

Knowledge Discovery in Databases

With a view towards machine learning

Luciano Melodia M.A.

Evolutionary Data Management, Friedrich-Alexander University Erlangen-Nürnberg

Summer semester 2021



Who and for Whom?

- Luciano Melodia M.A.
 - Evolutionary Data Management.
 - `luciano.melodia@fau.de`.
- Master course in Computer Science:
 - For partial fulfillment of a module with 5 ECTS.
 - Examination: Oral exam together with another lecture, such as:
 - DBRN (Datenbanken in Rechnernetzen).
 - DSS (Datenstromsysteme, German only).
 - DW (Data Warehousing).
 - TAS (Transaktionssysteme, also in Englisch).
 - Other programs:
 - M.Sc. International Information Systems.
 - M.Sc. Computational Engineering.
 - M.Sc. Mechanical Engineering.
 - M.Sc. Information- and Communication Technology.
 - ...

Prerequisites

- IDB (Implementierung von Datenbanksystemen, every WS).
- IDS (Introduction to Database Systems)
 - With parts on implementation:
 - Index structures.
 - Query processing.
 - ...

Lecture

- Time and place (during SARS-CoV2 pandemic via Zoom):
 - Tuesday, 8:15 - 9:45 c.t.
 - Room: K2-119.
 - Zoom-room: <https://fau.zoom.us/j/2164886110>.
- Lecture notes will be available as PDF via StudOn.
- Additional information is also available in StudOn.
- What about **feedback, questions, remarks** or **complaints**?
 - Always welcome.
 - Drop me a line: luciano.melodia@fau.de.
 - Consultation hours: Thursday, 5-6 p.m.
 - Room: 08.132.
 - Course will be evaluated electronically.

Lecture

- The lecture is based on the following book:
 - Jiawei Han, Micheline Kamber, and Jian Pei: *Data Mining – Concepts and Technologies*, 3rd ed. Waltham, M.A.: Morgan Kaufmann, 2012 (The Morgan Kaufmann Series in Data Management Systems).
 - Copies are available at the TNZB.
 - Slides have been downloaded from Jiawei Han's web server and modified.
 - Originally these slides were created by Prof. Dr.-Ing. Klaus Meyer-Wegener.
 - We thank Leonie Rumi and Noah Becker for proofreading and valuable comments on the structure of the lecture.
- Also interesting and related textbooks are:
 - Hongbo Du: *Data Mining Techniques and Applications*, 1st ed. Cengage Learning, 2010.
 - Ian Witten, Eibe Frank, and Mark Hall: *Data Mining: Practical Machine Learning Tools and Techniques*, 3rd ed. Burlington, M.A.: Morgan Kaufmann, 2011 (The Morgan Kaufmann Series in Data Management Systems).
 - ...

Our Goal

- Introduce you to the principles of data mining.
This is the core of knowledge discovery in databases.
- We'll make you familiar with:
 - Some KDD pipelines.
 - The challenge of data mining on large sets of data.
 - Technologies available for data analysis.
 - Systems offering these technologies.
 - Some applications . . .

Companion Lecture: Deep Learning

- First time hold in winter term 2017/2018, now offered again at:
 - Thursday, 8-9:45 a.m. c.t.
 - Room: H4.
 - Pattern Recognition Lab.
 - Lecture is in Englisch.
- Topics covered:
 - Classification using Neural Networks, probably in much more detail.
 - No databases, no indexes, no clustering . . .
- This lecture gives a broader overview, thus is less detailed.

Modules

- A module is the **unit of examination** at FAU and must have at least 5 ECTS.
- Thus, KDD alone is a too small lecture.
 - Just 2.5 ECTS, which corresponds to
75 hours of work,
30 for attendance and
45 hours for own studies and exam preparation.
- **Combine KDD with another lecture** of the same size, such as:
 - DBRN (Datenbanken in Rechnernetzen).
 - DSS (Datenstromsysteme).
 - DW (Data Warehousing).
 - TAS or TRASYS (Transaktionssysteme).
- Examination: a single oral exam for **both lectures** together.
30 minutes with approximately 15 minutes per lecture.

Exceptions for modules

- Combinations with DW (Data Warehousing):
 - DW has its own written exam.
 - Thus, you'll need a supplementary oral exam for KDD alone.
This exam will be also a 30 minutes exam.
- **Medical Engineering** Master program:
 - Modules of 2.5 ECTS are still allowed.
 - KDD can be chosen in
Studienrichtung Medizinische Bild- und Datenverarbeitung
M5 Medizintechnische Vertiefungsmodule (BDV).
Examination number 392229.
 - Examination: an oral exam for 30 minutes.

Exam

- **Registration** for examination:
 - At the examination office (dt. Prüfungsamt) or rather in MeinCampus.
 - Period (probably): May 15 – Mai 30, 2021 (three weeks).
 - Please denote your examination number!
- **Appointment** for oral exams:
 - In StudOn, or visit our homepage: Lehrstuhl → Lehre → Prüfungsinformationen.
 - Or visit <https://www.cs6.tf.fau.de/lehre/pruefungsinformationen/>.

Thank you for your attention.
Any questions about orga?

Ask them now, or again, drop me a line:
✉ luciano.melodia@fau.de.