

Projet Accidentologie



Rémi THINEY - Datascientest Feb24
github.com/remithiney/Projet_Accidentologie

Plan

1. Introduction
2. Problématique
3. Exploration des données
4. Modélisation
5. Résultats

Introduction



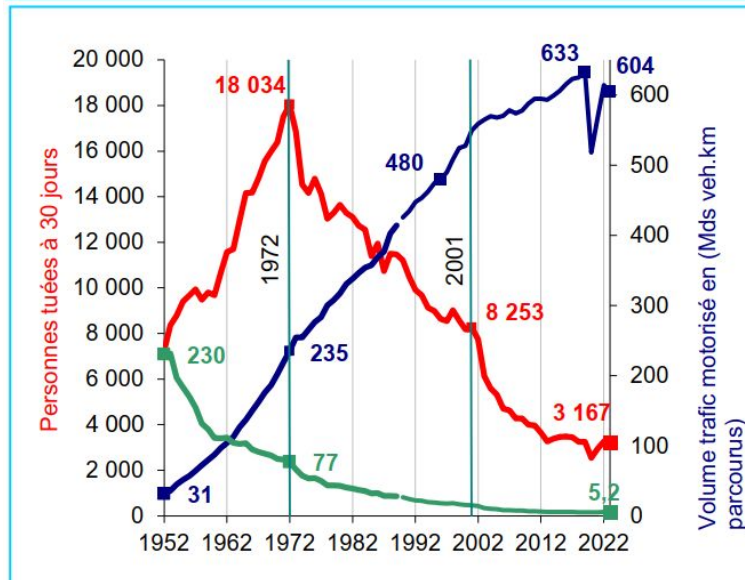
Autopromotion en relation avec le projet

- Master développement informatique appliqué à la gestion des entreprises
- Contractuel au service immigration de la Somme - droit au séjour
- Enseignant de la sécurité routière et responsable pédagogique d'agence

La sécurité routière enjeu majeur de société

- Réduction de **83%** des décès routiers.
- Pour **95%** de hausse du trafic.
- Soit **98%** de diminution du rapport décès/Mds km parcourus.

Évolution comparée de la mortalité et de la circulation routière entre 1952 et 2023*



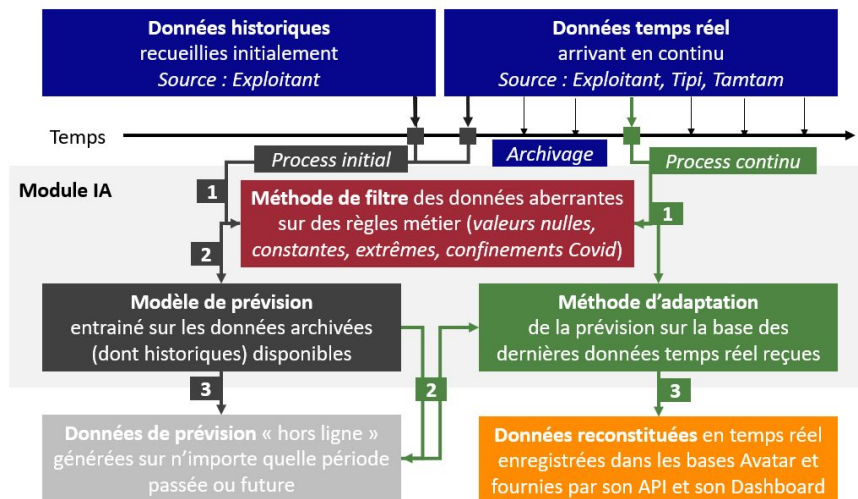
Les données de trafic fournies par le SDES ont été rebasées en 2024 pour les années allant de 1990 à 2023

*2023 : données du trafic provisoires

Machine et Deep Learning au service de la sécurité routière.

- Prédiction des risques.
- Analyse de trafic.
- Aide à la conduite.
- Voir conduite autonome.
- Détection des infractions.

= Opportunités d'améliorer notre compréhension VOIR interventions directes !



