Bibliography

[1] N. Adam, and J. Wortman. Security-control methods for statistical databases. *ACM Computing Surveys*, 21(4), pp. 515–556, 1989.

[2] G. Adomavicius, and A. Tuzhilin. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge and Data Engineering*, 17(6), pp. 734–749, 2005.

[3] R. C. Agarwal, C. C. Aggarwal, and V. V. V. Prasad. A tree projection algorithm for generation of frequent item sets. *Journal of parallel and Distributed Computing*, 61(3), pp. 350–371, 2001. Also available as *IBM Research Report*, RC21341, 1999.

[4] R. C. Agarwal, C. C. Aggarwal, and V. V. V. Prasad. Depth-first generation of long patterns. *ACM KDD Conference*, pp. 108–118, 2000. Also available as "Depth-first generation of large itemsets for association rules." *IBM Research Report*, RC21538, 1999.

[5] C. Aggarwal. Outlier analysis. *Springer*, 2013.

[6] C. Aggarwal. Social network data analytics. *Springer*, 2011.

[7] C. Aggarwal, and P. Yu. The igrid index: reversing the dimensionality curse for similarity indexing in high-dimensional space. *KDD Conference*, pp. 119–129, 2000.

[8] C. Aggarwal, and P. Yu. On static and dynamic methods for condensation-based privacy-preserving data mining. *ACM Transactions on Database Systems (TODS)*, 33(1), 2, 2008.

[9] C. Aggarwal. On unifying privacy and uncertain data models. *IEEE International Conference on Data Engineering*, pp. 386–395, 2008.

[10] C. Aggarwal. On k-anonymity and the curse of dimensionality, *Very Large Databases Conference*, pp. 901–909, 2005.

[11] C. Aggarwal. On randomization, public information and the curse of dimensionality. *IEEE International Conference on Data Engineering*, pp. 136–145, 2007.

[12] C. Aggarwal. Privacy and the dimensionality curse. *Privacy-Preserving Data Mining: Models and Algorithms*, Springer, pp. 433–460, 2008.

C. C. Aggarwal, *Data Mining: The Textbook*, DOI 10.1007/978-3-319-14142-8

[13] C. Aggarwal, X. Kong, Q. Gu, J. Han, and P. Yu. Active learning: a survey. *Data Classification: Algorithms and Applications*, CRC Press, 2014.

[14] C. Aggarwal. Instance-based learning: A survey. *Data Classification: Algorithms and Applications*, CRC Press, 2014.

[15] C. Aggarwal. Redesigning distance-functions and distance-based applications for high-dimensional data. *ACM SIGMOD Record*, 30(1), pp. 13–18, 2001.

[16] C. Aggarwal, and P. Yu. Mining associations with the collective strength approach. *ACM PODS Conference*, pp. 863–873, 1998.

[17] C. Aggarwal, A. Hinneburg, and D. Keim. On the surprising behavior of distance-metrics in high-dimensional space. *ICDT Conference*, pp. 420–434, 2001.

[18] C. Aggarwal. Managing and mining uncertain data. *Springer*, 2009.

[19] C. Aggarwal, C. Procopiuc, J. Wolf, P. Yu, and J. Park. Fast algorithms for projected clustering. *ACM SIGMOD Conference*, pp. 61–72, 1999.

[20] C. Aggarwal, J. Han, J. Wang, and P. Yu. On demand classification of data streams. *ACM KDD Conference*, pp. 503–508, 2004.

[21] C. Aggarwal. On change diagnosis in evolving data streams. *IEEE Transactions on Knowledge and Data Engineering*, 17(5), pp. 587–600, 2005.

[22] C. Aggarwal, and P. S. Yu. Finding generalized projected clusters in high dimensional spaces. *ACM SIGMOD Conference*, pp. 70–81, 2000.

[23] C. Aggarwal, and S. Parthasarathy. Mining massively incomplete data sets by conceptual reconstruction. *ACM KDD Conference*, pp. 227–232, 2001.

[24] C. Aggarwal. Outlier ensembles: position paper. *ACM SIGKDD Explorations*, 14(2), pp. 49–58, 2012.

[25] C. Aggarwal. On the effects of dimensionality reduction on high dimensional similarity search. *ACM PODS Conference*, pp. 256–266, 2001.

[26] C. Aggarwal, and H. Wang. Managing and mining graph data. *Springer*, 2010.

[27] C. Aggarwal, C. Procopiuc, and P. Yu. Finding localized associations in market basket data. *IEEE Transactions on Knowledge and Data Engineering*, 14(1), pp. 51–62, 2002.

[28] D. Agrawal, and C. Aggarwal. On the design and quantification of privacy-preserving data mining algorithms. *ACM PODS Conference*, pp. 247–255, 2001.

[29] C. Aggarwal, and P. Yu. Privacy-preserving data mining: models and algorithms. *Springer*, 2008.

[30] C. Aggarwal. Managing and mining sensor data. *Springer*, 2013.

[31] C. Aggarwal, and C. Zhai. Mining text data. *Springer*, 2012.

[32] C. Aggarwal, and C. Reddy. Data clustering: algorithms and applications, *CRC Press*, 2014.

[33] C. Aggarwal. Data classification: algorithms and applications. *CRC Press*, 2014.

[34] C. Aggarwal, and J. Han. Frequent pattern mining. *Springer*, 2014.

[35] C. Aggarwal. On biased reservoir sampling in the presence of stream evolution. *VLDB Conference*, pp. 607–618, 2006.

[36] C. Aggarwal. A framework for clustering massive-domain data streams. *IEEE ICDE Conference*, pp. 102–113, 2009.

[37] C. Aggarwal, and P. Yu. Online generation of association rules. *ICDE Conference*, pp. 402–411, 1998.

[38] C. Aggarwal, Z. Sun, and P. Yu. Online generation of profile association rules. *ACM KDD Conference*, pp. 129–133, 1998.

[39] C. Aggarwal, J. Han, J. Wang, and P. Yu. A framework for clustering evolving data streams, *VLDB Conference*, pp. 81–92, 2003.

[40] C. Aggarwal. Data streams: models and algorithms. *Springer*, 2007.

[41] C. Aggarwal, J. Wolf, and P. Yu. A new method for similarity indexing of market basket data. *ACM SIGMOD Conference*, pp. 407–418, 1999.

[42] C. Aggarwal, N. Ta, J. Wang, J. Feng, and M. Zaki. Xproj: A framework for projected structural clustering of XML documents. *ACM KDD Conference*, pp. 46–55, 2007.

[43] C. Aggarwal. A human-computer interactive method for projected clustering. *IEEE Transactions on Knowledge and Data Engineering*, 16(4). pp. 448–460. 2004.

[44] C. Aggarwal, and N. Li. On node classification in dynamic content-based networks. *SDM Conference*, pp. 355–366, 2011.

[45] C. Aggarwal, A. Khan, and X. Yan. On flow authority discovery in social networks. *SDM Conference*, pp. 522–533, 2011.

[46] C. Aggarwal, and P. Yu. Outlier detection for high dimensional data. *ACM SIGMOD Conference*, pp. 37–46, 2011.

[47] C. Aggarwal, and P. Yu. On classification of high-cardinality data streams. *SDM Conference*, 2010.

[48] C. Aggarwal, and P. Yu. On clustering massive text and categorical data streams. *Knowledge and information systems*, 24(2), pp. 171–196, 2010.

[49] C. Aggarwal, Y. Xie, and P. Yu. On dynamic link inference in heterogeneous networks. *SDM Conference*, pp. 415–426, 2011.

[50] C. Aggarwal, Y. Xie, and P. Yu. On dynamic data-driven selection of sensor streams. *ACM KDD Conference*, pp. 1226–1234, 2011.

[51] C. Aggarwal. On effective classification of strings with wavelets. *ACM KDD Conference*, pp. 163–172, 2002.

[52] C. Aggarwal. On abnormality detection in spuriously populated data streams. *SDM Conference*, pp. 80–91, 2005.

[53] R. Agrawal, K.-I. Lin, H. Sawhney, and K. Shim. Fast similarity search in the presence of noise, scaling, and translation in time-series databases. *VLDB Conference*, pp. 490–501, 1995.

[54] R. Agrawal, and J. Shafer. Parallel mining of association rules. *IEEE Transactions on Knowledge and Data Engineering*, 8(6), pp. 962–969, 1996. Also appears as *IBM Research Report*, RJ10004, January 1996.

[55] R. Agrawal, T. Imielinski, and A. Swami. Mining association rules between sets of items in large databases. *ACM SIGMOD Conference*, pp. 207–216, 1993.

[56] R. Agrawal, and R. Srikant. Fast algorithms for mining association rules. *VLDB Conference*, pp. 487–499, 1994.

[57] R. Agrawal, H. Mannila, R. Srikant, H. Toivonen, and A. I. Verkamo. Fast discovery of association rules. *Advances in knowledge discovery and data mining*, 12, pp. 307–328, 1996.

[58] R. Agrawal, J. Gehrke, D. Gunopulos, and P. Raghavan. Automatic subspace clustering of high dimensional data for data mining applications. *ACM SIGMOD Conference*, pp. 94–105, 1998.

[59] R. Agrawal, and R. Srikant. Mining sequential patterns. *IEEE International Conference on Data Engineering*, pp. 3–14, 1995.

[60] R. Agrawal, and R. Srikant. Privacy-preserving data mining. *ACM SIGMOD Conference*, pp. 439–450, 2000.

[61] M. Agyemang, K. Barker, and R. Alhajj. A comprehensive survey of numeric and symbolic outlier mining techniques. *Intelligent Data Analysis*, 10(6). pp. 521–538, 2006.

[62] R. Ahuja, T. Magnanti, and J. Orlin. Network flows: theory, algorithms, and applications. *Prentice Hall*, Englewood Cliffs, New Jersey, 1993.

[63] M. Al Hasan, and M. J. Zaki. A survey of link prediction in social networks. *Social network data analytics*, Springer, pp. 243–275, 2011.

[64] M. Al Hasan, V. Chaoji, S. Salem, and M. Zaki. Link prediction using supervised learning. *SDM Workshop on Link Analysis, Counter-terrorism and Security*, 2006.

[65] S. Anand, and B. Mobasher. Intelligent techniques for web personalization. *International conference on Intelligent Techniques for Web Personalization*, pp. 1–36, 2003.

[66] F. Angiulli, and C. Pizzuti. Fast Outlier detection in high dimensional spaces. *European Conference on Principles of Knowledge Discovery and Data Mining*, pp. 15–27, 2002.

[67] F. Angiulli, and F. Fassetti. Detecting distance-based outliers in streams of data. *ACM CIKM Conference*, pp. 811–820, 2007.

[68] L. Akoglu, H. Tong, J. Vreeken, and C. Faloutsos. Fast and reliable anomaly detection in categorical data. *ACM CIKM Conference*, pp. 415–424, 2012.

[69] R. Albert, and A. L. Barabasi. Statistical mechanics of complex networks. *Reviews of modern physics* 74, 1, 47, 2002.

[70] R. Albert, and A. L. Barabasi. Topology of evolving networks: local events and universality. *Physical review letters* 85, 24, pp. 5234–5237, 2000.

[71] P. Allison. Missing data. *Sage*, 2001.

[72] N. Alon, Y. Matias, and M. Szegedy. The space complexity of approximating the frequency moments. *ACM PODS Conference*, pp. 20–29, 1996.

[73] S. Altschul, T. Madden, A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D. Lipman. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic acids research*, 25(17), pp. 3389–3402, 1997.

