UNIVERSITÄT MANNHEIM

Data Mining II
Data Mining Cup – Let's Go!

Heiko Paulheim, André Melo

Requirements

- Final grade
  - 60 % written exam
  - 40 % project work
- Project work
  - work on DMC tasks
  - we meet every Tuesday to discuss the current progress
- Presentations
  - Three intermediate presentations
    - open questions, problems, current results (numbers in 10-fold CV)
  - one final presentation
  - everybody has to present once during those four presentations
- Final report
  - 10 pages per team
  - solutions, results, lessons learned

DMC Timeline

- Today: First look at the task, organization
- Do's and dont's
  - Brought to you by a DMC veteran
- Presentations:
- 17.04.18 DMC intermediate presentation
- 24.04.18 DMC intermediate presentation
  - 01.05.18 Holiday
- 08.05.18 DMC intermediate presentation
- 15.05.18 DMC intermediate presentation
- 17.05.18 DMC results submission
- 22./29.05.18 DMC final presentation

DMC Presentation Schedule

Intermediate Presentations (15’ slots)

- 17.04.18 Team 1-6
- 24.04.18 Team 3-8
- 08.05.18 Team 1+2,5-8
- 15.05.18 Team 1-4,7+8

Final Presentations

- 22.05.18 Team 1-4
- 29.05.18 Team 5-8

Project Grading

- Projects will be graded based on
  - Innovation of ideas created and pursued
  - Intermediate and final presentations
  - Quality of the final report

- We will have eight teams, but joint meetings
  - You are allowed to use ideas from the other teams
    - but you have to mark them in the final report
  - And you send us your slides of each intermediate presentation
    - so that we can track the origin of ideas

Individual Grading

- In each team, there may be smaller sub teams working on different tasks
  - Each slide must have a tag with the contributors' names

- Peer grading
  - At the end of the project, you will give grades to your team mates
  - Your grades will be kept secretly
  - We only use them to confirm (and, if necessary, adjust) our assessment

Let's Get Started with the Task

- You have looked at the data
- ...and read the task

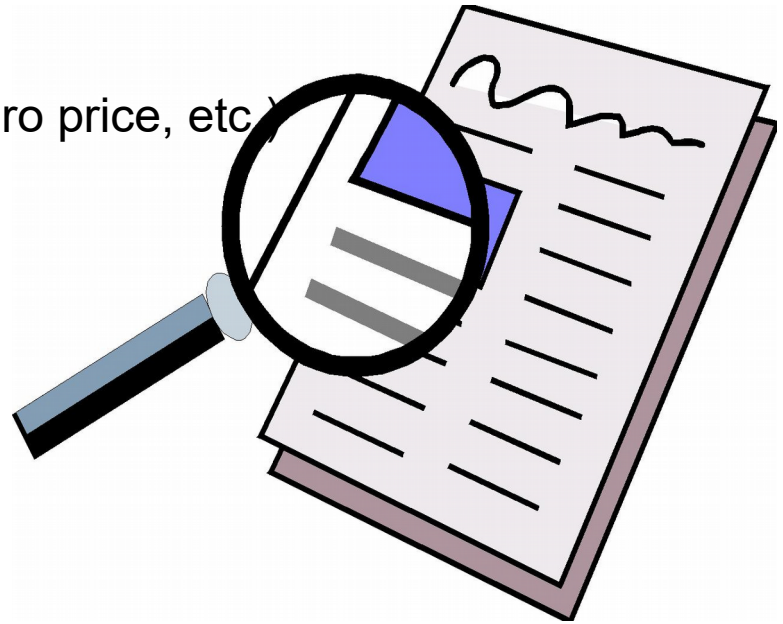
- Question: What does the data look like?
- Question: How do we evaluate the results?

Evaluation

- DMC comes with a fixed metric
- Always use this metric!
  - We will use Oct-Dec for training, Jan for testing
  - You must not use the January file for training, except for the prices!
- i.e., in each intermediate presentation,
  you'll report the numbers for January
  - We will provide some (fictional) test sets for January
  - You report
    - Macro average score
    - Standard deviation

What does the Data Look Like?

- Detailed questions include, but are not limited to
  - Are there unseen products/editions in the test set?
  - How is the distribution of products/editions/orders?
  - How many distinct products etc. are there?
  - How many categories are there? What is their distribution?
  - How strongly do the prices vary?
  - Can we observe any anomalies (e.g., zero price, etc.)
- Plus: what is the performance of a default model?



Now You Know what to Do!