



Modul	Data Science II, Datenvisualisierung und Überwachtes Lernen Data Science II, Data Visualisation and Supervised Learning
Modulnummer	PTI01821 Version: 1
Fakultät	Physikalische Technik / Informatik
Niveau	Bachelor
Dauer	1 Semester
Turnus	Sommersemester
Modulverantwortliche/-r	Prof. Dr. Jens Flemming Jens.Flemming@fh-zwickau.de
Dozent/-in(nen)	Prof. Dr. Mike Espig mike.espig@fh-zwickau.de Dozent/-in in: "Data Science II, Datenvisualisierung und Überwachtes Lernen" Prof. Dr. Jens Flemming Jens.Flemming@fh-zwickau.de Dozent/-in in: "Data Science II, Datenvisualisierung und Überwachtes Lernen"
Lehrsprache(n)	Deutsch - 80.00% in "Data Science II, Datenvisualisierung und Überwachtes Lernen" Englisch - 20.00% in "Data Science II, Datenvisualisierung und Überwachtes Lernen"
ECTS-Credits	10.00 Credits
Workload	300 Stunden
Lehrveranstaltungen	8.00 SWS (2.00 SWS Praktikum 6.00 SWS Vorlesung mit integr. Übung / seminaristische Vorlesung)
Selbststudienzeit	180.00 Stunden 180.00 Stunden Selbststudium - Data Science II, Datenvisualisierung und Überwachtes Lernen
Prüfungsvorleistung(en)	Softwareprojekt in "Data Science II, Datenvisualisierung und Überwachtes Lernen"
Prüfungsleistung(en)	mündliche Prüfungsleistung Modulprüfung Prüfungsdauer: 30 min Wichtigkeit: 100% in "Data Science II, Datenvisualisierung und Überwachtes Lernen"
Medienform	Keine Angabe

Lehrinhalte/Gliederung	<p>Im Modul werden verschiedene Methoden der Datenvisualisierung eingeführt sowie die Grundlagen des überwachten maschinellen Lernens behandelt. Ziel des überwachten Lernens ist die Entwicklung von Algorithmen, die anhand umfangreicher Beispieldatensätze automatisch Zuordnungen zwischen Ein- und Ausgabegrößen erlernen. Eingabegrößen können beispielsweise Bilder sein, Ausgabegrößen textuelle Beschreibungen der Bildinhalte.</p> <p>Zwei prominente Verfahrensklassen des überwachten maschinellen Lernens, lineare Regression und künstliche neuronale Netze, werden detailliert vorgestellt und an praxisnahen Aufgabenstellungen erprobt und untersucht.</p> <p>Datenvisualisierung:</p> <ul style="list-style-type: none"> • Einführung in das Python-Paket Matplotlib • Erstellen zwei- und dreidimensionaler Visualisierungen • Vorstellung weiterer Python-Pakete zur Datenvisualisierung je nach aktuellem Entwicklungsstand auf dem Gebiet, insbesondere Seaborn • Visualisierung von Geodaten • Erstellen animierter Visualisierungen <p>Grundlagen des überwachten Lernens:</p> <ul style="list-style-type: none"> • Abgrenzung Regression - Klassifikation • Hypothesen und Loss-Funktionen • Trainings-, Validierungs-, Testphase • Kreuzvalidierung • Hyperparameterwahl mittels Grid-Search • Bewerten der Vorhersagequalität trainierter Modelle • einfache Methoden zur Reduktion der Feature-Anzahl, insbesondere Hauptkomponentenanalyse (PCA) • Umsetzung mit dem Python-Paket Scikit-Learn <p>Lineare Regression:</p> <ul style="list-style-type: none"> • Grundlagen • Overfitting und Regularisierung • Umsetzung mit dem Python-Paket Scikit-Learn <p>Künstliche neuronale Netze:</p> <ul style="list-style-type: none"> • Grundlagen und Motivation • vorwärts gerichtete geschichtete neuronale Netze • Training mittels Gradientenverfahren • Overfitting und Regularisierung, insbesondere Regularisierung mittels Dropout-Schichten • faltende neuronale Netze (CNNs) • Klassifikation von Bilddaten • Trainieren und Anwenden neuronaler Netze mit den Python-Paketen Tensorflow und Keras
Qualifikationsziele	<p>Die Studierenden besitzen fundierte Kenntnisse zur Software-Bibliothek Matplotlib und zu verwandten Bibliotheken zur Datenvisualisierung und können diese praktisch anwenden.</p> <p>Die Studierenden beherrschen die theoretischen Grundlagen des überwachten maschinellen Lernens sowohl für Regressions- als auch für Klassifikationsprobleme. Sie können lineare Regressionen praktisch durchführen und deren Anwendungsgebiete und Grenzen einschätzen.</p> <p>Auf künstlichen neuronalen Netzen basierende Techniken des überwachten Lernens sind den Studierenden theoretisch wie praktisch vertraut. Sie können vorwärts gerichtete geschichtete und insbesondere faltende neuronale Netze auf Lernprobleme anwenden, deren Möglichkeiten und Grenzen bewerten, sowie diese effizient mit den Bibliotheken Tensorflow und Keras implementieren.</p> <p>Die Studierenden sind in der Lage, ihr Wissen durch eigenständige Arbeit mit Fachliteratur zu erweitern und sie besitzen ein geschärftes Bewusstsein für rechtliche Aspekte beim Sammeln und Verarbeiten von Daten.</p>
Sozial- und Selbstkompetenzen	Keine Angabe
Besondere Zulassungsvoraussetzung	keine

Empfohlene Voraussetzungen	<ul style="list-style-type: none"> • PTI01810 - Data Science I, Einführung in Python und Datenvorverarbeitung • PTI01710 - Mathematische Grundlagen I
Fortsetzungsmöglichkeiten	PTI01831, PTI01841, PTI01850, PTI01860, PTI01870
Literatur	<ul style="list-style-type: none"> • Vorlesungsskript • Géron: Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow, O'Reilly Media • Chollet: Deep Learning with Python, Manning • Mohri, Rostamizadeh, Talwalkar: Foundations of Machine Learning, MIT Press • Shalev-Shwartz, Ben-David: Understanding Machine Learning From Theory to Algorithms, Cambridge University Press • Bishop: Pattern Recognition and Machine Learning, Springer • Vapnik: Statistical Learning Theory, Springer • Vapnik: The Nature of Statistical Learning Theory, Springer • Györfi, Kohler, Krzyzak, Walk: A Distribution-Free Theory of Nonparametric Regression, Springer • Cucker, Zhou: Learning Theory An Approximation Theory Viewpoint, Cambridge Monographs • Barber: Bayesian Reasoning and Machine Learning, Cambridge University Press • Murphy: Machine Learning A Probabilistic Perspective, MIT Press
Hinweise	Die Modulprüfung kann in deutscher und englischer Sprache abgelegt werden.
Link zu Kurs/Lernressourcen im OPAL	
Zuordnung zum Curriculum	250 Data Science - Bachelor 2022 Vollzeit 250 Data Science - Bachelor 2022 Teilzeit