

BROOKLYN NINE-NINE

PROJET IA26

Samuel, Adam, Seydou et Ambrane

Données sur les accidents entre véhicules

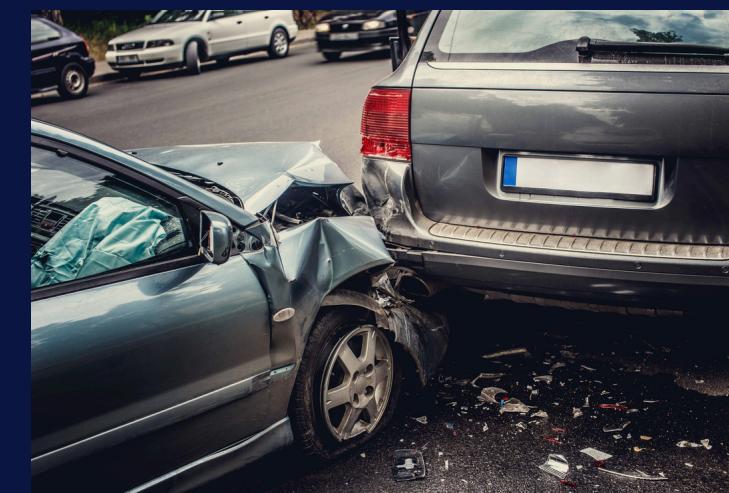


SOMMAIRE

1. Changement de dataset
2. Présentation du dataset et du contexte de notre projet
3. Méthodologie de travail
 - a. Préparation des données
 - b. Classifieurs et métriques
 - i. Random Forest
 - ii. Analyse en passant par des réseaux de neurones et Keras Tuner
 - c. Simulation d'appel
4. Résultats
5. Conclusion et perspectives

CHANGEMENT DE DATASET

- Contexte : le vélo, un réel engouement à Paris. Pratique à utiliser quand les transports en commun sont saturés.
- Problématique : prédire l'affluence des stations Vélib près d'un événement à Paris
- Difficultés rencontrées (à notre échelle, pour la personne qui met à jour le dataset, en termes de calendrier)



CONTEXTE

- Sécurité routière → enjeu public
- Des milliers d'accidents corporels : un accident survenu en voie publique, impliquant au moins un véhicule et ayant fait au moins une victime ayant une nécessité des soins

**Comprendre les facteurs qui
influencent la gravité des accidents**



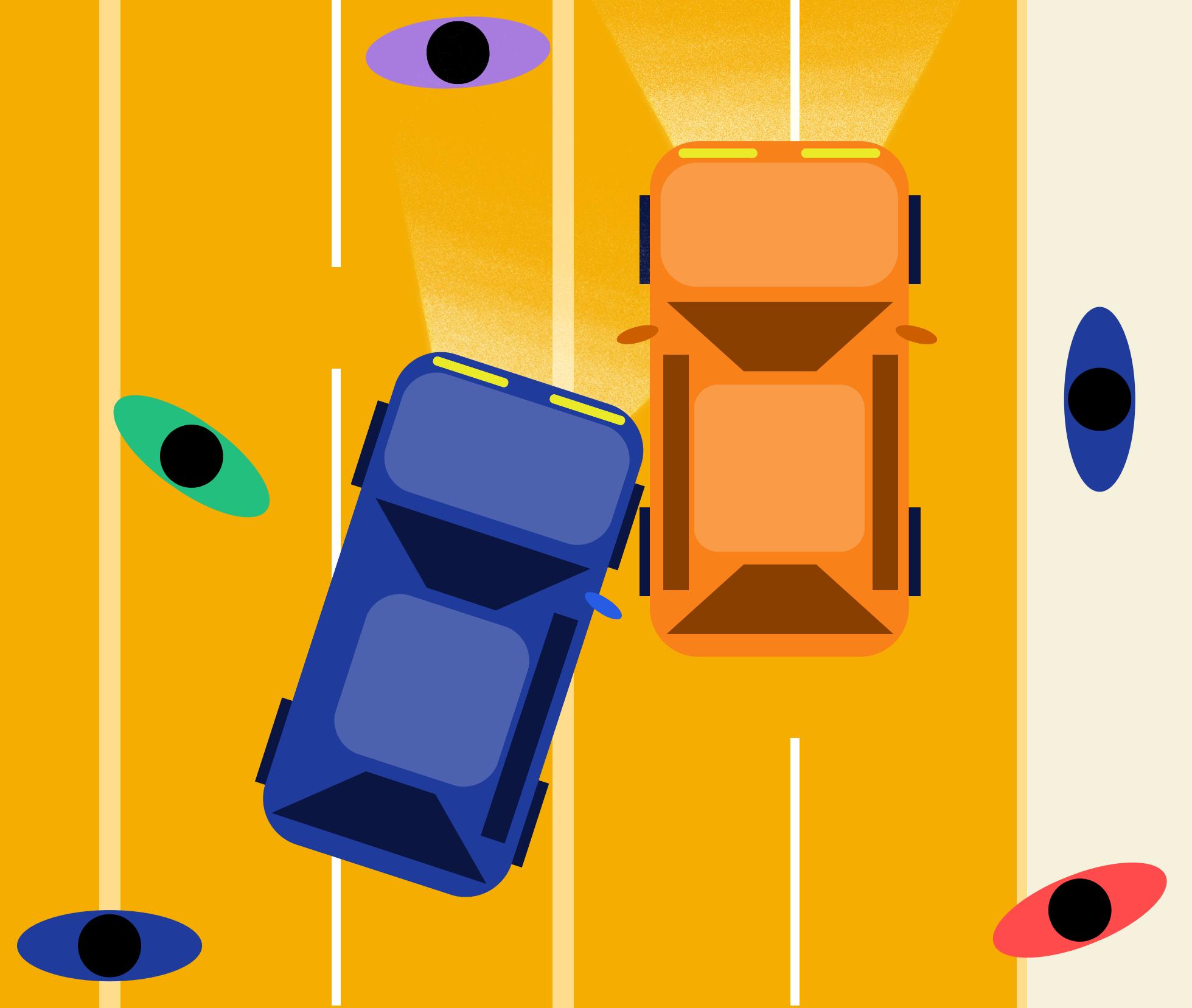


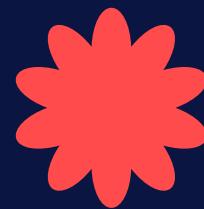
OBJECTIFS

Concevoir un modèle de prédiction capable d'estimer le niveau de gravité d'un accident corporel
→ à partir des données historiques d'accidents 2021- 2023

