

## DAFTAR PUSTAKA

- [1] Dara Sawitri, "Big Data Challenges And Opportunities In The Development Of Digital Technology," *Info Sains*, vol. 14, no. 02, hlm. 215–224, Jun 2024, doi: 10.54209/infosains.v14i02.4619.
- [2] M. Muniswamaiah, T. Agerwala, dan C. C. Tappert, "Big Data and Data Visualization Challenges," dalam *2023 IEEE International Conference on Big Data (BigData)*, Sorrento, Italy: IEEE, Des 2023, hlm. 6227–6229. doi: 10.1109/BigData59044.2023.10386491.
- [3] R. Indrakumari, T. Poongodi, P. Suresh, dan B. Balamurugan, "The growing role of integrated and insightful big and real-time data analytics platforms," dalam *Advances in Computers*, vol. 117, Elsevier, 2020, hlm. 165–186. doi: 10.1016/bs.adcom.2019.09.009.
- [4] S. Lokhande dan N. Khare, "An Outlook on Big Data and Big Data Analytics," *IJCA*, vol. 124, no. 11, hlm. 37–41, Agu 2015, doi: 10.5120/ijca2015905658.
- [5] Ms. Manali Sakpal, "Big Data : Analysis," *IJAR SCT*, hlm. 515–520, Jan 2024, doi: 10.48175/IJAR SCT-15074.
- [6] C. L. Jordan, R. Koochak, M. Roberts, A. Nalonnil, dan M. Honeychurch, "A Holistic Approach to Big Data and Data Analytics for Automated Reservoir Surveillance and Analysis," dalam *SPE Asia Pacific Oil & Gas Conference and Exhibition*, Adelaide, Australia: SPE, Okt 2022, hlm. D011S004R001. doi: 10.2118/210757-MS.
- [7] G. Malini, "Automation of Big Data Analytics Using Robotic Process Automation," *IJSRCSEIT*, hlm. 602–605, Apr 2021, doi: 10.32628/CSEIT2172124.
- [8] A. Mumuni dan F. Mumuni, "Automated data processing and feature engineering for deep learning and big data applications: A survey," *Journal of Information and Intelligence*, hlm. S2949715924000027, Jan 2024, doi: 10.1016/j.jiixd.2024.01.002.
- [9] M. Ramkumar, K. Malathi, dan K. Pavithra, "Optimizing Machine Learning Model Accuracy via OBNT Algorithm: Advanced Data Preprocessing Technique," dalam *2023 International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSES)*, Chennai, India: IEEE, Des 2023, hlm. 1–6. doi: 10.1109/ICSES60034.2023.10465344.
- [10] E. Hikmawati, N. U. Maulidevi, dan K. Surendro, "Adaptive rule: A novel framework for recommender system," *ICT Express*, vol. 6, no. 3, hlm. 214–219, Sep 2020, doi: 10.1016/j.icte.2020.06.001.

