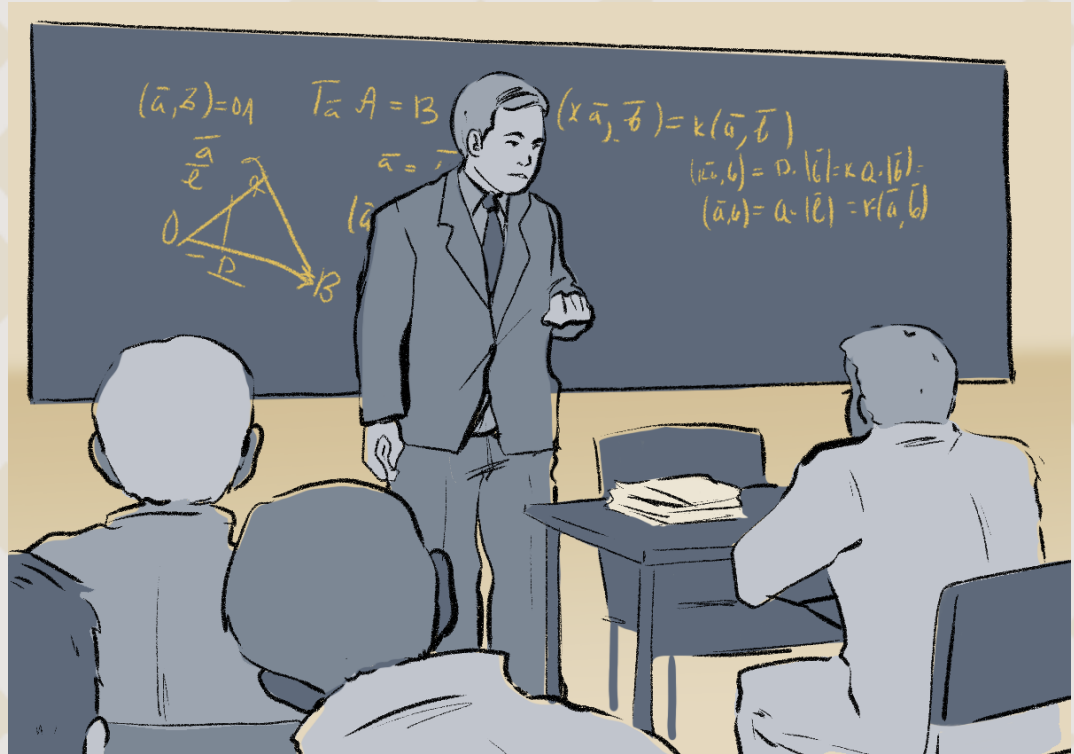




La Ciencia de Datos en 50 artículos for the working analyst





Con gran entusiasmo les presentamos "La ciencia de datos en 50 artículos for the working analyst" dedicada a los practitioners de la Ciencia de Datos. Esta es una parte de nuestra selección de los 150 artículos más relevantes en Ciencia de Datos, Machine Learning y Deep Learning.

El objetivo de este compendio es ayudar a aquellos que necesiten conocimiento y referencias previas para leer artículos state-of-the-art.

La Ciencia de Datos es una combinación armoniosa entre Computación, Estadística y otras áreas de las matemáticas, por lo que en esta lista encontrarán temas bastante diversos, aunque todos con el común denominador de ser importantes para la resolución de problemas data-driven.

¡Esperamos que encuentren esta lista útil en su estudio de esta área!
La elección de los artículos de esta edición no representa ni el estado del arte, ni un recuento histórico de área.

1.

Deerwester, S., Dumais, S., Landauer, T., Furnas, G. and Beck, L., 1988, January. Improving information-retrieval with latent semantic indexing. In Proceedings of the ASIS annual meeting (Vol. 25, pp. 36-40). 143
OLD MARLTON PIKE,
MEDFORD, NJ 08055-8750:
INFORMATION TODAY INC.

2.

Schapire, R.E., 1990. The strength of weak learnability. Machine learning, 5, pp.197-227.

3.

Friedman, J.H., 2001. Greedy function approximation: a gradient boosting machine. Annals of statistics, pp.1189-1232.

4.

Novembre, J., Johnson, T., Bryc, K., Kutalik, Z., Boyko, A.R., Auton, A., Indap, A., King, K.S., Bergmann, S., Nelson, M.R. and Stephens, M., 2008. Genes mirror geography within Europe. Nature, 456(7218), pp.98-101.

5.

Blei, D.M., Ng, A.Y. and Jordan, M.I., 2003. Latent dirichlet allocation. Journal of machine Learning research, 3(Jan), pp.993-1022.

