DAFTAR PUSTAKA

AlShboul, R. *et al.* (2018) 'A visualization cybersecurity method based on features' dissimilarity', *Computers and Security*, 77. doi: 10.1016/j.cose.2018.04.007.

Appiahene, P., Missah, Y. M. and Najim, U. (2020) 'Predicting Bank Operational Efficiency Using Machine Learning Algorithm: Comparative Study of Decision Tree, Random Forest, and Neural Networks', *Advances in Fuzzy Systems*, 2020. doi: 10.1155/2020/8581202.

Athapaththu, J. C. and Kulathunga, K. M. S. D. (2018) 'Factors affecting online purchase intention: A study of Sri Lankan online customers', *International Journal of Scientific and Technology Research*, 7(9).

Balasaraswathi, V. R., Sugumaran, M. and Hamid, Y. (2017) 'Feature selection techniques for intrusion detection using non-bio-inspired and bio-inspired optimization algorithms', *Journal of Communications and Information Networks*, 2(4). doi: 10.1007/s41650-017-0033-7.

Breiman, L. *et al.* (2017) *Classification and regression trees*, *CRC Press*. doi: 10.1201/9781315139470.

Buulolo, E. (2020) *Data Mining untuk Perguruan Tinggi*. Yogyakarta: Deepublish.

Chen, L. *et al.* (2021) 'Adaptive feature selection-based AdaBoost-KNN with direct optimization for dynamic emotion recognition in human-robot interaction', *IEEE Transactions on Emerging Topics in Computational Intelligence*. Institute of Electrical and Electronics Engineers Inc., 5(2), pp. 205–213. doi: 10.1109/TETCI.2019.2909930.

Chicco, D. and Jurman, G. (2020) 'The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation', *BMC Genomics*, 21(1). doi: 10.1186/s12864-019-6413-7.

Chouat, O. and Irawan, A. H. (2018) 'Implementation of Data Mining on Online Shop in Indonesia', in *IOP Conference Series: Materials Science and Engineering*. doi: 10.1088/1757-899X/407/1/012013.

Çiğşar, B. and Ünal, D. (2019) 'Comparison of Data Mining Classification Algorithms Determining the Default Risk', *Scientific Programming*, 2019. doi: 10.1155/2019/8706505.

Dipto, I. C. *et al.* (2020) 'Comparison of Different Machine Learning Algorithms for the Prediction of Coronary Artery Disease', *Journal of Data Analysis and Information Processing*, 08(02). doi: 10.4236/jdaip.2020.82003.

Fernández, A. *et al.* (2017) 'An insight into imbalanced Big Data classification: outcomes and challenges', *Complex & Intelligent Systems*, 3(2). doi:

10.1007/s40747-017-0037-9.

Gao, K., Khoshgoftaar, T. and Wald, R. (2014) 'Combining feature selection and ensemble learning for software quality estimation', in *Proceedings of the 27th International Florida Artificial Intelligence Research Society Conference (FLAIRS)*.

Goodfellow, I., Bengio, Y. and Courville, A. (2016) *Deep Learning*. Cambridge, MA, USA: The MIT Press.

Haixiang, G. *et al.* (2017) 'Learning from class-imbalanced data: Review of methods and applications', *Expert Systems with Applications*, 73, pp. 220–239. doi: 10.1016/j.eswa.2016.12.035.

Hao, R. *et al.* (2020) 'Bank direct marketing analysis based on ensemble learning', in *Journal of Physics: Conference Series*. doi: 10.1088/1742-6596/1627/1/012026.

Japkowicz, N. (2013) 'Assessment metrics for imbalanced learning', in Wiley-IEEE Press (ed.) *Imbalanced Learning: Foundations, Algorithms, and Applications*. 1st edn, pp. 187–206. doi: 10.1002/9781118646106.ch8.

Johnson, J. M. and Khoshgoftaar, T. M. (2019) 'Survey on deep learning with class imbalance', *Journal of Big Data*, 6(1). doi: 10.1186/s40537-019-0192-5.

Joloudari, J. H. *et al.* (2020) 'Early detection of the advanced persistent threat attack using performance analysis of deep learning', *IEEE Access*, 8. doi: 10.1109/ACCESS.2020.3029202.

Kabir, M. R., Ashraf, F. Bin and Ajwad, R. (2019) 'Analysis of different predicting model for online shoppers' purchase intention from empirical data', *2019 22nd International Conference on Computer and Information Technology (ICCIT)*, (March 2020). doi: 10.1109/ICCIT48885.2019.9038521.

Kamalov, F. and Thabtah, F. (2017) 'A Feature Selection Method Based on Ranked Vector Scores of Features for Classification', *Annals of Data Science*, 4(4). doi: 10.1007/s40745-017-0116-1.

