Federated learning (Aprendizaje Distribuido)

Muchas veces los datasets que se quieren procesar con Machine Learning son tan grandes que no pueden ser procesados con una sola máquina. En la computación distribuida clásica tenemos una plataforma central (Servidor) que almacena datos de manera distribuida en varios servidores esclavos. El problema que tenemos es que se debe realizar un envío de datos a un servidor central, estando el problema de que en caso de un ataque, un atacante puede quedarse escuchando para tomar los datos que se transportan al servidor. Otro problema es la latencia que hay de por medio, contando tanto el tiempo de transporte como el de procesado por parte del servidor.

Técnicas de protección de modelos de datos en Machine Learning:

- **Anonimización** (De las peores para machine learning al enmascarar los datos): Eliminación de datos, K-Anonimidad, etc... Vulnerable a ataques de enlazado, también puede generar datasets inútiles
- **Privacidad diferencial**: Consiste en añadir ruido a los datos, puede generar datasets inútiles.
- **Computación seguras entre múltiples partes**: Es la mejor manera de proteger machine learning, pero es excesivamente lenta. Hace que el calculo no dependa de nadie, pensado para entornos P2P. Sirve para solucionar problemas donde hay una función objetivo que se puede computar de forma colaborativa entre diferentes módulos sin necesidad de revelar información.