

- Abe, N., Zadrozny, B., & Langford, J. (2006). Outlier detection by active learning. In *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 767–772). New York: ACM Press.
- Adriaans, P., & Zantige, D. (1996). *Data mining*. Harlow, England: Addison-Wesley.
- Agrawal, R., & Srikant, R. (1994). Fast algorithms for mining association rules in large databases. In J. Bocca, M. Jarke, & C. Zaniolo (Eds.), *Proceedings of the International Conference on Very Large Data Bases* (pp. 478–499). Santiago, Chile. San Francisco: Morgan Kaufmann.
- Agrawal, R., Imielinski, T., & Swami, A. (1993a). Database mining: A performance perspective. *IEEE Transactions on Knowledge and Data Engineering*, 5(6), 914–925.
- . (1993b). Mining association rules between sets of items in large databases. In P. Buneman, & S. Jajodia (Eds.), *Proceedings of the ACM SIGMOD International Conference on Management of Data* (pp. 207–216). Washington, DC. New York: ACM Press.
- Aha, D. (1992). Tolerating noisy, irrelevant, and novel attributes in instance-based learning algorithms. *International Journal of Man-Machine Studies*, 36(2), 267–287.
- Almuallin, H., & Dietterich, T. G. (1991). Learning with many irrelevant features. In *Proceedings of the Ninth National Conference on Artificial Intelligence* (pp. 547–552). Anaheim, CA. Menlo Park, CA: AAAI Press.
- . (1992). Efficient algorithms for identifying relevant features. In *Proceedings of the Ninth Canadian Conference on Artificial Intelligence* (pp. 38–45). Vancouver, BC. San Francisco: Morgan Kaufmann.
- Andrews, S., Tsochantaridis, I., & Hofmann, T. (2003). Support vector machines for multiple-instance learning. In *Proceedings of the Conference on Neural Information Processing Systems* (pp. 561–568). Vancouver, BC. Cambridge, MA: MIT Press.
- Ankerst, M., Breunig, M. M., Kriegel, H.-P., & Sander, J. (1999). OPTICS: Ordering points to identify the clustering structure. In *Proceedings of the ACM SIGMOD International Conference on Management of Data* (pp. 49–60). New York: ACM Press.
- Appelt, D. (1999). An introduction to information extraction. *Artificial Intelligence Communications*, 12(3), 161–172.
- Arthur, D., & Vassilvitskii, S. (2007). K-means++: The advantages of careful seeding. In *Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms* (pp. 1027–1035). New Orleans. Philadelphia: Society for Industrial and Applied Mathematics.
- Asuncion, A., & Newman, D. J. (2007). *UCI Machine Learning Repository* [<http://www.ics.uci.edu/~mllearn/MLRepository.html>]. Irvine: University of California, School of Information and Computer Science.

- Asmis, E. (1984). *Epicurus' scientific method*. Ithaca, NY: Cornell University Press.
- Atkeson, C. G., Schaal, S. A., & Moore, A. W. (1997). Locally weighted learning. *AI Review*, 11, 11–71.
- Auer, P., & Ortner, R. (2004). A boosting approach to multiple instance learning. In *Proceedings of the European Conference on Machine Learning* (pp. 63–74). Pisa, Italy. Berlin: Springer-Verlag.
- Barnett, V., & Lewis, T. (1994). *Outliers in Statistical Data*. West Sussex, England: John Wiley, & Sons.
- Bay, S. D. (1999). Nearest neighbor classification from multiple feature subsets. *Intelligent Data Analysis*, 3(3), 191–209.
- Bay, S. D., & Schwabacher, M. (2003). Near linear time detection of distance-based outliers and applications to security. In *Proceedings of the Workshop on Data Mining for Counter Terrorism and Security*. San Francisco. Philadelphia: Society for Industrial and Applied Mathematics.
- Bayes, T. (1763). An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society of London*, 53, 370–418.
- Beck, J. R., & Schultz, E. K. (1986). The use of ROC curves in test performance evaluation. *Archives of Pathology and Laboratory Medicine*, 110, 13–20.
- Bergadano, F., & Gunetti, D. (1996). *Inductive logic programming: From machine learning to software engineering*. Cambridge, MA: MIT Press.
- Berry, M. J. A., & Linoff, G. (1997). *Data mining techniques for marketing, sales, and customer support*. New York: John Wiley.
- Beygelzimer, A., Kakade, S., & Langford, J. (2006). Cover trees for nearest neighbor. In *Proceedings of the 23rd International Conference on Machine Learning* (pp. 97–104). New York: ACM Press.
- Bifet, A., Holmes, G., Kirkby, R., & Pfahringer, B. (2010). MOA: Massive online analysis. *Journal of Machine Learning Research*, 9, 1601–1604.
- Bigus, J. P. (1996). *Data mining with neural networks*. New York: McGraw-Hill.
- Bishop, C. M. (1995). *Neural networks for pattern recognition*. New York: Oxford University Press.
- . (2006). *Pattern recognition and machine learning*. Springer-Verlag.
- BLI (Bureau of Labour Information) (1988). *Collective Bargaining Review (November)*. Ottawa: Labour Canada, Bureau of Labour Information.
- Blockeel, H., Page, D., & Srinivasan, A. (2005). Multi-instance tree learning. In *Proceedings of the 22nd International Conference on Machine Learning* (pp. 57–64). Bonn. New York: ACM Press.
- Blum, A., & Mitchell, T. (1998). Combining labeled and unlabeled data with co-training. In *Proceedings of the Eleventh Annual Conference on Computational Learning Theory* (pp. 92–100). Madison, WI. San Francisco: Morgan Kaufmann.
- Bouckaert, R. R. (1995). *Bayesian belief networks: From construction to inference*. Ph.D. Dissertation, Computer Science Department, University of Utrecht, The Netherlands.
- . (2004). Bayesian network classifiers in Weka. Working Paper 14/2004, Department of Computer Science, University of Waikato, New Zealand.
- . (2010). DensiTree: Making sense of sets of phylogenetic trees. *Bioinformatics*, 26(10), 1372–1373.
- Brachman, R. J., & Levesque, H. J. (Eds.) (1985). *Readings in knowledge representation*. San Francisco: Morgan Kaufmann.

- Brefeld, U., & Scheffer, T. (2004). Co-EM support vector learning. In R. Greiner, & D. Schuurmans (Eds.), *Proceedings of the Twenty-First International Conference on Machine Learning* (pp. 121–128). Banff, AB: New York: ACM Press.
- Breiman, L. (1996a). Stacked regression. *Machine Learning*, 24(1), 49–64.
- . (1996b). Bagging predictors. *Machine Learning*, 24(2), 123–140.
- . (1996c). [Bias, variance, and] arcing classifiers. Technical Report 460. Department of Statistics, University of California, Berkeley.
- . (1999). Pasting small votes for classification in large databases and online. *Machine Learning*, 36(1–2), 85–103.
- . (2001). Random forests. *Machine Learning*, 45(1), 5–32.
- Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification and regression trees*. Monterey, CA: Wadsworth.
- Brin, S., Motwani, R., Ullman, J. D., & Tsur, S. (1997). Dynamic itemset counting and implication rules for market basket data. *ACM SIGMOD Record*, 26(2), 255–264.
- Brin, S., & Page, L. (1998). The anatomy of a large-scale hypertext search engine. *Computer Networks and ISDN Systems*, 33, 107–117.
- Brodley, C. E., & Friedl, M. A. (1996). Identifying and eliminating mislabeled training instances. In *Proceedings of the Thirteenth National Conference on Artificial Intelligence* (pp. 799–805). Portland, OR: Menlo Park, CA: AAAI Press.
- Brownstown, L., Farrell, R., Kant, E., & Martin, N. (1985). *Programming expert systems in OPS5*. Reading, MA: Addison-Wesley.
- Buntine, W. (1992). Learning classification trees. *Statistics and Computing*, 2(2), 63–73.
- Burges, C. J. C. (1998). A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery*, 2(2), 121–167.
- Cabena, P., Hadjinian, P., Stadler, R., Verhees, J., & Zanasi, A. (1998). *Discovering data mining: From concept to implementation*. Upper Saddle River, NJ: Prentice-Hall.
- Cardie, C. (1993). Using decision trees to improve case-based learning. In P. Utgoff (Ed.), *Proceedings of the Tenth International Conference on Machine Learning* (pp. 25–32). Amherst, MA: San Francisco: Morgan Kaufmann.
- Califf, M. E., & Mooney, R. J. (1999). Relational learning of pattern-match rules for information extraction. In *Proceedings of the Sixteenth National Conference on Artificial Intelligence* (pp. 328–334). Orlando: Menlo Park, CA: AAAI Press.
- Cavnar, W. B., & Trenkle, J. M. (1994). N-Gram-based text categorization. In *Proceedings of the Third Symposium on Document Analysis and Information Retrieval* (pp. 161–175). Las Vegas: UNLV Publications/Reprographics.
- Ceglar, A., & Roddick, J. F. (2006). Association mining. *ACM Computing Surveys*, 38(2). New York: ACM Press.
- Cendrowska, J. (1987). PRISM: An algorithm for inducing modular rules. *International Journal of Man-Machine Studies*, 27(4), 349–370.
- Chakrabarti, S. (2003). *Mining the web: discovering knowledge from hypertext data*. San Francisco: Morgan Kaufmann.
- Chang, C.-C., & Lin, C.-J. (2001). LIBSVM: A library for support vector machines. Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
- Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence*, 16, 321–357.
- Cheeseman, P., & Stutz, J. (1995). Bayesian classification (AutoClass): Theory and results. In U. M. Fayyad, G. Piatetsky-Shapiro, P. Smyth, & R. Uthurusamy (Eds.), *Advances*