[74] M. R. Anderberg. Cluster Analysis for Applications. *Academic Press*, New York, 1973.

[75] P. Andritsos, P. Tsaparas, R. J. Miller, and K. C. Sevcik. LIMBO: Scalable clustering of categorical data. *EDBT Conference*, pp. 123–146, 2004.

[76] M. Ankerst, M. M. Breunig, H.-P. Kriegel, and J. Sander. OPTICS: ordering points to identify the clustering structure. *ACM SIGMOD Conference*, pp. 49–60, 1999.

[77] A. Apostolico, and C. Guerra. The longest common subsequence problem revisited. *Algorithmica*, 2(1–4), pp. 315–336, 1987.

[78] A. Azran. The rendezvous algorithm: Multiclass semi-supervised learning with markov random walks. *International Conference on Machine Learning*, pp. 49–56, 2007.

[79] A. Banerjee, S. Merugu, I. S. Dhillon, and J. Ghosh. Clustering with Bregman divergences. *Journal of Machine Learning Research*, 6, pp. 1705–1749, 2005.

[80] S. Basu, A. Banerjee, and R. J. Mooney. Semi-supervised clustering by seeding. *ICML Conference*, pp. 27–34, 2002.

[81] S. Basu, M. Bilenko, and R. J. Mooney. A probabilistic framework for semi-supervised clustering. *ACM KDD Conference*, pp. 59–68, 2004.

[82] R. J. Bayardo Jr. Efficiently mining long patterns from databases. *ACM SIGMOD*, pp. 85–93, 1998.

[83] R. J. Bayardo, and R. Agrawal. Data privacy through optimal k-anonymization. *IEEE International Conference on Data Engineering*, pp. 217–228, 2005.

[84] R. Beckman, and R. Cook. Outliers. *Technometrics*, 25(2), pp. 119–149, 1983.

[85] A. Ben-Hur, C. S. Ong, S. Sonnenburg, B. Scholkopf, and G. Ratsch. Support vector machines and kernels for computational biology. *PLoS computational biology*, 4(10), e1000173, 2008.

[86] M. Benkert, J. Gudmundsson, F. Hubner, and T. Wolle. Reporting flock patterns. *COMGEO*, 2008

[87] D. Berndt, and J. Clifford. Using dynamic time warping to find patterns in time series. *KDD Workshop*, 10(16), pp. 359–370, 1994.

[88] K. Beyer, J. Goldstein, R. Ramakrishnan, and U. Shaft. When is "nearest neighbor" meaningful? *International Conference on Database Theory*, pp. 217–235, 1999.

[89] V. Barnett, and T. Lewis. Outliers in statistical data. *Wiley*, 1994.

[90] M. Belkin, and P. Niyogi. Laplacian eigenmaps and spectral techniques for embedding and clustering. *NIPS*, pp. 585–591, 2001.

[91] M. Bezzi, S. De Capitani di Vimercati, S. Foresti, G. Livraga, P. Samarati, and R. Sassi. Modeling and preventing inferences from sensitive value distributions in data release. *Journal of Computer Security*, 20(4), pp. 393–436, 2012.

[92] L. Bergroth, H. Hakonen, and T. Raita. A survey of longest common subsequence algorithms. *String Processing and Information Retrieval*, 2000.

[93] S. Bhagat, G. Cormode, and S. Muthukrishnan. Node classification in social networks. *Social Network Data Analytics*, Springer, pp. 115–148. 2011.

[94] M. Bilenko, S. Basu, and R. J. Mooney. Integrating constraints and metric learning in semi-supervised clustering. *ICML Conference*, 2004.

[95] C. M. Bishop. Pattern recognition and machine learning. *Springer*, 2007.

[96] C. M. Bishop. Neural networks for pattern recognition. *Oxford University Press*, 1995.

[97] C. M. Bishop. Improving the generalization properties of radial basis function neural networks. *Neural Computation*, 3(4), pp. 579–588, 1991.

[98] D. Blei, A. Ng, and M. Jordan. Latent dirichlet allocation. *Journal of Machine Learning Research*, 3: pp. 993–1022, 2003.

[99] D. Blei. Probabilistic topic models. *Communications of the ACM*, 55(4), pp. 77–84, 2012.

[100] A. Blum, and T. Mitchell. Combining labeled and unlabeled data with co-training. *Proceedings of Conference on Computational Learning Theory*, 1998.

[101] A. Blum, and S. Chawla. Combining labeled and unlabeled data with graph mincuts. *ICML Conference*, 2001.

[102] C. Bohm, K. Haegler, N. Muller, and C. Plant. Coco: coding cost for parameter free outlier detection. *ACM KDD Conference*, 2009.

[103] K. Borgwardt, and H.-P. Kriegel. Shortest-path kernels on graphs. *IEEE International Conference on Data Mining*, 2005.

[104] S. Boriah, V. Chandola, and V. Kumar. Similarity measures for categorical data: A comparative evaluation. *SIAM Conference on Data Mining*, 2008.

[105] L. Bottou, and V. Vapnik. Local learning algorithms. *Neural Computation*, 4(6), pp. 888–900, 1992.

[106] L. Bottou, C. Cortes, J. S. Denker, H. Drucker, I. Guyon, L. Jackel, Y. LeCun, U. A. Müller, E. Säckinger, P. Simard, and V. Vapnik. Comparison of classifier methods: a case study in handwriting digit recognition. *International Conference on Pattern Recognition*, pp. 77–87, 1994.

[107] J. Boulicaut, A. Bykowski, and C. Rigotti. Approximation of frequency queries by means of free-sets. *Principles of Data Mining and Knowledge Discovery*, pp. 75–85, 2000.

[108] P. Bradley, and U. Fayyad. Refining initial points for k-means clustering. *ICML Conference*, pp. 91–99, 1998.

[109] M. Breunig, H.-P. Kriegel, R. Ng, and J. Sander. LOF: Identifying density-based local outliers. *ACM SIGMOD Conference*, 2000.

[110] L. Breiman, J. Friedman, C. Stone, and R. Olshen. Classification and regression trees. *CRC press*, 1984.

[111] L. Breiman. Random forests. *Machine Learning*, 45(1), pp. 5–32, 2001.

[112] L. Breiman. Bagging predictors. *Machine Learning*, 24(2), pp. 123–140, 1996.

[113] S. Brin, R. Motwani, and C. Silverstein. Beyond market baskets: generalizing association rules to correlations. *ACM SIGMOD Conference*, pp. 265–276, 1997.

[114] S. Brin, and L. Page. The anatomy of a large-scale hypertextual web search engine. *Computer Networks*, 30(1–7), pp. 107–117, 1998.

[115] B. Bringmann, S. Nijssen, and A. Zimmermann. Pattern-based classification: A unifying perspective. *arXiv preprint, arXiv:1111.6191*, 2011.

[116] C. Brodley, and P. Utgoff. Multivariate decision trees. *Machine learning*, 19(1), pp. 45–77, 1995.

[117] Y. Bu, L. Chen, A. W.-C. Fu, and D. Liu. Efficient anomaly monitoring over moving object trajectory streams. *ACM KDD Conference*, pp. 159–168, 2009.

[118] M. Bulmer. Principles of Statistics. *Dover Publications*, 1979.

[119] H. Bunke. On a relation between graph edit distance and maximum common subgraph. *Pattern Recognition Letters*, 18(8), pp. 689–694, 1997.

[120] H. Bunke, and K. Shearer. A graph distance metric based on the maximal common subgraph.*Pattern recognition letters*, 19(3), pp. 255–259, 1998.

[121] W. Buntine. Learning Classification Trees. *Artificial intelligence frontiers in statistics*. Chapman and Hall, pp. 182–201, 1993.

[122] T. Burnaby. On a method for character weighting a similarity coefficient employing the concept of information. *Mathematical Geology*, 2(1), 25–38, 1970.

[123] D. Burdick, M. Calimlim, and J. Gehrke. MAFIA: A maximal frequent itemset algorithm for transactional databases. *IEEE International Conference on Data Engineering*, pp. 443–452, 2001.

[124] C. Burges. A tutorial on support vector machines for pattern recognition. *Data mining and knowledge discovery*, 2(2), pp. 121–167, 1998.

[125] T. Calders, and B. Goethals. Mining all non-derivable frequent itemsets. *Principles of Knowledge Discovery and Data Mining*, pp. 74–86, 2002.

[126] T. Calders, C. Rigotti, and J. F. Boulicaut. A survey on condensed representations for frequent sets. In *Constraint-based mining and inductive databases*, pp. 64–80, Springer, 2006.

[127] S. Chakrabarti. Mining the Web: Discovering knowledge from hypertext data. *Morgan Kaufmann*, 2003.

[128] S. Chakrabarti, B. Dom, and P. Indyk. Enhanced hypertext categorization using hyperlinks. *ACM SIGMOD Conference*, pp. 307–318, 1998.

[129] S. Chakrabarti, S. Sarawagi, and B. Dom. Mining surprising patterns using temporal description length. *VLDB Conference*, pp. 606–617, 1998.

[130] K. P. Chan, and A. W. C. Fu. Efficient time series matching by wavelets.*IEEE International Conference on Data Engineering*, pp. 126–133, 1999.

[131] V. Chandola, A. Banerjee, and V. Kumar. Anomaly detection: A survey. *ACM Computing Surveys*, 41(3), 2009.

[132] V. Chandola, A. Banerjee, and V. Kumar. Anomaly detection for discrete sequences: A survey. *IEEE Transactions on Knowledge and Data Engineering*, 24(5), pp. 823–839, 2012.

[133] O. Chapelle. Training a support vector machine in the primal. *Neural Computation*, 19(5), pp. 1155–1178, 2007.

[134] C. Chatfield. The analysis of time series: an introduction. *CRC Press*, 2003.

[135] A. Chaturvedi, P. Green, and J. D. Carroll. K-modes clustering, *Journal of Classification*, 18(1), pp. 35–55, 2001.

[136] N. V. Chawla, N. Japkowicz, and A. Kotcz. Editorial: Special issue on learning from imbalanced data sets. *ACM SIGKDD Explorations Newsletter*, 6(1), 1–6, 2004.

[137] N. V. Chawla, K. W. Bower, L. O. Hall, and W. P. Kegelmeyer. SMOTE: synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research (JAIR)*, 16, pp. 321–356, 2002.

[138] N. Chawla, A. Lazarevic, L. Hall, and K. Bowyer. SMOTEBoost: Improving prediction of the minority class in boosting. *PKDD*, pp. 107–119, 2003.

[139] N. V. Chawla, D. A. Cieslak, L. O. Hall, and A. Joshi. Automatically countering imbalance and its empirical relationship to cost. *Data Mining and Knowledge Discovery*, 17(2), pp. 225–252, 2008.

[140] K. Chen, and L. Liu. A survey of multiplicative perturbation for privacy-preserving data mining. *Privacy-Preserving Data Mining: Models and Algorithms*, Springer, pp. 157–181, 2008.

[141] L. Chen, and R. Ng. On the marriage of L_p-norms and the edit distance. *VLDB Conference*, pp. 792–803, 2004.

[142] W. Chen, Y. Wang, and S. Yang. Efficient influence maximization in social networks. *ACM KDD Conference*, pp. 199–208, 2009.

[143] W. Chen, C. Wang, and Y. Wang. Scalable influence maximization for prevalent viral marketing in large-scale social networks. *ACM KDD Conference*, pp. 1029–1038, 2010.

[144] W. Chen, Y. Yuan, and L. Zhang. Scalable influence maximization in social networks under the linear threshold model. *IEEE International Conference on Data Mining*, pp. 88–97, 2010.

