

## **Data Mining (ELMIM422)**

**Academic Year :** 2014 – 2015

**Instructor :** Isabelle Linden

**Email:** [isabelle.linden@unamur.be](mailto:isabelle.linden@unamur.be)

**Credits:** 5 ECTS

### **Learning Goals**

#### Knowledge and Scientific reasoning

At the end of the course the student will be able

- to explain and use various data mining techniques,
- to explain the relevancy and to use various tools for presenting data,
- to explore the relevant use of various techniques in order to manage them.

Assessment: reports, talks and final exam

#### Written and oral communication

Each student shall be able to communicate in an organized, clear, and persuasive manner both by oral and written communication. Written communication will respect scientific standard.

Assessment: reports, talks and final exam

### **Course Description**

In addition to information about its own activities, an organisation generates large data sets. These data provide strategic information if they are properly stored, analysed and handled. The course of data mining and business intelligence explores various techniques and tools which help to make data "speak" in a business-orientated perspective.

### **Reference Textbook**

*Mohammed J. Zaki and Wagner Meira Jr., Data Mining and Analysis, fundamental concepts and algorithms, Cambridge University Press, 2014*

### **Step-Wise topics:**

1. Introduction to Data Mining
2. Data analysis foundations: (reminder) descriptive analysis, data cleaning, dimension reduction
3. Classification: application examples, probabilistic classification, decision trees and rules, linear discriminant analysis, logistic regression, support vector machine
4. Classification assessment: performance measures and graphs, classifier evaluation, bias-variance decomposition

5. Frequent pattern mining : itemset mining, summarizing itemsets
  6. Pattern and rule assessment
  7. Clustering: application examples: representative-based clustering, hierarchical clustering, density-based clustering
  8. Clustering evaluation, validation and interpretation
- + 2 special lectures:
1. data management and privacy
  2. Scientific writing standard

**Course material:**

Slides are available on the Webcampus platform

**Homework:**

Two assignments will be realised: a classification study (written report and oral presentation) and paper presentation. The description of deliverables and assessment criteria are available on the Webcampus platform.

**Evaluation Components:**

The course grade will be based on the homework and oral presentations (50%) and an oral individual final exam (50%)