

Data Mining II - Advanced Topics in Data Mining

(in German: Data Mining II - Advanced Topics in Data Mining)

Module-ID: FIN-INF-120455

Link:	LSF
Responsibility:	Myra Spiliopoulou
Lecturer:	Myra Spiliopoulou
Classes:	Vorlesung DM 2 Übung DM 2 COURTESY TRANSLATION: • Lecture class DM2 • Exercise class DM2
Applicability in curriculum:	- M.Sc. INF: Informatik - M.Sc. INGINF: Informatik - M.Sc. WIF: Informatik - M.Sc. DKE: Learning Methods and Models for Data Science - M.Sc. DE: Fachliche Spezialisierung - M.Sc. VC: Computer Science

Abbreviation	Credit Points	Semester	Term	Duration	Language	Level
DM2	6	Winter	ab 1.	1 Semester	english	Master

Intended learning outcomes:

When successfully completing this module, the students:

- comprehend why temporal data need different learning algorithms and evaluation procedures than used on static data
- comprehend the behaviour of supervised, unsupervised and semi-supervised learning algorithms on temporal data

- can design and apply simple learning algorithms and workflows on temporal data and interpret the induced models
- can evaluate models - once and in continuous evaluation, since both are needed in temporal learning

and have thus acquired skills they need in order to design and evaluate temporal learning algorithms themselves

Content:

Block 1: Data Streams

- Basics
- Stream clustering: methods and evaluation approaches
- Stream classification: learning methods and concept drift detectors; evaluation approaches
- Semi-supervised stream learning: emthods and evaluation approaches

Block 2: Time series

- Basics
- Prediction methods
- Evaluation of predictors

Workload:

56 h Präsenz + 124 h selbstständige Arbeit **COURTESY TRANSLATION:** 56 h in class + 124 h self-study

Pre-examination requirements:	Type of examination:	Teaching method / lecture hours per week (SWS):
	• Klausur 120 Minuten • Prüfungszulassungsvoraussetzung: Erreichen einer minimalen Anzahl von Punkten durch Votierung auf Übungsaufgaben COURTESY TRANSLATION: • Written exam of the form 'Klausur' with a duration of 120 min • Prerequisite for the written exam: a minimum number of points must be achieved during the exercise classes; this procedure is called 'Votierung'	• Vorlesung (2 SWS) • Übung (2 SWS) COURTESY TRANSLATION: • Lecture (2 hours per week of the semester) • Exercise (2 hours per week of the semester)

Prerequisites according to examination regulations:	Recommended prerequisites:
keine	• Familiarity with learning algorithms for static tabular data • Familiarity with methods for the evaluation of models induced on static data

Media:	Literature:
---------------	--------------------

Block 1:

- Gama, J. (2010) *Knowledge discovery from data streams*. CRC Press.
- Silva, J. A., Faria, E. R., Barros, R. C., Hruschka, E. R., Carvalho, A. C. D., & Gama, J. (2013). Data stream clustering: A survey. *ACM Computing Surveys (CSUR)*, 46(1), 1-31.
- Gama, J., Žliobaitė, I., Bifet, A., Pechenizkiy, M., & Bouchachia, A. (2014). A survey on concept drift adaptation. *ACM computing surveys (CSUR)*, 46(4), 1-37.
- Lemaire, V., Salperwyck, C., & Bondu, A. (2014, July). A survey on supervised classification on data streams. In *European Business Intelligence Summer School* (pp. 88-125). Springer, Cham.
- Žliobaitė, I., Bifet, A., Read, J., Pfahringer, B., & Holmes, G. (2015). Evaluation methods and decision theory for classification of streaming data with temporal dependence. *Machine Learning*, 98(3), 455-482.
- Murilo Gomes, H., Grzenda, M., Mello, R., Read, J., Nguyen, M. H. L., and Bifet, A. (2022). A survey on semi-supervised learning for delayed partially labelled data streams. *ACM Computing Surveys*, 55(4). - **for Block 1D**

Block 2:

- Parmezan, A. R. S., Souza, V. M., & Batista, G. E. (2019). Evaluation of statistical and machine learning models for time series prediction: Identifying the state-of-the-art and the best conditions for the use of each model. *Information sciences*, 484, 302-337.
- Cismondi, F. C., Fialho, A., Vieira, S., Reti, S., Sousa, J., Finkelstein, S. (2013). Missing data in medical databases: Impute, delete or classify? *Artificial Intelligence in Medicine*.
<http://dx.doi.org/10.1016/j.artmed.2013.01.003>.
- Further citations at the individual units: see the slidesets of the units.

Comments:

Revision #1

Created 2025-10-09 12:59:12 UTC by Fakultät für Informatik

Updated 2025-10-09 12:59:12 UTC by Fakultät für Informatik