Problème de classification :

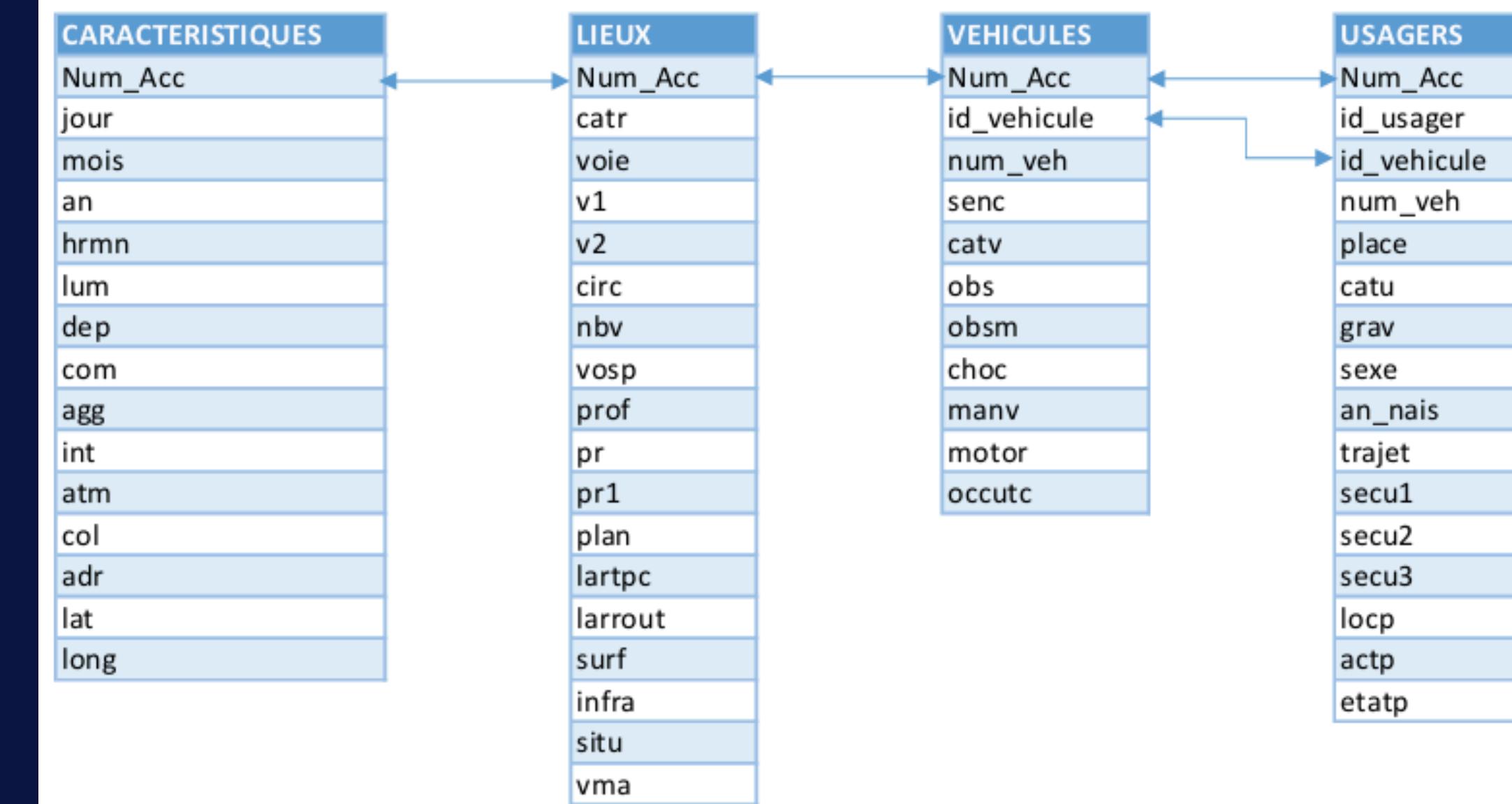
- 1 : indemne
- 2 : blessé léger
- 3 : blessé hospitalisé
- 4 : tué





NOS DONNÉES

“ACCIDENTS CORPORELS DE LA CIRCULATION ROUTIÈRE”

