

DAFTAR PUSTAKA

- Aftab, F., Bazai, S. U., Marjan, S., Baloch, L., Aslam, S., Amphawan, A., & Neo, T. K. (2023). A Comprehensive Survey on Sentiment Analysis Techniques. *International Journal of Technology*, 14(6), 1288–1298. <https://doi.org/10.14716/IJTECH.V14I6.6632>
- Agustina, N., Hadhiwibowo, A., Saputra, I. K. A. S., Cahyati, W., Indriarti, Z. P., & Rustandi, B. (2023). Sentiment Analysis of Air Quality to Identify Public Perspectives in Indonesia Based on Machine Learning. *Proceeding of 2023 17th International Conference on Telecommunication Systems, Services, and Applications, TSSA 2023*. <https://doi.org/10.1109/TSSA59948.2023.10366899>
- Alam, S., & Yao, N. (2019). The impact of preprocessing steps on the accuracy of machine learning algorithms in sentiment analysis. *Computational and Mathematical Organization Theory*, 25(3), 319–335. <https://doi.org/10.1007/S10588-018-9266-8>
- Alasadi, S. A., & Bhaya, W. S. (2017). Review of data preprocessing techniques in data mining. *Journal of Engineering and Applied Sciences*, 12(16), 4102–4107. <https://doi.org/10.3923/JEASCI.2017.4102> .4107
- Ali, H., Salleh, M. N. B. M., Talpur, K., & Ullah, A. (2019). A review on data preprocessing methods for class imbalance problem. *International Journal of Engineering & Technology*, 8(3), 390–397. <https://doi.org/0.14419/ijet.v8i3.29508>
- Azevedo, A., & Santos, M. F. (2008). KDD, semma and CRISP-DM: A parallel overview. <https://www.scopus.com/record/display.uri?eid=2-s2.0-58449118356&origin=inward&txGid=f62c40add7d3a8abef0b1d8910ed0450>.
- Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer. <https://link.springer.com/book/9780387310732>

- Błaszczyński, J., & Stefanowski, J. (2015). Neighbourhood sampling in bagging for imbalanced data. *Neurocomputing*, 150(PB), 529–542. <https://doi.org/10.1016/J.NEUCOM.2014.07.064>
- Bolboacă, S. D., Jäntschi, L., Sestraş, A. F., Sestraş, R. E., & Pamfil, D. C. (2011). Pearson-fisher chi-square statistic revisited. *Information (Switzerland)*, 2(3), 528–545. <https://doi.org/10.3390/info2030528>
- Campbell, C., & Ying, Y. (2011). *Support Vector Machines for Classification*. Springer, Cham. https://doi.org/10.1007/978-3-031-01552-6_1
- Chandra, W., Suprihatin, B., & Resti, Y. (2023). Median-KNN Regressor-SMOTE-Tomek Links for Handling Missing and Imbalanced Data in Air Quality Prediction. *Symmetry*, 15. <https://doi.org/10.3390/sym15040887>
- Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research*, 16, 321–357. <https://doi.org/10.1613/JAIR.953>
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences Second Edition*.
- Collier-Oxandale, A., Feenstra, B., Papapostolou, V., Zhang, H., Kuang, M., Der Boghossian, B., & Polidori, A. (2020). Field and laboratory performance evaluations of 28 gas-phase air quality sensors by the AQ-SPEC program. *Atmospheric Environment*, 220, 117092. <https://doi.org/10.1016/J.ATMOSENV.2019.117092>
- Dang, N. C., Moreno-García, M. N., & De la Prieta, F. (2020). Sentiment Analysis Based on Deep Learning: A Comparative Study. *Electronics 2020, Vol. 9, Page 483*, 9(3), 483. <https://doi.org/10.3390/ELECTRONICS9030483>
- Databoks. (2023, November 28). *Ada 27 Juta Pengguna Twitter di Indonesia, Terbanyak ke-4 Global*. <https://databoks.katadata.co.id/teknologi-telekomunikasi/statistik/75dd4b36866dc54/ada-27-juta-pengguna-twitter-di-indonesia-terbanyak-ke-4-global>

