

# KDDM2: Course Organisation

Roman Kern  
2020-10-01

Roman Kern, ISDS, TU Graz  
2020-10-01

KDDM2: Course Organisation  
Overall Goal

www.tugraz.at

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

> Welcome to the course KDDM2!  
> In the slide notes you will find some additional information and further resources.

> Main goal for KDDM2 is then to conduct a project, where the theoretical foundation is required and new insights and skills are acquired “on-the-go.”

Roman Kern, ISDS, TU Graz  
2020-10-01

KDDM2: Course Organisation  
Lecturer

www.tugraz.at

Name: Roman Kern  
Office: Inffeldgasse 13, 5th Floor, Room 012  
Office hours: By appointment  
Phone: +43-316/873-30860  
E-Mail: rkern@tugraz.at  
Homepage: <https://courses.isds.tugraz.at/rkern/about/>

Roman Kern, ISDS, TU Graz  
2020-10-01

KDDM2: Course Organisation  
Language

www.tugraz.at

- Lectures in **English**
- Available on the course web page
- Communication in **German/English**
  - Answers of the homework can be in either language
- Student presentations in **English**

Roman Kern, ISDS, TU Graz  
2020-10-01

- 1 Motivation
- 2 Course Organization
- 3 Projects

5 Roman Kern, ISDS, TU Graz  
2020-10-01

www.tugraz.at

## Motivation

Why should one be interested in KDDM2?

6 Roman Kern, ISDS, TU Graz  
2020-10-01

www.tugraz.at

Motivation  
Demand

- 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/>

> In fact, there are a couple of **former participants** of KDDM2, who are now working as data scientist in the industry.  
> Ranging from media, to energy, to retail.

7 Roman Kern, ISDS, TU Graz  
2020-10-01

www.tugraz.at

Motivation  
Definition

### 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.

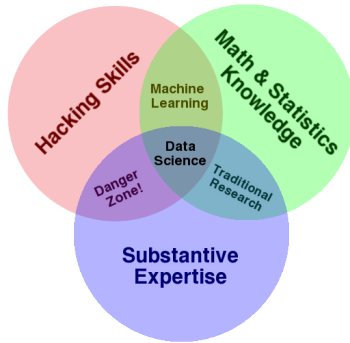
<http://www-01.ibm.com/software/data/infosphere/data-scientist/>

> The role of the data scientist connects the technique expertise with the **domain expertise**.  
> Without knowledge of the domain, data science cannot be successful.  
> On top of it, a data scientist needs to be **able to communicate**.  
> First, the importance of the work, and then the results to **bring the insights into action**.

8 Roman Kern, ISDS, TU Graz  
2020-10-01

Motivation  
Motivation

www.tugraz.at



- > Do not underestimate the important of domain expertise (substantive expertise) of **subject-matter experts**.
- > The goal of KDDM2 is to bring participants out of the **danger zone** (b/w hacking skills and domain expertise).
- > With hacking skills one refers to be able to code.
- > For example, one might be able to apply an of-the-shelf machine learning library on some data...
- > ... and there will be results, but will these results be any good ...
- > i.e., are the **requirements of the algorithms** being met?

9 Roman Kern, ISDS, TU Graz  
2020-10-01

Motivation  
Technologies

www.tugraz.at

Practical aspects of KDDM2:

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

10 Roman Kern, ISDS, TU Graz  
2020-10-01

www.tugraz.at

## Course Organisation

When & What

11 Roman Kern, ISDS, TU Graz  
2020-10-01

Course Organization  
Course Calendar

www.tugraz.at

- Please register until the 02.10.2020
- The course will take place
- ... online, i.e., slides and videos will be posted on the web page
- ... additionally, there is an online Q&A session
  - Thursday, 14:00 - 15:00
  - <https://tugraz.webex.com/meet/rkern>
  - ... feel free to join and discuss!

- > Please make use of the Q&A session if there are open questions:
- > ... selection of project
- > ... problems during the project execution
- > ... expectations on the outcome/presentation/etc.

12 Roman Kern, ISDS, TU Graz  
2020-10-01

> If you **do not want** to have your work being presented at the web page, please inform the lecturer.

- Course website:  
<https://courses.isds.tugraz.at/rkern/courses/kddm2/>
- Presentations (of projects) will be made available on the course website
- Suggestions of the practical projects and access to data sets

13

Roman Kern, ISDS, TU Graz  
2020-10-01

> The details of the grading, the deadlines and the distribution of point can be found on the **course website**.

- There is a **homework** to be done
  - Which will be made available in the TeachCenter
- Grading is also based on the **practical projects** and **presentation**
  - ... soundness of the approach
  - ... the outcome of the projects
  - ... the conducted evaluation
  - ... the presentation of the results

14

Roman Kern, ISDS, TU Graz  
2020-10-01

## Projects

Practical part of the course

15

Roman Kern, ISDS, TU Graz  
2020-10-01

> A **project plan** is available on the course homepage.

- 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 need to be assessed (evaluated)

16

Roman Kern, ISDS, TU Graz  
2020-10-01

### 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

17

Roman Kern, ISDS, TU Graz  
2020-10-01

### Work Plan: Submission

- Submit your video presentation & source code
- ... upload to TUG Cloud
- Deadline: **See homepage**

18

Roman Kern, ISDS, TU Graz  
2020-10-01

- 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)?

19

Roman Kern, ISDS, TU Graz  
2020-10-01

- The language of the presentation is English
- 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)

20

Roman Kern, ISDS, TU Graz  
2020-10-01

# The End