Commentaires Classification

# 8\_

## Nettoyage des données

Comment traiter les valeurs manquantes sans impacter la classification.

## Analyse de la pertinence des attributs

Savoir si on prend en compte toutes les colonnes. Voir si des attributs sont corrélés pour en supprimer certains. Pour réduire le training set.

## Robustesse

Cmmt se comporte le modèle si les données on du bruit ou des données manquantes.

## Transformation

On peut discrétiser les données en fct de l’importance des données.

## Interprétabilité

Taille des règles, compréhension du modèle.

# Arbre de décision

## 9\_

Si alors sinon

## 12\_

Glouton (greedy) choix optimal à chaque étape.

## 13\_

But : calculer la quantité de données relative à l’information à gérer.

## 14\_

Principal défaut : sensibilité et sur-apprentissage

Bagging (Bootstrap Aggregating)

-> Bootstrapping

-> Aggregating

Forets aléatoires

-> Échantillonnage des attributs pour décorréler les arbres construits

-> Pas d’élagage sur- apprentissage compensé par le bagging

# Classification Bayésienne

P(18°|n)  <= => R> dnorm(18, means = 25, sd = 4)