAKESPEARE from LEAF benchmark suite



DATA PARTITIONING STRATEGIES

- *IID PARTITIONING*

  Each client has a similar distribution.

- *NON-IID PARTITIONING*

  Clients received data from a limited subset of classes (Nc)

CLIENT PARTITIONING STRATEGIES

- *UNIFORM PARTICIPATION*

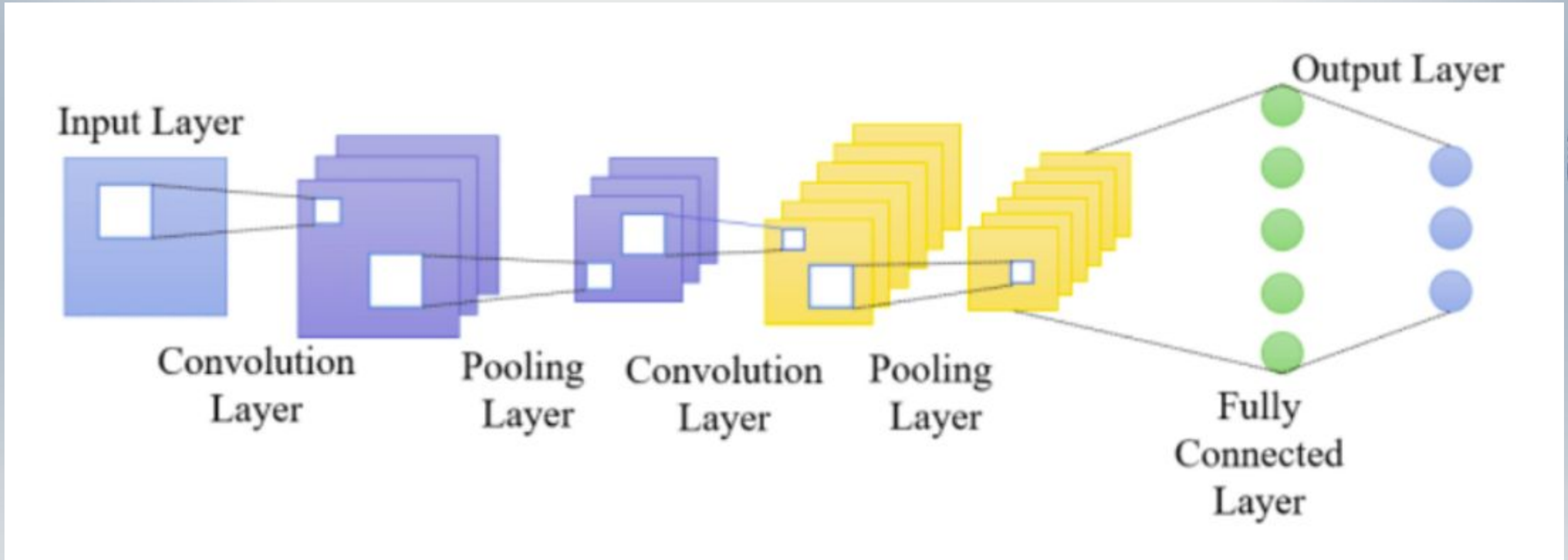
  Equal probability of being selected.

- *SKEWED PARTICIPATION*

  Clients are sampled in each round using a Dirichlet distribution.

