Valid from 2024.HS

Module description: Statistical Data Mining

Module Code	t.BA.WI.STDM.19HS
ECTS Credits	4
Language of Instruction/Examination	German
Organizational Unit	IDP
Module Coordinator	Martin Frey
Legal Framework	The module description is part of the legal basis in addition to the general academic regulations. It is binding. During the first week of the semester a written and communicated supplement can specify the module description in more detail.
Module Characteristic	Type 2a 4 consecutive lecture lessons per semester week and class
Module Description	The module introduces the basic principles of statistical data mining/machine learning. Methods from supervised and unsupervised learning are covered and applied to specific case studies.
Module Content	• Unsupervised Learning Similarity- and distance measurements, outlier detection. A selection of typical methods for data reduction, such as Principal Component Analysis (PCA), Multidimensional Scaling, t-SNE, and UMAP. A selection of well-known and modern clustering methods, such as k-Means Clustering, hierarchical Clustering, density-based, and model-based Clustering. Supervised Learning: Basics, model selection, cross-validation Evaluation and performance evaluation of classifiers. Variable importance. A selection of well-known and modern methods, such as Bayes Classifier, Nearest Neighbor Classifier, Support Vector Machines, Logistic Regression, Decision Trees, and Random Forest. Ensemble methods (Bagging and Boosting).
Prerequisite Knowledge	ExpD, Wahr and GStat

Learning Objectives (Competences)

Students...	Competencies	Taxonomies
You know the basics of the data mining process.	M, F	K1, K2, K3
You are familiar with unsupervised learning methods and their most important characteristics. You can recognize which problems they are suitable for and can apply them to solve new problems.	F	K1, K2
You are familiar with supervised learning methods and their most importantcharacteristics. You can recognize which problems they are suitable for and can apply them to solve new problems.	F, M	K4, K5
You can implement and interpret data mining methods for specific tasks in a programming language (R or Python).	M, F	K1, K2, K3

Module description: Statistical Data Mining

Performance Assessment

End-of-module exam	Assessment	Length (min.)	Weighting	Form
written exam	Grade	90	60	acc. to module agreement

Performance assessment during the semester	Assessment	Length (min.)	Weighting	Form
Presentation *Summary of a presented method including an exercise.*	Grade		8	
report	Grade		8	
report	Grade		8	
report	Grade		8	
Prediction Challenge *Creating a prediction model*	Grade		8	

Classroom Attendance Requirement

None

In consultation with the lecturer. Presentation takes place on site.

Learning material

- An introduction to statistical learning : with applications in R; James, Gareth; Witten, Daniela; Hastie, Trevor; Tibshirani, Robert; Boston: Springer Second edition; 2022
- An introduction to statistical learning with applications in Python; James, Gareth; Witten, Daniela; Hastie, Trevor; Tibshirani, Robert; Taylor Jonathan; Cham, Switzerland Springer 2023

Comments

The exact requirements for the semester tasks are communicated in writing by the lecturers at the beginning of the lecture.