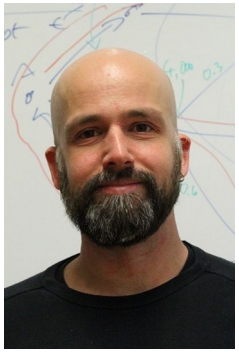
UNIVERSITÄT MANNHEIM

Data Mining II
Organization

Heiko Paulheim, Nicolas Heist

Hello



- Heiko Paulheim
- Professor for Data Science
- Research Interests:
  - Semantic Web and Linked Open Data
  - Data Mining with Linked Open Data
  - Ontology Matching
  - Data Quality and Data Cleaning
- Consultation: Tuesdays, 9-10am
  - Please make an appointment via e-mail to Ms. Lermer
- Heiko will teach the lectures

Hello

- M.Sc. Nicolas Heist
- Graduate Research Associate
- Research Interests:
  - Semantic Web Technologies
  - Knowledge Graphs and Linked Data
- eMail: nico@informatik.uni-mannheim.de
- Nico will teach the exercises and co-supervise the projects

Course Organization

- Lecture
  - addresses advanced data mining topics
  - builds on Data Mining I lecture contents!
- Project Work
  - we will take part in the Data Mining Cup 2020
  - with eight teams
    - the two best performing teams submit their solutions
  - regular presentations of your approaches
  - paper and final presentation
- Exercise
  - weekly with warm up on DMC tasks from previous years

Requirements

- Final exam
  - 100 % written exam
  - project is not graded, but mandatory!
- Project work
  - work on DMC tasks
- Presentations
  - up to three intermediate presentations
    - open questions, problems, current results (numbers!)
  - everybody has to present once during those presentations
- Final report
  - 10 pages
  - solutions, results, lessons learned

The Data Mining Cup

- An annual competition
  - for students
  - run since 2002
  - participation from all over the world
  - max. two teams per institution (i.e., university)
  - 2019: 149 participating teams from 28 countries
- Timeline
  - DMC registration today (!)
  - tasks are published on March 19th
  - submissions are due on April 23rd (internal submission: April 22nd)
- Further information: http://www.data-mining-cup.de/

The Data Mining Cup

- 2017: both Uni Mannheim teams among top 10 (out of 202)
- 2018: team from Uni Mannheim scores 2nd place (out of 197)
- 2019: team from Uni Mannheim scores 10th place (out of 149)
- Prices are awarded at a conference in Berlin in June
  - Top 10 teams are invited to present their solutions

Schedule

- 18.2. Introduction & Data Preprocessing
- 25.2. Ensembles
- 3.3. Time Series
- 10.3. Neural Networks & Deep Learning
- 17.3. Hyperparameter Tuning
- 24.3. DMC Session 1
- 31.3. DMC Session 2
- 7.4. *Easter Break*
- 14.4. *Easter Break*
- 21.4. DMC Session 3
- 28.4. Anomaly Detection
- 5.5. Model Verification

DMC task published on 19.3.

final DMC submission 23.4.

Deadlines at a Glance

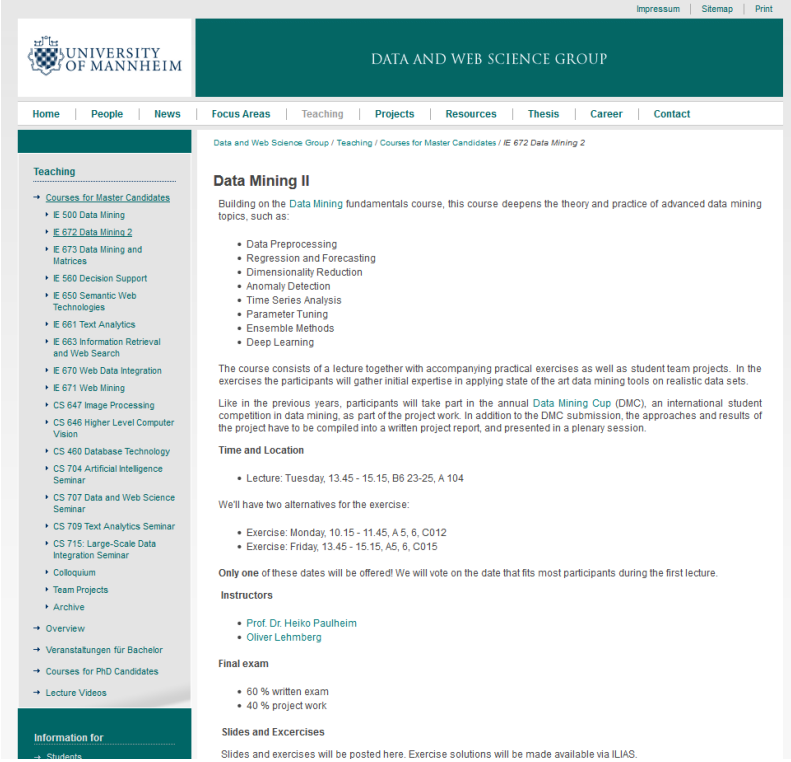
- today: DMC registration
- March 19th: you know the DMC tasks and your team
- April 21st: submission of your DMC solution to Nico and Heiko
- April 23rd: official submission of your DMC solution
- May 24th: submission of your final report