Machine learning: projet Avatar du Cerema

Problématique

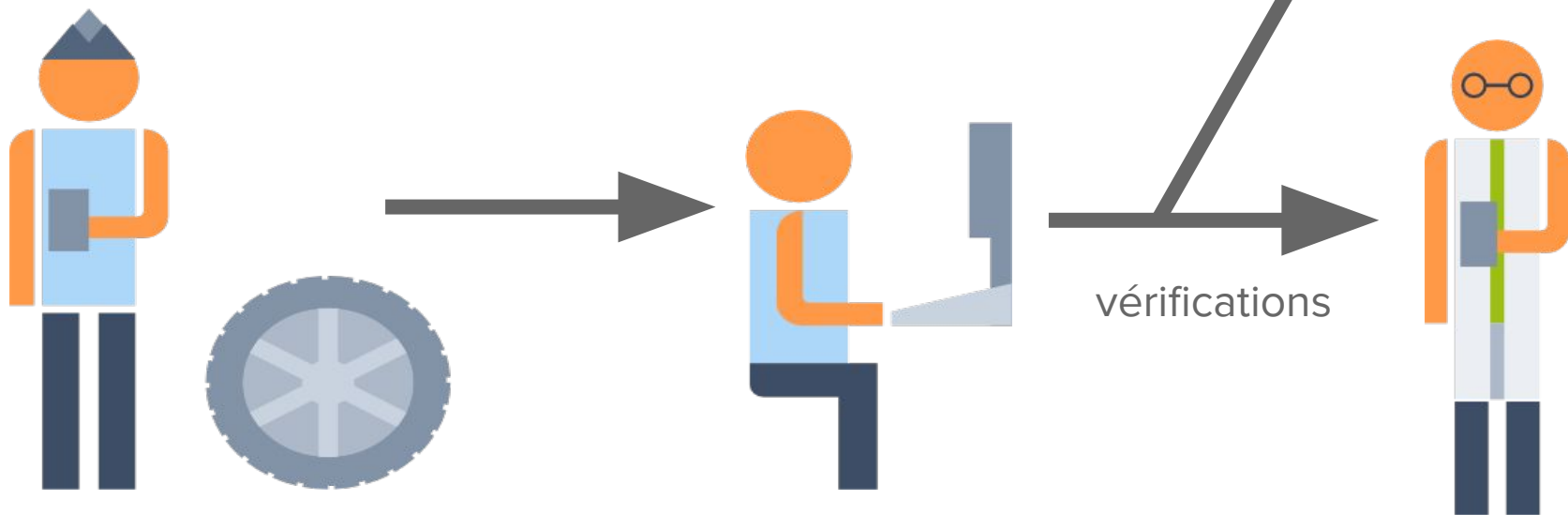


Objectifs du projet

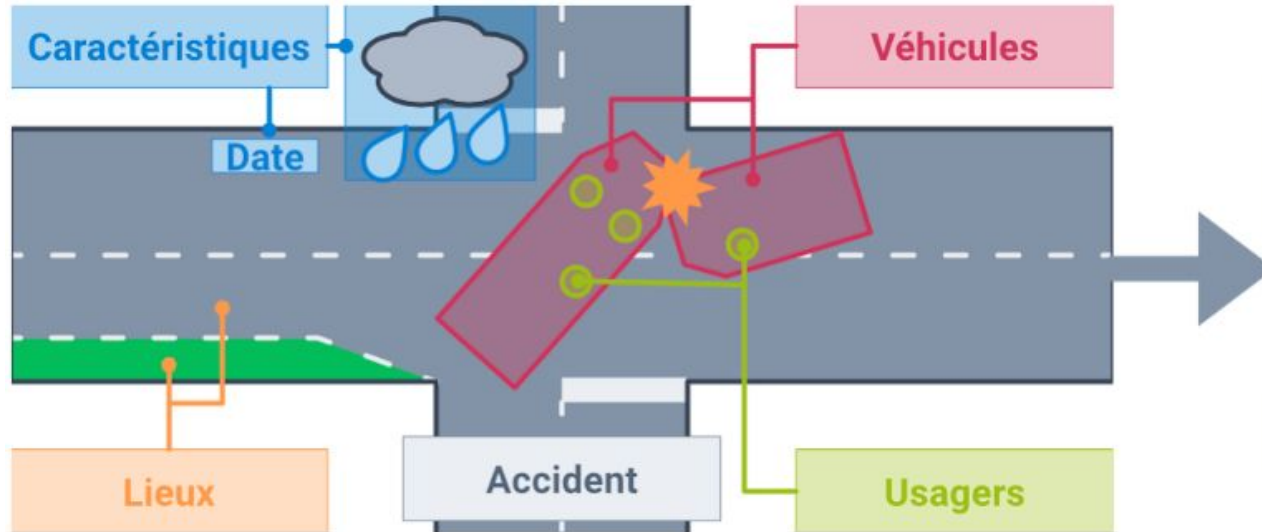
- Mieux comprendre les facteurs de risques.
- Prédire la gravité d'un accident routier pour un usager impliqué.
- Comprendre les enjeux de la mise en place d'un projet de machine learning.

Source des données

