

Modulbezeichnung: Data Analysis

Modulcode:

ECTS-Credits: 5

Modulverantwortliche(r): Professor Dr. Gerhard Dikta

Lehr- und Lernmethoden:

Vorlesung:	2	SWS	Vor- und Nachbereitung:	57	Zeitstunden
Übung:	1	SWS	Hausarbeiten / Referate u. a.:	60	Zeitstunden
Praktikum:		SWS	Gesamte Arbeitsbelastung:	150	Zeitstunden

Lernergebnisse:

Fachlich: Die Studierenden verstehen und bewerten aktuelle Konzepte des Themenfeldes Data Analysis. Die Studierenden wenden Verfahren aus den Gebieten Natural Language Processing und Image Processing in realen Problemen an, um strukturierte Daten aus unstrukturierten Datensätzen zu erzeugen. Zudem wenden die Studierenden Techniken zur Analyse und Reduktion strukturierter Daten an und verwenden Methoden zur Visualisierung der gewonnenen Informationen. Sie verwenden die dabei gängigen Werkzeuge. Zum Abschluss des Moduls können die Teilnehmer/innen Datensätze analysieren, Strategien zur Analyse dieser Datensätze entwickeln und selbst implementieren und die erzielten Ergebnisse adäquat darstellen.

Überfachlich: Die Studierenden entwickeln ein Projekt in Teamarbeit und stellen die Ergebnisse der Gruppe vor. Sie setzen sich kritisch mit den aktuellen technischen Möglichkeiten auseinander.

Beitrag zum Gesamtqualifikationsziel des Studiengangs: Die Modulinhalte umfassen wichtige Anwendungsgebiete mathematischer Verfahren in der Informatik. Sie führen Lerninhalte der Mathematik und der Informatik anhand praktischer Anwendungen zusammen.

Inhaltbeschreibung:

Vorlesung: Data Analysis ist ein grundlegender Schritt, um strukturierte Informationen aus einer unstrukturierten Datenbasis zu gewinnen. Ziel ist die Destillation einer Datenbasis, die für weitere Anwendungen geeignet ist, z.B. als Datenbasis für Machine Learning. Das Modul stellt Basisinhalte sowie gängige Techniken des Themenbereiches Data Analysis vor. Einzelne Themenfelder sind dabei:

- Pre-Processing von Datensätzen
- Natural Language Processing, Image Processing
- Fourier Analyse, Regularisierung
- Datenreduktion
- Visualisierung

Bachelor-Studiengang: Scientific Programming

Die Vorlesung umfasst die praktische Nutzung der vorgestellten Techniken und bereitet die Studierenden auf die eigene Implementierung einer realen Anwendung aus dem Bereich Data Analysis vor.

Projekt: Im Rahmen des Projekts wird ein reales Problem von den Studierenden vollständig umgesetzt und präsentiert.

Eingangsvoraussetzungen: Lineare Algebra, Stochastik

Art der Prüfung: 75% Schriftliche Klausur im Umfang von 2 Stunden, 25% Projekt

Literatur und Lernunterlagen:

(DSP) Tamal Bose: Digital Signal and Image Processing, ISBN 9780471327271

(ISL) James, G., Witten, D., Hastie, T., Tibshirani: An Introduction to Statistical Learning (with Applications in R), ISBN 9781461471370, <http://www-bcf.usc.edu/~gareth/ISL/index.html>

(DIP) Rafael C. Gonzalez, Richard E. Woods: Digital Image Processing, ISBN 9780131687288, http://www.imageprocessingplace.com/DIP-3E/dip3e_main_page.htm

(DAAG) J. Maindonald, W. John Braun: Data Analysis and Graphics Using R, Cambridge, ISBN 9780521762939, <http://www.cambridge.org/catalogue/catalogue.asp?isbn=9780521762939>

(DM) C. Aggarwal: Data Mining: The Textbook, ISBN 9783319141411, <http://www.springer.com/gp/book/9783319141411>