

DAFTAR PUSTAKA

- Ahmad, M. I., & Anwari, Moh. K. (2024). Computational linguistics and natural language processing techniques for semantic field extraction in Arabic online news. *Studies in English Language and Education*, 11(3), 1685–1709. <https://doi.org/10.24815/siele.v11i3.38090>
- Ahmad, S. B. S., Rafie, M., & Ghorabie, S. M. (2021). Spam detection on Twitter using a support vector machine and users' features by identifying their interactions. *Multimedia Tools and Applications*, 80(8). <https://doi.org/10.1007/s11042-020-10405-7>
- Alom, Z., Carminati, B., & Ferrari, E. (2020). A deep learning model for Twitter spam detection. *Online Social Networks and Media*, 18. <https://doi.org/10.1016/j.osnem.2020.100079>
- Argina, A. M. (2020). Penerapan Metode Klasifikasi K-Nearest Neighbor pada Dataset Penderita Penyakit Diabetes. *Indonesian Journal of Data and Science*, 1(2). <https://doi.org/10.33096/ijodas.v1i2.11>
- Binsaeed, K., Stringhini, G., & Youssef, A. E. (2020). Detecting Spam in Twitter Microblogging Services: A Novel Machine Learning Approach based on Domain Popularity. *International Journal of Advanced Computer Science and Applications*, 11(11). <https://doi.org/10.14569/IJACSA.2020.0111103>
- Chandrashekar, G., & Sahin, F. (2014). A survey on feature selection methods. *Computers and Electrical Engineering*, 40(1). <https://doi.org/10.1016/j.compeleceng.2013.11.024>
- Dharma, A. S., Sitilonga, O. Y., & Manurung, H. J. (2019). Perbandingan Algoritma Naive Bayes, ID3, Dan TAN Pada Klasifikasi SMS Spam. *Journal of Maritime and Education*, 1(2).
- Elgeldawi, E., Sayed, A., Galal, A. R., & Zaki, A. M. (2021). Hyperparameter tuning for machine learning algorithms used for arabic sentiment analysis. *Informatics*, 8(4). <https://doi.org/10.3390/informatics8040079>
- Fahmi Al Hafidz, A., Yulia Puspaningrum, E., & Lina Nurlaili, A. (2025). PENERAPAN LIGHTGBM MENGGUNAKAN EKSTRAKSI FITUR

- RUANG WARNA HSV UNTUK KLASIFIKASI REMPAH RIMPANG.
JATI (Jurnal Mahasiswa Teknik Informatika), 9(3), 5492–5498.
<https://doi.org/10.36040/jati.v9i3.14306>
- Fakih, A., Hamzami, M. A., Hadiano, M. R., & Alifah, N. I. S. (2025). Perbandingan Akurasi Algoritma C4.5 dan K-NN Untuk Prediksi Kelulusan Mahasiswa Penerima Beasiswa. *Jurnal Komputer Antartika*, 3(1), 18–25.
<https://doi.org/10.70052/jka.v3i1.623>
- Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, 27(8). <https://doi.org/10.1016/j.patrec.2005.10.010>
- Fazil, M., & Abulaish, M. (2018). A Hybrid Approach for Detecting Automated Spammers in Twitter. *IEEE Transactions on Information Forensics and Security*, 13(11). <https://doi.org/10.1109/TIFS.2018.2825958>
- Fu, F., Jiang, J., Shao, Y., & Cuix, B. (2018). An experimental evaluation of large scale GBDT Systems. *Proceedings of the VLDB Endowment*, 12(11). <https://doi.org/10.14778/3342263.3342273>
- Gadde, S., Lakshmanarao, A., & Satyanarayana, S. (2021). SMS Spam Detection using Machine Learning and Deep Learning Techniques. *2021 7th International Conference on Advanced Computing and Communication Systems, ICACCS 2021*.
<https://doi.org/10.1109/ICACCS51430.2021.9441783>
- Gumelar, G., Ain, Q., Marsuciati, R., Agustanti Bambang, S., Sunyoto, A., & Syukri Mustafa, M. (2021). Kombinasi Algoritma Sampling dengan Algoritma Klasifikasi untuk Meningkatkan Performa Klasifikasi Dataset Imbalance. *SISFOTEK : Sistem Informasi Dan Teknologi*.
- Gupta, H., Jamal, M. S., Madisetty, S., & Desarkar, M. S. (2018). A framework for real-time spam detection in Twitter. *2018 10th International Conference on Communication Systems and Networks, COMSNETS 2018, 2018-January*.
<https://doi.org/10.1109/COMSNETS.2018.8328222>
- Iqbal, T., Elahi, A., Wijns, W., & Shahzad, A. (2022). Exploring Unsupervised Machine Learning Classification Methods for Physiological Stress Detection.

