**ML Internship Task 6**

**Objective:** The goal of this task is to explore and implement Federated Learning (FL) techniques to train models on datasets from the EU and USA networks separately. You will apply FL to train models collaboratively across multiple local datasets without centralizing the data. The aim is to assess the effectiveness of FL in handling data from different regions and compare its performance with centralized learning approaches.

**Datasets:**

* EU network dataset
* USA network dataset

**Requirements:**

1. **Understanding Federated Learning:**
   * Research and read about Federated Learning, focusing on its principles, benefits, and common techniques. Look for reliable sources such as academic papers, textbooks, and reputable online articles.
   * Understand concepts such as Federated Averaging, client-server architecture, and privacy preservation in FL.
2. **Data Preparation:**
   * Use the EU and USA datasets provided.
   * Preprocess the datasets as necessary (e.g., handling missing values, normalization).
3. **Federated Learning Implementation:**
   * **Setup Federated Learning:**
     + Implement a Federated Learning framework suitable for your task (e.g., TensorFlow Federated, PySyft).
     + Configure a server to aggregate model updates and multiple clients to train on local data.
   * **EU Network Dataset:**
     + Split the EU dataset into several local subsets representing different clients.
     + Train a Federated Learning model using these local subsets.
     + Evaluate the aggregated model’s performance on a validation set from the EU dataset.
   * **USA Network Dataset:**
     + Similarly, split the USA dataset into local subsets.
     + Train a Federated Learning model on these subsets.
     + Evaluate the aggregated model’s performance on a validation set from the USA dataset.
4. **Model Training and Evaluation:**
   * Train Federated Learning models on both the EU and USA datasets separately.
   * Compare the performance of the FL models with a centralized model trained on the entire dataset (if applicable).
   * Use appropriate evaluation metrics (accuracy, precision, recall, F1-score) for model performance assessment.
5. **Comparison and Analysis:**
   * Compare the performance of Federated Learning models with the performance of centralized models in terms of accuracy, precision, recall, and F1-score.
   * Analyze the benefits and challenges of using Federated Learning compared to centralized learning approaches.
   * Provide visualizations (e.g., training/validation loss curves, accuracy plots) to support your analysis.
6. **Documentation:**
   * Provide a detailed report documenting the Federated Learning implementation, evaluation results, and performance comparisons.
   * Include code snippets, relevant plots, and a discussion of the Federated Learning strategy used.
   * Provide insights and recommendations based on your findings.

**Deliverables:**

1. **Results Report:** Include performance comparisons, visualizations, and a detailed discussion of Federated Learning versus centralized learning.
2. **Code:** Well-documented and organized codebase for Federated Learning implementation.
3. **Final Model:** The trained models saved in a format that can be loaded and used for future predictions.

**Deadline to Report:** [31st July 2024]