

## DAFTAR PUSTAKA

- Balahadia, F., Vinluan, A., Gonzales, D., & Ballera, M. (2021). Application of Spatiotemporal Analysis and Knowledge Discovery for Databases in the Bureau of Fire Protection as Incident Report System: Tool for Improving Fire Services. *International Journal of Computing Sciences Research*, 519–533. <https://doi.org/10.25147/ijcsr.2017.001.1.56>
- Burgess, J., & Baym, N. K. (2020). *Twitter: A Biography*. NYU Press. <https://books.google.co.id/books?id=VLHWDwAAQBAJ>
- Chowdhary, K. R. (2020). Fundamentals of artificial intelligence. Dalam *Fundamentals of Artificial Intelligence*. Springer India. <https://doi.org/10.1007/978-81-322-3972-7>
- Duong, H. T., & Nguyen-Thi, T. A. (2021). A review: preprocessing techniques and data augmentation for sentiment analysis. *Computational Social Networks*, 8(1). <https://doi.org/10.1186/s40649-020-00080-x>
- Esbensen, K. H., & Geladi, P. (2020). 2.02 - Principal Component Analysis: Concept, Geometrical Interpretation, Mathematical Background, Algorithms, History, Practice. Dalam *Comprehensive Chemometrics: Chemical and Biochemical Data Analysis, Second Edition: Four Volume Set* (Vol. 2). <https://doi.org/10.1016/B978-0-444-64165-6.05002-3>
- Fa'rifah, R. Y., & Pramesti, D. (2022). Cluster Analysis of Inclusive Economic Development Using K-Means Algorithm. *Jurnal Varian*, 5(2). <https://doi.org/10.30812/varian.v5i2.1894>
- Filonchyk, M., & Peterson, M. P. (2023). An integrated analysis of air pollution from US coal-fired power plants. *Geoscience Frontiers*, 14(2), 101498. <https://doi.org/10.1016/j.gsf.2022.101498>
- Firas, O. (2023). A combination of SEMMA & CRISP-DM models for effectively handling big data using formal concept analysis based knowledge discovery: A data mining approach. *World Journal of Advanced Engineering*

*Technology and Sciences*, 8(1).  
<https://doi.org/10.30574/wjaets.2023.8.1.0147>

Gbadoubissa, J. E. Z., Ari, A. A. A., & Gueroui, A. M. (2020). Efficient k-means based clustering scheme for mobile networks cell sites management. *Journal of King Saud University - Computer and Information Sciences*, 32(9).  
<https://doi.org/10.1016/j.jksuci.2018.10.015>

Gewers, F. L., Ferreira, G. R., De Arruda, H. F., Silva, F. N., Comin, C. H., Amancio, D. R., & Costa, L. D. F. (2021). Principal component analysis: A natural approach to data exploration. *ACM Computing Surveys*, 54(4).  
<https://doi.org/10.1145/3447755>

Grohe, M. (2020). Word2vec, node2vec, graph2vec, X2vec: Towards a Theory of Vector Embeddings of Structured Data. *Proceedings of the ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems*.  
<https://doi.org/10.1145/3375395.3387641>

Gurajala, S., Dhaniyala, S., & Matthews, J. N. (2019). Understanding Public Response to Air Quality Using Tweet Analysis. *Social Media and Society*, 5(3). <https://doi.org/10.1177/2056305119867656>

Hartigan, A., & Wong, M. A. (1979). A K-Means Clustering Algorithm. *Journal of the Royal Statistical Society*, 28(1).

Hasan, A. A., & Fang, H. (2021). Data Mining in Education: Discussing Knowledge Discovery in Database (KDD) with Cluster Associative Study. *ACM International Conference Proceeding Series*.  
<https://doi.org/10.1145/3469213.3471319>

Hickman, L., Thapa, S., Tay, L., Cao, M., & Srinivasan, P. (2022). Text Preprocessing for Text Mining in Organizational Research: Review and Recommendations. *Organizational Research Methods*, 25(1), 114–146.  
<https://doi.org/10.1177/1094428120971683>

- Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, 24(6). <https://doi.org/10.1037/h0071325>
- Jakob, M., & Steckel, J. C. (2022). The Political Economy of Coal: Obstacles to Clean Energy Transitions. Dalam *The Political Economy of Coal: Obstacles to Clean Energy Transitions*. <https://doi.org/10.4324/9781003044543>
- Janiesch, C., Zschech, P., & Heinrich, K. (2021). Machine learning and deep learning. *Electronic Markets*, 31(3). <https://doi.org/10.1007/s12525-021-00475-2>
- Karami, A., Lundy, M., Webb, F., & Dwivedi, Y. K. (2020). Twitter and Research: A Systematic Literature Review through Text Mining. *IEEE Access*, 8. <https://doi.org/10.1109/ACCESS.2020.2983656>
- Kelly, J., Myllyvirta, L., Tattari, V., & Hasan, K. (2023, September 12). *Air quality impacts of the Banten-Suralaya complex*. Centre for Research on Energy and Clean Air. <https://energyandcleanair.org/publication/air-quality-impacts-of-the-banten-suralaya-complex/>
- Kherwa, P., & Bansal, P. (2020). Topic Modeling: A Comprehensive Review. *EAI Endorsed Transactions on Scalable Information Systems*, 7(24). <https://doi.org/10.4108/eai.13-7-2018.159623>
- Lubis, A. R., & Nasution, M. K. M. (2023). Twitter Data Analysis and Text Normalization in Collecting Standard Word. *Journal of Applied Engineering and Technological Science*, 4(2). <https://doi.org/10.37385/jaets.v4i2.1991>
- Maharana, K., Mondal, S., & Nemade, B. (2022). A review: Data pre-processing and data augmentation techniques. *Global Transitions Proceedings*, 3(1). <https://doi.org/10.1016/j.gltp.2022.04.020>
- Mahat, H., Hussein, S. W., Saleh, Y., Hashim, M., Nayan, N., Said, Z. M., & Kurniawan, E. (2023). Social Media as a Medium for Disseminating Community Awareness of Environmental Issues in Malaysia. *TEM Journal*, 12(3). <https://doi.org/10.18421/TEM123-47>

- Mahmoud, H. A. H., Hafez, A. M., & Alabdulkreem, E. (2023). Language-Independent Text Tokenization Using Unsupervised Deep Learning. *Intelligent Automation and Soft Computing*, 35(1). <https://doi.org/10.32604/iasc.2023.026235>
- Manisalidis, I., Stavropoulou, E., Stavropoulos, A., & Bezirtzoglou, E. (2020). Environmental and Health Impacts of Air Pollution: A Review. Dalam *Frontiers in Public Health* (Vol. 8). Frontiers Media S.A. <https://doi.org/10.3389/fpubh.2020.00014>
- Murphy, K. P. (2012). Machine Learning: A Probabilistic Perspective. Dalam *The MIT Press*. The MIT Press.
- Myllyvirta, L., Kelly, J., & Uusivuori, Erika. (2023, Juli 18). *Health benefits of Just Energy Transition and coal phase-out in Indonesia*. Centre for Research on Energy and Clean Air. <https://energyandcleanair.org/publication/health-benefits-of-just-energy-transition-and-coal-phase-out-in-indonesia/>
- Naghizadeh, A., & Metaxas, D. N. (2020). Condensed silhouette: An optimized filtering process for cluster selection in K-means. *Procedia Computer Science*, 176. <https://doi.org/10.1016/j.procs.2020.08.022>
- Ngamlana, N. B., Malherbe, W., Gericke, G., & Coetzer, R. L. J. (2024). The effect of coal-fired power plants on ambient air quality in Mpumalanga province, South Africa, 2014–2018. *International Journal of Environmental Health Research*. <https://doi.org/10.1080/09603123.2024.2350600>
- Salsabila, N. A., Winatmoko, Y. A., Septiandri, A. A., & Jamal, A. (2018). Colloquial Indonesian Lexicon. *Proceedings of the 2018 International Conference on Asian Language Processing, IALP 2018*. <https://doi.org/10.1109/IALP.2018.8629151>
- Sarker, I. H. (2021). Machine Learning: Algorithms, Real-World Applications and Research Directions. Dalam *SN Computer Science* (Vol. 2, Nomor 3). <https://doi.org/10.1007/s42979-021-00592-x>