MODEL ARCHITECTURES

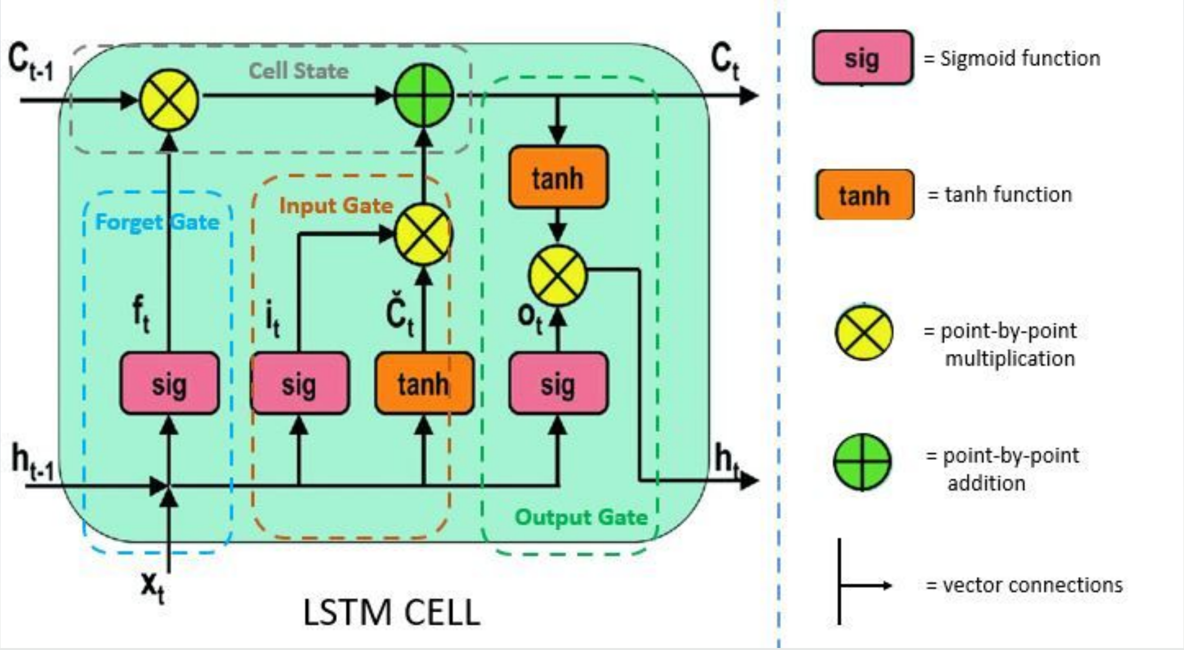
CIFAR-100's model: LeNet-5



https://www.geeksforgeeks.org/lenet-5-architecture/

MODEL ARCHITECTURES

SHAKESPEARE's model: LSTM



https://www.pluralsight.com/resources/blog/guides/introduction-to-lstm-units-in-rnn

CENTRALIZED TRAINING BASELINE

- Reference point: Training with centralized data.

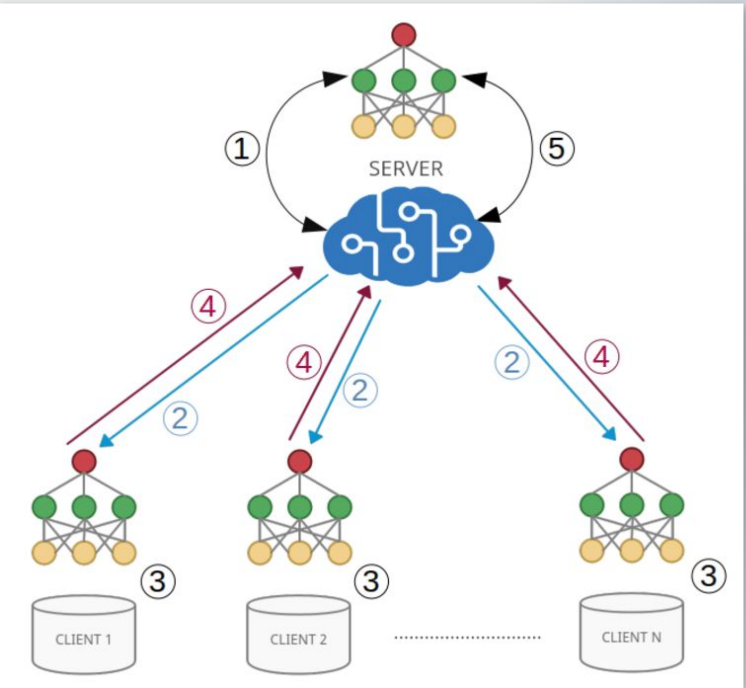
Dataset	Accuracy (%)	Loss
CIFAR-100	49.24	2.12
Shakespeare	55.26	2.15

Table 2. Result Centralized Training

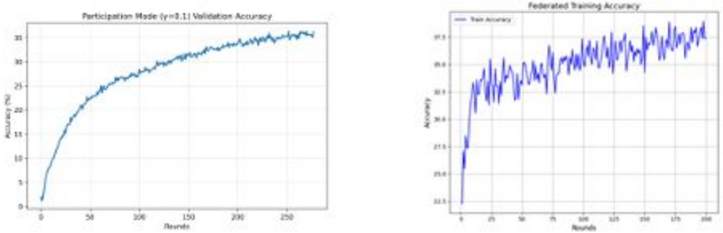
- Hyperparameter tuning for optimal performance.