- Faridzi, Z. Al, Pramesti, D., & Fa'Rifah, R. Y. (2023). A Comparison of Oversampling and Undersampling Methods in Sentiment Analysis Regarding Indonesia Fuel Price Increase Using Support Vector Machine. *ICADEIS 2023 - International Conference on Advancement in Data Science, E-Learning and Information Systems: Data, Intelligent Systems, and the Applications for Human Life, Proceeding*.
<https://doi.org/10.1109/ICADEIS58666.2023.10270851>
- Fathurahman, A. D., Pramesti, D., & Fakhurroja, H. (2024). *Sentiment Analysis of Presidential Debate Videos on YouTube in the 2024 Indonesian Presidential Elections*. 545–550. <https://doi.org/10.1109/icodsa62899.2024.10651794>
- Fawcet, T. (2004). *ROC Graphs: Notes and Practical Considerations for Researchers*.
https://www.researchgate.net/publication/284043217_ROC_Graphs_Notes_and_Practical_Considerations_for_Researchers
- Gharehchopogh, F. S., & Khalifelu, Z. A. (2011, November). Analysis and evaluation of unstructured data: Text mining versus natural language processing. *2011 5th International Conference on Application of Information and Communication Technologies, AICT 2011*.
<https://doi.org/10.1109/ICAICT.2011.6111017>
- Hoar, B., Ramachandran, R., Levis, M., Sparck, E., Wu, K., & Liu, C. (2023). *Summative Student Course Review Tool Based on Machine Learning Sentiment Analysis to Enhance Life Science Feedback Efficacy*.
<https://arxiv.org/abs/2301.06173v1>
- Hsu, C. W., & Lin, C. J. (2002). A comparison of methods for multiclass support vector machines. *IEEE Transactions on Neural Networks*, 13(2), 415–425.
<https://doi.org/10.1109/72.991427>
- IQAIR. (2023). *Negara Paling Berpolusi di Dunia 2023 - Ranking PM2.5 | IQAir*.
<https://www.iqair.com/id/world-most-polluted-countries>
- Isabelle, I., & Elisseeff, A. (2003). Document details - An introduction to variable and feature selection. <https://www.scopus.com/record/display.uri?eid=2-s2.0->

33745561205&origin=inward&txGid=52a9e09c5264a507a0aad8f3994e8f6
1, 3.

- Karagulian, F., Barbieri, M., Kotsev, A., Spinelle, L., Gerboles, M., Lagler, F., Redon, N., Crunaire, S., & Borowiak, A. (2019). Review of the Performance of Low-Cost Sensors for Air Quality Monitoring. *Atmosphere* 2019, Vol. 10, Page 506, 10(9), 506. <https://doi.org/10.3390/ATMOS10090506>
- Khurniawan, F. S., & Ruldeviyani, Y. (2020). Twitter Sentiment Analysis: Case Study on the Revision of the Indonesia's Corruption Eradication Commission (KPK) Law 2019. *2020 International Conference on Data Science and Its Applications, ICoDSA* 2020. <https://doi.org/10.1109/ICODSA50139.2020.9212851>
- Kruskal, W. H., & Wallis, W. A. (1952). Use of Ranks in One-Criterion Variance Analysis. *Journal of the American Statistical Association*, 47, 583. <https://doi.org/10.2307/2280779>
- Kumar, A., Dabas, V., & Hooda, P. (2020). Text classification algorithms for mining unstructured data: a SWOT analysis. *International Journal of Information Technology (Singapore)*, 12(4), 1159–1169. <https://doi.org/10.1007/S41870-017-0072-1/METRICS>
- Lamba, M., & Madhusudhan, M. (2022). Text Pre-Processing. Dalam *Text Mining for Information Professionals: An Uncharted Territory* (1 ed., hlm. 79–103). Springer International Publishing. https://doi.org/10.1007/978-3-030-85085-2_3
- Le, T., Vo, M. T., Vo, B., Lee, M. Y., & Baik, S. W. (2019). *A Hybrid Approach Using Oversampling Technique and Cost-Sensitive Learning for Bankruptcy Prediction*. 2, 1–12. <https://doi.org/http://dx.doi.org/10.1155/2019/8460934>
- Levine, T. R., & Hullett, C. R. (2002). Eta Squared, Partial Eta Squared, and Misreporting of Effect Size in Communication Research. *Human Communication Research*, 28, 612–625. <https://doi.org/10.1111/j.1468-2958.2002.tb00828.x>

