PRIVACY-PRESERVING USING FEDERATED LEARNING

[1]Ayan khan, [2] Ritesh Gholpe, [3] Harshal Bhute
[4]Mayur Durugkar, [5]Ashutosh Welekar
Under the Guidance of
Prof. Supriya Jawale

Abstract : In recent years, data privacy has become a major concern in machine learning systems due to the centralized collection of sensitive data. Traditional approaches require transferring user data to a central server, which increases the risk of data leakage and security breaches. To address these challenges, Federated Learning has emerged as an effective privacy-preserving technique.This paper presents a system that implements Federated Learning using a client-server architecture, where multiple clients train a machine learning model locally without sharing raw data. The central server aggregates model updates using the Federated Averaging (FedAvg) algorithm. The proposed system ensures data privacy while maintaining model performance.The implementation is carried out using Python and the Flower framework. Experimental results demonstrate that the system achieves efficient model training while preserving user data privacy.

IndexTerms **- Federated Learning, Privacy Preservation, Machine Learning, FedAvg, Distributed Systems**

1.INTRODUCTION

With the rapid growth of digital technologies, large volumes of data are generated daily from various sources such as mobile devices, healthcare systems, and financial platforms. Machine learning models rely heavily on this data to provide accurate predictions and insights.
Traditional machine learning approaches follow a centralized architecture where data is collected and stored on a central server. Although effective, this approach raises serious concerns regarding data privacy, security, and ownership. Sensitive information such as personal data, medical records, and financial transactions may be exposed during data transmission or storage.
To overcome these challenges, Federated Learning has been introduced as a decentralized approach to machine learning. This concept was proposed by H. Brendan McMahan at Google. It allows multiple clients to collaboratively train a model while keeping data locally on their devices.
In this paper, we implement a privacy-preserving machine learning system using Federated Learning, where clients train models locally and a central server aggregates updates.

2.LITERATURE REVIEW

Several studies have explored privacy-preserving machine learning techniques. Traditional centralized learning methods expose sensitive data, making them vulnerable to attacks.
Federated Learning, introduced by H. Brendan McMahan, enables decentralized model training without sharing raw data. Researchers have further enhanced this approach by integrating techniques such as differential privacy and secure aggregation.
Recent works focus on improving communication efficiency and handling non-IID data distribution across clients. However, challenges such as system heterogeneity and communication overhead still exist.
This study builds upon these existing approaches and implements a practical Federated Learning system using the FedAvg algorithm.

3. METHODOLOGY

The proposed system follows a client-server architecture for implementing Federated Learning.

3.1 System Architecture

- Multiple clients with local datasets
- Central server for aggregation
- Communication through model updates

3.2 Working Process

1. Server initializes global model
2. Model is sent to clients
3. Clients train locally
4. Clients send updated weights
5. Server aggregates using FedAvg
6. Updated model is redistributed

3.3 Algorithm Used (FedAvg)

$$w = \sum_{k=1}^{n} \frac{n_k}{n} w_k$$

Where:

- w_k= weights from client
- n_k= data size of client

4. IMPLEMENTATION

The system is implemented using:

- Python
- Flower (flwr) framework
- Scikit-learn
- NumPy, Pandas

Key Components:

- Client-side training module
- Server aggregation module
- Communication interface

Each client trains a local model and sends updates to the server, which performs aggregation.

5. RESULTS AND DISCUSSION

The system was evaluated based on:

- Model accuracy
- Loss reduction
- Communication rounds

Observations:

- Accuracy improves with each round
- Data privacy is preserved
- Communication overhead exists but manageable

The results show that Federated Learning provides a balance between performance and privacy.

6. APPLICATIONS

The proposed system can be applied in:

- Healthcare (secure patient data)
- Finance (fraud detection)
- Mobile applications (personalized services)
- IoT systems

7. CONCLUSION

This paper presents a privacy-preserving machine learning system using Federated Learning. The approach ensures that sensitive data remains on local devices while enabling collaborative model training.

The implementation demonstrates that Federated Learning is an effective solution for secure and distributed machine learning systems.

8. FUTURE SCOPE

Future improvements include:

- Integration of differential privacy
- Use of deep learning models
- Real-world deployment
- Secure aggregation techniques

9.REFERENCES

1. McMahan et al., “Communication-Efficient Learning of Deep Networks from Decentralized Data”
2. Research papers on Federated Learning
3. Flower Framework Documentation
4. Scikit-learn Documentation