

DAFTAR PUSTAKA

- Adi, S. (2022). Komparasi Metode Support Vector Machine (Svm), K-Nearest Neighbors (Knn), Dan Random Forest (Rf) Untuk Prediksi Penyakit Gagal Jantung. *Jurnal Ilmiah Matematika*, 10(02), 258–268.
- Agarwal, S. (2014). Data mining: Data mining concepts and techniques. *Proceedings - 2013 International Conference on Machine Intelligence Research and Advancement, ICMIRA 2013*, 203–207. <https://doi.org/10.1109/ICMIRA.2013.45>
- Aini, S. H. A., Sari, Y. A., & Arwan, A. (2018). Seleksi Fitur Information Gain untuk Klasifikasi Penyakit Jantung Menggunakan Kombinasi Metode K-Nearest Neighbor dan Naïve Bayes. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 2(9), 2546–2554. <http://j-ptiik.ub.ac.id>
- Andrade, J. J., Da Fonseca, L. G., Farage, M., & Marques, G. L. de O. (2020). Prediction of the Performance of Bituminous Mixes Using Adaptive Neuro-Fuzzy Inference Systems. *Revista Mundi Engenharia, Tecnologia e Gestão (ISSN: 2525-4782)*, 5(6), 1–14. <https://doi.org/10.21575/25254782rmetg2020vol5n61367>
- Angelina Sagita, Abdurahman Baizal, M. A. B. (2010). Analisa Pengaruh Metode Combine Sampling Dalam Churn Prediction Untuk Perusahaan Telekomunikasi. *Seminar Nasional Informatika, 2010(semnasIF)*, 14–22.
- Anis, M., & Ali, M. (2017). Investigating the Performance of Smote for Class Imbalanced Learning: A Case Study of Credit Scoring Datasets. *European Scientific Journal, ESJ*, 13(33), 340. <https://doi.org/10.19044/esj.2017.v13n33p340>
- Batista, G. E. A. P. A., & Monard, M. C. (2003). An analysis of four missing data treatment methods for supervised learning. *Applied Artificial Intelligence*, 17(5–6), 519–533. <https://doi.org/10.1080/713827181>
- Beckmann, M., Ebecken, N. F. F., & Pires de Lima, B. S. L. (2015). A KNN Undersampling Approach for Data Balancing. *Journal of Intelligent Learning Systems and Applications*, 07(04), 104–116. <https://doi.org/10.4236/jilsa.2015.74010>

- Béjot, Y., Bailly, H., Durier, J., & Giroud, M. (2016). Epidemiology of stroke in Europe and trends for the 21st century. *Presse Medicale*, 45(12), e391–e398. <https://doi.org/10.1016/j.lpm.2016.10.003>
- Benjamin, E. J., Blaha, M. J., Chiuve, S. E., Cushman, M., Das, S. R., Deo, R., De Ferranti, S. D., Floyd, J., Fornage, M., Gillespie, C., Isasi, C. R., Jim'nez, M. C., Jordan, L. C., Judd, S. E., Lackland, D., Lichtman, J. H., Lisabeth, L., Liu, S., Longenecker, C. T., ... Muntner, P. (2017). Heart Disease and Stroke Statistics'2017 Update: A Report from the American Heart Association. In *Circulation* (Vol. 135, Nomor 10). <https://doi.org/10.1161/CIR.0000000000000485>
- Berrar, D. (2018). Cross-validation. *Encyclopedia of Bioinformatics and Computational Biology: ABC of Bioinformatics*, 1–3(January 2018), 542–545. <https://doi.org/10.1016/B978-0-12-809633-8.20349-X>
- Beyan, C., & Fisher, R. (2015). Classifying imbalanced data sets using similarity based hierarchical decomposition. *Pattern Recognition*, 48(5), 1653–1672. <https://doi.org/10.1016/j.patcog.2014.10.032>
- Borkin, D., Némethová, A., Michal'çonok, G., & Maiorov, K. (2019). Impact of Data Normalization on Classification Model Accuracy. *Research Papers Faculty of Materials Science and Technology Slovak University of Technology*, 27(45), 79–84. <https://doi.org/10.2478/rput-2019-0029>
- Brownlee, J. (2020). Imbalanced Classification with Python - Choose Better Metrics, Balance Skewed Classes, and Apply Cost-Sensitive Learning. *Machine Learning Mastery*, 463. <https://books.google.com/books?hl=en&lr=&id=jaXJDwAAQBAJ&oi=fnd&pg=PP1&dq=%22intrusion+detection%22%7C%22anomaly+intrusion+detection%22%7C%22network+intrusion+detection%22%7C%22anomaly+based+network+intrusion+detection%22+%22imbalance+problem%22%7C%22imba>
- Cholil, S. R., Handayani, T., Prathivi, R., & Ardianita, T. (2021). Implementasi Algoritma Klasifikasi K-Nearest Neighbor (KNN) Untuk Klasifikasi Seleksi Penerima Beasiswa. *IJCIT (Indonesian Journal on Computer and Information Technology)*, 6(2), 118–127.

