

Modulbezeichnung:	Big Data Analytics	
Modulnummer: DLBECWBDA	Semester: --	Dauer: Minimaldauer 1 Semester
Modultyp: Wahlpflicht		Regulär angeboten im: WS, SS
Workload: 300 h		ECTS Punkte : 10
Zugangsvoraussetzungen: keine		Unterrichtssprache: Deutsch
Kurse im Modul: <ul style="list-style-type: none"> • Data Analytics und Big Data (DLBINGDABD01) • Statistical Computing (DLBDBSC01) 		Workload: Selbststudium: 220 h Selbstüberprüfung: 40 h Tutorien: 40 h
Kurskoordinatoren/Tutoren:: Siehe aktuelle Liste der Tutoren im Learning Management System		Modulverantwortliche(r): Dr. Amelie Eilken
Bezüge zu anderen Programmen:		Bezüge zu anderen Modulen im Programm:
<p>Qualifikations- und Lernziele des Moduls :</p> <p>Data Analytics und Big Data</p> <p>Die Studierenden</p> <ul style="list-style-type: none"> • können zwischen Informationen und Daten unterscheiden und kennen die Bedeutung dieser Begriffe für die Entscheidungsfindung. • können die Big Data-Problematik, insbesondere im Zusammenhang mit dem Internet of Things, herleiten und anhand von Beispielen beschreiben. • kennen Grundlagen aus der Statistik, die für die Analyse großer Datenbestände notwendig sind. • kennen den Prozess des Data Mining und können verschiedene Methoden darin einordnen. • kennen ausgewählte Methoden und Technologien, die im Big Data-Kontext angewendet werden und können sie an einfachen Beispielen anwenden. • kennen die rechtlichen Rahmenbedingungen für die Anwendung der Datenanalyse in Deutschland sowie international. • kennen die besonderen Chancen und Herausforderungen der Anwendung von Big Data-Analysen in der Industrie. <p>Statistical Computing</p> <p>Nach erfolgreichem Abschluss sind die Studierenden in der Lage,</p> <ul style="list-style-type: none"> • den Begriff Statistical Computing einzuordnen und abzugrenzen. • sich eine PC-Arbeitsumgebung zu schaffen, mit der Aufgaben aus dem Themengebiet Statistical Computing bearbeitet werden können. • einfache Programme mit der Programmiersprache R zu schreiben. • mit R Daten zu importieren und zu exportieren. • mit R verschiedene statistische Verfahren anzuwenden, von der deskriptiven Statistik über die Inferenzstatistik bis hin zur Varianz- und Regressionsanalyse. 		
Lehrinhalt des Moduls:		

Data Analytics und Big Data:

- Einführung in die Analyse von Daten
- Statistische Grundlagen
- Data Mining
- Big Data-Methoden und Technologien
- Rechtliche Aspekte der Datenanalyse
- Lösungsszenarien
- Anwendung von Big Data in der Industrie

Statistical Computing:

- Einstieg in das Statistical Computing
- Grundlagen der Programmierung mit R
- Auf Daten zugreifen
- Deskriptive Statistik
- Inferenzstatistik
- Varianzanalyse
- Regressionsanalyse

Lehrmethoden:	Siehe Kursbeschreibung	
Literatur:	Siehe Literaturliste der vorliegenden Kursbeschreibungen	
Anteil der Modulnote an der Gesamtabchlussnote des Programms : --	Prüfungszulassungsvoraussetzung:	Abschlussprüfungen:
	Siehe Kursbeschreibungen	<ul style="list-style-type: none"> • DLBINGDABD01: Schriftliche Ausarbeitung: Fallstudie (50%) • DLBDBSC01: Schriftliche Ausarbeitung: Fallstudie (50%)

