

DAFTAR PUSTAKA

- [1] I. Brown and C. Mues, "An experimental comparison of classification algorithms for imbalanced credit scoring data sets," *An experimental comparison of classification algorithms for imbalanced credit scoring data sets*, vol. 39, no. 3, pp. 3446–3453, 2012, doi: 10.1016/j.eswa.2011.09.033.
- [2] Y. Zou and C. Gao, "Extreme Learning Machine Enhanced Gradient Boosting for Credit Scoring," *Algorithms*, vol. 15, no. 5, 2022, doi: 10.3390/a15050149.
- [3] C. Qin, Y. Zhang, F. Bao, C. Zhang, P. Liu, and P. Liu, "XGBoost optimized by adaptive particle swarm optimization for credit scoring," *Math. Probl. Eng.*, vol. 2021, 2021, doi: 10.1155/2021/6655510.
- [4] O. Berman, Z. Drezner, and G. O. Wesolowsky, "The expropriation location problem," *Journal of the Operational Research Society*, vol. 54, no. 7, pp. 769–776, Jul. 2003, doi: 10.1057/palgrave.jors.2601548.
- [5] Y. Li and W. Chen, "A comparative performance assessment of ensemble learning for credit scoring," *Mathematics*, vol. 8, no. 10, pp. 1–19, 2020, doi: 10.3390/math8101756.
- [6] Andri, Darwin, N. P. Wong, Sutarman, and E. B. Nababan, "Evaluating Hyperparameter Tuning of Random Forest and CatBoost on Complex and Imbalanced Real-world Datasets," *Pertanika J. Sci. Technol.*, vol. 34, no. 1, pp. 283–315, 2026, doi: 10.47836/pjst.34.1.13.
- [7] J. Al Faysal *et al.*, "XGB-RF: A Hybrid Machine Learning Approach for IoT Intrusion Detection," *Telecom*, vol. 3, no. 1, pp. 52–69, 2022, doi: 10.3390/telecom3010003.
- [8] D. M. Alghazzawi, A. G. A. Alquraishie, S. K. Badri, and S. H. Hasan, "ERF-XGB: Ensemble Random Forest-Based XG Boost for Accurate Prediction and Classification of E-Commerce Product Review," *Sustainability (Switzerland)*, vol. 15, no. 9, 2023, doi: 10.3390/su15097076.
- [9] S. Chatterjee, D. Corbae, K. Dempsey, and J.-V. Ríos-Rull, "A Quantitative Theory of the Credit Score," *Econometrica*, vol. 91, no. 5, pp. 1803–1840, 2023, doi: 10.3982/ecta18771.
- [10] T. Bank, N. Segev, and M. O. Shaton, "Relationship Banking and Credit Scores: Evidence from a Natural Experiment," *SSRN Electronic Journal*, pp. 1–30, 2023, doi: 10.2139/ssrn.4567568.
- [11] M. Altalhan, A. Algarni, and M. Turki-Hadj Alouane, "Imbalanced Data Problem in Machine Learning: A Review," *IEEE Access*, vol. 13, no. December 2024, pp. 13686–13699, 2025, doi: 10.1109/ACCESS.2025.3531662.
- [12] P. Zhang, Y. Jia, and Y. Shang, "Research and application of XGBoost in imbalanced data," *Int. J. Distrib. Sens. Netw.*, vol. 18, no. 6, 2022, doi: 10.1177/15501329221106935.

