

Analyse de données - Data analysis

Niveau : Master 1 APE

Cursus : Master 1 APE

Semestre : semestre 7

Volume horaire : 18 heures de cours (CM) et 12 heures de pratique (TD)

Intervenant : Christine LARGERON

Contact : [email](#)

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

L'extraction automatique de connaissances (ECD - Knowledge Discovery in Databases : KDD) à partir de données peut être utilisée par exemple pour établir le profil d'un client, pour aider au diagnostic ou plus généralement pour l'aide à la décision. Après une étape de pré-traitement des données, le processus comporte une phase d'analyse et de fouille de données (Data analysis - 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 méthodes classiques d'analyse de données et de les appliquer sur des problèmes concrets.

Knowledge Discovery in Databases (KDD) can be applied for example to establish the profile of a customer, to automatically provide diagnosis or more generally for the decision support. After a data pre-processing, it includes a step of data mining whose objective is to search for new or hidden information from the data by using techniques from various fields such as statistics or artificial intelligence. The aim of this course is to present the process of automatic knowledge extraction as well as some classical methods of data analysis and to apply them to concrete problems.

Approche méthodologique :

Après la présentation des fondements théoriques de chaque méthode, celle-ci sera mise en œuvre sur un jeu de données en utilisant le logiciel Statistica et R. Les étudiants devront ensuite traiter de façon autonome un autre cas concret.

Plan de cours :

I- Introduction à l'ECD

Définitions

Data Mining

Processus d'ECD

II- Méthodes d'analyse de données

Pré-traitement des données

Analyse en composante principale (ACP) - TSNE

Analyse factorielle des correspondances (AFC – ACM)

Classification automatique (k means et méthodes hiérarchiques)

I- Introduction to ECD

Définitions
Data Mining
ECD process

II- Data analysis methods

Pre-processing of the data
Principal component analysis – (PCA) – TSNE
Correspondance Analysis (CA)
Clustering (k means et hierarchical methods)

Approche méthodologique

Après la présentation des fondements théoriques de chaque méthode, celle-ci sera mise en œuvre sur un jeu de données en utilisant le logiciel Statistica et R. Les étudiants devront ensuite traiter de façon autonome un autre cas concret.

After a presentation of the theoretical foundations of each method, it will be implemented on a dataset using the softwares Statistica and R. Then the students will have to deal with another concrete case.

Modalités d'évaluation :

- Etudes de cas et examen
- Case studies and examination

Support logiciel utilisé :

- Statistica – R

Compétences recherchées :

- comprendre les fondements mathématiques des méthodes
 - savoir quand utiliser chaque méthode
 - interpréter les tables et graphiques issus de ces méthodes
 - être capable de mener soi-même une telle étude.
-
- understand the mathematical foundations of the methods
 - know when each method can be applied
 - interpret correctly the tables and graphs resulting from these methods
 - be able to carry out such a analysis yourself.

Bibliographie / références : Donner quelques références.

- Bernadet, M., Largeron C., Richard, O. De la notion de probabilité à l'estimation, PUL
- Saporta G. Probabilités, analyse de données et statistiques, 622pages, Ed Technip, Paris, 2011
- <https://www.fichier-pdf.fr/2012/06/15/probabilites-analyse-de-donnees-et-statistiques/>