

ENSAE PARIS - INSTITUT POLYTECHNIQUE DE PARIS



MILLIMAN FRANCE

Projet de Statistiques Appliquées

Apports et perspectives du Federated Learning dans le domaine assurantiel

Mai 2025

AUTEURS :

ABE Kevin

BENABDESADOK Nayel

HOUNKPEVI Crespin

REN Alexandre

SUPERVISÉ PAR :

François HU, Head of AI Lab chez Milliman

Fallou NIAKH, Encadrant de la voie Actuariat à l'ENSAE Paris

Caroline HILLAIRET, Professeure et Responsable de la voie Actuariat

Remerciements

Nous souhaitons exprimer notre profonde gratitude à nos tuteurs, François Hu et Fallou Niakh, pour leur accompagnement bienveillant, leur disponibilité constante et la qualité de leurs conseils tout au long de ce projet.

Nous remercions également Mme Caroline Hillairet pour ses recommandations avisées, sa disponibilité et ses retours toujours constructifs, qui ont grandement enrichi notre travail.

Enfin, nous adressons nos sincères remerciements à l'ensemble de l'équipe pédagogique de l'ENSAE, dont l'excellence de l'enseignement nous a permis de mobiliser des méthodes avancées en machine learning, et d'explorer l'apprentissage fédéré dans un cadre à la fois concret, exigeant et stimulant.

Abstract

Dans le secteur assurantiel, la modélisation prédictive des sinistres se heurte à deux obstacles majeurs : l'hétérogénéité des bases de données entre compagnies d'assurance et la sensibilité juridique des informations clients. Face à ces contraintes, notre projet explore l'application du *Federated Learning* (FL), une méthode d'apprentissage collaboratif permettant d'entraîner un modèle global sans partage de données. Après avoir contextualisé le besoin d'un tel paradigme dans un cadre européen (France, Belgique, Italie), nous avons implémenté différents algorithmes de FL, notamment FedAvg, FedProx et FedOpt, et nous les avons comparés à des approches centralisées, dont le XGBoost.

Le traitement des classes déséquilibrées (via `class_weight='balanced'` et par data augmentation par SMOTENC) ainsi que l'uniformisation des variables inter-bases ont été essentiels pour garantir la cohérence des résultats. Les performances des modèles ont été évaluées selon des métriques robustes telles que l'AUC et le score F1. Nos résultats mettent en évidence les limites de FedAvg dans des environnements non-i.i.d., et soulignent l'intérêt de FedProx et FedOpt pour améliorer la stabilité et la convergence.

Ce travail démontre que le FL offre une alternative pertinente et éthique pour l'analyse prédictive dans l'assurance, tout en respectant les contraintes de confidentialité et de distribution des données.

Table des matières

Notations	6
1 Introduction et Motivation	1
1.1 Motivation du projet	1
1.2 Qu'est-ce que le Federated Learning?	1
1.3 Philosophie du projet	2
2 Principes Généraux du Federated Learning	3
2.1 Méthodologie et Notations	3
2.2 Modèle local de chaque client : La Régression Logistique	4
2.3 Méthode Globale d'Agrégation : le FedAvg	4
3 Données	5
3.1 Description des bases	5
3.2 Variables d'intérêt et transformations	6
3.3 Statistique descriptive	7
3.4 Normalisation des données	11
3.5 Traitement du déséquilibre de classes	11
3.6 Data augmentation	11
3.7 Significativité et analyse des variables	12
4 Implémentation du FedAvg	13
4.1 Paramétrage et choix computationnels	13
4.2 Un majorant pour nos performances : le XGBoost	13
4.3 Résultats	14
5 De nouvelles méthodes d'agrégation : FedProx et FedOpt	16
5.1 Federated Proximal — FedProx	16
5.1.1 Choix Optimal du paramètre proximal μ	17
5.2 Federated Optimization - FedOpt	17
5.2.1 Choix des Paramètres dans les Optimiseurs Adaptatifs β_1, β_2 et τ	18
6 Implémentation de FedProx et FedOpt et Résultats	19
7 Conclusion	20
Annexe	24

Liste des tableaux

1	Performance prédictive théorique de différents datasets les uns sur les autres	2
2	Quantité de données par base	6
3	Résumé des variables et de leur codage	7
4	Résultats de l'analyse du V de Cramér entre variables catégorielles	10
5	Quantité de données train après data augmentation	12
6	Odds ratios (en %) des variables explicatives sur la probabilité de sinistre	12
7	Comparaison des AUC obtenus avec XGBoost et régression logistique locale avant data augmentation	14
8	Comparaison des AUC obtenus avec XGBoost et régression logistique locale après data augmentation	14
9	Résultats de l'analyse du V de Cramér entre variables catégorielles après data augmentation	23
10	Résultats de la régression logistique sur la base belge	24
11	Résultats de la régression logistique sur la base européenne	24
12	Résultats de la régression logistique sur la base française	24
13	Évaluation croisée FedAvg pour chaque région source et cible.	24
14	Évaluation croisée FedProx pour chaque région source et cible.	25
15	Évaluation croisée FedOpt pour chaque région source et cible.	26

Table des figures

1	Schéma fonctionnel du mécanisme de FL	2
2	Distribution des Sinistres	7
3	Visualisation de la distribution de l'âge des conducteurs	8
4	Visualisation de la distribution du sexe des conducteurs	8
5	Visualisation de la distribution du carburant des conducteurs	9
6	Visualisation de la distribution de la puissance du véhicule	9
7	Visualisation de la distribution de la densité de population des sinistres	10
8	Matrice de corrélation de chaque base	10
9	Évolution du score AUC au fil des rounds pour FedAvg, par région	14
10	Évolution de la fonction de perte (loss) au fil des rounds pour FedAvg	14
11	Évolution des poids DriverAge par FedAvg	15
12	Évolution des poids DriverAge par FedOpt (Adam)	19
13	Distribution des Sinistres après data augmentation	21
14	Visualisation de la distribution de l'âge des conducteurs après data augmentation	21
15	Visualisation de la distribution du sexe des conducteurs après data augmentation	22
16	Visualisation de la distribution du carburant des conducteurs après data augmentation	22
17	Visualisation de la distribution de la puissance du véhicule après data augmentation	22
18	Visualisation de la distribution de la densité de population des sinistres après data augmentation	23
19	Matrice de corrélation de chaque base après data augmentation	23
20	Évolution du score AUC au fil des rounds pour FedProx, par région	25
21	Évolution de la fonction de perte (loss) au fil des rounds pour Fedprox	25
22	Évolution du score AUC au fil des rounds pour FedOpt, par région	26
23	Évolution de la fonction de perte (loss) au fil des rounds pour FedOpt	26

Notations

Symbole	Description
x_i^k	Variables explicatives pour l'individu i du client k
y_i^k	Variable binaire indiquant la présence (1) ou absence (0) de sinistre pour l'individu i du client k
f_i	Exposition au risque pour l'individu i (proportion d'année)
$\sigma(\cdot)$	Fonction sigmoïde : $\sigma(z) = \frac{1}{1+e^{-z}}$
w_k	Vecteur de poids du modèle local pour le client k
w_t^k	Poids du modèle local du client k à l'itération t
w^{FedAvg}	Poids du modèle global agrégé par FedAvg lors de la dernière agrégation
w_{t+1}^{FedAvg}	Poids du modèle global agrégé par FedAvg à l'itération $t+1$ (Après agrégation)
w^{FedOpt}	Poids du modèle global agrégé par FedOpt lors de la dernière agrégation
w_{t+1}^{FedOpt}	Poids du modèle global agrégé par FedOpt à l'itération $t+1$ (Après agrégation)
$\mathcal{L}(w_k)$	Fonction de vraisemblance associée au client k
$F_k(w_k)$	Fonction de perte locale pour le client k
$F_k^{\text{Prox}}(w_k)$	Fonction de perte locale avec régularisation proximale
n_k	Nombre d'observations du client k
K	Nombre total de clients (compagnies d'assurance)
E	Nombre d'itérations locales avant synchronisation avec le serveur
n_{round}	Nombre total de cycles d'agrégation, égale au nombre de mini-batches
α	Taux d'apprentissage local qui sera constant à toutes les itérations
μ	Coefficient de régularisation proximale dans FedProx
Δ_t^k	Différence entre poids locaux et globaux : $w_t^k - w_{\text{Dernière Agrégation}}^{\text{FedOpt}}$
Δ_t	Moyenne pondérée des différences Δ_t^k
m_t	Vecteur de moment accumulé dans FedOpt
u_t	Mesure adaptative des gradients dans FedOpt
β_1, β_2	Coefficients de pondération pour les moments dans FedOpt
τ	Terme de régularisation numérique dans FedOpt

1 Introduction et Motivation

1.1 Motivation du projet

Considérons un organisme souhaitant analyser la sinistralité automobile à l'échelle de l'ensemble du territoire européen. Pour se faire, il veut développer un modèle prédictif capable d'estimer la probabilité de survenance d'un sinistre pour un assuré. Dans cette perspective, l'organisme collabore avec plusieurs compagnies d'assurance implantées dans différents pays (France, Belgique, Espagne, etc.), chacune ayant des données relatives à sa propre clientèle.

Une approche intuitive consisterait à centraliser l'ensemble des données fournies par ces assureurs afin d'entraîner un modèle global de classification destiné à prédire la sinistralité individuelle. Toutefois, cette stratégie naïve se heurte rapidement à plusieurs limitations majeures :

- Les modèles d'apprentissage supervisé, ainsi que les fondements théoriques de l'inférence statistique, reposent sur l'hypothèse que les observations X_1, \dots, X_n sont **indépendantes et identiquement distribuées** (i.i.d.). Or, cette hypothèse est incompatible avec le contexte considéré. En effet, chaque compagnie d'assurance dessert une clientèle spécifique, souvent caractérisée par des critères socio-professionnels ou géographiques distincts. Cette hétérogénéité inter-clients induit des distributions divergentes entre assureurs, invalidant l'application directe d'un modèle global fondé sur l'hypothèse i.i.d.
- Le partage des données clients entre compagnies d'assurance est difficilement envisageable en raison de leur **caractère sensible** : données médicales, historique des sinistres, infractions, revenus ou informations de géolocalisation. Ces données sont strictement encadrées par des réglementations telles que le **RGPD**, rendant la centralisation des informations juridiquement et techniquement complexe.

Notre projet de Statistiques Appliquées s'inscrit dans la résolution de ce problème, en explorant un paradigme émergent en assurance : le *Federated Learning* (apprentissage fédéré). L'ensemble du code est disponible publiquement sur GitHub : https://github.com/nayelsdk/Federated_Learning_Milliman.

1.2 Qu'est-ce que le Federated Learning?

L'apprentissage fédéré (FL) est une méthode d'entraînement collaboratif de modèles, largement adoptée dans des secteurs comme la cybersécurité, le cloud computing, la santé et la finance. Le principe reste identique malgré des spécificités propres à chaque domaine : plusieurs entités entraînent ensemble un modèle global, sans jamais partager leurs données locales. Par exemple, dans le secteur médical, plusieurs hôpitaux peuvent entraîner un modèle de diagnostic commun sans jamais échanger les dossiers patients, améliorant ainsi la détection de maladies rares tout en respectant la confidentialité.

Dans le domaine du risque climatique, l'apprentissage fédéré montre un fort potentiel pour la détection précoce des feux de forêt. Un exemple marquant est le modèle *IOFireNet*, qui combine des données locales issues de capteurs et d'images satellites pour détecter et localiser les incendies avec précision, tout en préservant la confidentialité des données. Ce modèle repose sur des techniques avancées de traitement d'image intégrées dans une architecture fédérée, permettant à chaque entité participante (par exemple, des services d'incendie nationaux ou des centres météorologiques) de contribuer à l'entraînement sans divulguer ses données sensibles¹. Ces approches permettent d'améliorer la précision des systèmes d'alerte tout en assurant la souveraineté des données de chaque pays, un facteur clé dans les contextes transnationaux.

1. C. Pei et al., "IOFireNet : Federated Deep Learning for Wildfire Smoke Detection and Localization in Smart IoT Environment," *IEEE Internet of Things Journal*, 2024. Disponible sur <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC11621580/>

En somme, cette méthodologie vise à offrir de bonnes performances sur l'ensemble des participants. Par exemple, si une entreprise A entraîne un modèle uniquement sur ses propres données, celui-ci sera inefficace sur les données d'une entreprise B ou C. L'objectif du FL est d'assurer une performance homogène sur l'ensemble des entités :

Coefficients / Base de Données	A	B	C
A	✓✓	×	×
B	×	✓✓	×
C	×	×	✓✓
FL	✓	✓	✓

TABLE 1 – Performance prédictive théorique de différents datasets les uns sur les autres

La littérature distingue deux formes principales : l'**HFL**, où les entités ont des caractéristiques similaires sur des individus différents, et la **VFL**, où elles détiennent des données complémentaires sur les mêmes individus. Dans notre projet, comme les assurés diffèrent d'une entité à l'autre, nous sommes en **HFL**.

