

Fouille de données - Data mining

Niveau : Master 2

Cursus : Data Science et Management de l'Innovation (DSMI)

Semestre : 9

Volume horaire : 30 heures de cours

Intervenant : Christine LARGERON

Contact : [email](#)

Objectifs du cours : Description (5 à 10 lignes).

L'extraction automatique de connaissances (ECD - Knowledge Discovery in Data Bases : KDD) à partir de données peut être utilisée par exemple pour faire une typologie de produits en catégories, pour établir le profil d'un client, pour prédire le sujet d'un document qu'il s'agisse d'un message électronique ou d'un article scientifique ou plus généralement pour l'aide à la décision. Après une étape de pré-traitement des données, le processus comporte une phase de fouille de données (Data Mining) dont l'objectif est de rechercher des informations nouvelles ou cachées à partir des données en faisant appel à des techniques issues de domaines divers tels que les statistiques ou l'intelligence artificielle. Le but de cet enseignement est de présenter le processus d'extraction automatique de connaissance ainsi que quelques algorithmes classiques de fouille de données et de les appliquer sur des problèmes concrets en relation avec des problématiques de veille, notamment veille marketing, technologique ou encore veille image.

Plan de cours :

I- Introduction à l'ECD

Définitions

Data Mining

Processus d'ECD

II- Méthodes de fouille de données

Analyse en composante principale (ACP) - TSNE

Classification automatique (k means et méthodes hiérarchiques)

Les règles d'association

Les arbres de décision (Cart, C5) – Random forest

Text mining

Social mining

Support logiciel utilisé :

Statistica et R (igraph et TM)

Prérequis (à mentionner uniquement s'il y en a) :

Saporta G. Probabilités, analyse des données et statistique, troisième édition, 622pages, Editions Technip, Paris, 2011

<https://www.fichier-pdf.fr/2012/06/15/probabilites-analyse-de-donnees-et-statistiques/>