

# Semaine 1 : Fondamentaux de l'apprentissage fédéré et de l'IA en santé

## Lundi : Introduction au Federated Learning (FL)

**Objectif :** Comprendre les bases du Federated Learning (FL) et ses principes de fonctionnement.

- Concepts clés à comprendre :
  - Définition du Federated Learning.
  - Différences entre apprentissage centralisé et décentralisé.
  - Avantages du FL, notamment pour la confidentialité des données.
- Pratique :
  - Mise en place de *Flower* ou *FedML*, bibliothèques populaires pour le FL.
  - Installation de Python et configuration de l'environnement de travail avec PyTorch ou TensorFlow.
  - Création de ton premier modèle FL de base.

## Jeudi : Les enjeux médicaux de l'IA en santé

**Objectif :** Comprendre les enjeux spécifiques de l'IA et du FL en santé.

- Concepts clés à comprendre :
  - Le rôle du FL dans la protection des données personnelles en santé.
  - Les défis spécifiques de l'IA en médecine : qualité des données, diversité des sources, réglementation.
- Pratique :

- Examen des défis de la confidentialité et de la sécurité des données dans les applications médicales (ex : RGPD, HIPAA).
- Simulation de jeux de données médicaux synthétiques pour tests.

## **Samedi : Premiers pas avec les outils FL**

**Objectif :** Approfondir l'utilisation des outils FL pour l'apprentissage automatique.

- Pratique :
    - Installation complète de Flower ou FedML dans un environnement Python.
    - Implémentation d'un premier exemple simple en FL (par exemple, entraînement d'un modèle de régression avec données synthétiques).
- 

## **Semaine 2 : Données médicales et confidentialité**

### **Lundi : Types de données médicales et leur traitement**

**Objectif :** Appréhender les types de données utilisées en médecine et leur gestion.

- Concepts clés à comprendre :
  - Types de données : EHR (Electronic Health Records), DICOM (images médicales), notes cliniques.
  - Problématiques associées : données sensibles, accessibilité, et qualité des données.
- Pratique :
  - Découverte des jeux de données MIMIC-III, MedMNIST et autres.
  - Télécharger et prétraiter les jeux de données médicaux pour l'entraînement des modèles.

### **Jeudi : Confidentialité et sécurité dans le FL en santé**

**Objectif :** Se familiariser avec les concepts de confidentialité et de sécurité des données.

- Lecture :

- Introduction à la confidentialité différentielle et l'agrégation sécurisée dans le contexte du FL.
- Concepts clés à comprendre :
  - Confidentialité différentielle : un mécanisme pour garantir que les informations privées ne peuvent pas être extraites des modèles.
  - Agrégation sécurisée : garantir la sécurité des données lors de l'agrégation des résultats des modèles locaux.
- Pratique :
  - Implémentation d'un modèle simple avec confidentialité différentielle.
  - Utilisation de mécanismes d'agrégation sécurisée dans un environnement FL.

## **Samedi : Prétraitement des données sensibles**

**Objectif :** Préparer les données médicales pour l'apprentissage fédéré.

- Pratique :
    - Prétraitement des jeux de données (nettoyage, gestion des valeurs manquantes, normalisation des données).
    - Application de la confidentialité différentielle dans le prétraitement des données.
- 

## **Semaine 3 : Prédiction avec données tabulaires fédérées (EHR)**

### **Lundi : Introduction aux données tabulaires et aux modèles de prédiction**

**Objectif :** Appliquer le FL à la prédiction de maladies à partir de données structurées (EHR).

- Lecture :
  - Études de cas sur la prédiction de maladies (diabète, sepsis, etc.) à partir de données médicales structurées.

- Concepts clés à comprendre :
  - Structure des données EHR et leur utilisation pour la prédiction de maladies.
  - Méthodes d'analyse des données tabulaires.
- Pratique :
  - Exploration du jeu de données MIMIC-III et sélection des caractéristiques importantes pour la prédiction.

