

# Data mining

---

**Data mining** (the analysis step of the knowledge discovery in databases process,<sup>[1]</sup> or KDD), a relatively young and interdisciplinary field of computer science<sup>[2][3]</sup> is the process of discovering new patterns from large data sets involving methods at the intersection of artificial intelligence, machine learning, statistics and database systems.<sup>[2]</sup> The goal of data mining is to extract knowledge from a data set in a human-understandable structure<sup>[2]</sup> and involves database and data management, data preprocessing, model and inference considerations, interestingness metrics, complexity considerations, post-processing of found structure, visualization and online updating.<sup>[2]</sup>

The term is a buzzword, and is frequently misused to mean any form of large-scale data or information processing (collection, extraction, warehousing, analysis and statistics) but also generalized to any kind of computer decision support system including artificial intelligence, machine learning and business intelligence. In the proper use of the word, the key term is *discovery*, commonly defined as "detecting something new". Even the popular book "Data mining: Practical machine learning tools and techniques with Java"<sup>[4]</sup> (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.<sup>[5]</sup> Often the more general terms "(large scale) data analysis" or "analytics" or when referring to actual methods, artificial intelligence and machine learning are more appropriate.

The actual data mining task is the automatic or semi-automatic analysis of large quantities of data to extract previously unknown interesting patterns such as groups of data records (cluster analysis), unusual records (anomaly detection) and dependencies (association rule mining). This usually involves using database techniques such as spatial indexes. These patterns can then be seen as a kind of summary of the input data, and used in further analysis or for example in machine learning and predictive analytics. For example, the data mining step might identify multiple groups in the data, which can then be used to obtain more accurate prediction results by a decision support system. Neither the data collection, data preparation nor result interpretation and reporting are part of the data mining step, but do belong to the overall KDD process as additional steps.

The related terms *data dredging*, *data fishing* and *data snooping* refer to the use of data mining methods to sample parts of a larger population data set that are (or may be) too small for reliable statistical inferences to be made about the validity of any patterns discovered. These methods can, however, be used in creating new hypotheses to test against the larger data populations.

## Background

The manual extraction of patterns from data has occurred for centuries. Early methods of identifying patterns in data include Bayes' theorem (1700s) and regression analysis (1800s). The proliferation, ubiquity and increasing power of computer technology has increased data collection, storage and manipulations. As data sets have grown in size and complexity, direct hands-on data analysis has increasingly been augmented with indirect, automatic data processing. This has been aided by other discoveries in computer science, such as neural networks, cluster analysis, genetic algorithms (1950s), decision trees (1960s) and support vector machines (1990s). Data mining is the process of applying these methods to data with the intention of uncovering hidden patterns<sup>[6]</sup> in large data sets. It bridges the gap from applied statistics and artificial intelligence (which usually provide the mathematical background) to database management by exploiting the way data is stored and indexed in databases to execute the actual learning and discovery algorithms more efficiently, allowing such methods to be applied to larger data sets.

---

## Research and evolution

The premier professional body in the field is the Association for Computing Machinery's Special Interest Group on knowledge discovery and Data Mining (SIGKDD). Since 1989 they have hosted an annual international conference and published its proceedings,<sup>[7]</sup> and since 1999 have published a biannual academic journal titled "SIGKDD Explorations".<sup>[8]</sup>

Computer science conferences on data mining include:

- CIKM – ACM Conference on Information and Knowledge Management
- DMIN – International Conference on Data Mining
- DMKD – Research Issues on Data Mining and Knowledge Discovery
- ECDM – European Conference on Data Mining
- ECML-PKDD – European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases
- EDM – International Conference on Educational Data Mining
- ICDM – IEEE International Conference on Data Mining
- KDD – ACM SIGKDD Conference on Knowledge Discovery and Data Mining
- MLDM – Machine Learning and Data Mining in Pattern Recognition
- PAKDD – The annual Pacific-Asia Conference on Knowledge Discovery and Data Mining
- PAW – Predictive Analytics World
- SDM – SIAM International Conference on Data Mining (SIAM)
- SSTD – Symposium on Spatial and Temporal Databases

Data mining topics are present on most data management / database conferences.

## Process

The **knowledge discovery in databases (KDD) process** is commonly defined with the stages (1) Selection (2) Preprocessing (3) Transformation (4) *Data Mining* (5) Interpretation/Evaluation.<sup>[1]</sup> It exists however in many variations of this theme such as the CRoss Industry Standard Process for Data Mining (CRISP-DM) which defines six phases: (1) Business Understanding, (2) Data Understanding, (3) Data Preparation, (4) Modeling, (5) Evaluation, and (6) Deployment or a simplified process such as (1) Pre-processing, (2) Data mining, and (3) Results validation.

## Pre-processing

Before data mining algorithms can be used, a target data set must be assembled. As data mining can only uncover patterns actually present in the data, the target dataset must be large enough to contain these patterns while remaining concise enough to be mined in an acceptable timeframe. A common source for data is a data mart or data warehouse. Pre-processing is essential to analyze the multivariate datasets before data mining.

The target set is then cleaned. Data cleaning removes the observations containing noise and those with missing data.

Data mining involves six common classes of tasks:<sup>[1]</sup>

- Anomaly detection (Outlier/change/deviation detection) – The identification of unusual data records, that might be interesting or data errors and require further investigation.
- Association rule learning (Dependency modeling) – Searches for relationships between variables. For example a supermarket might gather data on customer purchasing habits. Using association rule learning, the supermarket can determine which products are frequently bought together and use this information for marketing purposes. This is sometimes referred to as market basket analysis.
- Clustering – is the task of discovering groups and structures in the data that are in some way or another "similar", without using known structures in the data.

- Classification – is the task of generalizing known structure to apply to new data. For example, an email program might attempt to classify an email as legitimate or spam.
- Regression – Attempts to find a function which models the data with the least error.
- Summarization – providing a more compact representation of the data set, including visualization and report generation.

## Results validation

The final step of knowledge discovery from data is to verify the patterns produced by the data mining algorithms occur in the wider data set. Not all patterns found by the data mining algorithms are necessarily valid. It is common for the data mining algorithms to find patterns in the training set which are not present in the general data set. This is called overfitting. To overcome this, the evaluation uses a test set of data on which the data mining algorithm was not trained. The learned patterns are applied to this test set and the resulting output is compared to the desired output. For example, a data mining algorithm trying to distinguish spam from legitimate emails would be trained on a training set of sample emails. Once trained, the learned patterns would be applied to the test set of emails on which it had not been trained. The accuracy of these patterns can then be measured from how many emails they correctly classify. A number of statistical methods may be used to evaluate the algorithm such as ROC curves.