6.

Diaconis, P., 2009. The markov chain monte carlo revolution. Bulletin of the American Mathematical Society, 46(2), pp.179-205.

7.

Shannon, C.E., 1948. A mathematical theory of communication. The Bell system technical journal, 27(3), pp.379-423.

8.

Lundberg, S.M. and Lee, S.I., 2017. A unified approach to interpreting model predictions. Advances in neural information processing systems, 30.

9.

Lehmann, E.L., 1993. The Fisher, Neyman-Pearson theories of testing hypotheses: one theory or two?. Journal of the American Statistical Association, 88(424), pp.1242-1249.

10.

Sur, P. and Candès, E.J., 2019. A modern maximum-likelihood theory for high-dimensional logistic regression. *Proceedings of the National Academy of Sciences*, 116(29), pp.14516-14525.

11.

Donoho, D.L., 2006. For most large underdetermined systems of linear equations the minimal ℓ_1 -norm solution is also the sparsest solution. *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences*, 59(6), pp.797-829.

12.

Aitken, A.C., 1936. IV.—On least squares and linear combination of observations. *Proceedings of the Royal Society of Edinburgh*, 55, pp.42-48.

13.

Shulman, B., 1998. Math-alive! Using original sources to teach mathematics in social context. *Problems, Resources, and Issues in Mathematics Undergraduate Studies*, 8(1), pp.1-14.

14.

Breiman, L., Friedman, J.H., Olshen, R.A. and Stone, C.J., 1984. Classification and regression trees. *Statistics/probability series*.

15.

Page, L., Brin, S., Motwani, R. and Winograd, T., 1998. The pagerank citation ranking: Bring order to the web. Technical report, Stanford University.

16.

Huber, P.J., 1992. Robust estimation of a location parameter. In *Breakthroughs in statistics: Methodology and distribution* (pp. 492-518). New York, NY: Springer New York.

17.

Johnson, W.B., Lindenstrauss, J. and Schechtman, G., 1986. Extensions of Lipschitz maps into Banach spaces. *Israel Journal of Mathematics*, 54(2), pp.129-138.

18.

Hofmann, T., Schölkopf, B. and Smola, A.J., 2008. *Kernel methods in machine learning*.

19.

Boser, B.E., Guyon, I.M. and Vapnik, V.N., 1992, July. A training algorithm for optimal margin classifiers. In Proceedings of the fifth annual workshop on Computational learning theory (pp. 144-152).

20.

Pearl, J., 1995. Causal diagrams for empirical research. *Biometrika*, 82(4), pp.669-688.

21.

Pearl, J., 2022. Fusion, propagation, and structuring in belief networks. In *Probabilistic and Causal Inference: The Works of Judea Pearl* (pp. 139-188).

22.

Forgy, E.W., 1965. Cluster analysis of multivariate data: efficiency versus interpretability of classifications. *biometrics*, 21, pp.768-769.

23.

Mahalanobis, P.C., 2018. On the generalized distance in statistics. *Sankhyā: The Indian Journal of Statistics, Series A* (2008-), 80, pp.S1-S7.

24.

Ng, A., Jordan, M. and Weiss, Y., 2001. On spectral clustering: Analysis and an algorithm. *Advances in neural information processing systems*, 14.

25.

Markov, A.A., 2006. An example of statistical investigation of the text Eugene Onegin concerning the connection of samples in chains. *Science in Context*, 19(4), pp.591-600.

26.

Hanley, J.A. and McNeil, B.J., 1982. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, 143(1), pp.29-36.

27.

Donoho, D., 2017. 50 years of data science. *Journal of Computational and Graphical Statistics*, 26(4), pp.745-766.

28.

Tukey, J.W., 1962. The future of data analysis. The annals of mathematical statistics, 33(1), pp.1-67.

29.

Bellman, R., 1952. On the theory of dynamic programming. Proceedings of the national Academy of Sciences, 38(8), pp.716-719.

30.

Watkins, C.J.C.H., 1989. Learning from delayed rewards.

31.

Kalbfleisch, J.D. and Schaubel, D.E., 2023. Fifty Years of the Cox Model. Annual Review of Statistics and its Application, 10, pp.1-23.

32.

Töscher, A., Jahrer, M. and Bell, R.M., 2009. The bigchaos solution to the netflix grand prize. Netflix prize documentation, pp.1-52.

33.

Tikhonov, A.N., 1963. Solution of incorrectly formulated problems and the regularization method. Sov Dok, 4, pp.1035-1038.

34.