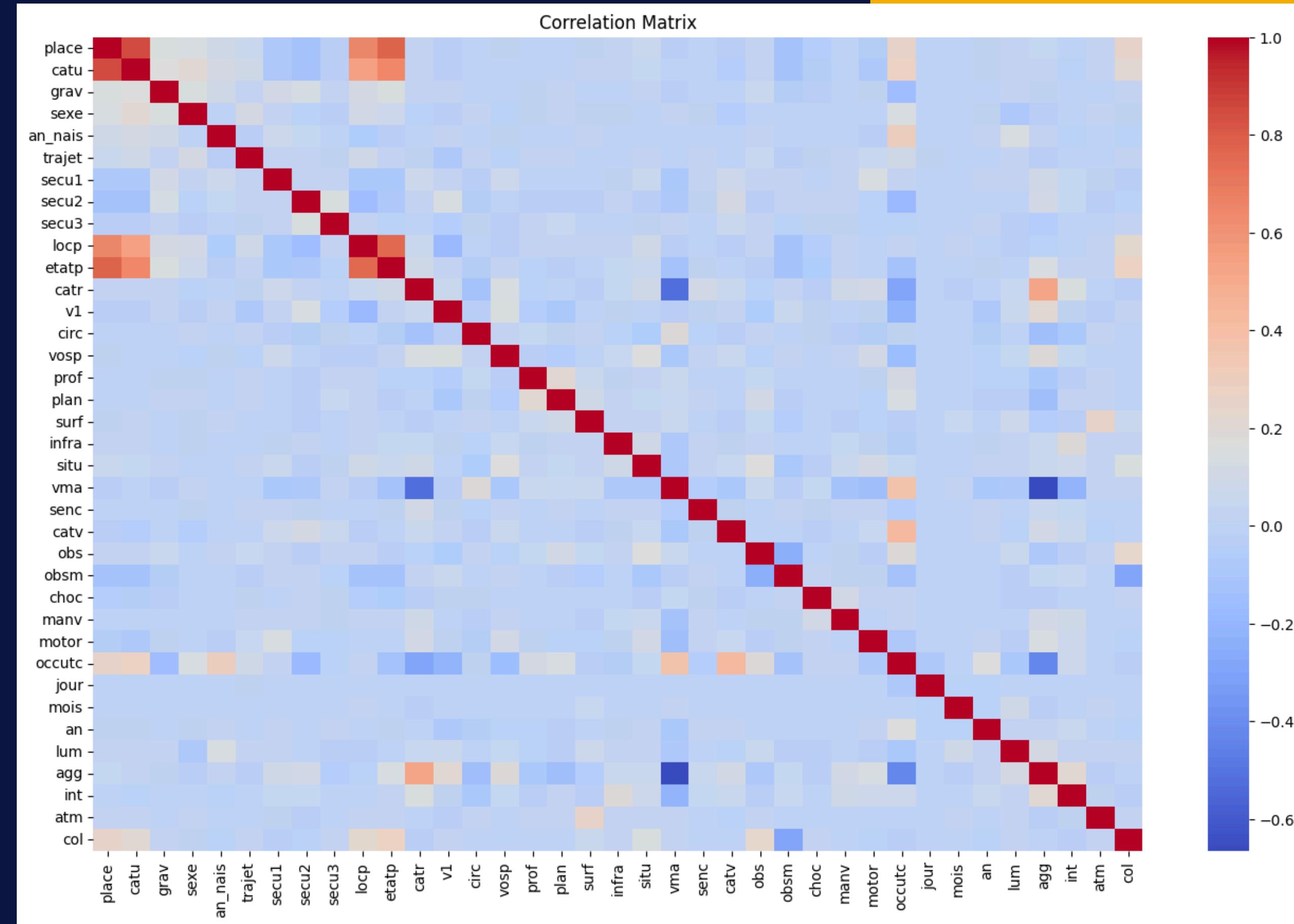




PRÉPARATION DES DONNÉES

1. Concaténation des données entre 2021 et 2023 pour chaque type de table
2. Fusionner les 4 tableaux dans une seule data base
3. Création de la matrice de corrélation

MATRICE DE CORRÉLATION



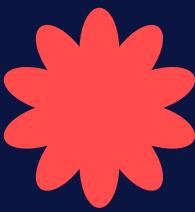
PRÉPARATION DES DONNÉES



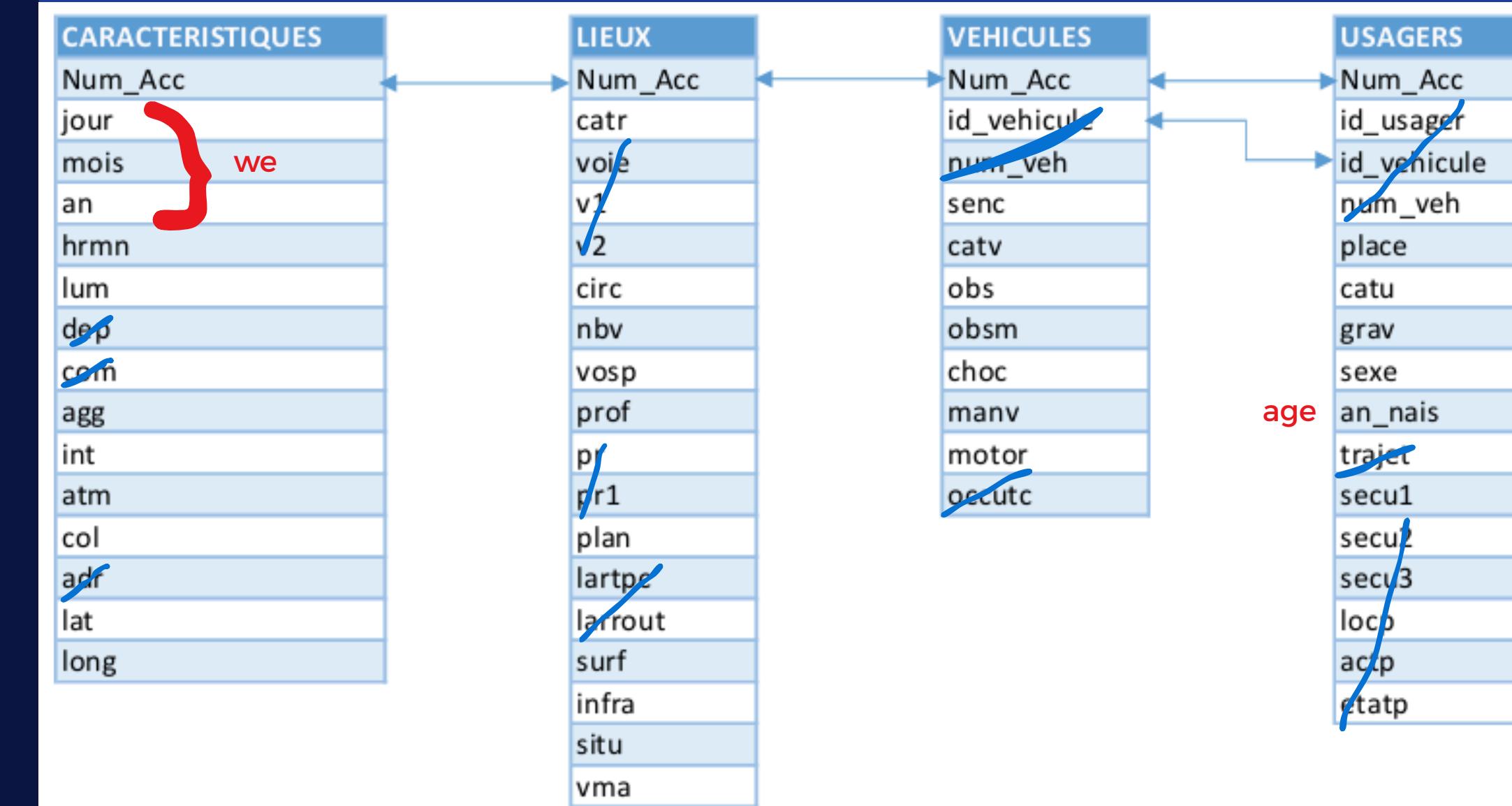
Données qu'on est capable de donner à l'oral au secours sur le lieu d'un accident



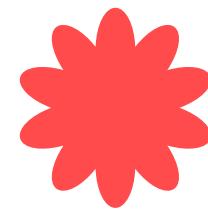
4. Sélection des colonnes pertinentes pour la prédiction
5. Revue du format de certaines colonnes
6. Création des fonctions :
 - one hot,
 - normalisation,
 - suppression des lignes incomplètes.



