

DAFTAR PUSTAKA

- Abdullah, D., Susilo, S., Ahmar, A. S., Rusli, R., dan Hidayat, R. (2022). The application of k-means clustering for province clustering in indonesia of the risk of the covid-19 pandemic based on covid-19 data. *Quality & Quantity*, 56(3), 1283–1291.
- Ahmed, M., Seraj, R., dan Islam, S. M. S. (2020). The k-means algorithm: A comprehensive survey and performance evaluation. *Electronics*, 9(8), 1295.
- Azmoodeh, M. (1990). *Abstract data types and algorithms*. Springer.
- Ball, G. H. (1965). Isodata, a novel method of data analysis and pattern classification. *stanford research institute*, AD-699616.
- Barnett, V., Lewis, T., et al. (1994). *Outliers in statistical data* (Vol. 3) (No. 1). Wiley New York.
- Berkhin, P. (2006). A survey of clustering data mining techniques. In *Grouping multidimensional data: Recent advances in clustering* (pp. 25–71). Springer.
- Berry, M. J., dan Linoff, G. S. (2004). *Data mining techniques: for marketing, sales, and customer relationship management*. John Wiley & Sons.
- Bezdek, J. C., dan Bezdek, J. C. (1981). Models for pattern recognition. *Pattern recognition with fuzzy objective function algorithms*, 1–13.
- Brodinová, Š., Filzmoser, P., Ortner, T., Breiteneder, C., dan Rohm, M. (2019). Robust and sparse k-means clustering for high-dimensional data. *Advances in Data Analysis and Classification*, 13, 905–932.
- Burton, P. W., dan Hall, T. R. (2014). Segmentation of the sumatran fault. *Geophysical Research Letters*, 41(12), 4149–4158.
- Caliński, T., dan Harabasz, J. (1974). A dendrite method for cluster analysis. *Communications in Statistics-theory and Methods*, 3(1), 1–27.
- Cao, F., Liang, J., Li, D., Bai, L., dan Dang, C. (2012). A dissimilarity measure for the k-modes clustering algorithm. *Knowledge-Based Systems*, 26, 120–127.



- Capó, M., Pérez, A., dan Lozano, J. A. (2020). An efficient k-means clustering algorithm for tall data. *Data mining and knowledge discovery*, 34, 776–811.
- Charrad, M., Ghazzali, N., Boiteau, V., dan Niknafs, A. (2014). Nbclust: an R package for determining the relevant number of clusters in a data set. *Journal of statistical software*, 61, 1–36.
- Chen, C., Wang, Y., Hu, W., dan Zheng, Z. (2020). Robust multi-view k-means clustering with outlier removal. *Knowledge-Based Systems*, 210, 106518.
- Cormen, T. H., Leiserson, C. E., Rivest, R. L., dan Stein, C. (2009). *Introduction to algorithms*. MIT press.
- Cuesta-Albertos, J. A., Gordaliza, A., dan Matrán, C. (1997). Trimmed k -means: an attempt to robustify quantizers. *The Annals of Statistics*, 25(2), 553–576.
- Davies, D. L., dan Bouldin, D. W. (1979, apr). A Cluster Separation Measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-I(2)*, 224–227. Retrieved from <https://doi.org/10.1109/TPAMI.1979.4766909><http://ieeexplore.ieee.org/document/4766909> doi: 10.1109/TPAMI.1979.4766909
- De Morsier, F., Tuia, D., Borgeaud, M., Gass, V., dan Thiran, J.-P. (2015). Cluster validity measure and merging system for hierarchical clustering considering outliers. *Pattern Recognition*, 48(4), 1478–1489.
- Dorabiala, O., Kutz, J. N., dan Aravkin, A. Y. (2022). Robust trimmed k-means. *Pattern Recognition Letters*, 161, 9–16.
- Dunn, J. C. (1973). A fuzzy relative of the isodata process and its use in detecting compact well-separated clusters. *Journal of Cybernetics*.
- Estivill-Castro, V. (2002). Why so many clustering algorithms: a position paper. *ACM SIGKDD explorations newsletter*, 4(1), 65–75.
- Ezugwu, A. E., Ikotun, A. M., Oyelade, O. O., Abualigah, L., Agushaka, J. O., Eke, C. I., dan Akinyelu, A. A. (2022). A comprehensive survey of clustering algorithms: State-of-the-art machine learning applications, taxonomy, challenges, and future research prospects. *Engineering Applications of Artificial Intelligence*, 110, 104743.



