

Prédiction de la gravité des accidents corporels routiers

Résumé

Dans ce compte-rendu, vous trouverez une étude sur les accidents corporels routiers de 2019 en France. Nous avons réalisé une étude descriptive des usagers impliqués dans ces accidents ainsi que la création d'un modèle de prédiction de la gravité de ces accidents.

Table des matières

1	Introduction	2
2	Études des accidents corporels routiers de 2019 en France	3
3	Prédiction de la gravité des accidents corporels routiers	4
4	Conclusion	5
A	Annexes	6
A.1	Jeux de données utilisés	6
A.2	Préparation des données	6
A.2.1	Jointure des 4 bases	6
A.2.2	Choix des variables	6
A.3	Statistiques descriptives sur les usagers	7
A.3.1	Statistiques globales	7
A.3.2	Croisement de la gravité avec d'autres variables	8
A.4	Classification des individus	12
A.4.1	Choix du nombre de variables explicatives	12
A.4.2	Méthode de choix des données train/test	12
A.4.3	Méthode de régression logistique	13
A.4.4	Méthode XGBoost	14
A.4.5	Méthode Random Forest	15
A.5	Dictionnaire des variables	16

1 Introduction

Enjeux majeurs des décisions politiques, les problèmes liés à la sécurité routière sont au cœur de tous les débats depuis maintenant quatre décennies.

Grâce aux innovations dans la conception et l'utilisation des voitures¹, couplées à de nombreuses mesures de sécurité², d'incontestables résultats ont été obtenus pour limiter le nombre et l'impact des accidents routiers dans notre vie quotidienne. Ainsi, par exemple, le nombre de tués sur la route est passé de 12510 en 1980 à 3480 en 2019³. Toutefois, ces mesures montrent leur limite en termes d'efficacité : le nombre de tués se stabilise par exemple autour de 3400/3500.

Notre étude porte sur les facteurs influençant la gravité des accidents corporels routiers. Nous allons essayer d'élaborer un modèle permettant de prédire la gravité d'un accident.

Pour cela, nous avons choisi un ensemble de base de données issues du site DataGouv sur les accidents corporels de la circulation routière de 2019. Les données utilisées sont présentées en annexe A.1.

Dans un premier temps, nous allons donc vous présenter une étude statistique des accidents corporels routiers en 2019, puis nous tenterons de mettre en place un modèle de prédiction de notre variable gravité.

1. On citera la ceinture de sécurité, les airbags, les renforts latéraux parmi tant d'autres.

2. On citera l'abaissement drastique du taux d'alcoolémie, l'instauration du permis à points, l'alourdissement du régime des contraventions, la politique de généralisations des radars et l'abaissement de la vitesse de 90 à 80km/h sur le réseau secondaire.

3. Statistiques de la sécurité routière

2 Études des accidents corporels routiers de 2019 en France

Dans un premier temps, nous avons étudié la répartition des usagers impliqués dans un accident corporel routier en 2019. Nous avons tout d'abord, étudié la répartition des usagers selon la gravité. Cette répartition est illustrée dans la table 1. Ensuite, nous avons étudié la répartition des usagers selon leur catégorie d'usagers. Cette répartition est illustrée dans la table 2. Puis nous avons étudié la répartition des accidents sur l'année 2019. Cette répartition est visible dans la figure ci-dessous :

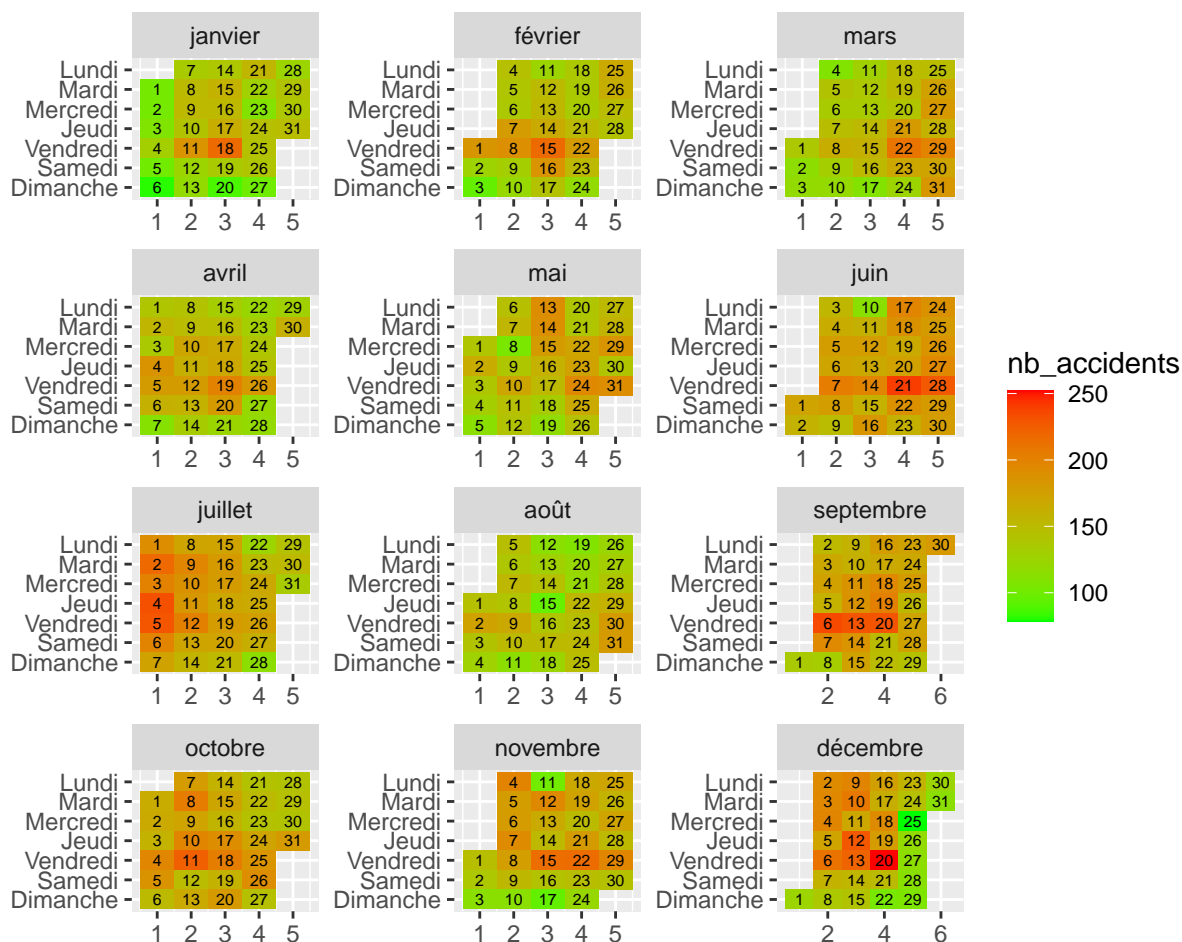


FIGURE 1 – Effectifs des accidents par jour sur l'année 2019

Source des données : Par les auteurs selon les données de 2019 du ministère de l'Intérieur.