NOS DONNÉES APRÈS TRAITEMENT



419 650 → 319 235



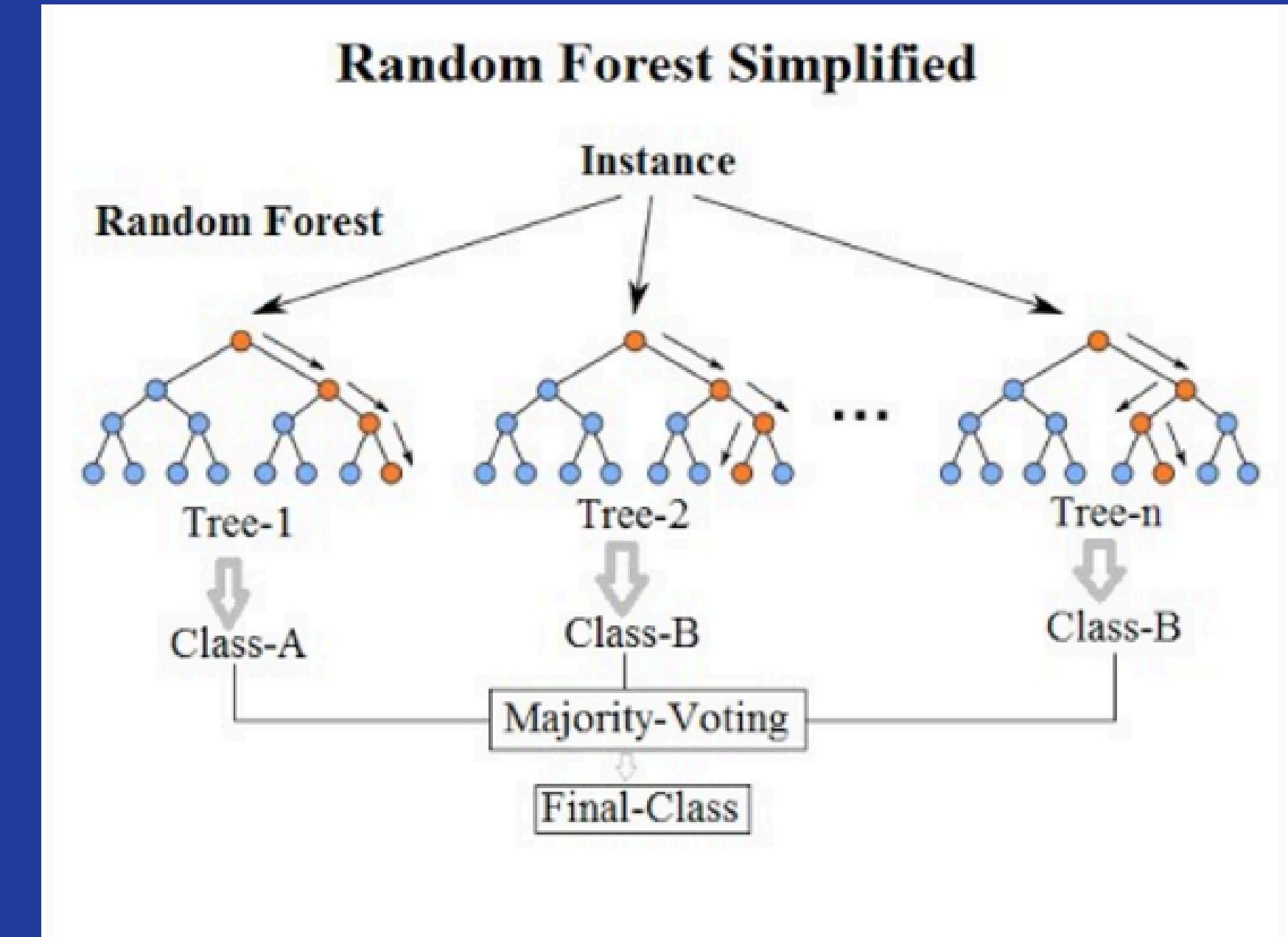
RANDOM FOREST

Avantages de Random Forest par rapport à K-Means et SVM pour la classification

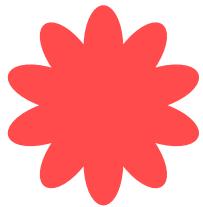
- Pas de notion de distances
- Explicabilité par la visualisation des arbres de décision

