

Data analysis

Level: Master 1 APE

Curriculum: Master 1 APE

Semester: Semester 7

Hours: 18 hours of lectures (CM) and 12 hours of practical work (TD)

Lecturer: Christine LARGERON

Contact: christine.largeron@univ-st-etienne.fr

Course objectives:

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.

Methodological approach:

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.

Course outline:

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)

Assessment methods :

- Case studies and examination

Software support :

- Statistica – R

**Skills required :**

- 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.

Bibliography / references:

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/>