- [13] E. Strelcenia and S. Prakoonwit, "A Survey on GAN Techniques for Data Augmentation to Address the Imbalanced Data Issues in Credit Card Fraud Detection," *Mach. Learn. Knowl. Extr.*, vol. 5, no. 1, pp. 304–329, 2023, doi: 10.3390/make5010019.
- [14] V. Çetin and O. Yıldız, "A comprehensive review on data preprocessing techniques in data analysis," *Pamukkale University Journal of Engineering Sciences*, vol. 28, no. 2, pp. 299–312, 2022, doi: 10.5505/pajes.2021.62687.
- [15] S. K. Ahmed, "How to choose a sampling technique and determine sample size for research: A simplified guide for researchers," Dec. 01, 2024, *Elsevier Ltd.* doi: 10.1016/j.oor.2024.100662.
- [16] C. Tang, D. Wang, A.-H. Tan, and C. Miao, "EEG-Based Emotion Recognition via Fast and Robust Feature Smoothing," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 10654 LNCS, pp. 83–92, 2017, doi: 10.1007/978-3-319-70772-3_8
- [17] A. Hana and Jay Lofstead, "Managing Randomness to Enable Reproducible Machine Learning," *Managing Randomness to Enable Reproducible Machine Learning*, pp. 15–20, Jun. 2022, doi: DOI:10.1145/3526062.3536353.
- [18] J. Graffelman and J. de Leeuw, "Improved Approximation and Visualization of the Correlation Matrix," *American Statistician*, vol. 77, no. 4, pp. 432–442, 2023, doi: 10.1080/00031305.2023.2186952.
- [19] P. Koukaras and C. Tjortjis, "Data Preprocessing and Feature Engineering for Data Mining: Techniques, Tools, and Best Practices," *AI (Switzerland)*, vol. 6, no. 10, 2025, doi: 10.3390/ai6100257.
- [20] P. O. Côté, A. Nikanjam, N. Ahmed, D. Humeniuk, and F. Khomh, "Data cleaning and machine learning: a systematic literature review," *Automated Software Engineering*, vol. 31, no. 2, 2024, doi: 10.1007/s10515-024-00453-w.
- [21] M. Lindauer and C. Doerr, "Automated Data Preparation for Machine Learning : A Survey," 2026.
- [22] M. Buddhadev, "An analytical approach to data quality: Techniques, challenges and tools in preprocessing," *International Journal of Cloud Computing and Database Management*, vol. 6, no. 1, pp. 47–50, 2025, doi: 10.33545/27075907.2025.v6.i1a.84.
- [23] ACAPS, "Data Cleaning," ACAPS Technical Note, Geneva, Switzerland, Apr. 2016. [Online]. Available: <http://www.acaps.org>.
- [24] C. Chrysostomou, "Efficient Data Cleaning and Preprocessing Techniques for Robust Machine Learning Models," 2023.
- [25] J. M. H. Pinheiro *et al.*, "The Impact of Feature Scaling in Machine Learning: Effects on Regression and Classification Tasks," *IEEE Access*, vol. 13, pp. 199903–199931, 2025, doi: 10.1109/ACCESS.2025.3635541.