Critère de Gini

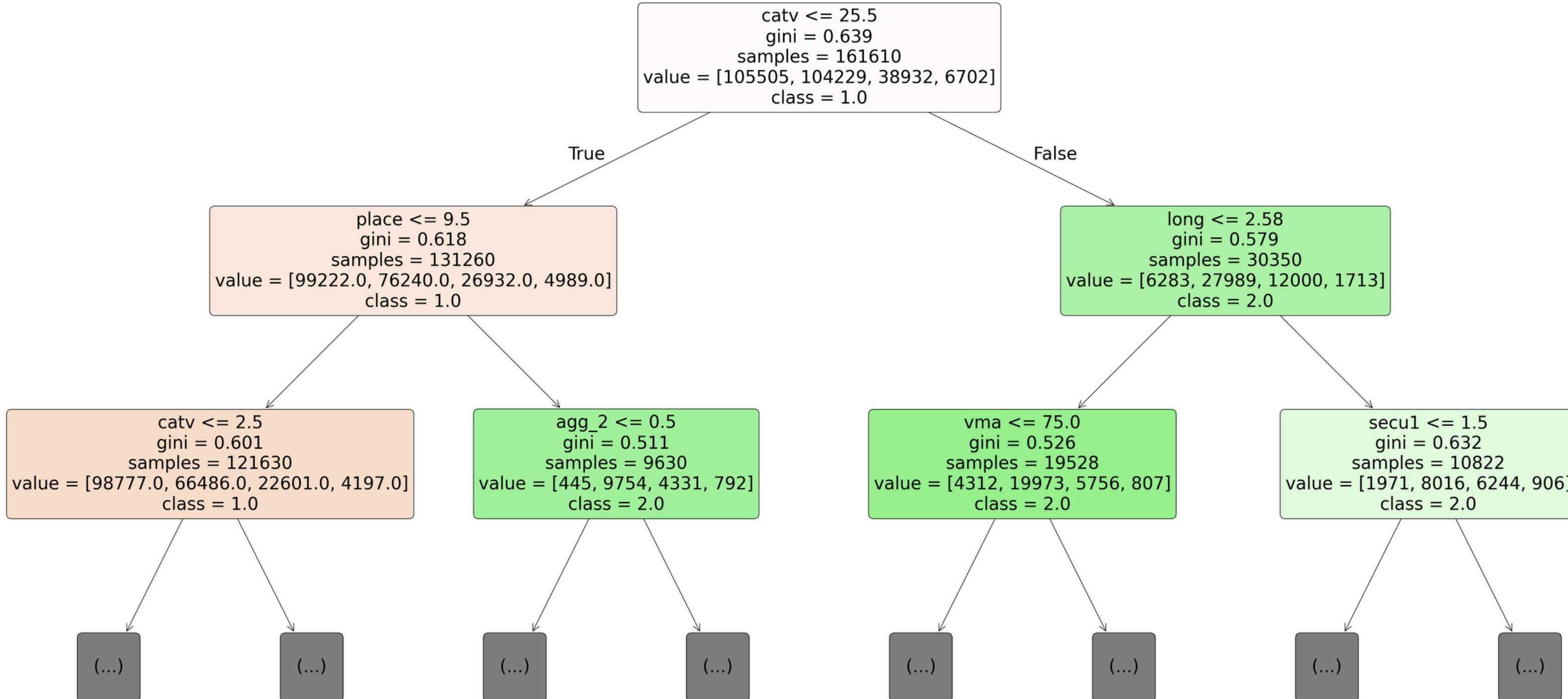
$$I_G(f) = \sum_{i=1}^m f_i(1 - f_i)$$



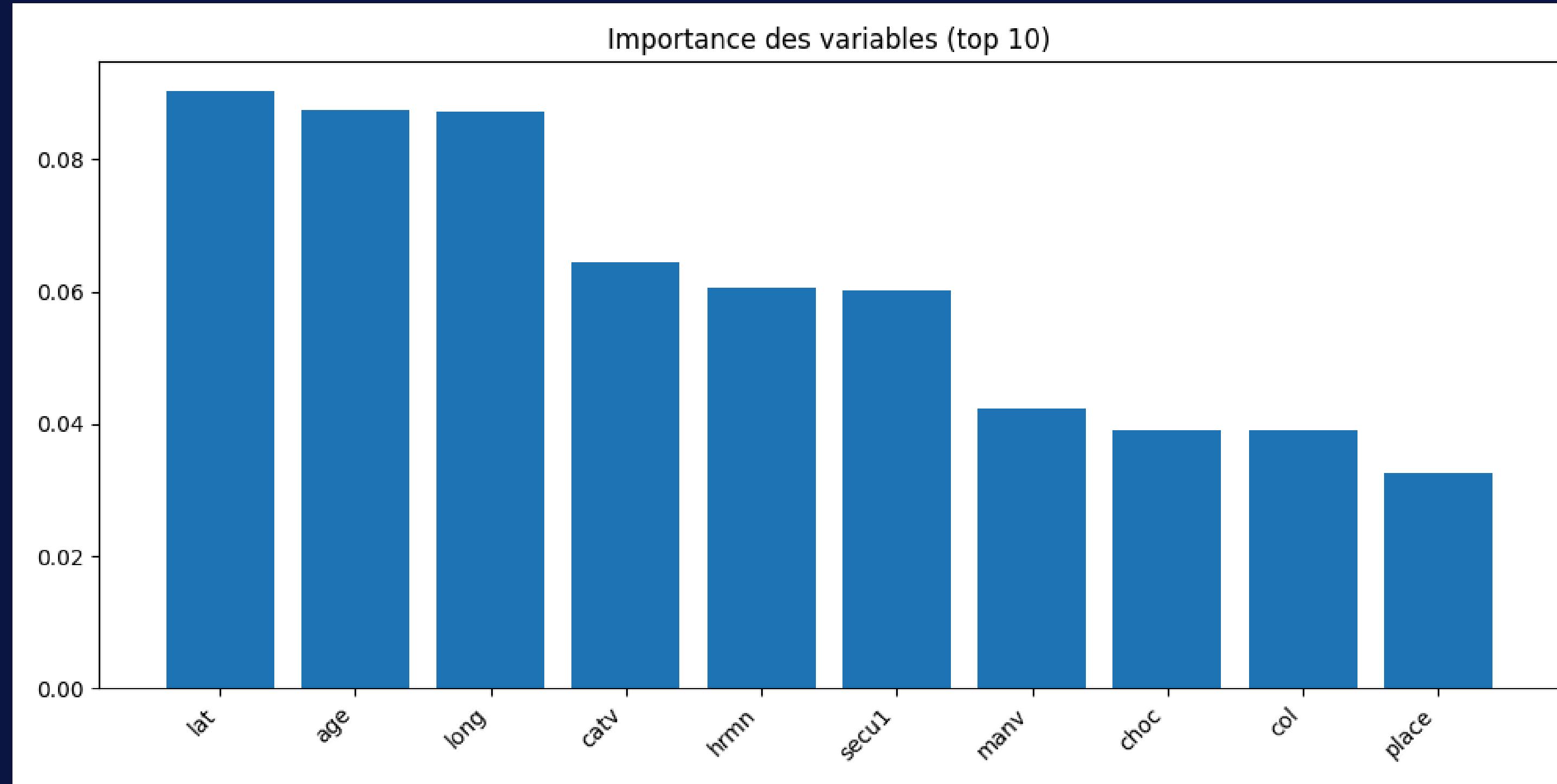
ZOOM SUR UN DES ARBRES DE DECISION DU RANDOM FOREST

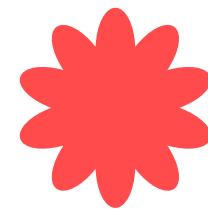


Arbre de décision #2



Mis en évidence des variables importantes pour la prédiction





RÉSULTATS

Un modèle plutôt bon:

- **Score F1 Macro: 0.53 forte disparité entre les classes**
- **Rappel Faible pour Blessé hospitalisé et Tué**