- Chormunge, S., & Jena, S. (2016). Efficient feature subset selection algorithm for high dimensional data. *International Journal of Electrical and Computer Engineering*, 6(4), 1880–1888. <https://doi.org/10.11591/ijece.v6i4.9800>
- Darojah, Z., Susetyoko, R., Ramadijanti, N., Elektronika, P., Surabaya, N., Korespondensi, P., & Network, N. (2023). *STRATEGI PENANGANAN IMBALANCE CLASS PADA MODEL KLASIFIKASI PENERIMA KARTU INDONESIA PINTAR KULIAH BERBASIS NEURAL NETWORK MENGGUNAKAN KOMBINASI SMOTE DAN ENN NEURAL NETWORK USING HYBRID SAMPLING TECHNIQUE COMBINING SMOTE AND ENN FOR IMBALANCED KARTU INDON.* 10(2), 457–466. <https://doi.org/10.25126/jtiik.2023106480>
- Delany, S. J. (2021). *k-Nearest Neighbour Classifiers - A Tutorial.* 54(6).
- Dogan, A., & Birant, D. (2021). Machine learning and data mining in manufacturing. *Expert Systems with Applications*, 166, 114060. <https://doi.org/10.1016/j.eswa.2020.114060>
- Enriko, I. K. A. (2019). Comparative study of heart disease diagnosis using top ten data mining classification algorithms. *ACM International Conference Proceeding Series*, 159–164. <https://doi.org/10.1145/3338188.3338220>
- Fan, Y., Liu, Y., Chen, H., & Ma, J. (2019). Data mining-based design and implementation of college physical education performance management and analysis system. *International Journal of Emerging Technologies in Learning*, 14(6), 87–97. <https://doi.org/10.3991/ijet.v14i06.10159>
- Feigin, V. L., Brainin, M., Norrving, B., Martins, S., Sacco, R. L., Hacke, W., Fisher, M., Pandian, J., & Lindsay, P. (2022). World Stroke Organization (WSO): Global Stroke Fact Sheet 2022. *International Journal of Stroke*, 17(1), 18–29. <https://doi.org/10.1177/17474930211065917>
- Ginting, A. K. B., Lydia, M. S., & Zamzami, E. M. (2021). Peningkatan Akurasi Metode K-Nearest Neighbor dengan Seleksi Fitur Symmetrical Uncertainty. *Jurnal Media Informatika Budidarma*, 5(4), 1714. <https://doi.org/10.30865/mib.v5i4.3254>
- Gu, Q., Cai, Z., Zhu, L., & Huang, B. (2008). Data mining on imbalanced data sets. *Proceedings - 2008 International Conference on Advanced Computer Theory*

- and Engineering, *ICACTE* 2008, 1020–1024.
<https://doi.org/10.1109/ICACTE.2008.26>
- Ha, J., Kambe, M., & Pe, J. (2011). Data Mining: Concepts and Techniques. In *Data Mining: Concepts and Techniques*. <https://doi.org/10.1016/C2009-0-61819-5>
- Haryawan, C., Muria, Y., & Ardhana, K. (2023). *ANALISA PERBANDINGAN TEKNIK OVERSAMPLING SMOTE*. 6(1), 73–78.
- He, H., & Ma, Y. (2013). *Imbalanced Learning: Foundation, Algorithms, and Applications*. John Wiley & Sons, Inc.,
<https://doi.org/http://dx.doi.org/10.1002/9781118646106>
- Henderi, H. (2021). Comparison of Min-Max normalization and Z-Score Normalization in the K-nearest neighbor (kNN) Algorithm to Test the Accuracy of Types of Breast Cancer. *IJIIS: International Journal of Informatics and Information Systems*, 4(1), 13–20.
<https://doi.org/10.47738/ijiis.v4i1.73>
- Hubert, M., & Vandervieren, E. (2008). An adjusted boxplot for skewed distributions. *Computational Statistics and Data Analysis*, 52(12), 5186–5201.
<https://doi.org/10.1016/j.csda.2007.11.008>
- Ihfa, R., & Harsanti, T. (2019). Komparasi Teknik Resampling Pada Pemodelan Regresi Logistik Biner Studi Kasus : Sikap Remaja terhadap Keperawanan di Indonesia Tahun 2017 Case Study : Youth Attitudes towards Virginity in Indonesia in 2017). *Pengembangan Official Statistics Dalam Mendukung Implementasi SDGs, 2017*, 863–870.
- Indrawati, A. (2021). Penerapan Teknik Kombinasi Oversampling Dan Undersampling Hybrid Oversampling and Undersampling Techniques To Handling Imbalanced Dataset. *JIKO(Jurnal Informatika dan Komputer)*, 4(1), 38–43. <https://doi.org/10.33387/jiko>
- Jiang, F., Jiang, Y., Zhi, H., Dong, Y., Li, H., Ma, S., Wang, Y., Dong, Q., Shen, H., & Wang, Y. (2017). Artificial intelligence in healthcare: Past, present and future. *Stroke and Vascular Neurology*, 2(4), 230–243.
<https://doi.org/10.1136/svn-2017-000101>
- Khushi, M., Shaukat, K., Alam, T. M., Hameed, I. A., Uddin, S., Luo, S., Yang, X.,