Cohorte : 57855 accidents.

Lecture : Il y a eu 252 accidents le 20 décembre 2019.

Ensuite, nous avons étudié la répartition des usagers selon la gravité croisée avec d'autres variables telles que la catégorie de l'usager, le type de trajet effectué, l'état de la surface, les conditions atmosphériques, etc.

La répartition des individus selon la gravité et leur catégorie est disponible dans la figure 2.

3 Prédiction de la gravité des accidents corporels routiers

4 Conclusion

A Annexes

A.1 Jeux de données utilisés

Les jeux de données utilisés sont disponibles sur DataGouv. Ces jeux de données décrivent les accidents corporels routiers ayant eu lieu en 2019.

Nous disposons de quatre bases de données :

- Base caractéristiques : correspondant aux caractéristiques générales de l'accident
- Base lieux : correspondant aux caractéristiques du lieu de l'accident
- Base véhicule : correspondant aux caractéristiques du véhicule prenant part à l'accident
- Base usagers : correspondant aux caractéristiques des usagers faisant partie de l'accident

A.2 Préparation des données

A.2.1 Jointure des 4 bases

Nous réalisons des jointures entre les différentes bases sur l'identifiant de l'accident mais également sur l'identifiant des véhicules impliqués. Nous n'avons finalement plus qu'un jeu de données à étudier dans lequel chaque ligne correspond à un usager impliqué dans un accident corporel routier en France en 2019.

A.2.2 Choix des variables

Cependant, en réalisant notre jointure, nous obtenons une base de données avec une soixantaine de variables. Nous choisissons donc de réaliser un premier tri sur ces variables et de garder les variables suivantes :

- L'identifiant de l'accident,
- La date,
- L'heure et minute,
- La luminosité,
- Le département,
- La commune,
- Si l'accident a eu lieu en agglomération ou non,
- Le type d'intersection,
- Les conditions atmosphériques,
- Le type de collision,
- La catégorie de la route,
- Le régime de circulation,
- Le nombre de voies,
- La présence d'une voie particulière,
- Le profil de la route,
- Le plan de la route,
- L'état de la surface,
- Les infrastructures mises en place,
- La situation de l'accident,
- La vitesse maximale autorisée,
- L'identifiant du véhicule,
- La gravité,
- La catégorie de l'utilisateur,
- Le sexe de l'utilisateur,
- Le type de trajet effectué,
- La catégorie du véhicule,
- L'obstacle fixe rencontré,

- L'obstacle mobile rencontré,
- Le type de manoeuvre effectué,
- Le moteur du véhicule.

A.3 Statistiques descriptives sur les usagers

A.3.1 Statistiques globales

A.3.1.1 Répartition des usagers par gravité

TABLE 1 – Effectifs des usagers par gravité

Var1	Freq
Indemne	54529
Tué	3413
Blessé hospitalisé	20329
Blessé léger	52586

A.3.1.2 Répartition des usagers par catégorie d'usagers

TABLE 2 – Effectifs des usagers par catégorie d'usagers

Var1	Freq
Conducteur	95854
Passager	23956
Piéton	11047

A.3.2 Croisement de la gravité avec d'autres variables

A.3.2.1 Répartition des usagers selon leur catégorie et la gravité

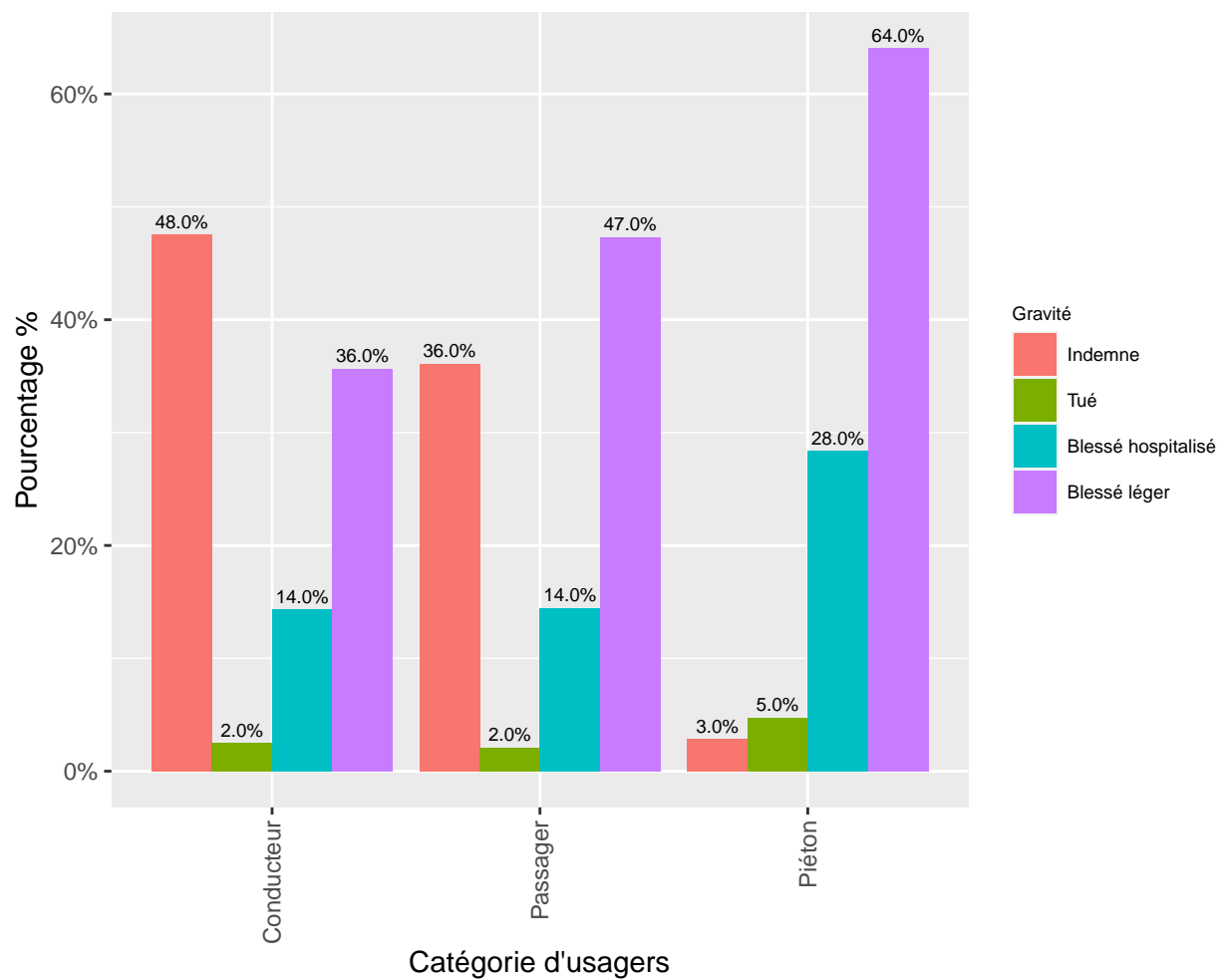


FIGURE 2 – Graphiques à bâtons sur les profils lignes, regroupement par catégorie d'usagers

Source des données : Par les auteurs selon les données de 2019 du ministère de l'Intérieur.