Kursnummer: DLBINGDABD01	Kursname: Data Analytics und Big Data	Gesamtstunden: 150 h ECTS Punkte: 5 ECTS
Kurstyp: Pflicht Kursangebot : Kursdauer : Minimaldauer 1 Semester		Zugangsvoraussetzungen: keine
Kurskoordinator(en) / Dozenten / Lektoren: Siehe aktuelle Liste der Tutoren im Learning Management System		Bezüge zu anderen Modulen: Siehe Modulbeschreibung
<p>Beschreibung des Kurses:</p> <p>Ziel des Kurses ist es, die Studierenden mit ausgewählten Methoden und Techniken der Datenanalyse im Kontext stetig wachsender, heterogener Datenmengen vertraut zu machen. Hierzu wird zunächst die grundsätzliche Relevanz von Big Data-Methoden anhand der historischen Entwicklung der Datenbestände motiviert. Entscheidend ist hier unter anderem die kontinuierliche Belieferung der Systeme mit Sensordaten aus dem Internet of Things. Es folgt eine kurze Einführung in die wesentlichen statistischen Grundlagen, bevor die einzelnen Schritte des Data Mining-Prozess thematisiert werden. In Abgrenzung zu diesen klassischen Verfahren werden dann ausgewählte Methoden vorgestellt, mit denen Datenbestände im Big Data-Kontext analysierbar gemacht werden können. Weil die Datenanalyse bestimmten gesetzlichen Rahmenbedingungen unterliegt, werden in diesem Kurs zudem rechtliche Aspekte wie der Datenschutz behandelt. Der Kurs schließt mit einem Überblick über den Praxiseinsatz von Big Data-Methoden und -Werkzeugen. Hierbei werden insbesondere die Anwendungsfelder im industriellen Kontext beleuchtet.</p> <p>Kursziele:</p> <p>Die Studierenden</p> <ul style="list-style-type: none"> • können zwischen Informationen und Daten unterscheiden und kennen die Bedeutung dieser Begriffe für die Entscheidungsfindung. • können die Big Data-Problematik, insbesondere im Zusammenhang mit dem Internet of Things, herleiten und anhand von Beispielen beschreiben. • kennen Grundlagen aus der Statistik, die für die Analyse großer Datenbestände notwendig sind. • kennen den Prozess des Data Mining und können verschiedene Methoden darin einordnen. • kennen ausgewählte Methoden und Technologien, die im Big Data-Kontext angewendet werden und können sie an einfachen Beispielen anwenden. • kennen die rechtlichen Rahmenbedingungen für die Anwendung der Datenanalyse in Deutschland sowie international. • kennen die besonderen Chancen und Herausforderungen der Anwendung von Big Data-Analysen in der Industrie. <p>Lehrmethoden:</p> <p>Die Lehrmaterialien enthalten Skripte, Video-Vorlesungen, Übungen, Podcasts, (Online-) Tutorien und Fallstudien. Sie sind so strukturiert, dass Studierende sie in freier Ortswahl und zeitlich unabhängig bearbeiten können.</p> <p>Inhalte des Kurses:</p> <p>1. Einführung in die Analyse von Daten</p> <p>1.1 Entscheidungen, Informationen, Daten</p> <p>1.2 Historische Entwicklung der Speicherung und Auswertung von Daten</p> <p>1.3 Big Data: Eigenschaften und Beispiele</p>		

1.4 Datenanalyse

1.5 Das Internet of Things als Treiber für Big Data

2. Statistische Grundlagen

2.1 Deskriptive Datenanalyse

2.2 Inferenzielle Datenanalyse

2.3 Explorative Datenanalyse

2.4 Multivariate Datenanalyse

3. Data Mining

3.1 Knowledge Discovery in Databases

3.2 Assoziationsanalyse

3.3 Korrelationsanalyse

3.4 Prognose

3.5 Clusteranalyse

3.6 Klassifikation

4. Big Data-Methoden und -Technologien

4.1 Technologiebausteine

4.2 MapReduce

4.3 Text- und semantische Analyse

4.4 Audio- und Videoanalyse

4.5 BASE und NoSQL

4.6 In-Memory-Datenbanken

4.7 Big-Data-Erfolgsfaktoren

5. Rechtliche Aspekte der Datenanalyse

5.1 Datenschutzgrundsätze in Deutschland

5.2 Anonymisierung und Pseudonymisierung

5.3 Internationale Datenanalyse

5.4 Leistungs- und Integritätsschutz

6. Lösungsszenarien

7. Anwendung von Big Data in der Industrie

7.1 Produktion und Logistik

7.2 Effizienzsteigerungen in der Supply Chain

7.3 Schlüsselfaktor Daten

7.4 Beispiele und Fazit

Literatur:

- Brandt, S. (2013): Datenanalyse für Naturwissenschaftler und Ingenieure. Mit statistischen Methoden und Java-Programmen. 5. Auflage, Springer, Wiesbaden. ISBN-13: 978-3-642376634.
- Dorschel, J. (Hrsg.) (2015): Praxishandbuch Big Data. Wirtschaft – Recht – Technik. Springer Gabler Wiesbaden. ISBN-13: 978-3-658072889.
- Gandomi, A./Haider, M. (2015): Beyond the hype. Big data concepts, methods, and analytics. In: International Journal of Information Management, 35. Jg., Heft 2, S. 137–144.
- Provost, F./Fawcett, T. (2013): Data science for business. What You Need to Know About Data Mining and Data-Analytic Thinking. O'Reilly, Sebastopol (CA). ISBN-13: 978-1449361327.
- Runkler, T. A. (2015): Data Mining. Modelle und Algorithmen intelligenter Datenanalyse. 2. Auflage, Springer Vieweg, Wiesbaden. ISBN-13: 978-3834816948

Prüfungsleistung:

Schriftliche Ausarbeitung: Fallstudie

Zeitaufwand Studierenden (in Std.): 150

Selbststudium (in Std.): 110
Selbstüberprüfung (in Std.): 20
Tutorien (in Std.): 20