- in *Knowledge Discovery and Data Mining* (pp. 153–180). Menlo Park, CA: AAAI Press.
- Chen, M. S., Jan, J., & Yu, P. S. (1996). Data mining: An overview from a database perspective. *IEEE Transactions on Knowledge and Data Engineering*, 8(6), 866–883.
- Chen, Y., Bi, J., & Wang, J. Z. (2006). MILES: Multiple-instance learning via embedded instance selection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(12), 1931–1947.
- Cherkauer, K. J., & Shavlik, J. W. (1996). Growing simpler decision trees to facilitate knowledge discovery. In E. Simoudis, J. W. Han, & U. Fayyad (Eds.), *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining* (pp. 315–318). Portland, OR: Menlo Park, CA: AAAI Press.
- Chevaleyre, Y., & Zucker, J.-D. (2001). Solving multiple-instance and multiple-part learning problems with decision trees and rule sets: Application to the mutagenesis problem. In *Proceedings of the Biennial Conference of the Canadian Society for Computational Studies of Intelligence* (pp. 204–214). Ottawa. Berlin: Springer-Verlag.
- Cleary, J. G., & Trigg, L. E. (1995). K*: An instance-based learner using an entropic distance measure. In A. Prieditis, & S. Russell (Eds.), *Proceedings of the Twelfth International Conference on Machine Learning* (pp. 108–114). Tahoe City, CA. San Francisco: Morgan Kaufmann.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20, 37–46.
- Cohen, W. W. (1995). Fast effective rule induction. In A. Prieditis, & S. Russell (Eds.), *Proceedings of the Twelfth International Conference on Machine Learning* (pp. 115–123). Tahoe City, CA. San Francisco: Morgan Kaufmann.
- Cooper, G. F., & Herskovits, E. (1992). A Bayesian method for the induction of probabilistic networks from data. *Machine Learning*, 9(4), 309–347.
- Cortes, C., & Vapnik, V. (1995). Support vector networks. *Machine Learning*, 20(3), 273–297.
- Cover, T. M., & Hart, P. E. (1967). Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, IT-13, 21–27.
- Cristianini, N., & Shawe-Taylor, J. (2000). *An introduction to support vector machines and other kernel-based learning methods*. Cambridge, UK: Cambridge University Press.
- Cypher, A. (Ed.). (1993). *Watch what I do: Programming by demonstration*. Cambridge, MA: MIT Press.
- Dasgupta, S. (2002). Performance guarantees for hierarchical clustering. In J. Kivinen, & R. H. Sloan (Eds.), *Proceedings of the Fifteenth Annual Conference on Computational Learning Theory* (pp. 351–363). Sydney. Berlin: Springer-Verlag.
- Dasu, T., Koutsofios, E., & Wright, J. (2006). Zen and the art of data mining. In *Proceedings of the KDD Workshop on Data Mining for Business Applications* (pp. 37–43). Philadelphia. Proceedings at: http://labs.accenture.com/kdd2006_workshop/dmba_proceedings.pdf
- Datta, S., Kargupta, H., & Sivakumar, K. (2003). Homeland defense, privacy-sensitive data mining, and random value distortion. In *Proceedings of the Workshop on Data Mining for Counter Terrorism and Security*. San Francisco. Philadelphia: Society for International and Applied Mathematics.
- Day, W. H. E., & Edelsbrunner, H. (1984). Efficient algorithms for agglomerative hierarchical clustering methods. *Journal of Classification*, 1(1), 7–24.
- Demiroz, G., & Guvenir, A. (1997). Classification by voting feature intervals. In M. van Someren, & G. Widmer (Eds.), *Proceedings of the Ninth European Conference on Machine Learning* (pp. 85–92). Prague. Berlin: Springer-Verlag.

- de Raedt, L. (2008). *Logical and relational learning*. New York: Springer-Verlag.
- Devroye, L., Györfi, L., & Lugosi, G. (1996). *A probabilistic theory of pattern recognition*. New York: Springer-Verlag.
- Dhar, V., & Stein, R. (1997). *Seven methods for transforming corporate data into business intelligence*. Upper Saddle River, NJ: Prentice-Hall.
- Diederich, J., Kindermann, J., Leopold, E., & Paass, G. (2003). Authorship attribution with support vector machines. *Applied Intelligence*, 19(1), 109–123.
- Dietterich, T. G. (2000). An experimental comparison of three methods for constructing ensembles of decision trees: Bagging, boosting, and randomization. *Machine Learning*, 40(2), 139–158.
- Dietterich, T. G., & Bakiri, G. (1995). Solving multiclass learning problems via error-correcting output codes. *Journal of Artificial Intelligence Research*, 2, 263–286.
- Dietterich, T. G., & Kong, E. B. (1995). Error-correcting output coding corrects bias and variance. In *Proceedings of the Twelfth International Conference on Machine Learning* (pp. 313–321). Tahoe City, CA. San Francisco: Morgan Kaufmann.
- Dietterich, T. G., Lathrop, R. H., & Lozano-Perez, T. (1997). Solving the multiple-instance problem with axis-parallel rectangles. *Artificial Intelligence Journal*, 89(1-2), 31–71.
- Domingos, P. (1997). Knowledge acquisition from examples via multiple models. In D. H. Fisher (Ed.), *Proceedings of the Fourteenth International Conference on Machine Learning* (pp. 98–106). Nashville. San Francisco: Morgan Kaufmann.
- . (1999). MetaCost: A general method for making classifiers cost-sensitive. In U. M. Fayyad, S. Chaudhuri, & D. Madigan (Eds.), *Proceedings of the Fifth International Conference on Knowledge Discovery and Data Mining* (pp. 155–164). San Diego. New York: ACM Press.
- Domingos, P., & Hulten, G. (2000). Mining high-speed data streams. In *International Conference on Knowledge Discovery and Data Mining* (pp. 71–80). Boston. New York: ACM Press.
- Dong, L., Frank, E., & Kramer, S. (2005). Ensembles of balanced nested dichotomies for multi-class problems. In *Proceedings of the Ninth European Conference on Principles and Practice of Knowledge Discovery in Databases* (pp. 84–95). Porto, Portugal. Berlin: Springer-Verlag.
- Dougherty, J., Kohavi, R., & Sahami, M. (1995). Supervised and unsupervised discretization of continuous features. In A. Friediris, & S. Russell (Eds.), *Proceedings of the Twelfth International Conference on Machine Learning* (pp. 194–202). Tahoe City, CA. San Francisco: Morgan Kaufmann.
- Drucker, H. (1997). Improving regressors using boosting techniques. In D. H. Fisher (Ed.), *Proceedings of the Fourteenth International Conference on Machine Learning* (pp. 107–115). Nashville. San Francisco: Morgan Kaufmann.
- Drummond, C., & Holte, R. C. (2000). Explicitly representing expected cost: An alternative to ROC representation. In R. Ramakrishnan, S. Stolfo, R. Bayardo, & I. Parsa (Eds.), *Proceedings of the Sixth International Conference on Knowledge Discovery and Data Mining* (pp. 198–207). Boston. New York: ACM Press.
- Duda, R. O., & Hart, P. E. (1973). *Pattern classification and scene analysis*. New York: John Wiley.
- Duda, R. O., Hart, P. E., & Stork, D. G. (2001). *Pattern Classification* (2nd ed.). New York: John Wiley.
- Dumais, S. T., Platt, J., Heckerman, D., & Sahami, M. (1998). Inductive learning algorithms and representations for text categorization. In *Proceedings of the ACM Seventh*