- Bulletin d'Analyse des Accidents Corporels
- 2005 à 2022



Classification

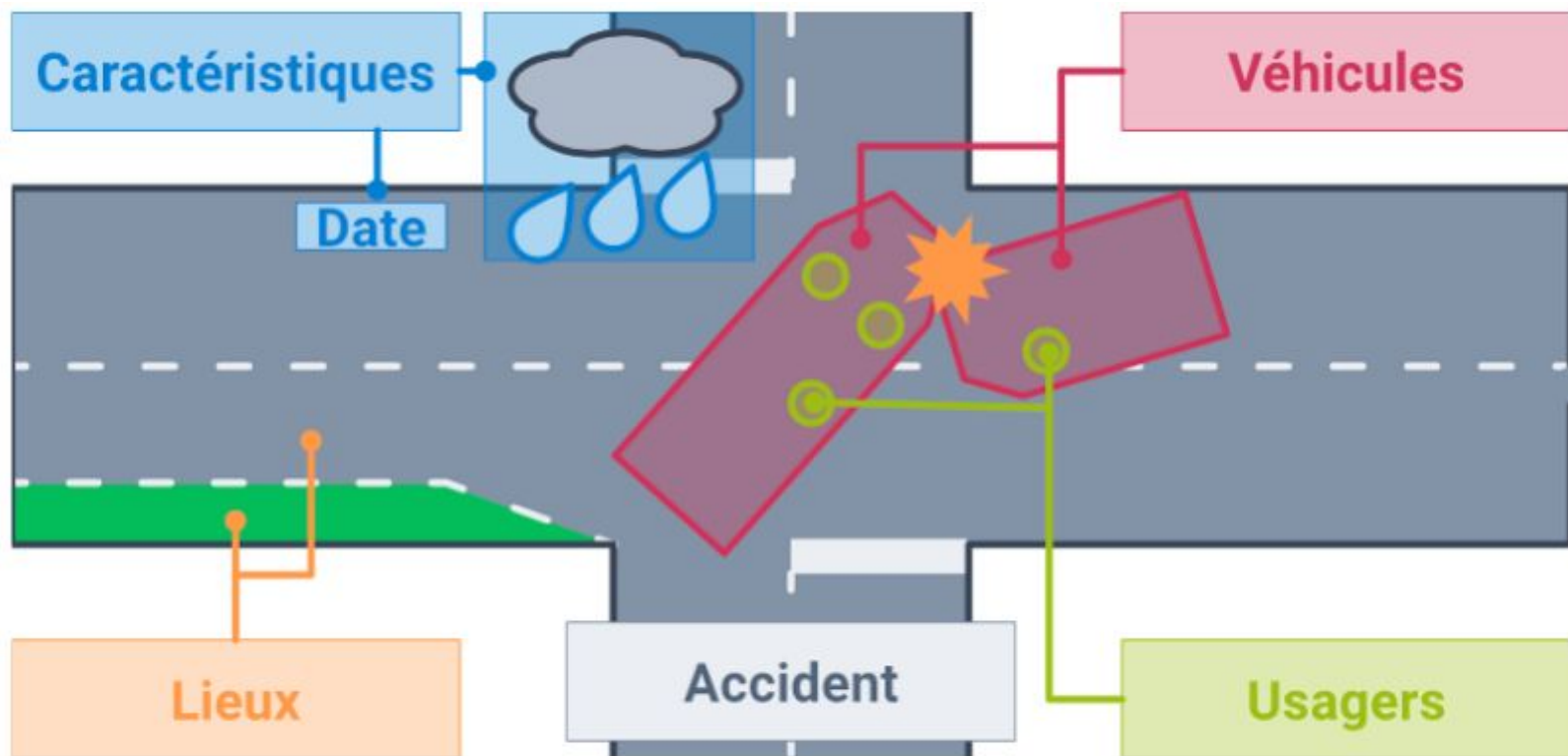


1. Indemne
2. Blessé léger
3. Blessé grave
4. Tué

Exploration des données



4 types de fichiers



Qualité des données

- La saisie des données évolue.
- La qualité des données évolue.
- Méthode de saisie de la variable cible change en 2018 !
- Nous étudions 2019 à 2022.

