

Diagnostic et modélisation des données Covid19

- 1. Objectif :** Développement d'un modèle performant pour prédire l'état épidémique d'une personne (Positive, Négative) à travers ses données cliniques disponibles. Métrique : recall (sensibilité) > 85%, score F1 > 60%
- 2. Exploration et visualisation :** identification de la target et features, identification des valeurs manquantes et aberrantes, visualisation et interprétation de la target et features, exploration des différentes relations possibles entre features et target.
- 3. Prétraitement des données :** Manipulation et préparation des données pour la modélisation :
Séparation des jeu d'entraînement et test, élimination ou imputation des valeurs manquantes, élimination ou traitement des valeurs aberrantes néfastes pour la modélisation (non informatives), encodages des variables quantitatives, sélection des features utiles, génération des nouvelles features utiles, standardisation des variables suivant le modèle choisi.
- 4. Modélisation :** Développer un modèle pour répondre à notre objectif :
Mettre en place les critères et les outils d'évaluation nécessaires, entraînement des différents modèles en optimisant leurs hyper-paramètres à travers GridSearchCV, RandomizedSearchCV et le prétraitement des données, Evaluation et comparaison des modèles suivant leurs performances pour répondre à notre objectif, interprétation du la courbe d'apprentissage, prédiction de notre jeu de test.