

## DAFTAR PUSTAKA

- [1] S. Dhankhad, E. A. Mohammed, and B. Far, "Supervised machine learning algorithms for credit card fraudulent transaction detection: A comparative study," *Proc. - 2018 IEEE 19th Int. Conf. Inf. Reuse Integr. Data Sci. IRI 2018*, pp. 122–125, 2018, doi: 10.1109/IRI.2018.00025.
- [2] R. A. Sowah *et al.*, "Decision Support System (DSS) for Fraud Detection in Health Insurance Claims Using Genetic Support Vector Machines (GSVMs)," *J. Eng. (United Kingdom)*, vol. 2019, 2019, doi: 10.1155/2019/1432597.
- [3] J. Li, Q. Lan, E. Zhu, Y. Xu, and D. Zhu, "A Study of Health Insurance Fraud in China and Recommendations for Fraud Detection and Prevention," *J. Organ. End User Comput.*, vol. 34, no. 4, pp. 1–19, 2022, doi: 10.4018/joeuc.301271.
- [4] J. Brownlee, *Imbalanced Classification with Python*. 2020.
- [5] H. He and Y. Ma, *He - Imbalanced Learning - 2014*. 2013.
- [6] L. Wang, Z. Zhang, X. Zhang, X. Zhou, P. Wang, and Y. Zheng, *A Deep-forest based approach for detecting fraudulent online transaction*, 1st ed., vol. 120. Elsevier Inc., 2021. doi: 10.1016/bs.adcom.2020.10.001.
- [7] J. O. Sinayobye, F. Kiwanuka, and S. Kaawaase Kyanda, "A state-of-the-art review of machine learning techniques for fraud detection research," *Proc. - Int. Conf. Softw. Eng.*, pp. 11–19, 2018, doi: 10.1145/3195528.3195534.
- [8] K. G. Al-Hashedi and P. Magalingam, "Financial fraud detection applying data mining techniques: A comprehensive review from 2009 to 2019," *Comput. Sci. Rev.*, vol. 40, 2021, doi: 10.1016/j.cosrev.2021.100402.
- [9] G. K. Kulatilleke, "Challenges and Complexities in Machine Learning based Credit Card Fraud Detection," pp. 1–17, 2022, [Online]. Available: <http://arxiv.org/abs/2208.10943>
- [10] R. A. Bauder and T. M. Khoshgoftaar, "Medicare fraud detection using random forest with class imbalanced big data," *Proc. - 2018 IEEE 19th Int. Conf. Inf. Reuse Integr. Data Sci. IRI 2018*, pp. 80–87, 2018, doi: 10.1109/IRI.2018.00019.
- [11] F. Thabtah, S. Hammoud, and F. Kamalov, "Data Imbalance in Classification : Experimental Evaluation," *Inf. Sci. (Ny)*, 2019, doi: 10.1016/j.ins.2019.11.004.
- [12] D. Devi, "A Review on Solution to Class Imbalance Problem : Undersampling Approaches," pp. 626–631, 2020.
- [13] M. R. P. Kumar and P. Jayagopal, "A preprocessing method combined with an ensemble framework for the multiclass imbalanced data classification," *Int. J. Comput. Appl.*, vol. 0, no. 0, pp. 1–8, 2019, doi: 10.1080/1206212X.2019.1700335.
- [14] R. Mohammed, J. Rawashdeh, and M. Abdullah, "Machine Learning with Oversampling and Undersampling Techniques : Overview Study and Experimental Results," pp. 243–248, 2020, doi: 10.1109/ICICS49469.2020.239556.
- [15] S. K. Shamitha and V. Ilango, "A time-efficient model for detecting fraudulent health insurance claims using Artificial neural networks," *2020 Int. Conf. Syst. Comput. Autom. Networking, ICSCAN 2020*, 2020, doi: 10.1109/ICSCAN49426.2020.9262298.
- [16] Y. Pristyanto, A. F. Nugraha, I. Pratama, A. Dahlan, and L. A. Wirasakti, "Dual Approach to Handling Imbalanced Class in Datasets Using Oversampling and Ensemble Learning Techniques," *Proc. 2021 15th Int. Conf. Ubiquitous Inf. Manag. Commun. IMCOM 2021*, 2021, doi: 10.1109/IMCOM51814.2021.9377420.
- [17] H. Faris *et al.*, "Improving financial bankruptcy prediction in a highly imbalanced class distribution using oversampling and ensemble learning: a case from the Spanish market," *Prog. Artif. Intell.*, vol. 9, no. 1, pp. 31–53, 2020, doi: 10.1007/s13748-019-00197-9.
- [18] P. Kumar, R. Bhatnagar, K. Gaur, and A. Bhatnagar, "Classification of Imbalanced