Cohorte : 130857 usagers.

Lecture : 64% des piétons sont des blessés légers.

A.3.2.2 Répartition des usagers selon leur trajet et la gravité

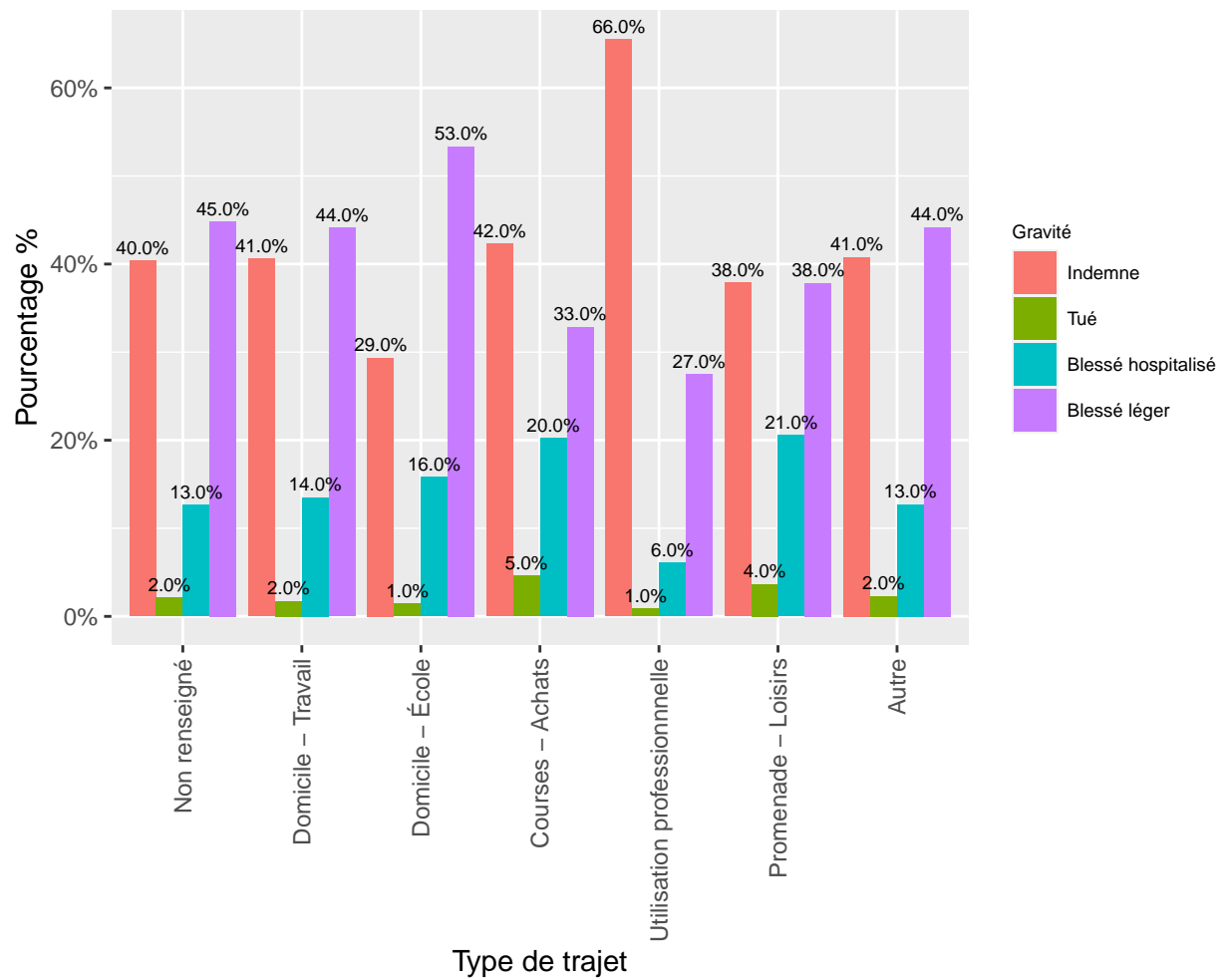


FIGURE 3 – Graphiques à bâtons sur les profils lignes, regroupement par type de trajet

Source des données : Par les auteurs selon les données de 2019 du ministère de l'Intérieur.

Cohorte : 130857 usagers.

Lecture : 65% des usagers ayant eu un accident lors d'un trajet d'utilisation professionnelle en ressortent indemnes.