- Fahim, A. (2021). K and starting means for k-means algorithm. *Journal of Computational Science*, 55, 101445.
- Fayyad, U. M., Piatetsky-Shapiro, G., Smyth, P., dan Uthurusamy, R. (1996). From data mining to knowledge discovery: an overview. In *Advances in knowledge discovery and data mining*.
- Fraley, C., dan Raftery, A. E. (1998). How many clusters? which clustering method? answers via model-based cluster analysis. *The computer journal*, 41(8), 578–588.
- Fukuyama, Y. (1989). A new method of choosing the number of clusters for fuzzy c-means method. In *Proc. 5th fuzzy system symp.* (pp. 247–250).
- Gan, G., Ma, C., dan Wu, J. (2020). *Data clustering: theory, algorithms, and applications*. SIAM.
- Gan, G., dan Ng, M. K.-P. (2017). K-means clustering with outlier removal. *Pattern Recognition Letters*, 90, 8–14.
- Ghazal, T. M. (2021). Performances of k-means clustering algorithm with different distance metrics. *Intelligent Automation & Soft Computing*, 30(2), 735–742.
- Ghorbani, H. (2019). Mahalanobis distance and its application for detecting multivariate outliers. *Facta Universitatis, Series: Mathematics and Informatics*, 583–595.
- Gonzalez, J. D., Yohai, V. J., dan Zamar, R. H. (2019). Robust clustering using tau-scales. *arXiv preprint arXiv:1906.08198*.
- Govender, P., dan Sivakumar, V. (2020). Application of k-means and hierarchical clustering techniques for analysis of air pollution: A review (1980–2019). *Atmospheric pollution research*, 11(1), 40–56.
- Guo, G., Chen, L., Ye, Y., dan Jiang, Q. (2016). Cluster validation method for determining the number of clusters in categorical sequences. *IEEE transactions on neural networks and learning systems*, 28(12), 2936–2948.
- Gupta, S., Kumar, R., Lu, K., Moseley, B., dan Vassilvitskii, S. (2017). Local search methods for k-means with outliers. *Proceedings of the VLDB Endowment*, 10(7), 757–768.



- Halkidi, M., Batistakis, Y., dan Vazirgiannis, M. (2001). On clustering validation techniques. *Journal of intelligent information systems*, 17, 107–145.
- Han, J., Pei, J., dan Tong, H. (2022). *Data mining: concepts and techniques*. Morgan kaufmann.
- Hassan, B. A., Tayfor, N. B., Hassan, A. A., Ahmed, A. M., Rashid, T. A., dan Abdalla, N. N. (2024). From a-to-z review of clustering validation indices. *Neurocomputing*, 601, 128198.
- Huang, A., et al. (2008). Similarity measures for text document clustering. In *Proceedings of the sixth new zealand computer science research student conference (nzcsrsc2008), christchurch, new zealand* (Vol. 4, pp. 9–56).
- Huber, P. J. (1992). Robust estimation of a location parameter. In *Breakthroughs in statistics: Methodology and distribution* (pp. 492–518). Springer.
- Huber, P. J. (2004). *Robust statistics* (Vol. 523). John Wiley & Sons.
- Hubert, M., dan Debruyne, M. (2010). Minimum covariance determinant. *Wiley interdisciplinary reviews: Computational statistics*, 2(1), 36–43.
- Ikotun, A. M., Ezugwu, A. E., Abualigah, L., Abuhajja, B., dan Heming, J. (2023). K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data. *Information Sciences*, 622, 178–210.
- Jain, A. K. (2010). Data clustering: 50 years beyond k-means. *Pattern recognition letters*, 31(8), 651–666.
- Jain, A. K., dan Dubes, R. C. (1988). *Algorithms for clustering data*. Prentice-Hall, Inc.
- Jain, A. K., Murty, M. N., dan Flynn, P. J. (1999). Data clustering: a review. *ACM computing surveys (CSUR)*, 31(3), 264–323.
- Johnson, R. A., Wichern, D. W., et al. (2002). *Applied multivariate statistical analysis*. Prentice hall Upper Saddle River, NJ.
- Kodinariya, T. M., Makwana, P. R., et al. (2013). Review on determining number of cluster in k-means clustering. *International Journal*, 1(6), 90–95.
- Koller, M., dan Mächler, M. (2014). *Definitions of ψ -functions available in robustbase*.



