

DAFTAR PUSTAKA

- [1] F. Aftab *et al.*, “A Comprehensive Survey on Sentiment Analysis Techniques,” *IJTech*, vol. 14, no. 6, p. 1288, Oct. 2023, doi: 10.14716/ijtech.v14i6.6632.
- [2] P. Baid, A. Gupta, and N. Chaplot, “Sentiment Analysis of Movie Reviews using Machine Learning Techniques,” *IJCA*, vol. 179, no. 7, pp. 45–49, Dec. 2017, doi: 10.5120/ijca2017916005.
- [3] F. Y. Pamuji and V. P. Ramadhan, “Komparasi Algoritma Random Forest dan Decision Tree untuk Memprediksi Keberhasilan Immunotherapy,” *JTMI*, vol. 7, no. 1, pp. 46–50, Jul. 2021, doi: 10.26905/jtmi.v7i1.5982.
- [4] V. P. Ramadhan and F. Y. Pamuji, “Analisis Perbandingan Algoritma Forecasting dalam Prediksi Harga Saham LQ45 PT Bank Mandiri Sekuritas (BMRI),” *JTMI*, vol. 8, no. 1, pp. 39–45, Jul. 2022, doi: 10.26905/jtmi.v8i1.6092.
- [5] A. Fernandez, S. Garcia, F. Herrera, and N. V. Chawla, “SMOTE for Learning from Imbalanced Data: Progress and Challenges, Marking the 15-year Anniversary,” *jair*, vol. 61, pp. 863–905, Apr. 2018, doi: 10.1613/jair.1.11192.
- [6] W.-C. Lin, C.-F. Tsai, Y.-H. Hu, and J.-S. Jhang, “Clustering-based undersampling in class-imbalanced data,” *Information Sciences*, vol. 409–410, pp. 17–26, Oct. 2017, doi: 10.1016/j.ins.2017.05.008.
- [7] W. Zhang, R. Ramezani, and A. Naeim, “WOTBoost: Weighted Oversampling Technique in Boosting for imbalanced learning,” in *2019 IEEE International Conference on Big Data (Big Data)*, Los Angeles, CA, USA: IEEE, Dec. 2019, pp. 2523–2531. doi: 10.1109/BigData47090.2019.9006091.
- [8] T. Zhu, Y. Lin, and Y. Liu, “Synthetic minority oversampling technique for multiclass imbalance problems,” *Pattern Recognition*, vol. 72, pp. 327–340, Dec. 2017, doi: 10.1016/j.patcog.2017.07.024.
- [9] J. Wei and K. Zou, “EDA: Easy Data Augmentation Techniques for Boosting Performance on Text Classification Tasks,” in *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, Hong Kong, China: Association for Computational Linguistics, 2019, pp. 6381–6387. doi: 10.18653/v1/D19-1670.
- [10] H. Wen and J. Zhao, “Sentiment Analysis of Imbalanced Comment Texts Under the Framework of BiLSTM,” in *2023 6th International Conference on Artificial Intelligence and Big Data (ICAIBD)*, Chengdu, China: IEEE, May 2023, pp. 312–319. doi: 10.1109/ICAIBD57115.2023.10206154.
- [11] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, “SMOTE: Synthetic Minority Over-sampling Technique,” 2011, doi: 10.48550/ARXIV.1106.1813.
- [12] Haibo He, Yang Bai, E. A. Garcia, and Shutao Li, “ADASYN: Adaptive synthetic sampling approach for imbalanced learning,” in *2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress*

- on Computational Intelligence*), Hong Kong, China: IEEE, Jun. 2008, pp. 1322–1328. doi: 10.1109/IJCNN.2008.4633969.
- [13] M. Wongkar and A. Angdressey, “Sentiment Analysis Using Naive Bayes Algorithm Of The Data Crawler: Twitter,” in *2019 Fourth International Conference on Informatics and Computing (ICIC)*, Semarang, Indonesia: IEEE, Oct. 2019, pp. 1–5. doi: 10.1109/ICIC47613.2019.8985884.
- [14] T. Widiyaningtyas, I. A. Elbaith Zaeni, and R. A. Farisi, “Sentiment Analysis Of Hotel Review Using N-Gram And Naive Bayes Methods,” in *2019 Fourth International Conference on Informatics and Computing (ICIC)*, Semarang, Indonesia: IEEE, Oct. 2019, pp. 1–5. doi: 10.1109/ICIC47613.2019.8985946.
