

DAFTAR PUSTAKA

- Aggrawal, C., 2015, Data Classification Algorithms and Applications, CRC Press.
- Abdoh, S.F., Abo Rizka, M. dan Maghraby, F.A., 2018, Cervical cancer diagnosis using random forest classifier with SMOTE and feature reduction techniques, *IEEE Access*, [Online] 659475–59485, tersedia di DOI:10.1109/ACCESS.2018.2874063.
- Alam, M.Z., Rahman, M.S. dan Rahman, M.S., 2019, A Random Forest based predictor for medical data classification using feature ranking, *Informatics in Medicine Unlocked*, [Online] 15 (January), 100180, tersedia di DOI:10.1016/j.imu.2019.100180.
- Amirruddin, A.D., Muharam, F.M., Ismail, M.H., Tan, N.P. dan Ismail, M.F., 2020, Hyperspectral spectroscopy and imbalance data approaches for classification of oil palm's macronutrients observed from frond 9 and 17, *Computers and Electronics in Agriculture*, [Online] 178105768, tersedia di DOI:10.1016/j.compag.2020.105768.
- Bandana, R., 2018, Sentiment analysis of movie reviews using heterogeneous features, *2018 2nd International Conference on Electronics, Materials Engineering and Nano-Technology, IEMENTech 2018*, [Online] 1–4, tersedia di DOI:10.1109/IEMENTECH.2018.8465346.
- Barandela, R., Sanchez, J. S., Garcia, V., & Rangel, E. (2003). Strategies for Learning in Class Imbalance Problem. *Pattern Recognition*, 849-851.
- Barua, S., Islam, M. dan Yao, X, 2014, 'MWMOTE — Majority Weighted Minority Oversampling Technique for Imbalanced Data Set Learning', *IEEE Transaction Knowledge Management*, 26(2), pp. 405–425.
- Bhowan, U., Johnston, M., dan Zhang, M., 2012, Developing New Fitness Functions in Genetic Programming for Classification with Unbalanced Data. In *IEEE Transactions on Systems, Man and Cybernetics-Part B*. Vol. 42. IEEE.
- Breiman, L., 2001, Random Forest, *Machine Learning*, 455-32, diakses 4 Desember 2022.
- Brownlee, J., 2020. Data Preparation for Machine Learning. San Francisco: Machine Learning Mastery.
- Chawla, N. V., Bowyer, K. W., Hall, and W.P Kegelmeyer., 2002. SMOTE: *Synthetic minority oversampling technique*, *J. Arief. Intell. Res.*, vol. 16, no.1, pp. 321-357.
- Cozma, M., Butnaru, A.M and Ionescu, R.T., 2018, Automated essay scoring with

string kernels and word embedding.

- Dikli, S., 2009, Automated Essay Scoring, *Turkish Online Journal of Distance Education*, 2, 1 Article 5, 49-62.
- Fawcett, T. and F. Pravost, 2006, An introduction to ROC analysis. *Pattern Recognition Letters*, 7(8), 861-874.
<https://doi.org/10.1016/J.PATREC.2005.10.010>
- Fernandez, A., Garcia, S., Galar, M., Prati, R.C., Krawczyk, B. dan Herrera, F., 2018, Learning from Imbalanced Data Sets, Springer, Switzerland.
- Goyal, A., Rathore, L. dan Sharma, A., 2021, SMO-RF: A machine learning approach by random forest for predicting class imbalancing followed by SMOTE, *Materials Today: Proceedings*, [Online] S2214785320406091, tersedia di DOI:10.1016/j.matpr.2020.12.891.
- Guo, X. (2016). On the Class Imbalanced Problem. Doi:10.1109/ICNC.2008.871.
- Harrington, P., 2012, *Machine Learning in Action*, New York: Manning Publications Co.
- Hayatin, N., Fatichah, C. & Purwitasari, D., 2015, Trending issue untuk Peningkatan Multi Dokumen, *Jurnal Ilmiah TEKnologi Informasi (JUTI)*, 13,1, 38-44.
- Hastie, T., Tibshirani, R. dan Friedman, J., 2017, *The Elements of Statistical Learning*, Second, Springer US.
- He, H., & Ma, Y. (2013). Imbalanced Learning: Foundations, Algorithms, and Applications. Canada: Wiley.
- Herwanto, G.B., Sari, Y., Prastowo, B.N., Bustoni, I.A. dan Hidayatulloh, I., 2018, UKARA: A Fast and Simple Automatic Short Answer Scoring System for Bahasa Indonesia, *Iceap 2018*, 2 (2), 48–53,
- Li, H., & Sun, J. (2012). Forecasting business failure: The use of nearest-neighbour support vectors and correcting imbalanced samples—Evidence from the Chinese hotel industry. *Tourism Management*, 33(3), 622-634.
- L. Breiman., 2014, Random Forests, *Machine learning*, vol. 45, no. 1, pp. 5-32.
- More, A.S. dan Rana, D.P., 2020, An Experimental Assessment of Random Forest Classification Performance Improvisation with Sampling and Stage Wise Success Rate Calculation, *Procedia Computer Science*, [Online] 1671711–1721, tersedia di DOI:10.1016/j.procs.2020.03.381.
- M. Allahyari., 2017, A Brief Survey of Text Mining: Classification, Clustering and Extraction Techniques, *KDD Bigdas*.

