

Big Data Science (C003802)

Course size (nominal values; actual values may depend on programme)
 Credits 5.0 Study time 150 h Contact hrs 62.5 h

Course offerings and teaching methods in academic year 2020-2021

A (semester 2)	English	guided self-study	15.0 h
		lecture	22.5 h
		self-reliant study activities	10.0 h
		seminar: practical PC room classes	15.0 h

Lecturers in academic year 2020-2021

Peralta, Daniel	WE02	lecturer-in-charge
Goetghebeur, Els	WE02	co-lecturer
Ley, Christophe	WE02	co-lecturer

Offered in the following programmes in 2020-2021

Master of Science in Statistical Data Analysis	crdts	offering
	5	A

Teaching languages

English

Keywords

Data visualisation, Data mining, Querying databases, 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).
- Data translation and conversion tools.
- IT paradigms for querying databases: MapReduce, NoSQL, in-memory approaches such as Spark.
- Processing unstructured text and graph data.
- Data visualisation techniques: visual analytics.
- Cluster analysis, Principal Component Analysis, Multidimensional Scaling.
- Relations between bias, variance and model complexity.
- Data mining techniques (penalisation methods, classification methods (e.g. support vector machines), bagging, boosting, random forests)

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).

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 databases in an accessible manner that provides insight into the research question.
- 4 Use state-of-the-art data mining tools to explore big data bases.
- 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

Guided self-study, group work, lecture, self-reliant study activities, seminar: practical PC room classes

Learning materials and price

Slides and scientific papers.

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. Cengage Learning

Course content-related study coaching

Evaluation methods

end-of-term evaluation and continuous assessment

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

Written examination

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

Written examination

Examination methods in case of permanent evaluation

Assignment

Possibilities of retake in case of permanent evaluation

examination during the second examination period is possible

Calculation of the examination mark

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