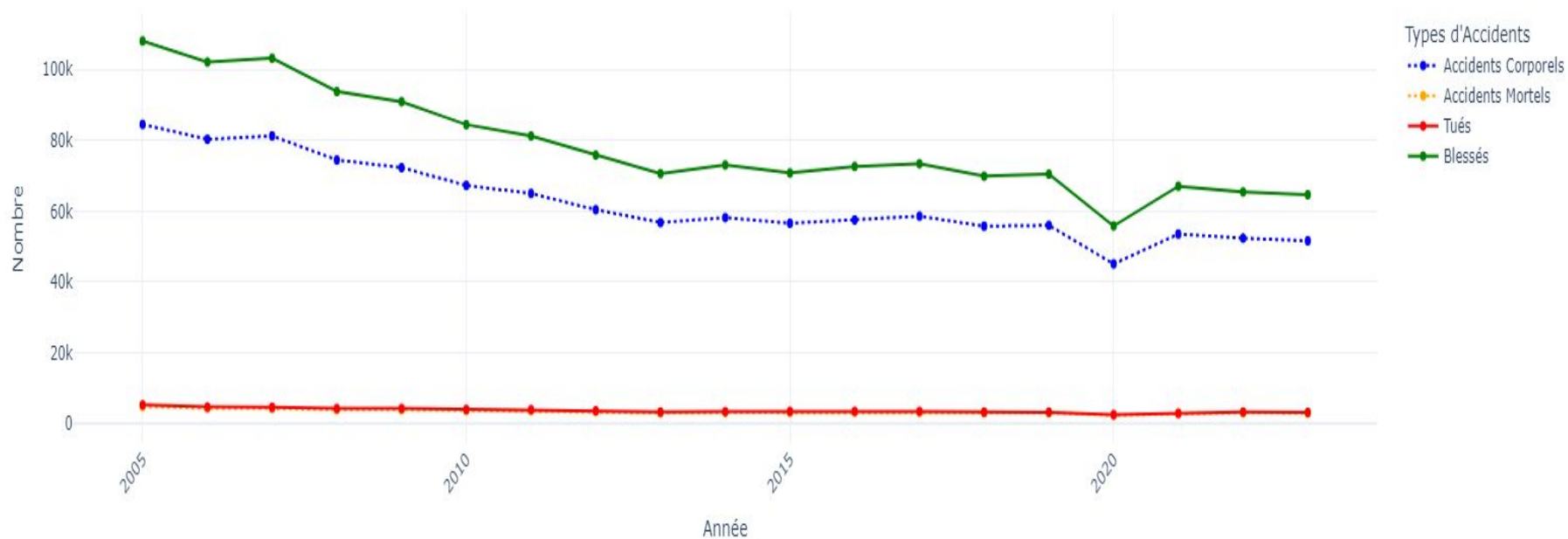
2005

2022



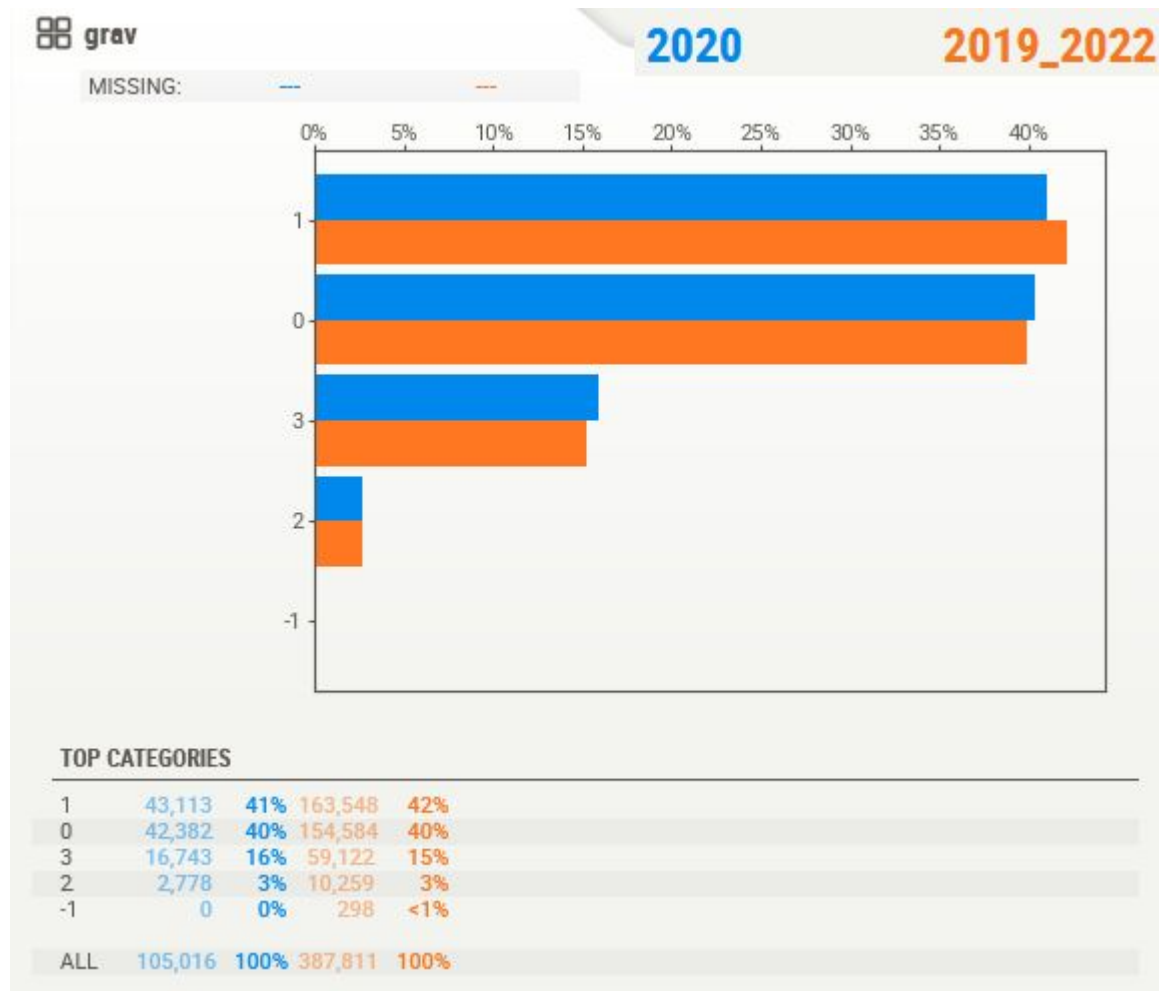
Le cas Covid-19

Évolution des Accidents Routiers en France Métropolitaine

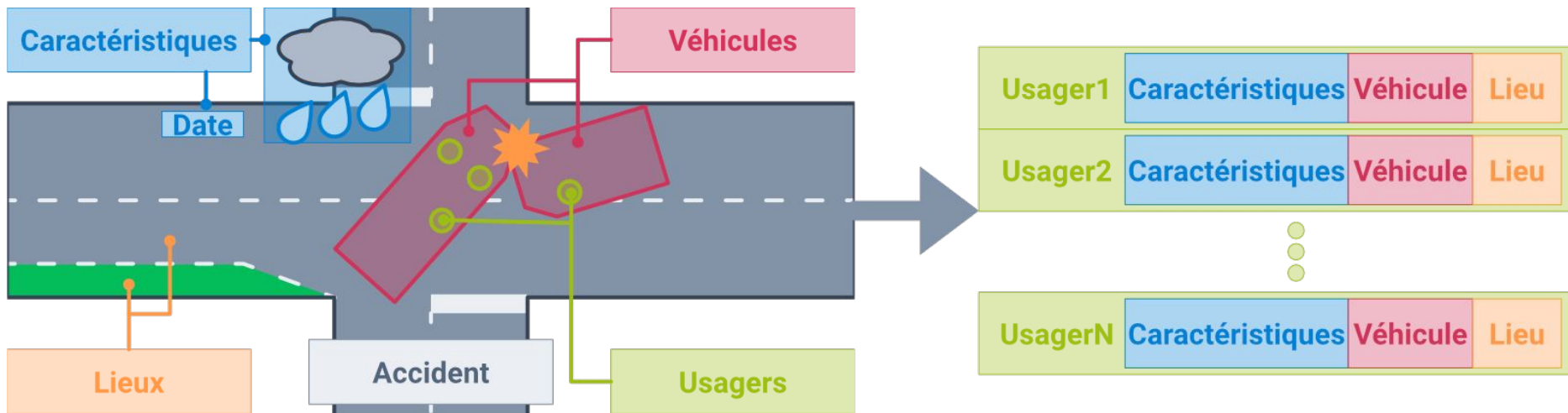


Le cas Covid-19

- Quelques mois de confinement, principalement en 2020.
- Pas d'impact 'majeur' sur la répartition des classes.

