

INF-BSc-AF-DLI-213: Grundlagen der Datenwissenschaft DLI					
Erweiterung von: INF-BSc-319: Grundlagen der Datenwissenschaft (Bachelorstudiengänge Informatik und Angewandte Informatik) ¹					
Studiengang: Bachelorstudiengang Angewandte Informatik					
Turnus jährlich	Dauer 1 Semester	Studienabschnitt 4.-6. Semester	Credits 5 ²	Aufwand 180 (60/120)	
1	Modulstruktur				
	Nr.	Element / Lehrveranstaltung	Typ	Credits	SWS
	1	Vorlesung Grundlagen der Datenwissenschaften	V	2	2
	2	Übungen zu Grundlagen der Datenwissenschaften	Ü	2	1
	3	Übungsprojekt zu Grundlagen der Datenwissenschaften	Proj.	1	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	<p>Lehrinhalte</p> <p>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.</p> <p>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.</p> <p>Content</p> <p>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.</p> <p>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</p>				

¹ Wird das Modul „Grundlagen der Datenwissenschaft DLI“ im Anwendungsfach Dienstleistungsinformatik gewählt, so ist die Wahl des Moduls „Grundlagen der Datenwissenschaft“ (INF-BSc-319) als Wahlmodul gem. Anhang A Abs. 3 Punkt c BPO AngInf nicht möglich.

² Vor dem Wintersemester 2019/20 hatte das Modul einen Umfang von 5 LP; das Element 3 einen Umfang von 2 LP.

4	<p>Kompetenzen</p> <p>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.</p> <p>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.</p> <p>Content</p> <p>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.</p> <p>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</p>		
5	<p>Prüfungen</p> <p><i>Modulprüfung:</i> Klausur oder mündliche Prüfung ^{BOSS-Nr.: -keine-}</p> <p><i>Studienleistungen:</i></p> <p>(1) aktive Mitarbeit in der Übung und Erreichen der Mindestpunktzahl bei den Übungsaufgaben ^{BOSS-Nr.: -keine-}</p> <p>(2) erfolgreiche Teilnahme am Übungsprojekt gemäß Ankündigung in der Veranstaltung ^{BOSS-Nr.: -keine-}</p> <p>Die Studienleistungen (1) und (2) sind Voraussetzung für die Teilnahme an der Modulprüfung.</p>		
6	<p>Prüfungsformen und -leistungen</p> <p><input checked="" type="checkbox"/> Modulprüfung <input type="checkbox"/> Teilleistungen</p>		
7	<p>Teilnahmevoraussetzungen</p> <p><i>Erfolgreich abgeschlossen:</i> –keine–</p> <p><i>Wünschenswerte Kenntnisse:</i> Grundlegende Kenntnisse der Mathematik (insbes. Statistik)</p>		
8	<p>Modultyp und Verwendbarkeit des Moduls</p> <p>Wahlpflichtmodul im Bachelorstudiengang Angewandte Informatik, Anwendungsfach Dienstleistungsinformatik</p> <p>Katalog Betriebliche Informationssysteme</p>		
9	<p>Modulbeauftragte/r (Studiendekan)</p>	<p>Zuständige Fakultät Informatik</p>	<p>Beschluss Fakultätsrat 20.04.2016 Änderung Fakultätsrat 19.09.2018, 11.12.2019</p>