- [11] E. Hikmawati, N. U. Maulidevi, dan K. Surendro, "Pruning Strategy on Adaptive Rule Model by Sorting Utility Items," *IEEE Access*, vol. 10, hlm. 91650–91662, 2022, doi: 10.1109/ACCESS.2022.3202307.
- [12] E. Hikmawati, N. U. Maulidevi, dan K. Surendro, "Rule-ranking method based on item utility in adaptive rule model," *PeerJ Computer Science*, vol. 8, hlm. e1013, Jun 2022, doi: 10.7717/peerj-cs.1013.
- [13] E. Hikmawati, N. U. Maulidevi, dan K. Surendro, "Improved Classification Accuracy by Feature Selection using Adaptive Support Method," dalam *Proceedings of the 2023 12th International Conference on Software and Computer Applications*, Kuantan Malaysia: ACM, Feb 2023, hlm. 171–176. doi: 10.1145/3587828.3587854.
- [14] E. Hikmawati, H. Nugroho, dan K. Surendro, "Improve the Quality of Recommender Systems based on Collaborative Filtering with Missing Data Imputation," dalam *Proceedings of the 2024 13th International Conference on Software and Computer Applications*, Bali Island Indonesia: ACM, Feb 2024, hlm. 75–80. doi: 10.1145/3651781.3651793.
- [15] P. K. Joshi, R. Prakash, dan A. K. Rai, "A Comprehensive Review of Machine Learning Application Across Different Domains," dalam *2024 2nd International Conference on Disruptive Technologies (ICDT)*, Greater Noida, India: IEEE, Mar 2024, hlm. 1266–1270. doi: 10.1109/ICDT61202.2024.10489222.
- [16] N. L. Rane, S. K. Mallick, Ö. Kaya, dan J. Rane, "Applications of machine learning in healthcare, finance, agriculture, retail, manufacturing, energy, and transportation: A review," dalam *Applied Machine Learning and Deep Learning: Architectures and Techniques*, Deep Science Publishing, 2024. doi: 10.70593/978-81-981271-4-3\_6.
- [17] N. Mishra, S. V. T. Sangeetha, D. K, dan V. Venugopal, "Data-Driven Models for Water Potability: A Machine Learning Approach," dalam *2024 Third International Conference on Intelligent Techniques in Control, Optimization and Signal Processing (INCOS)*, Krishnankoil, Virudhunagar district, Tamil Nadu, India: IEEE, Mar 2024, hlm. 1–6. doi: 10.1109/INCOS59338.2024.10527476.
- [18] Nanda Aulia Sofiah, Fanny Olivia, dan M. I. Jambak, "Predictive Analytics for Water Safety: Data Mining and Supervised Learning in Potability Classification," *ijcs*, vol. 13, no. 4, Jul 2024, doi: 10.33022/ijcs.v13i4.3884.
- [19] S. Saitala, "AI and Cloud Computing in Healthcare: Transforming Patient Care with AWS and Java," *IJRASET*, vol. 12, no. 9, hlm. 326–333, Sep 2024, doi: 10.22214/ijraset.2024.63684.

- [20] A. M, V. P, K. A. C, V. J, A. G, dan R. B. C, "Design and Analysis of Deep Learning based Water Potability Prediction," dalam *2023 International Conference on Self Sustainable Artificial Intelligence Systems (ICSSAS)*, Erode, India: IEEE, Okt 2023, hlm. 477–482. doi: 10.1109/ICSSAS57918.2023.10331807.
- [21] E. Okoro, "Potable Water Sources, Household Hygiene, and Sanitation Practices in Ikpoba Okha LGA, Edo State: Implications for Public Health and Sustainable Water Management Omoregie, Andrew Edosa.1 Omoregie Abieyuwa Peace2 Okoro, Enyinnaya Okoro.31 College of Medicine, University of Benin, Edo State, Nigeria2Institute of Child Health, University of Benin3Environmental Management and Toxicology Federal University Otuokeyeandrew. omoregie@uniben.edu.ng, abieyuwa.omoregie@uniben.edu.ng , okoroee@fuotuokeye.edu.ng, +2348034442722, +2348063583693Abstract Access to potable water and proper sanitation remains a pressing global issue, particularly in developing regions like Nigeria, where waterborne diseases are prevalent due to inadequate water a (Preprint)," 5 Oktober 2024, *JMIR Preprints*. doi: 10.2196/preprints.67225.
- [22] S. Saitala, "AI and Cloud Computing in Healthcare: Transforming Patient Care with AWS and Java," *IJRASET*, vol. 12, no. 9, hlm. 326–333, Sep 2024, doi: 10.22214/ijraset.2024.63684.
- [23] L. Hu dan L. Li, "Using Tree-Based Machine Learning for Health Studies: Literature Review and Case Series," *IJERPH*, vol. 19, no. 23, hlm. 16080, Des 2022, doi: 10.3390/ijerph192316080.
- [24] "Comparative Study of Tree-Based Methods for Concrete Compressive Strength Analysis," dalam *CONFERENCE PROCEEDING, ICCDETS*, Mei 2024. doi: 10.62919/mjch0842.
- [25] Y. Chen, S. Chen, Y. Yang, dan S. Lu, "Comparison of decision tree and ensemble algorithms," *ACE*, vol. 55, no. 1, hlm. 241–248, Jul 2024, doi: 10.54254/2755-2721/55/20241535.
- [26] C. Xianjun, "Research on the Application of Decision Tree Algorithm in Agricultural Economic Development," dalam *2024 Second International Conference on Data Science and Information System (ICDSIS)*, Hassan, India: IEEE, Mei 2024, hlm. 1–4. doi: 10.1109/ICDSIS61070.2024.10594692.
- [27] A. Salah dan D. Yevick, "A Random Forest Model for Predicting and Analyzing the Performance of CNT TFET with Highly Doped Pockets," *Advcd Theory and Sims*, hlm. 2400607, Sep 2024, doi: 10.1002/adts.202400607.

