SAINT-ETIENNE SCHOOL OF ECONOMICS

Data mining

Level: Master 2
Curriculum: Data Science and Innovation Management (DSMI)
Semester: 9
Hours: 30 hours of classes
Lecturer: Christine LARGERON
Contact: christine.largeron@univ-st-etienne.fr

Course objectives:
Automatic knowledge extraction (ECD - Knowledge Discovery in Data Bases: KDD) from data can be used, for example, to classify products into categories, to establish a customer profile, to predict the subject of a document, be it an e-mail message or a scientific article, or more generally to help scientific article or, more generally, for decision support.
After a data pre-processing stage, the process includes a data mining phase, the aim of which is to search for new or hidden information from the data, using techniques from a variety of fields such as statistics and artificial intelligence.
The aim of this course is to present the process of automatic knowledge extraction, as well as some classic algorithms, and to apply them to concrete problems related to market intelligence, in particular marketing, technology or image intelligence.

Course outline:
I- Introduction to EDC
        Definitions
        Data Mining
        The DCE process
II- Data mining methods
        Principal component analysis (PCA) - TSNE
        Automatic classification (k means and hierarchical methods)
        Association rules
        Decision trees (Cart, C5) - Random forest
        Text mining
        Social mining

Software support: Statistica and R (igraph and TM)

Prerequisites (to be mentioned only if any):
Saporta G. Probabilités, analyse des données et statistique, third edition, 622pages, Editions Technip, Paris, 2011
https://www.fichier-pdf.fr/2012/06/15/probabilites-analyse-de-donnees-et-statistiques