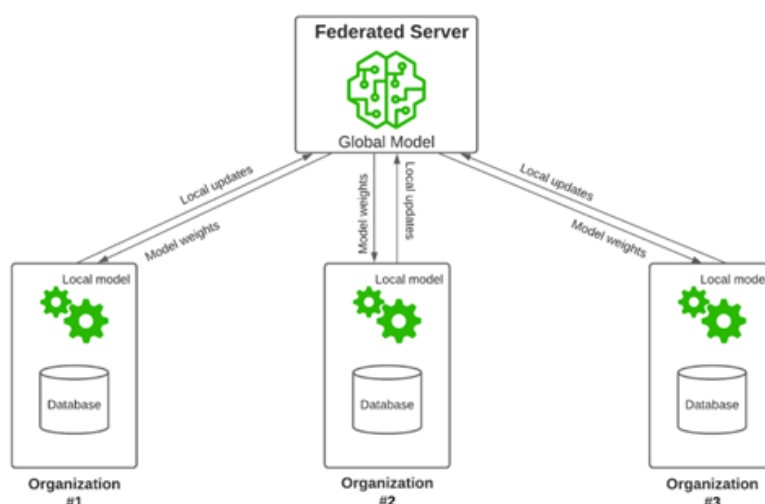


FIGURE 1 – Schéma fonctionnel du mécanisme de FL

1.3 Philosophie du projet

Notre projet, à visée exploratoire, ne cherche pas en premier lieu à comparer les performances du *Federated Learning* à d'autres méthodes d'apprentissage supervisé, bien que cette comparaison soit abordée dans une section ultérieure. L'objectif principal est plutôt d'évaluer la pertinence de ce paradigme dans le domaine assurantiel, en mettant en lumière sa robustesse face aux contraintes évoquées précédemment — notamment le non respect de l'hypothèse i.i.d. des données et les enjeux de confidentialité.

Par ailleurs, le secteur de l'assurance est fortement encadré par des réglementations visant à garantir l'égalité d'accès aux services et à éviter toute forme de discrimination, notamment fondée sur le genre, l'âge ou l'état de santé. Dans ce contexte, certaines variables sensibles — comme le genre — ne sont mobilisées dans notre analyse qu'à des fins pédagogiques et expérimentales, et en conformité avec les enjeux éthiques et réglementaires évoqués dans le cadre du projet².

2. Voir [Délibération n° 2009-373 du 26 octobre 2009 relative aux discriminations dans l'accès à l'assurance](#) publiée par la HALDE.

2 Principes Généraux du Federated Learning

2.1 Méthodologie et Notations

Pour mieux visualiser le fonctionnement de l'apprentissage fédéré ([1]), considérons un scénario dans lequel K compagnies d'assurance souhaitent collaborer pour entraîner un modèle prédictif commun. Chaque compagnie (appelée *client*) dispose localement d'une base de données contenant les informations de ses propres assurés. À partir de ces données locales, chaque client effectue un entraînement indépendant de son modèle et transmet ensuite les poids appris à un serveur central, appelé *Federated Server*. Ce serveur est chargé d'agréger les modèles locaux afin de mettre à jour le modèle global.

Lors de la phase d'entraînement local, nous adoptons une stratégie d'optimisation par *mini-batch gradient descent*, qui constitue un compromis entre deux approches extrêmes : l'approche *batch*, où les gradients sont calculés sur l'ensemble des données locales, et l'approche *stochastique*, qui repose sur des mises à jour à partir d'exemples individuels. L'utilisation de mini-batches permet non seulement de réduire le coût computationnel, mais aussi d'introduire un bruit bénéfique dans les gradients, favorisant ainsi la généralisation du modèle.

Le processus d'entraînement se déroule de manière itérative, avec des itérations locales indexées par $t \in \mathbb{N}$. On note E le nombre d'itérations locales effectuées par chaque client avant une synchronisation globale avec le serveur. Par exemple, si $E = 5$, chaque client mettra à jour ses poids pendant 5 itérations sur ses mini-batches avant de transmettre ses nouveaux paramètres au serveur pour agrégation. Ce cycle est répété n_{round} fois. Ainsi, chaque client effectue au total $n_{\text{round}} \cdot E$ itérations locales.

Afin de structurer cette dynamique d'entraînement, chaque jeu de données local est préalablement divisé en mini-batches. Le nombre de mini-batches est alors égal à n_{round} , ce qui permet d'assurer une couverture complète et équilibrée des données locales au fil des itérations d'entraînement.

Algorithm 1 Pseudo-code de l'Apprentissage Fédéré

```

1: Initialisation : Le serveur central initialise les poids globaux  $w_0$ 
2: for  $t = 1$  à  $n_{\text{round}}$  do
3:   Le serveur envoie les poids courants  $w_{t-1}$  aux clients sélectionnés
4:   for chaque client  $i \in \{1, \dots, K\}$  en parallèle do
5:     Le client  $i$  échantillonne un mini-batch  $B_t^i$  local (différent à chaque round)
6:     for  $e = 1$  à  $E$  do ▷ Itérations locales
7:       Mise à jour locale des poids :  $w \leftarrow w - \alpha \nabla F_k(w; B_t^i)$    Avg et Opt, pour Prox : (7)
8:     end for
9:     Le client  $i$  retourne les poids locaux  $w_t^i$  au serveur
10:  end for
11:  Agrégation globale :  $w_t \leftarrow \text{Aggregate}(w_t^1, \dots, w_t^K)$  ▷ FedAvg, FedProx, FedOpt
12: end for

```

Voici les différentes variables que nous introduisons à ce stade de l'étude :

- Pour chaque assureur k , nous disposons de n_k observations notées $\mathbf{x}_1^k, \dots, \mathbf{x}_{n_k}^k$ correspondant aux variables explicatives, ainsi que de $y_1^k, \dots, y_{n_k}^k$ (qui sera appelée *Sinistre*), représentant notre variable booléenne cible (0 ou 1). Cette dernière répond à la question suivante : « *L'individu i a-t-il réalisé un sinistre pendant sa période d'exposition au risque, compte tenu des informations disponibles ?* ».
- $\mathbf{w}^{\text{agregated}}$ et \mathbf{w}_k désignent respectivement les vecteurs de poids associés aux modèles global et local (pour l'assureur k). Ces vecteurs appartiennent à \mathbb{R}^d , où d est le nombre de covariables utilisées pour

prédire notre variable d'intérêt. Dans notre cas, $d = 5$. Les covariables retenues seront détaillées dans la section consacrée à la description de la base de données.

- $F_k(w_k)$ représente la fonction de perte locale associée à l'assureur k . Dans notre étude, il s'agit d'un modèle de régression logistique que nous allons développer dans la partie suivante.

2.2 Modèle local de chaque client : La Régression Logistique

Dans le cadre d'un projet assurantiel, les modèles de régression sont largement privilégiés en raison de leur simplicité d'utilisation et de leur grande interprétabilité, qualités qui sont moins évidentes pour les modèles de réseaux de neurones.

L'objectif est de prédire la probabilité de survenance d'un sinistre automobile, notée $P(y_i^k = 1 | x_i^k, w_k)$. Chaque observation est également associée à une durée d'exposition au risque sur une période d'une année, représentée par la variable $\text{Exposure} \in [0, 1]$.

Par exemple, si $\text{Exposure}_j = 0.5$, cela signifie que l'assuré j a été exposé au risque pendant 50% de l'année. Ainsi, la probabilité prédite sans prise en compte de l'exposition est $P(y_j^k = 1 | x_j^k, w_k) = 40\%$, alors que la probabilité réelle, pondérée par l'exposition est :

$$P_{\text{réel}}(y_j^k = 1 | x_j^k, w_k) = \text{Exposure}_j \times 40\% = 20\%.$$

Le calcul de la probabilité pour chaque observation s'exprime comme suit :

$$P(y_i^k = 1 | x_i^k, w_k) = \sigma\left((x_i^k)^T w_k\right) \times f_i = \frac{1}{1 + e^{-(x_i^k)^T w_k}} \times f_i \quad (1)$$

où $\sigma(\cdot)$ désigne la fonction sigmoïde, et $f_i = \text{Exposure}_i \in [0, 1]$.

De plus, on peut remarquer que la variable aléatoire $y_i^k | x_i^k, w_k$ suit une loi de Bernoulli, dont la probabilité de succès est donnée plus haut, par l'équation (1) :

$$P(y_i^k | x_i^k, w_k) = \left[\sigma\left((x_i^k)^T w_k\right) f_i\right]^{y_i^k} \cdot \left[1 - \sigma\left((x_i^k)^T w_k\right) f_i\right]^{1-y_i^k} \quad (2)$$

En supposant l'indépendance des observations, et par définition de la fonction de vraisemblance, nous avons :

$$\mathcal{L}(w_k) = \prod_{i=1}^{n_k} P(y_i^k | x_i^k, w_k) = \prod_{i=1}^{n_k} \left[\sigma\left((x_i^k)^T w_k\right) f_i\right]^{y_i^k} \left[1 - \sigma\left((x_i^k)^T w_k\right) f_i\right]^{1-y_i^k} \quad (3)$$

La fonction de coût correspond alors à la négative log-vraisemblance normalisée :

$$F_k(w_k) = -\frac{1}{n_k} \sum_{i=1}^{n_k} \left[y_i^k \log\left(\sigma\left((x_i^k)^T w_k\right) f_i\right) + (1 - y_i^k) \log\left(1 - \sigma\left((x_i^k)^T w_k\right) f_i\right) \right] \quad (4)$$

Nous utiliserons en premier lieu cette fonction coût dans nos modèles locaux. Maintenant que nous savons ce que font les *clients* de façon indépendantes, il faudrait savoir comment nous allons agréger les poids au *Serveur*.

2.3 Méthode Globale d'Agrégation : le FedAvg

La méthode **FedAvg** [1] est l'une des premières approches qui vient à l'esprit lorsqu'on envisage l'agrégation de données dans un contexte d'apprentissage fédéré. Elle repose sur une moyenne pondérée des poids w_k des modèles locaux, permettant ainsi de consolider les connaissances acquises de manière distribuée.

À la fin de chaque cycle d'entraînement, correspondant à E itérations locales, les poids des différents modèles sont transmis au *serveur central* où une agrégation, partielle (une partie des clients seulement réalisent l'agrégation et tous les clients reçoivent tout de même w_{t+1}^{FedAvg}) ou totale, est réalisée. Dans notre étude, nous disposons de trois jeux de données distincts, qui seront détaillés ultérieurement. En conséquence, nous optons exclusivement pour des **mise à jour totales**.

Lors d'une itération $t + 1$ (itération d'agrégation, juste à la suite des E itérations locales), la règle d'agrégation est formalisée par l'expression suivante :

$$w_{t+1}^{\text{FedAvg}} = \sum_{i=1}^K \frac{n_i}{\sum_j n_j} w_t^i \quad (5)$$

où :

— w_t^i désigne les poids du modèle local issu du i -ème dataset à l'itération t ,

Cette méthode garantit que les contributions des différents modèles locaux sont proportionnelles à la taille de leurs ensembles de données respectifs, assurant ainsi une agrégation équitable et efficace dans la mesure où les datasets soient relativement bien équilibrés.

Conclusion partielle

À partir de cette formalisation, un premier modèle de **Federated Learning** basée sur la méthode **FedAvg** peut être réalisé. Mais avant cela, une exploration préalable s'impose pour mieux appréhender nos jeux de données.

3 Données

3.1 Description des bases

Les bases de données **beMTPL**, **euMTPL** et **freMTPL** sont conçues pour analyser les risques liés aux contrats d'assurance automobile et aux sinistres. Elles permettent aux compagnies d'assurance d'évaluer la sinistralité des conducteurs et des véhicules. Ces bases contiennent des informations sur les assurés, les véhicules, les contrats d'assurance, ainsi que sur les sinistres survenus (notre variable binaire y_i^k d'intérêt). On y retrouve notamment des variables clés telles que le nombre de sinistres (**ClaimNb**), le montant des sinistres (**ClaimAmount**), l'exposition (**Exposure**), ainsi que des caractéristiques du conducteur (**âge**, **sexe**) et du véhicule (**puissance**, **âge**, **carburant**, etc.).