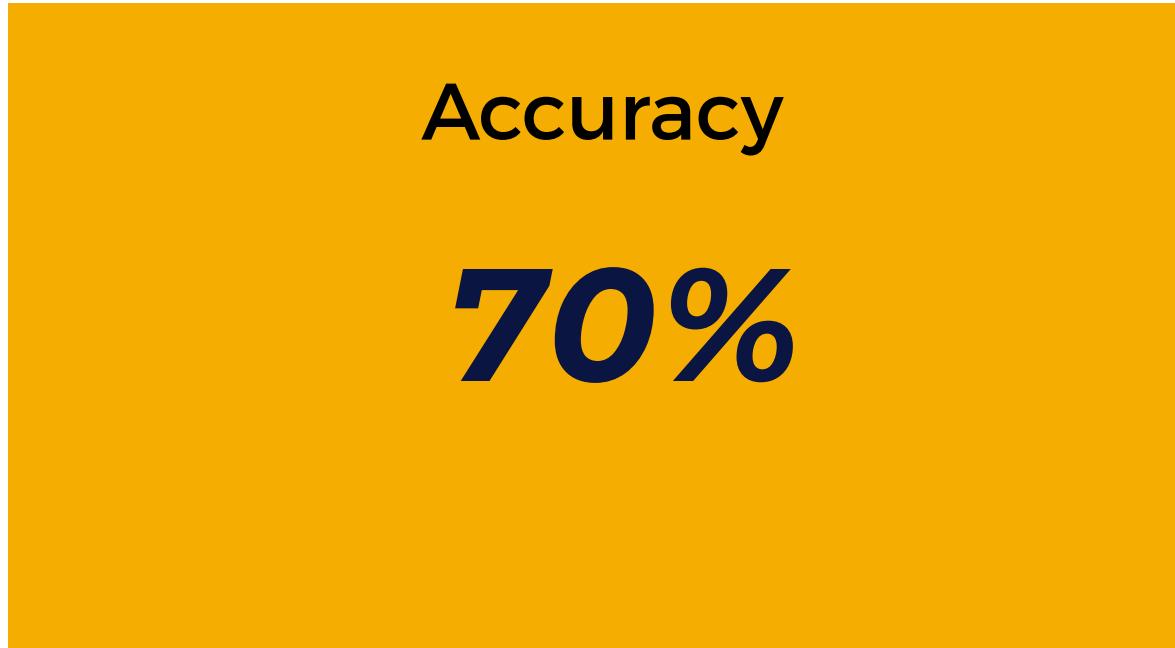
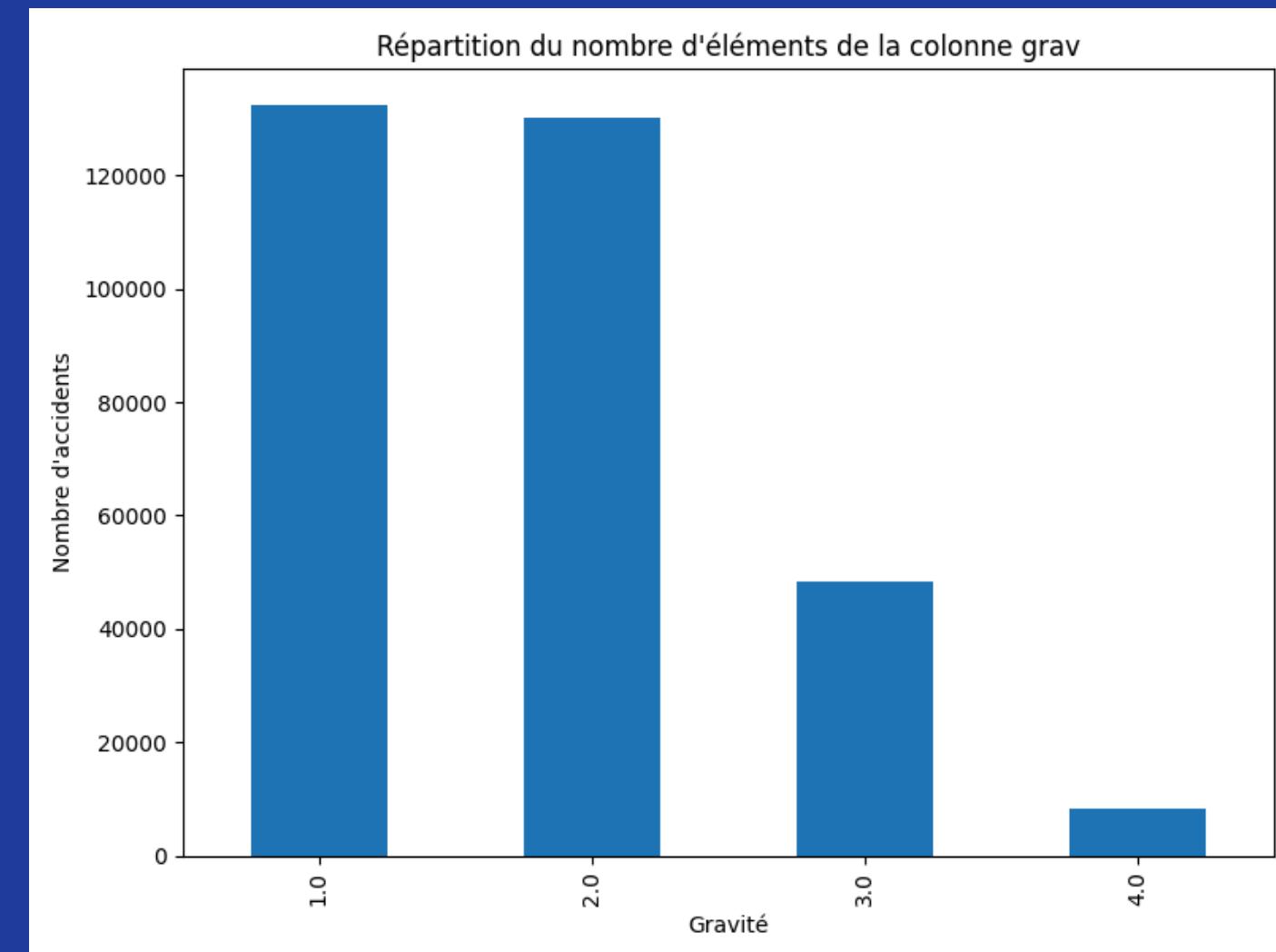
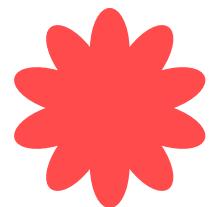