A.3.2.3 Répartition des usagers selon la condition atmosphérique et la gravité

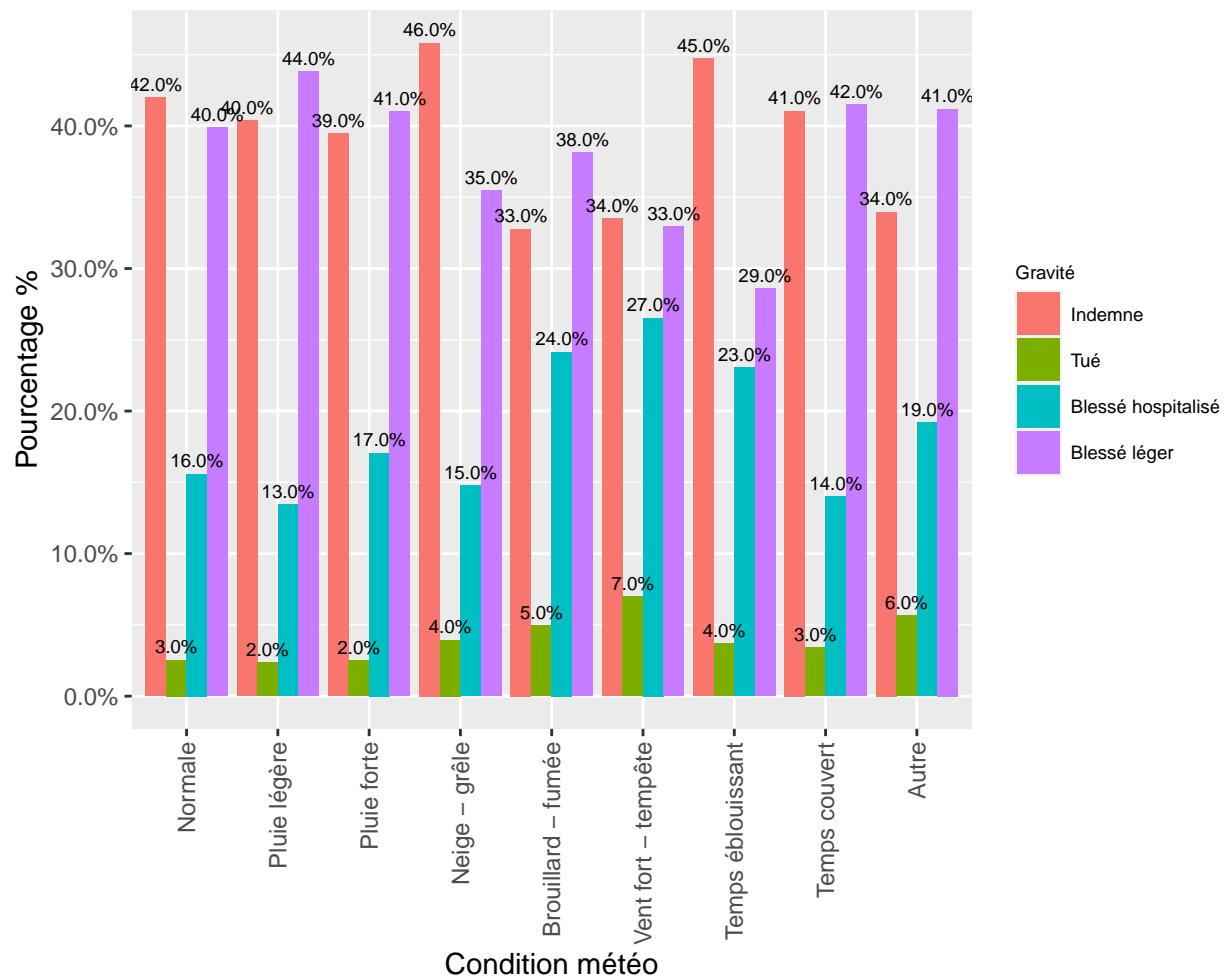


FIGURE 4 – Graphiques à bâtons sur les profils lignes, regroupement par type de trajet

Source des données : Par les auteurs selon les données de 2019 du ministère de l'Intérieur.

Cohorte : 130857 usagers.

Lecture : 42% des usagers ayant eu un accident alors que la condition atmosphérique était normale en ressortent indemnes.

A.3.2.4 Répartition des usagers selon la catégorie de route et la gravité

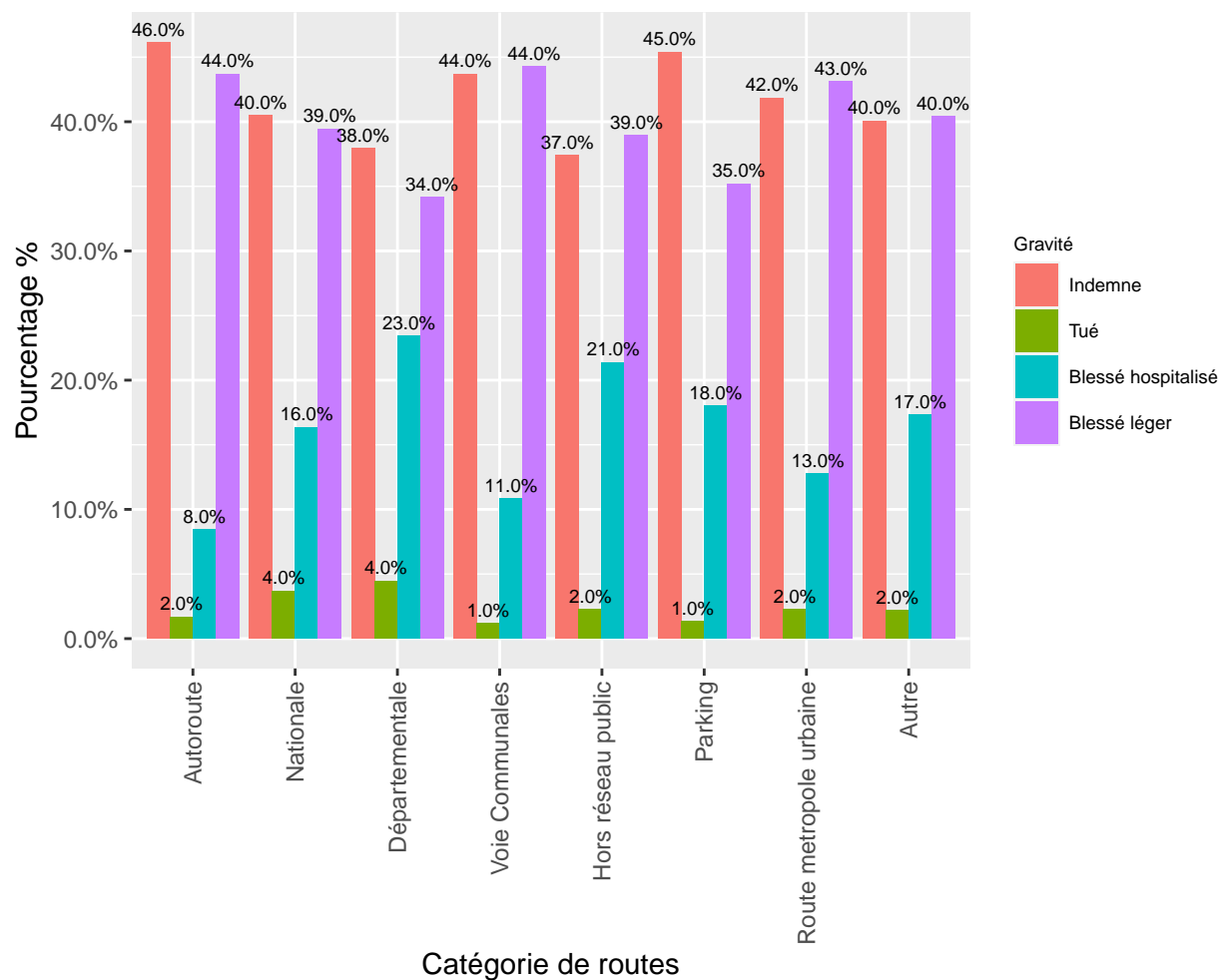


FIGURE 5 – Graphiques à bâtons sur les profils lignes, regroupement par type de trajet

Source des données : Par les auteurs selon les données de 2019 du ministère de l'Intérieur.