- [28] H. A. Salman, A. Kalakech, dan A. Steiti, "Random Forest Algorithm Overview," *Babylonian Journal of Machine Learning*, vol. 2024, hlm. 69–79, Jun 2024, doi: 10.58496/BJML/2024/007.
- [29] UG student, Dept. of Electronics and Computers Engineering, Sreenidhi Institute of Science and Technologies, Telangana, India dan . Swarna S, "Machine Learning Based Heart Disease Prediction Using Random Forest," *IJSREM*, vol. 08, no. 04, hlm. 1–5, Apr 2024, doi: 10.55041/IJSREM31320.
- [30] L. Lyu, Y. Hu, N. Wang, X. Zhou, dan M. Fang, "Application of Deep Learning and Transfer Learning in Continuous Missing Value Imputation of Water Quality Data," dalam *2022 IEEE 8th International Conference on Computer and Communications (ICCC)*, Chengdu, China: IEEE, Des 2022, hlm. 1211–1216. doi: 10.1109/ICCC56324.2022.10065657.
- [31] M. Pazo, S. Gerassis, M. Araújo, I. Margarida Antunes, dan X. Rigueira, "Enhancing water quality prediction for fluctuating missing data scenarios: A dynamic Bayesian network-based processing system to monitor cyanobacteria proliferation," *Science of The Total Environment*, vol. 927, hlm. 172340, Jun 2024, doi: 10.1016/j.scitotenv.2024.172340.
- [32] D. Sierra-Porta, "Assessing the impact of missing data on water quality index estimation: a machine learning approach," *Discov Water*, vol. 4, no. 1, hlm. 11, Mar 2024, doi: 10.1007/s43832-024-00068-y.
- [33] S. Kapare dan M. Anu, "Study of Machine Learning Tactics for Missing Data Imputation and Classification," dalam *2024 MIT Art, Design and Technology School of Computing International Conference (MITADTSOCiCon)*, Pune, India: IEEE, Apr 2024, hlm. 1–4. doi: 10.1109/MITADTSOCiCon60330.2024.10575776.
- [34] J. Xue, "Review on data imputation methods in machine learning," *J. Phys.: Conf. Ser.*, vol. 2646, no. 1, hlm. 012034, Des 2023, doi: 10.1088/1742-6596/2646/1/012034.
- [35] H. Nugroho dan K. Surendro, "A Comprehensive Bibliometric Analysis of Missing Value Imputation," *IEEE Access*, vol. 12, hlm. 14819–14846, 2024, doi: 10.1109/ACCESS.2024.3357533.
- [36] D. A. Dablain dan N. V. Chawla, "The Hidden Influence of Latent Feature Magnitude When Learning with Imbalanced Data," 2024, *arXiv*. doi: 10.48550/ARXIV.2407.10165.
- [37] Z. Liu, Y. Hu, T. Pang, Y. Zhou, P. Ren, dan Y. Yang, "Model Balancing Helps Low-data Training and Fine-tuning," 2024, *arXiv*. doi: 10.48550/ARXIV.2410.12178.