[145] D. Chen, C.-T. Lu, Y. Chen, and D. Kou. On detecting spatial outliers. *Geoinformatica*, 12: pp. 455–475, 2008.

[146] T. Cheng, and Z. Li. A hybrid approach to detect spatialtemporal outliers. *International Conference on Geoinformatics*, pp. 173–178, 2004.

[147] T. Cheng, and Z. Li. A multiscale approach for spatio-temporal outlier detection. *Transactions in GIS*, 10(2), pp. 253–263, March 2006.

[148] Y. Cheng. Mean shift, mode seeking, and clustering. *IEEE Transactions on PAMI*, 17(8), pp. 790–799, 1995.

[149] H. Cheng, X. Yan, J. Han, and C. Hsu. Discriminative frequent pattern analysis for effective classification. *ICDE Conference*, pp. 716–725, 2007.

[150] F. Y. Chin, and G. Ozsoyoglu. Auditing and inference control in statistical databases. *IEEE Transactions on Software Enginerring*, 8(6), pp. 113–139, April 1982.

[151] B. Chiu, E. Keogh, and S. Lonardi. Probabilistic discovery of time series motifs. *ACM KDD Conference*, pp. 493–498, 2003.

[152] F. Chung. Spectral Graph Theory. *Number 92 in CBMS Conference Series in Mathematics, American Mathematical Society*, 1997.

[153] V. Ciriani, S. De Capitani di Vimercati, S. Foresti, and P. Samarati. k-anonymous data mining: A survey. *Privacy-preserving data mining: models and algorithms*, Springer, pp. 105–136, 2008.

[154] C. Clifton, M. Kantarcioglu, J. Vaidya, X. Lin, and M. Y. Zhu. Tools for privacy preserving distributed data mining. *ACM SIGKDD Explorations Newsletter*, 4(2), pp. 28–34, 2002.

[155] N. Cristianini, and J. Shawe-Taylor. An introduction to support vector machines and other kernel-based learning methods. *Cambridge University Press*, 2000.

[156] W. Cochran. Sampling techniques. *John Wiley and Sons*, 2007.

[157] D. Cohn, L. Atlas, and R. Ladner. Improving generalization with active learning. *Machine Learning*, 5(2), pp. 201–221, 1994.

[158] D. Cohn, Z. Ghahramani, and M. Jordan. Active learning with statistical models. *Journal of Artificial Intelligence Research*, 4, pp. 129–145, 1996.

[159] D. Comaniciu, and P. Meer. Mean shift: A robust approach toward feature space analysis. *IEEE Transactions on PAMI*, 24(5), pp. 603–619, 2002.

[160] D. Cook, and L. Holder. Graph-based data mining. *IEEE Intelligent Systems*, 15(2), pp. 32–41, 2000.

[161] R. Cooley, B. Mobasher, and J. Srivastava. Data preparation for mining world wide web browsing patterns. *Knowledge and information systems*, 1(1), pp. 5–32, 1999.

[162] L. P. Cordella, P. Foggia, C. Sansone, and M. Vento. A (sub)graph isomorphism algorithm for matching large graphs. *IEEE Transactions on Pattern Mining and Machine Intelligence*, 26(10), pp. 1367–1372, 2004.

[163] H. Shang, Y. Zhang, X. Lin, and J. X. Yu. Taming verification hardness: an efficient algorithm for testing subgraph isomorphism. *Proceedings of the VLDB Endowment*, 1(1), pp. 364–375, 2008.

[164] J. R. Ullmann. An algorithm for subgraph isomorphism. *Journal of the ACM*, 23: pp. 31–42, January 1976.

[165] G. Cormode, and S. Muthukrishnan. An improved data stream summary: the count-min sketch and its applications. *Journal of Algorithms*, 55(1), pp. 58–75, 2005.

[166] S. Cost, and S. Salzberg. A weighted nearest neighbor algorithm for learning with symbolic features. *Machine Learning*, 10(1), pp. 57–78, 1993.

[167] T. Cover, and P. Hart. Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, 13(1), pp. 21–27, 1967.

[168] D. Cutting, D. Karger, J. Pedersen, and J. Tukey. Scatter/gather: A cluster-based approach to browsing large document collections. *ACM SIGIR Conference*, pp. 318–329, 1992.

[169] M. Dash, K. Choi, P. Scheuermann, and H. Liu. Feature selection for clustering-a filter solution. *ICDM Conference*, pp. 115–122, 2002.

[170] M. Deshpande, and G. Karypis. Item-based top-n recommendation algorithms. *ACM Transactions on Information Systems (TOIS)*, 22(1), pp. 143–177, 2004.

[171] I. Dhillon. Co-clustering documents and words using bipartite spectral graph partitioning, *ACM KDD Conference*, pp. 269–274, 2001.

[172] I. Dhillon, S. Mallela, and D. Modha. Information-theoretic co-clustering. *ACM KDD Conference*, pp. 89–98, 2003.

[173] I. Dhillon, Y. Guan, and B. Kulis. Kernel k-means: spectral clustering and normalized cuts. *ACM KDD Conference*, pp. 551–556, 2004.

[174] P. Domingos. MetaCost: A general framework for making classifiers cost-sensitive. *ACM KDD Conference*, pp. 155–164, 1999.

[175] P. Domingos. Bayesian averaging of classifiers and the overfitting problem. *ICML Conference*, pp. 223–230, 2000.

[176] P. Domingos, and G. Hulten. Mining high-speed data streams. *ACM KDD Conference*, pp. 71–80. 2000.

[177] P. Clark, and T. Niblett. The CN2 induction algorithm. *Machine Learning*, 3(4), pp. 261–283, 1989.

[178] W. W. Cohen. Fast effectve rule induction. *ICML Conference*, pp. 115–123, 1995.

[179] L. H. Cox. Suppression methodology and statistical disclosure control. *Journal of the American Statistical Association*, 75(370), pp. 377–385, 1980.

[180] E. Cohen, M. Datar, S. Fujiwara, A. Gionis, P. Indyk, R. Motwani, and C. Yang. Finding interesting associations without support pruning. *IEEE Transactions on Knowledge and Data Engineering*, 13(1), pp. 64–78, 2001.

[181] T. Dalenius, and S. Reiss. Data-swapping: A technique for disclosure control. *Journal of statistical planning and inference*, 6(1), pp. 73–85, 1982.

[182] G. Das, and H. Mannila. Context-based similarity measures for categorical databases. *PKDD Conference*, pp. 201–210, 2000.

[183] B. V. Dasarathy. Nearest neighbor (NN) norms: NN pattern classification techniques. *IEEE Computer Society Press*, 1990,

[184] S. Deerwester, S. Dumais, T. Landauer, G. Furnas, and R. Harshman. Indexing by latent semantic analysis. *JASIS*, 41(6), pp. 391–407, 1990.

[185] C. Ding, X. He, and H. Simon. On the equivalence of nonnegative matrix factorization and spectral clustering. *SDM Conference*, pp. 606–610, 2005.

[186] J. Domingo-Ferrer, and J. M. Mateo-Sanz. Practical data-oriented microaggregation for statistical disclosure control. *IEEE Transactions on Knowledge and Data Engineering*, 14(1), pp. 189–201, 2002.

[187] P. Domingos, and M. Pazzani. On the optimality of the simple bayesian classifier under zero-one loss. *Machine Learning*, 29(2–3), pp. 103–130, 1997.

[188] W. Du, and M. Atallah. Secure multi-party computation: A review and open problems. *CERIAS Tech. Report*, 2001-51, Purdue University, 2001.

[189] R. Duda, P. Hart, and D. Stork. Pattern classification. *John Wiley and Sons*, 2012.

[190] C. Dwork. Differential privacy: A survey of results. *Theory and Applications of Models of Computation*, Springer, pp. 1–19, 2008.

[191] C. Dwork. A firm foundation for private data analysis. *Communications of the ACM*, 54(1), pp. 86–95, 2011.

[192] D. Easley, and J. Kleinberg. Networks, crowds, and markets: Reasoning about a highly connected world. *Cambridge University Press*, 2010.

[193] C. Elkan. The foundations of cost-sensitive learning. *IJCAI*, pp. 973–978, 2001.

[194] R. Elmasri, and S. Navathe. *Fundamentals of Database Systems*. Addison-Wesley, 2010.

[195] L. Ertoz, M. Steinbach, and V. Kumar. A new shared nearest neighbor clustering algorithm and its applications. *Workshop on Clustering High Dimensional Data and its Applications*, pp. 105–115, 2002.

[196] P. Erdos, and A. Renyi. On random graphs. *Publicationes Mathematicae Debrecen*, 6, pp. 290–297, 1959.

[197] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu. A density-based algorithm for discovering clusters in large spatial databases with noise. *ACM KDD Conference*, pp. 226–231, 1996.

[198] M. Ester, H. P. Kriegel, J. Sander, M. Wimmer, and X. Xu. Incremental clustering for mining in a data warehousing environment. *VLDB Conference*, pp. 323–333, 1998.

[199] S. Even, O. Goldreich, and A. Lempel. A randomized protocol for signing contracts. *Communications of the ACM*, 28(6), pp. 637–647, 1985.

[200] A. Evfimievski, R. Srikant, R. Agrawal, and J. Gehrke. Privacy preserving mining of association rules. *Information Systems*, 29(4), pp. 343–364, 2004.

[201] M. Faloutsos, P. Faloutsos, and C. Faloutsos. On power-law relationships of the internet topology. *ACM SIGCOMM Computer Communication Review*, pp. 251–262, 1999.

[202] C. Faloutsos, and K. I. Lin. Fastmap: A fast algorithm for indexing, data-mining and visualization of traditional and multimedia datasets. *ACM SIGMOD Conference*, pp. 163–174, 1995.

[203] W. Fan, S. Stolfo, J. Zhang, and P. Chan. AdaCost: Misclassification cost sensitive boosting. *ICML Conference*, pp. 97–105, 1999.

[204] T. Fawcett. ROC Graphs: Notes and Practical Considerations for Researchers. *Technical Report HPL-2003-4*, Palo Alto, CA, HP Laboratories, 2003.

[205] X. Fern, and C. Brodley. Random projection for high dimensional data clustering: A cluster ensemble approach. *ICML Conference*, pp. 186–193, 2003.

[206] C. Fiduccia, and R. Mattheyses. A linear-time heuristic for improving network partitions. In *IEEE Conference on Design Automation*, pp. 175–181, 1982.

[207] R. Fisher. The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, 7: pp. 179–188, 1936.

[208] P. Flajolet, and G. N. Martin. Probabilistic counting algorithms for data base applications. *Journal of Computer and System Sciences*, 31(2), pp. 182–209, 1985.

[209] G. W. Flake. Square unit augmented, radially extended, multilayer perceptrons. *Neural Networks: Tricks of the Trade*, pp. 145–163, 1998.

[210] F. Fouss, A. Pirotte, J. Renders, and M. Saerens. Random-walk computation of similarities between nodes of a graph with application to collaborative recommendation. *IEEE Transactions on Knowledge and Data Engineering*, 19(3), pp. 355–369, 2007.

[211] S. Forrest, C. Warrender, and B. Pearlmutter. Detecting intrusions using system calls: alternate data models. *IEEE ISRSP*, 1999.

[212] S. Fortunato. Community Detection in Graphs. *Physics Reports*, 486(3–5), pp. 75–174, February 2010.