- [15] D. U. Suhendra, J. Jondri, and I. Indwiarti, “Sentiment Analysis of Hate Speech on Twitter Public Figures with AdaBoost and XGBoost Methods,” *mib*, vol. 6, no. 3, p. 1484, Jul. 2022, doi: 10.30865/mib.v6i3.4394.
- [16] D. V. Ramadhanti, R. Santoso, and T. Widiharih, “PERBANDINGAN SMOTE DAN ADASYN PADA DATA IMBALANCE UNTUK KLASIFIKASI RUMAH TANGGA MISKIN DI KABUPATEN TEMANGGUNG DENGAN ALGORITMA K-NEAREST NEIGHBOR,” *J.Gauss*, vol. 11, no. 4, pp. 499–505, Feb. 2023, doi: 10.14710/j.gauss.11.4.499-505.
- [17] K. Afifah, I. N. Yulita, and I. Sarathan, “Sentiment Analysis on Telemedicine App Reviews using XGBoost Classifier,” in *2021 International Conference on Artificial Intelligence and Big Data Analytics*, Bandung, Indonesia: IEEE, Oct. 2021, pp. 22–27. doi: 10.1109/ICAIBDA53487.2021.9689762.
- [18] B. Verma and R. S. Thakur, “Sentiment Analysis Using Lexicon and Machine Learning-Based Approaches: A Survey,” in *Proceedings of International Conference on Recent Advancement on Computer and Communication*, vol. 34, B. Tiwari, V. Tiwari, K. C. Das, D. K. Mishra, and J. C. Bansal, Eds., in *Lecture Notes in Networks and Systems*, vol. 34, Singapore: Springer Singapore, 2018, pp. 441–447. doi: 10.1007/978-981-10-8198-9_46.
- [19] M. Wankhade, A. C. S. Rao, and C. Kulkarni, “A survey on sentiment analysis methods, applications, and challenges,” *Artif Intell Rev*, vol. 55, no. 7, pp. 5731–5780, Oct. 2022, doi: 10.1007/s10462-022-10144-1.
- [20] W. Medhat, A. Hassan, and H. Korashy, “Sentiment analysis algorithms and applications: A survey,” *Ain Shams Engineering Journal*, vol. 5, no. 4, pp. 1093–1113, Dec. 2014, doi: 10.1016/j.asej.2014.04.011.
- [21] H. Zhang, “The Optimality of Naive Bayes,” in *Proceedings of the Seventeenth International Florida Artificial Intelligence Research Society Conference (FLAIRS 2004)*, Florida, USA: The AAAI Press.
- [22] A. Rahman and Md. S. Hossen, “Sentiment Analysis on Movie Review Data Using Machine Learning Approach,” in *2019 International Conference on Bangla Speech and Language Processing (ICBSLP)*, Sylhet, Bangladesh: IEEE, Sep. 2019, pp. 1–4. doi: 10.1109/ICBSLP47725.2019.201470.
- [23] M. F. Fakhrezi, Adian Fatchur Rochim, and Dinar Mutiara Kusomo Nugraheni, “Comparison of Sentiment Analysis Methods Based on Accuracy Value Case Study: Twitter Mentions of Academic Article,” *J. RESTI*

- (*Rekayasa Sist. Teknol. Inf.*), vol. 7, no. 1, pp. 161–167, Feb. 2023, doi: 10.29207/resti.v7i1.4767.
- [24] T. M. Barros, P. A. Souza Neto, I. Silva, and L. A. Guedes, “Predictive Models for Imbalanced Data: A School Dropout Perspective,” *Education Sciences*, vol. 9, no. 4, p. 275, Nov. 2019, doi: 10.3390/educsci9040275.
 - [25] K. Davagdorj, J. S. Lee, V. H. Pham, and K. H. Ryu, “A Comparative Analysis of Machine Learning Methods for Class Imbalance in a Smoking Cessation Intervention,” *Applied Sciences*, vol. 10, no. 9, p. 3307, May 2020, doi: 10.3390/app10093307.
 - [26] B. D. Boe, “Use Cases for Unstructured Data”, [Online]. Available: <https://www.odcms.org/2014/08/cases-unstructured-data/>
 - [27] D. A. Maluf and P. B. Tran, “Managing Unstructured Data With Structured Legacy Systems,” in *2008 IEEE Aerospace Conference*, Big Sky, MT, USA: IEEE, Mar. 2008, pp. 1–5. doi: 10.1109/AERO.2008.4526666.
 - [28] D. Khurana, A. Koli, K. Khatter, and S. Singh, “Natural language processing: state of the art, current trends and challenges,” *Multimed Tools Appl*, vol. 82, no. 3, pp. 3713–3744, Jan. 2023, doi: 10.1007/s11042-022-13428-4.
