
Books: (c2009-2012)

Text mining : applications and theory. Wiley, 2010.
QA 76.9 D343 T49 2010

The top ten algorithms in data mining. CRC Press,
c2009.QA 76.9D343 T66 2009

Torgo, Luis. Data mining with R : learning with case studies. Chapman & Hall/CRC, 2011.
QA 76.9 D343 T67 2011

Tuffery, Stephane. Data mining and statistics for decision making. Wiley, c2011.
QA 76.9 D343 T84 2011

Witten, I. H. Data mining : practical machine learning tools and techniques. Morgan Kaufmann/Elsevier, c2011.
QA 76.9 D343 W58 2011

Zhang, Zhongfei. Multimedia data mining : a systematic introduction to concepts and theory. CRC Press, c2009. QA 76.575 Z43 2009

e-Books: (c2011-2013)

Advances in machine learning and data mining for astronomy. CRC Press, c2012.

Contrast data mining concepts, algorithms, and applications. CRC Press, 2013.

Dua, Sumeet. Data mining for bioinformatics. CRC Press, 2013.

Mirkin, B. G. Clustering a data recovery approach. CRC Press, 2013.

Muhamad Amin, Anang Hudaya. Internet-scale pattern recognition new techniques for voluminous data sets and data clouds. CRC Press, 2013.

Talia, Domenico. Service-oriented distributed knowledge discovery. CRC Press, c2013.

Witten, I. H. Data mining practical machine learning tools and techniques. Morgan Kaufmann/Elsevier, c2011.

Woo, Andrew. Shadow algorithms data miner. CRC Press, 2012.

Wu, James. Foundations of predictive analytics. CRC Press, c2012.

Online Subscriptions:

ACM Digital Library—a vast collection of citations and full text from ACM journal and newsletter articles and conference proceedings

IEEE Xplore— Provides full-text access to the world's highest-quality technical literature in electrical engineering, computer science, and electronics published by the Institute of Electrical and Electronics Engineers

Science Direct—the world's largest electronic collection of science, technology and medicine full text and bibliographic information.

Springerlink – one of the world's leading interactive databases for high-quality STM journals, book series, books, reference works and the Online Archives Collection. SpringerLink is a powerful central access point for researchers and scientists

Disclaimer:

This pathfinder contains suggested materials on Data Mining that are available at the College of Engineering Library II. However, some references were not included.

We welcome suggestions for new pathfinder topics.

**University of the Philippines Diliman
COLLEGE OF ENGINEERING
LIBRARY II**

UP Alumni Engineers Centennial Hall
(Engineering Library & Computer Science Bldg.)
Velasquez St., Diliman, Quezon City

Phone: (632) 981-8500 local 3251 to 3252
Fax: (632) 434-8638
Email: library@engglib.upd.edu.ph
Website: www.engglib.upd.edu.ph



University of the Philippines Diliman COLLEGE OF ENGINEERING LIBRARY II

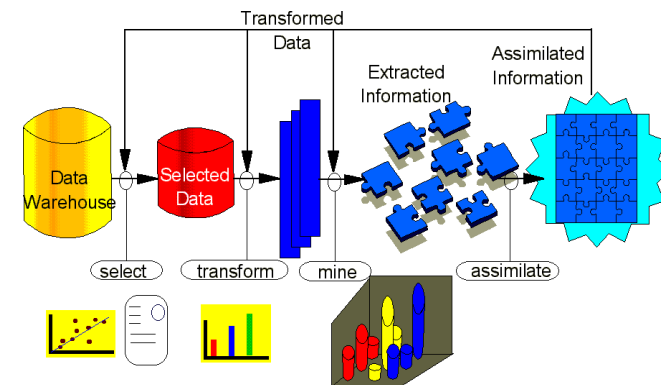


Image URL: http://www.yorku.ca/lbianchi/nats1700/data_mining.gif

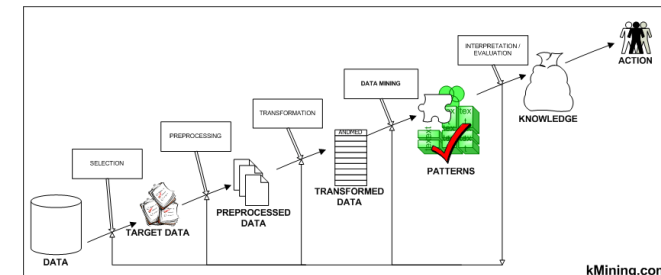


Image URL <http://student.dcu.ie/~aldeirm2/data-mining.gif>

DATA MINING

PATHFINDER

DATA MINING

is the development of computational algorithms for the identification or extraction of structure from data. This is done in order to help reduce, model, understand, or analyze the data. Tasks supported by data mining include prediction, segmentation, dependency modeling, summarization, and change and deviation detection. Database systems have brought digital data capture and storage to the mainstream of data processing, leading to the creation of large data warehouses. These are databases whose primary purpose is to gain access to data for analysis and decision support.