- Kondo, Y., Salibian-Barrera, M., dan Zamar, R. (2016). Rskc: An r package for a robust and sparse k-means clustering algorithm. *Journal of Statistical Software*, 72, 1–26.
- Lance, G. N., dan Williams, W. T. (1966). Computer programs for hierarchical polythetic classification (“similarity analyses”). *The Computer Journal*, 9(1), 60–64.
- Lenssen, L., dan Schubert, E. (2024). Medoid silhouette clustering with automatic cluster number selection. *Information Systems*, 120, 102290.
- Liu, Y., Li, Z., Xiong, H., Gao, X., Wu, J., dan Wu, S. (2013). Understanding and enhancement of internal clustering validation measures. *IEEE transactions on cybernetics*, 43(3), 982–994.
- Lloyd, S. (1982). Least squares quantization in pcm. *IEEE transactions on information theory*, 28(2), 129–137.
- MacQueen, J., et al. (1967). Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth berkeley symposium on mathematical statistics and probability* (Vol. 1, pp. 281–297).
- Mattjik, A. A., Sumertajaya, I., Wibawa, G. N. A., dan Hadi, A. F. (2011). *Sidik peubah ganda dengan menggunakan sas*.
- Miraftabzadeh, S. M., Colombo, C. G., Longo, M., dan Foadelli, F. (2023). K-means and alternative clustering methods in modern power systems. *IEEE Access*.
- Mohd, M. R. S., Herman, S. H., dan Sharif, Z. (2017). Application of k-means clustering in hot spot detection for thermal infrared images. In *2017 IEEE symposium on computer applications & industrial electronics (iscae)* (pp. 107–110).
- Møller, S. F., von Frese, J., dan Bro, R. (2005). Robust methods for multivariate data analysis. *Journal of Chemometrics: A Journal of the Chemometrics Society*, 19(10), 549–563.
- Munir, R. (2012). Matematika diskrit revisi kelima. *Indformatika, Bandung*.
- Novianti, P., Setyorini, D., dan Rafflesia, U. (2017). K-means cluster analysis in