If the learned patterns do not meet the desired standards, then it is necessary to reevaluate and change the pre-processing and data mining. If the learned patterns do meet the desired standards then the final step is to interpret the learned patterns and turn them into knowledge.

## Standards

There have been some efforts to define standards for the data mining process, for example the 1999 European Cross Industry Standard Process for Data Mining (CRISP-DM 1.0) and the 2004 Java Data Mining standard (JDM 1.0). Development on successors of these processes (CRISP-DM 2.0 and JDM 2.0) was active in 2006, but has stalled since. JDM 2.0 was withdrawn without reaching a final draft.

For exchanging the extracted models – in particular for the use in predictive analytics – the key standard is the Predictive Model Markup Language (PMML), which is an XML-based language developed by the Data Mining Group (DMG) and supported as exchange format by many data mining applications. As the name suggests it only covers prediction models, a particular data mining task of high importance to business applications, however extensions to for example cover subspace clustering have been proposed independently of the DMG.<sup>[9]</sup>

## Notable uses

### Games

Since the early 1960s, with the availability of oracles for certain combinatorial games, also called tablebases (e.g. for 3x3-chess) with any beginning configuration, small-board dots-and-boxes, small-board-hex, and certain endgames in chess, dots-and-boxes, and hex; a new area for data mining has been opened. This is the extraction of human-useable strategies from these oracles. Current pattern recognition approaches do not seem to fully acquire the high level of abstraction required to be applied successfully. Instead, extensive experimentation with the tablebases, combined with an intensive study of tablebase-answers to well designed problems and with knowledge of prior art, i.e. pre-tablebase knowledge, is used to yield insightful patterns. Berlekamp in dots-and-boxes etc. and John Nunn in chess endgames are notable examples of researchers doing this work, though they were not and are not involved in tablebase generation.

## Business

Data mining in customer relationship management applications can contribute significantly to the bottom line. Rather than randomly contacting a prospect or customer through a call center or sending mail, a company can concentrate its efforts on prospects that are predicted to have a high likelihood of responding to an offer. More sophisticated methods may be used to optimize resources across campaigns so that one may predict to which channel and to which offer an individual is most likely to respond—across all potential offers. Additionally, sophisticated applications could be used to automate the mailing. Once the results from data mining (potential prospect/customer and channel/offer) are determined, this "sophisticated application" can either automatically send an e-mail or regular mail. Finally, in cases where many people will take an action without an offer, uplift modeling can be used to determine which people will have the greatest increase in responding if given an offer. Data clustering can also be used to automatically discover the segments or groups within a customer data set.

Businesses employing data mining may see a return on investment, but also they recognize that the number of predictive models can quickly become very large. Rather than one model to predict how many customers will churn, a business could build a separate model for each region and customer type. Then instead of sending an offer to all people that are likely to churn, it may only want to send offers to loyal customers. Finally, it may want to determine which customers are going to be profitable over a window of time and only send the offers to those that are likely to be profitable. In order to maintain this quantity of models, they need to manage model versions and move to *automated data mining*.

Data mining can also be helpful to human-resources departments in identifying the characteristics of their most successful employees. Information obtained, such as universities attended by highly successful employees, can help HR focus recruiting efforts accordingly. Additionally, Strategic Enterprise Management applications help a company translate corporate-level goals, such as profit and margin share targets, into operational decisions, such as production plans and workforce levels.<sup>[10]</sup>

Another example of data mining, often called the market basket analysis, relates to its use in retail sales. If a clothing store records the purchases of customers, a data-mining system could identify those customers who favor silk shirts over cotton ones. Although some explanations of relationships may be difficult, taking advantage of it is easier. The example deals with association rules within transaction-based data. Not all data are transaction based and logical or inexact rules may also be present within a database.

Market basket analysis has also been used to identify the purchase patterns of the Alpha consumer. Alpha Consumers are people that play a key role in connecting with the concept behind a product, then adopting that product, and finally validating it for the rest of society. Analyzing the data collected on this type of user has allowed companies to predict future buying trends and forecast supply demands.

Data mining is a highly effective tool in the catalog marketing industry. Catalogers have a rich history of customer transactions on millions of customers dating back several years. Data mining tools can identify patterns among customers and help identify the most likely customers to respond to upcoming mailing campaigns.

Data mining for business applications is a component which needs to be integrated into a complex modelling and decision making process. Reactive business intelligence (RBI) advocates a holistic approach that integrates data mining, modeling and interactive visualization, into an end-to-end discovery and continuous innovation process powered by human and automated learning.<sup>[11]</sup> In the area of decision making the RBI approach has been used to mine the knowledge which is progressively acquired from the decision maker and self-tune the decision method accordingly.<sup>[12]</sup>

Related to an integrated-circuit production line, an example of data mining is described in the paper "Mining IC Test Data to Optimize VLSI Testing."<sup>[13]</sup> In this paper the application of data mining and decision analysis to the problem of die-level functional test is described. Experiments mentioned in this paper demonstrate the ability of applying a system of mining historical die-test data to create a probabilistic model of patterns of die failure. These patterns are then utilized to decide in real time which die to test next and when to stop testing. This system has been shown,

---

based on experiments with historical test data, to have the potential to improve profits on mature IC products.

## Science and engineering

In recent years, data mining has been used widely in the areas of science and engineering, such as bioinformatics, genetics, medicine, education and electrical power engineering.

In the study of human genetics, sequence mining helps address the important goal of understanding the mapping relationship between the inter-individual variation in human DNA sequences and variability in disease susceptibility. In lay terms, it is to find out how the changes in an individual's DNA sequence affect the risk of developing common diseases such as cancer. This is very important to help improve the diagnosis, prevention and treatment of the diseases. The data mining method that is used to perform this task is known as multifactor dimensionality reduction.<sup>[14]</sup>

In the area of electrical power engineering, data mining methods have been widely used for condition monitoring of high voltage electrical equipment. The purpose of condition monitoring is to obtain valuable information on the insulation's health status of the equipment. Data clustering such as self-organizing map (SOM) has been applied on the vibration monitoring and analysis of transformer on-load tap-changers (OLTCS). Using vibration monitoring, it can be observed that each tap change operation generates a signal that contains information about the condition of the tap changer contacts and the drive mechanisms. Obviously, different tap positions will generate different signals. However, there was considerable variability amongst normal condition signals for exactly the same tap position. SOM has been applied to detect abnormal conditions and to estimate the nature of the abnormalities.<sup>[15]</sup>

Data mining methods have also been applied for dissolved gas analysis (DGA) on power transformers. DGA, as a diagnostics for power transformer, has been available for many years. Methods such as SOM has been applied to analyze data and to determine trends which are not obvious to the standard DGA ratio methods such as Duval Triangle.<sup>[15]</sup>

A fourth area of application for data mining in science/engineering is within educational research, where data mining has been used to study the factors leading students to choose to engage in behaviors which reduce their learning<sup>[16]</sup> and to understand the factors influencing university student retention.<sup>[17]</sup> A similar example of the social application of data mining is its use in expertise finding systems, whereby descriptors of human expertise are extracted, normalized and classified so as to facilitate the finding of experts, particularly in scientific and technical fields. In this way, data mining can facilitate Institutional memory.