FEDERATED LEARNING EXPERIMENTS

- *FedAvg* algorithm for FL.
- 100 clients, 10% participate per round.
- Local training steps *J = {4, 8, 16}*.
- *Uniform vs. Skewed* participation.



IMPACT OF CLIENT PARTICIPATION



(a) CIFAR-100 Skewed (b) Shakespeare Skewed

Figure 2. Skewed Partecipation (γ=0.1)

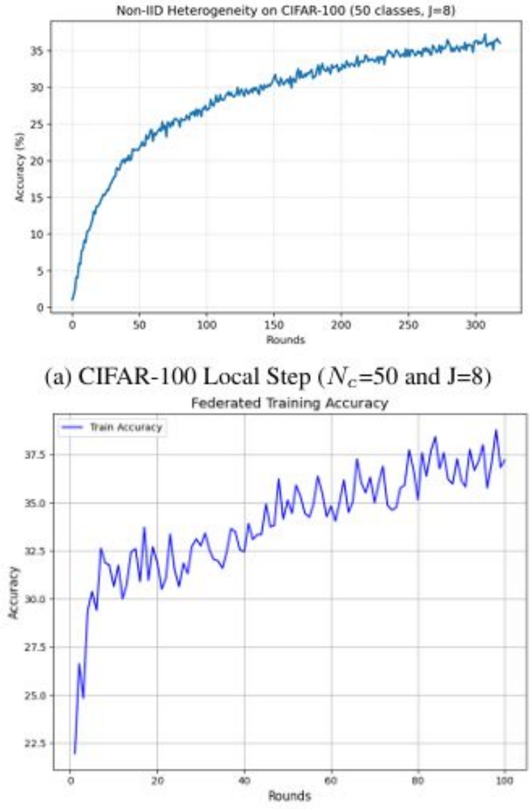
Dataset	Skewness (γ)	Accuracy (%)	Loss
CIFAR-100	0.1	37.55	2.82
	0.5	37.49	2.75
	1.0	39.71	2.76
Shakespeare	0.1	38.58	2.99
	0.5	39.05	2.97
	1.0	40.51	2.76

Table 3. Impact of Client Participation Strategies (Uniform vs Skewed

- *Uniform* participation → *Balanced* updates.
- *Skewed* participation → *Faster* convergence but *biased* updates.

IMPACT OF LOCAL TRAINING STEPS

- More local steps → *Fewer communication rounds.*
- *J = 8* is optimal: improves accuracy, avoids overfitting.
- *J = 16* → Risk of *client drift.*

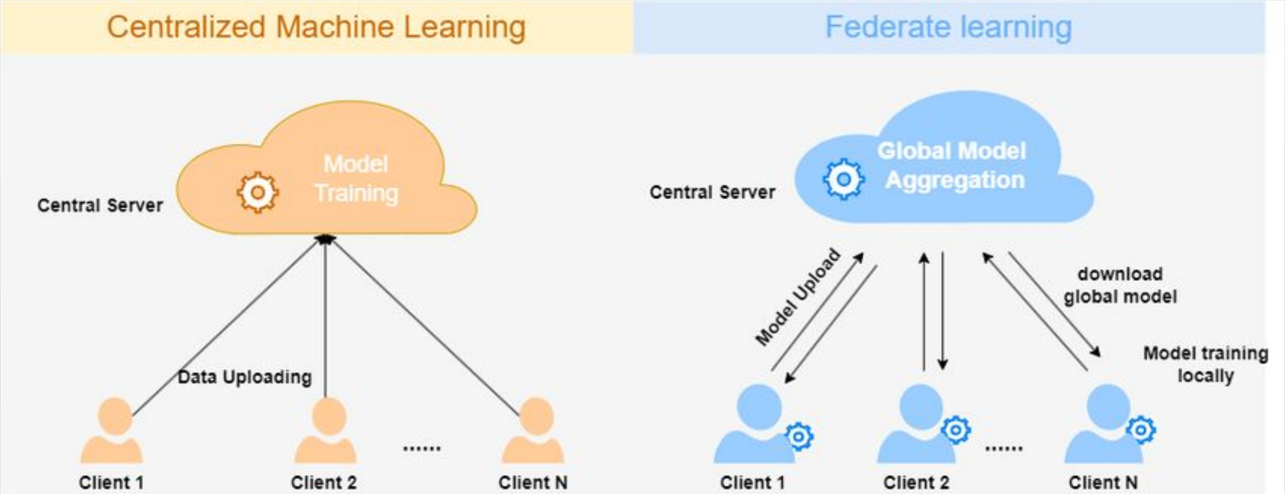


(a) CIFAR-100 Local Step (N_c=50 and J=8)

