Technology
Arts Sciences
TH Köln

Course
DDML - Data Mining

Version: 1 | Last Change: 12.02.2021 13:46 | Draft: 0 | Status: vom verantwortlichen Dozent freigegeben

General information

Long name	Data Mining
Approving CModule	DML_BaET
Responsible	Prof. Dr. Beate Rhein Professor Fakultät IME
Level	Bachelor
Semester in the year	winter semester
Duration	Semester
Hours in self-study	78
ECTS	5
Professors	Prof. Dr. Beate Rhein Professor Fakultät IME
Requirements	From Mathematics 1 and 2 the ability to construct mathematical models as well as knowledge of differential calculus and linear algebra is required.
Language	German
Separate final exam	Yes

Final exam

Details

Depending on the number of participants:
For a small number of participants: combination of exam or oral examination and evaluation of the mini-project.
For many participants, examination by written examination; mini-project as prerequisite for participation in the examination.

In the written or oral examination, the methods, procedures, pitfalls and legal foundations of data mining are examined.

In the mini-project the ability to act independently and on one's own responsibility and the use of suitable software will be tested.

Minimum standard

Basic knowledge of the general approach to data mining, the procedures covered and their limitations.

Exam Type

Depending on the number of participants:
For a small number of participants: combination of exam or oral examination and evaluation of the mini-project.
For many participants, examination by written examination; mini-project as prerequisite for participation in the examination.

In the written or oral examination, the methods, procedures, pitfalls and legal foundations of data mining are examined.

In the mini-project the ability to act independently and on one's own responsibility and the use of suitable software will be tested.

Lecture / Exercises

Learning goals

Knowledge

Introduction to a suitable software, e.g. Python
Introduction to descriptive statistics and possibly also probability calculation
Supervised learning:
- Classification procedure: Procedure, performance measures, application of a method of instance-based learning, e.g. k-nearest-neighbor and a method of model-based learning, e.g. decision trees
- Possibly regression analysis: about machine learning and classical
Unsupervised learning:
- Cluster analysis: k-means, possibly also DBSCAN
Preprocessing of the data:
- Handling Damaged / Missing Data
- Runaway or noise - problems
- Scaling
- Visualization of data

- Possible dimension reduction
- Assessment of data quality
- possibly look at different types of data records, make reference to NoSql databases, Outlook on current research, e.g. image recognition, Natural Language Processing, Reinforcement Learning

Skills

Be able to name and apply a suitable method and overall approach to tasks
Select and evaluate a suitable performance measure
Apply Privacy Policy

Expenditure classroom teaching

Type	Attendance (h/Wk.)
Lecture	2
Exercises (whole course)	0
Exercises (shared course)	2
Tutorial (voluntary)	0

Separate exam

none