Other examples of applying data mining method applications are biomedical data facilitated by domain ontologies,<sup>[18]</sup> mining clinical trial data,<sup>[19]</sup> traffic analysis using SOM,<sup>[20]</sup> et cetera.

In adverse drug reaction surveillance, the Uppsala Monitoring Centre has, since 1998, used data mining methods to routinely screen for reporting patterns indicative of emerging drug safety issues in the WHO global database of 4.6 million suspected adverse drug reaction incidents.<sup>[21]</sup> Recently, similar methodology has been developed to mine large collections of electronic health records for temporal patterns associating drug prescriptions to medical diagnoses.<sup>[22]</sup>

## Spatial data mining

Spatial data mining is the application of data mining methods to spatial data. Spatial data mining follows along the same functions in data mining, with the end objective to find patterns in geography. So far, data mining and Geographic Information Systems (GIS) have existed as two separate technologies, each with its own methods, traditions and approaches to visualization and data analysis. Particularly, most contemporary GIS have only very basic spatial analysis functionality. The immense explosion in geographically referenced data occasioned by developments in IT, digital mapping, remote sensing, and the global diffusion of GIS emphasizes the importance of developing data driven inductive approaches to geographical analysis and modeling.

---

Data mining, which is the partially automated search for hidden patterns in large databases, offers great potential benefits for applied GIS-based decision-making. Recently, the task of integrating these two technologies has become critical, especially as various public and private sector organizations possessing huge databases with thematic and geographically referenced data begin to realize the huge potential of the information hidden there. Among those organizations are:

- offices requiring analysis or dissemination of geo-referenced statistical data
- public health services searching for explanations of disease clusters
- environmental agencies assessing the impact of changing land-use patterns on climate change
- geo-marketing companies doing customer segmentation based on spatial location.

### **Challenges**

Geospatial data repositories tend to be very large. Moreover, existing GIS datasets are often splintered into feature and attribute components, that are conventionally archived in hybrid data management systems. Algorithmic requirements differ substantially for relational (attribute) data management and for topological (feature) data management.<sup>[23]</sup> Related to this is the range and diversity of geographic data formats, that also presents unique challenges. The digital geographic data revolution is creating new types of data formats beyond the traditional "vector" and "raster" formats. Geographic data repositories increasingly include ill-structured data such as imagery and geo-referenced multi-media.<sup>[24]</sup>

There are several critical research challenges in geographic knowledge discovery and data mining. Miller and Han<sup>[25]</sup> offer the following list of emerging research topics in the field:

- **Developing and supporting geographic data warehouses** – Spatial properties are often reduced to simple aspatial attributes in mainstream data warehouses. Creating an integrated GDW requires solving issues in spatial and temporal data interoperability, including differences in semantics, referencing systems, geometry, accuracy and position.
- **Better spatio-temporal representations in geographic knowledge discovery** – Current geographic knowledge discovery (GKD) methods generally use very simple representations of geographic objects and spatial relationships. Geographic data mining methods should recognize more complex geographic objects (lines and polygons) and relationships (non-Euclidean distances, direction, connectivity and interaction through attributed geographic space such as terrain). Time needs to be more fully integrated into these geographic representations and relationships.
- **Geographic knowledge discovery using diverse data types** – GKD methods should be developed that can handle diverse data types beyond the traditional raster and vector models, including imagery and geo-referenced multimedia, as well as dynamic data types (video streams, animation).

In four annual surveys of data miners,<sup>[26]</sup> data mining practitioners consistently identified that they faced three key challenges more than any others:

- Dirty Data
- Explaining Data Mining to Others
- Unavailability of Data / Difficult Access to Data

In the 2010 survey data miners also shared their experiences in overcoming these challenges.<sup>[27]</sup>

### **Sensor data mining**

Wireless sensor networks can be used for facilitating the collection of data for spatial data mining for a variety of applications such as air pollution monitoring.<sup>[28]</sup> A characteristic of such networks is that nearby sensor nodes monitoring an environmental feature typically register similar values. This kind of data redundancy due to the spatial correlation between sensor observations inspires the techniques for in-network data aggregation and mining. By measuring the spatial correlation between data sampled by different sensors, a wide class of specialized algorithms

can be developed to develop more efficient spatial data mining algorithms.<sup>[29]</sup>

## Visual data mining

The process of turning from analogical into digital, large data sets have been generated, collected and stored discovering statistical patterns, trends and information which is hidden in data, in order to build predictive patterns. A study found that Visual Data Mining is faster and much more intuitive than traditional data mining.<sup>[30][31]</sup>

## Music data mining

Data mining techniques and in particular co-occurrence analysis has been used to discover relevant similarities among music corpora (radio list, CD databases) for the purpose of classifying music into genres in an objective manner.<sup>[32]</sup>

## Surveillance

Prior data mining to stop terrorist programs under the U.S. government include the Total Information Awareness (TIA) program, Secure Flight (formerly known as Computer-Assisted Passenger Prescreening System (CAPPs II)), Analysis, Dissemination, Visualization, Insight, Semantic Enhancement (ADVISE),<sup>[33]</sup> and the Multi-state Anti-Terrorism Information Exchange (MATRIX).<sup>[34]</sup> These programs have been discontinued due to controversy over whether they violate the US Constitution's 4th amendment, although many programs that were formed under them continue to be funded by different organizations, or under different names.<sup>[35]</sup>

Two plausible data mining methods in the context of combating terrorism include "pattern mining" and "subject-based data mining".

### Pattern mining

"Pattern mining" is a data mining method that involves finding existing patterns in data. In this context *patterns* often means association rules. The original motivation for searching association rules came from the desire to analyze supermarket transaction data, that is, to examine customer behavior in terms of the purchased products. For example, an association rule "beer  $\Rightarrow$  potato chips (80%)" states that four out of five customers that bought beer also bought potato chips.

