

DATA SCIENCE

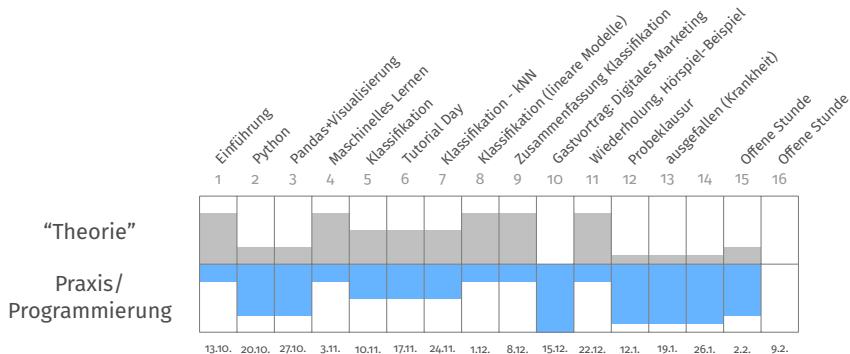
VORLESUNG 14

PROF. DR. CHRISTIAN BOCKERMANN

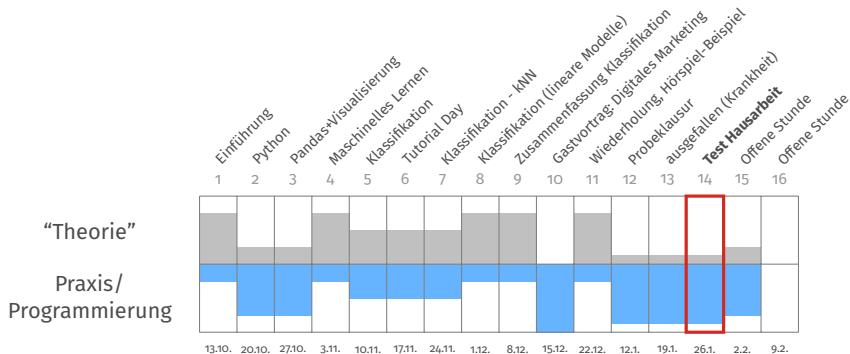
HOCHSCHULE BOCHUM

WINTERSEMESTER 2020/2021

Wo sind wir heute?



Wo sind wir heute?



Vorlesung (restliches Wintersemester)

- Keine Inhaltsfolien mehr (nur Präsenzfolien)
- Präsenzfolien zu Clustering, Regression (Anwendungen)
- Vertiefung der Themen im kommenden Semester
- Präsenzfolien wie gewohnt dienstags ab 13 Uhr
- Wiederholung, Übungen zu den bisherigen Inhalten

Hausarbeit

Ablauf Hausarbeit

- Aufgabenstellung als Aufgabe im Moodle Kurs
- Freischaltung / Abgabe über Moodle Kurs
- Zeitlicher Rahmen:

Ausgabe Thema: 26.2.2021 um 13:30 Uhr

Finale Abgabe: 5.3.2021 um 23:59 Uhr

- Bearbeitung als Jupyter-Notebook (Verzeichnis Hausarbeit), Abgabe des PDF-Exports via Moodle

Ablauf Hausarbeit

- Aufgabenstellung als Aufgabe im Moodle Kurs
- Freischaltung / Abgabe über Moodle Kurs
- Zeitlicher Rahmen:

Ausgabe Thema: 26.2.2021 um 13:30 Uhr

Finale Abgabe: 5.3.2021 um 23:59 Uhr

- Bearbeitung als Jupyter-Notebook (Verzeichnis Hausarbeit), Abgabe des PDF-Exports via Moodle



Am 5.3.2021 wird um 23:59 ihr Hausarbeit-Verzeichnis auf read-only gesetzt. Danach sind keine Änderungen am Notebook mehr möglich!

Fragemöglichkeiten während der Hausarbeit

- Während “Ausgabephase” – 26.2. 13:30 - 15:00 Uhr
- Montag, den 1.3. um 14:00 Uhr
- Mittwoch, den 3.3. um 13:00 Uhr
- Freitag, den 5.3. um 13:00 Uhr

Zu obigen Zeiten kurz (ca. 15min) im Vorlesungsraum, falls Bedarf.
Weitere Sprechstunden auf Anfrage.

Aufgabenstellung Hausarbeit

Zusammenhängende Aufgabenstellung für Analyse anhand eines Fallbeispiels:

1. Aufgaben zu Datenvorverarbeitung (Python Basics)
 - Filtern von Elementen aus Listen, Erstellen neuer Listen
 - Eigene Funktionen, Anwendung auf Listen
2. Aufgaben zu Statistiken, Aggregaten, Plots mit Pandas
 - Daten in DataFrame laden, Berechnung von Spalten
 - Filtern von Daten, Einfache Statistiken, ggf. Plots
3. Aufgaben zur Klassifikation
 - Daten laden, X und y bestimmen, Modell trainieren
 - Trainings- und/oder Test-Fehler, ggf. Parameter-Optimierung

Gestaltung der Hausarbeit als “Dossier”

- Vorteil von Notebooks:
Code und Beschreibung/Ergebnisse in **einem** Dokument

Gestaltung der Hausarbeit als “Dossier”

- Vorteil von Notebooks:
Code und Beschreibung/Ergebnisse in **einem** Dokument
- Nutzen Sie die Vorteile von **Markdown** um ihr Notebook zu strukturieren (Überschriften, fett/kursiv, Aufzählungen,...)

Gestaltung der Hausarbeit als “Dossier”

- Vorteil von Notebooks:
Code und Beschreibung/Ergebnisse in **einem** Dokument
- Nutzen Sie die Vorteile von **Markdown** um ihr Notebook zu strukturieren (Überschriften, fett/kursiv, Aufzählungen,...)
- Beschreiben Sie Ihr Vorgehen mit Text, dann Code:

`Zuerst brauchen wir eine Funktion, die für ein Kunden-Tupel den X-Wert berechnet.`

`Danach können wir mit einer Schleife die Werte für alle Kunden berechnen.`