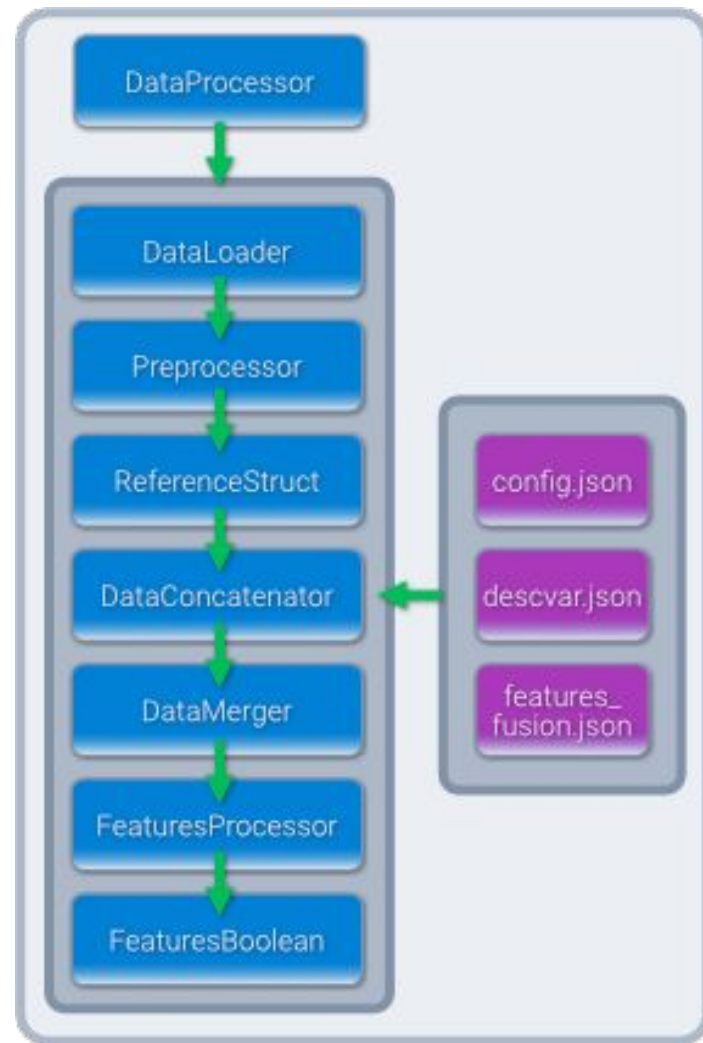


Prétraitement initial : Objectif

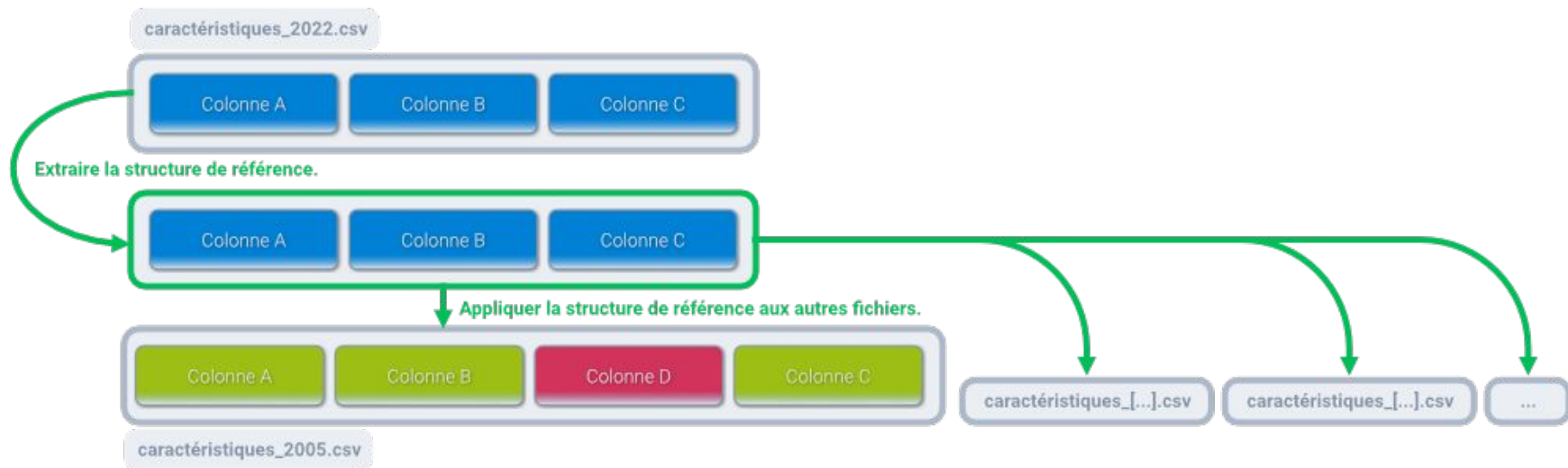


Prétraitement initial

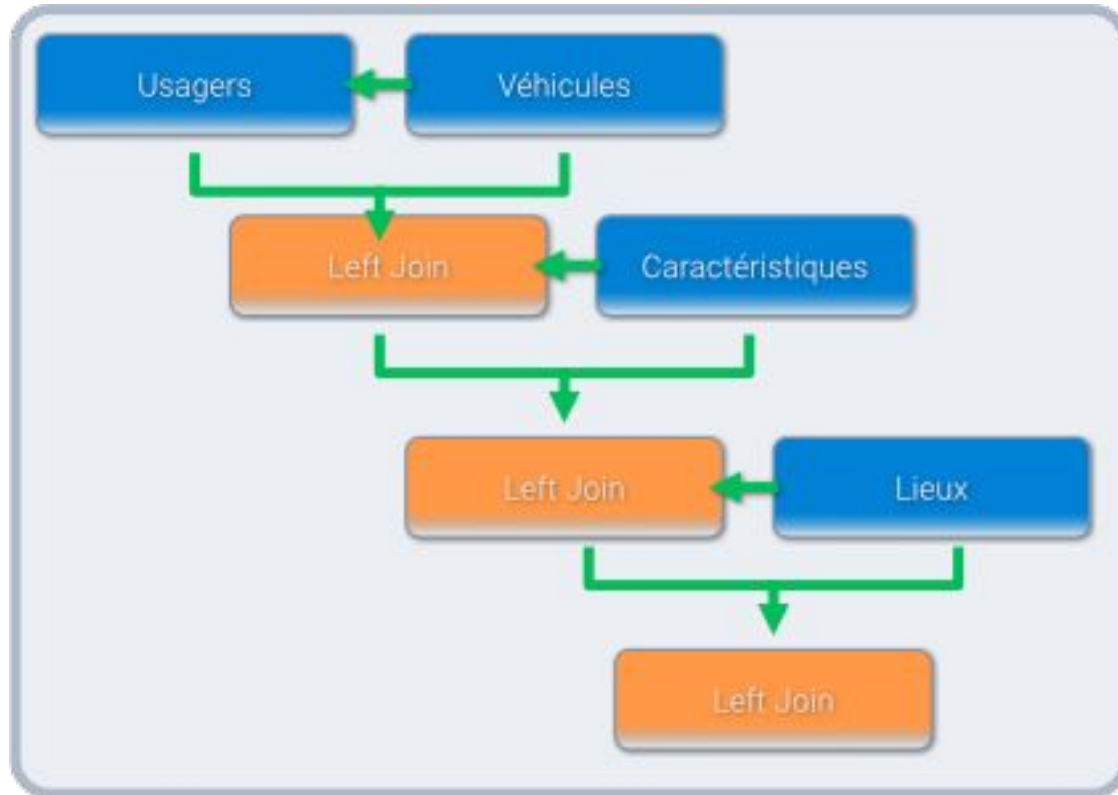
- Chargement des données
- Extraction d'une structure de référence
- Concaténation des données
- Fusion des données
- Features Engineering