Table 1: Rapport de classification de la Random Forest sur les données test

Classe	Précision	Rappel	F1-score	Support
1.0 (Indemne)	0.75	0.83	0.79	26 723
2.0 (Blessé léger)	0.68	0.69	0.69	25 826
3.0 (Blessé hospitalisé)	0.56	0.44	0.49	9 602
4.0 (Tué)	0.55	0.09	0.15	1 692

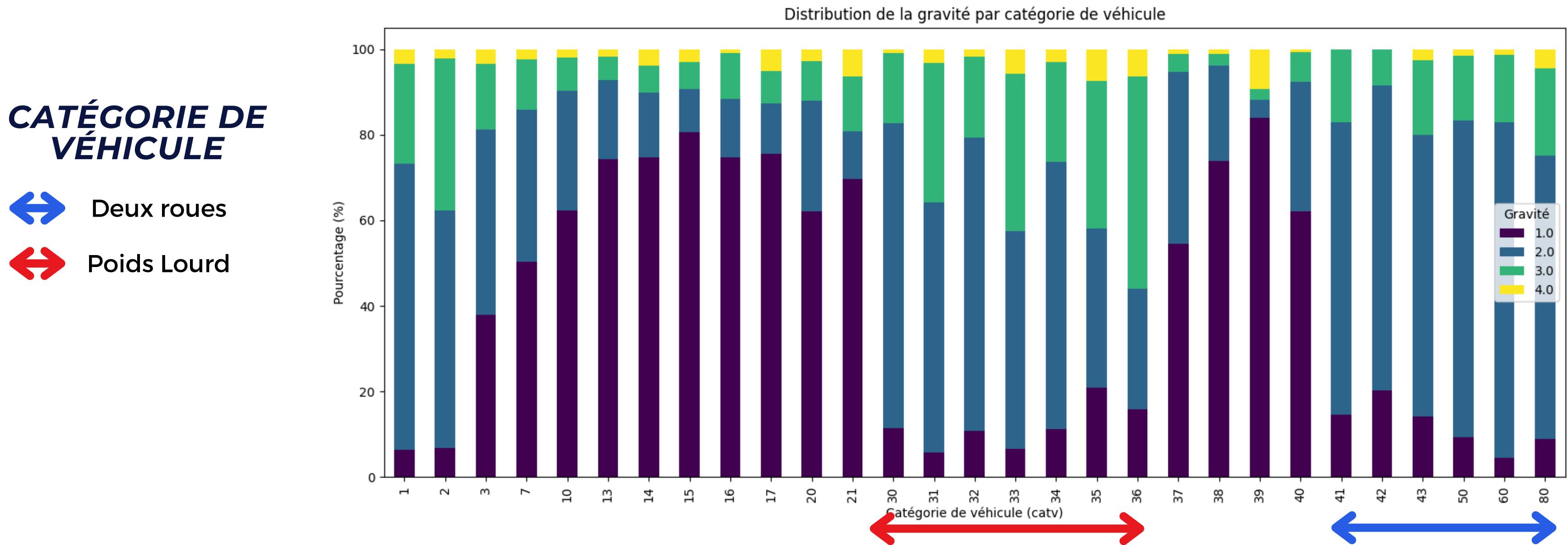




AMÉLIORATIONS

- Régression
- Entraînement sur un jeu de donnée mieux réparti $\frac{1}{3}$ de 2, $\frac{1}{3}$ de 3 et $\frac{1}{3}$ de 4
- Changement du jeu de donnée pour se concentrer sur des sous groupes aux dynamiques propres

