

Syllabus
BAE3341 Data Engineering und Data Science
Prof. Dr.-Ing. Heiko Thimm
Sommersemester 2024

Niveau	Bachelor												
Credits	3												
SWS	2												
Workload	90 Stunden												
Voraussetzungen	MNS1091 Mathematik 1 BAE1131 Einführung in die Informatik BAE1141 Projektseminar Programmierung BAE1090 Quantitative Methoden I BAE2080 Quantitative Methoden II BAE2430 Informationstechnologie												
Uhrzeit	s. LSF												
Raum	s. LSF												
Starttermin	s. LSF												
Lehrende(r)	<table border="1"><tr><td>Name</td><td>Prof. Dr.-Ing. Heiko Thimm</td></tr><tr><td>Büro</td><td>T1.3.29</td></tr><tr><td>Virtuelles Büro</td><td>Virtueller Hörsaal Prof. Thimm</td></tr><tr><td>Kolloquium</td><td>Montags, 13:45-15:15 Uhr bzw. nach Vereinbarung</td></tr><tr><td>Telefon</td><td>07231 28-6451</td></tr><tr><td>Email</td><td>heiko.thimm@hs-pforzheim.de (bevorzugte Kommunikationsform)</td></tr></table>	Name	Prof. Dr.-Ing. Heiko Thimm	Büro	T1.3.29	Virtuelles Büro	Virtueller Hörsaal Prof. Thimm	Kolloquium	Montags, 13:45-15:15 Uhr bzw. nach Vereinbarung	Telefon	07231 28-6451	Email	heiko.thimm@hs-pforzheim.de (bevorzugte Kommunikationsform)
Name	Prof. Dr.-Ing. Heiko Thimm												
Büro	T1.3.29												
Virtuelles Büro	Virtueller Hörsaal Prof. Thimm												
Kolloquium	Montags, 13:45-15:15 Uhr bzw. nach Vereinbarung												
Telefon	07231 28-6451												
Email	heiko.thimm@hs-pforzheim.de (bevorzugte Kommunikationsform)												

Kurzbeschreibung

In dieser Lehrveranstaltung werden grundlegende Konzepte, Methoden und Vorgehensweisen des als „Data Engineering und Data Science“ bezeichneten interdisziplinären Wissensgebiets aus theoretischer und praktischer Sicht behandelt. Der Schwerpunkt liegt dabei auf der Vermittlung des von der Informatik bereit gestellten Instrumentariums zur erfolgreichen Durchführung von Data Engineering und Data Science Projekten. Hierzu wird das CRISP-DM Vorgehensmodell mit seinen verschiedenen Phasen und Aufgabenstellungen vorgestellt. Ebenso werden ausgewählte Machine Learning Methoden und Modelle behandelt. Im begleitenden Labor (BAE3342) werden zu verschiedenen Themen des Lehrstoffs praktische Übungsaufgaben mit ausgewählten Softwaretools sowie Programmieraufgaben in Python durchgeführt.

Gliederung der Veranstaltung (vorläufig)

1. Einführung Data Engineering und Data Science
2. Datenmanagementsysteme und Datenbanken
3. Big Data und Big Data Systeme
4. Datenqualität und Datenaufbereitung
5. Data Science – Überblick und Grundlagen
6. Maschinelles Lernen – Grundprinzipien wichtiger Verfahren
7. Datenverständnis und Datenvorbereitung
8. Modellierung und Evaluation
9. Ausgewählte Methoden und Modelle
10. Kritische Aspekte und Diskussion