 - [29] S. Angra and S. Ahuja, “Machine learning and its applications: A review,” in *2017 International Conference on Big Data Analytics and Computational Intelligence (ICBDAC)*, Chirala, Andhra Pradesh, India: IEEE, Mar. 2017, pp. 57–60. doi: 10.1109/ICBDACI.2017.8070809.
 - [30] J. Brownlee, *Master Machine Learning Algorithms*, 1.1. 2016.
 - [31] “Machine learning: the power and promise of computers that learn by example,” *MACHINE LEARNING*.
 - [32] M. Thangaraj and M. Sivakami, “Text Classification Techniques: A Literature Review,” *IJIKM*, vol. 13, pp. 117–135, 2018, doi: 10.28945/4066.
 - [33] P. Tsangaratos and I. Ilia, “Comparison of a logistic regression and Naïve Bayes classifier in landslide susceptibility assessments: The influence of models complexity and training dataset size,” *CATENA*, vol. 145, pp. 164–179, Oct. 2016, doi: 10.1016/j.catena.2016.06.004.
 - [34] H. C. Husada and A. S. Paramita, “Analisis Sentimen Pada Maskapai Penerbangan di Platform Twitter Menggunakan Algoritma Support Vector Machine (SVM),” *teknika*, vol. 10, no. 1, pp. 18–26, Feb. 2021, doi: 10.34148/teknika.v10i1.311.
 - [35] S. Alam, H. Danang, and B. Suharjo, “User Opinion Polarization on IPDN Jatinangor,” *International Journal of Progressive Sciences and Technologies (IJPSAT)*, vol. 43, no. 2, pp. 306–313, Mar. 2024.
 - [36] B. LIU, *SENTIMENT ANALYSIS AND OPINION MINING*. S.I.: SPRINGER INTERNATIONAL PU, 2012.
 - [37] Kauffmann, Peral, Gil, Ferrández, Sellers, and Mora, “Managing Marketing Decision-Making with Sentiment Analysis: An Evaluation of the Main Product Features Using Text Data Mining,” *Sustainability*, vol. 11, no. 15, p. 4235, Aug. 2019, doi: 10.3390/su11154235.
 - [38] S. Bagui, C. Wilber, and K. Ren, “Analysis of Political Sentiment From Twitter Data:,” *NLPRE*, vol. 1, no. 1–2, p. 23, 2020, doi: 10.2991/nlpr.d.201013.001.

- [39] M. A. Saddam, E. K. Dewantara, and A. Solichin, "Sentiment Analysis of Flood Disaster Management in Jakarta on Twitter Using Support Vector Machines," *Sinkron*, vol. 8, no. 1, pp. 470–479, Jan. 2023, doi: 10.33395/sinkron.v8i1.12063.
- [40] B. Liu, *Sentiment Analysis and Opinion Mining*. Morgan & Claypool Publishers, 2012.
- [41] C. J. Pannucci and E. G. Wilkins, "Identifying and Avoiding Bias in Research:," *Plastic and Reconstructive Surgery*, vol. 126, no. 2, pp. 619–625, Aug. 2010, doi: 10.1097/PRS.0b013e3181de24bc.
- [42] N. V. Chawla, "Data Mining for Imbalanced Datasets: An Overview," in *Data Mining and Knowledge Discovery Handbook*, O. Maimon and L. Rokach, Eds., New York: Springer-Verlag, 2005, pp. 853–867. doi: 10.1007/0-387-25465-X_40.
- [43] L. Dou, F. Yang, L. Xu, and Q. Zou, "A comprehensive review of the imbalance classification of protein post-translational modifications," *Briefings in Bioinformatics*, vol. 22, no. 5, p. bbab089, Sep. 2021, doi: 10.1093/bib/bbab089.
- [44] N. V. Chawla, A. Lazarevic, L. O. Hall, and K. W. Bowyer, "SMOTEBoost: Improving Prediction of the Minority Class in Boosting," in *Knowledge Discovery in Databases: PKDD 2003*, vol. 2838, N. Lavrač, D. Gamberger, L. Todorovski, and H. Blockeel, Eds., in Lecture Notes in Computer Science, vol. 2838. , Berlin, Heidelberg: Springer Berlin Heidelberg, 2003, pp. 107–119. doi: 10.1007/978-3-540-39804-2_12.
