

INF-BSc-319: Grundlagen der Datenwissenschaft (GDW)					BOSS-Nr. 88700	
Englischer Modultitel: Foundations of Data Science						
Studiengänge: Bachelorstudiengang Informatik , Bachelorstudiengang Angewandte Informatik						
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		Vorlesung	2	2
2	Übungen zu Grundlagen der Datenwissenschaften		Übung	2	1	
2	Lehrveranstaltungs-sprache: Deutsch und Englisch					
3	<p>Lehrinhalte</p> <p>Im Zeitalter von „Big Data“ stehen die Informatik und die Datenverarbeitung vor vielfältigen Herausforderungen. Die zunehmende Komplexität der Anfragen an Datenbanken erfordert den Einsatz fortschrittlicher Methoden des maschinellen Lernens und des Data Minings. Darüber hinaus bringen die enormen Datenmengen algorithmische Fragestellungen mit sich, die es zu bewältigen gilt. Diese Vorlesung zielt darauf ab, die Grundlagen moderner Ansätze zur Datenverarbeitung sowie zur Wissensgewinnung aus großen Datensätzen zu vermitteln. Zunächst liegt der Fokus auf der Datenvorverarbeitung und -akquise, beispielsweise durch APIs oder Web-Scraping, sowie auf verschiedenen Daten- und Dateiformaten. Im Anschluss wird die parallele Verarbeitung großer Datenmengen behandelt; dabei werden Herausforderungen bei der Parallelisierung diskutiert und Berechnungsmodelle wie Map-Reduce, Think-like-a-vertex und (parallele) Datenstromverarbeitung eingeführt, um eine effiziente Bearbeitung großer Datensätze zu ermöglichen. Diese Konzepte werden anschließend anhand konkreter Anwendungen vertieft. Dabei werden unter anderem Probleme wie die Skalierbarkeit von Suchmaschinen sowie Graph-Probleme anhand des PageRank-Algorithmus von Google analysiert. Zudem wird das Frequent Itemset Mining behandelt, ebenso wie Techniken zur Ähnlichkeitssuche wie MinHash, Locality-Sensitive Hashing (LSH) und Hierarchical Navigable Small World (HNSW). Ein besonderer Schwerpunkt der Vorlesung liegt auf den algorithmischen Aspekten der Datenwissenschaften.</p>					
4	<p>Kompetenzen</p> <ul style="list-style-type: none"> • Die Studierenden können grundlegende Konzepte der Datenverarbeitung und Wissensgewinnung aus großen Datensätzen erläutern. • Die Studierenden sind in der Lage, geeignete Methoden zur Datenvorverarbeitung und -akquise auszuwählen und anzuwenden. • Die Studierenden verstehen die Herausforderungen der parallelen Verarbeitung großer Datenmengen und können entsprechende Berechnungsmodelle erklären und einsetzen. • Die Studierenden können algorithmische Probleme im Kontext von Suchmaschinen und Graphen erkennen und Lösungen anwenden. • Die Studierenden sind fähig, Techniken zur Ähnlichkeitssuche zu beschreiben und deren Anwendung in praktischen Szenarien zu demonstrieren. • Die Studierenden reflektieren kritisch über die Herausforderungen der Datenqualität in der Datenwissenschaft und deren Auswirkungen auf die Ergebnisse von Analysen und Modellen. • Die Studierenden reflektieren über die gesellschaftlichen Konsequenzen, die sich aus einer umfangreichen Datennutzung ergeben. 					
5	<p>Prüfungen</p> <ul style="list-style-type: none"> • Modulprüfung: Klausur oder mündliche Prüfung gemäß Ankündigung in der Veranstaltung BOSS-Nr. 88791 • Studienleistung: erfolgreiche Teilnahme an den Übungen gemäß Ankündigung in der Veranstaltung BOSS-Nr. 89941 <p>Die Studienleistung ist Voraussetzung für die Teilnahme an der Modulprüfung.</p>					

6	Prüfungsformen und -leistungen [x] Modulprüfungen [] Teilleistung		
7	Teilnahmevoraussetzungen <ul style="list-style-type: none"> • Erfolgreich abgeschlossen: keine • Wünschenswerte Kenntnisse: Programmierkenntnisse, Grundlegende Kenntnisse der Statistik 		
8	Modultyp und Verwendbarkeit des Moduls Wahlmodul in den Bachelor-Studiengängen Informatik und Angewandte Informatik		
9	Modulbeauftragte/r Prof. Dr. E. Schubert und JProf. Dr. T. Liebig	Zuständige Fakultät: Informatik	Beschluss Fakultätsrat: 03.09.2025