In the context of pattern mining as a tool to identify terrorist activity, the National Research Council provides the following definition: "Pattern-based data mining looks for patterns (including anomalous data patterns) that might be associated with terrorist activity — these patterns might be regarded as small signals in a large ocean of noise."<sup>[36][37][38]</sup> Pattern Mining includes new areas such a Music Information Retrieval (MIR) where patterns seen both in the temporal and non temporal domains are imported to classical knowledge discovery search methods.

### Subject-based data mining

"Subject-based data mining" is a data mining method involving the search for associations between individuals in data. In the context of combating terrorism, the National Research Council provides the following definition: "Subject-based data mining uses an initiating individual or other datum that is considered, based on other information, to be of high interest, and the goal is to determine what other persons or financial transactions or movements, etc., are related to that initiating datum."<sup>[37]</sup>

---

## Knowledge grid

Knowledge discovery on the Grid typically refers to conducting knowledge discovery in an open environment using Grid computing concepts, allowing users to integrate data from various online data sources as well make use of remote resources for executing their data mining tasks. The earliest examples was the Discovery Net [39][40] developed at Imperial College London which won the "Most Innovative Data Intensive Application Award" at the ACM SC02 (Supercomputing 2002) conference and exhibition, based on a demonstration of a fully interactive distributed knowledge discovery application for a bioinformatics application. Other examples also include work conducted by researchers at the University of Calabria who developed a Knowledge Grid architecture for distributed knowledge discovery, based on grid computing.<sup>[41][42]</sup>

## Reliability

Data mining can be misused, and can unintentionally produce results which appear significant but do not actually predict future behavior and cannot be reproduced on a new sample of data. See Data snooping, Data dredging.

## Privacy concerns and ethics

Some people believe that data mining itself is ethically neutral.<sup>[43]</sup> It is important to note that the term data mining has no ethical implications. The term is often associated with the mining of information in relation to peoples' behavior. However, data mining is a statistical method that is applied to a set of information, or a data set. Associating these data sets with people is an extreme narrowing of the types of data that are available in today's technological society. Examples could range from a set of crash test data for passenger vehicles, to the performance of a group of stocks. These types of data sets make up a great proportion of the information available to be acted on by data mining methods, and rarely have ethical concerns associated with them. However, the ways in which data mining can be used can raise questions regarding privacy, legality, and ethics.<sup>[44]</sup> In particular, data mining government or commercial data sets for national security or law enforcement purposes, such as in the Total Information Awareness Program or in ADVISE, has raised privacy concerns.<sup>[45][46]</sup>

Data mining requires data preparation which can uncover information or patterns which may compromise confidentiality and privacy obligations. A common way for this to occur is through data aggregation. Data aggregation is when the data are accrued, possibly from various sources, and put together so that they can be analyzed.<sup>[47]</sup> This is not data mining per se, but a result of the preparation of data before and for the purposes of the analysis. The threat to an individual's privacy comes into play when the data, once compiled, cause the data miner, or anyone who has access to the newly compiled data set, to be able to identify specific individuals, especially when originally the data were anonymous.

It is recommended that an individual is made aware of the following before data are collected:

- the purpose of the data collection and any data mining projects,
- how the data will be used,
- who will be able to mine the data and use them,
- the security surrounding access to the data, and in addition,
- how collected data can be updated.<sup>[47]</sup>

In the United States, privacy concerns have been somewhat addressed by their congress via the passage of regulatory controls such as the Health Insurance Portability and Accountability Act (HIPAA). The HIPAA requires individuals to be given "informed consent" regarding any information that they provide and its intended future uses by the facility receiving that information. According to an article in Biotech Business Week, "In practice, HIPAA may not offer any greater protection than the longstanding regulations in the research arena," says the AAHC. More importantly, the rule's goal of protection through informed consent is undermined by the complexity of consent forms that are required of patients and participants, which approach a level of incomprehensibility to average

---

individuals."<sup>[48]</sup> This underscores the necessity for data anonymity in data aggregation practices.

One may additionally modify the data so that they are anonymous, so that individuals may not be readily identified.<sup>[47]</sup> However, even de-identified data sets can contain enough information to identify individuals, as occurred when journalists were able to find several individuals based on a set of search histories that were inadvertently released by AOL.<sup>[49]</sup>

## Software

### Free libre open-source data-mining software and applications

1. Carrot2 – Text and search results clustering framework.
2. Chemicalize.org – A chemical structure miner and web search engine.
3. ELKI – A university research project with advanced cluster analysis and outlier detection methods written in the Java language.
4. GATE – Natural language processing and language engineering tool.
5. JHepWork – Java cross-platform data analysis framework developed at ANL.
6. KNIME – The Konstanz Information Miner, a user friendly and comprehensive data analytics framework.
7. NLTK or Natural Language Toolkit – A suite of libraries and programs for symbolic and statistical natural language processing (NLP) for the Python language.
8. Orange – A component-based data mining and machine learning software suite written in the Python language.
9. R – A programming language and software environment for statistical computing, data mining and graphics. It is part of the GNU project.
10. RapidMiner – An environment for machine learning and data mining experiments.
11. UIMA – The UIMA (Unstructured Information Management Architecture) is a component framework for analyzing unstructured content such as text, audio and video, originally developed by IBM.
12. Weka – A suite of machine learning software written in the Java language.

In 2010, the open source R language overtook other tools to become the tool used by more data miners (43%) than any other.<sup>[26]</sup>

### Commercial data-mining software and applications

- IBM InfoSphere Warehouse – in-database data mining platform provided by IBM.
- IBM SPSS Modeler – data mining software provided by IBM.
- KXEN Infinite Insight - data mining software provided by KXEN
- Microsoft Analysis Services - data mining software provided by Microsoft
- SAS Enterprise Miner – data mining software provided by the SAS Institute.
- STATISTICA Data Miner – data mining software provided by StatSoft.
- Oracle Data Mining - data mining software by Oracle.

