



# ***Les accidents de la Route en France***

Projet Fil Rouge - **Partie II**

---

Modélisation et  
Interprétation des prédictions

# 01. Contexte et Objectifs

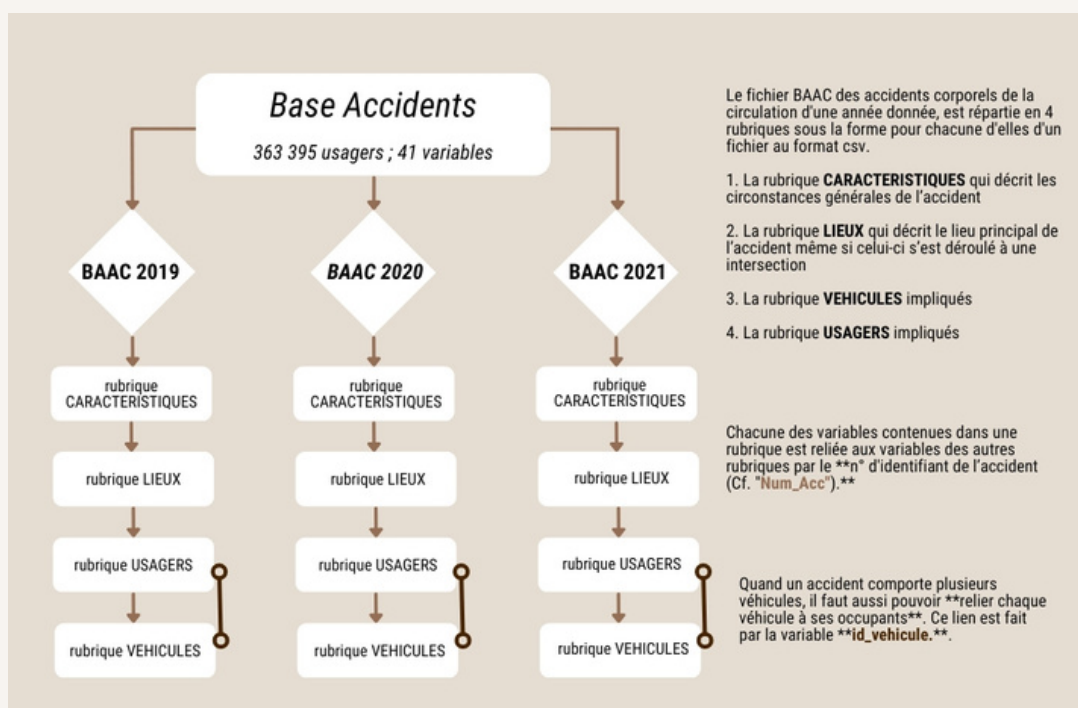
L'Observatoire national interministériel de la sécurité routière met à disposition chaque année depuis 2005, des bases de données des accidents corporels de la circulation routière.

Pour chaque accident corporel (soit un accident survenu sur une voie ouverte à la circulation publique, impliquant au moins un véhicule et ayant fait au moins une victime ayant nécessité des soins), des saisies d'information décrivant l'accident sont effectuées par l'unité des forces de l'ordre (police, gendarmerie, etc.) qui est intervenue sur le lieu de l'accident.

Ces saisies sont rassemblées dans une fiche intitulée bulletin d'analyse des accidents corporels. L'ensemble de ces fiches constitue le fichier national des accidents corporels de la circulation dit « Fichier BAAC » administré par l'Observatoire national interministériel de la sécurité routière "ONISR".

Les bases de données, extraites du fichier BAAC, répertorient **l'intégralité des accidents corporels de la circulation, intervenus durant une année précise en France métropolitaine, dans les départements d'Outre-mer (Guadeloupe, Guyane, Martinique, La Réunion et Mayotte depuis 2012) et dans les autres territoires d'outre-mer** (Saint-Pierre-et-Miquelon, Saint-Barthélemy, Saint-Martin, Wallis-et-Futuna, Polynésie française et Nouvelle-Calédonie; disponible qu'à partir de 2019 dans l'open data) avec une description simplifiée.

Cela comprend des informations de localisation de l'accident, telles que renseignées ainsi que des informations concernant les caractéristiques de l'accident et son lieu, les véhicules impliqués et leurs victimes.