Lernziele der Veranstaltung und deren Beitrag zu den Programmzielen

Programmziele	Lernziele der Veranstaltung
Nach Abschluss des Programms sind die Studierenden in der Lage, ...	Nach Abschluss der Veranstaltung sind die Studierenden in der Lage, ...
1 Fachwissen	
1.5 ...ihr solides Grundwissen in Mathematik nachzuweisen.	... relevante Grundkenntnisse der Statistik im Rahmen von Data Science Aufgabenstellungen anzuwenden.
1.6 ...betriebswirtschaftliche Probleme mit quantitativen Methoden und auf der Grundlage fundierter Kompetenzen bei der Datenrecherche lösen zu können.	... zu bewerten welches Potential in vorliegenden oder noch zu erstellenden Datensammlungen zur Lösung betriebswirtschaftlicher Problemstellungen steckt und welchen Beitrag Data Science zur Hebung dieses Potentials leisten kann.
1.7 ...ihr solides Grundwissen in Informatik nachzuweisen.	... die Kennzeichen von Big Data, die Bewertungskriterien von Datenqualität und die grundlegenden Prinzipien und Methoden von Data Engineering und Data Science fundiert zu beschreiben und anzuwenden. Sie können außerdem das CRISP-DM Vorgehensmodell für Data Science Projekte erklären.
1.8 ...ein fundiertes Expertenwissen in ihrer Spezialisierung nachzuweisen.	(Studiengang EI/ID) ... auf Data Science Anwendungen beruhende Innovationsideen zu entwickeln und entsprechende Anforderungsanalysen und Projektskizzen systematisch zu beschreiben.
2 Digitale Kompetenzen	
2.1 ...relevante, in der betrieblichen Praxis eingesetzte IT-Softwaretools und deren Funktionen zu kennen und zu verstehen und verfügen über ein Grundverständnis für digitale Technologien.	... für die Durchführung von Data Engineering und Data Science Projekten zur Verfügung stehende Softwaretools und Programmiersprachen zu benennen und anzuwenden.
2.2 ...die im betrieblichen Umfeld vorzufindenden Informationssysteme effektiv zur Problemlösung zu nutzen.	...die in betrieblichen Informationssystemen vorliegenden Datenbeständen auf deren Nutzbarkeit für Data Science Projekte zu bewerten. Sie kennen die im Rahmen der Datenvorbereitung durchzuführenden Aufgabenstellungen.
2.4 ...im professionellen Umfeld digitale Technologien verantwortungsbewusst einzusetzen.	... die Datenqualität von Datensammlungen zu bewerten und notwendige Aufgabenstellungen zur Qualitätsverbesserung und zur Datenvorbereitung gemäß des CRISP-DM Vorgehensmodells systematisch durchzuführen.
3 Kritisches Denken und analytische Fähigkeiten	
3.1 ...geeignete Methoden kompetent zu verwenden und auf komplexe Fragestellungen anzuwenden.	... aus dem Instrumentarium von Data Engineering und Data Science geeignete Methoden und Vorgehensweisen auszuwählen und anzuwenden um komplexe Fragestellungen zu lösen. Sie beherrschen die Methoden zur Bewertung der Datenqualität von Datensammlungen.
3.2 ...Ergebnisse umfassend zu interpretieren, kritisch zu reflektieren und eigene ganzheitliche Lösungsalternativen für komplexe Fragestellungen zu erarbeiten.	... Ergebnisse von Data Science Anwendungen zielgruppenspezifisch in geeigneten Grafiken darzustellen und die Aussagefähigkeit der Ergebnisse kritisch zu hinterfragen
4 Ethisches Bewusstsein und Nachhaltigkeit	
...fundierte Lösungsstrategien in den Bereichen Ethik, nachhaltige Entwicklung und gesellschaftliche Verantwortung zu entwickeln und auf typische wirtschaftliche Entscheidungsprobleme anzuwenden.	... die aus ethischer Sicht kritischen Aspekte von Data Science Anwendungen zu erklären und verfügbare Lösungen und zukünftige Projektideen hinsichtlich dieser Aspekte kritisch zu bewerten. Es wird außerdem auf das Potential von Data Science und Machine Learning für den Klimaschutz und für Umweltschutzmaßnahmen eingegangen.

5 Kommunikations- und Teamfähigkeit	
5.1 ...komplexe Sachverhalte in klarer schriftlicher Form auszudrücken.	... Aufgabenstellungen in Arbeitsblättern zu bearbeiten

Lehr- und Lernkonzept

Die Themen werden in seminaristischer Unterrichtsform behandelt. Die im Rahmen eines interaktiven Lehrgesprächs behandelten Folien werden den Studierenden als Lehrunterlage zur Verfügung gestellt. Die Lehrunterlage gliedert sich in 8-10 Kapitel. Am Ende von jedem Kapitel befindet sich eine Liste mit Hinweisen auf multimediale Lehrmedien zur Wiederholung und zur Vertiefung des behandelten Lehrstoffs. Es ist dort ebenfalls beschrieben, inwieweit die praktischen Übungsaufgaben des parallel stattfindenden Labors (BAE3342 - Praktische Übungen zu Data Engineering und Data Science) selbstständig vorzubereiten sind. Hierzu werden Verweise auf multimediale Lehrmedien, Lehrbuchkapitel und ggf. relevante Internetquellen gegeben. Im Labor wird durch kurze schriftliche und/oder mündliche Tests der Wissensstand der Studierenden ermittelt. Die Testergebnisse fließen als Prüfungsvorleistung in die Bewertung der Prüfungsleistung mit ein.

Als Leistungsnachweis wird am Ende der Vorlesungszeit (in der Regel in der letzten Vorlesungswoche) eine Prüfung der Prüfungsform PLK/PLR/PLP erbracht. Die für das jeweilige Semester geltende Prüfungsform wird zu Beginn der Vorlesungszeit festgelegt. Aus der bei dieser Prüfung erreichten Punktzahl und der im begleitenden Labor erreichten Punktzahl (Übungsaufgaben, Präsentation Lösungsweg) wird die Gesamtnote ermittelt.

Der Lehrende steht jederzeit als Gesprächspartner zur Verfügung und gibt Unterstützung und Ratschläge. Die Kommunikation erfolgt im persönlichen Gespräch.

