

# Course Organisation and Project Presentation

## Knowledge Discovery and Data Mining 2 (VU) (706.715)

Roman Kern

Institute for Interactive Systems and Data Science, TU Graz

2019-03-14

# Overall Goal

Bring the theoretical knowledge acquired in KDDM1 into practical application,  
... or, what it is like to be a data scientist?

Name: Roman Kern

Office: Know-Center,

Inffeldgasse 13, 5th Floor, Room 012

Office hours: By appointment

Phone: +43-316/873-30860

E-Mail: rkern@tugraz.at

Homepage: <http://kti.tugraz.at/staff/rkern/about/>

- Lectures in English
- Communication in German/English
- If in German: please informally (Du)!
- Student presentations in English

# Outline

- 1 Motivation
- 2 Course Organization
- 3 Projects

# Motivation

Why should one be interested in KDDM2?

- Job as data scientist
- “Data Scientist: The Sexiest Job of the 21st Century”
- <http://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century/>

“Start with the fact that there are no university programs offering degrees in data science”

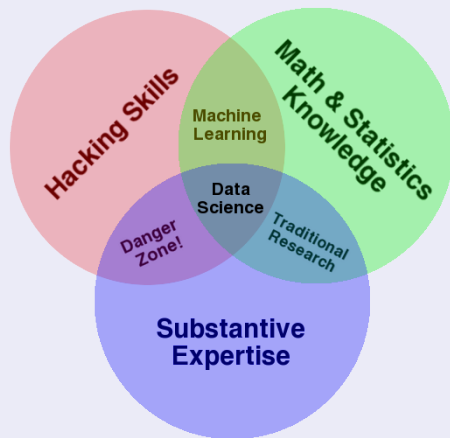
## What is a data scientist?

Data scientists are inquisitive: exploring, asking questions, doing “what if” analysis, questioning existing assumptions and processes. Armed with data and analytical results, a top-tier data scientist will then communicate informed conclusions and recommendations across an organization’s leadership structure.

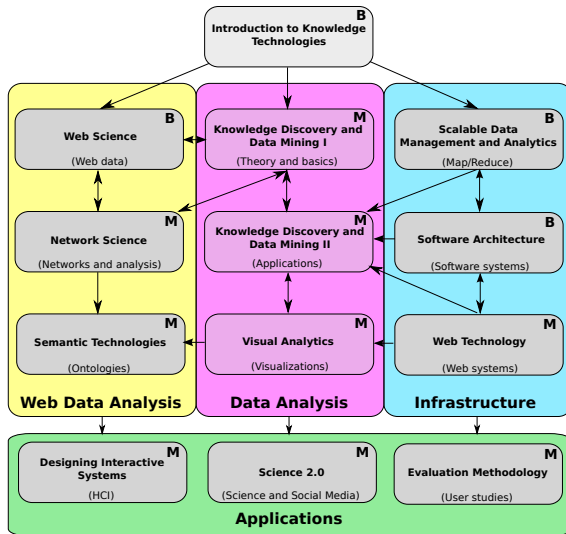
<http://www-01.ibm.com/software/data/infosphere/data-scientist/>



## Why KDDM2?



- Play with cool technologies
- ... in an hands-on approach
- Discussion & feedback
- Reports from the field



+ Projects, Bachelor Thesis, Master Projects, Master Thesis, PhD Thesis

# Course Organisation

## When & What

- Please register until the 07.03.2019
- The course will take place
- ... Thursday, 12:15 - 13:45
- ... in HS i8
- Note: Please have an eye on the calendar

- Course website: <http://kti.tugraz.at/staff/rkern/courses/kddm2>
- Posters/reports/presentations will be made available on the course website
- Description of the practical projects and access to data sets

- There is no written exam
- Therefore grading is based on the practical projects:
- ... soundness of the approach
- ... the outcome of the projects
- ... the conducted evaluation
- ... the presentation of the results

# Projects

Practical part of the course



# Overview Projects

- There are a number of practical projects
- ... from various stages of the KDD process
- Group of single students
- ... or groups of two people
  - ▶ ⇒ with bigger scope
- The focus is more on the approach, rather than the final results
  - ▶ ... but the results should be assessed (evaluated)

A project plan is available on the course homepage

## Work Plan: Group Registration

- Important: Please report groups/project
- ... by sending an e-mail to `rkern@tugraz.at`
- Deadline: **See homepage**
- Please add a [KDDM2] to the mail subject
- Students without project assignment will be unregistered

## Work Plan: Poster/Report Submission

- Submit your poster/report/presentation
- ... by sending an e-mail to `rkern@tugraz.at`
- Deadline: **See homepage**
- ... the poster will be printed (and paid) for you

If you do not manage this deadline, you are free to print the poster yourself and bring it to the poster session (and send it).

## Work Plan: Code Submission

- After the poster presentation submit your (zipped) source code/etc.
- ... by sending an e-mail to `rkern@tugraz.at`
- Deadline: **See homepage**

- There three options to present your results
- **Poster presentation:** prepare a A1 poster (portrait) and present at poster session
- **Written report:** write a 4-6 page report on project and submit via E-Mail
- **Oral presentation:** give an oral presentation, submit the slides (need to register a week before)

Templates available on the course homepage

## Guidelines on presentations

- What is the problem?
- Why did you choose your approach?
- How does the approach look like?
- How have you tackled the problem?
- How does the data look like (what are its properties)?
- What are your evaluation results (is the problem solved)?
- What have you learnt (new insights)?
- Did something unexpected happen?
- Would the solution apply to other scenarios (and how well)?

# Practical Aspects

- The language of the presentation is English (poster, report)
- Free to choose any programming language
- Free (to an extent) in the choice of data set
- The code is yours (free to share it via an open-source license)

# The End

Next: Ensemble Methods