Kisworini, R. Y. and Setiawan, M. A. (2020) 'Peningkatan Performa Naive Bayes Dengan Seleksi Atribut Menggunakan Chi Square Untuk Klasifikasi Loyalitas Pelanggan GRAB', *Journal of Informatics, Information System, Software Engineering and Applications (INISTA)*, 2(2). doi: 10.20895/inista.v2i2.127.

Leevy, J. L. *et al.* (2018) 'A survey on addressing high-class imbalance in big data', *Journal of Big Data*, 5(1). doi: 10.1186/s40537-018-0151-6.

Li, J. *et al.* (2017) 'Feature selection: A data perspective', *ACM Computing Surveys*, 50(6). doi: 10.1145/3136625.

Mazini, M., Shirazi, B. and Mahdavi, I. (2019) 'Anomaly network-based intrusion detection system using a reliable hybrid artificial bee colony and AdaBoost algorithms', *Journal of King Saud University - Computer and Information Sciences*, 31(4). doi: 10.1016/j.jksuci.2018.03.011.

Mulyati, S., Yulianti, Y. and Saifudin, A. (2017) 'Penerapan Resampling dan Adaboost untuk Penanganan Masalah Ketidakseimbangan Kelas Berbasis Naïve Bayes pada Prediksi Churn Pelanggan', *Jurnal Informatika Universitas Pamulang*, 2(4). doi: 10.32493/informatika.v2i4.1440.

Nurmasani, A. and Pristyanto, Y. (2021) 'Algoritme Stacking untuk Klasifikasi Penyakit Jantung pada Dataset Imbalanced Class', *Pseudocode*, 8(1). doi: 10.33369/pseudocode.8.1.21-26.

Online Shoppers Purchasing Intention Dataset (2018) *UCI Machine Learning Repository*. Available at: https://archive.ics.uci.edu/ml/datasets/Online+Shoppers+Purchasing+Intention+D ataset.

Putra, F. M. and Sitanggang, I. S. (2020) 'Classification model of air quality in Jakarta using decision tree algorithm based on air pollutant standard index', in *IOP Conference Series: Earth and Environmental Science*. doi: 10.1088/1755-1315/528/1/012053.

Quinlan, J. R. (1993) *C4.5: Programs for Machine Learning*. Morgan Kaufmann.

Rajeswari, S. and Suthendran, K. (2019) 'C5.0: Advanced Decision Tree (ADT) classification model for agricultural data analysis on cloud', *Computers and Electronics in Agriculture*, 156, pp. 530–539. doi: 10.1016/j.compag.2018.12.013.

Ross Quinlan (2017) *Is See5/C5.0 Better Than C4.5?*, *RuleQuest Research*. Available at: https://rulequest.com/see5-comparison.html#:~:text=Decision trees%3A faster%2C smaller&text=0 produce trees with similar,are noticeably smaller and C5.

Rustam, Z., Nadhifa, F. and Acar, M. (2018) 'Comparison of SVM and FSVM for predicting bank failures using chi-square feature selection', in *Journal of Physics: Conference Series*. doi: 10.1088/1742-6596/1108/1/012115.

Sakar, C. O. *et al.* (2019) 'Real-time prediction of online shoppers' purchasing intention using multilayer perceptron and LSTM recurrent neural networks', *Neural Computing and Applications*. Springer London, 31(10), pp. 6893–6908. doi: 10.1007/s00521-018-3523-0.

Santoso, N., Wibowo, W. and Himawati, H. (2019) 'Integration of synthetic minority oversampling technique for imbalanced class', *Indonesian Journal of Electrical Engineering and Computer Science*. doi: 10.11591/ijeecs.v13.i1.pp102-108.

Sarkar, S., Verma, A. and Maiti, J. (2018) 'Prediction of Occupational Incidents Using Proactive and Reactive Data: A Data Mining Approach', in *Industrial Safety Management*. doi: 10.1007/978-981-10-6328-2_6.

Seiffert, C. *et al.* (2008) 'Resampling or reweighting: A comparison of boosting implementations', in *Proceedings - International Conference on Tools with Artificial Intelligence, ICTAI*. doi: 10.1109/ICTAI.2008.59.

65

Shahraki, A., Abbasi, M. and Haugen, Ø. (2020) 'Boosting algorithms for network intrusion detection: A comparative evaluation of Real AdaBoost, Gentle AdaBoost and Modest AdaBoost', *Engineering Applications of Artificial Intelligence*, 94. doi: 10.1016/j.engappai.2020.103770.

Sudarto, Zarlis, M. and Sirait, P. (2016) 'Integrasi Density Based Feature Selection dan Adaptive Boosting (Adaboost) dalam Menangani Ketidakseimbangan Kelas', *JSM (Jurnal SIFO Mikroskil)*, 17(2), pp. 193–206.

Sulistiani, H. and Tjahyanto, A. (2017) 'Comparative Analysis of Feature Selection Method to Predict Customer Loyalty', *IPTEK Journal of Engineering*, 3(1). doi: 10.12962/joe.v3i1.2257.