## **Jeudi : Implémentation de modèles de prédiction avec FL**

**Objectif :** Construire un modèle de prédiction à l'aide de Federated Learning.

- Pratique :
  - Implémentation d'un modèle de régression logistique ou d'un réseau de neurones multicouche (MLP) avec Flower ou TensorFlow Federated.
  - Entraînement du modèle sur des données décentralisées (simulation).
- Intégration de techniques de confidentialité, comme la confidentialité différentielle pour sécuriser les données.

## **Samedi : Evaluation du modèle et analyse des résultats**

**Objectif :** Evaluer la performance du modèle de prédiction en FL.

- Pratique :
  - Evaluation du modèle à l'aide de métriques comme la précision, la perte (loss) et l'aire sous la courbe ROC.
  - Visualisation des résultats de l'entraînement : courbes de convergence, taux de précision, etc.

---

# **Semaine 4 : NLP médical et apprentissage fédéré**

**Lundi : Introduction au traitement du langage naturel (NLP) médical**

**Objectif :** Apprendre à utiliser le Federated Learning pour traiter des données textuelles médicales.

- Lecture :
  - Introduction aux modèles de NLP spécialisés en médecine, comme BioBERT et ClinicalBERT.
- Concepts clés à comprendre :
  - NLP médical et applications : extraction d'entités, classification de pathologies, résumé automatique.
  - Modèles de langage pré-entraînés pour les données médicales.
- Pratique :
  - Télécharger et prétraiter un jeu de données de notes cliniques (ex. MIMIC-III, i2b2).
  - Introduction à la tokenisation et préparation des données pour le fine-tuning d'un modèle.

## **Jeudi : Fine-tuning de ClinicalBERT en FL**

**Objectif :** Appliquer le Federated Learning pour affiner un modèle pré-entraîné sur des données médicales décentralisées.

- Pratique :
  - Fine-tuning de ClinicalBERT avec les données décentralisées sur un environnement FL.
  - Entraînement fédéré pour des tâches comme la classification de pathologies et l'extraction d'entités nommées (NER).

## **Samedi : Analyse des performances du modèle NLP fédéré**

**Objectif :** Evaluer l'efficacité des modèles de NLP dans un cadre fédéré.

- Pratique :
  - Evaluation du modèle de NLP à l'aide de métriques comme l'exactitude, la précision, et le rappel.

- Comparaison entre le modèle centralisé et fédéré.
- 

## **Semaine 5 : Vision par ordinateur en santé et Federated Learning**

### **Lundi : Introduction à la vision par ordinateur en santé**

**Objectif :** Comprendre l'application du Federated Learning à des images médicales.

- Lecture :
  - Résumé des articles de recherche sur le FL appliqué à la radiologie et aux images médicales (ex. ChestX-ray14).
- Concepts clés à comprendre :
  - Problématiques de la vision par ordinateur en santé : prétraitement des images, augmentation des données, etc.
- Pratique :
  - Exploration des jeux de données d'images médicales (NIH Chest X-ray, MedMNIST).

### **Jeudi : Implémentation d'un CNN fédéré pour la classification d'images médicales**

**Objectif :** Appliquer un modèle de réseau de neurones convolutif (CNN) pour classifier des images médicales en FL.

- Pratique :
  - Implémentation d'un modèle CNN (par exemple ResNet) dans un cadre fédéré avec Flower.
  - Entraînement sur un jeu de données d'images médicales (ex. MedMNIST ou ChestX-ray).

### **Samedi : Comparaison des résultats FL et centralisé**

**Objectif :** Comparer la performance d'un modèle de vision par ordinateur en apprentissage fédéré et centralisé.

- Pratique :
    - Evaluation des résultats des modèles en termes de précision, de perte et de vitesse d'apprentissage.
    - Visualisation des performances dans un environnement FL par rapport à un modèle centralisé.
- 

## **Semaine 6 à 8 : Sécurité, Robustesse, Projet complet et contributions open-source**

### **Lundi et Jeudi : Sécurité et robustesse du Federated Learning**

**Objectif :** Maîtriser les concepts de sécurité, d'attaques et de défenses dans un environnement FL.

### **Samedi : Projet complet – Pipeline FL santé**

**Objectif :** Construire un projet de bout en bout appliquant FL à un cas médical concret (prédiction, NLP, imagerie).

---