- Scheibel, W., Trapp, M., Limberger, D., & Döllner, J. (2020). A taxonomy of treemap visualization techniques. *VISIGRAPP 2020 - Proceedings of the 15th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*, 3. <https://doi.org/10.5220/0009153902730280>
- Shahapure, K. R., & Nicholas, C. (2020). Cluster quality analysis using silhouette score. *Proceedings - 2020 IEEE 7th International Conference on Data Science and Advanced Analytics, DSAA 2020*. <https://doi.org/10.1109/DSAA49011.2020.00096>
- Shi, C., Wei, B., Wei, S., Wang, W., Liu, H., & Liu, J. (2021). A quantitative discriminant method of elbow point for the optimal number of clusters in clustering algorithm. *Eurasip Journal on Wireless Communications and Networking*, 2021(1). <https://doi.org/10.1186/s13638-021-01910-w>
- Shu, X., & Ye, Y. (2023). Knowledge Discovery: Methods from data mining and machine learning. *Social Science Research*, 110. <https://doi.org/10.1016/j.ssresearch.2022.102817>
- Shutaywi, M., & Kachouie, N. N. (2021). Silhouette analysis for performance evaluation in machine learning with applications to clustering. *Entropy*, 23(6). <https://doi.org/10.3390/e23060759>
- Sinaga, K. P., & Yang, M. S. (2020). Unsupervised K-means clustering algorithm. *IEEE Access*, 8. <https://doi.org/10.1109/ACCESS.2020.2988796>
- Singh, A. K., Mittal, S., Malhotra, P., & Srivastava, Y. V. (2020). Clustering Evaluation by Davies-Bouldin Index(DBI) in Cereal data using K-Means. *Proceedings of the 4th International Conference on Computing Methodologies and Communication, ICCMC 2020*. <https://doi.org/10.1109/ICCMC48092.2020.ICCMC-00057>
- Sivakumar, S., Videla, L. S., Rajesh Kumar, T., Nagaraj, J., Itnal, S., & Haritha, D. (2020). Review on Word2Vec Word Embedding Neural Net. *Proceedings - International Conference on Smart Electronics and Communication, ICOSEC 2020*. <https://doi.org/10.1109/ICOSEC49089.2020.9215319>

- Solanki, Y. S., Chakrabarti, P., Jasinski, M., Leonowicz, Z., Bolshev, V., Vinogradov, A., Jasinska, E., Gono, R., & Nami, M. (2021). A hybrid supervised machine learning classifier system for breast cancer prognosis using feature selection and data imbalance handling approaches. *Electronics (Switzerland)*, 10(6). <https://doi.org/10.3390/electronics10060699>
- Statista. (2025, Februari 5). *Most popular social networks worldwide as of February 2025, by number of monthly active users*. Statista. <https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/>
- Taha, A. Y., Tiun, S., Rahman, A. H. A., & Sabah, A. (2021). Multilabel Over-sampling and Under-sampling with Class Alignment for Imbalanced Multilabel Text Classification. *Journal of Information and Communication Technology*, 20(3). <https://doi.org/10.32890/JICT2021.20.3.6>
- Tao, Y., Liu, W., Huang, Z., & Shi, C. (2022). Thematic analysis of reviews on the air quality of tourist destinations from a sentiment analysis perspective. *Tourism Management Perspectives*, 42. <https://doi.org/10.1016/j.tmp.2022.100969>
- Tao, Y., Zhang, F., Shi, C., & Chen, Y. (2019). Social media data-based sentiment analysis of tourists' air quality perceptions. *Sustainability (Switzerland)*, 11(18). <https://doi.org/10.3390/su11185070>
- Thabtah, F., Hammoud, S., Kamalov, F., & Gonsalves, A. (2020). Data imbalance in classification: Experimental evaluation. *Information Sciences*, 513, 429–441. <https://doi.org/10.1016/j.ins.2019.11.004>
- Vayansky, I., & Kumar, S. A. P. (2020). A review of topic modeling methods. *Information Systems*, 94. <https://doi.org/10.1016/j.is.2020.101582>
- Wang, B., Wang, N., & Chen, Z. (2021). Research on air quality forecast based on web text sentiment analysis. *Ecological Informatics*, 64. <https://doi.org/10.1016/j.ecoinf.2021.101354>

- Weng, Z., Song, Y., Cheng, C., Tong, D., Xu, M., Wang, M., & Xie, Y. (2023). Possible underestimation of the coal-fired power plants to air pollution in China. *Resources, Conservation and Recycling*, 198, 107208. <https://doi.org/10.1016/J.RESCONREC.2023.107208>
- Wijaya, Y. A., Kurniady, D. A., Setyanto, E., Tarihoran, W. S., Rusmana, D., & Rahim, R. (2021). Davies Bouldin Index Algorithm for Optimizing Clustering Case Studies Mapping School Facilities. *TEM Journal*, 10(3). <https://doi.org/10.18421/TEM103-13>
- Yilmaz, S., & Toklu, S. (2020). A deep learning analysis on question classification task using Word2vec representations. *Neural Computing and Applications*, 32(7). <https://doi.org/10.1007/s00521-020-04725-w>
- Zheng, S., Wang, J., Sun, C., Zhang, X., & Kahn, M. E. (2019). Air pollution lowers Chinese urbanites' expressed happiness on social media. Dalam *Nature Human Behaviour* (Vol. 3, Nomor 3, hlm. 237–243). Nature Publishing Group. <https://doi.org/10.1038/s41562-018-0521-2>
- Zhou, Q., & Sun, B. (2024). Adaptive K-means clustering based under-sampling methods to solve the class imbalance problem  
*Data and Information Management*, 8(3).  
<https://doi.org/10.1016/j.dim.2023.100064>