Cohorte : 130857 usagers.

Lecture : 42% des usagers ayant eu un accident alors que la condition atmosphérique était normale en ressortent indemnes.

A.3.2.5 Répartition des usagers selon la surface et la gravité

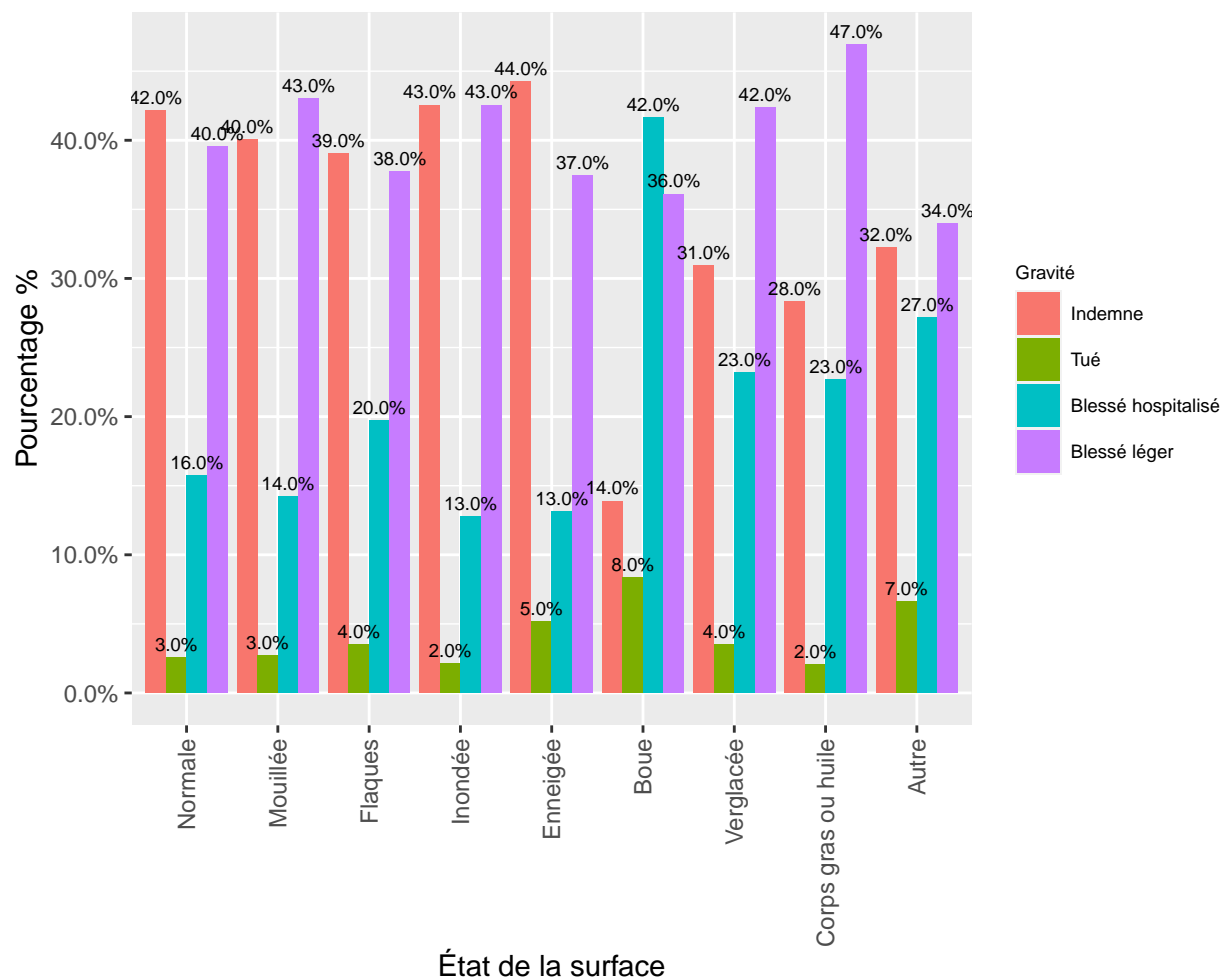


FIGURE 6 – Graphiques à bâtons sur les profils lignes, regroupement par type de trajet

Source des données : Par les auteurs selon les données de 2019 du ministère de l'Intérieur.

Cohorte : 130857 usagers.

Lecture : 42% des usagers ayant eu un accident alors que la condition atmosphérique était normale en ressortent indemnes.

A.4 Classification des individus

A.4.1 Choix du nombre de variables explicatives

A.4.2 Méthode de choix des données train/test

En vue de l'apprentissage des nos algorithmes de machine learning pour la prédiction, il est essentiel d'utiliser une méthode qui va choisir les données d'apprentissage et les données de test.

Ici nous divisons aléatoirement notre base en 2, le tableau ci-dessous décrit cette division.

TABLE 3 – Répartition des observation en vue de l'apprentissage

Base	Nb_obs	Pourcentage
train	91599	70
test	39258	30

A.4.3 Méthode de régression logistique

La méthode de régression logistique multinomiale donne une erreur de prédiction de $r\%$ déterminée par cross-validation. On représente la répartition des bonnes et mauvaises prédictions sur le graphe suivant.

A.4.4 Méthode XGBoost

La méthode XGBoost donne un taux d'erreur de 0.36%.

On représente la répartition des bonnes et mauvaises prédictions sur le graphe suivant.