Santosa, F. and Symes, W.W., 1986. Linear inversion of band-limited reflection seismograms. SIAM journal on scientific and statistical computing, 7(4), pp.1307-1330.

35.

Mockus, J., 2005, September. The Bayesian approach to global optimization. In System Modeling and Optimization: Proceedings of the 10th IFIP Conference New York City, USA, August 31–September 4, 1981 (pp. 473-481). Berlin, Heidelberg: Springer Berlin Heidelberg.

36.

Lee Rodgers, J. and Nicewander, W.A., 1988. Thirteen ways to look at the correlation coefficient. The American Statistician, 42(1), pp.59-66.

37.

Suddath, J.H., Kidd, R.H. and Reinhold, A.G., 1967. A Linearized Error Analysis of Onboard Primary Navigation Systems for the Apollo Lunar Module (Vol. 4027). National Aeronautics and Space Administration.

38.

Philip, G.M. and Watson, D.F., 1986. Matheronian geostatistics—Quo vadis?. *Mathematical geology*, 18, pp.93-117.

39.

Dence, T., 1997. Cubics, chaos and Newton's method. *The Mathematical Gazette*, 81(492), pp.403-408.

40.

Courant, R., 1943. Variational methods for the solution of problems of equilibrium and vibrations.

41.

Gittins, J.C., 1979. Bandit processes and dynamic allocation indices. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 41(2), pp.148-164.

42.

Salton, G. and Buckley, C., 1988. Term-weighting approaches in automatic text retrieval. *Information processing & management*, 24(5), pp.513-523.

43.

Jelinek, F., Bahl, L. and Mercer, R., 1975. Design of a linguistic statistical decoder for the recognition of continuous speech. *IEEE Transactions on Information Theory*, 21(3), pp.250-256.

44.

Jordan, M.I., Ghahramani, Z., Jaakkola, T.S. and Saul, L.K., 1999. An introduction to variational methods for graphical models. *Machine learning*, 37, pp.183-233.

45.

Breiman, L., 2001. Random forests. *Machine learning*, 45, pp.5-32.

46.

Kullback, S. and Leibler, R.A., 1951. On information and sufficiency. The annals of mathematical statistics, 22(1), pp.79-86.

47.

Bayes, T., 1763. LII. An essay towards solving a problem in the doctrine of chances. By the late Rev. Mr. Bayes, FRS communicated by Mr. Price, in a letter to John Canton, AMFR S. Philosophical transactions of the Royal Society of London, (53), pp.370-418.

48.

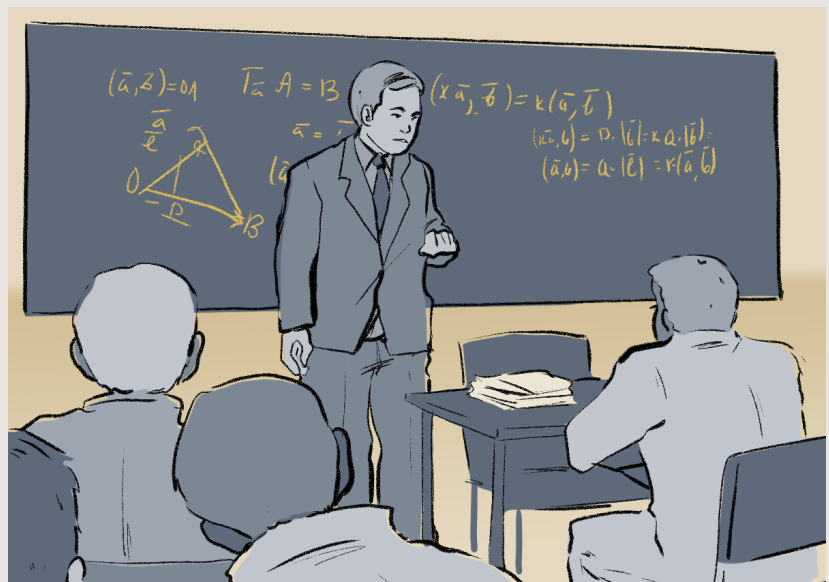
Anderson, O.D., 1977. The Box-Jenkins approach to time series analysis. RAIRO-Operations Research, 11(1), pp.3-29.

49.

Geman, S., Bienenstock, E. and Doursat, R., 1992. Neural networks and the bias/variance dilemma. Neural computation, 4(1), pp.1-58.

50.

Fisher, R.A., 1936. The use of multiple measurements in taxonomic problems. Annals of eugenics, 7(2), pp.179-188.





colegio-bourbaki.com