- & Reyes, M. C. (2021). A Comparative Performance Analysis of Data Resampling Methods on Imbalance Medical Data. *IEEE Access*, 9, 109960–109975. <https://doi.org/10.1109/ACCESS.2021.3102399>
- Kohsasih, K. L., & Situmorang, Z. (2022). Comparative Analysis of C4.5 and Naïve Bayes Algorithms in Predicting Cerebrovascular Disease. *Jurnal Informatika*, 9(1), 13–17.
- Kulkarni, A., Chong, D., & Batarseh, F. A. (2020). Foundations of data imbalance and solutions for a data democracy. In *Data Democracy: At the Nexus of Artificial Intelligence, Software Development, and Knowledge Engineering*. Elsevier Inc. <https://doi.org/10.1016/B978-0-12-818366-3.00005-8>
- Langhorne, P., Audebert, H. J., Cadilhac, D. A., Kim, J., & Lindsay, P. (2020). Stroke systems of care in high-income countries: what is optimal? *The Lancet*, 396(10260), 1433–1442. [https://doi.org/10.1016/S0140-6736\(20\)31363-5](https://doi.org/10.1016/S0140-6736(20)31363-5)
- Lee, S. J., & Siau, K. (2001). A review of data mining techniques. *Industrial Management and Data Systems*, 101(1), 41–46. <https://doi.org/10.1108/02635570110365989>
- Lingga, L. (2013). *All About Stroke: Hidup Sebelum dan Pasca Stroke*. Elex Media Komputindo.
- Mandrekar, J. N. (2010). Receiver operating characteristic curve in diagnostic test assessment. *Journal of Thoracic Oncology*, 5(9), 1315–1316. <https://doi.org/10.1097/JTO.0b013e3181ec173d>
- Marcot, B. G., & Hanea, A. M. (2021). What is an optimal value of k in k-fold cross-validation in discrete Bayesian network analysis? *Computational Statistics*, 36(3), 2009–2031. <https://doi.org/10.1007/s00180-020-00999-9>
- Mishra, A., Gupta, B. B., Perakovi, D., & Peñalvo, F. J. G. (2021). A Survey on Data mining classification approaches. *CEUR Workshop Proceedings*, 3080(July), 264–268.
- MIT. (2016). *Secondary Analysis of Electronic Health Records*. Springer Nature. <https://doi.org/10.1007/978-3-319-43742-2>
- Mualfah, D., Fadila, W., & Firdaus, R. (2022). Teknik SMOTE untuk Mengatasi Imbalance Data pada Deteksi Penyakit Stroke Menggunakan Algoritma Random Forest. *Jurnal CoSciTech (Computer Science and Information*