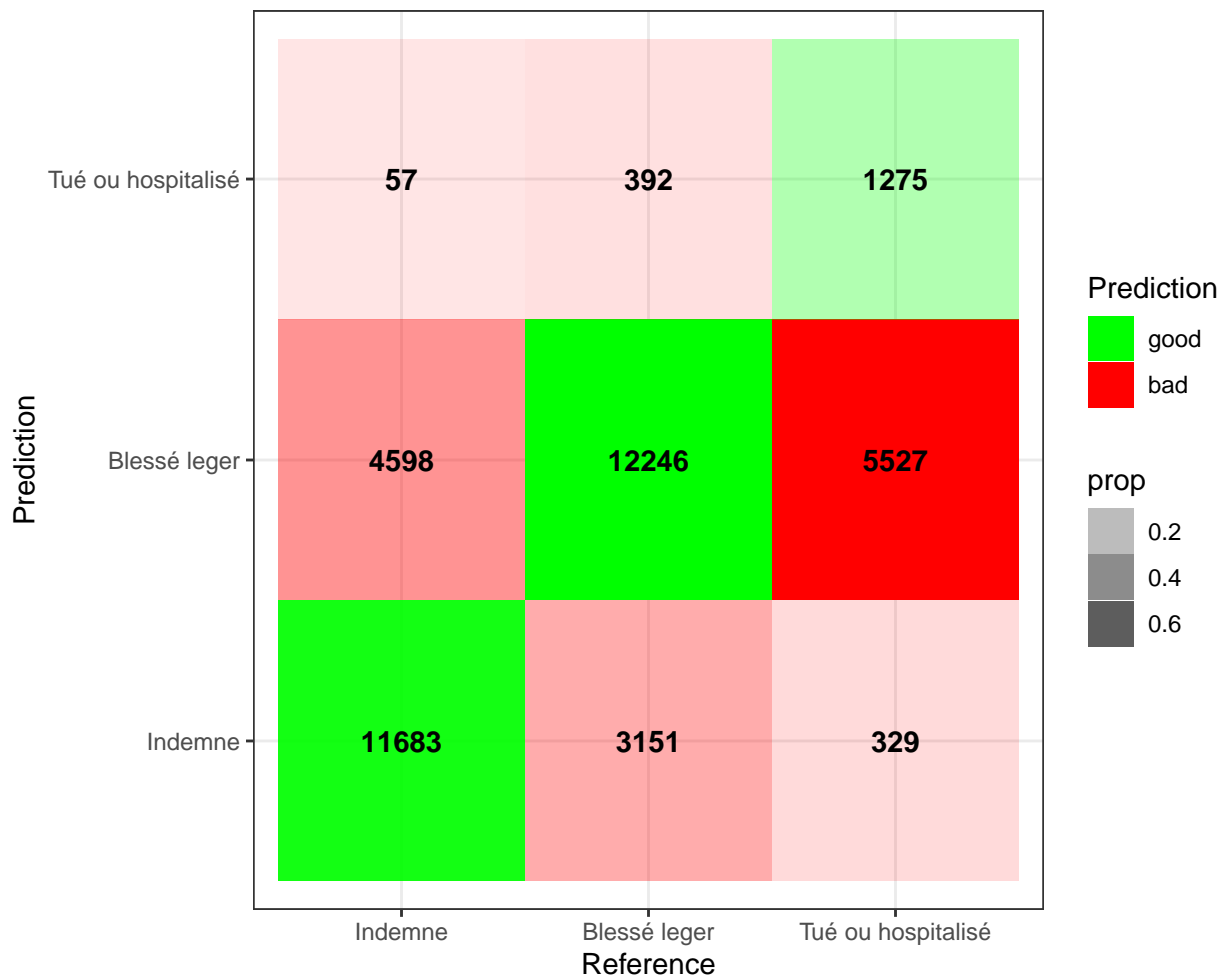


FIGURE 7 – Matrice de confusion représentant les résultats de la prédiction XGBoost

Source des données : Par les auteurs selon les données de 2019 du ministère de l'Intérieur.

Lecture : Le modèle prédictif donne 13479 bonnes prédictions pour la modalité "Indemne."

A.4.5 Méthode Random Forest

La méthode Random Forest donne un taux d'erreur de 0.33%.

On représente la répartition des bonnes et mauvaises prédictions sur le graphe suivant.

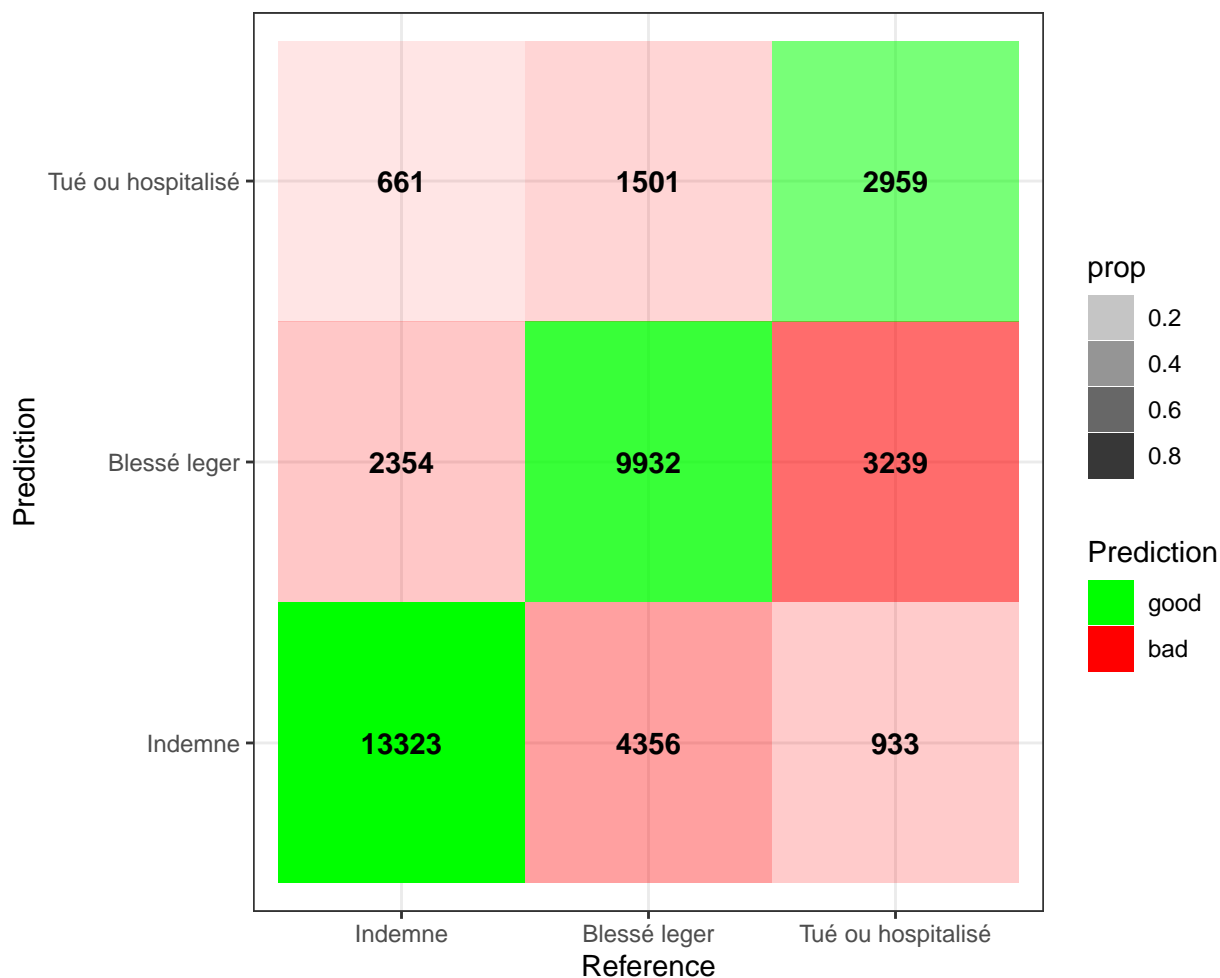


FIGURE 8 – Matrice de confusion représentant les résultats de la prédiction Random Forest

Source des données : Par les auteurs selon les données de 2019 du ministère de l'Intérieur.