[213] A. Frank, and A. Asuncion. UCI Machine Learning Repository, Irvine, CA: University of California, School of Information and Computer Science, 2010. `http://archive.ics.uci.edu/ml`

[214] E. Frank, M. Hall, and B. Pfahringer. Locally weighted naive bayes. *Proceedings of the Nineteenth conference on Uncertainty in Artificial Intelligence*, pp, 249–256, 2002.

[215] Y. Freund, and R. Schapire. A decision-theoretic generalization of online learning and application to boosting. *Computational Learning Theory*, pp. 23–37, 1995.

[216] J. Friedman. Flexible nearest neighbor classification. *Technical Report, Stanford University*, 1994.

[217] J. Friedman, R. Kohavi, and Y. Yun. Lazy decision trees. *Proceedings of the National Conference on Artificial Intelligence*, pp. 717–724, 1996.

[218] B. Fung, K. Wang, R. Chen, and P. S. Yu. Privacy-preserving data publishing: A survey of recent developments. *ACM Computing Surveys (CSUR)*, 42(4), 2010.

[219] G. Gan, C. Ma, and J. Wu. Data clustering: theory, algorithms, and applications. *SIAM*, 2007.

[220] V. Ganti, J. Gehrke, and R. Ramakrishnan. CACTUS: Clustering categorical data using summaries. *ACM KDD Conference*, pp. 73–83, 1999.

[221] M. Garey, and D. S. Johnson. Computers and intractability: A guide to the theory of NP-completeness. *New York, Freeman*, 1979.

[222] H. Galhardas, D. Florescu, D. Shasha, and E. Simon. AJAX: an extensible data cleaning tool. *ACM SIGMOD Conference* 29(2), pp. 590, 2000.

[223] J. Gao, and P.-N. Tan. Converting output scores from outlier detection algorithms into probability estimates. *ICDM Conference*, pp. 212–221, 2006.

[224] M. Garofalakis, R. Rastogi, and K. Shim. SPIRIT: Sequential pattern mining with regular expression constraints. *VLDB Conference*, pp. 7–10, 1999.

[225] T. Gartner, P. Flach, and S. Wrobel. On graph kernels: Hardness results and efficient alternatives. *COLT: Kernel 2003 Workshop Proceedings*, pp. 129–143, 2003.

[226] Y. Ge, H. Xiong, Z.-H. Zhou, H. Ozdemir, J. Yu, and K. Lee. Top-Eye: Top-k evolving trajectory outlier detection. *CIKM Conference*, pp. 1733–1736, 2010.

[227] J. Gehrke, V. Ganti, R. Ramakrishnan, and W.-Y. Loh. BOAT: Optimistic decision tree construction. *ACM SIGMOD Conference*, pp. 169–180, 1999.

[228] J. Gehrke, R. Ramakrishnan, and V. Ganti. Rainforest-a framework for fast decision tree construction of large datasets. *VLDB Conference*, pp. 416–427, 1998.

[229] D. Gibson, J. Kleinberg, and P. Raghavan. Clustering categorical data: an approach based on dynamical systems. *The VLDB Journal*, 8(3), pp. 222–236, 2000.

[230] M. Girvan, and M. Newman. Community structure in social and biological networks. *Proceedings of the National Academy of Sciences*, 99(12), pp. 7821–7826.

[231] S. Goil, H. Nagesh, and A. Choudhary. MAFIA: Efficient and scalable subspace clustering for very large data sets. *ACM KDD Conference*, pp. 443–452, 1999.

[232] D. W. Goodall. A new similarity index based on probability. *Biometrics*, 22(4), pp. 882–907, 1966.

[233] K. Gouda, and M. J. Zaki. Genmax: An efficient algorithm for mining maximal frequent itemsets. *Data Mining and Knowledge Discovery*, 11(3), pp. 223–242, 2005.

[234] A. Goyal, F. Bonchi, and L. V. S. Lakshmanan. A data-based approach to social influence maximization. *VLDB Conference*, pp. 73–84, 2011.

[235] A. Goyal, F. Bonchi, and L. V. S. Lakshmanan. Learning influence probabilities in social networks. *ACM WSDM Conference*, pp. 241–250, 2011.

[236] R. Gozalbes, J. P. Doucet, and F. Derouin. Application of topological descriptors in QSAR and drug design: history and new trends. *Current Drug Targets-Infectious Disorders*, 2(1), pp. 93–102, 2002.

[237] M. Gupta, J. Gao, C. Aggarwal, and J. Han. Outlier detection for temporal data. Morgan and Claypool, 2014.

[238] S. Guha, R. Rastogi, and K. Shim. ROCK: A robust clustering algorithm for categorical attributes. *Information Systems*, 25(5), pp. 345–366, 2000.

[239] S. Guha, R. Rastogi, and K. Shim. CURE: An efficient clustering algorithm for large databases. *ACM SIGMOD Conference*, pp. 73–84, 1998.

[240] S. Guha, A. Meyerson, N. Mishra, R. Motwani, and L. O'Callaghan. Clustering data streams: Theory and practice. *IEEE Transactions on Knowledge and Data Engineering*, 15(3), pp. 515–528, 2003.

[241] D. Gunopulos, and G. Das. Time series similarity measures and time series indexing. *ACM SIGMOD Conference*, pp, 624, 2001.

[242] V. Guralnik, and G. Karypis. A scalable algorithm for clustering sequential data. *IEEE International Conference on Data Engineering*, pp. 179–186, 2001.

[243] V. Guralnik, and G. Karypis. Parallel tree-projection-based sequence mining algorithms. *Parallel Computing*, 30(4): pp. 443–472, April 2004. Also appears in *European Conference in Parallel Processing*, 2001.

[244] D. Gusfield. Algorithms on strings, trees and sequences. *Cambridge University Press*, 1997.

[245] I. Guyon (Ed.). Feature extraction: foundations and applications. *Springer*, 2006.

[246] I. Guyon, and A. Elisseeff. An introduction to variable and feature selection. *Journal of Machine Learning Research*, 3, pp. 1157–1182, 2003.

[247] M. Halkidi, Y. Batistakis, and M. Vazirgiannis. Cluster validity methods: part I. *ACM SIGMOD record*, 31(2), pp. 40–45, 2002.

[248] M. Halkidi, Y. Batistakis, and M. Vazirgiannis. Clustering validity checking methods: part II. *ACM SIGMOD Record*, 31(3), pp. 19–27, 2002.

[249] E. Han, and G. Karypis. Centroid-based document classification: analysis and experimental results. *ECML Conference*, pp. 424–431, 2000.

[250] J. Han, M. Kamber, and J. Pei. Data mining: concepts and techniques. *Morgan Kaufmann*, 2011.

[251] J. Han, G. Dong, and Y. Yin. Efficient mining of partial periodic patterns in time series database. *International Conference on Data Engineering*, pp. 106–115, 1999.

[252] J. Han, J. Pei, and Y. Yin. Mining frequent patterns without candidate generation. *ACM SIGMOD Conference*, pp. 1–12, 2000.

[253] J. Han, H. Cheng, D. Xin, and X. Yan. Frequent pattern mining: current status and future directions. *Data Mining and Knowledge Discovery*, 15(1), pp. 55–86, 2007.

[254] J. Haslett, R. Brandley, P. Craig, A. Unwin, and G. Wills. Dynamic graphics for exploring spatial data with application to locating global and local anomalies. *The American Statistician*, 45: pp. 234–242, 1991.

[255] T. Hastie, and R. Tibshirani. Discriminant adaptive nearest neighbor classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(6), pp. 607–616, 1996.

[256] T. Hastie, R. Tibshirani, and J. Friedman. The elements of statistical learning. *Springer*, 2009.

[257] V. Hautamaki, V. Karkkainen, and P. Franti. Outlier detection using k-nearest neighbor graph. *International Conference on Pattern Recognition*, pp. 430–433, 2004.

[258] T. H. Haveliwala. Topic-sensitive pagerank. *World Wide Web Conference*, pp. 517-526, 2002.

[259] D. M. Hawkins. Identification of outliers. *Chapman and Hall*, 1980.

[260] S. Haykin. Kalman filtering and neural networks. *Wiley*, 2001.

[261] S. Haykin. Neural networks and learning machines. *Prentice Hall*, 2008.

[262] X. He, D. Cai, and P. Niyogi. Laplacian score for feature selection. *Advances in Neural Information Processing Systems*, 18, 507, 2006.

[263] Z. He, X. Xu, J. Huang, and S. Deng. FP-Outlier: Frequent pattern-based outlier detection. *COMSIS*, 2(1), pp. 103–118, 2005.

[264] Z. He, X. Xu, and S. Deng. Discovering cluster-based local outliers, *Pattern Recognition Letters*, Vol 24(9–10), pp. 1641–1650, 2003.

[265] M. Henrion, D. Hand, A. Gandy, and D. Mortlock. CASOS: A subspace method for anomaly detection in high-dimensional astronomical databases. *Statistical Analysis and Data Mining*, 2012.
Online first: `http://onlinelibrary.wiley.com/enhanced/doi/10.1002/sam.11167/`

[266] A. Hinneburg, C. Aggarwal, and D. Keim. What is the nearest neighbor in high-dimensional space? *VLDB Conference*, pp. 506–516, 2000.

[267] A. Hinneburg, and D. Keim. An efficient approach to clustering in large multimedia databases with noise. *ACM KDD Conference*, pp. 58–65, 1998.

[268] A. Hinneburg, D. A. Keim, and M. Wawryniuk. HD-Eye: Visual mining of high-dimensional data. *Computer Graphics and Applications*, 19(5), pp. 22–31, 1999.

[269] A. Hinneburg, and H. Gabriel. DENCLUE 2.0: Fast clustering based on kernel-density estimation. *Intelligent Data Analysis*, *Springer*, pp. 70–80, 2007.

[270] D. S. Hirschberg. Algorithms for the longest common subsequence problem. *Journal of the ACM (JACM)*, 24(4), pp. 664–675, 1975.

[271] T. Hofmann. Probabilistic latent semantic indexing. *ACM SIGIR Conference*, pp. 50–57, 1999.

[272] T. Hofmann. Latent semantic models for collaborative filtering. *ACM Transactions on Information Systems (TOIS)*, 22(1), pp. 89–114, 2004.

[273] M. Holsheimer, M. Kersten, H. Mannila, and H. Toivonen. A perspective on databases and data mining, *ACM KDD Conference*, pp. 150–155, 1995.

[274] S. Hofmeyr, S. Forrest, and A. Somayaji. Intrusion detection using sequences of system calls. *Journal of Computer Security*, 6(3), pp. 151–180, 1998.

[275] D. Hosmer Jr., S. Lemeshow, and R. Sturdivant. Applied logistic regression. *Wiley*, 2013.

[276] J. Huan, W. Wang, and J. Prins. Efficient mining of frequent subgraphs in the presence of isomorphism. *IEEE ICDM Conference*, pp. 549–552, 2003.

[277] Z. Huang, X. Li, and H. Chen. Link prediction approach to collaborative filtering. *ACM/IEEE-CS joint conference on Digital libraries*, pp. 141–142, 2005.

[278] Z. Huang, and M. Ng. A fuzzy k-modes algorithm for clustering categorical data. *IEEE Transactions on Fuzzy Systems*, 7(4), pp. 446–452, 1999.

[279] G. Hulten, L. Spencer, and P. Domingos. Mining time-changing data streams. *ACM KDD Conference*, pp. 97–106, 2001.

[280] J. W. Hunt, and T. G. Szymanski. A fast algorithm for computing longest common subsequences. *Communications of the ACM*, 20(5), pp. 350–353, 1977.

