DAFTAR PUSTAKA

Adetunji, A. B., Akande, O. N., Ajala, F. A., Oyewo, O., Akande, Y. F., & Oluwadara, G. (2021). House Price Prediction using Random Forest Machine Learning Technique. *Procedia Computer Science*, *199*, 806–813. https://doi.org/10.1016/j.procs.2022.01.100

Almutairi, M. M., Yamin, M., & Halikias, G. (2021). An Analysis of Data Integration Challenges from Heterogeneous Databases. *Proceedings of the 2021 8th International Conference on Computing for Sustainable Global Development, INDIACom 2021*, *c*, 352–356. https://doi.org/10.1109/INDIACom51348.2021.00061

Arafat, S., Sun, N., Weese, M. L., & Martinez, W. G. (2023). Boundary Peeling: Outlier Detection Method Using One-Class Peeling. *arXiv*, 1–29. https://arxiv.org/abs/2309.05630

Bala, B., & Behal, S. (2024). A Brief Survey of Data Preprocessing in Machine Learning and Deep Learning Techniques. *8th International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud), I-SMAC 2024 - Proceedings*, 1755–1762. https://doi.org/10.1109/I-SMAC61858.2024.10714767

Bergstra, J., & Bengio, Y. (2012). Random Search for Hyper-Parameter Optimization. *Journal of Machine Learning Research*, *13*, 281–305.

Berrar, D. (2018). Cross-validation. *Encyclopedia of Bioinformatics and Computational Biology: ABC of Bioinformatics*, *1–3*(April), 542–545. https://doi.org/10.1016/B978-0-12-809633-8.20349-X

Bintoro, P., Ratnasari, Wihardjo, E., Putri, I. P., & Asari, A. (2024). *Pengantar Machine Learning* (A. Asari, Ed.; 1 ed.). Mafy Media Literasi Indonesia.

Bozdal, M., Ileri, K., & Ozkahraman, A. (2024). Comparative Analysis of Dimensionality Reduction Techniques. *Journal of Supercomputing*, *80*(1), 1059–1079. https://doi.org/10.1007/s11227-023-05511-w

Bressert, E. (2012). *SciPy and NumPy* (R. Roumeliotis & M. Blanchette, Ed.). O'Reilly Media.

Carvalho, M., Pinho, A. J., & Brás, S. (2025). Resampling Approaches to Handle Class Imbalance: A Review from a Data Perspective. *Journal of Big Data*, *12*(1). https://doi.org/10.1186/s40537-025-01119-4

Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, *13-17-Augu*, 785–794. https://doi.org/10.1145/2939672.2939785

Colliot, O. (2023). A Non-technical Introduction to Machine Learning. Dalam O. (eds) Colliot (Ed.), *Machine Learning for Brain Disorders* (1). Humana Press. https://doi.org/https://doi.org/10.1007/978-1-0716-3195-9_1

Dabool, H., Alashwal, H., Alnuaimi, H., Alhouqani, A., Alkaabi, S., & Al Ahbabi, A. (2024). Comparative Analysis of Hyperparameter Tuning Methods in Classification Models For Ensemble Learning. *ACAI 2024 - 2024 7th International Conference on Algorithms, Computing and Artificial Intelligence*, 1–5. https://doi.org/10.1109/ACAI63924.2024.10899492

Geerts, M., Broucke, S. Vanden, & Weerdt, J. De. (2023). A Survey of Methods and Input Data Types for House Price Prediction. *ISPRS International Journal of Geo-Information*, *12*(5), 200. https://doi.org/10.3390/ijgi12050200

Genuer, R., & Poggi, J.-M. (2020). *Random Forests with R* (R. Gentlemen, K. Hornik, & G. Parmigiani, Ed.). Springer Nature Switzerland AG. https://doi.org/10.1007/978-3-030-56485-8_1

Geron, A. (2019). *Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow* (N. Tache, Ed.; 2 ed.). O'Reilly Media.

Gnanadurai, J. B., & Vinothina, V. (2023). Optimization of Outlier Nodes of DBSCAN using PSO in Zone-Based Wireless Sensor Networks. *2023 Fifth International Conference on Electrical, Computer and Communication Technologies (ICECCT)*, 1–5. https://doi.org/10.1109/ICECCT56650.2023.10179691

Hamami, F., & Dahlan, I. A. (2024). Regression Modeling for House Price Prediction in Java Island. *ICSINTESA 2024 - 2024 4th International Conference of Science and Information Technology in Smart Administration: The Collaboration of Smart Technology and Good Governance for Sustainable Development Goals*, 402–406. https://doi.org/10.1109/ICSINTESA62455.2024.10747910

Hasan, M. K., Alam, M. A., Roy, S., Dutta, A., Jawad, M. T., & Das, S. (2021). Missing Value Imputation Affects the Performance of Machine Learning: A Review and Analysis of the Literature (2010–2021). *Informatics in Medicine Unlocked*, *27*, 100799. https://doi.org/10.1016/j.imu.2021.100799

Hernes, M., Tutak, P., & Siewiera, M. (2024). Prediction of Residential Real Estate Price on Primary Market Using Machine Learning. *Procedia Computer Science*, *246*(C), 3142–3147. https://doi.org/10.1016/j.procs.2024.09.358

