

ALGORITHME POUR CALIBRER UNE FORÊT ALÉATOIRE

- 1) Lancer une première forêt aléatoire pour avoir une première évaluation de la performance prédictive du modèle.
 - 1) Ici, si on dispose d'un grosse base (par exemple 1 000 000 individus, 50 variables), penser à lancer la forêt sur un sous-échantillon de 20 000 – 30 000 individus → gros gain de temps et faible perte en terme de qualité de modèle
- 2) Optimiser un ou plusieurs paramètres de manière à maximiser la performance du modèle
 - 1) Dans l'exemple on a optimiser le paramètre « nodedepth »
 - 2) En général, lorsque la performance d'un modèle est faible (faible R^2 ou autre critère d'évaluation), diminuer la taille des arbre construits améliore la performance du modèle. (souvent le cas en assurance auto, assurance RC, assurance habitation, etc ...)