

<b>Modulbezeichnung: Statistisches Data Mining</b>																		
<b>Modulkürzel</b>	t.BA.WI.STDM.19HS																	
<b>ECTS Credits</b>	4																	
<b>Unterrichts- und Prüfungssprache</b>	Deutsch																	
<b>Verantwortliche OE</b>	IDP																	
<b>Modulverantwortung</b>	Martin Frey																	
<b>Rechtliche Grundlagen</b>	Die Modulbeschreibung ist neben Rahmenprüfungsordnung und Studienordnung Teil der Rechtsgrundlage. Sie ist verbindlich. Eine in der ersten Unterrichtswoche des Semesters schriftlich festgehaltene und kommunizierte Modulvereinbarung kann die Modulbeschreibung präzisieren. Die Modulvereinbarung ersetzt nicht die Modulbeschreibung.																	
<b>Modulprägung</b>	Typ 2a  4 Lektionen Vorlesung aufeinanderfolgend pro Semesterwoche und Klasse																	
<b>Beschreibung des Moduls</b>	Das Modul führt in die Grundprinzipien des Statistischen Data Mining/Machinelles Lernen ein. Es werden Methoden aus dem überwachten und unüberwachten Lernen behandelt und an konkreten Fallbeispielen angewendet.																	
<b>Inhalte des Moduls</b>	<ul style="list-style-type: none"> <li>Unüberwachtes Lernen: Ähnlichkeits- und Distanzmasse, Ausreisserdetektion. Eine Auswahl bekannter Methoden zur Datenreduktion, wie Principal Component Analysis (PCA), Multidimensional scaling, t-SNE und UMAP. Eine Auswahl bekannter und moderner Clustering-Methoden, wie K-means Clustering, hierarchical Clustering, dichte- und modellbasiertes Clustering. Überwachtes Lernen: Grundlagen, Modellselektierung, Cross-Validation Auswertung und Performance Evaluation von Klassifikatoren. Variable Importance. Eine Auswahl bekannter und moderner Methoden, wie Bayes Classifier, Nearest Neighbor Classifier, Support Vector Machines, Logistic Regression, Decision Trees und Random Forest. Ensemble Methoden (Bagging und Boosting).</li> </ul>																	
<b>Vorkenntnisse</b>	ExpD, Wahr und GStat																	
<b>Lernziele (Kompetenzen)</b>	<table border="1"> <thead> <tr> <th><b>Die Studierenden...</b></th> <th><b>Kompetenzen</b></th> <th><b>Taxonomiestufen</b></th> </tr> </thead> <tbody> <tr> <td>Sie kennen die Grundlagen des Data-Mining Processes.</td> <td>M, F</td> <td>K1, K2, K3</td> </tr> <tr> <td>Sie kennen Verfahren aus dem unüberwachten Lernens (unsupervised learning) und Ihre wichtigsten Eigenschaften. Sie können erkennen, für welche Probleme sie geeignet sind und können diese anwenden, um neue Probleme zu lösen.</td> <td>F</td> <td>K1, K2</td> </tr> <tr> <td>Sie kennen Verfahren aus dem überwachten Lernens (supervised learning) und Ihre wichtigsten Eigenschaften. Sie können erkennen, für welche Probleme sie geeignet sind und können diese anwenden, um neue Probleme zu lösen.</td> <td>F, M</td> <td>K4, K5</td> </tr> <tr> <td>Sie können Berechnungen für konkrete Fragestellungen in einer Programmiersprache (R oder Python) durchführen und interpretieren.</td> <td>M, F</td> <td>K1, K2, K3</td> </tr> </tbody> </table>			<b>Die Studierenden...</b>	<b>Kompetenzen</b>	<b>Taxonomiestufen</b>	Sie kennen die Grundlagen des Data-Mining Processes.	M, F	K1, K2, K3	Sie kennen Verfahren aus dem unüberwachten Lernens (unsupervised learning) und Ihre wichtigsten Eigenschaften. Sie können erkennen, für welche Probleme sie geeignet sind und können diese anwenden, um neue Probleme zu lösen.	F	K1, K2	Sie kennen Verfahren aus dem überwachten Lernens (supervised learning) und Ihre wichtigsten Eigenschaften. Sie können erkennen, für welche Probleme sie geeignet sind und können diese anwenden, um neue Probleme zu lösen.	F, M	K4, K5	Sie können Berechnungen für konkrete Fragestellungen in einer Programmiersprache (R oder Python) durchführen und interpretieren.	M, F	K1, K2, K3
<b>Die Studierenden...</b>	<b>Kompetenzen</b>	<b>Taxonomiestufen</b>																
Sie kennen die Grundlagen des Data-Mining Processes.	M, F	K1, K2, K3																
Sie kennen Verfahren aus dem unüberwachten Lernens (unsupervised learning) und Ihre wichtigsten Eigenschaften. Sie können erkennen, für welche Probleme sie geeignet sind und können diese anwenden, um neue Probleme zu lösen.	F	K1, K2																
Sie kennen Verfahren aus dem überwachten Lernens (supervised learning) und Ihre wichtigsten Eigenschaften. Sie können erkennen, für welche Probleme sie geeignet sind und können diese anwenden, um neue Probleme zu lösen.	F, M	K4, K5																
Sie können Berechnungen für konkrete Fragestellungen in einer Programmiersprache (R oder Python) durchführen und interpretieren.	M, F	K1, K2, K3																

## Modulbezeichnung: Statistisches Data Mining

<b>Leistungsnachweis</b>	<b>Modulendprüfung</b>	<b>Bewertung</b>	<b>Dauer (Min.)</b>	<b>Gewichtung</b>	<b>Form</b>	
	schriftliche Prüfung	Note	90	60	gem. Modulvereinbarung	
	<b>Leistungsnachweise während dem Semester</b>					
	Präsentation <i>Zusammenfassung einer vorgestellten vorgestellten Methode inklusive Übungsaufgabe.</i>		Note		8	
	Bericht		Note		8	
	Bericht		Note		8	
	Bericht		Note		8	
Vorhersage Challenge <i>Erstellen eines Vorhersage-Modell</i>		Note		8		
<b>Präsenzverpflichtung im Kontaktstudium</b>	Keine  In Absprache mit dem Dozenten/der Dozentin. Präsentation findet vor Ort statt.					
<b>Lernmaterialien</b>	<ul style="list-style-type: none"> <li>• An introduction to statistical learning : with applications in R; James, Gareth; Witten, Daniela; Hastie, Trevor; Tibshirani, Robert; Boston: Springer Second edition; 2022</li> <li>• An introduction to statistical learning with applications in Python; James, Gareth; Witten, Daniela; Hastie, Trevor; Tibshirani, Robert; Taylor Jonathan; Cham, Switzerland Springer 2023</li> </ul>					
<b>Bemerkungen</b>	Die genauen Anforderungen für die Leistungsnachweises werden von den Dozenten und Dozentinnen zu Beginn der Vorlesung schriftlich kommuniziert.					