

Data Mining

Lobis, Ulrich; ulrich.lobis@uibk.ac.at / Wang-Kathrein, Joseph; joseph.wang@uibk.ac.at

Data Mining kann als "process of discovering interesting patterns and knowledge from large amounts of data [...] [where] data sources can include databases, data warehouses, the Web, other information repositories, or data that are streamed into the system dynamically"(Han/Kamber 2005, Kap. 1.10., S. 39) betrachtet werden. Es handelt sich um einen Sammelbegriff, der eng mit *Machine Learning* und mit *Data Science* verknüpft ist. Üblicherweise wird *Data Mining* auf bestehende Daten angewandt. Für die DH als Teil der bevorzugt textorientierten Geisteswissenschaften sind vor allem *Text Mining* und *Web Mining* von großem Interesse.

Für die Anwendung von *Data Mining* stehen zahlreiche spezialisierte, kommerzielle und freie Softwarelösungen zur Verfügung, wobei in erster Linie auf *Java*, *R* und *Python* zu verweisen ist.

Literatur:

- Han, Jiawei; Kamber, Micheline; Pai, Jian: Data Mining: Concepts and Techniques. Amsterdam [u.a.]: 2005.

Software:

Natural Language Toolkit (nltk), scikit-learn, RapidMiner Studio, Weka, R

Themen:

Datenanalyse

Zitiervorschlag:

Lobis, Ulrich; Wang-Kathrein, Joseph. 2021. Data Mining. In: KONDE Weißbuch. Hrsg. v. Helmut W. Klug unter Mitarbeit von Selina Galka und Elisabeth Steiner im HRSM Projekt "Kompetenznetzwerk Digitale Edition". URL: <https://gams.uni-graz.at/o:konde.48>