- Data: Review of Methods and Applications," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 1099, no. 1, p. 012077, 2021, doi: 10.1088/1757-899x/1099/1/012077.
- [19] S. S. Kotekani and I. Velchamy, "An Effective Data Sampling Procedure for Imbalanced Data Learning on Health Insurance Fraud Detection," *J. Comput. Inf. Technol.*, vol. 28, no. 4, pp. 269–285, 2020, doi: 10.20532/cit.2020.1005216.
- [20] P. Mrozek, J. Panneerselvam, and O. Bagdasar, "Efficient resampling for fraud detection during anonymised credit card transactions with unbalanced datasets," *Proc. - 2020 IEEE/ACM 13th Int. Conf. Util. Cloud Comput. UCC 2020*, pp. 426–433, 2020, doi: 10.1109/UCC48980.2020.00067.
- [21] Z. S. Rubaidi, B. Ben Ammar, and M. Ben Aouicha, "Fraud Detection Using Large-scale Imbalance Dataset," *Int. J. Artif. Intell. Tools*, vol. 2250037, 2022, doi: 10.1142/s0218213022500373.
- [22] J. M. Johnson and T. M. Khoshgoftaar, "The Effects of Data Sampling with Deep Learning and Highly Imbalanced Big Data," *Inf. Syst. Front.*, vol. 22, no. 5, pp. 1113–1131, 2020, doi: 10.1007/s10796-020-10022-7.
- [23] C. A. Ramezan, T. A. Warner, and A. E. Maxwell, "Evaluation of Sampling and Cross-Validation Tuning Strategies for Regional-Scale Machine Learning Classification," 2019, doi: 10.3390/rs11020185.
- [24] S. K. Mukhiya and U. Ahmed, "Hands-On Exploratory Data Analysis with Python,"
- [25] J. Gong, H. Zhang, and W. Du, "Research on integrated learning fraud detection method based on combination classifier fusion (thbagging): A case study on the foundational medical insurance dataset," *Electron.*, vol. 9, no. 6, 2020, doi: 10.3390/electronics9060894.
- [26] T. Nyitrai and V. Miklós, "Socio-Economic Planning Sciences The effects of handling outliers on the performance of bankruptcy prediction models," *Socioecon. Plann. Sci.*, no. August 2017, pp. 0–1, 2018, doi: 10.1016/j.seps.2018.08.004.
- [27] X. Zhu, "Sample size calculation for Mann-Whitney U test with five methods," vol. 8, no. 3, pp. 184–195, 2021.
- [28] C. Deckert, A. G. Equivalence, and S. Ges, "Investigating the effect of binning on causal discovery," pp. 2574–2581, 2019.
- [29] R. A. Bauder and T. M. Khoshgoftaar, "The effects of varying class distribution on learner behavior for medicare fraud detection with imbalanced big data," *Heal. Inf. Sci. Syst.*, vol. 6, no. 1, 2018, doi: 10.1007/s13755-018-0051-3.
- [30] H. Wang and X. Liu, "Undersampling bankruptcy prediction: Taiwan bankruptcy data," *PLoS One*, vol. 16, no. 7 July, pp. 1–17, 2021, doi: 10.1371/journal.pone.0254030.
- [31] T. Hasanin, T. M. Khoshgoftaar, J. Leevy, and N. Seliya, "Investigating Random Undersampling and Feature Selection on Bioinformatics Big Data," *2019 IEEE Fifth Int. Conf. Big Data Comput. Serv. Appl.*, pp. 346–356, 2019, doi: 10.1109/BigDataService.2019.00063.
- [32] N. A. Verdikha, T. B. Adji, and A. E. Permanasari, "Study of Undersampling Method: Instance Hardness Threshold with Various Estimators for Hate Speech Classification," *IJITEE (International J. Inf. Technol. Electr. Eng.)*, vol. 2, no. 2, 2018, doi: 10.22146/ijitee.42152.
- [33] A. Kabir, C. Ruiz, and S. A. Alvarez, "Mixed Bagging: A Novel Ensemble Learning Framework for Supervised Classification Based on Instance Hardness," *Proc. - IEEE Int. Conf. Data Mining, ICDM*, vol. 2018-Novem, pp. 1073–1078, 2018, doi: 10.1109/ICDM.2018.00137.
- [34] L. K. Xin, "Prediction of Depression among Women Using Random Oversampling and Random Forest," 2021.
- [35] G. Kovács, "An empirical comparison and evaluation of minority oversampling techniques on a large number of imbalanced datasets," *Appl. Soft Comput. J.*, vol. 83, p. 105662, 2019, doi:

10.1016/j.asoc.2019.105662.

- [36] H. Mengash, *Advances in Data Science , Cyber Security and IT Applications*. 2019. doi: 10.1007/978-3-030-36365-9.
- [37] C. V. Priscilla, "Influence of Optimizing XGBoost to handle Class Imbalance in Credit Card Fraud Detection," no. Icssit 2020, pp. 1309–1315, 2023.
- [38] S. He, B. Li, H. Peng, J. U. N. Xin, and E. Zhang, "An Effective Cost-Sensitive XGBoost Method for Malicious URLs Detection in Imbalanced Dataset," *IEEE Access*, vol. 9, pp. 93089–93096, 2021, doi: 10.1109/ACCESS.2021.3093094.
- [39] O. Alshboul, A. Shehadeh, G. Almasabha, and A. S. Almuflih, "Extreme Gradient Boosting-Based Machine Learning Approach for Green Building Cost Prediction," *Sustain.*, vol. 14, no. 11, 2022, doi: 10.3390/su14116651.
- [40] A. Luque, A. Carrasco, A. Martín, and A. De, "The impact of class imbalance in classification performance metrics based on the binary confusion matrix," *Pattern Recognit.*, vol. 91, pp. 216–231, 2019, doi: 10.1016/j.patcog.2019.02.023.