# Objectifs

---

**L'objectif de ce projet est de prédire la gravité de l'ensemble des accidents routiers en France intervenus entre 2019 et 2021.**



Une première étape est d'étudier et appliquer des méthodes pour nettoyer le jeu de données.

Une fois le jeu de données propre, une deuxième étape est d'extraire les caractéristiques qui semblent être pertinentes pour estimer la gravité des accidents.

Ensuite, à partir de ses résultats, l'objectif est de travailler un scoring des zones à risque en fonction des informations météorologiques, l'emplacement géographique (coordonnées GPS, images satellite, ...) ...

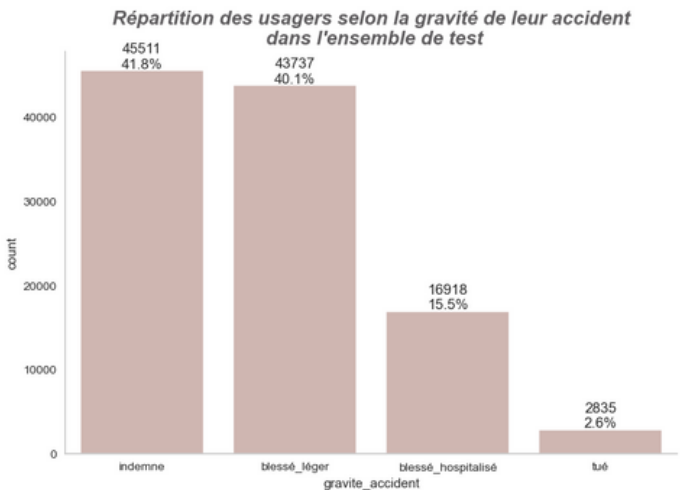
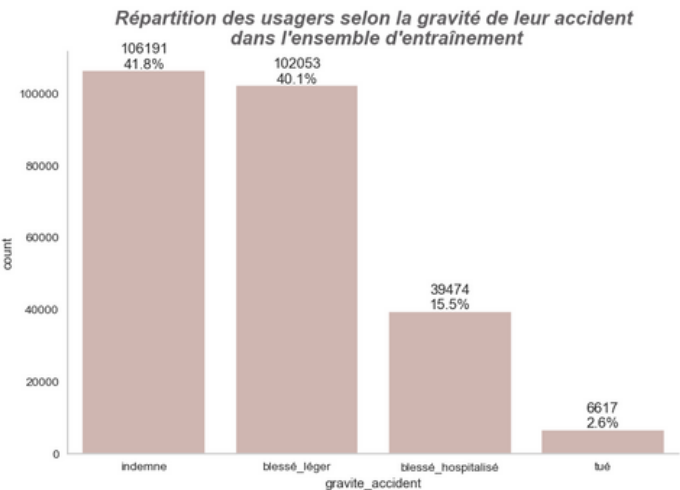
Une fois l'entraînement du modèle effectué, nous allons comparer notre modèle avec les données historiques.

## Objectif de la partie II

***L'objectif de ce second rapport est d'effectuer la modélisation de la gravité des accidents de la route en France et d'interpréter les prédictions obtenues afin de répondre à la question "Pourquoi le modèle a-t-il prédit cette valeur ?".***



# 02. En cours de construction.



Performances Globales							
Modèles	Paramètres Optimisés	Balanced Accuracy	Weighted Precision	Weighted Recall	Weighted F1-score	Average AUC	Score Géolocal
XGBoost Cost-Sensitive Learning	Non	55%	64%	58%	60%	82%	69%
XGBoost SMOTETomek	Non	51%	63%	59%	60%	80%	69%
XGBoost ClusterCentroids	Non	28%	55%	6%	5%	57%	16%
XGBoost Cost-Sensitive Learning	hgboost	52%	64%	62%	63%	83%	71%
XGBoost Cost-Sensitive Learning	GridSearchCV	51%	63%	61%	62%	81%	70%
BalancedRandomForestClassifier	Non	54%	63%	56%	58%	80%	68%
BalancedRandomForestClassifier	RandomizedSearchCV	54%	64%	56%	59%	82%	68%
LightGBMClassifier	Optuna	54%	64%	60%	61%	82%	70%