Literatur und Kursmaterialien

Die im Lehrgespräch (Lehrveranstaltung BAE 3341) behandelten Folien werden im eLearning System als Lehrskript zur Verfügung gestellt. Das Lehrskript enthält umfassende Angaben und Verweise auf weiterführende Lehrmaterialien einschließlich Links auf Lernunterlagen und Kurse zum Selbststudium.

Es werden unter anderem die nachfolgenden Lehrbücher als Begleitlektüre und zur Wiederholung/Vertiefung des Lehrstoffs empfohlen. Sie stehen in begrenzter Anzahl in der Hochschulbibliothek und teilweise auch als elektronische Lehrbücher zur Verfügung:

- McKinney, W.: Datenanalyse mit Python - Auswertung von Daten mit Pandas, NumPy und IPython, O'Reilly, 2015
- Müller A.C., Guido, S.: Einführung in Machine Learning mit Python, O'Reilly, 2017
- Grus, J.: Einführung in Data Science, Grundprinzipien der Datenanalyse mit Python, O'Reilly, 2019
- Steyer, R.: Programmierung Grundlagen – mit Beispielen in Python, Herdt Campus, 2018

Leistungsnachweis

Der Leistungsnachweis für die Lehrveranstaltung BAE 3341 (Vorlesung bzw. seminaristisches Lehrgespräch) erfolgt im Rahmen einer Klausur (60 min) oder in Absprache mit den Teilnehmern im Rahmen einer Projektarbeit. Im Falle einer Projektarbeit können dabei die Übungsaufgaben des zugehörigen Labors (BA3342) bereits Bestandteil der Projektarbeit darstellen.

Der Leistungsnachweis für die Lehrveranstaltung BAE 3342 (Labor) erfolgt durch die Bearbeitung von Übungsaufgaben, die Bestandteil einer Projektarbeit sein können. Die Vorgehensweise zur Lösung der Aufgabe ist schriftlich nach den Vorgaben des Lehrteams zu dokumentieren und den Laborteilnehmern im Rahmen einer Kurzpräsentation mit live Demo vorzustellen.

Es können in beiden Lehrveranstaltungen des Moduls zusammen maximal 100 Punkte erreicht werden: 60 Punkte in der Lehrveranstaltung BAE3341 in Prüfungen der Prüfungsform PLK/PLR/PLP, 40 Punkte in der Lehrveranstaltung BAE3342 in Prüfungen der Prüfungsform PLL/PLP. Zur Ermittlung der Gesamtnote wird das übliche 100-Punkte-Notenraster angelegt.

Zeitplan

Der Terminplan wird in den ersten beiden Vorlesungswochen mit den Teilnehmern abgestimmt und im weiteren Verlauf des Semesters bei Bedarf ggf. angepasst.

Den aktuellen Terminplan entnehmen Sie bitte dem LSF.

Akademische Integrität und studentische Verantwortung

Bitte beteiligen Sie sich aktiv an der Lehrveranstaltung. Außerhalb der Lehrveranstaltung nutzen Sie bitte meine Sprechstunde, zu der ich Sie nach vorheriger Terminvereinbarung gerne in meinem Büro T1.3.29 empfangen.

Für das Labor gilt Anwesenheitspflicht von der Sie in begründeten Fällen und in Absprache mit dem Lehrteam entbunden werden können.

Verhaltensregeln für Studierende

[Link zu den Verhaltensregeln für Online-Lehre](#)

Selbstverständnis als Lehrende/r

Die Lehrveranstaltung soll zur selbständigen Erarbeitung und Vertiefung des Lehrstoffs motivieren. Hierzu ist eine Kombination aus theoretischen und umsetzungsorientierten Teilen der Lehrveranstaltung geplant. Durch aktive Beteiligung der Studierenden gewinnt die Lehrveranstaltung weiter an Anschaulichkeit und Praxisrelevanz.

Für kritische Anmerkungen zur Lehrveranstaltung und für Verbesserungsvorschläge habe ich jederzeit ein offenes Ohr.

Sonstige Informationen

Sie müssen zwingend im selben Semester sowohl an der Lehrveranstaltung BAE 33241 (Seminar/Lehrgespräch) als auch an der Lehrveranstaltung BAE3342 (Labor) teilnehmen.

Im E-Learning System finden Sie für beide Lehrveranstaltungen des Fokusfachs einen gemeinsamen Kurs.

Es ist geplant, die Vorkenntnisse der Teilnehmer mittels eines elektronischen Fragebogens vor dem Vorlesungsbeginn zu erheben. Ein Link zu diesem Fragebogen wird rechtzeitig im E-Learning System veröffentlicht.

Die Teilnahme an der o.g. Befragung und eine im E-Learning System durchzuführende verbindliche Anmeldung zum Labor bis zum Ende der ersten Vorlesungswoche ist zwingende Voraussetzung für die Teilnahme an der Lehrveranstaltung.

Lesen Sie regelmäßig die Ankündigungen des Lehrteams in den Foren der entsprechenden Moodle Kurse.

Sprache: Deutsch