According to Rixer's Annual Data Miner Survey in 2010, IBM SPSS Modeler, STATISTICA Data Miner and R received the strongest satisfaction ratings.<sup>[26]</sup>

---

## Marketplace surveys

Several researchers and organizations have conducted reviews of data mining tools and surveys of data miners. These identify some of the strengths and weaknesses of the software packages. They also provide an overview of the behaviors, preferences and views of data miners. Some of these reports include:

- Annual Rixer Analytics Data Miner Surveys.<sup>[26]</sup>
- Forrester Research 2010 Predictive Analytics and Data Mining Solutions report.<sup>[50]</sup>
- Gartner 2008 "Magic Quadrant" report.<sup>[51]</sup>
- Haughton et al.'s 2003 Review of Data Mining Software Packages in *The American Statistician*.<sup>[52]</sup>
- Robert A. Nisbet's 2006 Three Part Series of articles "Data Mining Tools: Which One is Best For CRM?"<sup>[53]</sup>
- 2011 Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery in<sup>[54]</sup>

## References

- [1] Fayyad, Usama; Gregory Piatetsky-Shapiro, and Padhraic Smyth (1996). "From Data Mining to Knowledge Discovery in Databases" (<http://www.kdnuggets.com/gpspubs/aimag-kdd-overview-1996-Fayyad.pdf>). . Retrieved 2008-12-17.
- [2] "Data Mining Curriculum" (<http://www.sigkdd.org/curriculum.php>). ACM SIGKDD. 2006-04-30. . Retrieved 2011-10-28.
- [3] Clifton, Christopher (2010). "Encyclopædia Britannica: Definition of Data Mining" (<http://www.britannica.com/EBchecked/topic/1056150/data-mining>). . Retrieved 2010-12-09.
- [4] Ian H. Witten; Eibe Frank; Mark A. Hall (30 January 2011). *Data Mining: Practical Machine Learning Tools and Techniques* (3 ed.). Elsevier. ISBN 978-0-12-374856-0.
- [5] R.R. Bouckaert; E. Frank; M.A. Hall; G. Holmes; B. Pfahringer; P. Reutemann; I.H. Witten (2010). "WEKA Experiences with a Java open-source project". *Journal of Machine Learning Research* **11**: 2533–2541. "the original title, "Practical machine learning", was changed [...] The term "data mining" was prepended primarily for marketing reasons."
- [6] Kantardzic, Mehmed (2003). *Data Mining: Concepts, Models, Methods, and Algorithms*. John Wiley & Sons. ISBN 0471228524. OCLC 50055336.
- [7] Proceedings (<http://www.kdd.org/conferences.php>), International Conferences on Knowledge Discovery and Data Mining, ACM, New York.
- [8] SIGKDD Explorations (<http://www.kdd.org/explorations/about.php>), ACM, New York.
- [9] Günemann, S.; Kremer, H.; Seidl, T. (2011). "An extension of the PMML standard to subspace clustering models". *Proceedings of the 2011 workshop on Predictive markup language modeling - PMML '11*. pp. 48. doi:10.1145/2023598.2023605. ISBN 9781450308373.
- [10] Ellen Monk, Bret Wagner (2006). *Concepts in Enterprise Resource Planning, Second Edition*. Thomson Course Technology, Boston, MA. ISBN 0-619-21663-8. OCLC 224465825.
- [11] Roberto Battiti and Mauro Brunato, *Reactive Business Intelligence. From Data to Models to Insight* (<http://www.reactivebusinessintelligence.com/>), Reactive Search Srl, Italy, February 2011. ISBN 978-88-905795-0-9.
- [12] Battiti, Roberto; Andrea Passerini (2010). "Brain-Computer Evolutionary Multi-Objective Optimization (BC-EMO): a genetic algorithm adapting to the decision maker". *IEEE Transactions on Evolutionary Computation* **14** (15): 671–687. Bibcode url =<http://rtm.science.unitn.it/~battiti/archive/bcemo.pdf>; doi:10.1109/TEVC.2010.2058118.
- [13] Tony Fountain, Thomas Dietterich & Bill Sudka (2000) *Mining IC Test Data to Optimize VLSI Testing* (<http://web.engr.oregonstate.edu/~tgd/publications/kdd2000-dlft.pdf>), in Proceedings of the Sixth ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. (pp. 18–25). ACM Press.
- [14] Xingquan Zhu, Ian Davidson (2007). *Knowledge Discovery and Data Mining: Challenges and Realities*. Hershey, New York. p. 18. ISBN 978-159904252-7.
- [15] A.J. McGrail, E. Gulski et al.. "Data Mining Techniques to Asses the Condition of High Voltage Electrical Plant". *CIGRE WG 15.11 of Study Committee 15*.
- [16] R. Baker. "Is Gaming the System State-or-Trait? Educational Data Mining Through the Multi-Contextual Application of a Validated Behavioral Model". *Workshop on Data Mining for User Modeling 2007*.
- [17] J.F. Superby, J-P. Vandamme, N. Meskens. "Determination of factors influencing the achievement of the first-year university students using data mining methods". *Workshop on Educational Data Mining 2006*.
- [18] Xingquan Zhu, Ian Davidson (2007). *Knowledge Discovery and Data Mining: Challenges and Realities*. Hershey, New York. pp. 163–189. ISBN 978-159904252-7.
- [19] Xingquan Zhu, Ian Davidson (2007). *Knowledge Discovery and Data Mining: Challenges and Realities*. Hershey, New Your. pp. 31–48. ISBN 978-159904252-7.
- [20] Yudong Chen, Yi Zhang, Jianming Hu, Xiang Li. "Traffic Data Analysis Using Kernel PCA and Self-Organizing Map". *Intelligent Vehicles Symposium, 2006 IEEE*.
- [21] Bate A, Lindquist M, Edwards IR, Olsson S, Orre R, Lansner A, De Freitas RM. A Bayesian neural network method for adverse drug reaction signal generation. *Eur J Clin Pharmacol*. 1998 Jun;54(4):315–21.

