Data Mining II
Organization

Heiko Paulheim, Nicolas Heist
Hello

• Heiko Paulheim
• Professor for Data Science
• Research Interests:
  – Knowledge Graphs on the Web and their Applications
  – Data Quality and Data Cleaning on Knowledge Graphs
  – Using Knowledge Graphs in Data Mining
  – Societal Impact of Artificial Intelligence
• 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 2023
  – you will work in teams
    • the two best performing teams submit their solutions
  – regular presentations of your approaches
  – paper and final presentation

• Exercise
  – weekly, includes 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)
  – 2022: 78 participating teams from 23 countries

• Timeline: tba! Last year’s dates:
  – DMC registration starts on March 1\textsuperscript{st}
  – tasks are published on April 12th
  – submissions are due on June 28\textsuperscript{th} (internal submission: June 21\textsuperscript{st})

• 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)
• 2020: team from Uni Mannheim scores 8th place (out of 162)
• 2022: team from Uni Mannheim scores 8th place (out of 78)
• Prices are awarded in a ceremony in July
Schedule

- 21.2. Introduction & Data Preprocessing
- 28.2. Ensembles
- 7.3. Time Series
- 14.3. Neural Networks & Deep Learning
- 22.3. Anomaly Detection
- 28.3. Hyperparameter Tuning
- Easter Break
- 18.4. Model Verification and Interpretation
- 25.4. DMC session
- 2.5. DMC session
- 9.5. DMC session
- 16.5. DMC session
- 23.5. DMC session
Deadlines

• Some deadlines are still subject to Data Mining Cup announcements
• Final report and first DMC solutions due:
  – June 2\textsuperscript{nd}, 2023
• Final DMC solutions due:
  – One week before DMC deadline
Course Organization

• Lecture Webpage: Slides, Announcements
  – hint: look at version tags!

• Additional Material
Video Recordings of 2021

- We will provide lecture videos in ILIAS
- In case you prefer not to or cannot attend the lecture, you are advised to watch those lectures

Data Mining II
Data Preprocessing

Heiko Paulheim
Exercises

• We offer two exercise groups
  – You only need to attend one
  – Room: A5, 6, C 012

• Monday, 12:00 – 13:30
• Monday, 13:45 – 15:15
Literature & Slide Sources

• Pang-Ning Tan, Michael Steinbach, Vipin Kumar: Introduction to Data Mining, Pearson / Addison Wesley.
  – 10 copies in university library

  – several copies in university library
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?