- López, V., Fernández, A., Moreno-Torres, J. G., & Herrera, F. (2012). Analysis of preprocessing vs. cost-sensitive learning for imbalanced classification. Open problems on intrinsic data characteristics. *Expert Systems with Applications*, 39(7), 6585–6608. <https://doi.org/10.1016/J.ESWA.2011.12.043>
- Maharani, W., & Effendy, V. (2022). Big five personality prediction based in Indonesian tweets using machine learning methods. *International Journal of Electrical and Computer Engineering (IJECE)*, 12(2), 1973–1981. <https://doi.org/10.11591/ijece.v12i2.pp1973-1981>
- Manisalidis, I., Stavropoulou, E., Stavropoulos, A., & Bezirtzoglou, E. (2020). Environmental and Health Impacts of Air Pollution: A Review. *Frontiers in Public Health*, 8, 505570. <https://doi.org/10.3389/FPUBH.2020.00014/BIBTEX>
- Margaritis, D., Keramydas, C., Papachristos, I., & Lambropoulou, D. (2021). Calibration of Low-cost Gas Sensors for Air Quality Monitoring. *Aerosol and Air Quality Research*, 21(11), 210073. <https://doi.org/10.4209/AAQR.210073>
- Meng, J., Shi, L., & Luan, Y. (2014). Plant microRNA-Target Interaction Identification Model Based on the Integration of Prediction Tools and Support Vector Machine. *PLOS ONE*, 9(7), e103181. <https://doi.org/10.1371/JOURNAL.PONE.0103181>
- Min, J., Yunxiu, H., Yong, S., Fengxiang, J., & Ting, L. (2021). Research on Analysis of Evaluation Influence Factors of Air Quality Opinions Sentiment Value and Quantification Method. *Water, Air, and Soil Pollution*, 232(7), 1–17. <https://doi.org/10.1007/S11270-021-05197-X/METRICS>
- Nafas. (2024a). *NAFAS Indonesia*. <https://nafas.co.id/nafas-buka-data>
- Nafas. (2024b). *NAFAS website*. <https://caz.nafas.co.id/>
- Naseem, U., Razzak, I., & Eklund, P. W. (2021). A survey of pre-processing techniques to improve short-text quality: a case study on hate speech detection on twitter. *Multimedia Tools and Applications*, 80(28–29), 35239–35266. <https://doi.org/10.1007/S11042-020-10082-6/METRICS>

- Nti, I. K., Nyarko-Boateng, O., Adekoya, F. A., & Weyori, B. A. (2021). An empirical assessment of different kernel functions on the performance of support vector machines. *Bulletin of Electrical Engineering and Informatics*, 10(6), 3403–3411. <https://doi.org/10.11591/EEL.V10I6.3046>
- Nurdeni, D. A., Budi, I., & Santoso, A. B. (2021). Sentiment Analysis on Covid19 Vaccines in Indonesia: From the Perspective of Sinovac and Pfizer. *3rd 2021 East Indonesia Conference on Computer and Information Technology, EIconCIT* 2021, 122–127. <https://doi.org/10.1109/EICONCIT50028.2021.9431852>
- Prastiani, Fakhurroja, H., & Hamami, F. (2023). Social Media Sentiment Analysis for Local Water Company Customers Using a Support Vector Machine Algorithm. *10th International Conference on ICT for Smart Society, ICISS 2023 - Proceeding*. <https://doi.org/10.1109/ICISS59129.2023.10291991>
- Ramachandran, D., & Parvathi, R. (2019). Analysis of Twitter Specific Preprocessing Technique for Tweets. *Procedia Computer Science*, 165, 245–251. <https://doi.org/10.1016/J.PROCS.2020.01.083>
- Ridwan, A., Nuha, H. H., & Dharayani, R. (2022). Sentiment Analysis of Floods on Twitter Social Media Using the Naive Bayes Classifier Method with the N-Gram Feature. *2022 International Conference on Data Science and Its Applications, ICoDSA* 2022, 114–118. <https://doi.org/10.1109/ICODSA55874.2022.9862827>
- Roiger, R. J. (2017). Data Mining: A Tutorial-Based Primer, Second Edition. Dalam *Data Mining: A Tutorial-Based Primer, Second Edition* (2nd Edition). CRC Press. <https://doi.org/10.1201/9781315382586/DATA-MINING-RICHARD-ROIGER/ACCESSIBILITY-INFORMATION>
- Sabri, I. A. A., Man, M., Bakar, W. A. W. A., & Rose, A. N. M. (2019). Web Data Extraction Approach for Deep Web using WEIDJ. *Procedia Computer Science*, 163, 417–426. <https://doi.org/10.1016/J.PROCS.2019.12.124>
- Salsabila, N. A., Ardhito Winatmoko, Y., Akbar Septiandri, A., & Jamal, A. (2018). Colloquial Indonesian Lexicon. *Proceedings of the 2018 International*