- [22] Norén GN, Bate A, Hopstadius J, Star K, Edwards IR. Temporal Pattern Discovery for Trends and Transient Effects: Its Application to Patient Records. *Proceedings of the Fourteenth International Conference on Knowledge Discovery and Data Mining SIGKDD 2008*, pages 963–971. Las Vegas NV, 2008.
- [23] Healey, R., 1991, Database Management Systems. In Maguire, D., Goodchild, M.F., and Rhind, D., (eds.), *Geographic Information Systems: Principles and Applications* (London: Longman).
- [24] Câmara, A. S. and Raper, J., (eds.), 1999, *Spatial Multimedia and Virtual Reality*, (London: Taylor and Francis).
- [25] Miller, H. and Han, J., (eds.), 2001, *Geographic Data Mining and Knowledge Discovery*, (London: Taylor & Francis).
- [26] Karl Rexer, Heather Allen, & Paul Gearan (2010) *2010 Data Miner Survey Summary* (<http://www.rexeranalytics.com/Data-Miner-Survey-Results-2010.html>), presented at Predictive Analytics World, Oct. 2010.
- [27] Karl Rexer, Heather Allen, & Paul Gearan (2010) *2010 Data Miner Survey – Overcoming Data Mining's Top Challenges* ([http://www.rexeranalytics.com/Overcoming\\_Challenges.html](http://www.rexeranalytics.com/Overcoming_Challenges.html)).
- [28]
- This citation will be automatically completed in the next few minutes. You can jump the queue or expand by hand ([http://en.wikipedia.org/wiki/Template:cite\\_doi/\\_10.3390.2fs8063601?preload=Template:Cite\\_doi/preload&editintro=Template:Cite\\_doi/editintro&action=edit](http://en.wikipedia.org/wiki/Template:cite_doi/_10.3390.2fs8063601?preload=Template:Cite_doi/preload&editintro=Template:Cite_doi/editintro&action=edit))
- [29] This citation will be automatically completed in the next few minutes. You can jump the queue or expand by hand ([http://en.wikipedia.org/wiki/Template:cite\\_doi/\\_10.1109.2fjse.2010.2056916?preload=Template:Cite\\_doi/preload&editintro=Template:Cite\\_doi/editintro&action=edit](http://en.wikipedia.org/wiki/Template:cite_doi/_10.1109.2fjse.2010.2056916?preload=Template:Cite_doi/preload&editintro=Template:Cite_doi/editintro&action=edit))
- [30] ZHAO Kaidi and LIU Bing (University of Illinois), M. TIRPARK Thomas and WEIMIN Xiao (Motorola Labs): A Visual Data Mining Framework for Convenient Identification of Useful Knowledge (<http://dl.acm.org/citation.cfm?id=1106390>)
- [31] A. KEIM Daniel, Information Visualization and Visual Data Mining (<http://citeseer.ist.psu.edu/viewdoc/summary?doi=10.1.1.135.7051>)
- [32] Pachet, F., Westermann, G. and Laigre, D. Musical Data Mining for Electronic Music Distribution (<http://www.csl.sony.fr/downloads/papers/2001/pachet01c.pdf>), Proceedings of the 1st WedelMusic Conference, pp. 101-106, Firenze, Italy, 2001.
- [33] Government Accountability Office, *Data Mining: Early Attention to Privacy in Developing a Key DHS Program Could Reduce Risks*, GAO-07-293, Washington, D.C.: February 2007.
- [34] Secure Flight Program report (<http://www.msnbc.msn.com/id/20604775/>), MSNBC.
- [35] "Total/Terrorism Information Awareness (TIA): Is It Truly Dead?" ([http://w2.eff.org/Privacy/TIA/20031003\\_comments.php](http://w2.eff.org/Privacy/TIA/20031003_comments.php)). *Electronic Frontier Foundation (official website)*. 2003. . Retrieved 2009-03-15.
- [36] R. Agrawal et al., *Fast discovery of association rules*, in *Advances in knowledge discovery and data mining* pp. 307–328, MIT Press, 1996.
- [37] National Research Council, *Protecting Individual Privacy in the Struggle Against Terrorists: A Framework for Program Assessment*, Washington, DC: National Academies Press, 2008.
- [38] Stephen Haag et al. (2006). *Management Information Systems for the information age*. Toronto: McGraw-Hill Ryerson. p. 28. ISBN 0-07-095569-7. OCLC 63194770.
- [39] Ghanem, M.; Guo, Y.; Rowe, A.; Wendel, P. (2002). "Grid-based knowledge discovery services for high throughput informatics". *Proceedings 11th IEEE International Symposium on High Performance Distributed Computing*. pp. 416. doi:10.1109/HPDC.2002.1029946. ISBN 0-7695-1686-6.
- [40] Ghanem, M.; Curcin, V.; Wendel, P.; Guo, Y. (2009). "Building and Using Analytical Workflows in Discovery Net". *Data Mining Techniques in Grid Computing Environments*. pp. 119. doi:10.1002/9780470699904.ch8. ISBN 9780470699904.
- [41] Mario Cannataro; Domenico Talia (January 2003). "The knowledge grid: An Architecture for Distributed Knowledge Discovery" (<http://grid.deis.unical.it/papers/pdf/CACM2003.pdf>). *Communications of the ACM* **46** (1): 89–93. doi:10.1145/602421.602425. . Retrieved October 17, 2011.
- [42] Domenico Talia; Paolo Trunfio (July 2010). "How distributed data mining tasks can thrive as knowledge services" (<http://grid.deis.unical.it/papers/pdf/CACM2010.pdf>). *Communications of the ACM* **53** (7): 132–137. doi:10.1145/1785414.1785451. . Retrieved October 17, 2011.
- [43] William Seltzer. *The Promise and Pitfalls of Data Mining: Ethical Issues* (<http://www.amstat.org/committees/ethics/linksdir/Jsm2005Seltzer.pdf>)..
- [44] Chip Pitts (March 15, 2007). "The End of Illegal Domestic Spying? Don't Count on It" ([http://www.washingtonspectator.com/articles/20070315surveillance\\_1.cfm](http://www.washingtonspectator.com/articles/20070315surveillance_1.cfm)). *Washington Spectator*..
- [45] K.A. Taipale (December 15, 2003). "Data Mining and Domestic Security: Connecting the Dots to Make Sense of Data" (<http://www.strl.org/cite.cgi?volume=5&article=2>). *Columbia Science and Technology Law Review* **5** (2). OCLC 45263753. SSRN 546782. .
- [46] John Resig, Ankur Teredesai (2004). "A Framework for Mining Instant Messaging Services" (<http://citeseer.ist.psu.edu/resig04framework.html>). *In Proceedings of the 2004 SIAM DM Conference*.
- [47] *Think Before You Dig: Privacy Implications of Data Mining & Aggregation* (<http://www.nascio.org/publications/documents/NASCIO-dataMining.pdf>), NASCIO Research Brief, September 2004.
- [48] Biotech Business Week Editors. (June 30, 2008). BIOMEDICINE; HIPAA Privacy Rule Impedes Biomedical Research. Biotech Business Week. Retrieved 17 Nov 2009 from LexisNexis Academic.
- [49] AOL search data identified individuals (<http://www.securityfocus.com/brief/277>), SecurityFocus, August 2006.
- [50] James Kobielski (1 July 2008) *The Forrester Wave: Predictive Analytics and Data Mining Solutions, Q1 2010* ([http://www.forrester.com/rb/Research/wave&trade;\\_predictive\\_analytics\\_and\\_data\\_mining\\_solutions,q{id}/56077/t/2](http://www.forrester.com/rb/Research/wave&trade;_predictive_analytics_and_data_mining_solutions,q{id}/56077/t/2)), Forrester Research.