(b) Shakespeare Local Step (N_c=50 and J=8)

Figure 3. Local Training Step (N_c=50 and J=8)

KEY FINDINGS



Best trade-off in FL :

→ *Uniform participation* (more stable updates), *J = 8* (balances local training & communication), and proper *client selection.*

Personal Contribution: *Two-Phase* Federated Learning

- *Literature*: Modifications to the algorithm to mitigate client drift.
- *Our Approach*: Modifications to the training pipeline, introducing an intermediate model exchange before aggregation.
- Goal: Improve generalization and mitigate client drift.

Two-Phase Training Implementation

- *Phase 1*: Local training (J/2 epochs).
- Model *shuffling*: Clients receive different models.
- *Phase 2*: Additional local training (J/2 epochs).
- Final *aggregation*: Standard Fed Avg.

J: number of local epochs
J/2: half of J relative to the baseline used for each experiment

Two-Phase Training Performance

J	Standard (%)	Two-Phase (%)	Difference
4	38.25	**39.16**	+0.91
8	38.15	37.48	-0.67
16	38.85	**40.23**	+1.38

Table 5. Two-Phase Training Results (IID Setting)

N_c	J	Standard (%)	Two-Phase (%)	Difference
1	4	38.93	**40.20**	+1.27
1	8	39.62	38.40	-1.22
1	16	38.23	36.88	-1.35
5	4	37.48	**38.47**	+0.99
5	8	39.17	**40.17**	+1.00
5	16	40.01	**40.44**	+0.43
10	4	37.80	36.15	-1.65
10	8	39.85	38.29	-1.56
10	16	39.56	**40.83**	+1.27
50	4	37.48	**40.56**	+3.08
50	8	38.80	**39.07**	+0.27
50	16	39.28	**39.74**	+0.46

Table 6. Two-Phase Training Results (Non-IID Setting)

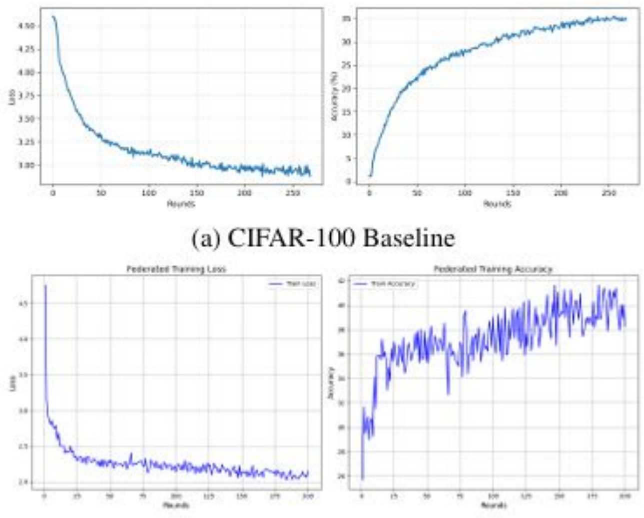
- *Improvement in accuracy*: Up to +3.08% improvement in heterogeneous settings.

Limitations

- Increased Communication Overhead: Two-phase training results in higher communication costs.
- Privacy Concerns: The current approach makes clients more vulnerable to various attack strategies (e.g., white-box Membership Inference Attack).

Key Takeaways

- Non-IID data negatively impacts FL models.
- Optimal local steps (J) balance accuracy & communication.
- Two-phase training mitigates client drift but increases overhead.



(a) CIFAR-100 Baseline

(b) Shakespeare Baseline

Figure 1. Baseline performance

THANK YOU FOR YOUR ATTENTION!

Fiata Rosamaria: s329502
Galtieri Chiara: s331345
Taormina Nicolò: s331853