

Federated learning

Henrique MARQUES REIS

Travail de Bachelor 2023

Filière Informatique et systèmes de communication - Orientation Ingénierie des données

Professeur: Stefano CARRINO

Expert: Eliane MAALOUF

Description

Dans le cadre du projet travail de Bachelor, ce projet consiste à exploiter une infrastructure scalable (créée lors du P3) capable de supporter du federated learning.

Le federated learning est un paradigme d'entraînement d'intelligence artificielle visant à entraîner les modèles dans les machines des utilisateurs. Ceci permet l'accès à des données de l'utilisateur sans que celui-ci n'ait besoin de les partager.

On peut également utiliser ce paradigme pour paralléliser l'entraînement d'une IA.

Le but de ce projet est d'effectuer, dans un premier temps, un travail de recherche sur les différents algorithmes de federated learning existants et de les tester, puis d'exploiter au moins un de ces algorithmes dans un cas réel.

Ce travail consiste à étudier, tester et évaluer plusieurs algorithmes de federated learning et de les comparer.

Déroulement

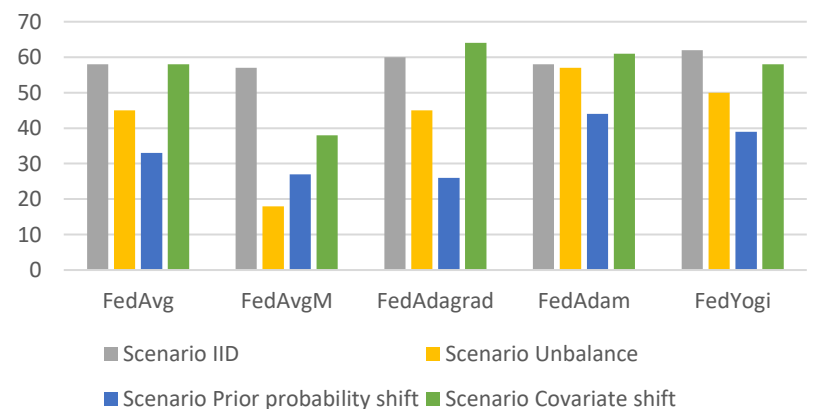
Différentes phases du projet:

- État de l'art : cette phase consiste à étudier tout ce qui existe sur le sujet du travail
- Analyse : suite à l'état de l'art, l'analyse permet de voir quel choix faire pour atteindre les objectifs fixés
- Réalisations sur jeux de données «simples» : avant de passer à un cas réel et plus compliqué, le federated learning a été implémenté sur un cas simple pour comprendre son fonctionnement
- Réalisations cas réel : Finalement implémenter le federated learning sur un jeu de données réel et plus complexe

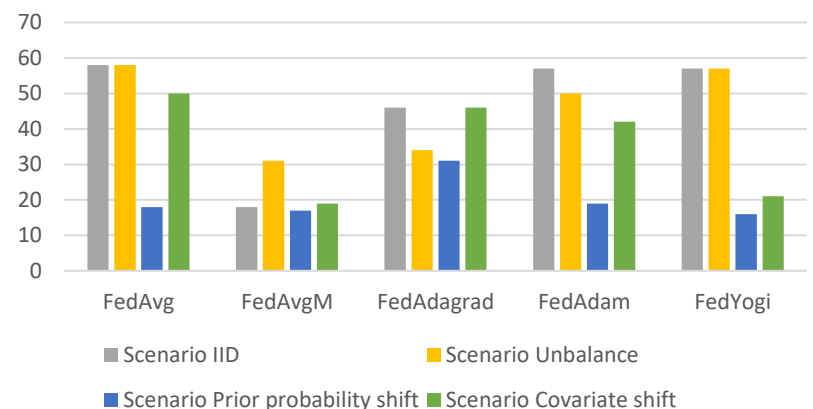
Résultats

Sur les graphes ci-dessous, vous trouverez les différentes « accuracy », ou précisions en français, des modèles créés en utilisant le federated learning sur un jeu de données simples. Le modèle « classique » créé pour ce jeu de données a obtenu une précision de 58%, or on voit sur ces résultats que certains algorithmes ont même réussi à le surpasser.

Précisions des modèles pour 4 clients



Précisions des modèles pour 10 clients



Discussion : Conclusions et perspectives

Tous les objectifs ont été atteints et même si certains modèles obtenus sur les données réelles n'ont pas donné de bons résultats, l'implémentation des données sur l'infrastructure a été une très bonne expérience pour comprendre les défis qui viennent avec un grand ensemble de données dans un environnement de federated learning.

Pour conclure, le federated learning est une technique très intéressante qui est toujours en cours de développement et qui sera très utile à l'avenir, avec l'émergence des préoccupations concernant la confidentialité des données et l'augmentation des données décentralisées. Ce projet a été une très bonne opportunité de découvrir cette technique et de comprendre les défis qui l'accompagnent.