- [51] Gareth Herschel (1 July 2008) *Magic Quadrant for Customer Data-Mining Applications* (<http://mediaproducts.gartner.com/reprints/sas/vol5/article3/article3.html>), Gartner Inc.
- [52] Dominique Haughton, Joel Deichmann, Abdolreza Eshghi, Selin Sayek, Nicholas Teebagy, & Heikki Topi (2003) *A Review of Software Packages for Data Mining* (<http://www.jstor.org/pss/30037299>), *The American Statistician*, Vol. 57, No. 4, pp. 290–309.
- [53] Robert Nisbet (2006) *Data Mining Tools: Which One is Best for CRM? Part 1* (<http://www.information-management.com/specialreports/20060124/1046025-1.html>), *Information Management Special Reports*, January 2006.
- [54] Ralf Mikut; Markus Reischl (September/October 2011). "Data Mining Tools" (<http://onlinelibrary.wiley.com/doi/10.1002/widm.24/abstract>). *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 1 (5): 431–445. doi:10.1002/widm.24. . Retrieved October 21, 2011.

## Further reading

- Cabena, Peter, Pablo Hadjnian, Rolf Stadler, Jaap Verhees and Alessandro Zanasi (1997). *Discovering Data Mining: From Concept to Implementation*. Prentice Hall, ISBN 0-13-743980-6.
- Feldman, Ronen and James Sanger. *The Text Mining Handbook*. Cambridge University Press, ISBN 978-0-521-83657-9.
- Guo, Yike and Robert Grossman, editors (1999). *High Performance Data Mining: Scaling Algorithms, Applications and Systems*. Kluwer Academic Publishers.
- Hastie, Trevor, Robert Tibshirani and Jerome Friedman (2001). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, ISBN 0-387-95284-5.
- Liu, Bing (2007). *Web Data Mining: Exploring Hyperlinks, Contents and Usage Data*. Springer, ISBN 3-540-37881-2.
- Murphy, Chris (May 16, 2011). "Is Data Mining Free Speech?". *InformationWeek* (UMB): 12.
- Nisbet, Robert, John Elder, Gary Miner (2009). *Handbook of Statistical Analysis & Data Mining Applications*. Academic Press/Elsevier. ISBN 9780123747655
- Poncelet, Pascal, Florent Masséglio and Maguelonne Teisseire, editors (October 2007). "Data Mining Patterns: New Methods and Applications", *Information Science Reference*. ISBN 978-1-59904-162-9.
- Pang-Ning Tan, Michael Steinbach and Vipin Kumar (2005). *Introduction to Data Mining*. ISBN 0-321-32136-7
- Sergios Theodoridis, Konstantinos Koutroumbas (2009). *Pattern Recognition*, 4th Edition. Academic Press. ISBN 978-1-59749-272-0.
- Weiss and Indurkhy. *Predictive Data Mining*. Morgan Kaufmann.
- Ian H. Witten; Eibe Frank; Mark A. Hall (30 January 2011). *Data Mining: Practical Machine Learning Tools and Techniques* (3 ed.). Elsevier. ISBN 978-0-12-374856-0. (See also Free Weka software.)
- Ye, N. (2003). *The Handbook of Data Mining*. Mahwah, New Jersey: Lawrence Erlbaum.

## External links

- Data Mining Software ([http://www.dmoz.org/Computers/Software/Databases/Data\\_Mining/](http://www.dmoz.org/Computers/Software/Databases/Data_Mining/)) at the Open Directory Project