Ces jeux de données sont accessibles publiquement via le site de CASData (<https://cas.uqam.ca/pub/web/CASdatasets-manual.pdf>).

freMTPL

La base **freMTPL** représente le marché français de l'assurance automobile, spécifiquement la garantie *Responsabilité Civile Automobile* (Motor Third-Party Liability). Elle résulte de la fusion des jeux de données **freMTPLfreq** et **freMTPLsev** (ou leurs extensions **freMTPL2freq** et **freMTPL2sev**). Chaque contrat y est décrit par ses caractéristiques de risque, le nombre de sinistres observés, ainsi que leur coût total ou détaillé sur une période principalement annuelle.

beMTPL et euMTPL

La base **beMTPL** (Belgian Motor Third-Party Liability) contient des données sur 163 212 assurés uniques, incluant véhicules, conducteurs et sinistres, avec une exposition mesurée en fraction d'année, ce qui permet une analyse précise des fréquences de sinistres.

La base **euMTPL**, de portée européenne, adopte une structure similaire aux autres jeux de données, couvrant trois années d'expérience à travers plusieurs pays. Collectées au début des années 2000, ces données permettent d'analyser l'évolution temporelle du risque automobile en Europe.

Cependant, une analyse attentive montre que le dataset est en réalité principalement centré sur l'Italie. En effet, la variable province, codée sous forme de deux lettres, correspond aux différentes provinces italiennes (voir la liste complète : https://fr.wikipedia.org/wiki/Province_d%27Italie). Pour davantage de clarté, nous continuerons quand même de l'appeler base européenne.

Nous pouvons en premier lieu porter un regard le nombre d'observations de chaque dataset et leur proportion, variables importantes à prendre en compte pour le **FedAvg** :

Base	Nombre d'observations
freMTPL - France	1 091 182
beMTPL - Belgique	163 212
euMTPL - Europe/Italie	2 373 197

TABLE 2 – Quantité de données par base

Remarque : Nous pouvons voir que la base européenne est prépondérante par rapport aux autres, cela est à prendre en compte pour évaluer notre modèle ainsi que nos résultats de FL.

3.2 Variables d'intérêt et transformations

Les trois bases, bien qu'elles se rapportent à la même activité (risque automobile), ne présentent pas exactement les mêmes données. Il était donc nécessaire d'agglomérer et de mettre en cohésion les bases en prenant les mêmes covariables $(x_i^k)_{i=1}^{n_k}$ d'intérêts. Certaines variables n'étaient toutefois pas présentes dans toutes les bases, ce qui nous a amenés à les générer ou à les récupérer à partir d'autres sources.

- **DriverAge** : Aucune modification n'a été apporté, cette variable est commune aux trois bases. Toutefois on permet de rappeler que l'étude de cette variable est purement théorique et ne doit pas être utiliser pour déterminer la prime d'assurance selon les législations européennes de 2010.
- **Gender** : La variable Gender n'étant pas présente dans la base française, mais disponible dans les bases belge et européenne, nous avons initialement envisagé de la générer artificiellement en utilisant la répartition homme/femme en France (49%-51%). Cependant, l'analyse de la significativité de cette variable dans une régression logistique appliquée à la base française a révélé qu'elle n'était pas significative aux seuils standards α , ce qui est cohérent avec son caractère artificiel.

Pour affiner l'analyse, nous avons reconstitué la variable Gender dans la base française en nous basant sur d'autres critères. Plus précisément, nous avons cherché à estimer la probabilité qu'un assuré déclare un sinistre en fonction de son genre. Pour cela, nous avons attribué le genre des assurés français sinistrés de manière à respecter les paramètres suivants : $p_h = \frac{\text{Nombre d'hommes sinistrés}}{\text{Nombre total d'hommes}} = 0.06$ et $p_f = \frac{\text{Nombre de femmes non sinistrés}}{\text{Nombre total de femmes}} = 0.03$. Nous avons d'abord généré aléatoirement une variable Gender avec une répartition de 50% d'hommes et 50% de femmes, puis ajusté les affectations de genre pour que les taux conditionnels soient respectés. Cette reconstitution, volontairement biaisée, a pour but de créer une variable Gender ayant un léger impact sur la sinistralité, afin de tester sa contribution

dans les modèles prédictifs.

- **Power** : La variable Power est présente dans les trois bases. Toutefois son codage est différent dans freMPTLfreq(numérique) et freMTPL2freq(alphabétique). Afin d'avoir une cohérence entre les bases, nous avons procédé à une correspondance des différentes valeurs.
- **Fuel_type** : Fuel_type est présent dans les trois bases mais il y a hétérogénéité entre les trois bases nous avons dû uniformiser les données pour ne garder que la variable binaire Diesel ou Regular
- **Density** : La variable Density, représentant le nombre d'habitants par km², n'était initialement disponible que dans la base française. Toutefois, les bases belge et européenne contenaient une variable indiquant la région de résidence de l'assuré. En exploitant cette information, nous avons pu associer à chaque région une densité de population issue de bases de données officielles belges et européennes.

Nous avons ainsi procédé à une jointure entre les bases d'assurés et les données de densité régionales, permettant de reconstituer la variable Density pour les bases belge et européenne.

- **Sinistre** : La base européenne ne possédait pas de variable portant sur la présence de sinistre mais on avait le coût des sinistres. Nous dirons que si le montant est nul, alors il n'y a pas eu de sinistre (= 0), sinon il y a eu sinistre (= 1).

Variable	Description	Codage
DriverAge	Âge du conducteur principal	Entier (ex. : 45)
Gender	Sexe du conducteur	Binaire : Male = 1, Female = 0
Power	Puissance du véhicule (en kW)	Entier (ex. : 85)
Fuel_type	Type de carburant	Binaire : Diesel = 1, Regular = 0
Density	Densité de population de la zone de résidence	Entier (ex. : 1200 hab/ km ²)
Sinistre	Présence d'un sinistre	Binaire : Oui = 1, Non = 0

TABLE 3 – Résumé des variables et de leur codage

3.3 Statistique descriptive

Dans cette section, nous présentons les statistiques descriptives des trois bases de données utilisées. L'objectif est de mieux comprendre la structure des données, détecter d'éventuelles anomalies et évaluer les distributions des variables avant d'appliquer toute méthode de modélisation. Cette analyse permet également d'identifier d'éventuelles corrélations ou colinéarités entre les variables, afin d'orienter de manière plus rigoureuse la construction des modèles de régression dans les sections suivantes.

Sinistre

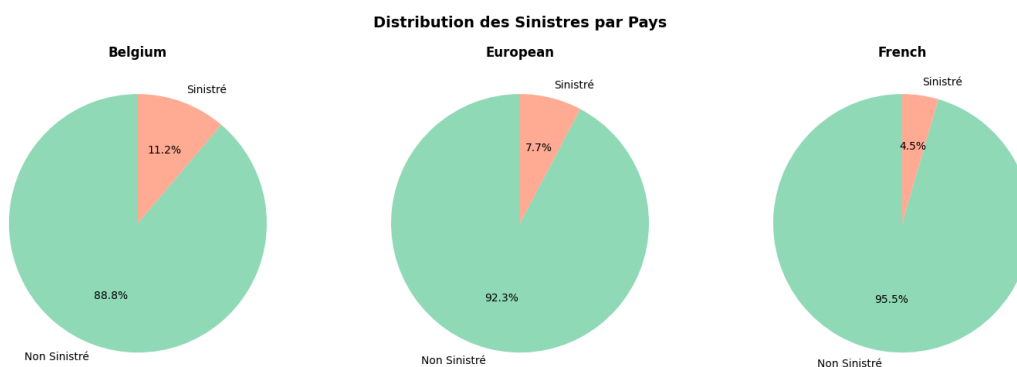
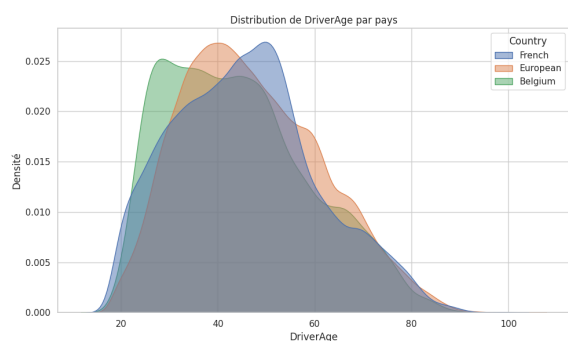
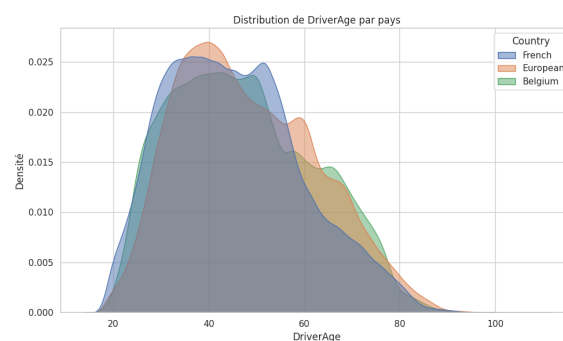


FIGURE 2 – Distribution des Sinistres

DriverAge



(a) Distribution de l'âge des conducteurs parmi les sinistres

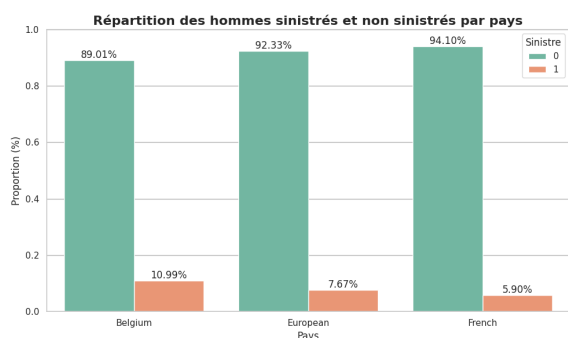


(b) Distribution de l'âge des conducteurs parmi les non sinistres

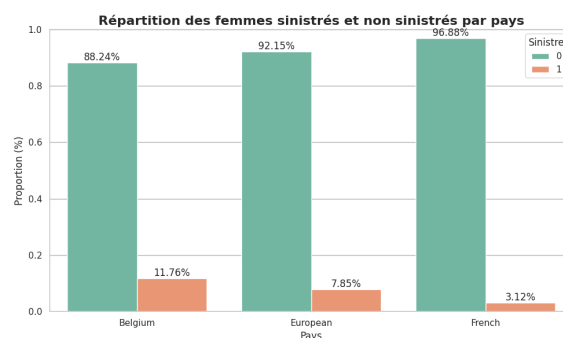
FIGURE 3 – Visualisation de la distribution de l'âge des conducteurs

On observe que, pour la base européenne (courbe orange), les deux distributions (sinistres et non sinistres) sont globalement similaires, suggérant que la répartition de l'âge est relativement stable quel que soit le statut sinistre. En revanche, des différences sont présentes dans les bases françaises et belges. En France, on remarque qu'il y a proportionnellement plus de conducteurs non sinistres âgés de moins de 45 ans, ce qui pourrait indiquer une sinistralité légèrement plus élevée chez les conducteurs plus âgés. À l'inverse, en Belgique, on observe une plus forte proportion de sinistres parmi les moins de 45 ans, ce qui suggère une tendance opposée. Ainsi, ces différences selon les pays rendent difficile toute conclusion généralisable quant à l'effet de l'âge sur la survenue d'un sinistre. Une analyse statistique approfondie serait nécessaire pour mieux cerner ce lien potentiel.

