Data Mining

2025 - 2026

Is part of the next programmes:

- M0012004 Master of Computer Science: Software Engineering
- M0012005 Master of Computer Science: Data Science and Artificial Intelligence
- M0012006 Master of Computer Science: Computer Networks
- M0048004 Master of Computer Science: Software Engineering
- M0048005 Master of Computer Science: Data Science and Artificial Intelligence
- M0048006 Master of Computer Science: Computer Networks
- M0090004 Master of Teaching in Science and Technology: Computer Science
- U0001008 Courses open to exchange students in Sciences

Course Code:	2001WETGDT
Study Domain:	Computer Science
Semester:	2E SEM
Contact Hours:	26

Credits:	6
Study Load (hours):	168
Contract Restrictions:	No contract restriction
Language of Instructions:	ENG
Lecturer(s):	T Bart Goethals C Toon Calders
Examperiod:	exam in the 2nd semester

1. Prerequisites *

speaking and writing of:

- English

specific prerequisites for this course

*** Programming experience**
*** Data structures, Algorithms, Complexity**
*** Introduction to Artificial Intelligence**

2. Learning outcomes *

- Display a comprehensive understanding of different data mining tasks, including classification, clustering, outlier detection, pattern mining, and recommenders.
- Reproduce the main characteristics and limitations of algorithms for addressing data mining tasks.
- Select, based on a problem description of a data mining problem, the most appropriate combination of algorithms to solve it.

- Analyze the models resulting from a data mining exercise and identify threaths to validity such as model bias, under- and overfitting.
- Develop and execute a data mining workflow on a real-life dataset to solve a data-driven analysis problem.

3. Course contents *

After a short introduction to data mining, we study and discuss several advanced data mining techniques. The data mining techniques that will be addressed are divided into the following categories:

- Classification:
  - Standard classification techniques, such as k-nearest neighbors, decision trees, Bayesian classifiers, logistic regression, as well as techniques for combining classifiers in ensembles (bagging and boosting, random forest)
  - important concepts: likelihood, regularization
  - common issues: under- and overfitting
  - evaluation measures and techniques for classifiers: Accuracy, precision/recall, AUC, hold-out, cross validation
- Clustering: k-means and k-medoids, density based clustering (DBSCAN), Expectation-Maximization-based clustering, hierarchical clustering
- Outlier detection
- Pattern mining: frequent itemset mining, subgroup discovery
- Recommender systems: user-user, item-item recommendation, collaborative filtering

During the coverage of these topics, several foundational concepts in machine learning and data mining will be treated, such as bias-variance decomposition, maximum likelihood learning, minimal description length principle, etc.
The course will also contain a practical component; students have to make projects individually.

4. International dimension *

- This course stimulates international and intercultural competences.
- Students use course materials in a foreign language.
- Students write papers in a foreign language.

5. Teaching method and planned learning activities

5.1 Used teaching methods *

Class contact teaching

- Lectures
- Seminars/Tutorials
- Guest lectures

Personal work

Assignments

- Individually

Project

- Individually

5.2 Planned learning activities and teaching methods

The course consists of theory lectures combined with practical assignments. 4 practical assignments will be given throughout the year, with main topics:

- pattern mining
- classification
- clustering
- recommenders

5.3 Facilities for working students *

Classroom activities

- Lectures: recording available via video link on Blackboard
- Seminars/tutorials: alternative assignment possible

- Practica: free to choose the group division

Others

Recordings of previous year made available

6. Assessment method and criteria *

6.1 Used assessment methods *

Examination

- Written examination without oral presentation
- - Closed book

Other assessment methods

- Project
- Presentation

6.2 Assessment criteria *

The course is evaluated with a written exam and four individual projects. In order to pass the course, students need to obtain at least 40% on both components, and a passing grade in general. Hence, only a limited form of compensation between the two components is possible. In case the student passes the course, the assigned grade will be the average of the two components, otherwise the lowest of the two grades will be assigned.

- written exam (50%): **closed book exam** with questions testing the theoretical topics covered in the course. The evaluation is based on the understanding of the data mining techniques and algorithms, their theoretical foundations, capabilities and limitations.
- individual assignments (50%): The students have to carry out 4 assignments related to pattern mining, recommender systems, classification, and clustering respectively. The project will be evaluated on the basis of the appropriateness of the techniques used, quality of the designed data mining workflow, correctness

of the conclusions drawn, and accuracy and quality of the solution. The results will have to be presented in a report and submission of the code.

Copied Code - Code created by Generative AI

For implementations supporting the project, students may copy code from the internet, but must do so explicit attribution of the source where it is copied from. Students may use generative AI tools for the code adopted in their project, but must do so with explicit acknowledgment, i.e. it must be clearly indicated which code was created by which tools. Students may be asked to elaborate on how they used generative AI.

Support from generative AI

Students may use generative AI tools as for writing reports, similar to initial search engines such as Google and for checking grammar and spelling. Students may be asked to elaborate on how they used generative AI.

7. Study material

7.1 Required reading *

- "Introduction to Data Mining" by Pang-Ning Tan, Michael Steinbach, Vipin Kumar (Addison-Wesley)
- Course notes: lecture slides, exercises, solutions to selected exercises

7.2 Optional reading

The following study material can be studied voluntarily :

- Research papers made available via Blackboard.
- Jiawei Han, Micheline Kamber, and Jian Pei, Data Mining: Concepts and Techniques, 3rd edition, Morgan Kaufmann, 2011.

8. Contact information *

- Prof. dr. Bart Goethals (bart.goethals@uantwerpen.be)
- Prof. dr. Toon Calders (toon.calders@uantwerpen.be)

9. Tutoring

Questions can be addressed to the teacher and assistants by email or after making an appointment.