# Article Sources and Contributors

**Data mining** Source: <http://en.wikipedia.org/w/index.php?oldid=479861836> Contributors: 16@r, 1yesfan, 48states, 9Nak, A m sheldon, AVM, Aacruzr, Aaron Brenneman, Aaronbrick, Aaronzat, Abrech, Acather96, Adam McMaster, Adam78, Adamminstead, Adrian.walker, Aetheling, Ahoerstemeier, Aiwing, Akhil joey, Alan Liefting, AlexAnglin, Alfo, Altenmann, Alusayman, Amossin, Andonic, Andre Engels, Andrei Stroe, Andres.santana, Andrevan, AndrewHZ, Andrzejrauch, Andy Marchbanks, Angela, Anilkumar 0587, Ansell, Antandrus, Antonrojo, Ap, Apogr, Ar5144-06, Aranea Mortem, ArielGold, Arpabr, Articdawg, Aseld, Atallcostsky, Atamir, Atheweatherman, Athernar, AtholM, Atlantia, Axeman89, BMF81, BRUTE, Badgernet, Barek, Barticus88, Beefyt, Beetstra, Bennose, Bensin, Bhikubhadwa, Blake-, Blanchardb, BluePlateSpecial, Bmhkim, Bmicom, Boleslav Bobcik, Bonadea, Bongomatic, Bongwarrior, Bovinecone, Bryan Derksen, Btcoal, Bubba73, Bull3r, Burakordu, Burkeangirl, CRGreathouse, CWY2190, Caesura, Calliopejen1, Calum MacÜisdean, Can't sleep, clown will eat me, Candace Gillholley, Capitalist, Capricorn42, CesarB, CfIm001, Cgfjppg, Charvest, Chenxlee, Chenzheruc, Chire, Chivista, Chowbok, Chrisguyot, Chzz, Cis411, Cmccormick8, CoderGnome, Colewaldron, Comatose51, Cometstyles, CommodiCast, Comp8956, Compo, Compunoe, Confuss, Connormah, Cronk28, Ctacmo, Cureden, Curuzx, Cutter, D vandyke67, DH85868993, Dagoldman, DanDoughty, Dank, Daqu, Data.mining, DataExp, Davgrig04, David Eppstein, David Haslam, Dawn Bard, Dcoetzee, DePiep, Deanabb, Debuntu, Deepred6502, Dekisugi, Delaszk, Delldot, Deltabeignet, Den fjärrtraden ankan, Denoir, Deville, Dhart, Dicklyon, Differentview, Dlyons493, Dmmd123, Dodabe, Don4of4, Doug4, Download, Dr Oldekop, DrCroco, Dreamyshade, Drknightbatman, Drowne, DryaUnda, Dudukeda, DueSouth, DyingIce, ERI employee, Ecmathouse, Edayapattiarun, Ejrrjs, Elsendero, Elseviereditormath, Emarsee, Emilescheepers444, Enggakshat, EngineerScotty, EntropyAS, EoGuy, Epic, Equilibrioception, Er.piyushkp, Ericmortenson, ErinRea, Essjay, Estard, Etz Haim, Eyalurus, Evafamax, Fahadsadah, Falcon8765, Fang Aili, Fauxstar, Fcueto, Feraudhy, Filemon, Filip\*, Fly by Night, Flyskippy1, Foober, Fourthcourse, Foxyliyah, Frap, Fratrep, Funkykeith777, Furykef, Fuzheado, Gadfiim, Gaius Cornelius, Gargvikram07, Gary King, GeorgeBills, Giftite, Giggy, Glane23, Glenn Maddox, Gmelli, Gogo Dodo, GoingBatty, Gomm, Gpswiki, Graciella, GraemeL, Graham87, GregAsche, GregorB, Ground Zero, Gcschoryn, Guerrerrcarlos, Gurubrahma, Gwyatt-agastie, Gzkn, H005, Haakon, Hadleywickham, HamburgerRadio, Hanifbbz, Headbomp, Hede2000, Hefaisatos, Helwr, Henrygb, Hherbert, Hike395, Hu12, HughJorgan, Hulek, Hut 8.5, Hydrargrymus, Hynespn, Ianhowlett, Inoshika, Ipeirots, Isomorphic, IvanLanin, Ixfd64, JCLately, JHP, JJ Harrison, Jackverr, Jamiemac, Janvo, Jasonem, Jay, Jayrde, Jefgodesky, Jeltz, Jerryobject, Jesuja, Jet57, Jfitz, Jfroelich, Jguthaaaz, Jim1138, Jimmaths, Jimmy Pitt, Jkosik1, Jks, Joanlofe, Joerg Kurt Wegner, John Broughton, John Vandenberg, Joinarnold, JonHarder, Jonkerz, Josephjhthomas, Joy, Jpbown, Jpcedenog, Jpoelma13, Jrtaylor19, Juliustch, Just Another Dan, Kadambarid, Kainaw, Katanada, Katieh5584, Katyare, Kbdank71, Kexperi, KeyStroke, Kgflieischmann, Killian441, Kjtobo, Kkarimi, Kmettler, Kotsiantis, Krexer, Kur, Kusal one, Kxlai, Kyleamiller, LOL, LaUs3r, Lambiam, Lament, Lark137, LarryGilbert, Lawrykid, LeighvsOptimvsMaximvs, Leiluo, Leo505, Leonardo61, Lisasolomonalford, Liwaste, Looper5920, Loren.wilton, Lovro, Luke145, M.r santosh kumar, MER-C, MK8, Mabdul, Mackseem, MainFrame, Malo, Mannfredo, Manufan 11, MarcoTolo, Mark Klamberg, Mark.hornick, Marner, Martin Jensen, Mathstat, Mat90855, Mattelsen, Matusz, Maureen, Mdd, Melcombe, Mhahsler, Michael Hardy, Michal burda, Mike Rosoff, Mike Schwartz, Mike78465, Mikeputnam, Mindmatrix, Mkch, Modify, Monkeyman, Moshiurbd, Mpaye, MrOllie, Mshechet, Mswake, Mydogategodshat, Mátyás, N8chz, NHRHS2010, Nabeth, Naeemmalik036, Nano5656, Nathanaleshwilyild, NawlinWiki, Neile, Netra Nahar, Netsnipe, Nezza 4 eva, Ngorman, Nilfanion, Nirvanalulu, Nixdorf, Nineagle, Nowozin, Nuno Tavares, Nuwanmenuka, ONEder Boy, Ocarbonne, Odo Benus, Ojigiri, OlafvanD, OlavN, Oleg Alexandrov, Oli2140, Onasraou, Onco p53, Onebravemonkey, OverlordQ, OwenX, Oxymoron83, Pakcw, Palapa, Parkshi Basrur, Paul Foxworthy, Pdfpdf, Pgan002, Ph92, Phantomsteve, Pharaoh of the Wizards, Philip Trueman, Philopedia, Philwelch, PhnomPencil, Pinethicket, Pingku, Pinkadelica, Pjoe, Pmrauer, Pmbhagat, Predictor, Priyan782, Propheci, Pyromaniaman, Quanticle, Qwfp, R'n'B, RJASE1, RTG, Raand, Rabarbaro70, Radagast83, RadioFan, Ralf Klinkenberg, Ralf Mikut, Ramkumar Krishnan, Raomohsinhan, Raymondwinn, Rd232, Reaper Eternal, RepubCarrier, Rich Farmbrough, RichardF, Ripper234, Rjanag, Rjwilmsi, Rknasc, RoboBaby, RonFredericks, Ronz, Rosannel, Roxy1984, Ruggaad, Rusty spatula, Ryanscraper, SOMart, Sanya r, SarekOFVulcan, Sarnholm, Sbacle, Schul253, Scientio, Seaphoto, Sebastianmm, Sepreece, Serenity-Fr, Sergio.ballesterro, Shantavira, ShaunMacPherson, Shaw76, Shoesss, Shwapnil, Simon Lacoste-Julien, Simsong, SiobhanHansa, Siroxo, Sjakkalle, SkerHawx, Slumph, Slon02, Smallman12q, Somewherepurple, Spencer, Spiritia, Splash, SpuriousQ, Starnestommy, Statethatiamin, StaticGull, Stekre, Stephenpace, Sterdeus, Stheodor, Stspasha, Sunray, Sunsetsky, Supertouch, Swift-Epic (Rectefy), T789, TFinn734, Talgalii, Tareq300, Tdelamater, Tedickey, Terryholmsby, TestPilot, Texterp, Tgeairn, The Anome, The Evil IP Address, The Rambling Man, The Thing That Should Not Be, Thefriedone, ThreeDee912, Thumperward, Thundertide, Tiffany9027, TigerShark, Timdew, Tlevine, Tobias Bergemann, Tom harrison, Tomwsulcer, Tony1, Toohuman1, Tpbradbury, Traroth, Tristanb, Trusilver, Tslocum, Twerbrou, TyrantX, Ubuntu2, Udufruduhu, Uksas, UngerJ, Unyoyega, Valerie928, Valodzka, Van helsing, Vangelis12, Varlaam, Verloren, Versus22, Verticalsearch, VerySmartNiceGuy, Veyklevar, Vgy7ujn, Vijay.babu.k, Vitamin b, Vlad gerchikov, Vonaurum, Vonkje, W Nowicki, Waggers, Walter Görilitz, Wavelength, Webzie, Weregerbil, Whizzdumb, WikHead, Wiki0709, Wikiant, Wikipi, WinterSpw, WojPob, Woohookitty, X7q, Yecril, Ypoulion, Zeno of Elea, ZimZalaBim, Zzuuzz, 1266 anonymous edits

## License

Creative Commons Attribution-Share Alike 3.0 Unported  
[/creativecommons.org/licenses/by-sa/3.0/](http://creativecommons.org/licenses/by-sa/3.0/)