Source: McGraw-Hill's Access Science
Encyclopedia of Science and Technology
(<http://www.accessscience.com>)

Data mining consists of five major elements:

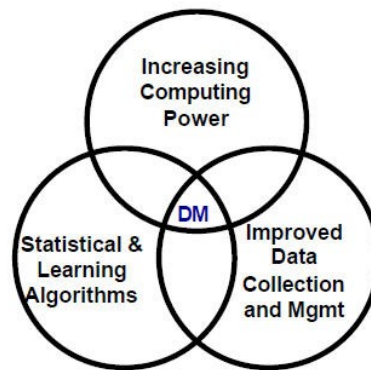
- Extract, transform, and load transaction data onto the data warehouse system.
- Store and manage the data in a multidimensional database system.
- Provide data access to business analysts and information technology professionals.
- Analyze the data by application software.
- Present the data in a useful format, such as a graph or table.

Different levels of analysis are available:

- *Artificial neural networks:* Non-linear predictive models that learn through training and resemble biological neural networks in structure.

- *Genetic algorithms:* Optimization techniques that use processes such as genetic combination, mutation, and natural selection in a design based on the concepts of natural evolution.
- *Decision trees:* Tree-shaped structures that represent sets of decisions. These decisions generate rules for the classification of a dataset
- *Nearest neighbor method:* A technique that classifies each record in a dataset based on a combination of the classes of the k record(s) most similar to it in a historical dataset (where $k > 1$). Sometimes called the k -nearest neighbor technique.
- *Rule induction:* The extraction of useful if-then rules from data based on statistical significance.
- *Data visualization:* The visual interpretation of complex relationships in multidimensional data. Graphics tools are used to illustrate data relationships.

(Source:<http://www.anderson.ucla.edu/faculty/jason.frand/teacher/technologies/palace/datamining.htm>)



<http://www.datawarehouseolution.net/wp-content/uploads/2009/02/Introduction-of-Data-Mining.jpg>

Books: (c2009-c2012)

- Data mining : know it all. Elsevier/Morgan Kaufmann, c2009. QA 76.9 D343 D38 2009
- Geographic data mining and knowledge discovery. CRC Press, c2009. G 70.2 G436 2009
- Guidici, Paolo. Applied data mining for business and industry. Wiley, c2009. QA 76.9 D434 G58 2009
- Kamath, Chandrika. Scientific data mining : a practical perspective. Society for Industrial and Applied Mathematics, c2009. QA 76.9 D343 K36 2009
- Cluster analysis. Wiley, c2011. QA 278 E94 2011
- Du, Hongbo. Data mining techniques and applications : an introduction. Cengage Learning, c2010. QA 76.9 D343 D82
- Dua, Sumeet. Data mining and machine learning in cybersecurity. CRC Press, c2011. QA 76.9 D343 D83 2011
- Han, Jiawei. Data mining : concepts and techniques. Elsevier, c2012. QA 76.9 D343 H36 2012
- Long, Bo. Relational data clustering : models, algorithms, and applications. Chapman & Hall/CRC, c2010. QA 76.9 D343 L66 2010
- Machine interpretation of patterns : image analysis and data mining Sankar K. Pal. World Scientific, 2010. TK 7882 P3 M33 2010
- Managing and mining uncertain data. Springer, c2009. QA 76.9 D343 M36 2009
- Mining software specifications : methodologies and applications. CRC Press, 2011. QA 76.9 D343 M56 2011
- Mitsa, Theophano. Temporal data mining. Chapman & Hall/CRC, c2010. QA 76.9 D343 M58 2010
- Pharmaceutical data mining : approaches and applications for drug discovery. Wiley, c2010. RM 300 P53 2010
- Privacy-aware knowledge discovery : novel applications and new techniques. CRC Press, 2011. QA 76.9 D314 P75 2011
- Suh, Sang C. Practical applications of data mining. Jones & Bartlett, c2012. QA 76.9 D343 S84 2012