# Data Mining

**Lecturer**
Mahmoud SAKR (Coordinator)

**Course mnemonic**
INFO-H423

**ECTS credits**
5 credits

**Language(s) of instruction**
English

**Course period**
First term

**Campus**
Solbosch

## Course content

Our world is getting more and more driven by data. The amounts of data collected is massively growing. Data Mining (a.k.a Knowledge Discovery in Database) is the science of extracting implicit, non-trivial, and potentially useful information from data. Three main tasks that data mining methods try to perform are: (1) to model the regularities in the data in the form of compact models and patterns, (2) to spot the irregularities, (3) to predict the future.

**The course will cover the following topics:**

- Classification.
- Model validation and data preparation
- Clustering
- Time series forecasting
- Spatial and Spatiotemporal data mining
- Frequent pattern and association rule mining.
- Stream data mining
- Outlier mining
- Applications.

## Objectives (and/or specific learning outcomes)

**This is a first course in Data Mining, with the following goals:**

- To introduce the fundamental concepts and techniques of data mining
- To develop skills of using recent data mining software for solving practical problems
- To establish the main characteristics and limitations of algorithms for addressing data mining tasks

- To select the most appropriate combination of algorithms to solve a data mining problem
- To develop and execute a data mining workflow on real-life datasets, and to solve a data-driven analysis problem
- To identify promising business applications of data mining

## Pre-requisits and co-requisits

### Required knowledge and skills

This course assumes that participants have background on the following topics. There will be no formal check of this background at the inscription time. It is up to the participants to see and decide themselves.

- Good knowledge of programming, with hands on experience in atleast one programming language
- General knowledge of data structures, algorithms, and complexity analysis
- General knowledge of databases and SQL

## Teaching method and learning activities

The course activities include lectures, exercise sessions, invited industry talk, industry visit if COVID-19 allows, and a group project. The lectures aim at building a fundamental understanding of popular data mining techniques strengths and limitations, as well as their associated computational complexity issues. The exercise sessions use Rapidminer as a tool for quickly prototyping data mining applications, suitable for engineers. The industry activities aim at connecting to the application world and get an idea about the state of practice. Finally the project is a chance to compile and further extend on the studied topics in a real big data mining application.

### References, bibliography and recommended reading

**Textbook:**

- Charu C. Aggarwal. Data Mining: The Textbook. Springer, 2015. (Available for download through Cible+).

**Other useful readings:**

- David J. Hand, Heikki Mannila, Padhraic Smyth. Principles of Data Mining. MIT Press, 2001.
- Delmater Rhonda, Hancock Monte. Data Mining Explained. Digital Press, 2001.
- Pang-Ning Tan, Michael Steinbach, Vipin Kumar. Introduction to Data Mining. Pearson Education (Addison Wesley), 2006.
- Rob J Hyndman and George Athanasopoulos. Forecasting: Principles and Practice [https://otexts.com/fpp2/] (2nd ed). Monash University, Australia

## Course notes

Université virtuelle

## Other information

### Place(s) of teaching

Solbosch

### Contact(s)

Lecturer: Mahmoud SAKR <mahmoud.sakr@ulb.be>

Assistant: Jean-Philippe HUBINONT <jean-philippe.hubinont@ulb.ac.be>

## Evaluation method(s)

written examination and Project

### Determination of the mark (including the weighting of partial marks)

The grade of the course will be split as follows:

> Group project 40%

> Written exam 60%

## Main language(s) of evaluation

English

## Programmes

### Programmes proposing this course at the Brussels School of Engineering

MA-IRIF | Master of science in Computer Science and Engineering | finalité Professional/unit 2 and finalité Big Data Management and Analytics (Erasmus Mundus)/unit 1 and MS-BGDA | Specialized Master in data science, Big data | unit U

### Programmes proposing this course at the Solvay Brussels School of Economics and Management

MS-BGDA | Specialized Master in data science, Big data | unit U

### Programmes proposing this course at the faculty of Sciences

MA-INFO | Master in Computer science | finalité Professional/unit 2 and MS-BGDA | Specialized Master in data science, Big data | unit U