Prétraitement initial : structure de référence

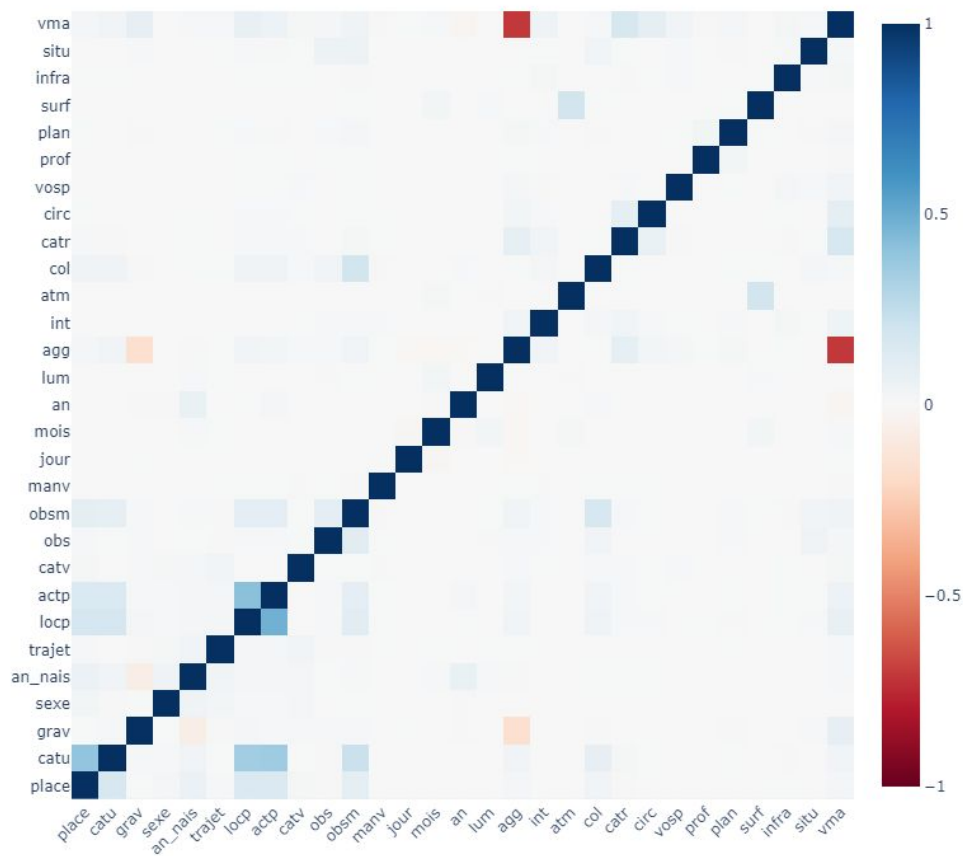


Prétraitement initial : fusion



Corrélation des features

Matrice de corrélation mixte (numérique & catégoriel)