Gender



(a) Proportion d'homme sinistré et non sinistré



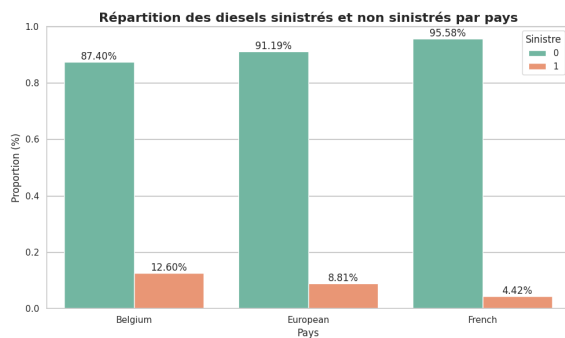
(b) Proportion de femme sinistrée et non sinistré

FIGURE 4 – Visualisation de la distribution du sexe des conducteurs

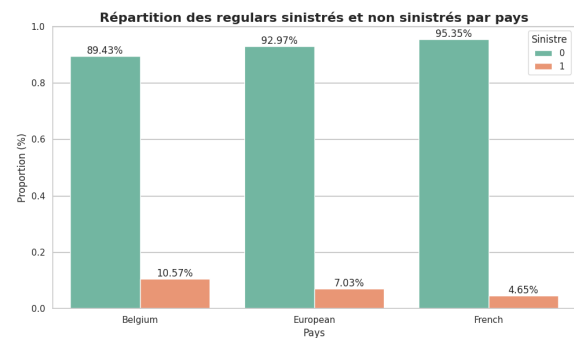
On constate que pour la base belge nous avons 10.99% d'hommes sinistres et 11.76% de femmes sinistrées et pour la base européenne 7.67% d'hommes sinistres et 7.85% de femmes sinistrées. Les proportions de femmes et d'hommes restent relativement similaires avec une différence entre sinistres et non sinistres de 0.8% au maximum. Par conséquent, il est difficile de conclure que le sexe a un impact significatif sur la survenance d'un sinistre. Une analyse statistique plus approfondie serait nécessaire pour déterminer s'il existe une réelle dépendance entre le sexe et la probabilité d'avoir un sinistre.

La base française est différente des deux autres et nous avons effectivement une influence du genre sur la survenance d'un sinistre mais cela est volontaire, comme précisé dans la section précédente.

Fuel type



(a) Distribution de Diesel sinistré et non sinistré

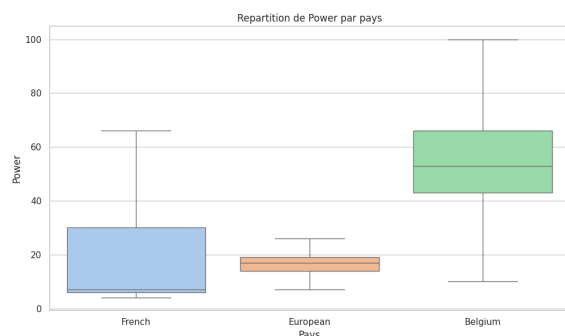


(b) Distribution de Regular sinistré et non sinistré

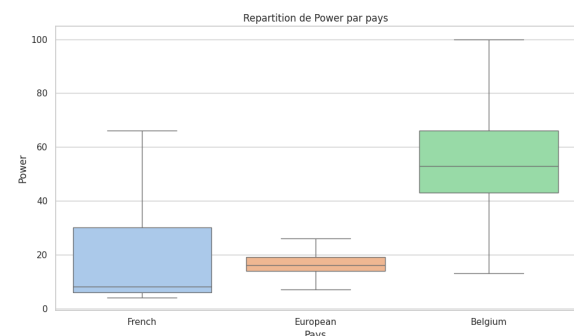
FIGURE 5 – Visualisation de la distribution du carburant des conducteurs

On constate que pour la base française la répartition est identique parmi les sinistrés et non sinistrés. Cependant on observe une légère différence pour les deux autres bases. On a une différence de 2% environ. Ainsi on a donc pour la base européenne et belge que les assurés ayant une voiture avec un carburant diesel sont légèrement plus susceptibles d'avoir un sinistre.

Power



(a) Boxplot de la puissance du véhicule des sinistrés

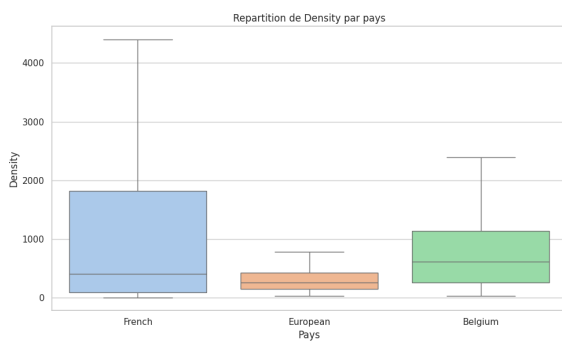


(b) Boxplot de la puissance du véhicule des non sinistrés

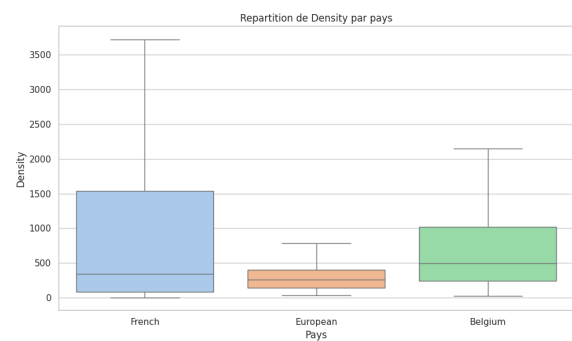
FIGURE 6 – Visualisation de la distribution de la puissance du véhicule

La puissance du véhicule est très asymétrique avec une grande concentration sur les petites valeurs et quelques très grandes valeurs, ce qui rend difficile l'analyse. Ainsi nous avons choisi de ne pas afficher les outliers dans le boxplot pour plus de visibilité. On remarque que les distributions ne sont pas les mêmes entre les bases, on a donc une hétérogénéité entre les pays mais cependant parmi le groupe de sinistrés il n'y a presque pas de différence. Par ailleurs, à cause de la présence de valeurs très élevées, on se retrouve avec des données très concentrées autour de la valeur 0 après la normalisation MinMaxScaler spécifiée plus bas.

Density



(a) Boxplot de la densité de population des sinistres

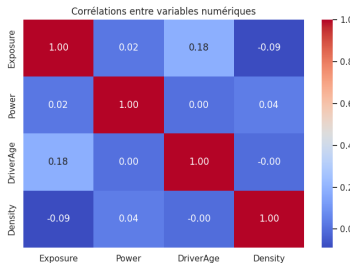


(b) Boxplot de la densité de population des sinistres

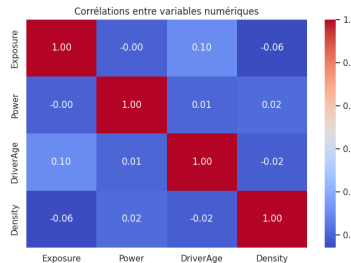
FIGURE 7 – Visualisation de la distribution de la densité de population des sinistres

Pour les mêmes raisons que pour la variable Power, les outliers ne sont pas affichés. Dans la base française, on observe un boxplot légèrement plus étalé pour les conducteurs ayant eu un sinistre, bien que la médiane reste similaire entre les deux groupes. Pour les bases belge et européenne, les distributions sont quasiment identiques quel que soit le statut sinistre. Ainsi, il n'apparaît pas de corrélation claire ni systématique entre la densité de population et la survenue d'un sinistre dans l'ensemble des pays. On observe comme pour la variable Power que les différentes distributions sont hétérogènes entre les pays.

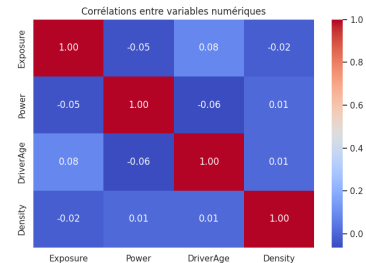
Corrélation des variables



(a) France



(b) Belge



(c) Européenne

FIGURE 8 – Matrice de corrélation de chaque base

On observe une très légère corrélation entre Exposure et DriverAge (0.18 dans la base française) mais toutes les autres relations entre variables sont quasiment inexistantes (≤ 0.10). Il n'existe donc aucun lien fort entre les variables continues mesurées.

Jeu de données	Paire de variables	Cramer's V	Force de corrélation	p-value	Significatif
French	Fuel type - Sinistre	0.006	Faible	0.000	Oui
French	Fuel type - Sex	0.001	Faible	0.4178	Non
French	Sinistre - Sex	0.067	Faible	0.000	Oui
Belgium	Fuel type - Sinistre	0.030	Faible	0.000	Oui
Belgium	Fuel type - Sex	0.103	Modérée	0.000	Oui
Belgium	Sinistre - Sex	0.011	Faible	0.000	Oui
European	Fuel type - Sinistre	0.033	Faible	0.000	Oui
European	Fuel type - Sex	0.153	Modérée	0.000	Oui
European	Sinistre - Sex	0.003	Faible	0.000	Oui

TABLE 4 – Résultats de l'analyse du V de Cramér entre variables catégorielles

Le V de Cramér indique que l'association entre les différentes variables catégorielles est très faible dans l'ensemble, avec une valeur maximale observée de seulement 0,1. Bien que certaines p-values soient statistiquement significatives, l'intensité de la corrélation reste négligeable. Cela suggère qu'il n'existe pas de dépendance entre les variables.

Par conséquent, les variables explicatives ne présentent pas de multicolinéarité, ce qui permet d'envisager l'application d'un modèle de régression logistique.

3.4 Normalisation des données

Afin de garantir une échelle homogène entre les variables continues, indispensable pour la méthode de régression logistique avec régularisation, nous avons appliqué une normalisation via la méthode MinMaxScaler aux variables DriverAge, Power et Density.

Cette normalisation est définie par la formule suivante avec X un vecteur : $X_{MinMaxScaled} = \frac{X - X_{min}}{X_{max} - X_{min}}$.

Cette transformation permet de les ramener dans un intervalle commun (ici $[0, 1]$), facilite l'interprétation et améliore la stabilité des algorithmes d'apprentissage.

3.5 Traitement du déséquilibre de classes

Nous observons que la grande majorité des assurés n'ont déclaré aucun sinistre, tandis qu'une minorité (4.5%, 7.7% et 11.2%) a été impliquée dans au moins un événement (Voir Statistique descriptive 2). Cette distribution illustre un déséquilibre typique dans les données d'assurance, où les sinistres sont des événements relativement rares. Ce phénomène, appelé *déséquilibre de classes*, doit impérativement être pris en compte lors de la modélisation pour éviter un apprentissage biaisé en faveur de la classe majoritaire.

Pour y remédier, nous utilisons le paramètre `class_weight='balanced'` dans nos modèles de classification. Ce paramètre permet d'ajuster automatiquement l'importance accordée à chaque classe en fonction de sa fréquence dans l'échantillon, de manière à compenser le déséquilibre structurel.

Par ailleurs, les métriques classiques comme la précision peuvent se révéler trompeuses dans ce contexte. Un modèle trivial prédisant systématiquement l'absence de sinistre (classe majoritaire) obtiendrait une précision proche de 90 %, sans pour autant offrir une réelle capacité de discrimination. C'est pourquoi nous privilégions des indicateurs mieux adaptés aux données déséquilibrées, tels que la courbe ROC (Receiver Operating Characteristic) et le score F1, qui tiennent compte des déséquilibres entre faux positifs et faux négatifs.

3.6 Data augmentation

Afin de corriger plus efficacement le déséquilibre important entre les classes avec seulement 11.2% au maximum d'observations (`Sinistre=1`), nous avons recours à une méthode de *data augmentation* adaptée : SMOTENC (Synthetic Minority Over-sampling Technique for Nominal and Continuous features). Elle crée de nouvelles observations en interpolant entre des individus minoritaires et leurs plus proches voisins de même classe, uniquement sur les variables continues, tout en sélectionnant les modalités des variables catégorielles aléatoirement.

SMOTENC est appliqué uniquement sur l'échantillon d'entraînement, après séparation stratifiée des données. Cette approche garantit une évaluation plus réaliste de la capacité du modèle à généraliser, sans influence des données augmentées sur la phase de test. L'application de SMOTENC nous a permis de ramener le déséquilibre à un ratio d'environ 70/30 dans les données d'entraînement, améliorant ainsi la représentativité de la classe minoritaire et, par conséquent, la capacité du modèle à en apprendre la

dynamique. Le tableau suivant récapitule les tailles des échantillons d'entraînement après augmentation (train uniquement) :

Base	Nombre d'observations (train après augmentation)
freMTPL - France	892 451
beMTPL - Belgique	124 354
euMTPL - Europe/Italie	1 878 003

TABLE 5 – Quantité de données train après data augmentation

Les analyses statistiques (voir Annexe 13) révèlent que les distributions des variables explicatives (âge du conducteur, puissance du véhicule, densité, etc.) restent globalement stables après l'augmentation. La principale évolution concerne naturellement la variable cible Sinistre, dont la proportion passe de 10% à 30% dans le train, conformément à l'objectif de rééquilibrage.

3.7 Significativité et analyse des variables

Pour chaque base, la régression logistique permet d'obtenir des coefficients estimés associés à chaque covariable et tous les coefficients obtenus dans les trois modèles de régression logistique sont statistiquement significatifs, avec des *p-values* proches de zéro ($p < 0,001$). Les détails des résultats de chaque régression sont disponibles en annexe.