- Conference on Asian Language Processing, IALP 2018*, 226–229.
<https://doi.org/10.1109/IALP.2018.8629151>
- Santos, C. N. Dos, & Bayser, M. G. de. (2014, September). (PDF) *Deep Convolutional Neural Networks for Sentiment Analysis of Short Texts*.
https://www.researchgate.net/publication/274380447_Deep_Convolutional_Neural_Networks_for_Sentiment_Analysis_of_Short_Texts
- Sokhin, T., & Butakov, N. (2018). Semi-automatic sentiment analysis based on topic modeling. *Procedia Computer Science*, 136, 284–292.
<https://doi.org/10.1016/J.PROCS.2018.08.286>
- Sugiyono. (2007). Statistik untuk penelitian. Dalam <https://inlislite.uin-suska.ac.id/opac/detail-opac?id=30140>.
- Talib, R., Kashif, M. H., Ayesha, S., & Fatima, F. (2016). Text Mining: Techniques, Applications and Issues. *International Journal of Advanced Computer Science and Applications*, 7(11). <https://doi.org/10.14569/IJACSA.2016.071153>
- Tan, K. L., Lee, C. P., & Lim, K. M. (2023). A Survey of Sentiment Analysis: Approaches, Datasets, and Future Research. *Applied Sciences* 2023, Vol. 13, Page 4550, 13(7), 4550. <https://doi.org/10.3390/APP13074550>
- Tao, Y., Zhang, F., Shi, C., & Chen, Y. (2019). Social Media Data-Based Sentiment Analysis of Tourists' Air Quality Perceptions. *Sustainability* 2019, Vol. 11, Page 5070, 11(18), 5070. <https://doi.org/10.3390/SU11185070>
- Taye, M. M. (2023). Understanding of Machine Learning with Deep Learning: Architectures, Workflow, Applications and Future Directions. *Computers* 2023, Vol. 12, Page 91, 12(5), 91. <https://doi.org/10.3390/COMPUTERS12050091>
- United Nations Environment Programme (UNEP), & Climate and Clean Air Coalition (CCAC). (2019). Air Pollution in Asia and the Pacific: Science-based solutions (Summary + Full report). Dalam <https://www.ccacoalition.org/resources/air-pollution-asia-and-pacific-science-based-solutions-summary-full-report>.

- Vidya, N. A., Fanany, M. I., & Budi, I. (2015). Twitter Sentiment to Analyze Net Brand Reputation of Mobile Phone Providers. *Procedia Computer Science*, 72, 519–526. <https://doi.org/10.1016/j.procs.2015.12.159>
- Wadhwa, K., Mehra, D., Gosain, H., & Haritash, A. K. (2023). Sentiment Analysis of Air Quality Perception in Major Metro Cities of India. *2023 14th International Conference on Computing Communication and Networking Technologies, ICCCNT 2023*. <https://doi.org/10.1109/ICCCNT56998.2023.10307448>
- Wankhade, M., Rao, A. C. S., & Kulkarni, C. (2022). A survey on sentiment analysis methods, applications, and challenges. *Artificial Intelligence Review* 2022 55:7, 55(7), 5731–5780. <https://doi.org/10.1007/S10462-022-10144-1>
- X. (2023). *X for Websites | Products | Twitter Developer Platform*. <https://developer.x.com/en/products/x-for-websites>
- Xu, X., Dong, D., Wang, Y., & Wang, S. (2019). The impacts of different air pollutants on domestic and inbound tourism in China. *International Journal of Environmental Research and Public Health*, 16. <https://doi.org/10.3390/ijerph16245127>
- Yao, K., & Zheng, Y. (2023). Fundamentals of Machine Learning. Dalam K. Yao & Y. Zheng (Ed.), *Nanophotonics and Machine Learning: Concepts, Fundamentals, and Applications* (hlm. 77–112). Springer International Publishing. https://doi.org/10.1007/978-3-031-20473-9_3
- Zhang, Q., Yang, L. T., Chen, Z., & Li, P. (2018). A survey on deep learning for big data. *Information Fusion*, 42, 146–157. <https://doi.org/10.1016/J.INFFUS.2017.10.006>