

Syllabus  
**BAE3342 Praktische Übungen zu Data Engineering und Data Science**  
Prof. Dr.-Ing. Heiko Thimm  
Sommersemester 2024

<b>Niveau</b>	Bachelor	
<b>Credits</b>	3	
<b>SWS</b>	2	
<b>Workload</b>	90 Stunden	
<b>Voraussetzungen</b>	MNS1091 Mathematik 1 BAE1131 Einführung in die Informatik BAE1141 Projektseminar Programmierung BAE1090 Quantitative Methoden I BAE2080 Quantitative Methoden II BAE2430 Informationstechnologie	
<b>Uhrzeit</b>	s. LSF	
<b>Raum</b>	s. LSF	
<b>Starttermin</b>	s. LSF	
<b>Lehrende(r)</b>	<b>Name</b>	Prof. Dr.-Ing. Heiko Thimm mit Unterstützung durch die wissenschaftliche Mitarbeiterin Frau Vanessa Schmidt
	<b>Büro</b>	T1.3.29 (Prof. Thimm) T1.3.02 (Frau Schmidt)
	<b>Virtuelles Büro</b>	<a href="#">Virtueller Hörsaal Prof. Thimm</a>
	<b>Kolloquium</b>	Montags, 13:45-15:15 Uhr bzw. nach Vereinbarung
	<b>Telefon</b>	07231 28-6451
	<b>Email</b>	<a href="mailto:heiko.thimm@hs-pforzheim.de">heiko.thimm@hs-pforzheim.de</a> <a href="mailto:vanessa.schmidt@hs-pforzheim.de">vanessa.schmidt@hs-pforzheim.de</a> (bevorzugte Kommunikationsform)

## Kurzbeschreibung

In dieser Lehrveranstaltung werden ausgewählte Inhalte der Lehrveranstaltung BAE3341-Data Engineering und Data Science anhand von Systemdemos und praktischen Übungen am Rechner veranschaulicht und vertieft. In den Übungen erlernen die Studierenden ebenfalls erste praktische Kenntnisse in der Nutzung gängiger Tools und der Anwendungen bewährter Vorgehensweisen und Methoden für typische Aufgaben des Data Engineerings und Data Science. Dabei ist auch die angeleitete Entwicklung von Python Programmen für Data Science Problemstellungen Gegenstand der Lehrveranstaltung. Auch ausgewählte Methoden und Modelle des Machine Learning sollen dabei zum Einsatz kommen und in Aufgaben mit Übungsdatensammlungen angewendet werden.

Die Übungen zur Durchführung explorativer Datenanalysen werden mit dem lizenzkostenfreien Tool Orange durchgeführt (<https://orangedatamining.com/>). Für die Programmierübungen und Projektarbeit wird Google Colab verwendet (<https://colab.research.google.com/>). Die Übungen/Projektarbeit sollte möglichst auf einem eigenen Rechner (PC, Laptop, Notebook) durchgeführt werden.

Es wird von den Studierenden ein hohes Maß an Eigenmotivation zum Erlernen entsprechender Programmierkenntnisse im Selbststudium erwartet.

Die Übungsaufgaben sind von den TeilnehmerInnen eigenständig auf Basis vorgegebener Literaturhinweise und sonstiger Lernmaterialien vorzubereiten. Anhand von regelmäßigen notenrelevanten schriftlichen und/oder mündlichen Wissenstests wird der aktuelle Kenntnisstand der TeilnehmerInnen ermittelt.

Abhängig von den Vorkenntnissen der TeilnehmerInnen können anstatt der Übungsaufgaben in Absprache mit dem Lehrteam ggf. auch andere Projektaufgaben durchgeführt werden.

## Gliederung der Veranstaltung (vorläufig)

- Übungsaufgabe(n) Datenmanagementsysteme und Datenbanken
- Übungsaufgabe(n) zu den CRISP-DM Phasen Datenverständnis und Datenvorbereitung
- Übungsaufgabe(n) zu den CRISP-DM Phasen Modellierung und Evaluation
- Übungsaufgaben zu ausgewählten Methoden und Modelle des Data Science einschließlich Machine Learning

## 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, ...
<b>1 Fachwissen</b>	
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 Prinzi-