Lecture : Le modèle prédictif donne 13323 bonnes prédictions pour la modalité "Indemne."

A.5 Dictionnaire des variables

— Base caractéristiques :

- Num_acc : Numéro d'identifiant de l'accident
- Jour mois : Jour et mois de l'accident
- Hrmn : Heures et minutes de l'accident
- Lum : Conditions d'éclairage dans lesquelles l'accident s'est produit
 - 1 – Plein jour
 - 2 – Crépuscule ou aube
 - 3 – Nuit sans éclairage public
 - 4 – Nuit avec éclairage public non allumé
 - 5 – Nuit avec éclairage public allumé
- Com : Le numéro de commune est un code donné par l'INSEE. Le code est composé du code INSEE du département suivi par 3 chiffres
- Agg : Localisation de l'accident
 - 1 – Hors agglomération
 - 2 – En agglomération
- Int : Le type d'intersection où l'accident s'est déroulé
 - 1 – Hors intersection
 - 2 – Intersection en X
 - 3 – Intersection en T
 - 4 – Intersection en Y
 - 5 – Intersection à plus de 4 branches
 - 6 – Giratoire
 - 7 – Place
 - 8 – Passage à niveau
 - 9 – Autre intersection
- Atm : Le type de condition atmosphérique
 - 1 – Non renseigné
 - 1 – Normale
 - 2 – Pluie légère
 - 3 – Pluie forte
 - 4 – Neige - Grêle
 - 5 – Brouillard - Fumée
 - 6 – Vent fort - Tempête
 - 7 – Temps éblouissant
 - 8 – Temps couvert
 - 9 – Couvert
- Col : Le type de collision
 - 1 – Non renseigné
 - 1 – Deux véhicules - frontale
 - 2 – Deux véhicules - par l'arrière
 - 3 – Deux véhicules - par le côté
 - 4 – Trois véhicules et plus - en chaîne
 - 5 – Trois véhicules et plus - collisions multiples
 - 6 – Autre collision
 - 7 – Sans collision

— Base lieux :

- Catr : Catégorie de route
 - 1 – Autoroute
 - 2 – Route nationale
 - 3 – Route départementale
 - 4 – Voie communale

- 5 – Hors réseau public
- 6 – Parc de stationnement ouvert à la circulation publique
- 7 – Routes de métropole urbaine
- 9 – Autre
- Voie : Numéro de la route
- Circ : Régime de circulation
 - 1 – Non renseigné
 - 1 – À sens unique
 - 2 – Bidirectionnelle
 - 3 – À chaussées séparées
 - 4 – Avec voies d'affectation variable
- Nbv : Nombre total de voies de circulation
- Vosp : Signale l'existence d'une voie réservée, indépendamment du fait que l'accident ait lieu ou non sur cette voie
 - 1 – Non renseigné
 - 0 – Sans objet
 - 1 – Piste cyclable
 - 2 – Bande cyclable
 - 3 – Voie réservée
- Prof : Profil en long décrit la déclivité de la route à l'endroit de l'accident
 - 1 – Non renseigné
 - 1 – Plat
 - 2 – Pente
 - 3 – Sommet de côte
 - 4 – Bas de côte
- Plan : Tracé en plan
 - 1 – Non renseigné
 - 1 – Partie rectiligne
 - 2 – En courbe à gauche
 - 3 – En courbe à droite
 - 4 – En "s"
- Larout : Largeur de la chaussée affectée à la circulation des véhicules, ne sont pas compris les bandes d'arrêts d'urgences, les TPC et les places de stationnement (en m)
- Surf : État de la surface
 - 1 – Non renseigné
 - 1 – Normale
 - 2 – Mouillée
 - 3 – Flaques
 - 4 – Inondée
 - 5 – Enneigée
 - 6 – Boue
 - 7 – Verglacée
 - 8 – Corps gras - huile
 - 9 – Autre
- Infra : Aménagement, infrastructure
 - 1 – Non renseigné
 - 0 – Aucun
 - 1 – Souterrain - tunnel
 - 2 – Pont - autopont
 - 3 – Bretelle d'échangeur ou de raccordement
 - 4 – Voie ferrée
 - 5 – Carrefour aménagé
 - 6 – Zone piétonne

7 – Zone de péage

8 – Chantier

9 – Autres

— Situ : Situation de l'accident

-1 – Non renseigné

0 – Aucun

1 – Sur chaussée

2 – Sur bandes d'arrêt d'urgence

3 – Sur accotement

4 – Sur trottoir

5 – Sur piste cyclable

6 – Sur autre voie spéciale

8 – Autres

— VMA : Vitesse maximale autorisée sur le lieu et au moment de l'accident

— **Base véhicule :**

— Id_vehicule : Identifiant unique du véhicule repris pour chacun des usagers occupant ce véhicule (y compris les piétons qui sont rattachés aux véhicules qui les ont heurtés) – Code numérique

— Catv : Catégorie du véhicule : 00 – Indéterminable

01 – Bicyclette

02 – Cyclomoteur <50cm³

03 – Voiturette (Quadricycle à moteur carrossé) (anciennement “voiturette ou tricycle à moteur”)

04 – Référence inutilisée depuis 2006 (scooter immatriculé)

05 – Référence inutilisée depuis 2006 (motocyclette)

06 – Référence inutilisée depuis 2006 (side-car)

07 – VL seul

08 – Référence inutilisée depuis 2006 (VL + caravane)

09 – Référence inutilisée depuis 2006 (VL + remorque)

10 – VU seul 1,5T <= PTAC <= 3,5T avec ou sans remorque (anciennement VU seul 1,5T <= PTAC <= 3,5T)

11 – Référence inutilisée depuis 2006 (VU (10) + caravane)

12 – Référence inutilisée depuis 2006 (VU (10) + remorque)

13 – PL seul 3,5T <PTCA <= 7,5T

14 – PL seul > 7,5T

15 – PL > 3,5T + remorque

16 – Tracteur routier seul

17 – Tracteur routier + semi-remorque

18 – Référence inutilisée depuis 2006 (transport en commun)

19 – Référence inutilisée depuis 2006 (tramway)