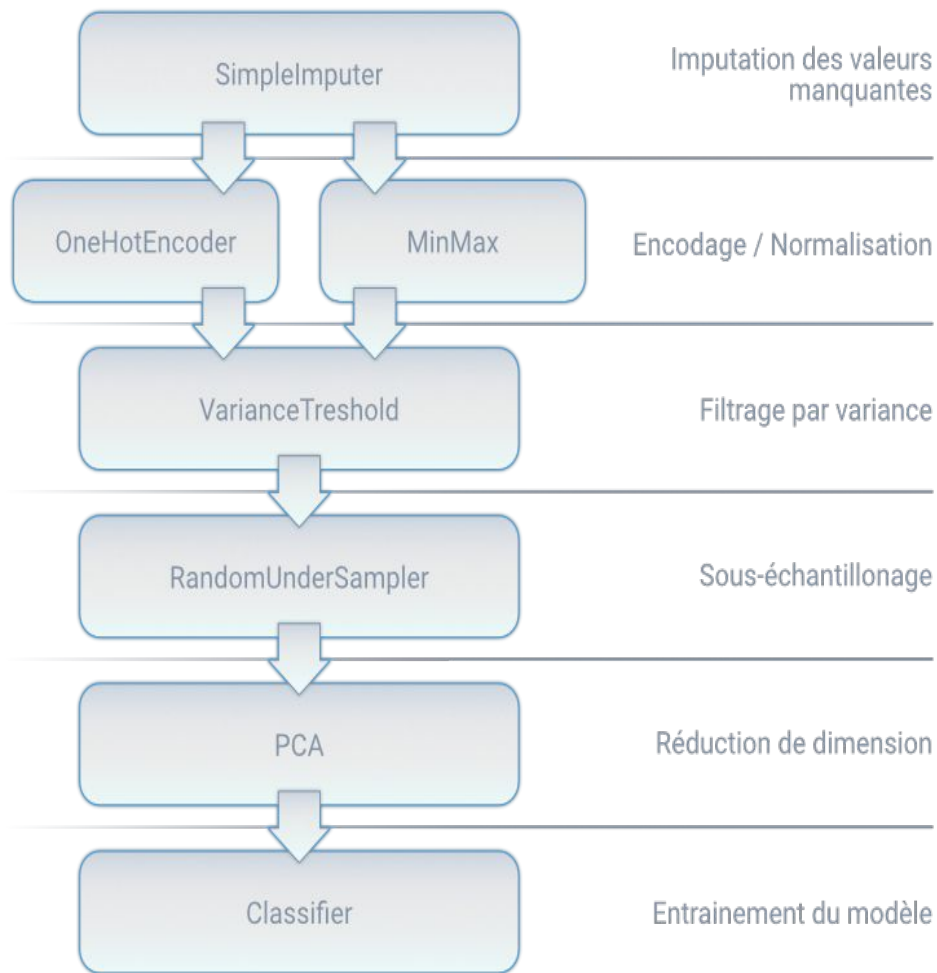


Modélisation



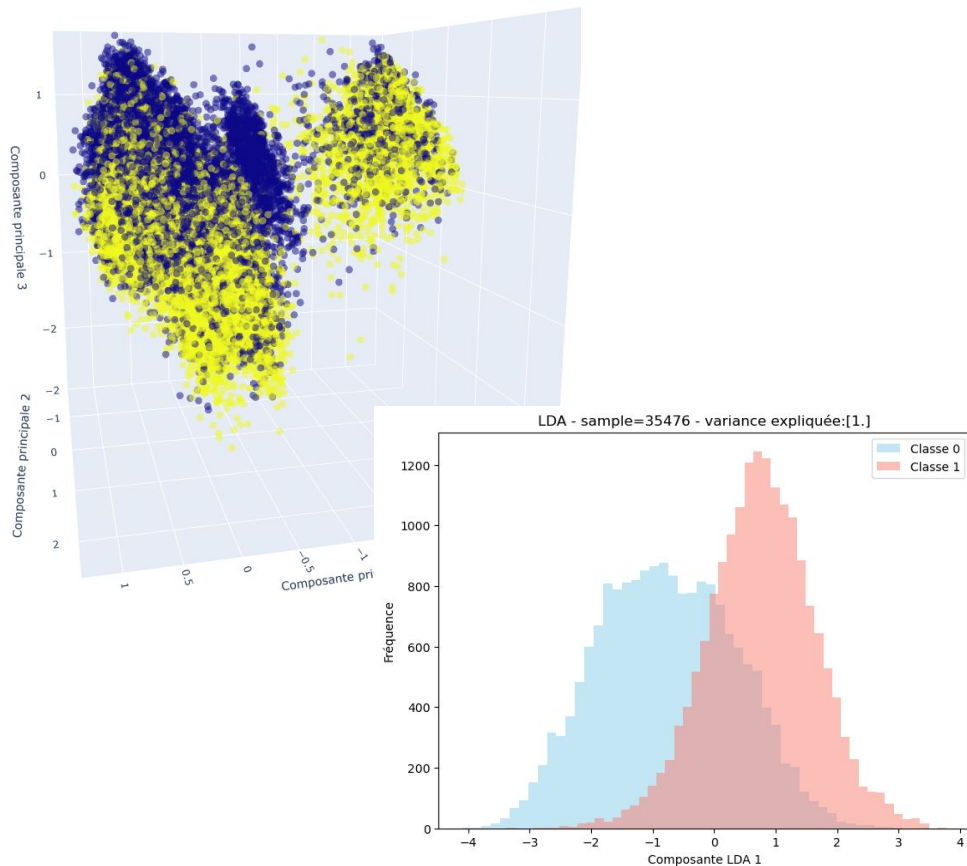
Pipeline

- Imputation des valeurs manquantes
- Encodage et standardisation
- Feature selection : Seuil de variance
- Rééquilibrer les classes (RUS)
- Réduction de dimension (PCA)
- Optimisation HParams :
(RandomizedSearch - [SKFold])
- Evaluation
- Sauvegarde + run MLFlow
- SHAP (si pas PCA)



Choix du modèle









- Logistic Regression
- Decision Tree
- AdaBoosting
- RandomForest
- XGBoost



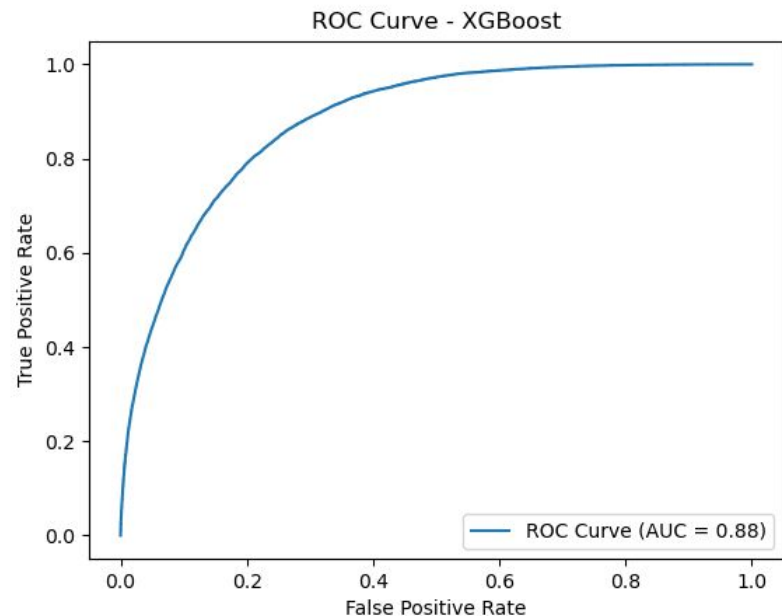
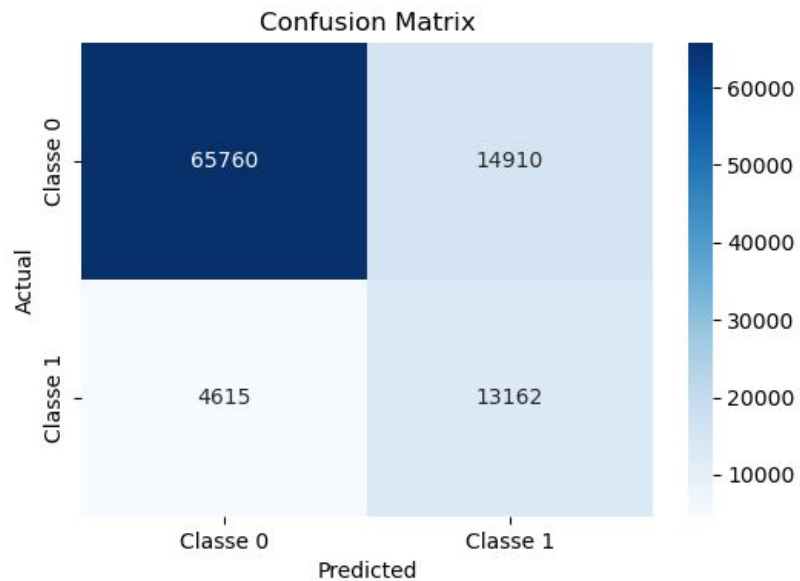
Résultats