Kursnummer: DLBDBSC01	Kursname: Statistical Computing	Gesamtstunden: 150 h ECTS Punkte: 5 ECTS
Kurstyp: Pflicht Kursangebot : Kursdauer : Minimaldauer 1 Semester		Zugangsvoraussetzungen: keine
Kurskoordinator(en) / Dozenten / Lektoren: Siehe aktuelle Liste der Tutoren im Learning Management System		Bezüge zu anderen Modulen: Siehe Modulbeschreibung
<p>Beschreibung des Kurses:</p> <p>Das Statistical Computing verbindet die Konzepte und Methoden der Statistik mit den Werkzeugen aus der Informatik. Das Ergebnis sind unter anderem Statistik-Programme und -Programmiersprachen, die viele nützliche Funktionen zur Analyse von digital verfügbaren Datenquellen bieten. In diesem Kurs wird den Studierenden die Programmiersprache R vermittelt, um damit anschließend statistische Verfahren (z. B. Regressionsanalyse, Varianzanalyse) anwenden zu können. Im Rahmen einer Fallstudie sollen die erworbenen Kompetenzen eingesetzt werden, um Zusammenhänge aus komplexen Datenquellen zu gewinnen und graphisch darzustellen.</p> <p>Kursziele:</p> <p>Nach erfolgreichem Abschluss des Kurses sind die Studierenden in der Lage,</p> <ul style="list-style-type: none"> • den Begriff Statistical Computing einzuordnen und abzugrenzen. • sich eine PC-Arbeitsumgebung zu schaffen, mit der Aufgaben aus dem Themengebiet Statistical Computing bearbeitet werden können. • einfache Programme mit der Programmiersprache R zu schreiben. • mit R Daten zu importieren und zu exportieren. • mit R verschiedene statistische Verfahren anzuwenden, von der deskriptiven Statistik über die Inferenzstatistik bis hin zur Varianz- und Regressionsanalyse. <p>Lehrmethoden:</p> <p>Die Lehrmaterialien enthalten Skripte, Video-Vorlesungen, Übungen, Podcasts, (Online-) Tutorien und Fallstudien. Sie sind so strukturiert, dass Studierende sie in freier Ortswahl und zeitlich unabhängig bearbeiten können.</p> <p>Inhalte des Kurses:</p> <p>1. Einstieg in das Statistical Computing</p> <p>1.1 Begriffsbestimmung und Abgrenzung</p> <p>1.2 Statistik-Programm vs. Statistik-Programmiersprache</p> <p>1.3 Einrichtung der Arbeitsumgebung</p> <p>2. Grundlagen der Programmierung mit R</p> <p>2.1 R als Taschenrechner</p> <p>2.2 Zuweisungen und Variablen</p> <p>2.3 Vektoren und Matrizen</p> <p>2.4 Logik</p> <p>2.5 Funktionen</p> <p>2.6 Datentypen und Datenstrukturen</p>		

3. Auf Daten zugreifen

3.1 Daten eingeben

3.2 Import und Export von externen Dateien

3.3 Datenmanagement in R

4. Deskriptive Statistik

4.1 Univariate deskriptive Statistik

4.2 Bivariate deskriptive Statistik

5. Inferenzstatistik

5.1 Verteilungen

5.2 Stichproben

5.3 t-Tests

6. Varianzanalyse

6.1 Grundlagen und Abgrenzung zum t-Test

6.2 Einfaktorielle Varianzanalyse

6.3 Zweifaktorielle Varianzanalyse

7. Regressionsanalyse

7.1 Korrelation

7.2 Lineare Regression

7.3 Weitere Modelle und Verfahren

Literatur:

- Ligges, U. (2008): Programmieren mit R. 3. Auflage, Springer, Heidelberg.
- Luhmann, M. (2015): R für Einsteiger. Einführung in die Statistiksoftware für die Sozialwissenschaften. Beltz, Weinheim, Basel.
- Toomey, D. (2017): Jupyter for Data Science. Exploratory analysis, statistical modeling, machine learning, and data visualization with Jupyter. Packt Publishing, Birmingham, UK.
- Vanderplas, J. (2017): Data Science mit Python. Das Handbuch für den Einsatz von Python, Jupyter, NumPy, Pandas, Matplotlib, Scikit-Learn. mitp, Frechen.
- Wollschläger, D. (2015): Grundlagen der Datenanalyse mit R. Eine anwendungsorientierte Einführung. Springer Spektrum, Berlin Heidelberg.

Prüfungsleistung:

Schriftliche Ausarbeitung: Fallstudie

Zeitaufwand Studierende (in Std.): 150

Selbststudium (in Std.): 110

Selbstüberprüfung (in Std.): 20

Tutorien (in Std.): 20