



INTRODUCTION À DATA SCIENCE

MACHINE LEARNING AND DATA MINING

Responsable(s): Alexandre SAIDI, Céline HARTWEG-HELBERT, Marc JACOB

| Cours : 16.0 | TD : 0.0 | TP : 0.0 | Autonomie : 0.0 | BE : 12.0 | Projet : 0.0 | Langue du cours : FR

Objectifs de la formation

Donner aux élèves un aperçu et la pratique des bases de cette discipline. Un tour non exhaustif des méthodes les plus utilisées sur des données transactionnelles (ce sont en général des données non-image car les données images nécessitent d'autres traitements).

Mots-clés : Fouille de données, Extraction de connaissances à partir de données, Data Science, Machine learning.

Programme

- La problématique scientifique de l'apprentissage artificiel (Machine Learning)
- Formalisation et fondements Algorithmiques / Statistiques
- Quelques applications typiques
- Apprentissage supervisé, Non-(et semi-) supervisé,
- Classification, régression, association (Introduction à la fouille de textes) - Arbres de décision / de régression
- Méthodes statistiques (Bayes, SVM, Méthodes à base de noyaux)
- Méthodes et indicateurs d'évaluation statistiques
- Méthodes de Classification (Clustering)

Compétences

- Initiation & pratique de Data Science : en présence des données, l'élève devrait être capable de choisir les méthodes applicables. Il devrait ensuite pouvoir évaluer les résultats et choisir la meilleure méthode.
- Rappel des éléments statistiques (de base).
- Apporter un outil d'aide à la décision.

Travail en autonomie

Objectifs : Permettre aux élèves de pratiquer le Data science à travers les méthodes les plus employées.
L'outil utilisé est WEKA.
Le meilleur outil de nos jours serait Python (+ les bibliothèques de Data Science) mais le public de ce MOD n'est pas forcément Informaticien. Il risque

Méthodes :

Bibliographie

u.m. Fayyad & al. From Data Mining to Knowledge Discovery in Databases. 1996.
stan matWin & al. Challenges in Computational Statistics and Data Mining. Jan Mielniczuk, 2016.
i.h. Witten, e. Frank. Data Mining - practical ML Tools & Techniques. 2005.

Contrôle des connaissances

Test écrit + BEs.
Savoir : 66%
Savoir Faire (rendus des BEs) : 33%