- Technology*), 3(2), 107–113. <https://doi.org/10.37859/coscitech.v3i2.3912>
- Muqtiit WS, A., & Nooraeni, R. (2020). Penerapan Metode Resampling Dalam Mengatasi Imbalanced Data Pada Determinan Kasus Diare Pada Balita Di Indonesia (Analisis Data Sdki 2017). *Jurnal MSA (Matematika dan Statistika serta Aplikasinya)*, 8(1), 19. <https://doi.org/10.24252/msa.v8i1.13452>
- Nasution, D. A., Khotimah, H. H., & Chamidah, N. (2019). Perbandingan Normalisasi Data untuk Klasifikasi Wine Menggunakan Algoritma K-NN. *Computer Engineering, Science and System Journal*, 4(1), 78. <https://doi.org/10.24114/cess.v4i1.11458>
- P2PTM Kemenkes RI. (2018). *Apa itu Stroke?*
- P2PTM Kemenkes RI. (2022). *Tingkatan Kualitas dan Layanan Stroke Lewat Transformasi Kesehatan.*
- Pandurangan, P. R. (2021). *Data Mining Approaches to Predict the Factors that Affect the Agriculture International Journal of Computer Sciences and Engineering Open Access Data Mining Approaches to Predict the Factors that Affect the Agriculture Growth using Stochastic Model. November, 17–23.*
- Patro, S. G. K., & sahu, K. K. (2015). Normalization: A Preprocessing Stage. *Iarjset*, 20–22. <https://doi.org/10.17148/iarjset.2015.2305>
- Pereira, R. M., Costa, Y. M. G., & Silla, C. N. (2020). MLTL: A multi-label approach for the Tomek Link undersampling algorithm: MLTL: The Multi-Label Tomek Link. *Neurocomputing*, 383, 95–105. <https://doi.org/10.1016/j.neucom.2019.11.076>
- Pham, H. (2006). Springer Handbook of Engineering Statistics. In *Springer London*. Springer London. https://doi.org/10.1007/springerreference_60775
- Pristyanto, Y. (2019). Penerapan Metode Ensemble Untuk Meningkatkan Kinerja Algoritme Klasifikasi Pada Imbalanced Dataset. *Jurnal Teknoinfo*, 13(1), 11. <https://doi.org/10.33365/jti.v13i1.184>
- Puspitawuri, A., Santoso, E., & Dewi, C. (2019). Diagnosis Tingkat Risiko Penyakit Stroke Menggunakan Metode K-Nearest Neighbor dan Naïve Bayes. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 3(4), 3319–3324. <https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/4916>
- Ramadhan, N. G., & Khoirunnisa, A. (2021). Klasifikasi Data Malaria

- Menggunakan Metode Support Vector Machine. *Jurnal Media Informatika Budidarma*, 5(4), 1580. <https://doi.org/10.30865/mib.v5i4.3347>
- RISKESDAS. (2018). Laporan Nasional RISKESDAS 2018. In *Kementerian Kesehatan RI*. <https://www.kemkes.go.id/article/view/19093000001/penyakit-jantung-penyebab-kematian-terbanyak-ke-2-di-indonesia.html>
- Said, H., Matondang, N. H., & Irmanda, H. N. (2022). Penerapan Algoritma K-Nearest Neighbor Untuk Memprediksi Kualitas Air Yang Dapat Dikonsumsi. *Techno.Com*, 21(2), 256–267. <https://doi.org/10.33633/tc.v21i2.5901>
- Saifudin, A., Pamulang, U., Wahono, R. S., Dian, U., & Semarang, N. (2015). Pendekatan Level Data untuk Menangani Ketidakseimbangan Kelas pada Prediksi Cacat Software. *Journal of Software Engineering*, 1(2), 76–85.
- Shalabi, L. Al, Shaaban, Z., & Kasasbeh, B. (2006). Data Mining: A Preprocessing Engine. *Journal of Computer Science*, 2(9), 735–739. <https://doi.org/10.3844/jcssp.2006.735.739>
- Siringoringo, R. (2018). Klasifikasi data tidak Seimbang menggunakan algoritma SMOTE dan k-nearest neighbor. *Jurnal ISD*, 3(1), 44–49.
- Torbey, M. T., & Selim, M. H. (2013). *The Stroke Book* (Second). Cambridge University Press.
- Uysal, A. K. (2018). On Two-Stage Feature Selection Methods for Text Classification. *IEEE Access*, 6, 43233–43251. <https://doi.org/10.1109/ACCESS.2018.2863547>
- Vluymans, S. (2019). Learning from imbalanced data. *Studies in Computational Intelligence*, 807(9), 81–110. https://doi.org/10.1007/978-3-030-04663-7_4
- Witten, I. H., Frank, E., & Hall, M. A. (2011). Data Mining. In *Data Mining*. <https://doi.org/10.1016/C2009-0-19715-5>
- Xu, Z., Shen, D., Nie, T., & Kou, Y. (2020). A hybrid sampling algorithm combining M-SMOTE and ENN based on Random forest for medical imbalanced data. *Journal of Biomedical Informatics*, 107(May 2019), 103465. <https://doi.org/10.1016/j.jbi.2020.103465>
- Yang, J., Li, Y., Liu, Q., Li, L., Feng, A., Wang, T., Zheng, S., Xu, A., & Lyu, J. (2020). Brief introduction of medical database and data mining technology in

big data era. *Journal of Evidence-Based Medicine*, 13(1), 57–69.
<https://doi.org/10.1111/jebm.12373>