- [26] W. Li, A. Halimi, X. Jiang, J. Vaidya, and E. Ayday, "Privacy-Preserving Model and Preprocessing Verification for Machine Learning," 2025.
- [27] J. E. Black, J. K. Kueper, and T. S. Williamson, "An introduction to machine learning for classification and prediction," *Fam. Pract.*, vol. 40, no. 1, pp. 200–204, 2023, doi: 10.1093/fampra/cmac104.
- [28] F. Y. Osisanwo, J. E. T. Akinsola, O. Awodele, J. O. Hinmikaiye, O. Olakanmi, and J. Akinjobi, "Supervised Machine Learning Algorithms: Classification and Comparison," *International Journal of Computer Trends and Technology*, vol. 48, no. 2, pp. 128–138, 2017, doi: 10.14445/22312803/IJCTT-V48P126.
- [29] A. F. A. H. Alnuaimi and T. H. K. Albaldawi, "An overview of machine learning classification techniques," *BIO Web Conf.*, vol. 97, pp. 1–24, 2024, doi: 10.1051/bioconf/20249700133.
- [30] S. Naeem, A. Ali, S. Anam, and M. M. Ahmed, "An Unsupervised Machine Learning Algorithms: Comprehensive Review," *International Journal of Computing and Digital Systems*, vol. 13, no. 1, pp. 911–921, 2023, doi: 10.12785/ijcds/130172.
- [31] Q. Gui, H. Zhou, N. Guo, and B. Niu, "A survey of class-imbalanced semi-supervised learning," *Mach. Learn.*, vol. 113, no. 8, pp. 5057–5086, 2024, doi: 10.1007/s10994-023-06344-7.
- [32] J. Yang *et al.*, "Deep reinforcement learning for multi-class imbalanced training: applications in healthcare," *Mach. Learn.*, vol. 113, no. 5, pp. 2655–2674, 2024, doi: 10.1007/s10994-023-06481-z.
- [33] I. D. Mienye and Y. Sun, "A Survey of Ensemble Learning: Concepts, Algorithms, Applications, and Prospects," *IEEE Access*, vol. 10, no. August, pp. 99129–99149, 2022, doi: 10.1109/ACCESS.2022.3207287.
- [34] P. Mahajan, S. Uddin, F. Hajati, and M. A. Moni, "Ensemble Learning for Disease Prediction: A Review," *Healthcare (Switzerland)*, vol. 11, no. 12, 2023, doi: 10.3390/healthcare11121808.
- [35] H. Zhao, W. Liu, Y. Wang, and L. Wu, "Comparative analysis of algorithmic approaches in ensemble learning: bagging vs. boosting," *Sci. Rep.*, vol. 15, no. 1, pp. 1–12, 2025, doi: 10.1038/s41598-025-15971-0.
- [36] T. Aboneh, A. Rorissa, and R. Srinivasagan, "Stacking-Based Ensemble Learning Method for Multi-Spectral Image Classification," *Technologies (Basel)*, vol. 10, no. 1, pp. 1–18, 2022, doi: 10.3390/technologies10010017.
- [37] H. Salman, A. Kalakech, and A. Steiti, "Random Forest Algorithm Overview," *Babylonian Journal of Machine Learning*, vol. 2024, pp. 69–79, 2024, doi: 10.58496/BJML/2024/007.
- [38] X. Y. Liew, N. Hameed, and J. Clos, "An investigation of XGBoost-based algorithm for breast cancer classification," *Machine Learning with Applications*, vol. 6, no. September, p. 100154, 2021, doi: 10.1016/j.mlwa.2021.100154.

- [39] B. F. Azevedo, A. M. A. C. Rocha, and A. I. Pereira, "Hybrid approaches to optimization and machine learning methods: a systematic literature review," *Mach. Learn.*, vol. 113, no. 7, pp. 4055–4097, Jul. 2024, doi: 10.1007/s10994-023-06467-x.
- [40] S. Zian, S. A. Kareem, and K. D. Varathan, "An Empirical Evaluation of Stacked Ensembles with Different Meta-Learners in Imbalanced Classification," *IEEE Access*, vol. 9, pp. 87434–87452, 2021, doi: 10.1109/ACCESS.2021.3088414.
- [41] S. Sathyanarayanan, "Confusion Matrix-Based Performance Evaluation Metrics," *African Journal of Biomedical Research*, pp. 4023–4031, Nov. 2024, doi: 10.53555/ajbr.v27i4s.4345.
- [42] C. Bliem, "Estimation of the (re-)integration like-lihood into the Austrian labour market: a random forest approach Subject Area: Information Systems," 2020.
- [43] S. Alam, M. S. Ayub, S. Arora, and M. A. Khan, "An investigation of the imputation techniques for missing values in ordinal data enhancing clustering and classification analysis validity," *Decision Analytics Journal*, vol. 9, Dec. 2023, doi: 10.1016/j.dajour.2023.100341.
- [44] S. M. Memon, R. Wamala, and I. H. Kabano, "A comparison of imputation methods for categorical data," *Inform. Med. Unlocked*, vol. 42, Jan. 2023, doi: 10.1016/j.imu.2023.101382.