Les coefficients, une fois transformés par exponentiation, donnent une variation relative (en pourcentage) de la cote (odds) de survenue d'un sinistre pour une unité d'augmentation de la variable concernée. Ces résultats ont été résumés dans le tableau ci-dessous (cf. Tableau 6), permettant une lecture comparative directe entre les trois bases.

Le tableau ci-dessous reprend l'effet marginal interprété en pourcentage :

$$\text{Var (\%)} = 100 \times (e^{\beta} - 1), \quad \text{où } \beta \text{ est le coefficient estimé.}$$

Variable	Base Belge (%)	Base Européenne (%)	Base Française (%)
Power	71,09	611,00	-56,27
DriverAge	-65,54	-27,79	13,66
Density	105,37	15,95	19,06
Sex	-3,75	-4,74	94,88
Fuel_type	18,27	26,24	-3,26

TABLE 6 – Odds ratios (en %) des variables explicatives sur la probabilité de sinistre

Globalement, les facteurs influençant le risque de sinistre varient selon les pays. La puissance du véhicule, bien que fortement corrélée au risque en Belgique et surtout en Europe, est curieusement associée à un moindre risque en France. L'âge du conducteur suit une tendance logique : plus le conducteur est âgé, moins il est risqué, sauf en France. La densité de population est un facteur de risque clair et cohérent, soulignant l'effet des environnements urbains. Conduire un véhicule diesel (vs. regular) est associé à un risque plus élevé en Belgique et en Europe, mais à un risque légèrement plus faible en France. Ensuite, pour la variable Sex, être un homme (vs. une femme) réduit légèrement le risque de sinistre en Belgique et en Europe, tandis que cet effet est positivement corrélée en France ce qui est attendu au vu de la construction de cette variable. Enfin, cette différence de signe entre pays est un indice précieux indiquant : soit des comportements et contextes différents, soit la variable reflète des réalités sociales ou techniques différentes, soit il y a des corrélations

locales non contrôlées. Cela justifie l'intérêt d'approches comme l'apprentissage fédéré, car elles permettent à chaque client de conserver ses spécificités, tout en apprenant un modèle commun.

4 Implémentation du FedAvg

4.1 Paramétrage et choix computationnels

Pour implémenter notre modèle FedAvg, comme mentionné dans le pseudo-code 1, nous avons réalisé plusieurs régressions logistiques à l'aide de mini-batches. Afin de capter les spécificités de chaque base de données, chaque mini-batch a été entraîné sur 60 % des données, puis évalué sur les 40 % restantes.

Nous avons initialement utilisé la classe `LogisticRegression()` de `sklearn`, avant d'opter pour `SGDClassifier`, qui permet de spécifier la fonction de perte – en l'occurrence, `log_loss` pour une régression logistique. Plusieurs paramètres importants ont été utilisés : `class_weight='balanced'` afin de corriger le déséquilibre entre les classes, et `warm_start=True`, qui permet d'interrompre l'apprentissage pour réinjecter de nouveaux poids après l'agrégation, facilitant ainsi la poursuite du processus d'entraînement dans le cadre fédéré.

Nous avons également appliqué une régularisation de type \mathcal{L}^2 via `penalty="l2"`, dans le but de limiter l'impact de certaines variables dominantes. En effet, nos tests préliminaires ont montré que l'intercept tendait à dominer fortement les autres coefficients, justifiant ainsi le recours à la régularisation.

4.2 Un majorant pour nos performances : le XGBoost

Comme évoqué précédemment, la précision n'est pas un indicateur pertinent dans un contexte de classes déséquilibrées, comme c'est le cas ici avec la variable `Sinistre`. Pour cette raison, nous privilégions l'utilisation de métriques robustes telles que la courbe ROC et l'aire sous la courbe (AUC). Afin d'évaluer la performance maximale théorique que l'on pourrait atteindre avec nos données – de qualité raisonnable mais imparfaite – nous avons construit un modèle XGBoost. Ce dernier nous sert de borne supérieure pour les performances attendues, en particulier en termes d'AUC.

Bien que l'objectif principal de notre projet ne soit pas de démontrer la supériorité du Federated Learning (FL) sur d'autres approches, mais plutôt de mettre en évidence sa robustesse face aux contraintes évoquées dans l'introduction (hétérogénéité des données, contraintes de confidentialité, etc.), il reste néanmoins instructif de comparer ses performances à un modèle plus performant et non fédéré tel que XGBoost.

Parallèlement, nous avons calculé l'AUC obtenue par des régressions logistiques locales, indépendamment sur chaque base, afin de disposer d'un point de comparaison avec le modèle fédéré.

Les résultats des trois bases avant et après data augmentation sont résumés dans le tableau 7 et 8. On constate que les performances globales avant augmentation restent modestes, avec un AUC maximal de 0.66 obtenu sur la base européenne. Ces résultats suggèrent que, même avec un modèle plus puissant comme XGBoost, les données disponibles présentent des limites importantes en termes de pouvoir prédictif. Cela conforte notre choix d'utiliser la méthode de data augmentation car nous observons une amélioration nette des modèles avec maintenant un AUC maximum de 0.72 pour la base française mais la plus grande amélioration est celle de la base belge passant de 0.59 à 0.70, cela montre que le data augmentation a permis de compenser le manque de données réelles et d'améliorer le modèle.

Base	AUC XGBoost	AUC Régression Logistique
freMTPL - France	0.6729	0.55
beMTPL - Belgique	0.5987	0.56
euMTPL - Europe/Italie	0.6605	0.57

TABLE 7 – Comparaison des AUC obtenus avec XGBoost et régression logistique locale avant data augmentation

Base	AUC XGBoost	AUC Régression Logistique
freMTPL - France	0.7200	0.6423
beMTPL - Belgique	0.7059	0.5907
euMTPL - Europe/Italie	0.6709	0.6436

TABLE 8 – Comparaison des AUC obtenus avec XGBoost et régression logistique locale après data augmentation

4.3 Résultats

L'analyse des résultats obtenus avec l'algorithme **FedAvg** révèle une performance contrastée selon les régions et les variables étudiées. Cette hétérogénéité se manifeste aussi bien dans les métriques globales (AUC, perte) que dans les coefficients estimés par le modèle.

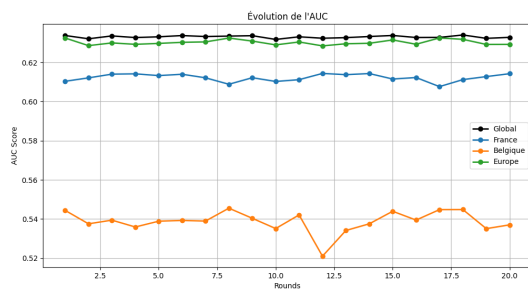


FIGURE 9 – Évolution du score AUC au fil des rounds pour FedAvg, par région

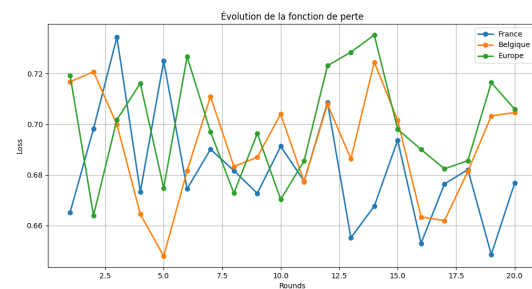


FIGURE 10 – Évolution de la fonction de perte (loss) au fil des rounds pour FedAvg

Le score AUC obtenu avec l'algorithme FedAvg reste globalement modeste, avec une valeur stabilisée autour de **0.619** dans le modèle global. Ce résultat est cohérent avec ceux obtenus dans les tableaux 7 et 8 via régression logistique locale (entre 0.55 et 0.57). En revanche, le modèle XGBoost atteint des performances bien plus élevées (jusqu'à **0.84** sur la base France), ce qui confirme que FedAvg est limité par la linéarité du modèle et l'hétérogénéité inter-clients.

Cette observation est cohérente avec les résultats de l'évaluation croisée (cf. [tableau en annexe](#)) : on observe une **amélioration de 6.44 % sur la Belgique** lorsque le modèle est issu de la France, mais une **dégradation de 7.37 %** lorsque le modèle est entraîné à partir des données belges elles-mêmes.

En parallèle, l'évolution de la *fonction de perte* met en évidence une **instabilité persistante**, en particulier pour les régions belge et européenne, traduisant un processus d'apprentissage plus difficile et une convergence potentiellement plus lente ou perturbée par l'hétérogénéité entre les clients.

L'une des principales limites de FedAvg dans notre contexte provient de la base belge, dont la taille relativement réduite rend l'utilisation de mini-batches instable. Avec environ 160 000 observations réparties sur 20 rounds d'agrégation, chaque mini-batch ne contient qu'environ 8 000 observations. Ce volume est insuffisant pour garantir un entraînement stable et efficace. Dès lors, restreindre les bases française et européenne pour équilibrer artificiellement la répartition (par exemple en imposant une règle de 33% par base) ne serait ni pertinent, ni avantageux sur le plan statistique.

Cette instabilité se manifeste par des poids locaux erratiques, entraînant une contribution peu fiable

du client belge lors de l'agrégation. En pratique, cela détériore non seulement les performances du client concerné, mais aussi celles du modèle global. À l'inverse, la base européenne, beaucoup plus volumineuse, exerce une influence prépondérante sur l'estimation des poids. Les coefficients du modèle global tendent alors à converger vers ceux du client Europe, traduisant une surreprésentation de cette base dans le modèle fédéré final. Ainsi, la contribution des petites bases se retrouve fortement diluée dans le modèle fédéré, au profit de la base dominante.

Concernant l'évolution des poids, rappelons que, pour chaque round, les points bleus, orange et verts correspondent respectivement aux coefficients locaux des bases française, belge et européenne, avant agrégation. Le point rouge représente quant à lui le coefficient global obtenu après agrégation via l'algorithme FedAvg.

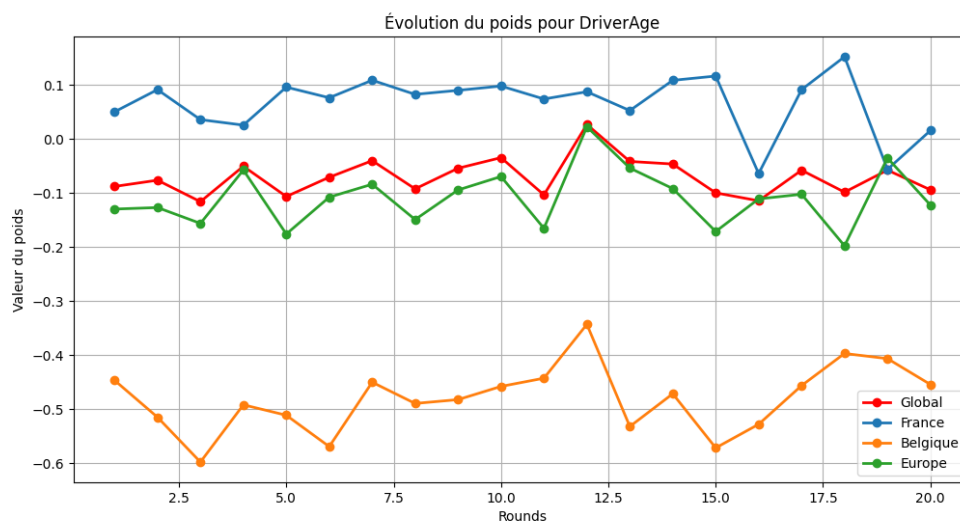


FIGURE 11 – Évolution des poids DriverAge par FedAvg

Tout d'abord, nous constatons bien le phénomène d'instabilité : l'évolution des coefficients locaux est globalement erratique, voire volatile, notamment pour la Belgique, et l'évolution des coefficients globaux est très proche de celle du client européen pour des raisons déjà évoquées précédemment. Chaque coefficient oscille autour d'une valeur cible propre à chaque client, mais cette instabilité rend leur interprétation plus délicate. À l'inverse, les coefficients globaux présentent une trajectoire beaucoup plus régulière et stable, ce qui témoigne de la capacité de l'agrégation à atténuer l'hétérogénéité des données locales. La courbe globale reste également centrée autour d'une valeur cible, mais avec une amplitude de variation nettement plus faible mais tout de même erratique.

Enfin, les valeurs des coefficients ne peuvent pas être comparées directement à celles présentées en section 3, en raison d'un changement de spécification : ici, la probabilité de sinistre a été pondérée par la variable *Exposure*. Toutefois, certaines observations méritent d'être soulignées :

- Le coefficient de la variable *Sexe* (client France) est proche de 0,6, ce qui correspond aux attentes, puisque cette variable a été introduite de manière biaisée dans ce but (cf. section *Variable d'intérêt*).
- Le coefficient de la variable *Power* (client Europe) est quant à lui très proche de zéro, alors qu'il était significativement plus élevé dans la régression logistique locale de la section 3. Ce résultat inattendu reste à ce stade inexpliqué.

