

Lehrform / Anzahl der SWS: 2VL + 2UE

Kreditpunkte: 6

Turnus: i.d.R. angeboten in jedem SoSe

Inhaltliche Voraussetzungen: PI1, Mathe 1+2

Sprache: Englisch

Lehrende: Dr. Hendrik Heuer, Dr. Juliane Jarke

Studiengang	(Primäre) Modul(e), ggf. Schwerpunkt(e)	Semester
Informatik (Bachelor VF)	IBVA	ab 3. Sem.
Digitale Medien (Bachelor)	B-MA-2	ab 3. Sem.
Wirtschaftsinformatik (Bachelor)	WI-W/11	ab 3. Sem.
Informatik (Master)	<i>General Studies</i> (auf Antrag)	ab 1. Sem.
Digitale Medien (Master)	B-MA-2	ab 3. Sem.

Lernziele / Learning Outcome:

During this course, you will work in small groups on independent projects. Each group will have to

- formulate a research question
- pick and potentially collect a dataset
- pick a suitable operationalisation and method
- find and justify the best machine learning model
- describe your approach and findings in a report

Inhalte / Contents:

From medical decision support systems to automatic language translation, from sorting and prioritizing news on social networks to autonomous cars: Machine learning is woven into the fabric of daily life. Applying machine learning, data science aims to extract knowledge or insights from data.

The class will provide an introduction to data science and applied machine learning. For this, the programming language Python will be used (and taught). You will learn about the difference between supervised and unsupervised machine learning, and three machine learning tasks:

- classification (e.g. k-NN, Decision Trees, Support Vector Machines)
- regression (Linear Regression, Logistic Regression)
- clustering (k-means, dimensionality reduction with PCA and t-SNE)

We will explore natural language processing for text mining and computer vision. Evaluation, as an integral part of data science, will be taught as well as data processing and data mining. To communicate our findings, we will also look at different visualization techniques.

Unterlagen (Skripte, Literatur, Programme):

- Doing Data Science, Cathy O'Neil & Rachel Schutt
 - Introduction to Machine Learning with Python, Andreas C. Müller & Sarah Guido
 - Building Machine Learning Systems with Python, Luis Pedro Coelho & Willi Richert
-

Form der Prüfung: i.d.R. Übungsaufgaben und Fachgespräch

Arbeitsaufwand:	180h
Vorlesungen	24h
Übungen	20h
Python lernen	6h
Exposé	18h
Exposé Peer Feedback	6h
Progress Presentation	12h
Progress Presentation Peer Feedback	10h
Final Presentation	12h
Final Report	24h
Final Project	48h

Weitere Hinweise: Lectures from last year can be retrieved online: <http://hen-drik.de/datascience2021/>