Checklist de base : Machine Learning

# Analyse de forme

Variable target : SARS-Cov-2 exam result

Lignes/colonnes : 5644/111

Types des variables : 74 quantitatives, 37 qualitatives

Analyse des valeurs manquantes :

* Deux groupes se distinguent :
  + ~ 76 % : Tests concernant d’autres types de virus
  + ~89 % : Analyse de paramètres sanguins

# Analyse de fond

Visualisation de la target :

* 10 % de positifs
* 90% de négatifs (très déséquilibré !)

Très important lors de l’exploration de données : TRACER les histogrammes des variables continues

Signification des variables :

* Variables standardisées, asymétriques, test sanguins
* Age quantile : difficile à interpréter, les données ont été traitées ou transformées mathématiquement
* Variables qualitatives : binaires (0, 1) detected/not detected, viral, Rhinovirus qui semble très élevé

Relation variables/target :

Idées à tester :

* Leucocyte
* Monocyte
* Platelets : plaquettes

🡪 Ces taux sont différents entre les personnes testées positives au Covid19 et celles testées négatives. Il faut voir si cela à un sens de considérer que cela est lié.