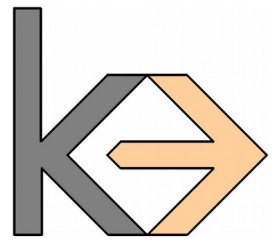




TECHNISCHE  
UNIVERSITÄT  
DARMSTADT



# Data Mining und maschinelles Lernen

Johannes Fürnkranz

`juffi@ke.tu-darmstadt.de`

# Inhalt

---

- Einführung in maschinelles Lernen und Data Mining
    - mit Schwerpunkt auf symbolisch/logischen Ansätzen
    - mathematisch/statistisch/symbolische Ansätze werden in Vorlesungen der Kollegen Jan Peters, Stefan Roth, u.a. behandelt
- Kombinationen sind durchaus sinnvoll

# Organisatorisches

- Termine
  - Vorlesung
    - **Mi** 18.05-19.45, S202|C205
  - Übung
    - **Di** 9.50-11.30, S202|C205
    - erste Übung: **3. 11.**
- Benotung:
  - Klausur am Ende des Semesters
  - Teilnahme am praktischen Projekt kann eine Notenverbesserung um 1 Notenstufe bringen
    - Klausur muß aber bestanden werden!

# Übungen

---

- Übungen
  - Lösung von Aufgaben zum besseren Verständnis und zur Prüfungsvorbereitung
  - Lösungen werden von Assistenten in den Übungsterminen diskutiert
- Zuständiger Assistent
  - Eneldo Loza (eneldo@ke.tu-darmstadt.de)

# Übungen

- Veröffentlichung der Übungsausgabe ist üblicherweise Dienstags auf der Homepage
- Besprechung der Übung jeweils eine Woche später
  - es wird von den Studierenden erwartet, daß die Übungsaufgaben **vor der Besprechung** selbständig bearbeitet werden
  - Besprechung sollte vor allem dazu dienen, offene Punkte anzusprechen
- Übungsaufgaben beziehen sich meist auf die Vorlesung des vorherigen Mittwochs