Lecture Contents

- Data Preprocessing (today!)
- Ensemble Learning
- Time Series Analysis
- Neural Networks and Deep Learning
- Parameter Tuning
- Anomaly Detection
- Model Evaluation, Verification, and Comparison

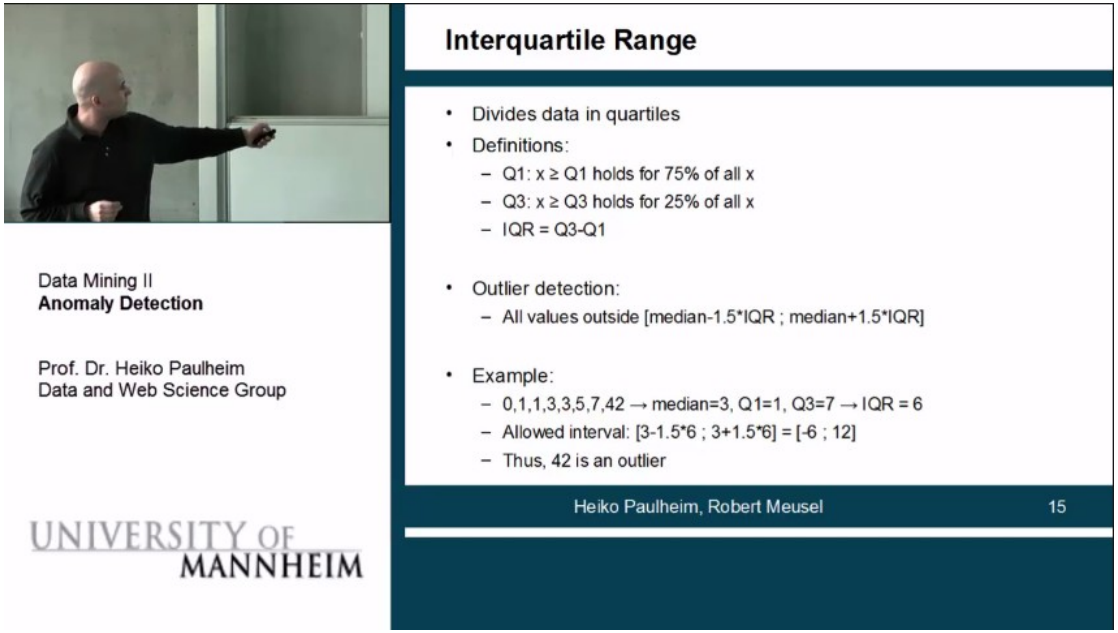
Course Organization

- Lecture Webpage: Slides, Announcements
  - http://dws.informatik.uni-mannheim.de/en/teaching/courses-for-master-candidates/ie-672-data-mining-2/
  - hint: look at version tags!
- Additional Material
  - ILIAS eLearning System, https://ilias.uni-mannheim.de/



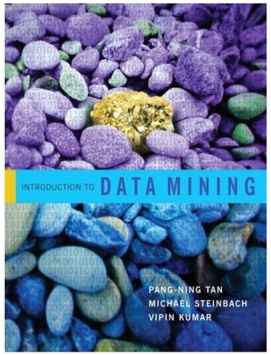
Video Recordings of Last Year's Lecture

- http://dws.informatik.uni-mannheim.de/en/teaching/lecture-videos/
  - Accessible from within university network and VPN

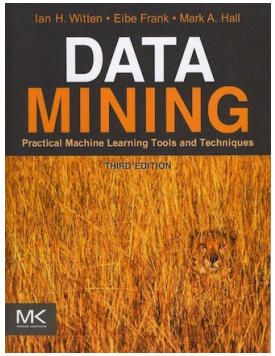


Literature & Slide Sources

- Pang-Ning Tan, Michael Steinbach, Vipin Kumar:
  Introduction to Data Mining,
  Pearson / Addison Wesley.
  - 10 copies in university library.
  - we provide scans of important chapters via ILIAS



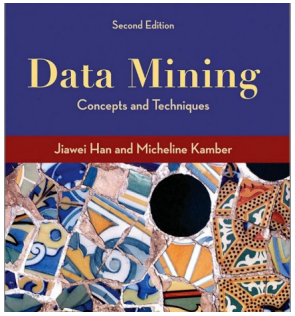
- Ian H. Witten, Eibe Frank, Mark A. Hall:
  Data Mining: Practical Machine Learning
  Tools and Techniques, 3rd Edition, Morgan Kaufmann.
  - several copies in university library
  - we provide scans of important chapters via ILIAS



Literature & Slide Sources

- Gregory Piatetsky-Shapiro, Gary Parker:
  KDNuggets Data Mining course:
  http://www.kdnuggets.com/data_mining_course/

- Jiawei Han and Micheline Kamber:
  Data Mining – Concepts and Techniques
  – free e-book access via university library



Questions?