- International Conference on Information and Knowledge Management* (pp. 148–155). Bethesda, MD. New York: ACM Press.
- Efron, B., & Tibshirani, R. (1993). *An introduction to the bootstrap*. London: Chapman and Hall.
- Egan, J. P. (1975). *Signal detection theory and ROC analysis. Series in Cognition and Perception*. New York: Academic Press.
- Ester, M., Kriegel, H.-P., Sander, J., & Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD-96)* (pp. 226–231). Menlo Park, CA: AAAI Press.
- Fan, R.-E., Chang, K.-W., Hsieh, C.-J., Wang, X.-R., & Lin, C.-J. (2008). LIBLINEAR: A library for large linear classification. *Journal of Machine Learning Research*, 9, 1871–1874.
- Fayyad, U. M., & Irani, K. B. (1993). Multi-interval discretization of continuous-valued attributes for classification learning. In *Proceedings of the Thirteenth International Joint Conference on Artificial Intelligence* (pp. 1022–1027). Chambéry, France. San Francisco: Morgan Kaufmann.
- Fayyad, U. M., & Smyth, P. (1995). From massive datasets to science catalogs: Applications and challenges. In *Proceedings of the Workshop on Massive Datasets* (pp. 129–141). Washington, DC: NRC, Committee on Applied and Theoretical Statistics.
- Fayyad, U. M., Piatetsky-Shapiro, G., Smyth, P., & Uthurusamy, R. (Eds.), (1996). *Advances in knowledge discovery and data mining*. Menlo Park, CA: AAAI Press/MIT Press.
- Fisher, D. (1987). Knowledge acquisition via incremental conceptual clustering. *Machine Learning*, 2(2), 139–172.
- Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. *Annual Eugenics*, 7 (part II), 179–188. Reprinted in *Contributions to Mathematical Statistics*, 1950. New York: John Wiley.
- Fix, E., & Hodges, J. L. Jr. (1951). Discriminatory analysis; non-parametric discrimination: Consistency properties. Technical Report 21-49-004(4), USAF School of Aviation Medicine, Randolph Field, TX.
- Flach, P. A., & Lachiche, N. (1999). Confirmation-guided discovery of first-order rules with Tertius. *Machine Learning*, 42, 61–95.
- Fletcher, R. (1987). *Practical methods of optimization* (2nd ed.). New York: John Wiley.
- Foulds, J., & Frank, E. (2008). Revisiting multiple-instance learning via embedded instance selection. In *Proceedings of the Australasian Joint Conference on Artificial Intelligence* (pp. 300–310). Auckland. Berlin: Springer-Verlag.
- . (2010). A review of multi-instance learning assumptions. *Knowledge Engineering Review*, 25(1), 1–25.
- Fradkin, D., & Madigan, D. (2003). Experiments with random projections for machine learning. In L. Getoor, T. E. Senator, P. Domingos, & C. Faloutsos (Eds.), *Proceedings of the Ninth International Conference on Knowledge Discovery and Data Mining* (pp. 517–522). Washington, DC. New York: ACM Press.
- Frank E. (2000). *Pruning decision trees and lists*. Ph.D. Dissertation, Department of Computer Science, University of Waikato, New Zealand.
- Frank, E., & Hall, M. (2001). A simple approach to ordinal classification. In L. de Raedt, & P. A. Flach (Eds.), *Proceedings of the Twelfth European Conference on Machine Learning* (pp. 145–156). Freiburg, Germany. Berlin: Springer-Verlag.

- Frank, E., Hall, M., & Pfahringer, B. (2003). Locally weighted Naïve Bayes. In U. Kjærulff, & C. Meek (Eds.), *Proceedings of the Nineteenth Conference on Uncertainty in Artificial Intelligence* (pp. 249–256). Acapulco. San Francisco: Morgan Kaufmann.
- Frank, E., Holmes, G., Kirkby, R., & Hall, M. (2002). Racing Committees for Large Datasets. In S. Lange, K. Satoh, & C. H. Smith (Eds.), *Proceedings of the Fifth International Conference on Discovery Science* (pp. 153–164). Lübeck, Germany. Berlin: Springer-Verlag.
- Frank, E., & Kramer, S. (2004). Ensembles of nested dichotomies for multi-class problems. In *Proceedings of the Twenty-First International Conference on Machine Learning* (pp. 305–312). Banff, AB. New York: ACM Press.
- Frank, E., Paynter, G. W., Witten, I. H., Gutwin, C., & Nevill-Manning, C. G. (1999). Domain-specific key phrase extraction. In *Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence* (pp. 668–673). Stockholm. San Francisco: Morgan Kaufmann.
- Frank, E., Wang, Y., Inglis, S., Holmes, G., & Witten, I. H. (1998). Using model trees for classification. *Machine Learning*, 32(1), 63–76.
- Frank, E., & Witten, I. H. (1998). Generating accurate rule sets without global optimization. In J. Shavlik (Ed.), *Proceedings of the Fifteenth International Conference on Machine Learning* (pp. 144–151). Madison, WI. San Francisco: Morgan Kaufmann.
- . (1999). Making better use of global discretization. In I. Bratko, & S. Dzeroski (Eds.), *Proceedings of the Sixteenth International Conference on Machine Learning* (pp. 115–123). Bled, Slovenia. San Francisco: Morgan Kaufmann.
- Frank, E., & Xu, X. (2003). Applying propositional learning algorithms to multi-instance data. Technical Report 06/03, Department of Computer Science, University of Waikato, New Zealand.
- Freitag, D. (2002). Machine learning for information extraction in informal domains. *Machine Learning*, 39(2/3), 169–202.
- Freund, Y., & Mason, L. (1999). The alternating decision tree learning algorithm. In I. Bratko, & S. Dzeroski (Eds.), *Proceedings of the Sixteenth International Conference on Machine Learning* (pp. 124–133). Bled, Slovenia. San Francisco: Morgan Kaufmann.
- Freund, Y., & Schapire, R. E. (1996). Experiments with a new boosting algorithm. In L. Saitta (Ed.), *Proceedings of the Thirteenth International Conference on Machine Learning* (pp. 148–156). Bari, Italy. San Francisco: Morgan Kaufmann.
- . (1999). Large margin classification using the perceptron algorithm. *Machine Learning*, 37(3), 277–296.
- Friedman, J. H. (1996). Another approach to polychotomous classification. Technical report, Department of Statistics, Stanford University, Stanford, CA.
- . (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, 29(5), 1189–1232.
- Friedman, J. H., Bentley, J. L., & Finkel, R. A. (1977). An algorithm for finding best matches in logarithmic expected time. *ACM Transactions on Mathematical Software*, 3(3), 209–266.
- Friedman, J. H., Hastie, T., & Tibshirani, R. (2000). Additive logistic regression: A statistical view of boosting. *Annals of Statistics*, 28(2), 337–374.
- Friedman, N., Geiger, D., & Goldszmidt, M. (1997). Bayesian Network Classifiers. *Machine Learning*, 29(2), 131–163.
- Fulton, T., Kasif, S., & Salzberg, S. (1995). Efficient algorithms for finding multiway splits for decision trees. In A. Prieditis, & S. Russell (Eds.), *Proceedings of the Twelfth*