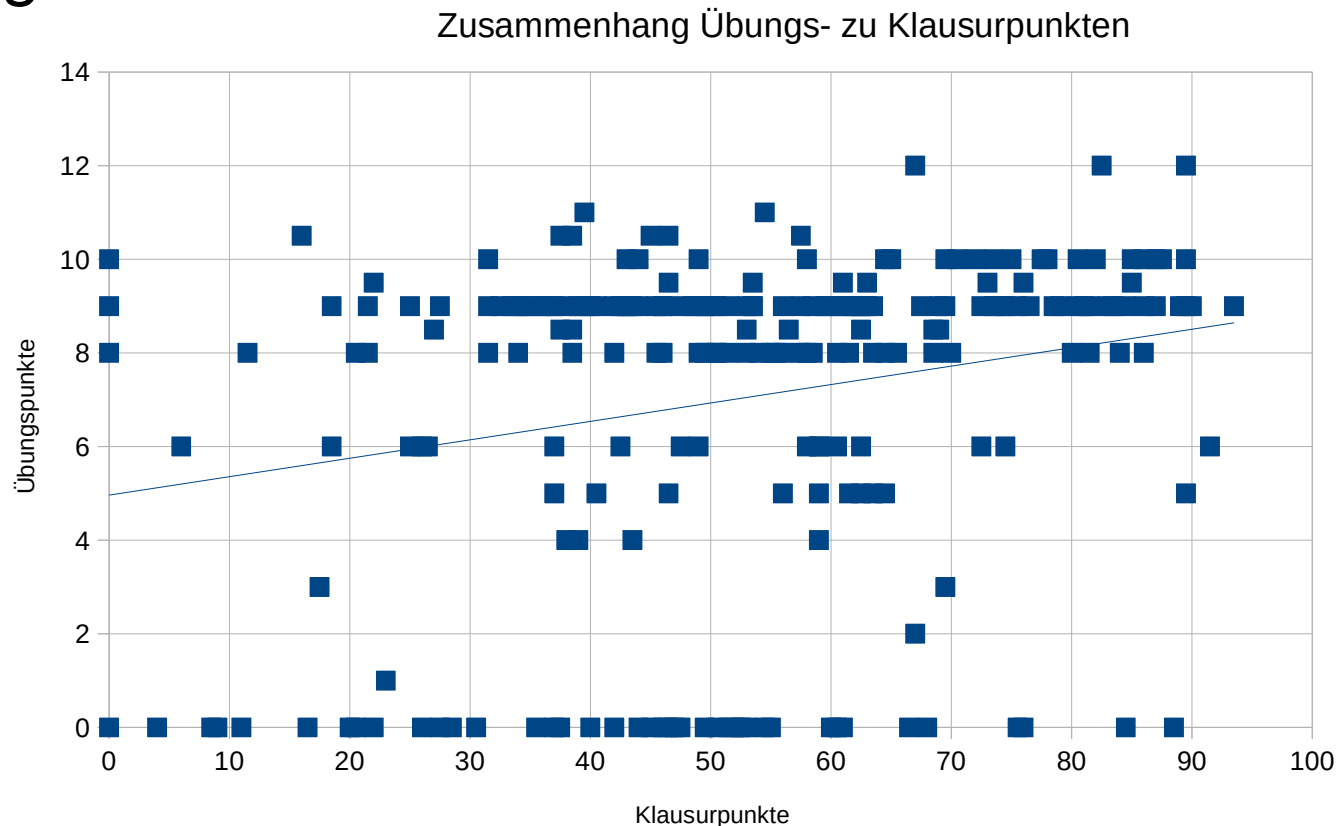
# Projekt

- Im Laufe des Semester erhalten Sie ein praktisches Projekt
  - das bis zum Semesterende gelöst werden sollen
  - Lösung in Gruppen (bis zu 3 Teilnehmern) ist erlaubt
- Thematik
  - Praktische Erprobung der vorgestellten Methoden mit Hilfe der Weka-Data Mining Library (open source, Java)
  - Programmierung eines einfachen Algorithmus
- Benotung
  - Erfolgreiche Teilnahme kann eine Verbesserung um bis zu einer Notenstufe führen
  - Nur bei bestandener Klausur

