Course Specifications
Valid as from the academic year 2018-2019

Lecturers in academic year 2018-2019
Saeyes, Yvan WE02 lecturer-in-charge
Peralta, Daniel WE02 co-lecturer

Offered in the following programmes in 2018-2019 crdts offering
Master of Science in Computer Science 6 A

Course size (nominal values; actual values may depend on programme)
Credits 6.0 Study time 165 h Contact hrs 62.5 h

Course offerings and teaching methods in academic year 2018-2019
A (semester 2) English seminar: practical PC room classes 15.0 h
self-reliant study activities 10.0 h
lecture 22.5 h
guided self-study 15.0 h

Teaching languages
English

Keywords
Regression, classification, model building, dimension reduction, big data

Position of the course
Familiarize the students with the most important methods to extract information from large databases in a statistical way. The students are expected to learn how to use these techniques correctly in applications and they acquire the skills to interpret obtained results in a statistically correct manner.
The students will also be introduced to big data and the problems this might impose.
This course builds on the content of ‘Analysis of continuous data’ and ‘Categorical data analysis’ and assumes the student has acquired the skills taught in ‘Statistical Computing’.

Contents

**Distributed Databases**
- noSQL database systems
- Distributed DBMS
  - Distributed data storage
  - Distributed query processing
  - Distributed transaction model
  - Homogeneous and heterogeneous solutions
  - Client-server distributed databases
- Parallel DBMS
  - Parallel DBMS architectures: shared memory, shared disk, shared nothing;
  - Speedup and scale-up, e.g., use of the MapReduce processing model
  - Data replication and weak consistency models

**Data Mining**
- Uses of data mining
- Data mining algorithms
- Associative and sequential patterns
- Data clustering
- Market basket analysis
- Data cleaning
- Data visualization

(Approved) 1
Initial competences

Required prerequisites: Have a thorough understanding of linear regression as taught in the course ‘Analysis of continuous data’.

Final competences

1. Explain the techniques used for data fragmentation, replication, and allocation during the distributed database design process. [Familiarity]
2. Evaluate simple strategies for executing a distributed query to select the strategy that minimizes the amount of data transfer. [Assessment]
3. Explain how the two-phase commit protocol is used to deal with committing a transaction that accesses databases stored on multiple nodes. [Familiarity]
4. Describe distributed concurrency control based on the distinguished copy techniques and the voting method. [Familiarity]
5. Describe the three levels of software in the client-server model. [Familiarity]
6. Compare and contrast different uses of data mining as evidenced in both research and application. [Assessment]
7. Explain the value of finding associations in market basket data. [Familiarity]
8. Characterize the kinds of patterns that can be discovered by association rule mining. [Assessment]
9. Describe how to extend a relational system to find patterns using association rules. [Familiarity]
10. Evaluate different methodologies for effective application of data mining. [Assessment]
11. Identify and characterize sources of noise, redundancy, and outliers in presented data. [Assessment]
12. Identify mechanisms (on-line aggregation, anytime behavior, interactive visualization) to close the loop in the data mining process. [Familiarity]
13. Describe why the various close-the-loop processes improve the effectiveness of data mining. [Familiarity]

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, lecture, self-reliant study activities, seminar: practical PC room classes.

Extra information on the teaching methods

Minerva will be used to ensure a smooth organisation and follow-up of the practical assignments.

Learning materials and price

A syllabus is available. Estimated cost: 10 EUR.

References


Course content-related study coaching

The exercises and practical assignments are supervised by the lecturer.

Evaluation methods

end-of-term evaluation

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

Oral examination, assignment.

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

Oral examination, assignment.

Examination methods in case of permanent evaluation

Possibilities of retake in case of permanent evaluation

not applicable

Extra information on the examination methods

Project work shows the student’s capability to analyse real data problems by using the methods taught in this course and interpret correctly the obtained results.

Oral examination based on a written report about the results of the student’s project work.

(Approved)
Calculation of the examination mark
100% examination