Ces résultats mettent en évidence que, bien que l'algorithme FedAvg parvienne à stabiliser certains coefficients, il demeure **partiellement sous-optimal dans des contextes marqués par une forte hétérogénéité**. Les

spécificités régionales ne sont pas entièrement prises en compte, et des biais structurels viennent perturber l'estimation des coefficients.

En comparaison avec les résultats obtenus via les régressions logistiques locales, FedAvg ne surpasse pas ces dernières, à l'exception notable de la base européenne, où l'agrégation semble produire de meilleurs résultats. Ces limites ont motivé le développement d'extensions de FedAvg, mieux adaptées aux environnements hétérogènes.

5 De nouvelles méthodes d'agrégation : FedProx et FedOpt

Dans le prolongement de notre étude, il a été nécessaire de tester d'autres approches d'agrégation fédérée afin d'analyser si celles-ci permettent une amélioration significative des performances. C'est dans cette optique que les méthodes **FedProx** (Federated Proximal) et **FedOpt** (Federated Optimization) ont été implémentées.

5.1 Federated Proximal — FedProx

L'algorithme **FedProx** reprend l'architecture de communication de **FedAvg**([2]), où un serveur central envoie un modèle global aux clients, qui réalisent ensuite un entraînement local avant de renvoyer leurs mises à jour. Toutefois, FedProx se distingue par une modification essentielle apportée à la fonction de coût locale. Cette adaptation a pour objectif de mieux gérer l'*hétérogénéité des données* entre clients — un défi central dans les scénarios fédérés avec données non-i.i.d.

Afin de limiter la divergence entre les paramètres locaux w_k et le modèle global issu de FedAvg à la dernière agrégation w^{FedAvg} , FedProx introduit un terme de régularisation dit *proximal*. Ce terme incite chaque client à conserver ses poids proches de ceux du modèle global courant, réduisant ainsi l'effet de dérive locale. La fonction de coût locale devient alors :

$$F_k^{\text{Prox}}(w_k) = -\frac{1}{n_k} \sum_{i=1}^{n_k} \left[y_i^k \log \left(\sigma((x_i^k)^T w_k) f_i \right) + (1 - y_i^k) \log \left(1 - \sigma((x_i^k)^T w_k) f_i \right) \right] + \frac{\mu}{2} \|w_k - w^{\text{FedAvg}}\|_2^2 \quad (6)$$

Le paramètre $\mu > 0$ régle l'intensité de cette contrainte de proximité. Une valeur élevée de μ contraint davantage le modèle local à rester proche du modèle global, ce qui peut stabiliser l'apprentissage dans les environnements très hétérogènes.

Ce terme proximal impacte directement la règle de mise à jour des poids lors de l'entraînement local. À chaque itération locale $e \in E$, la mise à jour des poids s'effectue selon :

$$w_{t+1}^k = w_t^k - \alpha \nabla F_k(w_k, B) - \mu(w_t^k - w^{\text{FedAvg}}) \quad (7)$$

où :

- $\nabla F_k(w_k, B)$ est le gradient local de la fonction de coût standard calculé sur un mini-batch B ,
- α est le taux d'apprentissage (learning rate) local à l'itération t que l'on prendra constant dans notre cas,
- μ est le coefficient de régularisation proximal,
- w^{FedAvg} est le modèle global reçu du serveur au début du round.

À la fin de la dernière itération, une agrégation type FedAvg est réalisée pour renvoyer les poids.

Une question pertinente nous vient alors : quelle valeur de μ choisir pour avoir les meilleurs résultats ?

5.1.1 Choix Optimal du paramètre proximal μ

Pour choisir un μ optimal, il convient d'être en mesure de modifier ce paramètre par client et par itération : $\mu = \mu_{t,k}$ (client k et itération t). Dans l'algorithme FedADM [5], une extension adaptative de FedProx, le paramètre proximal $\mu_{t,k}$ est ajusté dynamiquement en fonction de la dissimilarité entre le modèle local et le modèle global. Cette approche vise à mieux gérer l'hétérogénéité des données dans l'apprentissage fédéré.

Mesure de la dissimilarité : La dissimilarité entre le modèle local du client k et le modèle global à l'itération t est mesurée par la norme euclidienne (norme L2) : $\|w_k - w_{t,\text{Global}}\|_2$

- w_k : vecteur des paramètres du modèle local du client k
- $w_{t,\text{Global}}$: vecteur des paramètres du modèle global à l'itération t

Ajustement dynamique de $\mu_{t,k}$: Le paramètre proximal $\mu_{t,k}$ est mis à jour selon la formule suivante :

$$\mu_{t,k} = \mu_{t-1,k} + \alpha (\|w_k - w_{t,\text{Global}}\| - \xi)$$

où :

- $\mu_{t-1,k}$: valeur précédente du paramètre proximal pour le client k
- α : taux d'apprentissage contrôlant la vitesse d'ajustement
- ξ : seuil prédéfini de dissimilarité

Cette stratégie permet :

- **Si** $\|w_k - w_{t,\text{global}}\| > \xi$: augmentation de $\mu_{t,k}$, renforçant la régularisation pour rapprocher le modèle local du modèle global.
- **Si** $\|w_k - w_{t,\text{global}}\| < \xi$: diminution de $\mu_{t,k}$, permettant plus de flexibilité au modèle local.

Toutefois, par manque de temps, nous n'avons pas été en mesure d'implémenter ce paramétrage adaptatif de μ .

5.2 Federated Optimization - FedOpt

FedOpt est une extension de FedAvg. Elle ajuste différemment les poids de FedAvg aux différentes itérations.

Seul l'équation 5 diffère du modèle de base. Ici, les poids actualisés à l'itération $t + 1$ dépendent maintenant de l'historique des différents poids. Ces modifications sont appelées *Méthodes Adaptatives*. Les optimiseurs utilisés dans FedOpt incluent FedAdagrad, FedAdam et FedYogi ([4]). Chacun d'eux a des caractéristiques spécifiques pour améliorer la convergence.

Posons $\Delta_t^k := w_t^k - w_t^{\text{FedOpt}}$, la différence entre les poids du client k et le modèle global agrégé lors du round précédent. On définit alors la mise à jour moyenne pondérée des clients par $\Delta_t := \sum_{k=1}^K \frac{n_k}{\sum_{j=1}^K n_j} \Delta_t^k$. Cette quantité représente la direction moyenne selon laquelle les modèles locaux s'éloignent du modèle global, et constitue la base de la mise à jour dans FedOpt.

On introduit ensuite le vecteur de moment accumulé m_t , calculé à chaque itération t selon :

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) \Delta_t$$

où :

- $\beta_1 \in [0, 1[$ est un coefficient contrôlant l'influence des mises à jour précédentes m_{t-1} par rapport à la mise à jour actuelle Δ_t .

Le vecteur m_t permet de lisser les mises à jour du modèle global, en tenant compte de l'historique, ce qui améliore la stabilité et la convergence de l'optimisation, à l'instar des méthodes comme Adam ou Yogi.

Les différentes variantes de FedOpt diffèrent par la manière dont elles estiment la variance via la mise à jour du vecteur u_t , représentant une mesure adaptative des gradients :

- **FedAdagrad** : $u_t = u_{t-1} + (\Delta_t)^2$
- **FedYogi** : $u_t = u_{t-1} - (1 - \beta_2) \cdot \text{sign}(u_{t-1} - (\Delta_t)^2) \cdot (\Delta_t)^2$
- **FedAdam** : $u_t = \beta_2 u_{t-1} + (1 - \beta_2)(\Delta_t)^2$

Le modèle global est ensuite mis à jour selon la règle suivante :

$$w_{t+1}^{\text{FedOpt}} = w^{\text{FedOpt}} - \frac{\alpha}{\sqrt{u_t + \tau}} \cdot m_t \quad (8)$$

où :

- w^{FedOpt} les poids du modèle global agrégé lors du round précédent
- α est le taux d'apprentissage côté serveur à l'itération t ,
- u_t représente une moyenne mobile (ou adaptative) des carrés des mises à jour,
- $\beta_2 \in [0, 1[$ est un paramètre d'oubli contrôlant l'influence de l'historique sur u_t ,
- τ est un terme de régularisation numérique qui évite une division instable et contrôle le degré d'adaptabilité de l'algorithme.

Enfin, notons que les mises à jour effectuées par le serveur dans FedOpt sont invariantes à une mise à l'échelle constante du taux d'apprentissage local α_t , à condition que τ soit choisi de manière cohérente.

De même que pour FedProx, quelles valeurs de β_1 , β_2 et τ choisir ?

5.2.1 Choix des Paramètres dans les Optimiseurs Adaptatifs β_1 , β_2 et τ

Étant donné que les données de validation sont souvent inaccessibles dans les environnements d'apprentissage fédéré, les auteurs de l'article *Adaptive Federated Optimization* de Reddi et al. (2021) [4] proposent une méthode alternative pour le réglage des hyperparamètres :

"Nous sélectionnons η_l , η et τ par une recherche en grille. Bien que cela soit souvent fait en utilisant des données de validation dans des environnements centralisés, de telles données sont souvent inaccessibles en apprentissage fédéré, en particulier en apprentissage fédéré inter-appareils. Par conséquent, nous ajustons les paramètres en sélectionnant ceux qui minimisent la perte d'entraînement moyenne sur les 100 dernières itérations de formation."

Ainsi, les paramètres sont choisis en minimisant la perte d'entraînement moyenne sur les 100 dernières itérations, ce qui permet une évaluation plus robuste dans des environnements où les données de validation sont limitées. Pour ce qui est des **Recommandations Pratiques** :

- **Taux d'apprentissage côté serveur (η)** : Commencer avec des valeurs telles que 0.001 ou 0.01 et ajuster en fonction de la convergence observée.
- **Taux d'apprentissage côté client (η_l)** : Souvent fixé à des valeurs plus élevées que η , comme 0.1, mais nécessite une calibration en fonction de la variabilité des données locales.
- **Paramètres β_1 et β_2** : Les auteurs fixent $\beta_1 = 0.9$ et $\beta_2 = 0.99$ pour FedAdam et FedYogi, conformément aux pratiques courantes.
- **Paramètre τ** : Fixé à 10^{-3} dans les expériences, avec une robustesse observée à travers différentes tâches.

En ce sens, nous fixerons β_1 , β_2 et τ comme recommandé.

6 Implémentation de FedProx et FedOpt et Résultats

Nous conservons les mêmes hyperparamètres que dans l'implémentation de *FedAvg*.

Les résultats obtenus avec **FedProx** montrent des performances globalement similaires à celles de FedAvg, sans réelle amélioration notable. Le score AUC global observable de FedProx (cf [figure 20](#)) atteint en moyenne 0.619 avec FedAvg et 0.618 avec FedProx, ce qui indique une quasi-équivalence en termes de pouvoir prédictif agrégé.

Malgré l'introduction d'un terme de régularisation proximal ($\mu = 0.1$) censé stabiliser les mises à jour locales, FedProx ne parvient pas à stabiliser le modèle. De plus, le client belge possède des mêmes limitations que pour la FedAvg : son AUC reste inférieur à 0.55 et présente une grande variabilité au fil des rounds, comme en témoigne la courbe FedProx AUC Belgique [figure 20](#).

Concernant FedOpt, la courbe AUC [figure 22](#) montre une progression régulière sur l'ensemble des clients, notamment pour **l'Europe** où le modèle atteint **0.6319**, soit l'une des meilleures performances observées dans notre étude. La **Belgique**, bien qu'encore en retrait, bénéficie également d'une amélioration relative (+3.92 % vs. local).

Les **courbes de perte** (client_loss) [figure 23](#) confirment ce comportement : elles suivent une trajectoire *descendante plus douce et plus stable*, avec des oscillations modérées par rapport à FedProx. Cela témoigne de l'efficacité des optimiseurs adaptatifs (FedAdam utilisé ici) dans le lissage des mises à jour globales.

L'évolution des **coefficients** sous FedOpt adopte une forme particulière : les courbes suivent des *sinusoïdes décroissantes* [figure 12](#), traduisant une stabilisation progressive avec des phases d'oscillation amortie. Ce comportement est observé sur plusieurs variables, notamment Fuel_type, DriverAge ou Density, et témoigne d'un meilleur équilibre entre adaptation locale et convergence globale. Contrairement à FedAvg où les poids fluctuent brutalement, FedOpt réussit à lisser leur évolution sans blocage prématuré.