[281] Y. S. Hwang, and S. Y. Bang. An efficient method to construct a radial basis function neural network classifier. *Neural Networks*, 10(8), pp. 1495–1503, 1997.

[282] A. Inokuchi, T. Washio, and H. Motoda. An apriori-based algorithm on mining frequent substructures from graph data. *Principles on Knowledge Discovery and Data Mining*, pp. 13–23, 2000.

[283] H. V. Jagadish, A. O. Mendelzon, and T. Milo. Similarity-based queries. *ACM PODS Conference*, pp. 36–45, 1995.

[284] A. K. Jain, and R. C. Dubes. Algorithms for clustering data. *Prentice-Hall, Inc.*, 1998.

[285] A. Jain, M. Murty, and P. Flynn. Data clustering: A review. *ACM Computing Surveys (CSUR)*, 31(3):264–323, 1999.

[286] A. Jain, R. Duin, and J. Mao. Statistical pattern recognition: A review. *IEEE Transactions on Pattern Analysis and Machine Intelligence,*, 22(1), pp. 4–37, 2000.

[287] V. Janeja, and V. Atluri. Random walks to identify anomalous free-form spatial scan windows. *IEEE Transactions on Knowledge and Data Engineering*, 20(10), pp. 1378–1392, 2008.

[288] J. Rennie, and N. Srebro. Fast maximum margin matrix factorization for collaborative prediction. *ICML Conference*, pp. 713–718, 2005.

[289] G. Jeh, and J. Widom. SimRank: a measure of structural-context similarity. *ACM KDD Conference*, pp. 538–543, 2003.

[290] H. Jeung, M. L. Yiu, X. Zhou, C. Jensen, and H. Shen. Discovery of convoys in trajectory databases. *VLDB Conference*, pp. 1068–1080, 2008.

[291] T. Joachims. Making Large scale SVMs practical. *Advances in Kernel Methods, Support Vector Learning*, pp. 169–184, *MIT Press*, Cambridge, 1998.

[292] T. Joachims. Training Linear SVMs in Linear Time. *ACM KDD Conference*, pp. 217–226, 2006.

[293] T. Joachims. Transductive inference for text classification using support vector machines. *International Conference on Machine Learning*, pp. 200–209, 1999.

[294] T. Joachims. Transductive learning via spectral graph partitioning. *ICML Conference*, pp. 290–297, 2003.

[295] I. Jolliffe. Principal component analysis. *John Wiley and Sons*, 2005.

[296] M. Joshi, V. Kumar, and R. Agarwal. Evaluating boosting algorithms to classify rare classes: comparison and improvements. *IEEE ICDM Conference*, pp. 257–264, 2001.

[297] M. Kantarcioglu. A survey of privacy-preserving methods across horizontally partitioned data. *Privacy-Preserving Data Mining: Models and Algorithms*, Springer, pp. 313–335, 2008.

[298] H. Kashima, K. Tsuda, and A. Inokuchi. Kernels for graphs. In *Kernel Methods in Computational Biology*, MIT Press, Cambridge, MA, 2004.

[299] D. Karger, and C. Stein. A new approach to the minimum cut problem. *Journal of the ACM (JACM)*, 43(4), pp. 601–640, 1996.

[300] G. Karypis, E. H. Han, and V. Kumar. Chameleon: Hierarchical clustering using dynamic modeling. *Computer*, 32(8), pp, 68–75, 1999.

[301] G. Karypis, and V. Kumar. A fast and high quality multilevel scheme for partitioning irregular graphs. *SIAM Journal on scientific Computing*, 20(1), pp. 359–392, 1998.

[302] G. Karypis, R. Aggarwal, V. Kumar, and S. Shekhar. Multilevel hypergraph partitioning: applications in VLSI domain. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, 7(1), pp. 69–79, 1999.

[303] L. Kaufman, and P. J. Rousseeuw. Finding groups in data: an introduction to cluster analysis. *Wiley*, 2009.

[304] D. Kempe, J. Kleinberg, and E. Tardos. Maximizing the spread of influence through a social network. *ACM KDD Conference*, pp. 137–146, 2003.

[305] E. Keogh, S. Lonardi, and C. Ratanamahatana. Towards parameter-free data mining. *ACM KDD Conference*, pp. 206–215, 2004.

[306] E. Keogh, J. Lin, and A. Fu. HOT SAX: Finding the most unusual time series subsequence: Algorithms and applications. *IEEE ICDM Conference*, pp. 8, 2005.

[307] E. Keogh, and M. Pazzani. Scaling up dynamic time-warping for data mining applications. *ACM KDD Conference*, pp. 285–289, 2000.

[308] E. Keogh. Exact indexing of dynamic time warping. *VLDB Conference*, pp. 406–417, 2002.

[309] E. Keogh, K. Chakrabarti, M. Pazzani, and S. Mehrotra. Dimensionality reduction for fast similarity searching in large time series datanases. *Knowledge and Infomration Systems*, pp. 263–286, 2000.

[310] E. Keogh, S. Lonardi, and B. Y.-C. Chiu. Finding surprising patterns in a time series database in linear time and space. *ACM KDD Conference*, pp. 550–556, 2002.

[311] E. Keogh, S. Lonardi, and C. Ratanamahatana. Towards parameter-free data mining. *ACM KDD Conference*, pp. 206–215, 2004.

[312] B. Kernighan, and S. Lin. An efficient heuristic procedure for partitioning graphs. *Bell System Technical Journal*, 1970.

[313] A. Khan, N. Li, X. Yan, Z. Guan, S. Chakraborty, and S. Tao. Neighborhood-based fast graph search in large networks. *ACM SIGMOD Conference*, pp. 901–912, 2011.

[314] A. Khan, Y. Wu, C. Aggarwal, and X. Yan. Nema: Fast graph matching with label similarity. *Proceedings of the VLDB Endowment*, 6(3), pp. 181–192, 2013.

[315] D. Kifer, and J. Gehrke. Injecting utility into anonymized datasets. *ACM SIGMOD Conference*, pp. 217–228, 2006.

[316] L. Kissner, and D. Song. Privacy-preserving set operations. *Advances in Cryptology–CRYPTO*, pp. 241–257, 2005.

[317] J. Kleinberg. Authoritative sources in a hyperlinked environment. *Journal of the ACM (JACM)*, 46(5), pp. 604–632, 1999.

[318] S. Knerr, L. Personnaz, and G. Dreyfus. Single-layer learning revisited: a stepwise procedure for building and training a neural network. In J. Fogelman, editor, *Neurocomputing: Algorithms, Architectures and Applications*. Springer-Verlag, 1990.

[319] E. Knorr, and R. Ng. Algorithms for mining distance-based outliers in large datasets. *VLDB Conference*, pp. 392–403, 1998.

[320] E. Knorr, and R. Ng. Finding intensional knowledge of distance-based outliers. *VLDB Conference*, pp. 211–222, 1999.

[321] Y. Koren, R. Bell, and C. Volinsky. Matrix factorization techniques for recommender systems. *Computer*, 42(8), pp. 30–37, 2009.

[322] Y. Koren. Factorization meets the neighborhood: a multifaceted collaborative filtering model. *ACM KDD Conference*, pp. 426–434, 2008.

[323] Y. Koren. Collaborative filtering with temporal dynamics. *Communications of the ACM,*, 53(4), pp. 89–97, 2010.

[324] D. Kostakos, G. Trajcevski, D. Gunopulos, and C. Aggarwal. Time series data clustering. *Data Clustering: Algorithms and Applications*, CRC Press, 2013.

[325] J. Konstan. Introduction to recommender systems: algorithms and evaluation. *ACM Transactions on Information Systems*, 22(1), pp. 1–4, 2004.

[326] Y. Kou, C. T. Lu, and D. Chen. Spatial weighted outlier detection, *SIAM Conference on Data Mining*, 2006.

[327] A. Krogh, M. Brown, I. Mian, K. Sjolander, and D. Haussler. Hidden Markov models in computational biology: Applications to protein modeling. *Journal of molecular biology*, 235(5), pp. 1501–1531, 1994.

[328] J. B. Kruskal. Nonmetric multidimensional scaling: a numerical method. *Psychometrika*, 29(2), pp. 115–129, 1964.

[329] B. Kulis, S. Basu, I. Dhillon, and R. Mooney. Semi-supervised graph clustering: a kernel approach. *Machine Learning*, 74(1), pp. 1–22, 2009.

[330] S. Kulkarni, G. Lugosi, and S. Venkatesh. Learning pattern classification: a survey. *IEEE Transactions on Information Theory*, 44(6), pp. 2178–2206, 1998.

[331] M. Kuramochi, and G. Karypis. Frequent subgraph discovery. *IEEE International Conference on Data Mining*, pp. 313–320, 2001.

[332] L. V. S. Lakshmanan, R. Ng, J. Han, and A. Pang. Optimization of constrained frequent set queries with 2-variable constraints. *ACM SIGMOD Conference*, pp. 157–168, 1999.

[333] P. Langley, W. Iba, and K. Thompson. An analysis of Bayesian classifiers. *Proceedings of the National Conference on Artificial Intelligence*, pp. 223–228, 1992.

[334] A. Lazarevic, and V. Kumar. Feature bagging for outlier detection. *ACM KDD Conference*, pp. 157–166, 2005.

[335] K. LeFevre, D. J. DeWitt, and R. Ramakrishnan. Incognito: Efficient full-domain k-anonymity. *ACM SIGMOD Conference*, pp. 49–60, 2005.

[336] K. LeFevre, D. J. DeWitt, and R. Ramakrishnan. Mondrian multidimensional k-anonymity. *IEEE International Conference on Data Engineering*, pp. 25, 2006.

[337] J.-G. Lee, J. Han, and X. Li. Trajectory outlier detection: A partition-and-detect framework. *ICDE Conference*, pp. 140–149, 2008.

[338] J.-G. Lee, J. Han, and K.-Y. Whang. Trajectory clustering: a partition-and-group framework. *ACM SIGMOD Conference*, pp. 593–604, 2007.

[339] J.-G. Lee, J. Han, X. Li, and H. Gonzalez. TraClass: trajectory classification using hierarchical region-based and trajectory-based clustering. *Proceedings of the VLDB Endowment*, 1(1), pp. 1081–1094, 2008.

[340] W. Lee, and D. Xiang. Information theoretic measures for anomaly detection. *IEEE Symposium on Security and Privacy*, pp. 130–143, 2001.

[341] J. Leskovec, D. Huttenlocher, and J. Kleinberg. Predicting positive and negative links in online social networks. *World Wide Web Conference*, pp. 641–650, 2010.

[342] J. Leskovec, J. Kleinberg, and C. Faloutsos. Graphs over time: densification laws, shrinking diameters, and possible explanations. *ACM KDD Conference*, pp. 177–187, 2005.

[343] J. Leskovec, A. Rajaraman, and J. Ullman. Mining of massive datasets. *Cambridge University Press*, 2012.

[344] D. Lewis. Naive Bayes at forty: The independence assumption in information retrieval. *ECML Conference*, pp. 4–15, 1998.

[345] D. Lewis, and J. Catlett. Heterogeneous uncertainty sampling for supervised learning. *ICML Conference*, pp. 148–156, 1994.

[346] C. Li, Q. Yang, J. Wang, and M. Li. Efficient mining of gap-constrained subsequences and its various applications. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 6(1), 2, 2012.

[347] J. Li, G. Dong, K. Ramamohanarao, and L. Wong. Deeps: A new instance-based lazy discovery and classification system. *Machine Learning*, 54(2), pp. 99–124, 2004.

