**Data mining** is the process of discovering patterns in large [data sets](https://en.wikipedia.org/wiki/Data_set) involving methods at the intersection of [machine learning](https://en.wikipedia.org/wiki/Machine_learning), [statistics](https://en.wikipedia.org/wiki/Statistics), and [database systems](https://en.wikipedia.org/wiki/Database_system). Data mining is an [interdisciplinary](https://en.wikipedia.org/wiki/Interdisciplinary) subfield of [computer science](https://en.wikipedia.org/wiki/Computer_science) and [statistics](https://en.wikipedia.org/wiki/Statistics) with an overall goal to extract information (with intelligent methods) from a data set and transform the information into a comprehensible structure for further use. Data mining is the analysis step of the "knowledge discovery in databases" process or KDD. Aside from the raw analysis step, it also involves database and [data management](https://en.wikipedia.org/wiki/Data_management) aspects, [data pre-processing](https://en.wikipedia.org/wiki/Data_pre-processing), [model](https://en.wikipedia.org/wiki/Statistical_model) and [inference](https://en.wikipedia.org/wiki/Statistical_inference) considerations, interestingness metrics, [complexity](https://en.wikipedia.org/wiki/Computational_complexity_theory) considerations, post-processing of discovered structures, [visualization](https://en.wikipedia.org/wiki/Data_visualization), and [online updating](https://en.wikipedia.org/wiki/Online_algorithm).

The term "data mining" is a [misnomer](https://en.wikipedia.org/wiki/Misnomer), because the goal is the extraction of patterns and knowledge from large amounts of data, not the extraction (*mining*) of data itself. It also is a [buzzword](https://en.wikipedia.org/wiki/Buzzword) and is frequently applied to any form of large-scale data or [information processing](https://en.wikipedia.org/wiki/Information_processing) ([collection](https://en.wikipedia.org/wiki/Data_collection), [extraction](https://en.wikipedia.org/wiki/Information_extraction), [warehousing](https://en.wikipedia.org/wiki/Data_warehouse), analysis, and statistics) as well as any application of [computer decision support system](https://en.wikipedia.org/wiki/Decision_support_system), including [artificial intelligence](https://en.wikipedia.org/wiki/Artificial_intelligence) (e.g., machine learning) and [business intelligence](https://en.wikipedia.org/wiki/Business_intelligence). The book *Data mining: Practical machine learning tools and techniques with Java* (which covers mostly machine learning material) was originally to be named just *Practical machine learning*, and the term *data mining* was only added for marketing reasons.[[9]](https://en.wikipedia.org/wiki/Data_mining#cite_note-9) Often the more general terms (*large scale*) [*data analysis*](https://en.wikipedia.org/wiki/Data_analysis) and [*analytics*](https://en.wikipedia.org/wiki/Analytics) – or, when referring to actual methods, *artificial intelligence* and *machine learning* – are more appropriate.