- International Conference on Machine Learning* (pp. 244–251). Tahoe City, CA. San Francisco: Morgan Kaufmann.
- Fürnkranz, J. (2002). Round robin classification. *Journal of Machine Learning Research*, 2, 721–747.
- . (2003). Round robin ensembles. *Intelligent Data Analysis*, 7(5), 385–403.
- Fürnkranz, J., & Flach, P. A. (2005). ROC ‘n’ rule learning: Towards a better understanding of covering algorithms. *Machine Learning*, 58(1), 39–77.
- Fürnkranz, J., & Widmer, G. (1994). Incremental reduced-error pruning. In H. Hirsh, & W. Cohen (Eds.), *Proceedings of the Eleventh International Conference on Machine Learning* (pp. 70–77). New Brunswick, NJ. San Francisco: Morgan Kaufmann.
- Gaines, B. R., & Compton, P. (1995). Induction of ripple-down rules applied to modeling large data bases. *Journal of Intelligent Information Systems*, 5(3), 211–228.
- Gama, J. (2004). Large margin trees. *Machine Learning*, 55(3), 219–250.
- Gärtner, T., Flach, P. A., Kowalczyk, A., & Smola, A. J. (2002). Multi-instance kernels. In *Proceedings of the International Conference on Machine Learning* (pp. 179–186). Sydney. San Francisco: Morgan Kaufmann.
- Genkin, A., Lewis, D. D., & Madigan, D. (2007). Large-scale Bayesian logistic regression for text categorization. *Technometrics*, 49(3), 291–304.
- Gennari, J. H., Langley, P., & Fisher, D. (1990). Models of incremental concept formation. *Artificial Intelligence*, 40, 11–61.
- Ghani, R. (2002). Combining labeled and unlabeled data for multiclass text categorization. In C. Sammut, & A. Hoffmann (Eds.), *Proceedings of the Nineteenth International Conference on Machine Learning* (pp. 187–194). Sydney. San Francisco: Morgan Kaufmann.
- Gilad-Bachrach, R., Navot, A., & Tishby, N. (2004). Margin based feature selection: Theory and algorithms. In R. Greiner, & D. Schuurmans (Eds.), *Proceedings of the Twenty-First International Conference on Machine Learning* (pp. 337–344). Banff, AB. New York: ACM Press.
- Giraud-Carrier, C. (1996). FLARE: Induction with prior knowledge. In J. Nealon, & J. Hunt (Eds.), *Research and Development in Expert Systems XIII* (pp. 11–24). Cambridge, England: SGES Publications.
- Gluck, M., & Corter, J. (1985). Information, uncertainty and the utility of categories. In *Proceedings of the Annual Conference of the Cognitive Science Society* (pp. 283–287). Irvine, CA. Hillsdale, NJ: Lawrence Erlbaum.
- Goldberg, D. E. (1989). *Genetic algorithms in search, optimization and machine learning*. Reading, MA: Addison-Wesley.
- Good P. (1994). *Permutation tests: A practical guide to resampling methods for testing hypotheses*. New York: Springer-Verlag.
- Grossman, D., & Domingos, P. (2004). Learning Bayesian network classifiers by maximizing conditional likelihood. In R. Greiner, & D. Schuurmans (Eds.), *Proceedings of the Twenty-First International Conference on Machine Learning* (pp. 361–368). Banff, AB. New York: ACM Press.
- Groth, R. (1998). *Data mining: A hands-on approach for business professionals*. Upper Saddle River, NJ: Prentice-Hall.
- Guo, Y., & Greiner, R. (2004). *Discriminative model selection for belief net structures*. Canada: Department of Computing Science, TR04-22, University of Alberta.
- Gütlein, M., Frank, E., Hall, M., & Karwath, A. (2009). Large-scale attribute selection using wrappers. In *Proceedings of the IEEE Symposium on Computational Intelligence and Data Mining* (pp. 332–339). Nashville. Washington, DC: IEEE Computer Society.

- Guyon, I., Weston, J., Barnhill, S., & Vapnik, V. (2002). Gene selection for cancer classification using support vector machines. *Machine Learning*, 46(1–3), 389–422.
- Hall, M. (2000). Correlation-based feature selection for discrete and numeric class machine learning. In P. Langley (Ed.), *Proceedings of the Seventeenth International Conference on Machine Learning* (pp. 359–366). Stanford, CA: San Francisco: Morgan Kaufmann.
- Hall, M., Holmes, G., & Frank, E. (1999). Generating rule sets from model trees. In N. Y. Foo (Ed.), *Proceedings of the Twelfth Australian Joint Conference on Artificial Intelligence* (pp. 1–12). Sydney: Berlin: Springer-Verlag.
- Hall, M., & Frank, E. (2008). Combining Naïve Bayes and decision tables. In *Proceedings of the 21st Florida Artificial Intelligence Research Society Conference* (pp. 318–319). Miami: Menlo Park, CA: AAAI Press.
- Han, J., Pei, J., & Yin, Y. (2000). Mining frequent patterns without candidate generation. In *Proceedings of the ACM-SIGMOD International Conference on Management of Data* (pp. 1–12). Dallas: New York: ACM Press.
- Han, J., Pei, J., Yin, Y., & Mao, R. (2004). Mining frequent patterns without candidate generation: A frequent-pattern tree approach. *Data Mining and Knowledge Discovery*, 8(1), 53–87.
- Han, J., & Kamber, M. (2006). *Data mining: Concepts and techniques* (2nd ed.). San Francisco: Morgan Kaufmann.
- Hand, D. J. (2006). Classifier Technology and the Illusion of Progress. *Statistical Science*, 21(1), 1–14.
- Hand, D. J., Manilla, H., & Smyth, P. (2001). *Principles of Data Mining*. Cambridge, MA: MIT Press.
- Hartigan, J. A. (1975). *Clustering algorithms*. New York: John Wiley.
- Hastie, T., & Tibshirani, R. (1998). Classification by pairwise coupling. *Annals of Statistics*, 26(2), 451–471.
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning* (2nd ed.). New York: Springer-Verlag.
- Heckerman, D., Geiger, D., & Chickering, D. M. (1995). Learning Bayesian networks: The combination of knowledge and statistical data. *Machine Learning*, 20(3), 197–243.
- Hempstalk, K., Frank, E., & Witten, I. H. (2008). One-class classification by combining density and class probability estimation. In *Proceedings of the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases* (pp. 505–519). Antwerp: Berlin: Springer-Verlag.
- Hempstalk, K., & Frank, E. (2008). Discriminating against new classes: One-class versus multi-class classification. In *Proceedings of the Twenty-first Australasian Joint Conference on Artificial Intelligence*. Auckland (pp. 225–236). New York: Springer-Verlag.
- Ho, T. K. (1998). The random subspace method for constructing decision forests. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(8), 832–844.
- Hochbaum, D. S., & Shmoys, D. B. (1985). A best possible heuristic for the k -center problem. *Mathematics of Operations Research*, 10(2), 180–184.
- Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: applications to nonorthogonal problems. *Technometrics*, 12(1), 69–82.
- Holmes, G., & Nevill-Manning, C. G. (1995). Feature selection via the discovery of simple classification rules. In G. E. Lasker, & X. Liu (Eds.), *Proceedings of the International Symposium on Intelligent Data Analysis* (pp. 75–79). Baden-Baden, Germany: International Institute for Advanced Studies in Systems Research and Cybernetics. Baden-Baden.

- Windsor, Ont: International Institute for Advanced Studies in Systems Research and Cybernetics.
- Holmes, G., Pfahringer, B., Kirkby, R., Frank, E., & Hall, M. (2002). Multiclass alternating decision trees. In T. Elomaa, H. Mannila, & H. Toivonen (Eds.), *Proceedings of the Thirteenth European Conference on Machine Learning* (pp. 161–172). Helsinki. Berlin: Springer-Verlag.
- Holte, R. C. (1993). Very simple classification rules perform well on most commonly used datasets. *Machine Learning*, 11, 63–91.
- Huffman, S. B. (1996). Learning information extraction patterns from examples. In S. Wertmer, E. Riloff, & G. Scheler (Eds.), *Connectionist, statistical, and symbolic approaches to learning for natural language processing* (pp. 246–260). Berlin: Springer-Verlag.
- Jabbour, K., Riveros, J. F. V., Landsbergen, D., & Meyer, W. (1988). ALFA: Automated load forecasting assistant. *IEEE Transactions on Power Systems*, 3(3), 908–914.
- Jiang, L., & Zhang, H. (2006). Weightily averaged one-dependence estimators. In *Proceedings of the 9th Biennial Pacific Rim International Conference on Artificial Intelligence* (pp. 970–974). Guilin, China. Berlin: Springer-Verlag.
- John, G. H. (1995). Robust decision trees: Removing outliers from databases. In U. M. Fayyad, & R. Uthurusamy (Eds.), *Proceedings of the First International Conference on Knowledge Discovery and Data Mining* (pp. 174–179). Montreal. Menlo Park, CA: AAAI Press.
- . (1997). *Enhancements to the data mining process*. Ph.D. Dissertation, Computer Science Department, Stanford University, Stanford, CA.
- John, G. H., Kohavi, R., & Pflieger, P. (1994). Irrelevant features and the subset selection problem. In H. Hirsh, & W. Cohen (Eds.), *Proceedings of the Eleventh International Conference on Machine Learning* (pp. 121–129). New Brunswick, NJ. San Francisco: Morgan Kaufmann.
- John, G. H., & Langley, P. (1995). Estimating continuous distributions in Bayesian classifiers. In P. Besnard, & S. Hanks (Eds.), *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence* (pp. 338–345). Montreal. San Francisco: Morgan Kaufmann.
- Johns, M. V. (1961). An empirical Bayes approach to nonparametric two-way classification. In H. Solomon (Ed.), *Studies in item analysis and prediction*. Palo Alto, CA: Stanford University Press.
- Kass, R., & Wasserman, L. (1995). A reference Bayesian test for nested hypotheses and its relationship to the Schwarz criterion. *Journal of the American Statistical Association*, 90, 928–934.
- Keerthi, S. S., Shevade, S. K., Bhattacharyya, C., & Murthy, K. R. K. (2001). Improvements to Platt's SMO algorithm for SVM classifier design. *Neural Computation*, 13(3), 637–649.
- Kerber, R. (1992). Chimerge: Discretization of numeric attributes. In W. Swartout (Ed.), *Proceedings of the Tenth National Conference on Artificial Intelligence* (pp. 123–128). San Jose, CA. Menlo Park, CA: AAAI Press.
- Kibler, D., & Aha, D. W. (1987). Learning representative exemplars of concepts: An initial case study. In P. Langley (Ed.), *Proceedings of the Fourth Machine Learning Workshop* (pp. 24–30). Irvine, CA. San Francisco: Morgan Kaufmann.
- Kimball, R., & Ross, M. (2002). *The data warehouse toolkit* (2nd ed.). New York: John Wiley.
- Kira, K., & Rendell, L. (1992). A practical approach to feature selection. In D. Sleeman, & P. Edwards (Eds.), *Proceedings of the Ninth International Workshop on Machine Learning* (pp. 249–258). Aberdeen, Scotland. San Francisco: Morgan Kaufmann.