- earthquake epicenter clustering. *International Journal of Advances in Intelligent Informatics*, 3(2), 81–89.
- Oyelade, O. J., Oladipupo, O. O., dan Obagbuwa, I. C. (2010). Application of k means clustering algorithm for prediction of students academic performance. *arXiv preprint arXiv:1002.2425*.
- Pal, N. R., dan Bezdek, J. C. (1995). On cluster validity for the fuzzy c-means model. *IEEE Transactions on Fuzzy systems*, 3(3), 370–379.
- Pardoe, I. (2020). *Applied regression modeling*. John Wiley & Sons.
- Patil, C., dan Baidari, I. (2019). Estimating the optimal number of clusters k in a dataset using data depth. *Data Science and Engineering*, 4, 132–140.
- Qi, J., Yu, Y., Wang, L., dan Liu, J. (2016). K*-means: An effective and efficient k-means clustering algorithm. In *2016 IEEE International Conferences on Big Data and Cloud Computing (BDCloud), Social Computing and Networking (SocialCom), Sustainable Computing and Communications (SustainCom)(BDCloud-SocialCom-SustainCom)* (pp. 242–249).
- Rafflesia, U., Effendie, A. R., Sari, D. P., dan Rosadi, D. (2025). Improving cluster validity in k-means clustering using robust statistical estimators. *IAENG International Journal of Applied Mathematics*, -(1), -.
- Rafflesia, U., dan Rosadi, D. (2023). The application of K-means clustering and fuzzy C-means clustering analysis for modeling the spread of second wave coronavirus disease in Indonesia. In *Aip conference proceedings* (Vol. 2720, p. 020008). Retrieved from <http://aip.scitation.org/doi/abs/10.1063/5.0137755> doi: 10.1063/5.0137755
- Rafflesia, U., Rosadi, D., dan Sari, D. P. (2024). Multivariate outlier identification based on mahalanobis distance for cross-section and spatial data. In *Proceedings of the 8th international conference on quantitative sciences and its applications (icoqisia2024)* (Vol. September).
- Rafflesia, U., Rosadi, D., Sari, D. P., dan Effendie, A. R. (2025). Performance evaluation of k-means clustering algorithm using some robust distances: A case study on seismic data in sumatra. *Statistika: Statistics and Economy*

Journal, December(4), 1-22.

- Rafflesia, U., Rosadi, D., Sari, D. P., dan Novianti, P. (2025). Analysis of seismic data in sumatra using robust k-means clustering. *Journal of Applied Data Sciences*, 6(1), 391–404.
- Rafflesia, U., Rosadi, D., Utami, H., dan Sari, D. P. (2024). Robust k-means clustering for modeling the spread of coronavirus disease in indonesia. In *Aip conference proceedings* (Vol. 3024).
- Rezaee, M. R., Lelieveldt, B. P., dan Reiber, J. H. (1998). A new cluster validity index for the fuzzy c-mean. *Pattern recognition letters*, 19(3-4), 237–246.
- Ripley, B. D., dan Venables, W. N. (2002). *Modern applied statistics with s* (Vol. 4). Springer.
- Rousseeuw, P. J. (1985). Multivariate estimation with high breakdown point. *Mathematical statistics and applications*, 8(283-297), 37.
- Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20(C), 53–65. doi: 10.1016/0377-0427(87)90125-7
- Rousseeuw, P. J. (1991). Tutorial to robust statistics. *Journal of chemometrics*, 5(1), 1–20.
- Rousseeuw, P. J., dan Driessen, K. V. (1999). A fast algorithm for the minimum covariance determinant estimator. *Technometrics*, 41(3), 212–223.
- Schubert, E., dan Rousseeuw, P. J. (2021). Fast and eager k-medoids clustering: O(k) runtime improvement of the pam, clara, and clarans algorithms. *Information Systems*, 101, 101804.
- Sedgewick, R. (1990). *Algorithms in c++* (Vol. 1). Addison-Wesley Publishing Company.
- Shafeeq, A., dan Hareesha, K. (2012). Dynamic clustering of data with modified k-means algorithm. In *Proceedings of the 2012 conference on information and computer networks* (pp. 221–225).
- Shamrat, F. J. M., Tasnim, Z., Mahmud, I., Jahan, M. N., dan Nobel, N. I. (2020). Application of k-means clustering algorithm to determine the density of de-

