## **Conclusion**

- Y a-t-il des algorithmes/stratégies/critères pour identifier les bonnes valeurs des hyperparamètres (en ACP, en regroupement hiérarchique, arbres, méthodes d'ensemble)?
- Comment définir/construire les variables  $X_1,\dots,X_p$  (feature engineering) ?
- Comment identifier les données aberrantes et décider si on les enlève ou les garde?
- Comment scinder les données en entrainement vs validation?
- Les données sont-elles représentatives de la "population cible"?
- On fait quoi si pour certaines observations, des valeurs de certaines variables sont manquantes ?