- Kirkby, R. (2007). *Improving Hoeffding trees*. Ph.D. Dissertation, Department of Computer Science, University of Waikato, New Zealand.
- Kittler, J. (1978). Feature set search algorithms. In C. H. Chen (Ed.), *Pattern recognition and signal processing* (pp. 41–60). Amsterdam: Sijthoff an Noordhoff.
- Kivinen, J., Smola, A. J., & Williamson, R. C. (2002). Online learning with kernels. *IEEE Transactions on Signal Processing*, 52, 2165–2176.
- Kleinberg, J. (1998). Authoritative sources in a hyperlinked environment. In *Proceedings of the ACM-SIAM Symposium on Discrete Algorithms* (pp. 604–632). Extended version published in *Journal of the ACM* 46 (1999).
- Koestler, A. (1964). *The act of creation*. London: Hutchinson.
- Kohavi, R. (1995a). A study of cross-validation and bootstrap for accuracy estimation and model selection. In *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence* (pp. 1137–1143). Montreal. San Francisco: Morgan Kaufmann.
- . (1995b). The power of decision tables. In N. Lavrac, & S. Wrobel (Eds.), *Proceedings of the Eighth European Conference on Machine Learning* (pp. 174–189). Iráklion, Crete. Berlin: Springer-Verlag.
- . (1996). Scaling up the accuracy of Naïve Bayes classifiers: A decision-tree hybrid. In E. Simoudis, J. W. Han, & U. Fayyad (Eds.), *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining* (pp. 202–207). Portland, OR. Menlo Park, CA: AAAI Press.
- Kohavi, R., & John, G. H. (1997). Wrappers for feature subset selection. *Artificial Intelligence*, 97(1–2), 273–324.
- Kohavi, R., & Kunz, C. (1997). Option decision trees with majority votes. In D. Fisher (Ed.), *Proceedings of the Fourteenth International Conference on Machine Learning* (pp. 161–191). Nashville. San Francisco: Morgan Kaufmann.
- Kohavi, R., & Provost, F. (Eds.), (1998). Machine learning: Special issue on applications of machine learning and the knowledge discovery process. *Machine Learning*, 30(2/3), 127–274.
- Kohavi, R., & Sahami, M. (1996). Error-based and entropy-based discretization of continuous features. In E. Simoudis, J. W. Han, & U. Fayyad (Eds.), *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining* (pp. 114–119). Portland, OR. Menlo Park, CA: AAAI Press.
- Komarek, P., & Moore, A. (2000). A dynamic adaptation of AD-trees for efficient machine learning on large data sets. In P. Langley (Ed.), *Proceedings of the Seventeenth International Conference on Machine Learning* (pp. 495–502). Stanford, CA. San Francisco: Morgan Kaufmann.
- Kononenko, I. (1995). On biases in estimating multi-valued attributes. In *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence* (pp. 1034–1040). Montreal. San Francisco: Morgan Kaufmann.
- Koppel, M., & Schler, J. (2004). Authorship verification as a one-class classification problem. In R. Greiner, & D. Schuurmans (Eds.), *Proceedings of the Twenty-First International Conference on Machine Learning* (pp. 489–495). Banff, AB. New York: ACM Press.
- Kroegel, M.-A., & Wrobel, S. (2002). Feature selection for propositionalization. In *Proceedings of the International Conference on Discovery Science* (pp. 430–434). Lübeck, Germany. Berlin: Springer-Verlag.
- Kubat, M., Holte, R. C., & Matwin, S. (1998). Machine learning for the detection of oil spills in satellite radar images. *Machine Learning*, 30, 195–215.

- Kuncheva, L. I., & Rodriguez, J. J. (2007). An experimental study on rotation forest ensembles. In *Proceedings of the Seventh International Workshop on Multiple Classifier Systems* (pp. 459–468). Prague. Berlin/Heidelberg: Springer-Verlag.
- Kushmerick, N., Weld, D. S., & Doorenbos, R. (1997). Wrapper induction for information extraction. In *Proceedings of the Fifteenth International Joint Conference on Artificial Intelligence* (pp. 729–735). Nagoya, Japan. San Francisco: Morgan Kaufmann.
- Laguna, M., & Marti, R. (2003). *Scatter search: Methodology and implementations in C*. Dordrecht, The Netherlands: Kluwer Academic Press.
- Landwehr, N., Hall, M., & Frank, E. (2005). Logistic model trees. *Machine Learning*, 59(1–2), 161–205.
- Langley, P. (1996). *Elements of machine learning*. San Francisco: Morgan Kaufmann.
- Langley, P., Iba, W., & Thompson, K. (1992). An analysis of Bayesian classifiers. In W. Swartout (Ed.), *Proceedings of the Tenth National Conference on Artificial Intelligence* (pp. 223–228). San Jose, CA. Menlo Park, CA: AAAI Press.
- Langley, P., & Sage, S. (1994). Induction of selective Bayesian classifiers. In R. L. de Mantaras, & D. Poole (Eds.), *Proceedings of the Tenth Conference on Uncertainty in Artificial Intelligence* (pp. 399–406). Seattle. San Francisco: Morgan Kaufmann.
- Langley, P., & Simon, H. A. (1995). Applications of machine learning and rule induction. *Communications of the ACM*, 38(11), 55–64.
- Lavrac, N., Motoda, H., Fawcett, T., Holte, R., Langley, P., & Adriaans, P. (Eds.), (2004). Special issue on lessons learned from data mining applications and collaborative problem solving. *Machine Learning*, 57(1/2).
- Lawson, C. L., & Hanson, R. J. (1995). *Solving least squares problems*. Philadelphia: SIAM Publications.
- le Cessie, S., & van Houwelingen, J. C. (1992). Ridge estimators in logistic regression. *Applied Statistics*, 41(1), 191–201.
- Li, M., & Vitanyi, P. M. B. (1992). Inductive reasoning and Kolmogorov complexity. *Journal of Computer and System Sciences*, 44, 343–384.
- Lieberman, H. (Ed.), (2001). *Your wish is my command: Programming by example*. San Francisco: Morgan Kaufmann.
- Littlestone, N. (1988). Learning quickly when irrelevant attributes abound: A new linear-threshold algorithm. *Machine Learning*, 2(4), 285–318.
- . (1989). *Mistake bounds and logarithmic linear-threshold learning algorithms*. Ph. D. Dissertation, University of California, Santa Cruz.
- Liu, B. (2009) *Web data mining: Exploring hyperlinks, contents, and usage data*. Berlin: Springer-Verlag.
- Liu, B., Hsu, W., & Ma, Y. M. (1998). Integrating classification and association rule mining. In *Proceedings of the Fourth International Conference on Knowledge Discovery and Data Mining (KDD-98)* (pp. 80–86). New York. Menlo Park, CA: AAAI Press.
- Liu, H., & Setiono, R. (1996). A probabilistic approach to feature selection: A filter solution. In L. Saitta (Ed.), *Proceedings of the Thirteenth International Conference on Machine Learning* (pp. 319–327), Bari, Italy. San Francisco: Morgan Kaufmann.
- . (1997). Feature selection via discretization. *IEEE Transactions on Knowledge and Data Engineering*, 9(4), 642–645.
- Luan, J. (2002). Data mining and its applications in higher education. *New directions for institutional research*, 2002(113), 17–36.
- Mann, T. (1993). *Library research models: A guide to classification, cataloging, and computers*. New York: Oxford University Press.