	<p>pien 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.</p>
1.8	<p>...ein fundiertes Expertenwissen in ihrer Spezialisierung nachzuweisen.</p> <p>(Studiengang WI/ID) ... auf Data Science Anwendungen beruhende Innovationsideen zu entwickeln und entsprechende Anforderungsanalysen und Projektskizzen systematisch zu beschreiben.</p>
<b>2 Digitale Kompetenzen</b>	
2.1	<p>...relevante, in der betrieblichen Praxis eingesetzte IT-Softwaretools und deren Funktionen zu kennen und zu verstehen und verfügen über ein Grundverständnis für digitale Technologien.</p> <p>... für die Durchführung von Data Engineering und Data Science Projekten zur Verfügung stehende Softwaretools und Programmiersprachen zu benennen und anzuwenden.</p>
2.2	<p>...die im betrieblichen Umfeld vorzufindenden Informationssysteme effektiv zur Problemlösung zu nutzen.</p> <p>...die in betrieblichen Informationssystemen vorliegenden Datenbeständen auf deren Nutzbarkeit für Data Science Projekte zu bewerten. Sie kennen die im Rahmen der Datenvorbereitung durchzuführenden Aufgabenstellungen.</p>
2.4	<p>...im professionellen Umfeld digitale Technologien verantwortungsbewusst einzusetzen.</p> <p>... die Datenqualität von Datensammlungen zu bewerten und notwendige Aufgabenstellungen zur Qualitätsverbesserung und zur Datenvorbereitung gemäß des CRISP-DM Vorgehensmodells systematisch durchzuführen.</p>
<b>3 Kritisches Denken und analytische Fähigkeiten</b>	
3.1	<p>...geeignete Methoden kompetent zu verwenden und auf komplexe Fragestellungen anzuwenden.</p> <p>... 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.</p>
3.2	<p>...Ergebnisse umfassend zu interpretieren, kritisch zu reflektieren und eigene ganzheitliche Lösungsalternativen für komplexe Fragestellungen zu erarbeiten.</p> <p>... Ergebnisse von Data Science Anwendungen zielgruppenspezifisch in geeigneten Grafiken darzustellen und die Aussagefähigkeit der Ergebnisse kritisch zu hinterfragen.</p>
<b>4 Ethisches Bewusstsein und Nachhaltigkeit</b>	
	<p>...fundierte Lösungsstrategien in den Bereichen Ethik, nachhaltige Entwicklung und gesellschaftliche Verantwortung zu entwickeln und auf typische wirtschaftliche Entscheidungsprobleme anzuwenden.</p> <p>... 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.</p>
<b>5 Kommunikations- und Teamfähigkeit</b>	
5.1	<p>...komplexe Sachverhalte in klarer schriftlicher Form auszudrücken.</p> <p>... Aufgabenstellungen in Arbeitsblättern zu bearbeiten</p>

## Lehr- und Lernkonzept

Voraussetzung für eine erfolgreiche Teilnahme an dieser Lehrveranstaltung ist die Kenntnis der zugrundeliegenden Theorie. Diese wird in der Lehrveranstaltung BAE 3341, dem zugehörigen Vorlesungsskript, den Selbstlerneinheiten und den Unterlagen zu den Übungsaufgaben behandelt. Die Theorie sollte den Studierenden vor der Durchführung der jeweiligen Laborübung bekannt sein. Durch termingerechte schriftliche und/oder mündliche Wissenstests wird der Kenntnisstand der Teilnehmer erhoben.

In den Laborübungen sind typische praktische Data Engineering und Data Science Aufgabenstellungen zu bearbeiten. Die Übungsaufgaben bauen teilweise aufeinander auf, weshalb eine kontinuierliche Bearbeitung der Laboraufgaben eine unabdingbare Voraussetzung für den Lernerfolg darstellt. Zur Lösung der Übungsaufgaben müssen verschiedene Softwaretools eingesetzt werden. Es müssen ebenfalls eigene Programme in der Programmiersprache Python entwickelt werden. Hierzu werden teilweise Anleitungen zur Verfügung gestellt.

Zu jeder Übungsaufgabe ist eine kurze schriftliche Übungsdokumentation nach den Vorgaben des Lehrteams zu erstellen. Die Dokumentation ist fristgerecht zur Bewertung einzureichen. Von den Teilnehmern sind außerdem Präsentationen mit anschließender Diskussion/Fragerunde durchzuführen, die bewertet werden („Präsentationsrunden“ nach Abschluss ausgewählter Übungsaufgaben). In den Präsentationen sind der Lösungsweg (ggf. auch mögliche Alternativen) im Rahmen von live Demonstrationen und relevante Auszüge aus der Übungsdokumentation vorzustellen.

Die Lehrenden stehen als Gesprächspartner zur Verfügung und geben Unterstützung und Ratschläge. Es ist jedoch den Lehrenden ein großes Anliegen, dass die Übungsaufgaben selbstständig gelöst werden und dabei nur die zulässigen Informationsquellen zum Einsatz kommen.

Die Kommunikation erfolgt im persönlichen Gespräch, in Video Konferenzen und über das Moodle Forum im Kurs.

## **Literatur und Kursmaterialien**

Die im Lehrgespräch (Lehrveranstaltung BAE 3341) behandelten Folien werden im E-Learning 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