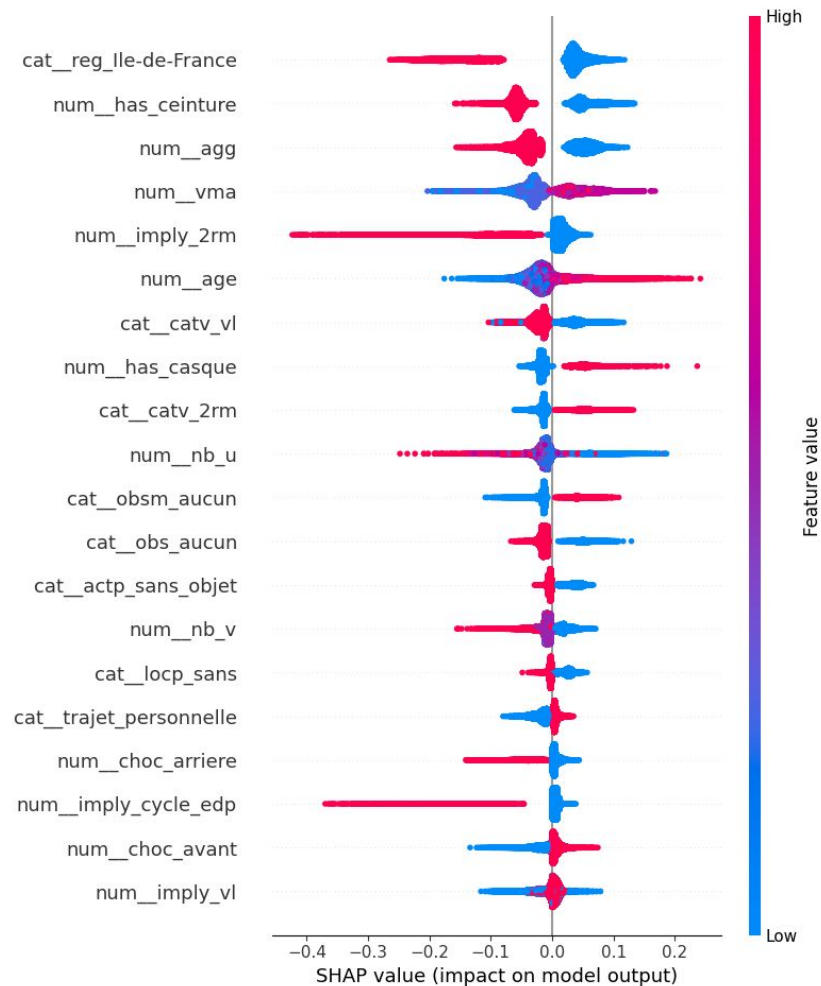
Résultats

				Metrics			
<input type="checkbox"/>	Run Name	Created	Duration	f1_score	recall 	accuracy	precision
<input type="checkbox"/>	 Evaluation_RandomForest...	 11 days ago	1.3min	0.846890448...	0.861001350...	-	-
<input type="checkbox"/>	 Evaluation_RandomForest...	 11 days ago	1.3min	0.846705451...	0.860869300...	-	-
<input type="checkbox"/>	 Evaluation_RandomForest...	 10 days ago	1.6min	0.845737547...	0.860402043...	-	-
<input type="checkbox"/>	 Evaluation_RandomForest...	 10 days ago	1.7min	0.845342846...	0.860188730...	-	-
<input type="checkbox"/>	 Evaluation_XGBoost	 16 days ago	2.4s	0.846675252...	0.860107468...	-	-
<input type="checkbox"/>	 Evaluation_XGBoost	 16 days ago	2.1s	0.846675252...	0.860107468...	-	-
<input type="checkbox"/>	 Evaluation_XGBoost	 16 days ago	2.4s	0.846675252...	0.860107468...	-	-
<input type="checkbox"/>	 Evaluation_XGBoost	 16 days ago	2.3s	0.846675252...	0.860107468...	-	-
<input type="checkbox"/>	 Evaluation_XGBoost	 17 days ago	2.2s	0.846675252...	0.860107468...	-	-
<input type="checkbox"/>	 Evaluation_XGBoost	 17 days ago	2.3s	0.846675252...	0.860107468...	-	-

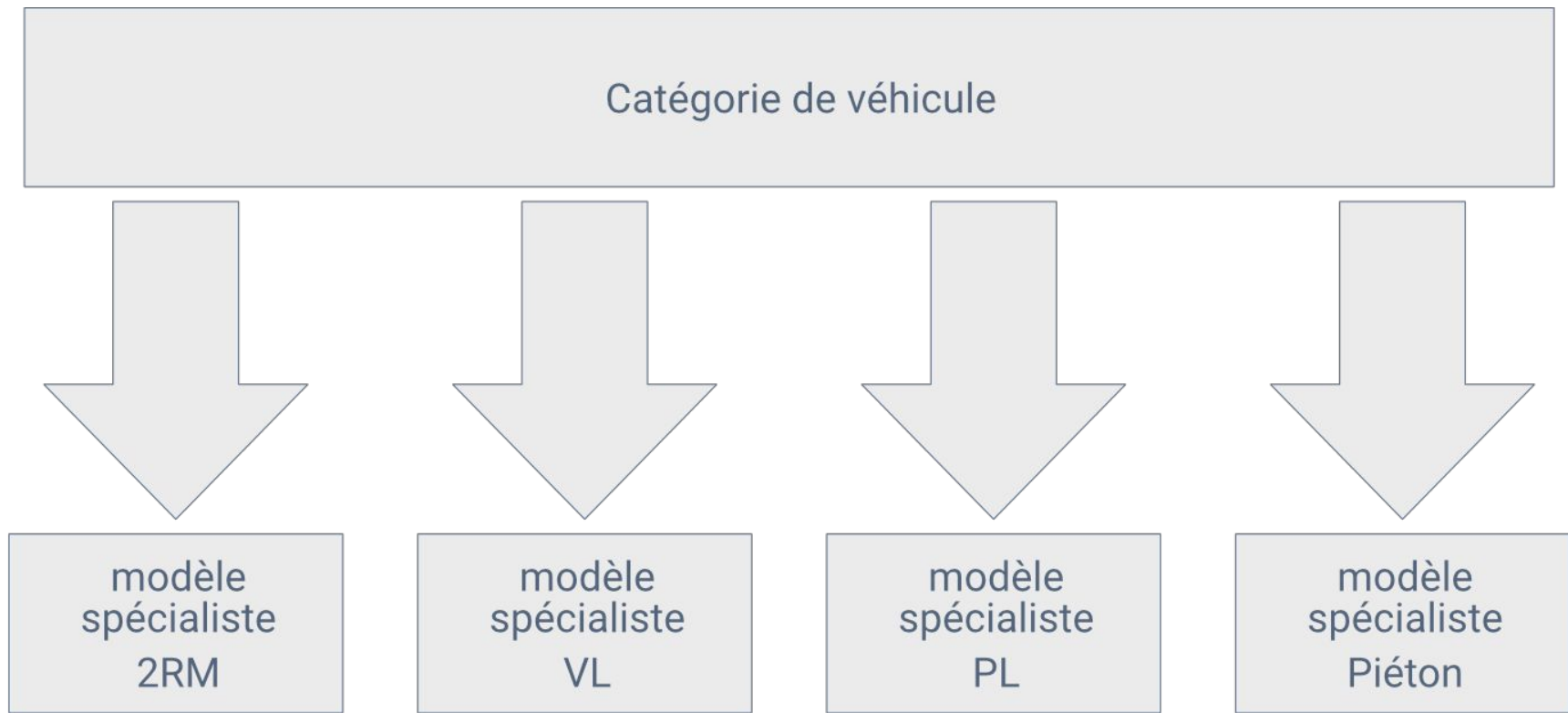
Métriques



Interprétation



Optimisation



Merci

