

BOSS-NR. ?????

Modul INF-ML-319: Grundlagen der Datenwissenschaft (GDW)

identisch mit

INF-BSc-319: Grundlagen der Datenwissenschaft (Bachelor Informatik / Angewandte Informatik)

INF-BL-319: Grundlagen der Datenwissenschaft (Informatik Lehramt Bachelor)

INF-BSc-AF-EC-319: Grundlagen der Datenwissenschaft (Bachelor Informatik / Angewandte Informatik)

Ist Teil von

INF-BSc-AF-DLI-213: Grundlagen der Datenwissenschaft DLI (Bachelor Informatik / Angewandte Informatik)

Studiengänge: Master Lehramt Informatik Gy/Ge und BK

Turnus jährlich	Dauer 1 Semester	Studienabschnitt ab 5. Semester	Credits 4	Aufwand 120 (45/75)
---------------------------	----------------------------	---	---------------------	-------------------------------

1	Modulstruktur				
	Nr.	Element / Lehrveranstaltung	Typ	Credits	SWS
	1	Grundlagen der Datenwissenschaften	V	2	2
	2	Übungen zu Grundlagen der Datenwissenschaften	Ü	2	1

2 **Lehrveranstaltungssprache:** Deutsch und Englisch, insbesondere soll der 2. Teil im Rahmen des International Summer Programs angeboten und daher auf Englisch gehalten werden

3 **Lehrinhalte**

Die Verarbeitung von Daten stellt im Zeitalter von „Big Data“ neue Herausforderungen an die Informatik. Anfragen an Datenbanken werden immer komplexer und benötigen Ansätze des Maschinellen Lernens und des Data Minings, und die Größe der Datenmengen stellt neue algorithmische Herausforderung an diese Techniken. Die Vorlesung soll die Grundlagen moderner Ansätze zur Verarbeitung von Daten und zum Extrahieren von Wissen aus ihnen abdecken. Das umfasst Fragestellungen zu Datenbankmanagementsystemen wie z.B. Datenbanken, Anfragesprachen, verteilte Dateisysteme, Map-Reduce und Datenstrommodelle als Hilfsmittel um Algorithmen zu entwerfen, die mit großen Datenmengen umgehen können. Dann soll anhand der Nächste-Nachbarn-Klassifikation Techniken zur Ähnlichkeitssuche wie z.B. minhash, locality-sensitive Hashing und spektrales Hashing angesprochen werden. Danach sollen weitere Basistechniken des Maschinellen Lernens und des Data Minings erläutert werden wie z.B. naive Bayes, Entscheidungsbäume, Stützvektormaschinen, die Clusteranalyse und das Finden von häufigen Teilmengen. Abschliessend sollen die Grundlagen von Suchmaschinen vermittelt werden wie z.B. Googles PageRank und Kleinbergs Hypertext-induced Topic Selection zur Berechnung von Hubs und Authorities.

Die Vorlesung basiert auf <http://www.mmds.org> und wird hauptsächlich algorithmische Aspekte der Datenwissenschaften betrachten. Grundlegende Konzepte der Mathematik und Statistik sollen aber auch vermittelt werden.

Content

In the age of "big data", data processing faces new challenges. Queries become more complex and often involve data mining and machine learning tasks, and the scale of the datasets requires new algorithmic approaches. This course will cover the foundations of modern data processing and mining. This includes topics from database management, such as databases, query languages, distributed file systems, map-reduce, and the streaming model as tools for creating algorithms that succeed on massive amounts of data. Starting from the nearest-neighbor learning approach, it will discuss similarity search, including the key techniques of minhashing, locality-sensitive hashing, and spectral hashing. Then it will continue with basic machine learning and data mining techniques such as naive Bayes, decision trees and support vector machines as well as frequent-itemset mining approaches and algorithms for clustering. Finally, it will lay out the foundations of the technology of search engines, including Google's PageRank and the hubs-and-authorities approach.

We will mainly focus on computational aspects of the data science, although math and statistics will also be touched. It is based on <http://www.mmds.org>

4 **Kompetenzen**

Ziel des Moduls ist es, dass Studierende eine ausreichenden Kompetenz zu vermitteln, die ihnen eine aktiven Lösungsgestaltung von alltäglich auftauchenden Problemen der Datenwissenschaften

	befähigt. Im Einzelnen: Verständnis dafür, was Daten sind, Kenntnis der grundlegenden und fortgeschrittenen Verfahren der verteilten Datenverarbeitung, Kenntnis der grundlegenden und fortgeschrittenen Verfahren des Maschinellen Lernens und des Data Minings zum Extrahieren von Wissen aus Daten. Insbesondere soll die Studierende durch ein grundlegendes Verständnis der Prinzipien der Datenwissenschaften die Fähigkeit erhalten, deren Möglichkeiten und Grenzen in bestimmten Anwendungsfeldern einschätzen zu können.	
5	Prüfungen <i>Modulprüfung:</i> Klausur oder mündliche Prüfung ^{BOSS-NR. ?????} <i>Studienleistung:</i> <ul style="list-style-type: none"> • aktive Mitarbeit in der Übung und Erreichen der Mindestpunktzahl bei den Übungsaufgaben ^{BOSS-NR. ????} Die Studienleistung ist Voraussetzung für die Teilnahme an der Modulprüfung.	
6	Prüfungsformen und -leistungen <input checked="" type="checkbox"/> Modulprüfung <input type="checkbox"/> Teilleistungen	
7	Teilnahmevoraussetzungen <i>Erfolgreich abgeschlossen:</i> -keine- <i>Wünschenswerte Kenntnisse:</i> Grundlegende Kenntnisse der Mathematik (insbes. Statistik)	
8	Modultyp und Verwendbarkeit des Moduls <ul style="list-style-type: none"> • Wahlmodul im Fach Informatik des Lehramtsmasterstudiengangs für Gymnasien und Gesamtschulen für Studierende, die gemäß der FSB-B_GyGe_inf i.d.F.v. 24.09.2014 ein Wahlmodul nicht im Lehramtsbachelorstudiengang erfolgreich absolviert haben • Wahlmodul im Fach Informatik des Lehramtsmasterstudiengangs für Berufskollegs 	
9	Modulbeauftragte/r Prof. Dr. E. Schubert	Zuständige Fakultät Informatik
		<small> Beschluss Fakultätsrat 10.03.2015 Änderung Fakultätsrat 19.09.2018, 27.10.2021, 18.10.2022 </small>