GHENT UNIVERSITY

Course Specifications

Valid as from the academic year 2024-2025

Data Mining (F000759)

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 1)	English	Gent	lecture
			independent work
			seminar

Lecturers in academic year 2024-2025

Benoit, Dries	EB23	lecturer-in-charge

Offered in the following programmes in 2024-2025	**crdts**	**offering**
Bachelor of Science in Business Engineering	5	A
Exchange programme in Economics and Business Administration	5	A
Preparatory Course Master of Science in Business Engineering	5	A

Teaching languages

English

Keywords

Univariate and multivariate regression analysis, classification, resampling methods, model selection and regularization, tree-based methods, unsupervised learning, R

Position of the course

Business processes have been digitalized at high pace in recent decades. This lead to a tremendous increase in information stored in databases. With datamining, analysts try to find relevant patterns in this huge source of information that help management in optimizing their decisions. Students will learn how to extract relevant information from databases. Students will learn how to apply a broad set of methods in a correct way to solve real-life business problems. They will also learn the skills of how to interpret the results from such analyses in a meaningful way. All exercises will be solved using the statistical programming language R. This course builds on the concepts introduced in the course "Statistics I".

Contents

* Relationship between bias and variance and model complexity
* Regression analysis
• simple linear regression
• multiple regression
• qualitative predictors
* Classification
• logistic regression analysis
• linear discriminant analysis
• k-nearest neighbors
* Resampling methods
• crossvalidation
• bootstrap
* Model selection and regularization
• variable selection
• shrinkage methods
• dimension reduction
* Unsupervised learning
• principal component analysis

- clustering

Initial competences

Students need to have a thorough understanding of the basic statistical principles and concepts as treated in the course "Statistics I".

Final competences

1 Apply datamining methods in a correct way, using the programming language R
2 Interpret the results in a correct way and communicate these to a non-technical audience
3 Choose a well-suited method to tackle a specific business problem, knowing what the advantages and disadvantages are of the chosen approach.

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

Extra information on the teaching methods

The learning management system Ufora will be used to coordinate the different diidactical approaches used in this course.
Active participation in lab sessions via dodona-platform is required.

Study material

Type: Handbook

Name: An Introduction to Statistical Learning: with Applications in R (2nd edition)
Indicative price: Free or paid by faculty
Optional: no
Language : English
ISBN : 978-1-07161-417-4
Online Available : Yes
Available in the Library : Yes
Available through Student Association : Yes
Usability and Lifetime within the Course Unit : intensive
Usability and Lifetime within the Study Programme : regularly
Additional information: Free (and legal) pdf can be downloaded from: https://www.statlearning.com/

References

James, G., Witten, D., Hastie, T. & Tibshirani, R. (2021) An Introduction to Statistical Learning with Applications in R, Springer (Version 2). (Note: a free, legal version of the book can be found online).

Course content-related study coaching

Students can rely on the teaching assistants and teacher of the course.
Communications and interaction through the Ufora platform (group discussions, exercises, ...). Example exam questions will be made available.

Assessment moments

end-of-term and continuous assessment

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

Written assessment

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

Written assessment

Examination methods in case of permanent assessment

Participation, Written assessment open-book

Possibilities of retake in case of permanent assessment

examination during the second examination period is not possible

Extra information on the examination methods

Periodic evaluation (written exam):
- Multiple choice questions

- Open questions

Non-periodic evaluation (assignment):

- Dodona exercises

Calculation of the examination mark

95% periodic evaluation (written exam)
5% non-periodic evalutation (assignment)
Consequences of not respecting deadlines: Not respecting deadlines of the individual work and / or group work without a well-motivated and substantiated reason (such as illness) leads to a zero for individual work and/or group work, respectively.