- [45] T. Le, M. T. Vo, B. Vo, M. Y. Lee, and S. W. Baik, "A Hybrid Approach Using Oversampling Technique and Cost-Sensitive Learning for Bankruptcy Prediction," *Complexity*, vol. 2019, pp. 1–12, Aug. 2019, doi: 10.1155/2019/8460934.
- [46] J. M. Choi, "A Selective Sampling Method for Imbalanced Data Learning on Support Vector Machines," Doctor of Philosophy, Iowa State University, Digital Repository, Ames, 2010. doi: 10.31274/etd-180810-2653.
- [47] A. K. Mishra, P. Roy, S. Bandyopadhyay, and S. K. Das, "Feature fusion based machine learning pipeline to improve breast cancer prediction," *Multimed Tools Appl*, vol. 81, no. 26, pp. 37627–37655, Nov. 2022, doi: 10.1007/s11042-022-13498-4.
- [48] Y. Fan, X. Cui, H. Han, and H. Lu, "Chiller fault detection and diagnosis by knowledge transfer based on adaptive imbalanced processing," *Science and Technology for the Built Environment*, vol. 26, no. 8, pp. 1082–1099, Sep. 2020, doi: 10.1080/23744731.2020.1757327.
- [49] "Compare over-sampling samplers," Imbalanced Learn. Accessed: Jul. 15, 2024. [Online]. Available: https://imbalanced-learn.org/stable/auto_examples/over-sampling/plot_comparison_over_sampling.html
- [50] C. Shah, *A hands-on introduction to data science*. Cambridge: Cambridge university press, 2020.

- [51] S. L. Ting, W. H. Ip, and A. H. C. Tsang, "Is Naïve Bayes a Good Classifier for Document Classification?," *International Journal of Software Engineering and Its Applications*, vol. 5, no. 3, pp. 37–46, Jul. 2011.
- [52] G. Salton and C. Buckley, "Term-weighting approaches in automatic text retrieval," *Information Processing & Management*, vol. 24, no. 5, pp. 513–523, Jan. 1988, doi: 10.1016/0306-4573(88)90021-0.
- [53] C. Sammut and G. I. Webb, Eds., "TF–IDF," in *Encyclopedia of Machine Learning*, Boston, MA: Springer US, 2011, pp. 986–987. doi: 10.1007/978-0-387-30164-8_832.
- [54] V. Amrizal, "PENERAPAN METODE TERM FREQUENCY INVERSE DOCUMENT FREQUENCY (TF-IDF) DAN COSINE SIMILARITY PADA SISTEM TEMU KEMBALI INFORMASI UNTUK MENGETAHUI SYARAH HADITS BERBASIS WEB (STUDI KASUS: HADITS SHAHIH BUKHARI-MUSLIM)," *J. Teknik inform.*, vol. 11, no. 2, pp. 149–164, Nov. 2018, doi: 10.15408/jti.v11i2.8623.
- [55] I. Markoulidakis, I. Rallis, I. Georgoulas, G. Kopsiaftis, A. Doulamis, and N. Doulamis, "Multiclass Confusion Matrix Reduction Method and Its Application on Net Promoter Score Classification Problem," *Technologies*, vol. 9, no. 4, p. 81, Nov. 2021, doi: 10.3390/technologies9040081.
- [56] S. Visa, B. Ramsay, and A. Ralescu, "Confusion Matrix-based Feature Selection," presented at the Midwest Artificial Intelligence and Cognitive Science Conference, 2011.
- [57] A. De Giorgio, G. Cola, and L. Wang, "Systematic review of class imbalance problems in manufacturing," *Journal of Manufacturing Systems*, vol. 71, pp. 620–644, Dec. 2023, doi: 10.1016/j.jmsy.2023.10.014.
- [58] S. Kwak, "Are Only p -Values Less Than 0.05 Significant? A p -Value Greater Than 0.05 Is Also Significant!," *J Lipid Atheroscler*, vol. 12, no. 2, p. 89, 2023, doi: 10.12997/jla.2023.12.2.89.
- [59] A. Kumar, "Level of Significance & Hypothesis Testing," Analytics Yogi. Accessed: Jun. 26, 2024. [Online]. Available: <https://vitalflux.com/level-of-significance-hypothesis-testing/>
- [60] A. Purwarianti and I. A. P. A. Crisdayanti, "Improving Bi-LSTM Performance for Indonesian Sentiment Analysis Using Paragraph Vector," in *2019 International Conference of Advanced Informatics: Concepts, Theory and Applications (ICAICTA)*, Yogyakarta, Indonesia: IEEE, Sep. 2019, pp. 1–5. doi: 10.1109/ICAICTA.2019.8904199.