- Ngurah, G., Nata, M. & Yudiasra, P.P., 2017, Preprocessing Text Mining Pada Email Box Bahasa Indonesia, In, *Konferensi Nasional Sistem & Informatika 2017*, STMIK STIKOM, Bali, pp. 479-483.
- Pennington, J., Socher, R. dan Manning, C.D., 2014, GloVe: Global Vectors for Word Representation, *Empirical Methods in Natural Language Processing (EMNLP)*, [Online], 2014 hlm. 1532–1543, tersedia di <http://www.aclweb.org/anthology/D14-1162>.
- Polat, K., 2019, A Hybrid Approach to Parkinson Disease Classification Using Speech Signal: The Combination of SMOTE and Random Forests, *2019 Scientific Meeting on Electrical-Electronics & Biomedical Engineering and Computer Science (EBBT)*, [Online], April 2019 IEEE, Istanbul, Turkey., hlm. 1–3, tersedia di DOI:10.1109/EBBT.2019.8741725, diakses 20 September 2022.
- Rajesh, K.N.V.P.S. dan Dhuli, R., 2018, Classification of imbalanced ECG beats using re-sampling techniques and AdaBoost ensemble classifier, *Biomedical Signal Processing and Control*, [Online] 41242–254, tersedia di DOI:10.1016/j.bspc.2017.12.004.
- Riassetiawan, M., Prastowo, B.N. dan Novindasari, I., 2018, SISTEM SKORING OTOMATIS UNTUK DATA JAWABAN ESAI DENGAN MENGGUNAKAN PENDEKATAN KOMPUTASI: CLUSTERING DAN CONVOLUTIONAL NEURAL Automatic Scoring System for Essay Answer Data Using Computational Approach :, *PROSIDING Ist National Conference on Educational Assessment and Policy (NCEAP 2018)*, (Nceap), 89–96,.
- Sakaguchi, K., Heilman, M. and Madnani, N., 2015, Effective Feature Integration for Automated Short Answer Scoring, *Conference of the North American Chapter of the Association for Computational Linguistic: Human Language Technologies*, 1049-1054.
- Saputra, I.P.G.H., 2017, Peringkasan Teks Otomatis Untuk Dokumen Bahasa Bali Berbasis Metode Ekstraktif, *Jurnal Ilmu Komputer*, X, 1, 33-38.
- Sulistyo, M.E., Saptono, R. & Asshidiq, A., 2015, Penilaian Ujian Bertipe Essay Menggunakan Metode Text Similarity, *Telematika*, 12,2, 146-158.
- Sun, J., Li, H., Fujita, H., Fu, B. dan Ai, W., 2020, Class-imbalanced dynamic financial distress prediction based on Adaboost-SVM ensemble combined with SMOTE and time weighting, *Information Fusion*, [Online] 54 (July 2019), 128–144, tersedia di DOI:10.1016/j.inffus.2019.07.006.
- Tan, P.N., Steinbach, M. dan Kumar, V., 2014, *Introduction to Data Mining*, Pearson, London.



- Valenti, S., Neri, F. and Cucchiarelli, A., 2017, An Overview of Current Research on Automated Essay Grading, *Journal of Information Technology Education: Research*, 2, 319-330.
- Wyner, A. J., 2017, Explaining the success of adaboost and random forests as interpolating classifiers, *Journal of Machine Learning Research*, 18, pp. 1–33.