MSE
0.378



RÉSEAUX DE NEURONES

Performance du Réseau de Neurones - Classification

60%

Accuracy

52%

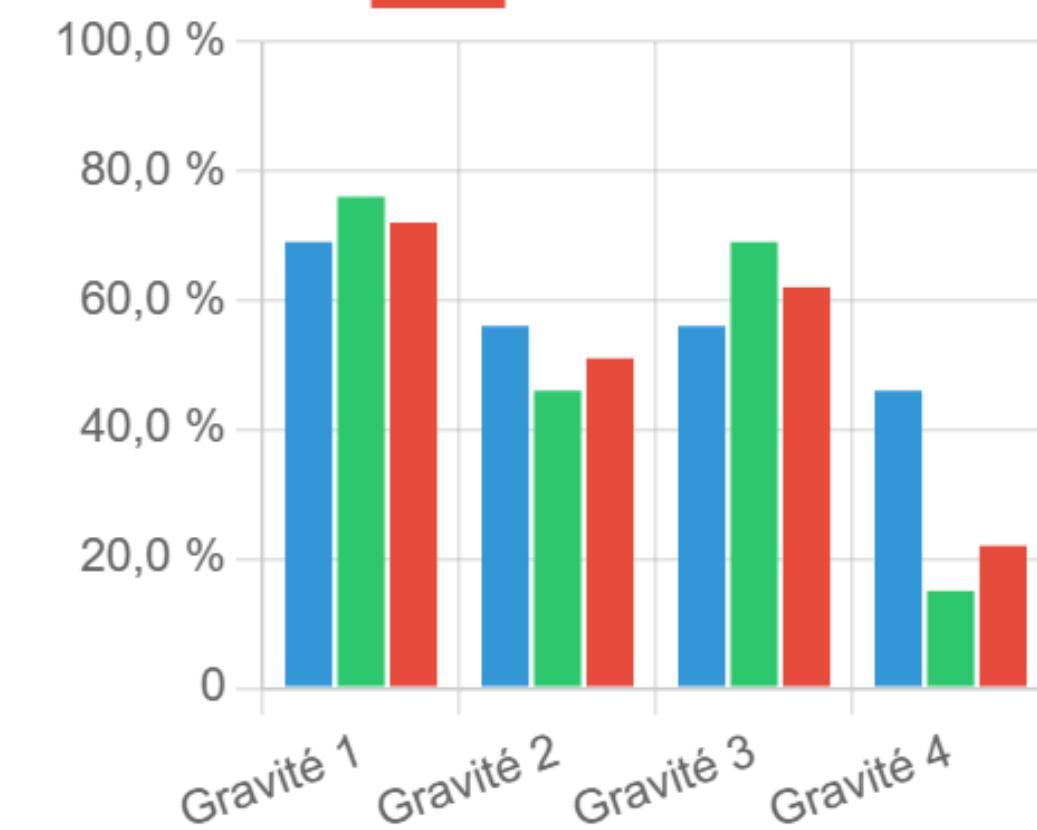
F1-Score Moyen

26,813

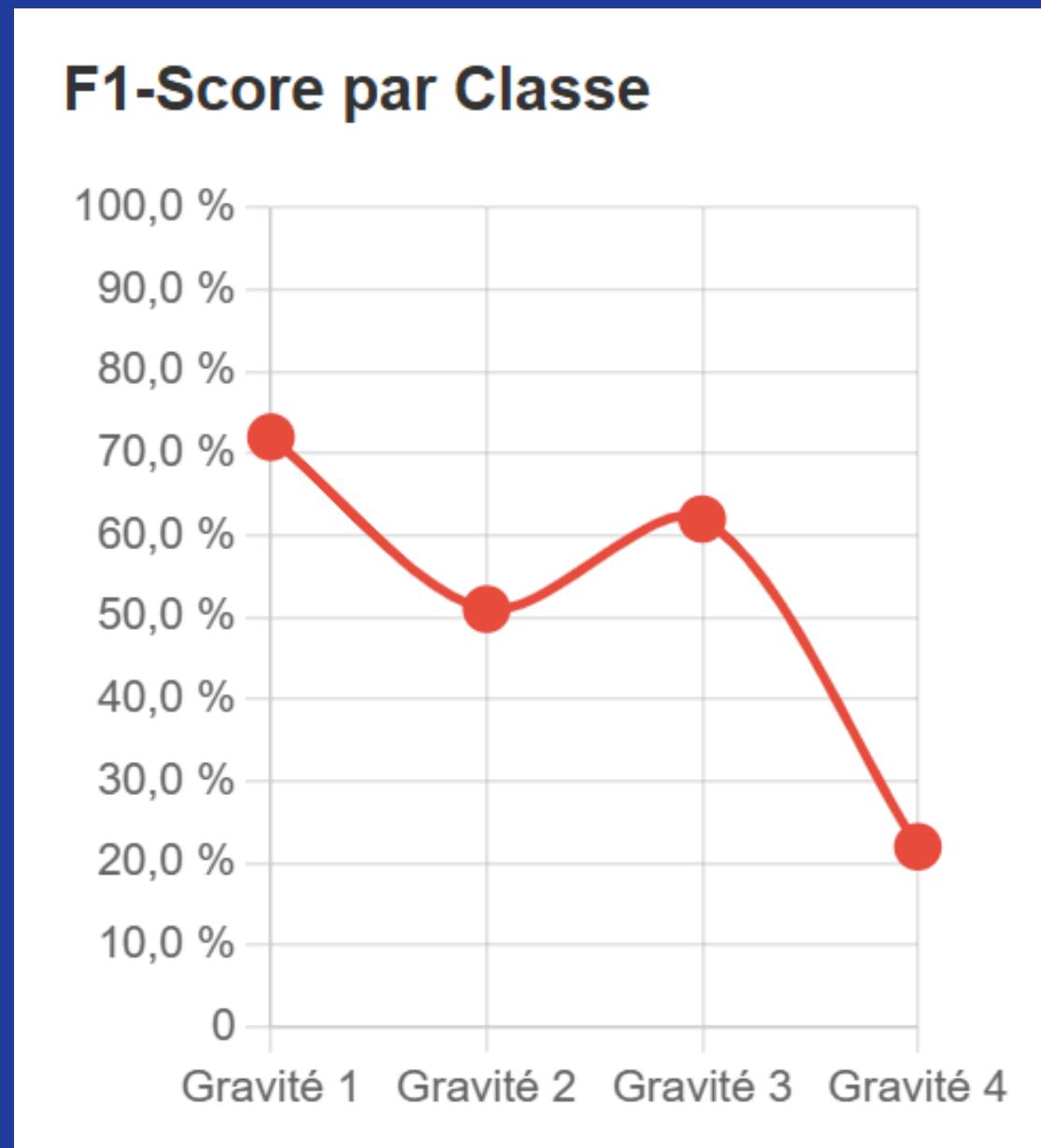
Échantillons Total

Métriques par Classe

Precision Recall
F1-Score



RÉSEAUX DE NEURONES



Matrice de Confusion

	Prédit G1	Prédit G2	Prédit G3	Prédit G4
Réel G1	6339	1147	862	31
Réel G2	2014	3876	2447	42
Réel G3	678	1730	5753	218
Réel G4	94	135	1201	246

SIMULATION D'APPEL & RÉSULTATS

Lors de la simulation suivante : un conducteur sur une voie à sens unique le 4 juin 2025 à 20h 19 rue Hallé 75014 a rencontré sur l'avant gauche de sa voiture un poteau et un cycliste alors que la route communale plate était mouillée et qu'il y avait une forte pluie

```
place catu sexe secu1 senc catv obs obsm choc manv motor catr \
 1   1   1   1   0   7   8   1   2   12   1   4

circ nbv vosp prof plan surf infra situ vma hrmn lum atm col \
 1   1   1   1   2   2   0   4   50   20   2   3   6

lat      long     we age agg_1 agg_2 int_1 int_2 int_3 int_4 \
48.830561 2.332691 False 22 False True  True False False False

int_5 int_6 int_7 int_8 int_9
False False False False False
```

La prédiction à partir du classifieur est :

[1.]

La prédiction à partir de la régression est :

[1.]

CONCLUSION ET PERSPECTIVES

- Après changement de dataset
- Traitement d'un problème de **classification** par une méthodologie complète
- Sans grande surprise : **localisation géographique et âge des utilisateurs** → **facteurs les + déterminants dans la gravité des accidents**
- Difficulté : **déséquilibre** dans la **représentation des classes dans les données d'apprentissage**
- **Simulation d'appel** démontre une certaine faisabilité opérationnelle
- **Perspectives pour la sécurité routière**
 - identification des zones à risque
 - optimisation des interventions de secours
 - aide à la décision pour les politiques de prévention
- Améliorer le modèle demanderait d'améliorer les **données** : + **précises** (notamment pour **accidents graves**)



MERCI

Contactez nous en cas d'accidents ;))