- mand of different kinds of jobs. *International Journal of Scientific & Technology Research*, 9(02), 2550–2557.
- Sharma, S. (1995). *Applied multivariate techniques*. John Wiley & Sons, Inc.
- Shekhar, S., Lu, C.-T., dan Zhang, P. (2003). A unified approach to detecting spatial outliers. *GeoInformatica*, 7, 139–166.
- Shi, W., dan Zeng, W. (2014). Application of k-means clustering to environmental risk zoning of the chemical industrial area. *Frontiers of Environmental Science & Engineering*, 8, 117–127.
- Shodiq, M. N., Barakbah, A. R., dan Harsono, T. (2015). Spatial analysis of earthquake distribution with automatic clustering for prediction of earthquake seismicity in indonesia. In *The fourth indonesian-japanese conference on knowledge creation and intelligent computing (kic)* (pp. 24–26).
- Shukla, S., dan Naganna, S. (2014). A review on k-means data clustering approach. *International Journal of Information & Computation Technology*, 4(17), 1847–1860.
- Siang, J. J. (2009). *Matematika diskrit dan aplikasinya pada ilmu komputer*. Andi.
- Sinaga, K. P., dan Yang, M. S. (2020). Unsupervised K-means clustering algorithm. *IEEE Access*, 8, 80716–80727. doi: 10.1109/ACCESS.2020.2988796
- Smiti, A. (2020). A critical overview of outlier detection methods. *Computer Science Review*, 38, 100306. Retrieved from <https://doi.org/10.1016/j.cosrev.2020.100306> doi: 10.1016/j.cosrev.2020.100306
- Steinley, D. (2006). K-means clustering: a half-century synthesis. *British Journal of Mathematical and Statistical Psychology*, 59(1), 1–34.
- Supranto, J. (2004). *Analisis multivariat: Arti dan interpretasi. cetakan pertama*. Rineka Cipta Press. Jakarta. 359hlm.
- Szpankowski, W. (2011). *Average case analysis of algorithms on sequences*. John Wiley & Sons.
- Vendramin, L., Campello, R. J., dan Hruschka, E. R. (2010). Relative clustering validity criteria: A comparative overview. *Statistical analysis and data mining: the ASA data science journal*, 3(4), 209–235.



- Wattimanela, H., Pasaribu, U., Indratno, S., dan Puspito, A. (2015). Earthquakes clustering based on the magnitude and the depths in molla province. In *Aip conference proceedings* (Vol. 1692, p. 020021).
- Weatherill, G., dan Burton, P. W. (2009). Delineation of shallow seismic source zones using k-means cluster analysis, with application to the aegean region. *Geophysical Journal International*, 176(2), 565–588.
- Witten, D. M., dan Tibshirani, R. (2010). A framework for feature selection in clustering. *Journal of the American Statistical Association*, 105(490), 713–726.
- Wu, K.-L., Yang, M.-S., dan Hsieh, J.-N. (2009). Robust cluster validity indexes. *Pattern Recognition*, 42(11), 2541–2550.
- Wunsch, D., dan Xu, R. (2008). *Clustering*. John Wiley & Sons.
- Xie, Xiong, Z.-Y., Dai, Q.-Z., Wang, X.-X., dan Zhang, Y.-F. (2020). A new internal index based on density core for clustering validation. *Information Sciences*, 506, 346–365.
- Xie, X. L., dan Beni, G. (1991). A validity measure for fuzzy clustering. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 13(08), 841–847.
- Xu, dan Tian, Y. (2015). A comprehensive survey of clustering algorithms. *Annals of Data Science*.
- Xu, D., dan Tian, Y. (2015). A comprehensive survey of clustering algorithms. *Annals of Data Science*, 2, 165–193.
- Xu, R., dan Wunsch, D. (2005). Survey of clustering algorithms. *IEEE Transactions on neural networks*, 16(3), 645–678.
- Zhang, Feng, Q., Huang, J., Guo, Y., Xu, J., dan Wang, J. (2021). A local search algorithm for k-means with outliers. *Neurocomputing*, 450, 230–241.
- Zhang, Lu, F., Liu, A., Guo, P., dan Liu, C. (2016). Application of k-means clustering algorithm for classification of nba guards. *International Journal of Science and Engineering Applications*, 5(1), 1–6.