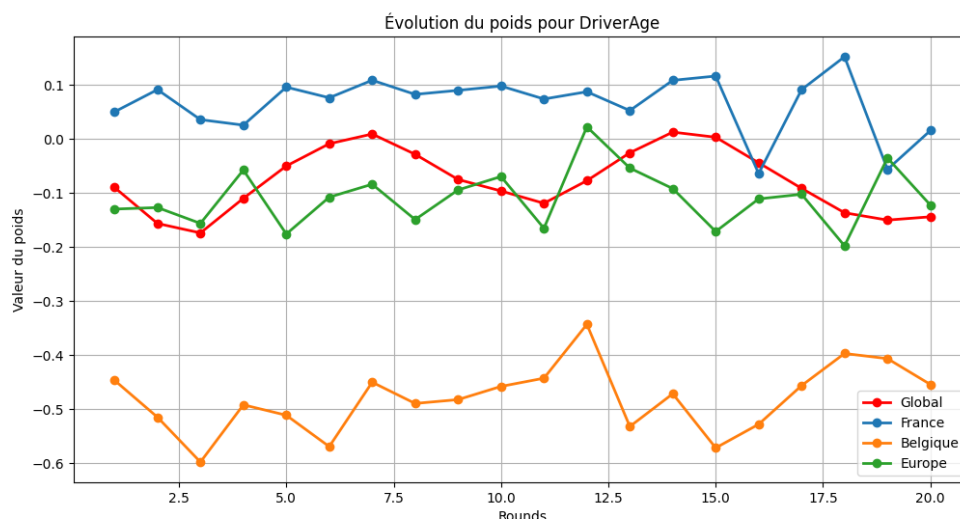


FIGURE 12 – Évolution des poids DriverAge par FedOpt (Adam)

On note également une **atténuation des biais régionaux** : par exemple, le poids de Sex reste élevé en France, mais n'est pas aussi dominant qu'en FedAvg. La variable Power, souvent neutre dans FedAvg, présente ici des valeurs plus cohérentes avec les effets locaux (notamment en Europe).

Enfin, bien que le modèle global reste encore largement influencé par le client Europe (en raison de sa taille), **FedOpt semble mieux prendre en compte les clients minoritaires**, notamment en réduisant l'écart de performance observé pour la Belgique.

7 Conclusion

Ce projet de Statistiques Appliquées nous a permis d'explorer les enjeux de l'apprentissage fédéré dans le secteur de l'assurance. Bien que les performances obtenues, notamment en termes d'AUC globale, soient inférieures aux attentes initiales, plusieurs enseignements précieux ont émergé.

Tout d'abord, il apparaît clairement que les **algorithmes d'agrégation** influencent significativement la qualité des résultats. *FedOpt* s'est distingué comme le candidat le plus performant en assurant une bonne stabilité, contrairement à *FedAvg* ou *FedProx*, dont les comportements sont plus simples mais parfois moins efficaces. Cependant, dans une optique « entreprise », ce gain de performance peut être contrebalancé par une moindre **interprétabilité**, élément central dans des environnements réglementés comme l'assurance.

En parallèle, notre étude a mis en lumière des **limites structurelles** liées à la taille et à la composition des jeux de données. La base belge, en particulier, s'est avérée trop restreinte pour supporter un grand nombre de rounds d'entraînement sans déséquilibrer la fédération.

Nous avons également observé une **perte d'information potentielle** lors de l'agrégation des poids : certaines spécificités locales peuvent être diluées dans la moyenne globale, en particulier en présence d'une **forte hétérogénéité inter-clients** (différences de distributions, d'exposition ou de taux de sinistres).

Concernant l'augmentation des données, une méthode d'*over-sampling* basée sur SMOTENC a été appliquée exclusivement sur les données d'entraînement dans le cas de la régression logistique. Cette stratégie a permis d'équilibrer les classes en générant artificiellement des observations pour la classe minoritaire. Toutefois, bien que techniquement applicable à des variables catégorielles binaires, cette approche reste peu intuitive dans notre contexte. Une alternative plus naturelle aurait été d'adopter une stratégie d'**under-sampling contrôlé**, permettant de réduire la classe majoritaire sans générer artificiellement de données, et d'agréger localement les poids de modèles entraînés sur des sous-échantillons.

En somme, ce projet met en évidence à la fois le **potentiel** et les **limites actuelles** de l'apprentissage fédéré appliqué à l'assurance. Il nécessite une adaptation fine aux caractéristiques des données, une gestion rigoureuse de l'hétérogénéité entre clients, et une réflexion sur les arbitrages entre performance, explicabilité, et équité. De nombreuses pistes restent ouvertes, notamment en matière de robustesse, de régularisation des modèles locaux, ou encore de pondération adaptative des contributions clients.

Annexe

Statistiques Descriptives - Data Augmentation

Cette section présente les statistiques descriptives après data augmentation par la méthode SMOTENC. Les résultats sont presque identiques et cela confirme bien par ailleurs la performance de SMOTENC à préserver les distributions des variables.

On observe que les proportions de sinistres correspondent bien à ceux qu'on cherchait après data augmentation mais elles restent cependant toujours déséquilibrées et donc les remarques sur les métriques de précision sont toujours valables. Les conclusions sont identiques par rapport aux bases sans data augmentation : il n'y a pas de multicollinéarité entre les variables et les distributions des variables entre les différentes bases sont hétérogènes.

Pour plus d'informations, nous vous conseillons d'aller à la partie de Statistiques descriptives(3.3).

Sinistre

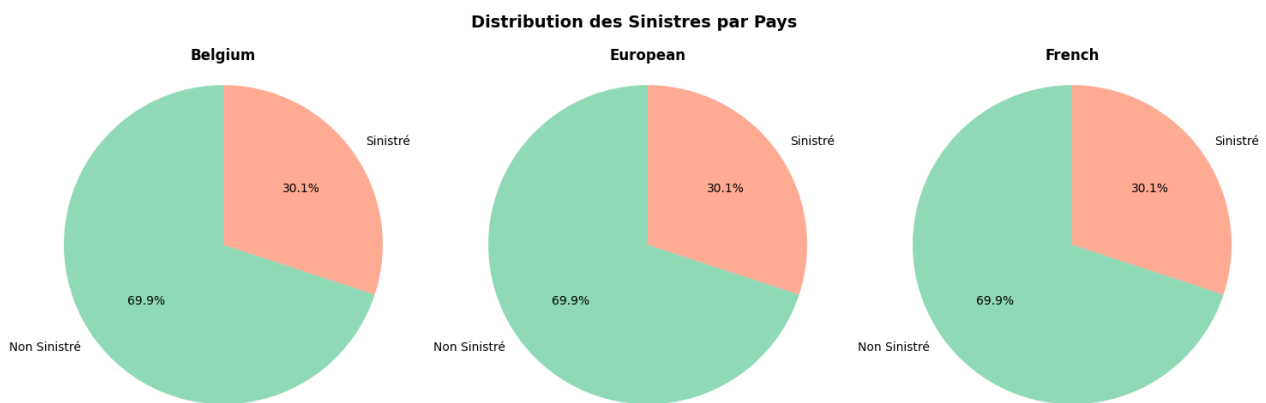
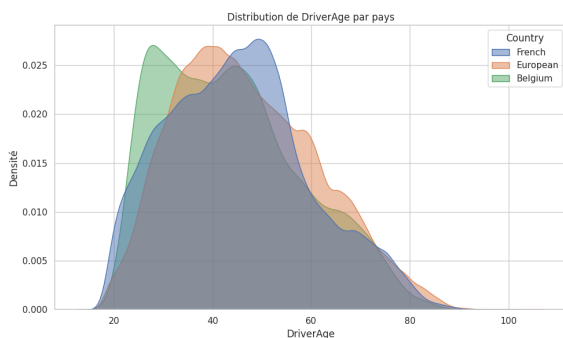
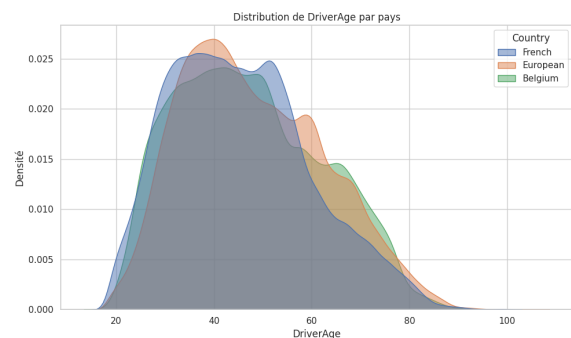


FIGURE 13 – Distribution des Sinistres après data augmentation

DriverAge



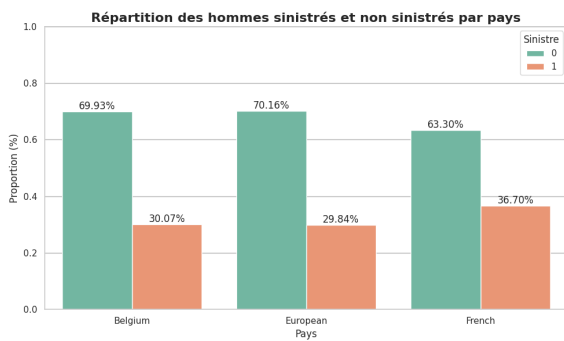
(a) Distribution de l'âge des conducteurs parmi les sinistres



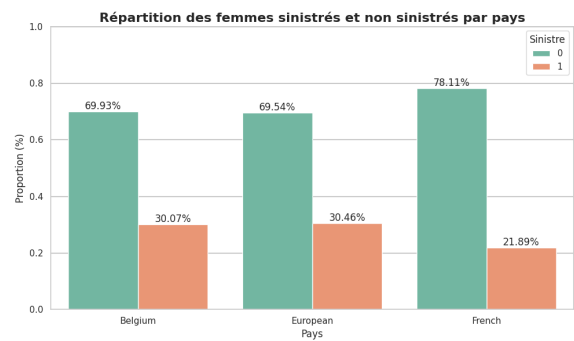
(b) Distribution de l'âge des conducteurs parmi les non sinistres

FIGURE 14 – Visualisation de la distribution de l'âge des conducteurs après data augmentation

Gender



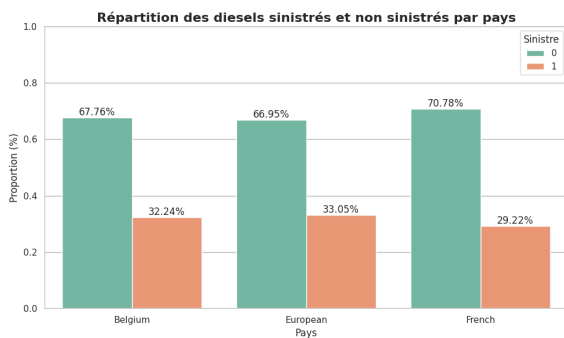
(a) Distribution de la proportion d'homme sinistré et non sinistré



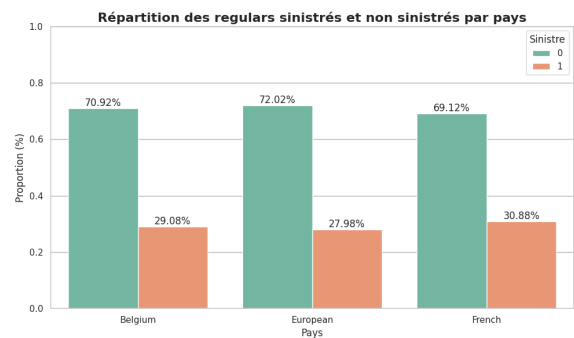
(b) Distribution de la proportion de femme sinistrée et non sinistrée

FIGURE 15 – Visualisation de la distribution du sexe des conducteurs après data augmentation

Fuel type



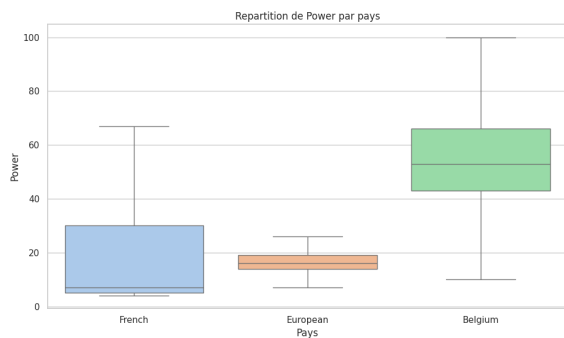
(a) Distribution du carburant Diesel selon les sinistres