Sun, Y. *et al.* (2007) 'Cost-sensitive boosting for classification of imbalanced data', *Pattern Recognition*, 40(12), pp. 3358–3378. doi: 10.1016/j.patcog.2007.04.009.

Sun, Y. and Liu, F. (2017) 'SMOTE-NCL: A re-sampling method with filter for network intrusion detection', in *2016 2nd IEEE International Conference on Computer and Communications (ICCC)*, pp. 1157–1161. doi: 10.1109/CompComm.2016.7924886.

Thakkar, A. and Lohiya, R. (2021) 'Attack classification using feature selection techniques: a comparative study', *Journal of Ambient Intelligence and Humanized Computing*, 12(1). doi: 10.1007/s12652-020-02167-9.

Thaseen, I. S., Kumar, C. A. and Ahmad, A. (2019) 'Integrated Intrusion Detection Model Using Chi-Square Feature Selection and Ensemble of Classifiers', *Arabian Journal for Science and Engineering*, 44(4). doi: 10.1007/s13369-018-3507-5.

Trivedi, S. K. (2020) 'A study on credit scoring modeling with different feature selection and machine learning approaches', *Technology in Society*, 63. doi: 10.1016/j.techsoc.2020.101413.

Tseng, C. J. *et al.* (2017) 'Integration of data mining classification techniques and ensemble learning to identify risk factors and diagnose ovarian cancer recurrence', *Artificial Intelligence in Medicine*, 78. doi: 10.1016/j.artmed.2017.06.003.

Vulandari, R. T. (2017) *Data Mining Teori dan Aplikasi Rapidminer*. 1st edn. Yogyakarta: Penerbit Gava Media.

Wagner, D., Chaipoopirutana, S. and Combs, H. (2019) 'A Study of Factors Influencing the Online Purchasing Intention toward Online Shopping in Thailand', *AtMA 2019 Proccedings*, (2013), pp. 277–292.

Wang, G. and Wu, N. (2020) 'A Comparative Study on Contract Recommendation Model: Using Macao Mobile Phone Datasets', *IEEE Access*, 8. doi: 10.1109/ACCESS.2020.2975029.
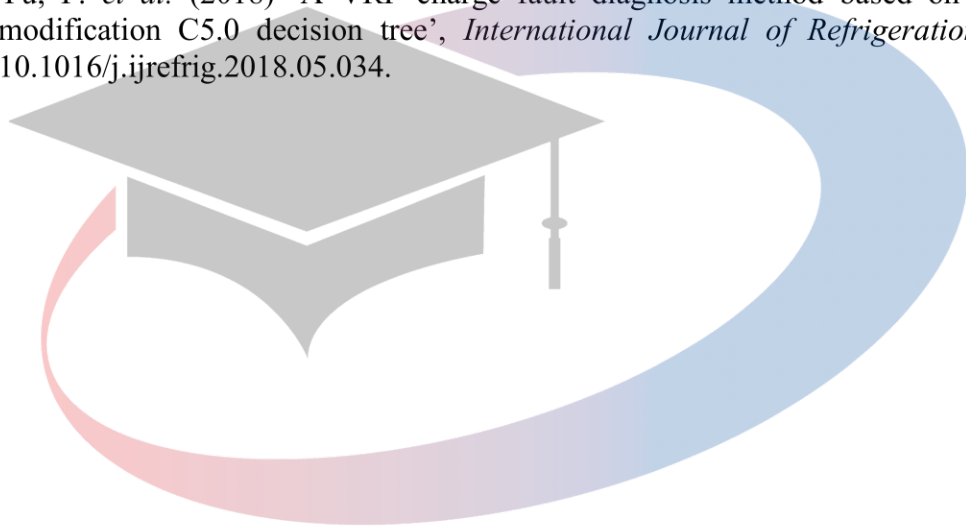
Wang, J. *et al.* (2019) 'An ensemble feature selection method for high-dimensional data based on sort aggregation', *Systems Science and Control Engineering*, 7(2). doi: 10.1080/21642583.2019.1620658.

Wu, X. *et al.* (2008) 'Top 10 algorithms in data mining', *Knowledge and Information Systems*, 14(1), pp. 1–37. doi: 10.1007/s10115-007-0114-2.

Yap, B. W. *et al.* (2014) 'An application of oversampling, undersampling, bagging and boosting in handling imbalanced datasets', in *Lecture Notes in Electrical Engineering*, pp. 13–22. doi: 10.1007/978-981-4585-18-7_2.

Yildirim, P. (2016) 'Pattern Classification with Imbalanced and Multiclass Data for the Prediction of Albendazole Adverse Event Outcomes', in *Procedia Computer Science*. doi: 10.1016/j.procs.2016.04.216.

Yu, F. *et al.* (2018) 'A VRF charge fault diagnosis method based on expert modification C5.0 decision tree', *International Journal of Refrigeration*. doi: 10.1016/j.ijrefrig.2018.05.034.