

# Data Mining

Data mining (the analysis step of the "Knowledge Discovery in Databases" process, or KDD), an interdisciplinary subfield of computer science, is the computational process of discovering patterns in large data sets involving methods at the intersection of artificial intelligence, machine learning, statistics, and database systems.

Knowledge item description	References
Beginner	
Understanding the concept of data mining	[2.1]
Understanding statistical limits on data mining	[2.1]
Big Data mining: MapReduce and Big Data software stack	[2.2], [3.1]
Competent	
Finding similar items	[2.3], [3.2]
Dimensionality Reduction	[2.11], [3.7]
Link Analysis	[2.5], [3.1]
Frequent Itemsets	[2.6]
Classification with massive datasets	[3.6]
Clustering	[2.7], [3.5]
Expert	
Mining Data Streams	[2.4]
Advertising on the Web	[2.8]
Recommendation Systems	[2.9], [3.4]
Mining Social-Network Graphs	[2.10], [3.3]

## References

#	Reference	Link
2.	Rajaraman, Leskovec, Ullman, Mining of Massive Datasets	TBD
2.1.	Data Mining	Chapter 1
2.2.	Map-Reduce and the New Software Stack	Chapter 2
2.3.	Finding Similar Items	Chapter 3
2.4.	Mining Data Streams	Chapter 4
2.5.	Link Analysis	Chapter 5
2.6.	Frequent Itemsets	Chapter 6
2.7.	Clustering	Chapter 7
2.8.	Advertising on the Web	Chapter 8
2.9.	Recommendation Systems	Chapter 9
2.10.	Mining Social-Network Graphs	Chapter 10
2.11.	Dimensionality Reduction	Chapter 11

3.	Coursera Class: Mining Massive Datasets	<a href="#">Link</a>
3.1.	Week 1	
3.2.	Week 2	
3.3.	Week 3	
3.4.	Week 4	
3.5.	Week 5	
3.6.	Week 6	
3.7.	Week 7	