**SD-701 : Projet de Data Mining**

*Analyse d’une base de données des accidents corporels de la circulation routières*

# Présentation du projet

Dans le cadre du projet SD 701 nous souhaitons analyser une base de données des accidents corporels de la circulation routières.

En appliquant différentes méthodes de DataMining sur un ensemble de jeux de données relatifs aux accidents de la route nous souhaitons déterminer les facteurs qui influencent la gravité d’un accident.

Le contexte de l’étude pourrait être la suivante :

Il s’avère que la plupart des personnes impliquées dans un accident de la route ne sont pas blessés ou ne présentent que des blessures artificielles. Cependant, il se peut que des personnes gravement blessées nécessitent une prise en charge à l’hôpital. Dans une situation de crise sanitaire, où tous les hôpitaux se retrouvent en surcharge de patients, il peut être utile de diminuer les accidents graves admis en soins intensifs. Pour ce faire, une campagne de prévention s’appuyant sur les facteurs influençant la gravité pourrait aider a diminuer le nombre d’accidents graves.

Pour connaître les facteurs influençant la gravité d’un accident, l’option retenu serait de diminuer la probabilité qu’une personne ait un accident sachant que cette dernière à déjà eu un accident. (Formule de Bayes). En ce sens connaître les facteurs influençant la gravité des accidents permettraient des diminuer la probabilité qu’une personne soit hospitalisée.

# Objectif du projet

L'objectif de ce projet est de déterminer les facteurs influençant la gravité des accidents. Sur la base de cette analyse l’agence régional de santé pourra alors fixer les orientations de sa future campagne nationale de lutte contre les accidents de la route.

# Présentation du « Dataset »

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 BAAC1" 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 ainsi que les départements d’Outre-mer (Guadeloupe, Guyane, Martinique, La Réunion et Mayotte depuis 2012) 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.

**Description des bases de données annuelles des accidents corporels de la circulation routière - Années de 2005 à 2018**

Source Dataset : data.gouv.fr (ministère de l’intérieur)

* <https://www.data.gouv.fr/fr/datasets/bases-de-donnees-annuelles-des-accidents-corporels-de-la-circulation-routiere-annees-de-2005-a-2019/>

Notre base de données est composée de **4** fichiers csv :

* ***Caractéristique.csv*** (58441 lignes)
  + Décrit les circonstances générales de l’accident (types de collisions, luminosité, date)
* ***Lieux.csv*** (58441 lignes)
  + Décrit le lieu principal de l’accident (catégorie route, nb de voies, régime de circulation, surface de la chaussée,…)
* ***Véhicules.csv*** (100711 lignes)
  + Décrit les véhicules impliqués dans l’accident (n° plaque immatriculations, type de véhicule, localisation du choc, manœuvre,…)
* ***Usagers.csv*** (132978 lignes)
  + Décrit les usagers impliqués dans l’accident (place de l’usager dans le véhicule, gravité, trajet de l’usager,…)

Chacune des variables contenues dans une rubrique est reliée aux variables des autres

rubriques. Le n° d'identifiant de l’accident (Cf. "***Num\_Acc***") présent dans ces 4 rubriques permet

d'établir un lien entre toutes les variables qui décrivent un accident. Quand un accident comporte

plusieurs véhicules, il est également possible de relier chaque véhicule à ses occupants. Ce lien est fait par la variable ***Num\_veh***.