- Marill, T., & Green, D. M. (1963). On the effectiveness of receptors in recognition systems. *IEEE Transactions on Information Theory*, 9(11), 11–17.
- Maron, O. (1998). *Learning from ambiguity*. Ph.D. Thesis, Massachusetts Institute of Technology, Cambridge, MA.
- Maron, O., & Lozano-Peréz, T. (1997). A framework for multiple-instance learning. In *Proceedings of the Conference on Neural Information Processing Systems* (pp. 570–576). Denver. Cambridge, MA: MIT Press.
- Martin, B. (1995). *Instance-based learning: Nearest neighbour with generalisation*. M.Sc. Thesis, Department of Computer Science, University of Waikato, New Zealand.
- McCallum A., & Nigam, K. (1998). A comparison of event models for Naïve Bayes text classification. In *Proceedings of the AAAI-98 Workshop on Learning for Text Categorization* (pp. 41–48). Madison, WI. Menlo Park, CA: AAAI Press.
- Medelyan, O., & Witten, I. H. (2008). Domain independent automatic keyphrase indexing with small training sets. *Journal of the American Society for Information Science and Technology*, 59, 1026–1040.
- Mehta, M., Agrawal, R., & Rissanen, J. (1996). SLIQ: A fast scalable classifier for data mining. In P. Apers, M. Bouzeghoub, & G. Gardarin (Eds.), *Proceedings of the Fifth International Conference on Extending Database Technology* (pp. 18–32). Avignon, France. New York: Springer-Verlag.
- Melville, P., & Mooney, R. J. (2005). Creating diversity in ensembles using artificial data. *Information Fusion*, 6(1), 99–111.
- Michalski, R. S., & Chilausky, R. L. (1980). Learning by being told and learning from examples: An experimental comparison of the two methods of knowledge acquisition in the context of developing an expert system for soybean disease diagnosis. *International Journal of Policy Analysis and Information Systems*, 4(2).
- Michie, D. (1989). Problems of computer-aided concept formation. In J. R. Quinlan (Ed.), *Applications of expert systems (Vol. 2)* (pp. 310–333). Wokingham, UK: Addison-Wesley.
- Minsky, M., & Papert, S. (1969). *Perceptrons*. Cambridge, MA: MIT Press.
- Mitchell, T. M. (1997). *Machine Learning*. New York: McGraw-Hill.
- Mitchell T. M., Caruana, R., Freitag, D., McDermott, J., & Zabowski, D. (1994). Experience with a learning personal assistant. *Communications of the ACM*, 37 (7), 81–91.
- Moore, A. W. (1991). *Efficient memory-based learning for robot control*. Ph.D. Dissertation, Computer Laboratory, University of Cambridge, UK.
- . (2000). The anchors hierarchy: Using the triangle inequality to survive high-dimensional data. In C. Boutilier, & M. Goldszmidt (Eds.), *Proceedings of the Sixteenth Conference on Uncertainty in Artificial Intelligence* (pp. 397–405). Stanford, CA. San Francisco: Morgan Kaufmann.
- Moore, A. W., & Lee, M. S. (1994). Efficient algorithms for minimizing cross validation error. In W. W. Cohen, & H. Hirsh (Eds.), *Proceedings of the Eleventh International Conference on Machine Learning* (pp. 190–198). New Brunswick, NJ. San Francisco: Morgan Kaufmann.
- . (1998). Cached sufficient statistics for efficient machine learning with large datasets. *Journal Artificial Intelligence Research*, 8, 67–91.
- Moore, A. W., & Pelleg, D. (2000). X-means: Extending k -means with efficient estimation of the number of clusters. In P. Langley (Ed.), *Proceedings of the Seventeenth International Conference on Machine Learning* (pp. 727–734). Stanford, CA. San Francisco: Morgan Kaufmann.

- Mutter, S., Hall, M., & Frank, E. (2004). Using classification to evaluate the output of confidence-based association rule mining. In *Proceedings of the Seventeenth Australian Joint Conference on Artificial Intelligence* (pp. 538–549). Cairns, Australia. Berlin: Springer-Verlag.
- Nadeau, C., & Bengio, Y. (2003). Inference for the generalization error. *Machine Learning*, 52(3), 239–281.
- Nahm, U. Y., & Mooney, R. J. (2000). Using information extraction to aid the discovery of prediction rules from texts. In *Proceedings of the Workshop on Text Mining at the Sixth International Conference on Knowledge Discovery and Data Mining* (pp. 51–58). Boston. Workshop proceedings at: <http://www.cs.cmu.edu/~dunja/WshKDD2000.html>
- Niculescu-Mizil, A., & Caruana, R. (2005). Predicting good probabilities with supervised learning. In *Proceedings of the 22nd International Conference on Machine Learning* (pp. 625–632). Bonn. New York: ACM Press.
- Nie, N. H., Hull, C., Jenkins, H., Steinbrenner, J. G. K., & Bent, D. H. (1970). *Statistical package for the social sciences*. New York: McGraw-Hill.
- Nigam, K., & Ghani, R. (2000). Analyzing the effectiveness and applicability of co-training. In *Proceedings of the Ninth International Conference on Information and Knowledge Management* (pp. 86–93). McLean, VA. New York: ACM Press.
- Nigam, K., McCallum, A. K., Thrun, S., & Mitchell, T. M. (2000). Text classification from labeled and unlabeled documents using EM. *Machine Learning*, 39(2/3), 103–134.
- Nilsson, N. J. (1965). *Learning machines*. New York: McGraw-Hill.
- Nisbet, R., Elder, J., & Miner, G. (2009). *Handbook of statistical analysis and data mining applications*. New York: Academic Press.
- Oates, T., & Jensen, D. (1997). The effects of training set size on decision tree complexity. In *Proceedings of the Fourteenth International Conference on Machine Learning* (pp. 254–262). Nashville. San Francisco: Morgan Kaufmann.
- Ohm, P. (2009). Broken promises of privacy: Responding to the surprising failure of anonymization. University of Colorado Law Legal Studies Research Paper No. 09-12, August.
- Omohundro, S. M. (1987). Efficient algorithms with neural network behavior. *Journal of Complex Systems*, 1(2), 273–347.
- Paynter G. W. (2000). *Automating iterative tasks with programming by demonstration*. Ph.D. Dissertation, Department of Computer Science, University of Waikato, New Zealand.
- Pearson, R. (2005). Mining imperfect data. *Society for Industrial and Applied Mechanics*, Philadelphia.
- Pei, J., Han, J., Mortazavi-Asi, B., Wang, J., Pinto, H., Chen, Q., et al. (2004). Mining sequential patterns by pattern-growth: The PrefixSpan approach. *IEEE Transactions on Knowledge and Data Engineering*, 16(11), 1424–1440.
- Piatetsky-Shapiro, G., & Frawley, W. J. (Eds.) (1991). *Knowledge discovery in databases*. Menlo Park, CA: AAAI Press/MIT Press.
- Platt, J. (1998). Fast training of support vector machines using sequential minimal optimization. In B. Schölkopf, C. Burges, & A. Smola (Eds.), *Advances in kernel methods: Support vector learning* (pp. 185–209). Cambridge, MA: MIT Press.
- Power, D. J. (2002). What is the true story about data mining, beer and diapers? *DSS News*, 3(23); see <http://www.dssresources.com/newsletters/66.php>.
- Provost, F., & Fawcett, T. (1997). Analysis and visualization of classifier performance: Comparison under imprecise class and cost distributions. In D. Heckerman, H. Mannila, D. Pregibon, & R. Uthurusamy (Eds.), *Proceedings of the Third International Conference*

