GHENT UNIVERSITY

Course Specifications

Valid as from the academic year 2024-2025

Big Data Science (C003802)

Course size *(nominal values; actual values may depend on programme)*

Credits 5.0 **Study time 150 h**

Course offerings and teaching methods in academic year 2024-2025

A (semester 2)	English	Gent	lecture
			independent work
			seminar

Lecturers in academic year 2024-2025

Peralta Cámara, Daniel	WE02	lecturer-in-charge
Goetghebeur, Els	WE02	co-lecturer

Offered in the following programmes in 2024-2025	**crdts**	**offering**
Master of Science in Statistical Data Analysis	5	A

Teaching languages

English

Keywords

Data visualisation, Data mining, Machine learning, Processing big data, Statistical learning.

Position of the course

This course offers a broad perspective on big data science and its role within academia and industry. It will focus in diverse aspects of data science, such as

- exploration: data visualisation and data mining;
- modelling: focused analysis of big data;
- computing: data capture, adaptation, storage and processing.

Contents

- The role of the data scientist.
- Epistemology of data science.
- Identifying data sources and biases.
- Data acquisition processes and data preparation (standardisation).
- Relations between bias, variance and model complexity.
- Predictive data mining techniques: penalisation methods, classification methods (e.g. support vector machines), bagging, boosting, random forests
- Cluster analysis, Principal Component Analysis, Multidimensional Scaling.
- Missing data
- Data translation and conversion tools.
- IT paradigms for parallelizing data analysis: MapReduce, in-memory approaches such as Spark.
- Data visualisation techniques: visual analytics.
- Processing unstructured text and graph data.
- Neural networks and deep learning

Initial competences

Basic data analysis skills and concepts (such as offered in the courses Principles of Statistical Data Analysis, and Statistical Modelling) and programming skills (such as offered in the courses Statistical Computing, and, Programming and Algorithms), preferably in Python.

Final competences

1 Have knowledge of methods and concepts for the processing of big data sets.

2 Query and process internal and external data sources that contain raw information, such as non-standardised data, unstructured text, ...

3 Visualise big datasets in an accessible manner that provides insight into the research question.
4 Use state-of-the-art data mining algorithms (e.g. classification, regression, dimensionality reduction...) to explore big datasets.
5 Understand how big data analyses may be subject to bias.
6 Express the uncertainty of big data analyses.
7 Collaborate with colleagues.
8 Adequately report the results from a big data analysis.

Conditions for credit contract

Access to this course unit via a credit contract is determined after successful competences assessment

Conditions for exam contract

This course unit cannot be taken via an exam contract

Teaching methods

Seminar, Lecture, Independent work

Study material

Type: Slides

Name: Big data science
Indicative price: Free or paid by faculty
Optional: no
Language : English
Available on Ufora : Yes
Online Available : Yes
Available in the Library : No
Available through Student Association : No
Additional information: The slides and any additional material for the course will be made available via Ufora. Most of the recommended books are freely available online. For the practical part of the course, it is recommended to bring your own laptop, but the heavy processing can be done on the UGent HPC infrastructure.

References

O'Neil. C., and Schutt. R. (2013) Doing Data Science: Straight Talk from the Frontline. O'Reilly.
Trochim. W. (2006) The Research Methods Knowledge Base. Cenage Learning
Hastie. T., Tibshirani. R., and Friedman J. (2009) The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer
Gareth. J., Witten. D., Hastie. T., and Tibshirani. R. (2013) An Introduction to Statistical Learning with Applications in R. Springer.

Course content-related study coaching

Assessment moments

end-of-term and continuous assessment

Examination methods in case of periodic assessment during the first examination period

Oral assessment, Assignment

Examination methods in case of periodic assessment during the second examination period

Oral assessment, Assignment

Examination methods in case of permanent assessment

Oral assessment, Assignment

Possibilities of retake in case of permanent assessment

examination during the second examination period is possible

Calculation of the examination mark

50% on group work and 50% on end-of-term exam.