[348] N. Li, T. Li, and S. Venkatasubramanian. t-closeness: Privacy beyond k-anonymity and ℓ-diversity. *IEEE International Conference on Data Engineering*, pp. 106–115, 2007.

[349] W. Li, J. Han, and J. Pei. CMAR: Accurate and efficient classification based on multiple class-association rules. *IEEE ICDM Conference*, pp. 369–376, 2001.

[350] Y. Li, M. Dong, and J. Hua. Localized feature selection for clustering. *Pattern Recognition Letters*, 29(1), 10–18, 2008.

[351] Z. Li, B. Ding, J. Han, and R. Kays. Swarm: Mining relaxed temporal moving object clusters. *Proceedings of the VLDB Endowment*, 3(1–2), pp. 732–734, 2010.

[352] Z. Li, B. Ding, J. Han, R. Kays, and P. Nye. Mining periodic behaviors for moving objects. *ACM KDD Conference*, pp. 1099–1108, 2010.

[353] D. Liben-Nowell, and J. Kleinberg. The link-prediction problem for social networks. *Journal of the American Society for Information Science and Technology*, 58(7), pp. 1019–1031, 2007.

[354] R. Lichtenwalter, J. Lussier, and N. Chawla. New perspectives and methods in link prediction. *ACM KDD Conference*, pp. 243–252, 2010.

[355] J. Lin, E. Keogh, S. Lonardi, and B. Chiu. Experiencing SAX: a novel symbolic representation of time series. *Data Mining and Knowledge Discovery*, 15(2), pp. 107–144, 2003.

[356] J. Lin, E. Keogh, S. Lonardi, and P. Patel. Finding motifs in time series. *Proceedings of the 2nd Workshop on Temporal Data*, 2002.

[357] B. Liu. Web data mining: exploring hyperlinks, contents, and usage data. *Springer*, New York, 2007.

[358] B. Liu, W. Hsu, and Y. Ma. Integrating classification and association rule mining. *ACM KDD Conference*, pp. 80–86, 1998.

[359] G. Liu, H. Lu, W. Lou, and J. X. Yu. On computing, storing and querying frequent patterns. *ACM KDD Conference*, pp. 607–612, 2003.

[360] H. Liu, and H. Motoda. Feature selection for knowledge discovery and data mining. *Springer*, 1998.

[361] J. Liu, Y. Pan, K. Wang, and J. Han. Mining frequent item sets by opportunistic projection. *ACM KDD Conference*, pp. 229–238, 2002.

[362] L. Liu, J. Tang, J. Han, M. Jiang, and S. Yang. Mining topic-level influence in heterogeneous networks. *ACM CIKM Conference*, pp. 199–208, 2010.

[363] D. Lin. An Information-theoretic Definition of Similarity. *ICML Conference*, pp. 296–304, 1998.

[364] R. Little, and D. Rubin. Statistical analysis with missing data. *Wiley*, 2002.

[365] F. T. Liu, K. M. Ting, and Z.-H. Zhou. Isolation forest. *IEEE ICDM Conference*, pp. 413–422, 2008.

[366] H. Liu, and H. Motoda. Computational methods of feature selection. *Chapman and Hall/CRC*, 2007.

[367] K. Liu, C. Giannella, and H. Kargupta. A survey of attack techniques on privacy-preserving data perturbation methods. *Privacy-Preserving Data Mining: Models and Algorithms*, Springer, pp. 359–381, 2008.

[368] B. London, and L. Getoor. Collective classification of network data. *Data Classification: Algorithms and Applications*, CRC Press, pp. 399–416, 2014.

[369] C.-T. Lu, D. Chen, and Y. Kou. Algorithms for spatial outlier detection, *IEEE ICDM Conference*, pp. 597–600, 2003.

[370] Q. Lu, and L. Getoor. Link-based classification. *ICML Conference*, pp. 496–503, 2003.

[371] U. von Luxburg. A tutorial on spectral clustering. *Statistics and computing*, 17(4), pp. 395–416, 2007.

[372] A. Machanavajjhala, D. Kifer, J. Gehrke, and M. Venkitasubramaniam. ℓ-diversity: privacy beyond k-anonymity. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 1(3), 2007.

[373] S. Macskassy, and F. Provost. A simple relational classifier. *Second Workshop on Multi-Relational Data Mining (MRDM) at ACM KDD Conference*, 2003.

[374] S. C. Madeira, and A. L. Oliveira. Biclustering algorithms for biological data analysis: a survey. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*. 1(1), pp. 24–45, 2004.

[375] N. Mamoulis, H. Cao, G. Kollios, M. Hadjieleftheriou, Y. Tao, and D. Cheung. Mining, indexing, and querying historical spatiotemporal data. *ACM KDD Conference*, pp. 236–245, 2004.

[376] G. Manku, and R. Motwani. Approximate frequency counts over data streams. *VLDB Conference*, pp. 346–357, 2002.

[377] C. Manning, P. Raghavan, and H. Schutze. Introduction to information retrieval. *Cambridge University Press*, Cambridge, 2008.

[378] M. Markou, and S. Singh. Novelty detection: a review, part 1: statistical approaches. *Signal Processing*, 83(12), pp. 2481–2497, 2003.

[379] G. J. McLachian. Discriminant analysis and statistical pattern recognition. *Wiler Interscience*, 2004.

[380] M. Markou, and S. Singh. Novelty detection: A review, part 2: neural network-based approaches. *Signal Processing*, 83(12), pp. 2481–2497, 2003.

[381] M. Mehta, R. Agrawal, and J. Rissanen. SLIQ: A fast scalable classifier for data mining, *EDBT Conference*, pp. 18–32, 1996.

[382] P. Melville, M. Saar-Tsechansky, F. Provost, and R. Mooney. An expected utility approach to active feature-value acquisition. *IEEE ICDM Conference*, 2005.

[383] A. K. Menon, and C. Elkan. Link prediction via matrix factorization. *Machine Learning and Knowledge Discovery in Databases*, pp. 437–452, 2011.

[384] B. Messmer, and H. Bunke. A new algorithm for error-tolerant subgraph isomprohism detection. *IEEE Transactions on Pattern Mining and Machine Intelligence*, 20(5), pp. 493–504, 1998.

[385] A. Meyerson, and R. Williams. On the complexity of optimal k-anonymization. *ACM PODS Conference*, pp. 223–228, 2004.

[386] R. Michalski, I. Mozetic, J. Hong, and N. Lavrac. The multi-purpose incremental learning system AQ15 and its testing application to three medical domains. *Proceedings of the AAAI*, pp. 1–41, 1986.

[387] C. Michael, and A. Ghosh. Two state-based approaches to program-based anomaly detection. *Computer Security Applications Conference*, pp. 21, 2000.

[388] H. Miller, and J. Han. Geographic data mining and knowledge discovery. *CRC Press*, 2009.

[389] T. M. Mitchell. Machine learning. *McGraw Hill International Edition*, 1997.

[390] B. Mobasher. Web usage mining and personalization. *Practical Handbook of Internet Computing, ed. Munindar Singh*, pp, 264–265, CRC Press, 2005.

[391] D. Montgomery, E. Peck, and G. Vining. Introduction to linear regression analysis. *John Wiley and Sons*, 2012.

[392] C. H. Mooney, and J. F. Roddick. Sequential pattern mining: approaches and algorithms. *ACM Computing Surveys (CSUR)*, 45(2), 2013.

[393] B. Moret. Decision trees and diagrams. *ACM Computing Surveys (CSUR)*, 14(4), pp. 593–623, 1982.

[394] A. Mueen, E. Keogh, Q. Zhu, S. Cash, and M. Westover. Exact discovery of time series motifs. *SDM Conference*, pp. 473–484, 2009.

[395] A. Mueen, and E. Keogh. Online discovery and maintenance of time series motifs. *ACM KDD Conference*, pp. 1089–1098, 2010.

[396] E. Muller, M. Schiffer, and T. Seidl. Statistical selection of relevant subspace projections for outlier ranking. *ICDE Conference*, pp, 434–445, 2011.

[397] E. Muller, I. Assent, P. Iglesias, Y. Mulle, and K. Bohm. Outlier analysis via subspace analysis in multiple views of the data. *IEEE ICDM Conference*, pp. 529–538, 2012.

[398] S. K. Murthy. Automatic construction of decision trees from data: A multi-disciplinary survey. *Data Mining and Knowledge Discovery*, 2(4), pp. 345–389, 1998.

[399] S. Nabar, K. Kenthapadi, N. Mishra, and R. Motwani. A survey of query auditing techniques for data privacy. *Privacy-Preserving Data Mining: Models and Algorithms*, Springer, pp. 415–431, 2008.

[400] D. Nadeau, and S. Sekine. A survey of named entity recognition and classification. *Lingvisticae Investigationes*, 30(1), 3–26, 2007.

[401] M. Naor, and B. Pinkas. Efficient oblivious transfer protocols. *SODA Conference*, pp. 448–457, 2001.

[402] A. Narayanan, and V. Shmatikov. How to break anonymity of the netflix prize dataset. *arXiv preprint cs/0610105*, 2006. `http://arxiv.org/abs/cs/0610105`

[403] G. Nemhauser, and L. Wolsey. Integer and combinatorial optimization. *Wiley*, New York, 1988.

[404] J. Neville, and D. Jensen. Iterative classification in relational data. *AAAI Workshop on Learning Statistical Models from Relational Data*, pp. 13–20, 2000.

[405] A. Ng, M. Jordan, and Y. Weiss. On spectral clustering analysis and an algorithm. *Advances in Neural Information Processing Systems*, pp. 849–856, 2001.

[406] R. T. Ng, L. V. S. Lakshmanan, J. Han, and A. Pang. Exploratory mining and pruning optimizations of constrained associations rules. *ACM SIGMOD Conference*, pp. 13–24, 1998.

[407] R. T. Ng, and J. Han. CLARANS: A method for clustering objects for spatial data mining. *IEEE Transactions on Knowledge and Data Engineering*, 14(5), pp. 1003–1016, 2002.

[408] M. Neuhaus, and H. Bunke. Automatic learning of cost functions for graph edit distance. *Information Sciences*, 177(1), pp. 239–247, 2007.

[409] M. Neuhaus, K. Riesen, and H. Bunke. Fast suboptimal algorithms for the computation of graph edit distance. *Structural, Syntactic, and Statistical Pattern Recognition*, pp. 163–172, 2006.

[410] K. Nigam, A. McCallum, S. Thrun, and T. Mitchell. Text classification with labeled and unlabeled data using EM. *Machine Learning*, 39(2), pp. 103–134, 2000.

[411] B. Ozden, S. Ramaswamy, and A. Silberschatz. Cyclic association rules. *International Conference on Data Engineering*, pp. 412–421, 1998.

[412] L. Page, S. Brin, R. Motwani, and T. Winograd. The PageRank citation engine: Bringing order to the web. *Technical Report*, 1999–0120, Computer Science Department, Stanford University, 1998.

[413] F. Pan, G. Cong, A. Tung, J. Yang, and M. Zaki. CARPENTER: Finding closed patterns in long biological datasets. *ACM KDD Conference*, pp. 637–642, 2003.

[414] T. Palpanas. Real-time data analytics in sensor networks. *Managing and Mining Sensor Data*, pp. 173–210, Springer, 2013.

[415] F. Pan, A. K. H. Tung, G. Cong, and X. Xu. COBBLER: Combining column and row enumeration for closed pattern discovery. *International Conference on Scientific and Statistical Database Management*, pp. 21–30, 2004.

