

Exploration des données à l'aide des arbres de décision

Grégoire Le Campion, Hugues Pécout

Atelier, mercredi 14 octobre 13h45-16h30

Présentation générale

Il existe en statistiques un certain nombre d'outils pour tenter de prédire, d'expliquer, classer des variables. Les méthodes les plus connues sont notamment l'analyse factorielle, ou encore les méthodes de régression.

Ces différentes méthodes bien qu'extrêmement intéressantes ont de nombreuses conditions qu'il n'est pas toujours aisé de remplir en SHS, et fournissent des résultats pas toujours simple à interpréter.

L'idée ici est de vous présenter une méthode alternative et tout-terrain : l'arbre de décision avec ses avantages et ses limites.

Environnement informatique

Tout environnement (Linux, Mac, Windows), en revanche nécessité d'avoir installer sur l'ordinateur un navigateur internet (firefox, chrome...). Dans l'optique d'un éventuel approfondissement du sujet (facultatif) R et Rstudio seront nécessaires.

Type de données traitées - droit d'accès

Données quantitatives et qualitatives, mais non textuelles, mise sous forme base de données.

Niveau requis

Aucun pré-requis nécessaire

Objectifs

Comprendre ce qu'est et permet un arbre de décision. Pouvoir utiliser cette méthode d'analyse à l'issue de l'atelier de manière autonome

Formule pédagogique

Atelier « bring your data »

Références bibliographiques

- T.Hastie, R.Tibshirani et J.Friedman : The Elementsof Statistical Learning : Data Mining, Inference, andPrediction.. Springer, 2nd ed., 2009
- T.Hastie, R.Tibshirani et J.Friedman : The Elementsof Statistical Learning : Data Mining, Inference, andPrediction. Springer, 2nd ed., 2009