- on *Knowledge Discovery and Data Mining* (pp. 43–48). Huntington Beach, CA: Menlo Park, CA: AAAI Press.
- Pyle, D. (1999). *Data preparation for data mining*. San Francisco, CA: Morgan Kaufmann.
- Quinlan, J. R. (1986). Induction of decision trees. *Machine Learning*, 1(1), 81–106.
- . (1992). Learning with continuous classes. In N. Adams, & L. Sterling (Eds.), *Proceedings of the Fifth Australian Joint Conference on Artificial Intelligence* (pp. 343–348). Hobart, Tasmania. Singapore: World Scientific.
- . (1993). *C4.5: Programs for machine learning*. San Francisco: Morgan Kaufmann.
- . (1996). Improved use of continuous attributes in C4.5. *Journal of Artificial Intelligence Research*, 4, 77–90.
- Ramon, J., & de Raedt, L. (2000). Multi instance neural networks. In *Proceedings of the ICML Workshop on Attribute-Value and Relational Learning* (pp. 53–60). Stanford, CA.
- Ray, S., & Craven, M. (2005). Supervised learning versus multiple instance learning: An empirical comparison. In *Proceedings of the International Conference on Machine Learning* (pp. 697–704). Bonn. New York: ACM Press.
- Read, J., Pfahringer, B., Holmes, G., & Frank, E. (2009). Classifier chains for multi-label classification. In *Proceedings of the 13th European Conference on Principles and Practice of Knowledge Discovery in Databases and 20th European Conference on Machine Learning* (pp. 254–269). Bled, Slovenia. Berlin: Springer-Verlag.
- Rennie, J. D. M., Shih, L., Teevan, J., & Karger, D. R. (2003). Tackling the poor assumptions of Naïve Bayes text classifiers. In T. Fawcett, & N. Mishra (Eds.), *Proceedings of the Twentieth International Conference on Machine Learning* (pp. 616–623). Washington, DC: Menlo Park, CA: AAAI Press.
- Ricci, F., & Aha, D. W. (1998). Error-correcting output codes for local learners. In C. Nédellec, & C. Rouveird (Eds.), *Proceedings of the European Conference on Machine Learning* (pp. 280–291). Chemnitz, Germany. Berlin: Springer-Verlag.
- Richards, D., & Compton, P. (1998). Taking up the situated cognition challenge with ripple-down rules. *International Journal of Human-Computer Studies*, 49(6), 895–926.
- Rifkin, R., & Klautau, A. (2004). In defense of one-vs.-all classification. *Journal of Machine Learning Research*, 5, 101–141.
- Ripley, B. D. (1996). *Pattern recognition and neural networks*. Cambridge, UK: Cambridge University Press.
- Rissanen, J. (1985). The minimum description length principle. In S. Kotz, & N. L. Johnson (Eds.), *Encyclopedia of Statistical Sciences* (Vol. 5) (pp. 523–527). New York: John Wiley.
- Rodriguez, J. J., Kuncheva, L. I., & Alonso, C. J. (2006). Rotation forest: A new classifier ensemble method. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(10), 1619–1630.
- Rousseeuw, P. J., & Leroy, A. M. (1987). *Robust regression and outlier detection*. New York: John Wiley.
- Russell, S., & Norvig, P. (2009). *Artificial intelligence: A modern approach* (3rd ed.). Upper Saddle River, NJ: Prentice-Hall.
- Sahami, M., Dumais, S., Heckerman, D., & Horvitz, E. (1998). A Bayesian approach to filtering junk e-mail. In *Proceedings of the AAAI-98 Workshop on Learning for Text Categorization* (pp. 55–62). Madison, WI. Menlo Park, CA: AAAI Press.
- Saitta, L., & Neri, F. (1998). Learning in the “real world.” *Machine Learning*, 30(2/3), 133–163.

- Salzberg, S. (1991). A nearest hyperrectangle learning method. *Machine Learning*, 6(3), 251–276.
- Schapire, R. E., Freund, Y., Bartlett, P., & Lee, W. S. (1997). Boosting the margin: A new explanation for the effectiveness of voting methods. In D. H. Fisher (Ed.), *Proceedings of the Fourteenth International Conference on Machine Learning* (pp. 322–330). Nashville, San Francisco: Morgan Kaufmann.
- Scheffer, T. (2001). Finding association rules that trade support optimally against confidence. In L. de Raedt, & A. Siebes (Eds.), *Proceedings of the Fifth European Conference on Principles of Data Mining and Knowledge Discovery* (pp. 424–435). Freiburg, Germany. Berlin: Springer-Verlag.
- Schölkopf, B., Bartlett, P., Smola, A. J., & Williamson, R. (1999). Shrinking the tube: A new support vector regression algorithm. *Advances in Neural Information Processing Systems*, 11, 330–336. Cambridge, MA: MIT Press.
- Schölkopf, B., Williamson, R., Smola, A., Shawe-Taylor, J., & Platt, J. (2000). Support vector method for novelty detection. *Advances in Neural Information Processing Systems*, 12, 582–588. Cambridge, MA: MIT Press.
- Schölkopf, B., & Smola, A. J. (2002). *Learning with kernels: Support vector machines, regularization, optimization, and beyond*. Cambridge, MA: MIT Press.
- Sebastiani, F. (2002). Machine learning in automated text categorization. *ACM Computing Surveys*, 34(1), 1–47.
- Seewald A. K. (2002). How to make stacking better and faster while also taking care of an unknown weakness. In *Proceedings of the Nineteenth International Conference on Machine Learning* (pp. 54–561). Sydney. San Francisco: Morgan Kaufmann.
- Seewald, A. K., & Fürnkranz, J. (2001). An evaluation of grading classifiers. In F. Hoffmann, D. J. Hand, N. M. Adams, D. H. Fisher, & G. Guimarães (Eds.), *Proceedings of the Fourth International Conference on Advances in Intelligent Data Analysis* (pp. 115–124). Cascais, Portugal. Berlin: Springer-Verlag.
- Shafer, R., Agrawal, R., & Metha, M. (1996). SPRINT: A scalable parallel classifier for data mining. In T. M. Vijayaraman, A. P. Buchmann, C. Mohan, & N. L. Sarda (Eds.), *Proceedings of the Second International Conference on Very Large Databases* (pp. 544–555). Mumbai (Bombay). San Francisco: Morgan Kaufmann.
- Shalev-Shwartz, S., Singer, Y., & Srebro, N. (2007). Pegasos: Primal estimated sub-gradient solver for SVM. In *Proceedings of the 24th international conference on Machine Learning* (pp. 807–814). New York: ACM Press.
- Shawe-Taylor, J., & Cristianini, N. (2004). *Kernel methods for pattern analysis*. Cambridge, UK: Cambridge University Press.
- Slonim, N., Friedman, N., & Tishby, N. (2002). Unsupervised document classification using sequential information maximization. In *Proceedings of the 25th International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 120–136). New York: ACM Press.
- Smola, A. J., & B. Schölkopf. (2004). A tutorial on support vector regression. *Statistics and Computing*, 14(3), 199–222.
- Soderland, S., Fisher, D., Aseltine, J., & Lehnert, W. (1995). Crystal: Inducing a conceptual dictionary. In *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence* (pp. 1314–1319). Montreal. Menlo Park, CA: AAAI Press.
- Srikant, R., & Agrawal, R. (1996). Mining sequential patterns: Generalizations and performance improvements. In P. M. Apers, M. Bouzeghoub, & G. Gardarin (Eds.), *Proceedings*