[416] C. Papadimitriou, H. Tamaki, P. Raghavan, and S. Vempala. Latent semantic indexing: A probabilistic analysis. *ACM PODS Conference*, pp. 159–168, 1998.

[417] N. Pasquier, Y. Bastide, R. Taouil, and L. Lakhal. Discovering frequent closed itemsets for association rules. *International Conference on Database Theory*, pp. 398–416, 1999.

[418] P. Patel, E. Keogh, J. Lin, and S. Lonardi. Mining motifs in massive time series databases. *IEEE ICDM Conference*, pp. 370–377, 2002.

[419] J. Pei, J. Han, H. Lu, S. Nishio, S. Tang, and D. Yang. H-mine: Hyper-structure mining of frequent patterns in large databases. *IEEE ICDM Conference*, pp. 441–448, 2001.

[420] J. Pei, J. Han, and R. Mao. CLOSET: An efficient algorithm for mining frequent closed itemsets. *ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery*, pp, 21–30, 2000.

[421] J. Pei, J. Han, B. Mortazavi-Asl, J. Wang, H. Pinto, Q. Chen, U. Dayal, and M. C. Hsu. Mining sequential patterns by pattern-growth: The prefixspan approach. *IEEE Transactions on Knowledge and Data Engineering*, 16(11), pp. 1424–1440, 2004.

[422] J. Pei, J. Han, and L. V. S. Lakshmanan. Mining frequent patterns with convertible constraints. *ICDE Conference*, pp. 433–442, 2001.

[423] D. Pelleg, and A. W. Moore. X-means: Extending k-means with efficient estimation of the number of clusters. *ICML Conference*, pp. 727–734, 2000.

[424] M. Petrou, and C. Petrou. Image processing: the fundamentals. *Wiley*, 2010.

[425] D. Pierrakos, G. Paliouras, C. Papatheodorou, and C. Spyropoulos. Web usage mining as a tool for personalization: a survey. *User Modeling and User-Adapted Interaction*, 13(4), pp, 311–372, 2003.

[426] D. Pokrajac, A. Lazerevic, and L. Latecki. Incremental local outlier detection for data streams. *Computational Intelligence and Data Mining Conference*, pp. 504–515, 2007.

[427] S. A. Macskassy, and F. Provost. Classification in networked data: A toolkit and a univariate case study. *Joirnal of Machine Learning Research*, 8, pp. 935–983, 2007.

[428] G. Qi, C. Aggarwal, and T. Huang. Link Prediction across networks by biased cross-network sampling. *IEEE ICDE Conference*, pp. 793–804, 2013.

[429] G. Qi, C. Aggarwak, and T. Huang. Online community detection in social sensing. *ACM WSDM Conference*, pp. 617–626, 2013.

[430] J. Quinlan. C4.5: programs for machine learning. *Morgan-Kaufmann Publishers*, 1993.

[431] J. Quinlan. Induction of decision trees. *Machine Learning*, 1, pp. 81–106, 1986.

[432] D. Rafiei, and A. Mendelzon. Similarity-based queries for time series data, *ACM SIGMOD Record*, 26(2), pp. 13–25, 1997.

[433] E. Rahm, and H. Do. Data cleaning: problems and current approaches, *IEEE Data Engineering Bulletin*, 23(4), pp. 3–13, 2000.

[434] R. Ramakrishnan, and J. Gehrke. Database Management Systems. *Osborne/McGraw Hill*, 1990.

[435] V. Raman, and J. Hellerstein. Potter's wheel: An interactive data cleaning system. *VLDB Conference*, pp. 381–390, 2001.

[436] S. Ramaswamy, R. Rastogi, and K. Shim. Efficient algorithms for mining outliers from large data sets. *ACM SIGMOD Conference*, pp. 427–438, 2000.

[437] M. Rege, M. Dong, and F. Fotouhi. Co-clustering documents and words using bipartite isoperimetric graph partitioning. *IEEE ICDM Conference*, pp. 532–541, 2006.

[438] E. S. Ristad, and P. N. Yianilos. Learning string-edit distance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 20(5), pp. 522–532, 1998.

[439] F. Rosenblatt. The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological review*, 65(6), 286, 1958.

[440] R. Salakhutdinov, and A. Mnih. *Probabilistic Matrix Factorization. Advances in Neural and Information Processing Systems*, pp. 1257–1264, 2007.

[441] G. Salton, and M. J. McGill. Introduction to modern information retrieval. *McGraw Hill*, 1986.

[442] P. Samarati. Protecting respondents identities in microdata release. *IEEE Transactions on Knowledge and Data Engineering*, 13(6), pp. 1010–1027, 2001.

[443] H. Samet. The design and analysis of spatial data structures. *Addison-Wesley*, Reading, MA, 1990.

[444] J. Sander, M. Ester, H. P. Kriegel, and X. Xu. Density-based clustering in spatial databases: The algorithm gdbscan and its applications. *Data Mining and Knowledge Discovery*, 2(2), pp. 169–194, 1998.

[445] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl. Item-based collaborative filtering recommendation algorithms. *World Wide Web Conference*, pp. 285–295, 2001.

[446] A. Savasere, E. Omiecinski, and S. B. Navathe. An efficient algorithm for mining association rules in large databases. *Very Large Databases Conference*, pp. 432–444, 1995.

[447] A. Savasere, E. Omiecinski, and S. Navathe. Mining for strong negative associations in a large database of customer transactions. *IEEE ICDE Conference*, pp. 494–502, 1998.

[448] C. Saunders, A. Gammerman, and V. Vovk. Ridge regression learning algorithm in dual variables. *ICML Conference*, pp. 515–521, 1998.

[449] B. Scholkopf, and A. J. Smola. Learning with kernels: support vector machines, regularization, optimization, and beyond. *Cambridge University Press*, 2001.

[450] B. Scholkopf, A. Smola, and K.-R. Muller. Nonlinear component analysis as a kernel eigenvalue problem. *Neural Computation*, 10(5), pp. 1299–1319, 1998.

[451] B. Scholkopf, and A. J. Smola. *Learning with Kernels*. MIT Press, Cambridge, MA, 2002.

[452] H. Schutze, and C. Silverstein. Projections for efficient document clustering. *ACM SIGIR Conference*, pp. 74–81, 1997.

[453] F. Sebastiani. Machine Learning in Automated Text Categorization. *ACM Computing Surveys*, 34(1), 2002.

[454] B. Settles. Active Learning. *Morgan and Claypool*, 2012.

[455] B. Settles, and M. Craven. An analysis of active learning strategies for sequence labeling tasks. *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 1069–1078, 2008.

[456] D. Seung, and L. Lee. Algorithms for non-negative matrix factorization. *Advances in Neural Information Processing Systems*, 13, pp. 556–562, 2001.

[457] H. Seung, M. Opper, and H. Sompolinsky. Query by committee. *Fifth annual workshop on Computational learning theory*, pp. 287–294, 1992.

[458] J. Shafer, R. Agrawal, and M. Mehta. SPRINT: A scalable parallel classifier for data mining. *VLDB Conference*, pp. 544–555, 1996.

[459] S. Shekhar, C. T. Lu, and P. Zhang. Detecting graph-based spatial outliers: algorithms and applications. *ACM KDD Conference*, pp. 371–376, 2001.

[460] S.Shekhar, C. T. Lu, and P. Zhang. A unified approach to detecting spatial outliers. *Geoinformatica*, 7(2), pp. 139–166, 2003.

[461] S. Shekhar, and S. Chawla. A tour of spatial databases. *Prentice Hall*, 2002.

[462] S. Shekhar, C. T. Lu, and P. Zhang. Detecting graph-based spatial outliers. *Intelligent Data Analysis*, 6, pp. 451–468, 2002.

[463] S. Shekhar, and Y. Huang. Discovering spatial co-location patterns: a summary of results. In *Advances in Spatial and Temporal Databases* , pp. 236–256, Springer, 2001.

[464] G. Sheikholeslami, S. Chatterjee, and A. Zhang. Wavecluster: A multi-resolution clustering approach for very large spatial databases. *VLDB Conference*, pp. 428–439, 1998.

[465] P. Shenoy, J. Haritsa, S. Sudarshan, G., Bhalotia, M. Bawa, and D. Shah. Turbo-charging vertical mining of large databases. *ACM SIGMOD Conference*, 29(2), pp. 22–35, 2000.

[466] J. Shi, and J. Malik. Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 22(8), pp. 888–905, 2000.

[467] R. Shumway, and D. Stoffer. Time-series analysis and its applications: With R examples, *Springer*, New York, 2011.

[468] M.-L. Shyu, S.-C. Chen, K. Sarinnapakorn, and L. Chang. A novel anomaly detection scheme based on principal component classifier, *ICDM Conference*, pp. 353–365, 2003.

[469] R. Sibson. SLINK: An optimally efficient algorithm for the single-link clustering method. The Computer Journal, 16(1), pp. 30–34, 1973.

[470] A. Siebes, J. Vreeken, and M. van Leeuwen. itemsets that compress. *SDM Conference*, pp. 393–404, 2006.

[471] B. W. Silverman. Density Estimation for Statistics and Data Analysis. *Chapman and Hall*, 1986.

[472] K. Smets, and J. Vreeken. The odd one out: Identifying and characterising anomalies. *SIAM Conference on Data Mining*, pp. 804–815, 2011.

[473] E. S. Smirnov. On exact methods in systematics. *Systematic Zoology*, 17(1), pp. 1–13, 1968.

[474] P. Smyth. Clustering sequences with hidden Markov models. *Advances in Neural Information Processing Systems*, pp. 648–654, 1997.

[475] E. J. Stollnitz, and T. D. De Rose. Wavelets for computer graphics: theory and applications. *Morgan Kaufmann*, 1996.

[476] R. Srikant, and R. Agrawal. Mining quantitative association rules in large relational tables. *ACM SIGMOD Conference*, pp. 1–12, 1996.

[477] J. Srivastava, R. Cooley, M. Deshpande, and P. N. Tan. Web usage mining: Discovery and applications of usage patterns from web data. *ACM SIGKDD Explorations Newsletter*, 1(2), pp. 12–23, 2000.

[478] I. Steinwart, and A. Christmann. Support vector machines. *Springer*, 2008.

[479] A. Strehl, and J. Ghosh. Cluster ensembles—a knowledge reuse framework for combining multiple partitions. *Journal of Machine Learning Research*, 3, pp. 583–617, 2003.

[480] G. Strang. An introduction to linear algebra. *Wellesley Cambridge Press*, 2009.

[481] G. Strang, and K. Borre. Linear algebra, geodesy, and GPS. *Wellesley Cambridge Press*, 1997.

[482] K. Subbian, C. Aggarwal, and J. Srivasatava. Content-centric flow mining for influence analysis in social streams. *CIKM Conference*, pp. 841–846, 2013.

[483] J. Sun, and J. Tang. A survey of models and algorithms for social influence analysis. *Social Network Data Analytics*, Springer, pp. 177–214, 2011.

[484] Y. Sun, J. Han, C. Aggarwal, and N. Chawla. When will it happen?: relationship prediction in heterogeneous information networks. *ACM international conference on Web search and data mining*, pp. 663–672, 2012.

[485] P.-N Tan, M. Steinbach, and V. Kumar. Introduction to data mining. *Addison-Wesley*, 2005.

[486] P. N. Tan, V. Kumar, and J. Srivastava. Selecting the right interestingness measure for association patterns. *ACM KDD Conference*, pp. 32–41, 2002.