- [38] M. Ayyannan, "Accuracy Enhancement of Machine Learning Model by Handling Imbalance Data," dalam *2024 International Conference on Expert Clouds and Applications (ICOECA)*, Bengaluru, India: IEEE, Apr 2024, hlm. 593–599. doi: 10.1109/ICOECA62351.2024.00109.
- [39] A. Chebli, S. Daas, dan T. Hafs, "Evaluating the impact of data imputation on model precision in machine learning," *SEES*, vol. 5, no. 2, hlm. e8310, Sep 2024, doi: 10.54021/seesv5n2-256.
- [40] A. J. Mohammed dan A. M. Abdulazeez, "Healthcare Diseases Classification Based on Machine Learning Algorithms: A Review," *Int. J. Inform. Inf. Sys. and Comp. Eng.*, vol. 5, no. 2, hlm. 218–252, Jul 2024, doi: 10.34010/injiiscom.v5i2.13581.
- [41] M.-A. Suci dan R. Ioana Lung, "A game theoretic decision forest for feature selection and classification," *Logic Journal of the IGPL*, hlm. jzae049, Mei 2024, doi: 10.1093/jigpal/jzae049.
- [42] M. N. Arefin dan A. K. M. Masum, "A Probabilistic Approach for Missing Data Imputation," *Complexity*, vol. 2024, hlm. 1–15, Jan 2024, doi: 10.1155/2024/4737963.
- [43] A. Finocchi *dkk.*, "Multiple imputation integrated to machine learning: predicting post-stroke recovery of ambulation after intensive inpatient rehabilitation," *Sci Rep*, vol. 14, no. 1, hlm. 25188, Okt 2024, doi: 10.1038/s41598-024-74537-8.
- [44] S. Xie, J. Zhang, dan A. T. Lawniczak, "Handling Data Imbalance In Linear Modelling of Fatality Rate of Auto Collision," dalam *Proceedings of the 2023 6th International Conference on Information Science and Systems*, Edinburgh United Kingdom: ACM, Agu 2023, hlm. 83–89. doi: 10.1145/3625156.3625169.
- [45] "Kaggle." [Daring]. Tersedia pada: <https://www.kaggle.com/datasets/nayanack/water-probability>
- [46] S. Chen dan C. Xu, "Predictive Mean Matching Imputation Procedure Based on Machine Learning Models for Complex Survey Data," *Journal of Data Science*, hlm. 456–468, 2024, doi: 10.6339/24-JDS1135.
- [47] L. R. Abramson, "Some drawbacks in using the mean probability," dalam *[1990] Proceedings. First International Symposium on Uncertainty Modeling and Analysis*, College Park, MD, USA: IEEE Comput. Soc. Press, 1991, hlm. 321–323. doi: 10.1109/ISUMA.1990.151272.
- [48] J.-B. Hardouin, R. Conroy, dan V. Sébille, "Imputation by the mean score should be avoided when validating a Patient Reported Outcomes questionnaire by a Rasch model in presence of informative missing data," *BMC Med Res Methodol*, vol. 11, no. 1, hlm. 105, Des 2011, doi: 10.1186/1471-2288-11-105.

- [49] R. C. Moore, D. P. W. Ellis, E. Fonseca, S. Hershey, A. Jansen, dan M. Plakal, "Dataset Balancing Can Hurt Model Performance," dalam *ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Rhodes Island, Greece: IEEE, Jun 2023, hlm. 1–5. doi: 10.1109/ICASSP49357.2023.10095255.
- [50] R. Schwartz dan G. Stanovsky, "On the Limitations of Dataset Balancing: The Lost Battle Against Spurious Correlations," 2022, *arXiv*. doi: 10.48550/ARXIV.2204.12708.