Hutami, A. R. (2018). *Aplikasi Neural Network Untuk Prediksi Harga Rumah di Yogyakarta Menggunakan Backpropagation (Studi Kasus: Rumah di Situs Online OLX.co.id)* [Universitas Islam Indonesia]. https://dspace.uii.ac.id/handle/123456789/6536

Izonin, I., Ilchyshyn, B., Tkachenko, R., Gregus, M., Shakhovska, N., & Strauss, C. (2022). Towards Data Normalization Task for the Efficient Mining of Medical Data. *Proceedings - International Conference on Advanced Computer Information Technologies, ACIT*, (2021), 480–484. https://doi.org/10.1109/ACIT54803.2022.9913112

Jackson, E., & Agrawal, R. (2019). Performance Evaluation of Different Feature Encoding Schemes on Cybersecurity Logs. *Conference Proceedings - IEEE SOUTHEASTCON*, *2019-April*, 1–9. https://doi.org/10.1109/SoutheastCon42311.2019.9020560

Jollife, I. T., & Cadima, J. (2016). Principal Component Analysis: A Review and Recent Developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *374*(2065). https://doi.org/10.1098/rsta.2015.0202

Jonathan, B., Rahim, Z., Barzani, A., & Oktavega, W. (2019). Evaluation of Mean Absolute Error in Collaborative Filtering for Sparsity Users and Items on Female Daily Network. *Proceedings - 1st International Conference on Informatics, Multimedia, Cyber and Information System, ICIMCIS 2019*, 41–44. https://doi.org/10.1109/ICIMCIS48181.2019.8985340

Kahlert, T., & Giza, K. (2016). *Visual Studio Code Tips & Tricks Vol. 1*. Microsoft Deutschland GmbH. http://download.microsoft.com/download/8/A/4/8A48E46A-C355-4E5C-8417-E6ACD8A207D4/VisualStudioCode-TipsAndTricks-Vol.1.pdf

Kamalov, F., Moussa, S., & Reyes, J. A. (2023). Data Transformation in Machine Learning: Empirical Analysis. *2023 International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies, 3ICT 2023*, 115–120. https://doi.org/10.1109/3ICT60104.2023.10391512

Khazari, A. S., Marisa, F., & Wijaya, I. D. (2017). Sistem Rekomendasi Penentuan Judul Skripsi Menggunakan Algoritma Decision Tree. *Jurnal Teknologi dan Manajemen Informatika*, *3*(1). https://doi.org/10.26905/jtmi.v3i1.1248

Kumar, R. S., Meera, P. S., Lavanya, V., & Hemamalini, S. (2025). Brown Bear Optimized Random Forest Model for Short Term Solar Power Forecasting. *Results in Engineering*, *25*(January), 104583. https://doi.org/10.1016/j.rineng.2025.104583

Liu, T., Gao, H., & Wu, J. (2020). Review of Outlier Detection Algorithms Based on Grain Storage Temperature Data. *Proceedings of 2020 IEEE International Conference on Artificial Intelligence and Computer Applications, ICAICA 2020*, 1045–1048. https://doi.org/10.1109/ICAICA50127.2020.9182588

Majeed, W. A., & Hamad, M. M. (2023). Outlier for High Dimensional Data A Review. *Proceedings - International Conference on Developments in eSystems Engineering, DeSE*, 765–770. https://doi.org/10.1109/DeSE60595.2023.10469669

Maxwell, A. E., Warner, T. A., & Fang, F. (2018). Implementation of Machine-Learning Classification in Remote Sensing: An Applied Review. *International Journal of Remote Sensing*, *39*(9), 2784–2817. https://doi.org/10.1080/01431161.2018.1433343

McKinney, W. (2012). *Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython* (M. Beaugureau, Ed.; 2 ed.). O'Reilly Media.

Monalisa, S. (2018). Analysis Outlier Data on RFM and LRFM Models to Determining Customer Loyalty with DBSCAN Algorithm. *Proceeding - 2018 International*

Symposium on Advanced Intelligent Informatics: Revolutionize Intelligent Informatics Spectrum for Humanity, SAIN 2018, 1–5. https://doi.org/10.1109/SAIN.2018.8673380

Montes, I. (2025). Neighbourhood Models Induced by the Euclidean Distance and the Kullback-Leibler Divergence. *Fuzzy Sets and Systems*, *529*(September 2025), 109745. https://doi.org/10.1016/j.fss.2025.109745

Montgomery, D. C., Peck, E. A., & Vining, G. G. (2012). *Introduction to Linear Regression Analysis* (D. J. Balding, N. A. C. Cressie, G. M. Fitzmaurice, H. Goldstein, I. M. Johnstone, G. Molenberghs, D. W. Scott, A. F. M. Smith, R. S. Tsay, & S. Weisberg, Ed.; 5 ed.). John Wiley & Sons.