[487] J. Tang, Z. Chen, A. W.-C. Fu, and D. W. Cheung. Enhancing effectiveness of outlier detection for low density patterns. *PAKDD Conference*, pp. 535–548, 2002.

[488] J. Tang, J. Sun, C. Wang, and Z. Yang. Social influence analysis in large-scale networks. *ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 807–816, 2009.

[489] B. Taskar, M. Wong, P. Abbeel, and D. Koller. Link prediction in relational data. *Advances in Neural Information Processing Systems*, 2003.

[490] J. Tenenbaum, V. De Silva, and J. Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290 (5500), pp. 2319–2323, 2000.

[491] K. Ting, and I. Witten. Issues in stacked generalization. *Journal of Artificial Intelligence Research*, 10, pp. 271–289, 1999.

[492] T. Mitsa. Temporal data mining. *CRC Press*, 2010.

[493] H. Toivonen. Sampling large databases for association rules. *VLDB Conference*, pp. 134–145, 1996.

[494] V. Vapnik. The nature of statistical learning theory. *Springer*, 2000.

[495] J. Vaidya. A survey of privacy-preserving methods across vertically partitioned data. *Privacy-Preserving Data Mining: Models and Algorithms*, Springer, pp. 337–358, 2008.

[496] V. Vapnik. Statistical learning theory. *Wiley*, 1998.

[497] V. Verykios, and A. Gkoulalas-Divanis. A Survey of Association Rule Hiding Methods for Privacy. *Privacy-Preserving Data Mining: Models and Algorithms*, Springer, pp. 267–289, 2008.

[498] J. S. Vitter. Random sampling with a reservoir. *ACM Transactions on Mathematical Software (TOMS)*, 11(1), pp. 37–57, 2006.

[499] M. Vlachos, M. Hadjieleftheriou, D. Gunopulos, and E. Keogh. Indexing multidimensional time-series with support for multiple distance measures. *ACM KDD Conference*, pp. 216–225, 2003.

[500] M. Vlachos, G. Kollios, and D. Gunopulos. Discovering similar multidimensional trajectories. *IEEE International Conference on Data Engineering*, pp. 673–684, 2002.

[501] T. De Vries, S. Chawla, and M. Houle. Finding local anomalies in very high dimensional space. *IEEE ICDM Conference*, pp. 128–137, 2010.

[502] A. Waddell, and R. Oldford. Interactive visual clustering of high dimensional data by exploring low-dimensional subspaces. *INFOVIS*, 2012.

[503] H. Wang, W. Fan, P. Yu, and J. Han. Mining concept-drifting data streams using ensemble classifiers. *ACM KDD Conference*, pp. 226–235, 2003.

[504] J. Wang, J. Han, and J. Pei. Closet+: Searching for the best strategies for mining frequent closed itemsets. *ACM KDD Conference*, pp. 236–245, 2003.

[505] J. Wang, Y. Zhang, L. Zhou, G. Karypis, and C. C. Aggarwal. Discriminating subsequence discovery for sequence clustering. *SIAM Conference on Data Mining*, pp. 605–610, 2007.

[506] W. Wang, J. Yang, and R. Muntz. STING: A statistical information grid approach to spatial data mining. *VLDB Conference*, pp. 186–195, 1997.

[507] J. S. Walker. Fast fourier transforms. *CRC Press*, 1996.

[508] S. Wasserman. Social network analysis: Methods and applications. *Cambridge University Press*, 1994.

[509] D. Watts, and D. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393 (6684), pp. 440–442, 1998.

[510] L. Wei, E. Keogh, and X. Xi. SAXually Explicit images: Finding unusual shapes. *IEEE ICDM Conference*, pp. 711–720, 2006.

[511] H. Wiener. Structural determination of paraffin boiling points. *Journal of the American Chemical Society*. 1(69). pp. 17–20, 1947.

[512] L. Willenborg, and T. De Waal. Elements of statistical disclosure control. *Springer*, 2001.

[513] D. Wolpert. Stacked generalization. *Neural Networks*, 5(2), pp. 241–259, 1992.

[514] X. Xiao, and Y. Tao. Anatomy: Simple and effective privacy preservation. *Very Large Databases Conference*, pp. 139–150, 2006.

[515] D. Xin, J. Han, X. Yan, and H. Cheng. Mining compressed frequent-pattern sets. *VLDB Conference*, pp. 709–720, 2005.

[516] Z. Xing, J. Pei, and E. Keogh. A brief survey on sequence classification. *SIGKDD Explorations Newsletter*, 12(1), pp. 40–48, 2010.

[517] H. Xiong, P. N. Tan, and V. Kumar. Mining strong affinity association patterns in data sets with skewed support distribution. *ICDM Conference*, pp. 387–394, 2003.

[518] K. Yaminshi, J. Takeuchi, and G. Williams. Online unsupervised outlier detection using finite mixtures with discounted learning algorithms, *ACM KDD Conference*,pp. 320–324, 2000.

[519] X. Yan, and J. Han. gSpan: Graph-based substructure pattern mining. *IEEE International Conference on Data Mining*, pp. 721–724, 2002.

[520] X. Yan, P. Yu, and J. Han. Substructure similarity search in graph databases. *ACM SIGMOD Conference*, pp. 766–777, 2005.

[521] X. Yan, P. Yu, and J. Han. Graph indexing: a frequent structure-based approach. *ACM SIGMOD Conference*, pp. 335–346, 2004.

[522] X. Yan, F. Zhu, J. Han, and P. S. Yu. Searching substructures with superimposed distance. *International Conference on Data Engineering*, pp. 88, 2006.

[523] J. Yang, and W. Wang. CLUSEQ: efficient and effective sequence clustering. *IEEE International Conference on Data Engineering*, pp. 101–112, 2003.

[524] D. Yankov, E. Keogh, J. Medina, B. Chiu, and V. Zordan. Detecting time series motifs under uniform scaling. *ACM KDD Conference*, pp. 844–853, 2007.

[525] N. Ye. A markov chain model of temporal behavior for anomaly detection. *IEEE Information Assurance Workshop*, pp. 169, 2004.

[526] B. K. Yi, H. V. Jagadish, and C. Faloutsos. Efficient retrieval of similar time sequences under time warping. *IEEE International Conference on Data Engineering*, pp. 201–208, 1998.

[527] B. K. Yi, N. Sidiropoulos, T. Johnson, H. V. Jagadish, C. Faloutsos, and A. Biliris. Online data mining for co-evolving time sequences. *International Conference on Data Engineering*, pp. 13–22, 2000.

[528] H. Yildirim, and M. Krishnamoorthy. A random walk method for alleviating the sparsity problem in collaborative filtering. *ACM conference on Recommender systems*, pp. 131–138, 2008.

[529] X. Yin, and J. Han. CPAR: Classification based on predictive association rules. *SIAM international conference on data mining*, pp. 331–335, 2003.

[530] S. Yu, and J. Shi. Multiclass spectral clustering. *International Conference on Computer Vision*, 2003.

[531] B. Zadrozny, J. Langford, and N. Abe. Cost-sensitive learning by cost-proportionate example weighting. *ICDM Conference*, pp. 435–442, 2003.

[532] R. Zafarani, M. A. Abbasi, and H. Liu. Social media mining: an introduction. *Cambridge University Press*, New York, 2014.

[533] H. Zakerzadeh, C. Aggarwal, and K. Barker. Towards breaking the curse of dimensionality for high-dimensional privacy. *SIAM Conference on Data Mining*, pp. 731–739, 2014.

[534] M. J. Zaki. Scalable algorithms for association mining. *IEEE Transactions on Knowledge and Data Engineering*, 12(3), pp. 372–390, 2000.

[535] M. J. Zaki. SPADE: An efficient algorithm for mining frequent sequences. *Machine learning*, 42(1–2), pp. 31–60, 2001. 31–60.

[536] M. J. Zaki, and M. Wagner Jr. Data mining and analysis: fundamental concepts and algorithms. *Cambridge University Press*, 2014.

[537] M. J. Zaki, S. Parthasarathy, M. Ogihara, and W. Li. New algorithms for fast discovery of association rules. *KDD Conference*, pp. 283–286, 1997.

[538] M. J. Zaki, and K. Gouda. Fast vertical mining using diffsets. *ACM KDD Conference*, pp. 326–335, 2003.

[539] M. J. Zaki, and C. Hsiao. CHARM: An efficient algorithm for closed itemset mining. *SIAM Conference on Data Mining*, pp. 457–473, 2002.

[540] M. J. Zaki, and C. Aggarwal. XRules: An effective algorithm for structural classification of XML data. *Machine Learning*, 62(1–2), pp. 137–170, 2006.

[541] B. Zenko. Is combining classifiers better than selecting the best one? *Machine Learning*, pp. 255–273, 2004.

[542] Y. Zhai, and B. Liu. Web data extraction based on partial tree alignment. *World Wide Web Conference*, pp. 76–85, 2005.

[543] D. Zhan, M. Li, Y. Li, and Z.-H. Zhou. Learning instance specific distances using metric propagation. *ICML Conference*, pp. 1225–1232, 2009.

[544] H. Zhang, A. Berg, M. Maire, and J. Malik. SVM-KNN: Discriminative nearest neighbor classification for visual category recognition. *Computer Vision and Pattern Recognition*, pp. 2126–2136, 2006.

[545] J. Zhang, Z. Ghahramani, and Y. Yang. A probabilistic model for online document clustering with application to novelty detection. *Advances in Neural Information Processing Systems*, pp. 1617–1624, 2004.

[546] J. Zhang, Q. Gao, and H. Wang. SPOT: A system for detecting projected outliers from high-dimensional data stream. *ICDE Conference*, 2008.

[547] D. Zhang, and G. Lu. Review of shape representation and description techniques. *Pattern Recognition*, 37(1), pp. 1–19, 2004.

[548] S. Zhang, W. Wang, J. Ford, and F. Makedon. Learning from incomplete ratings using nonnegative matrix factorization. *SIAM Conference on Data Mining*, pp. 549–553, 2006.

[549] T. Zhang, R. Ramakrishnan, and M. Livny. BIRCH: an efficient data clustering method for very large databases. *ACM SIGMOD Conference*, pp. 103–114, 1996.

[550] Z. Zhao, and H. Liu. Spectral feature selection for supervised and unsupervised learning. *ICML Conference*, pp. 1151–1157, 2007.

[551] D. Zhou, O. Bousquet, T. Lal, J. Weston, and B. Scholkopf. Learning with local and global consistency. *Advances in Neural Information Processing Systems*, 16(16), pp. 321–328, 2004.

[552] D. Zhou, J. Huang, and B. Scholkopf. Learning from labeled and unlabeled data on a directed graph. *ICML Conference*, pp. 1036–1043, 2005.

[553] F. Zhu, X. Yan, J. Han, P. S. Yu, and H. Cheng. Mining colossal frequent patterns by core pattern fusion. *ICDE Conference*, pp. 706–715, 2007.

[554] X. Zhu, Z. Ghahramani, and J. Lafferty. Semi-supervised learning using gaussian fields and harmonic functions. *ICML Conference*, pp. 912–919, 2003.

[555] X. Zhu, and A. Goldberg. Introduction to semi-supervised learning. *Morgan and Claypool*, 2009.

[556] `http://db.csail.mit.edu/labdata/labdata.html`.

[557] `http://www.itl.nist.gov/iad/mig/tests/tdt/tasks/fsd.html`.

[558] `http://sifter.org/~simon/journal/20061211.html`.

[559] `http://www.netflixprize.com/`.

Index

C. C. Aggarwal, *Data Mining: The Textbook*, DOI 10.1007/978-3-319-14142-8