# Übung und Projekt

Ziel: Lösung von Aufgaben zum besseren Verständnis und zur Prüfungsvorbereitung

→ Zusammenhang zwischen Teilnahme an Übung und Erfolg in Klausur

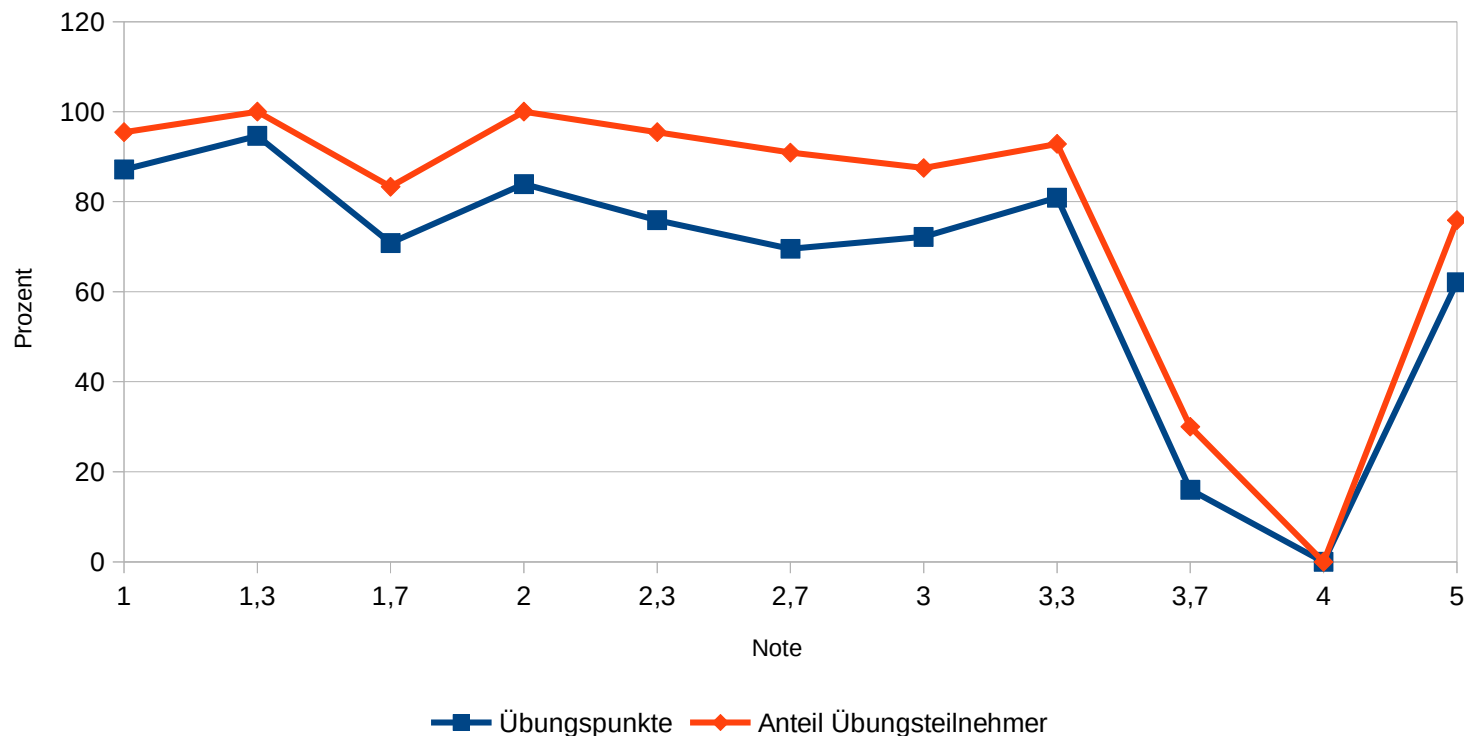


# Übung und Projekt

Ziel: Lösung von Aufgaben zum besseren Verständnis und zur Prüfungsvorbereitung

→ Zusammenhang zwischen Teilnahme an Übung und Erfolg in Klausur

Zusammenhang Note-Teilnahme/Übungspunkte





# Übung und Projekt

Ziel: Lösung von Aufgaben zum besseren Verständnis und zur Prüfungsvorbereitung

→ Zusammenhang zwischen Teilnahme an Übung und Erfolg in Klausur

Notendurchschnitt und Unterschiede			
	WS12/13	WS13/14	WS14/15
Endnote mit Übungsteilnahme (inkl. 5,0er)	2,81	2,88	3,10
Endnote mit Übungsteilnahme ohne Bonus	+0,48	+0,34	+0,35
Endnote ohne Übungsteilnahme	+1,00	+0,82	+1,09

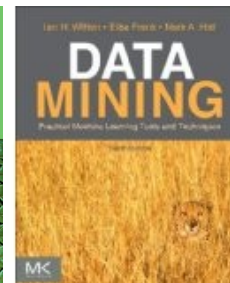
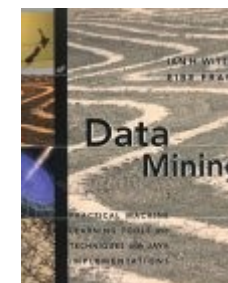
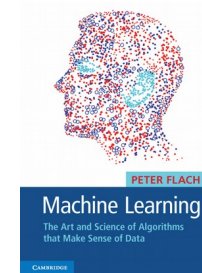
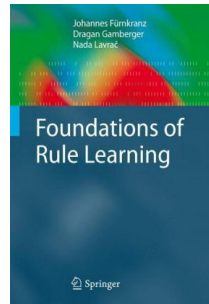
# Materialien

- Folien
  - Vorlesungssprache ist Deutsch,
  - but slides will be in English
- Vorlesungs-Homepage
  - <http://www.ke.informatik.tu-darmstadt.de/lehre/ws-15-16/mldm/>
- Diskussions-Forum
  - <http://www.fachschaft.informatik.tu-darmstadt.de/forum/viewforum.php?f=225>
  - Noch unter dem alten Vorlesungsnamen “Maschinelles Lernen: Symbolische Ansätze”.

# Bücher

Die folgenden Bücher können als gute Ergänzung für einige der in der Vorlesung gebrachten Kapitel dienen.

- Johannes Fürnkranz, Dragan Gamberger, Nada Lavrac: *Fundamentals of Rule Learning*, Springer-Verlag 2012
- Tom Mitchell: *Machine Learning*, McGraw-Hill, 1997
- Peter Flach: *Machine Learning*, Cambridge University Press 2012
- Ian Witten & Eibe Frank: *Data Mining*, Morgan Kaufmann, 1999  
Deutsche Ausgabe: Hansa, 2001  
2<sup>nd</sup> edition, 2005, 3<sup>rd</sup> edition 2011



# Software

- WEKA: Open Source Java Library  
<http://www.cs.waikato.ac.nz/ml/weka/>
- Extensive Documentation available there



# Weiterführende Veranstaltungen

---

- PR Praktikum Maschinelles Lernen und Data Mining
  - Üblicherweise im SS
  - Praktische Anwendung des Gelernten
  - Teilnahme an einem Wettbewerb
- Seminare
  - Finden jedes Semester statt
  - Diskussion aktueller Forschungsarbeiten auf dem Gebiet
- Forschungspraktika
  - Kann jederzeit begonnen werden
  - Individuelle Aufgabenstellungen, auch in Gruppen möglich, ähnlich zu BA oder MA