- of the Fifth International Conference on Extending Database Technology. Avignon, France. Lecture Notes in Computer Science. Vol. 1057 (pp. 3–17). London: Springer-Verlag.
- Stevens, S. S. (1946). On the theory of scales of measurement. *Science*, 103, 677–680.
- Stone, P., & Veloso, M. (2000). Multiagent systems: A survey from a machine learning perspective. *Autonomous Robots*, 8(3), 345–383.
- Stout, Q. F. (2008). Unimodal regression via prefix isotonic regression. *Computational Statistics and Data Analysis*, 53, 289–297.
- Su, J., Zhang, H., Ling, C. X., & Matwin, S. (2008). Discriminative parameter learning for Bayesian networks. In *Proceedings of the 25th International Conference on Machine Learning* (pp. 1016–1023). Helsinki. New York: ACM Press.
- Swets, J. (1988). Measuring the accuracy of diagnostic systems. *Science*, 240, 1285–1293.
- Ting, K. M. (2002). An instance-weighting method to induce cost-sensitive trees. *IEEE Transactions on Knowledge and Data Engineering*, 14(3), 659–665.
- Ting, K. M., & Witten, I. H. (1997a). Stacked generalization: When does it work? In *Proceedings of the Fifteenth International Joint Conference on Artificial Intelligence* (pp. 866–871). Nagoya, Japan. San Francisco: Morgan Kaufmann.
- . (1997b). Stacking bagged and dagged models. In D. H. Fisher (Ed.), *Proceedings of the Fourteenth International Conference on Machine Learning* (pp. 367–375). Nashville. San Francisco: Morgan Kaufmann.
- Turney, P. D. (1999). *Learning to extract key phrases from text*. Technical Report ERB-1057, Institute for Information Technology, National Research Council of Canada, Ottawa.
- U.S. House of Representatives Subcommittee on Aviation (2002). Hearing on aviation security with a focus on passenger profiling, February 27, 2002; see <http://www.house.gov/transportation/aviation/02-27-02/02-27-02memo.html>.
- Utgoff, P. E. (1989). Incremental induction of decision trees. *Machine Learning*, 4(2), 161–186.
- Utgoff, P. E., Berkman, N. C., & Clouse, J. A. (1997). Decision tree induction based on efficient tree restructuring. *Machine Learning*, 29(1), 5–44.
- Vafaie, H., & DeJong, K. (1992). Genetic algorithms as a tool for feature selection in machine learning. In *Proceedings of the International Conference on Tools with Artificial Intelligence* (pp. 200–203). Arlington, VA: IEEE Computer Society Press.
- van Rijsbergen, C. A. (1979). *Information retrieval*. London: Butterworths.
- Vapnik, V. (1999). *The nature of statistical learning theory* (2nd ed.). New York: Springer-Verlag.
- Vitter, J. S. (1985). Random sampling with a reservoir. *ACM Transactions on Mathematical Software*, 1(11), 37–57.
- Wang, J., & Zucker, J.-D. (2000). Solving the multiple-instance problem: A lazy learning approach. In *Proceedings of the International Conference on Machine Learning* (pp. 1119–1125). Stanford, CA. San Francisco: Morgan Kaufmann.
- Wang, J., Han, J., & Pei, J. (2003). CLOSET+: Searching for the best strategies for mining frequent closed itemsets. In *Proceedings of the International Conference on Knowledge Discovery and Data Mining* (pp. 236–245). Washington, DC. New York: ACM Press.
- Wang, Y., & Witten, I. H. (1997). Induction of model trees for predicting continuous classes. In M. van Someren, & G. Widmer (Eds.), *Proceedings of the of the Poster Papers of the European Conference on Machine Learning* (pp. 128–137). University of Economics, Faculty of Informatics and Statistics, Prague. Berlin: Springer.

- . (2002). Modeling for optimal probability prediction. In C. Sammut, & A. Hoffmann (Eds.), *Proceedings of the Nineteenth International Conference on Machine Learning* (pp. 650–657). Sydney. San Francisco: Morgan Kaufmann.
- Webb, G. I. (1999). Decision tree grafting from the all-tests-but-one partition. In *Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence* (pp. 702–707). San Francisco: Morgan Kaufmann.
- . (2000). MultiBoosting: A technique for combining boosting and wagging. *Machine Learning*, 40(2), 159–196.
- Webb, G. I., Boughton, J., & Wang, Z. (2005). Not so naïve Bayes: Aggregating one-dependence estimators. *Machine Learning*, 58(1), 5–24.
- Weidmann, N., Frank, E., & Pfahringer, B. (2003). A two-level learning method for generalized multi-instance problems. In *Proceedings of the European Conference on Machine Learning* (pp. 468–479). Cavtat, Croatia. Berlin: Springer-Verlag.
- Weiser, M., & Brown, J. S. (1997). The coming age of calm technology. In P. J. Denning, & R. M. Metcalfe (Eds.), *Beyond calculation: The next fifty years* (pp. 75–86). New York: Copernicus.
- Weiss, S. M., & Indurkha, N. (1998). *Predictive data mining: A practical guide*. San Francisco: Morgan Kaufmann.
- Wettschereck, D., & Dietterich, T. G. (1995). An experimental comparison of the nearest-neighbor and nearest-hyperrectangle algorithms. *Machine Learning*, 19(1), 5–28.
- Wild, C. J., & Seber, G. A. F. (1995). *Introduction to probability and statistics*. Department of Statistics, University of Auckland, New Zealand.
- Winston, P. H. (1992). *Artificial intelligence*. Reading, MA: Addison-Wesley.
- Witten, I. H. (2004). Text mining. In M. P. Singh (Ed.), *Practical handbook of Internet computing* (pp. 14-1–14-22). Boca Raton, FL: CRC Press.
- Witten, I. H., Bray, Z., Mahoui, M., & Teahan, W. (1999a). Text mining: A new frontier for lossless compression. In J. A. Storer, & M. Cohn (Eds.), *Proceedings of the Data Compression Conference* (pp. 198–207). Snowbird, UT. Los Alamitos, CA: IEEE Press.
- Witten, I. H., Moffat, A., & Bell, T. C. (1999b). *Managing gigabytes: Compressing and indexing documents and images* (2nd ed.). San Francisco: Morgan Kaufmann.
- Wolpert, D. H. (1992). Stacked generalization. *Neural Networks*, 5, 241–259.
- Wu, X. V., Kumar, J. R., Quinlan, J., Ghosh, Q., Yang, H., Motoda, G. J., et al. (2008). Top 10 algorithms in data mining. *Knowledge and Information Systems*, 14(1), 1–37.
- Wu, X., & Kumar, V. (Eds.), (2009). *The top ten algorithms in data mining*. New York: Chapman and Hall.
- Xu, X., & Frank, E. (2004). Logistic regression and boosting for labeled bags of instances. In *Proceedings of the 8th Pacific-Asia Conference on Knowledge Discovery and Data Mining* (pp. 272–281). Sydney. Berlin: Springer-Verlag.
- Yan, X., & Han, J. (2002). gSpan: Graph-based substructure pattern mining. In *Proceedings of the IEEE International Conference on Data Mining* (pp. 721–724). Maebashi City, Japan. Washington, DC: IEEE Computer Society.
- Yan, X., & Han, J. (2003). CloseGraph: Mining closed frequent graph patterns. In *Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 286–295). Washington, DC. New York: ACM Press.
- Yan, X., Han, J., & Afshar, R. (2003). CloSpan: Mining closed sequential patterns in large datasets. In *Proceedings of the SIAM International Conference on Data Mining* (pp. 166–177). San Francisco. Philadelphia: Society for Industrial and Applied Mathematics.

- Yang, Y., & Webb, G. I. (2001). Proportional k -interval discretization for Naïve Bayes classifiers. In L. de Raedt, & P. Flach (Eds.), *Proceedings of the Twelfth European Conference on Machine Learning* (pp. 564–575). Freiburg, Germany. Berlin: Springer-Verlag.
- Yang, Y., Guan, X., & You, J. (2002). CLOPE: A fast and effective clustering algorithm for transactional data. In *Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 682–687). Edmonton, AB. New York: ACM Press.
- Yurcik, W., Barlow, J., Zhou, Y., Raje, H., Li, Y., Yin, X., et al. (2003). Scalable data management alternatives to support data mining heterogeneous logs for computer network security. In *Proceedings of the Workshop on Data Mining for Counter Terrorism and Security*. San Francisco. Philadelphia: Society for International and Applied Mathematics.
- Zadrozny, B., & Elkan, C. (2002). Transforming classifier scores into accurate multiclass probability estimates. In *Proceedings of the Eighth ACM International Conference on Knowledge Discovery and Data Mining* (pp. 694–699). Edmonton, AB. New York: ACM Press.
- Zaki, M. J., Parthasarathy, S., Ogihara, M., & Li, W. (1997). New algorithms for fast discovery of association rules. In *Proceedings Knowledge Discovery in Databases* (pp. 283–286). Newport Beach, CA. Menlo Park, CA: AAAI Press.
- Zhang, H., Jiang, L., & Su, J. (2005). Hidden Naïve Bayes. In *Proceedings of the 20th National Conference on Artificial Intelligence* (pp. 919–924). Pittsburgh. Menlo Park, CA: AAAI Press.
- Zhang, T., Ramakrishnan, R., & Livny, M. (1996). BIRCH: An efficient data clustering method for very large databases. In *Proceedings of the ACM SIGMOD International Conference on Management of Data* (pp. 103–114). Montreal. New York: ACM Press.
- Zhang, T. (2004). Solving large scale linear prediction problems using stochastic gradient descent algorithms. In *Proceedings of the 21st International Conference on Machine Learning* (pp. 919–926). Banff, AB. Madison, WI: Omnipress.
- Zheng, F., & Webb, G. (2006). Efficient lazy elimination for averaged one-dependence estimators. In *Proceedings of the 23rd International Conference on Machine Learning* (pp. 1113–1120). New York: ACM Press.
- Zheng, Z., & Webb, G. (2000). Lazy learning of Bayesian rules. *Machine Learning*, 41(1), 53–84.
- Zhou, Z.-H., & Zhang, M.-L. (2007). Solving multi-instance problems with classifier ensemble based on constructive clustering. *Knowledge and Information Systems*, 11(2), 155–170.