Nandipati, M., Ogunsanya, M., & Desai, S. (2024). Predictive Models for 3D Inkjet Material Printer Using Automated Image Analysis and Machine Learning Algorithms. *Manufacturing Letters*, *41*, 810–821. https://doi.org/10.1016/j.mfglet.2024.09.101

Quang, T., Minh, N., Hy, D., & Bo, M. (2020). Housing Price Prediction via Improved Machine Learning Techniques. *Procedia Computer Science*, *174*(2019), 433–442. https://doi.org/10.1016/j.procs.2020.06.111

Rajeswari, A. M., Yalini, S. K., Janani, R., Rajeswari, N., & Deisy, C. (2018). A Comparative Evaluation of Supervised and Unsupervised Methods for Detecting Outliers. *Proceedings of the International Conference on Inventive Communication and Computational Technologies, ICICCT 2018*, (Icicct), 1068–1073. https://doi.org/10.1109/ICICCT.2018.8473123

Rani, M. J., & Singh, D. (2024). A Comparative Study and Analysis of Dimensionality Reduction Techniques on High Dimensional Datasets for Network Anomaly Detection. *Proceedings - 2024 4th International Conference on Soft Computing for Security Applications, ICSCSA 2024*, 220–227. https://doi.org/10.1109/ICSCSA64454.2024.00042

Roshinta, T. A., Dinata, I. F., Riatma, D. L., Syafi'i, M. A., Firdaus, N., & A'La, F. Y. (2023). A Comparison of Prediction Algorithms in Food Sales with Different K-Folds Cross-Validation. *Proceedings - 2023 6th International Conference on Computer and Informatics Engineering: AI Trust, Risk and Security Management (AI Trism), IC2IE 2023*, 314–318. https://doi.org/10.1109/IC2IE60547.2023.10331578

Sandunil, K., Bennour, Z., Ben Mahmud, H., & Giwelli, A. (2024). Effects of Tuning Decision Trees in Random Forest Regression on Predicting Porosity of a Hydrocarbon Reservoir. A Case Study: Volve Oil Field, North Sea. *Energy Advances*, *3*(9), 2335–2347. https://doi.org/10.1039/d4ya00313f

Saneep, K., Sundareswaran, K., Srinivasa Rao Nayak, P., & Puthusserry, G. V. (2025). State of Charge Estimation of Lithium-Ion Batteries Using PSO Optimized Random Forest Algorithm and Performance Analysis. *Journal of Energy Storage*, *114*, 115879. https://doi.org/10.1016/j.est.2025.115879

Schubert, E., Sander, J., Ester, M., Kriegel, H. P., & Xu, X. (2017). DBSCAN Revisited, Revisited: Why and How You Should (Still) Use DBSCAN. *ACM Transactions on Database Systems*, *42*(3), 1–21. https://doi.org/10.1145/3068335

Sheather, S. J. (2009). A Modern Approach to Regression with R. Dalam G. Casella, S. Fienberg, & I. Olkin (Ed.), *A modern approach to regression analysis* (1 ed.). Springer New York, NY.

Shlens, J. (2014). A Tutorial on Principal Component Analysis. *arXiv*. http://arxiv.org/abs/1404.1100

Singh, D., & Singh, B. (2020). Investigating the Impact of Data Normalization on Classification Performance. *Applied Soft Computing*, *97*, 105524. https://doi.org/10.1016/j.asoc.2019.105524

Tigga, O., Pal, J., & Mustafi, D. (2023). A Comparative Study of Multiple Linear Regression and K Nearest Neighbours using Machine Learning. *2023 5th International Conference on Electrical, Computer and Communication Technologies, ICECCT 2023*, 1–5. https://doi.org/10.1109/ICECCT56650.2023.10179713

Wang, A. X., Le, V. T., Trung, H. N., & Nguyen, B. P. (2025). Addressing Imbalance in Health Data: Synthetic Minority Oversampling Using Deep Learning. *Computers in Biology and Medicine*, *188*(January), 109830. https://doi.org/10.1016/j.compbiomed.2025.109830

Wang, F., Zou, Y., Zhang, H., & Shi, H. (2019). House Price Prediction Approach based on Deep Learning and ARIMA Model. *2019 IEEE 7th International Conference on Computer Science and Network Technology (ICCSNT)*, 303–307. https://doi.org/10.1109/ICCSNT47585.2019.8962443

Wang, Z., Hong, Y., Huang, L., Zheng, M., Yuan, H., & Zeng, R. (2025). A Comprehensive Review and Future Research Directions of Ensemble Learning Models for Predicting Building Energy Consumption. *Energy and Buildings*, *335*(November 2024), 115589. https://doi.org/10.1016/j.enbuild.2025.115589

Ying, K. H., Hu, W. Y., Chen, J. B., & Li, G. N. (2021). Research on Instance-Level Data Cleaning Technology. *Proceedings - 2021 International Conference on Artificial Intelligence, Big Data and Algorithms, CAIBDA 2021*, 238–242. https://doi.org/10.1109/CAIBDA53561.2021.00057

Zou, F. (2022). Research on Data Cleaning in Big Data Environment. *Proceedings - 2022 International Conference on Cloud Computing, Big Data and Internet of Things, 3CBIT 2022*, 145–148. https://doi.org/10.1109/3CBIT57391.2022.00037