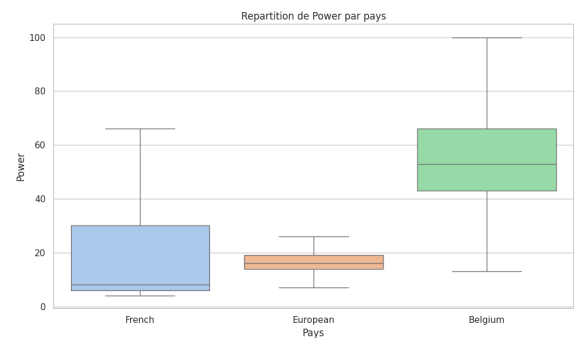
(b) Distribution du carburant Regular selon les sinistres

FIGURE 16 – Visualisation de la distribution du carburant des conducteurs après data augmentation

Power



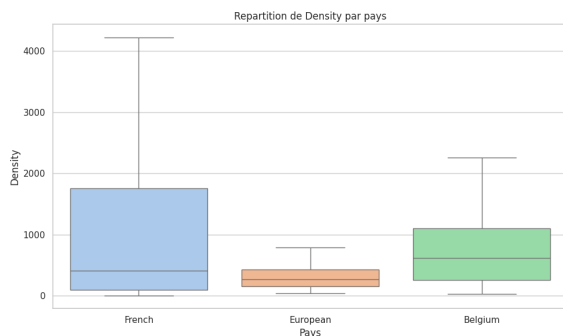
(a) Boxplot de la puissance du véhicule des sinistres



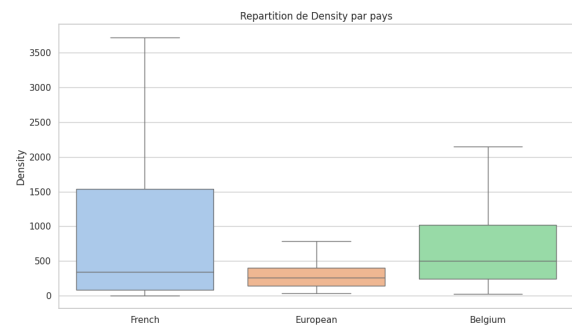
(b) Boxplot de la puissance du véhicule des non sinistres

FIGURE 17 – Visualisation de la distribution de la puissance du véhicule après data augmentation

Density



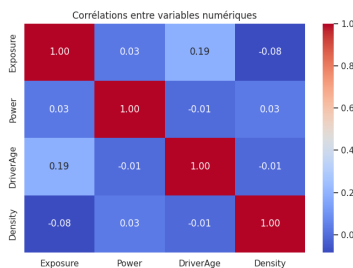
(a) Boxplot de la densité de population des sinistres



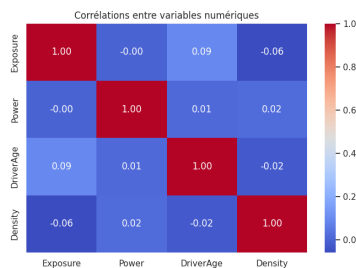
(b) Boxplot de la densité de population des sinistres

FIGURE 18 – Visualisation de la distribution de la densité de population des sinistres après data augmentation

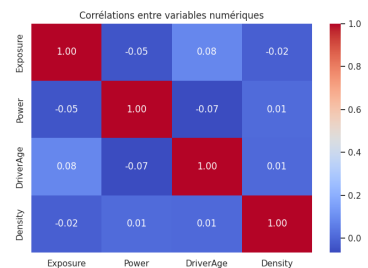
Corrélation des variables



(a) France



(b) Belge



(c) Européenne

FIGURE 19 – Matrice de corrélation de chaque base après data augmentation

TABLE 9 – Résultats de l'analyse du V de Cramér entre variables catégorielles après data augmentation

Jeu de données	Paire de variables	Cramer's V	Force de corrélation	p-value	Significatif
French	Fuel type - Sinistre	0.018	Faible	0.000	Oui
French	Fuel type - Sex	0.007	Faible	0.000	Oui
French	Sinistre - Sex	0.161	Modérée	0.000	Oui
Belgium	Fuel type - Sinistre	0.032	Faible	0.000	Oui
Belgium	Fuel type - Sex	0.109	Modérée	0.000	Oui
Belgium	Sinistre - Sex	0.000	Faible	0.09895	Non
European	Fuel type - Sinistre	0.054	Faible	0.000	Oui
European	Fuel type - Sex	0.151	Modérée	0.000	Oui
European	Sinistre - Sex	0.007	Faible	0.000	Oui

Résultats de la régression logistique sur les bases

Variable	Coef	Std. Err.	z	P> z	[0.025, 0.975]
Constante	-0.6425	0.017	-37.625	0.000	[-0.676, -0.609]
Power	0.5369	0.060	8.941	0.000	[0.419, 0.655]
DriverAge	-1.0614	0.026	-40.649	0.000	[-1.113, -1.010]
Fuel_type	0.1678	0.010	16.062	0.000	[0.147, 0.188]
Density	0.7202	0.032	22.724	0.000	[0.658, 0.782]
Sex	-0.0382	0.011	-3.418	0.001	[-0.060, -0.016]

TABLE 10 – Résultats de la régression logistique sur la base belge

Variable	Coef	Std. Err.	z	P> z	[0.025, 0.975]
Constante	-0.8672	0.006	-149.357	0.000	[-0.879, -0.856]
Power	1.9634	0.325	6.045	0.000	[1.327, 2.600]
DriverAge	-0.3254	0.008	-40.563	0.000	[-0.341, -0.310]
Fuel_type	0.2330	0.003	81.959	0.000	[0.227, 0.239]
Density	0.1480	0.005	30.865	0.000	[0.139, 0.157]
Sex	-0.0486	0.003	-18.389	0.000	[-0.054, -0.043]

TABLE 11 – Résultats de la régression logistique sur la base européenne

Variable	Coef	Std. Err.	z	P> z	[0.025, 0.975]
Constante	-1.1779	0.005	-225.763	0.000	[-1.188, -1.168]
Power	-0.8288	0.012	-66.505	0.000	[-0.853, -0.804]
DriverAge	0.1281	0.010	12.389	0.000	[0.108, 0.148]
Fuel_type	-0.0331	0.004	-9.034	0.000	[-0.040, -0.026]
Density	0.1745	0.011	15.658	0.000	[0.153, 0.196]
Sex	0.6671	0.004	179.147	0.000	[0.660, 0.674]

TABLE 12 – Résultats de la régression logistique sur la base française

FedAvg - Évaluation croisée, AUC/Loss, Évolution des poids

Source	Données test	AUC local	AUC FedAvg	Amélioration (%)
France	France	0.6242	0.6107	-2.16
France	Belgique	0.5063	0.5389	+6.44
France	Europe	0.6201	0.6298	+1.56
Belgique	France	0.5968	0.6107	+2.33
Belgique	Belgique	0.5818	0.5389	-7.37
Belgique	Europe	0.6379	0.6298	-1.27
Europe	France	0.5980	0.6107	+2.13
Europe	Belgique	0.5669	0.5389	-4.94
Europe	Europe	0.6371	0.6298	-1.15

TABLE 13 – Évaluation croisée FedAvg pour chaque région source et cible.



FedProx - Évaluation croisée, AUC/Loss, Évolution des poids

Source	Données test	AUC Local	AUC FedProx	Amélioration (%)
France	France	0.6242	0.6108	-1.34%
France	Belgique	0.5063	0.5384	+3.22%
France	Europe	0.6201	0.6298	+0.97%
Belgique	France	0.5968	0.6108	+1.41%
Belgique	Belgique	0.5818	0.5384	-4.34%
Belgique	Europe	0.6379	0.6298	-0.81%
Europe	France	0.5980	0.6108	+1.28%
Europe	Belgique	0.5669	0.5384	-2.85%
Europe	Europe	0.6371	0.6298	-0.74%

TABLE 14 – Évaluation croisée FedProx pour chaque région source et cible.

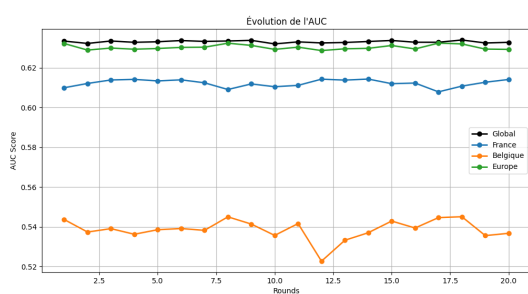


FIGURE 20 – Évolution du score AUC au fil des rounds pour FedProx, par région

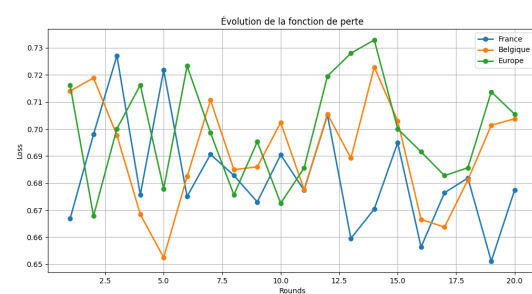


FIGURE 21 – Évolution de la fonction de perte (loss) au fil des rounds pour Fedprox



FedOpt - Évaluation croisée, AUC/Loss, Évolution des poids

Source	Données Test	AUC Local	AUC FedOpt	Amélioration(%)
France	France	0.6242	0.6098	-1.44%
France	Belgique	0.5063	0.5455	+3.92%
France	Europe	0.6201	0.6319	+1.18%
Belgique	France	0.5968	0.6098	+1.31%
Belgique	Belgique	0.5818	0.5455	-3.63%
Belgique	Europe	0.6379	0.6319	-0.60%
Europe	France	0.5980	0.6098	+1.18%
Europe	Belgique	0.5669	0.5455	-2.14%
Europe	Europe	0.6371	0.6319	-0.53%

TABLE 15 – Évaluation croisée FedOpt pour chaque région source et cible.

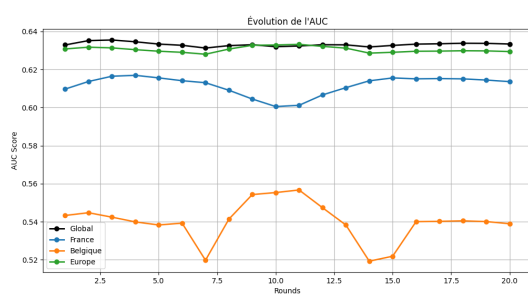


FIGURE 22 – Évolution du score AUC au fil des rounds pour FedOpt, par région

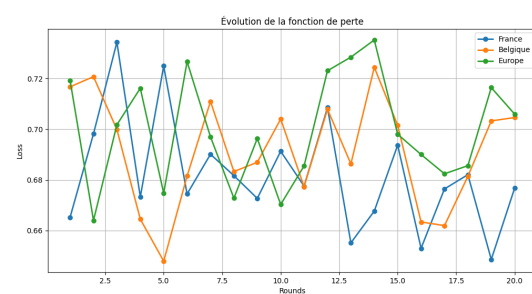


FIGURE 23 – Évolution de la fonction de perte (loss) au fil des rounds pour FedOpt



Références

- [1] Tian LI et al. « A Unified Linear Speedup Analysis of Federated Averaging and Nesterov FedAvg ». In : *arXiv preprint arXiv:2007.05690* (2020). URL : <https://arxiv.org/pdf/2007.05690>.
- [2] Tian LI et al. « FedProx: Federated Optimization in Heterogeneous Networks ». In : *arXiv preprint arXiv:1812.06127* (2020). URL : <https://flower.ai/docs/baselines/fedprox.html>.
- [3] H Brendan MCMAHAN et al. « Federated Optimization in Heterogeneous Networks ». In : *Proceedings of the 2020 MLSys Conference*. 2020. URL : https://proceedings.mlsys.org/paper_files/paper/2020/file/1f5fe83998a09396ebe6477d9475ba0c-Supplemental.pdf.
- [4] Sashank J REDDI et al. « Adaptive Federated Optimization ». In : *arXiv preprint arXiv:2003.00295* (2020). URL : <https://arxiv.org/pdf/2003.00295>.
- [5] Ning YANG et al. *FedADM: Adaptive Federated Learning via Dissimilarity Measure*. 2024. URL : <https://openreview.net/forum?id=IsHWcsk4Fz>.
- [6] Qiang YANG et al. « FedOpt: Towards Communication Efficiency and Privacy Preservation in Federated Learning ». In : *Applied Sciences* 10.8 (2020), p. 2864. URL : <https://www.mdpi.com/2076-3417/10/8/2864>.