20 – Engin spécial

21 – Tracteur agricole

30 – Scooter < 50 cm³

31 – Motocyclette > 50 cm³ et <= 125 cm³

32 – Scooter > 50 cm³ et <= 125 cm³

33 – Motocyclette > 125 cm³

34 – Scooter > 125 cm³

35 – Quad léger <= 50 cm³ (Quadricycle à moteur non carrossé)

36 – Quad lourd > 50 cm³ (Quadricycle à moteur non carrossé)

37 – Autobus

38 – Autocar

39 – Train

40 – Tramway

41 – 3RM <= 50 cm³

42 – 3RM > 50 cm³ <= 125 cm³

- 43 – 3RM > 125 cm³
- 50 – EDP à moteur
- 60 – EDP sans moteur 80 – VAE
- 99 – Autre véhicule
- Obs : Obstacle fixe heurté
 - 1 – Non renseigné
 - 0 – Sans objet
 - 1 – Véhicule en stationnement
 - 2 – Arbre
 - 3 – Glissière métallique
 - 4 – Glissière béton
 - 5 – Autre glissière
 - 6 – Bâtiment, mur, pile de pont
 - 7 – Support de signalisation verticale ou poste d'appel d'urgence
 - 8 – Poteau
 - 9 – Mobilier urbain
 - 10 – Parapet
 - 11 – Ilot, refuge, borne haute
 - 12 – Bordure de trottoir
 - 13 – Fossé, talus, paroi rocheuse
 - 14 – Autre obstacle fixe sur chaussée
 - 15 – Autre obstacle fixe sur trottoir ou accotement
 - 16 – Sortie de chaussée sans obstacle
 - 17 – Buse – tête d'aqueduc
- Obsm : Obstacle mobile heurté
 - 1 – Non renseigné
 - 0 – Aucun
 - 1 – Piéton
 - 2 – Véhicule
 - 4 – Véhicule sur rail
 - 5 – Animal domestique
 - 6 – Animal sauvage
 - 9 – Autre
- Manv : Manoeuvre principale avant l'accident
 - 1 – Non renseigné
 - 0 – Inconnue
 - 1 – Sans changement de direction
 - 2 – Même sens, même file
 - 3 – Entre 2 files
 - 4 – En marche arrière
 - 5 – A contresens
 - 6 – En franchissant le terre-plein central
 - 7 – Dans le couloir bus, dans le même sens
 - 8 – Dans le couloir bus, dans le sens inverse
 - 9 – En s'insérant 10 – En faisant demi-tour sur la chaussée
- Changeant de file**
 - 11 – A gauche
 - 12 – A droite
- Déporté**
 - 13 – A gauche
 - 14 – A droite
- Tournant**
 - 15 – A gauche
 - 16 – A droite

Dépassant

17 – A gauche

18 – A droite

Divers

19 – Traversant la chaussée

20 – Manœuvre de stationnement

21 – Manœuvre d'évitement

22 – Ouverture de porte

23 – Arrêté (hors stationnement)

24 – En stationnement (avec occupants)

25 – Circulant sur trottoir

26 – Autres manœuvres

— Motor : Type de motorisation du véhicule

-1 – Non renseigné

0 – Inconnue

1 – Hydrocarbures

2 – Hybride électrique

3 – Electrique

4 – Hydrogène

5 – Humaine

6 – Autre

— **Base usagers :**

— Catu : Catégorie d'utilisateur

1 – Conducteur

2 – Passager

3 – Piéton

— Grav: Gravité de blessure de l'utilisateur, les usagers accidentés sont classés en trois catégories de victimes plus les indemnes

1 – Indemne

2 – Tué

3 – Blessé hospitalisé

4 – Blessé léger

— Sexe : Sexe de l'utilisateur

1 – Masculin

2 – Féminin

— An_nais : Année de naissance de l'utilisateur

— Trajet : Motif du déplacement au moment de l'accident

-1 – Non renseigné

0 – Non renseigné

1 – Domicile – travail

2 – Domicile – école

3 – Courses – achats

4 – Utilisation professionnelle

5 – Promenade – loisirs

9 – Autre

— Secu1 : Le renseignement du caractère indique la présence et l'utilisation de l'équipement de sécurité

-1 – Non renseigné

0 – Aucun équipement

1 – Ceinture

2 – Casque

3 – Dispositif enfants

4 – Gilet réfléchissant

- 5 – Airbag (2RM/3RM)
- 6 – Gants (2RM/3RM)
- 7 – Gants + Airbag (2RM/3RM)
- 8 – Non déterminable
- 9 – Autre
- Secu2 : Le renseignement du caractère indique la présence et l'utilisation de l'équipement de sécurité
 - 1 – Non renseigné
 - 0 – Aucun équipement
 - 1 – Ceinture
 - 2 – Casque
 - 3 – Dispositif enfants
 - 4 – Gilet réfléchissant
 - 5 – Airbag (2RM/3RM)
 - 6 – Gants (2RM/3RM)
 - 7 – Gants + Airbag (2RM/3RM)
 - 8 – Non déterminable
 - 9 – Autre
- Secu3 : Le renseignement du caractère indique la présence et l'utilisation de l'équipement de sécurité
 - 1 – Non renseigné
 - 0 – Aucun équipement
 - 1 – Ceinture
 - 2 – Casque
 - 3 – Dispositif enfants
 - 4 – Gilet réfléchissant
 - 5 – Airbag (2RM/3RM)
 - 6 – Gants (2RM/3RM)
 - 7 – Gants + Airbag (2RM/3RM)
 - 8 – Non déterminable
 - 9 – Autre
- Locp : Localisation du piéton
 - 1 – Non renseigné
 - 0 – Sans objet
 - Sur chaussée :**
 - 1 – A + 50 m du passage piéton
 - 2 – A – 50 m du passage piéton
 - Sur passage piéton :**
 - 3 – Sans signalisation lumineuse
 - 4 – Avec signalisation lumineuse
 - Divers :**
 - 5 – Sur trottoir
 - 6 – Sur accotement
 - 7 – Sur refuge ou BAU
 - 8 – Sur contre allée
 - 9 – Inconnue
- Actp : Action du piéton
 - 1 – Non renseigné
 - Se déplaçant**
 - 0 – Non renseigné ou sans objet
 - 1 – Sens véhicule heurtant
 - 2 – Sens inverse du véhicule
 - Divers**
 - 3 – Traversant

- 4 – Masqué
- 5 – Jouant – courant
- 6 – Avec animal
- 9 – Autre
- A – Monte/descend du véhicule
- B – Inconnue
- Etatp : Cette variable permet de préciser si le piéton accidenté était seul ou non
 - 1 – Non renseigné
 - 1 – Seul
 - 2 – Accompagné
 - 3 – En groupe