Frontiers in Medical Technology, 4(March), 1–12.
<https://doi.org/10.3389/fmedt.2022.782756>

- Isa, I. G. T., & Junedi, B. (2022). Hyperparameter Tuning Epoch dalam Meningkatkan Akurasi Data Latih dan Data Validasi pada Citra Pengendara. *Prosiding Sains Nasional Dan Teknologi*, 12(1).
<https://doi.org/10.36499/psnst.v12i1.6697>
- Istifarinta, O. L., Prajitno, P., & Soejoko, D. S. (2022). Computer-Aided Detection (CAD) Deteksi Nodul Paru-Paru dari Computed Tomography (CT). *INTERVENSIONAL Journal of Medical Physics and Biophysics*, 9(1).
- Jain, A. K., Goel, D., Agarwal, S., Singh, Y., & Bajaj, G. (2020). Predicting Spam Messages Using Back Propagation Neural Network. *Wireless Personal Communications*, 110(1). <https://doi.org/10.1007/s11277-019-06734-y>
- Kaddoura, S., Chandrasekaran, G., Popescu, D. E., & Duraisamy, J. H. (2022). A systematic literature review on spam content detection and classification. *PeerJ Computer Science*, 8. <https://doi.org/10.7717/PEERJ-CS.830>
- Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T. Y. (2017). LightGBM: A highly efficient gradient boosting decision tree. *Advances in Neural Information Processing Systems, 2017-December*.
- Khurana, D., Koli, A., Khatter, K., & Singh, S. (2023). Natural language processing: state of the art, current trends and challenges. *Multimedia Tools and Applications*, 82(3). <https://doi.org/10.1007/s11042-022-13428-4>
- Kumar, N., Sonowal, S., & Nishant. (2020). Email Spam Detection Using Machine Learning Algorithms. *Proceedings of the 2nd International Conference on Inventive Research in Computing Applications, ICIRCA 2020*.
<https://doi.org/10.1109/ICIRCA48905.2020.9183098>
- Liu, P., Qian, W., Xu, D., Ren, B., & Cao, J. (2023). Multi-Modal Fake News Detection via Bridging the Gap between Modals. *Entropy*, 25(4), 1–17.
<https://doi.org/10.3390/e25040614>
- Mehbodniya, A., Rao, M. V., David, L. G., Joe Nige, K. G., & Vennam, P. (2022). Online product sentiment analysis using random evolutionary whale

- optimization algorithm and deep belief network. *Pattern Recognition Letters*, 159, 1–8. <https://doi.org/10.1016/j.patrec.2022.04.024>
- Mukaka, M. M. (2012). Statistics Corner: A guide to appropriate use of Correlation coefficient in medical research. In *Malawi Medical Journal* (Vol. 24, Issue 3). www.mmj.medcol.mw
- Natekin, A., & Knoll, A. (2013). Gradient boosting machines, a tutorial. *Frontiers in Neurorobotics*, 7(DEC). <https://doi.org/10.3389/fnbot.2013.00021>
- Navaney, P., Dubey, G., & Rana, A. (2018). SMS Spam Filtering Using Supervised Machine Learning Algorithms. *Proceedings of the 8th International Conference Confluence 2018 on Cloud Computing, Data Science and Engineering, Confluence 2018*. <https://doi.org/10.1109/CONFLUENCE.2018.8442564>
- Nugraha, A. C., & Irawan, M. I. (2023). Komparasi Deteksi Kecurangan pada Data Klaim Asuransi Pelayanan Kesehatan Menggunakan Metode Support Vector Machine (SVM) dan Extreme Gradient Boosting (XGBoost). *Jurnal Sains Dan Seni ITS*, 12(1). <https://doi.org/10.12962/j23373520.v12i1.107032>
- Nugraha, W., & Sasongko, A. (2022). Hyperparameter Tuning pada Algoritma Klasifikasi dengan Grid Search. *SISTEMASI: Jurnal Sistem Informasi*, 11(2).
- Ora, A. (2020). Spam Detection in Short Message Service Using Natural Language Processing and Machine Learning Techniques. *National College of Ireland*, Pp.1-27.
- Puspitasari, N., Septiarini, A., & Aliudin, A. R. (2023). METODE K-NEAREST NEIGHBOR DAN FITUR WARNA UNTUK KLASIFIKASI DAUN SIRIH BERDASARKAN CITRA DIGITAL. *PROSISKO: Jurnal Pengembangan Riset Dan Observasi Sistem Komputer*, 10(2). <https://doi.org/10.30656/prosisko.v10i2.6924>
- Riza Adrianti Supono, & Muhammad Azis Suprayogi. (2021). Perbandingan Metode TF-ABS dan TF-IDF Pada Klasifikasi Teks Helpdesk Menggunakan K-Nearest Neighbor. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 5(5). <https://doi.org/10.29207/resti.v5i5.3403>

- Rodrigues, A. P., Fernandes, R., Aakash, A., Abhishek, B., Shetty, A., Atul, K., Lakshmana, K., & Shafi, R. M. (2022). Real-Time Twitter Spam Detection and Sentiment Analysis using Machine Learning and Deep Learning Techniques. *Computational Intelligence and Neuroscience*, 2022. <https://doi.org/10.1155/2022/5211949>
- Salman Al-Tameemi, I. K., Feizi-Derakhshi, M. R., Pashazadeh, S., & Asadpour, M. (2023). Interpretable Multimodal Sentiment Classification Using Deep Multi-View Attentive Network of Image and Text Data. *IEEE Access*, 11(July), 91060–91081. <https://doi.org/10.1109/ACCESS.2023.3307716>
- Septiana Rizky, P., Haiban Hirzi, R., & Hidayaturohman, U. (2022). Perbandingan Metode LightGBM dan XGBoost dalam Menangani Data dengan Kelas Tidak Seimbang. *J Statistika: Jurnal Ilmiah Teori Dan Aplikasi Statistika*, 15(2). <https://doi.org/10.36456/jstat.vol15.no2.a5548>
- Srivastava, A., & Singh, P. (2024). Spam Detection Using Natural Language Processing. *Journal of Applied Science and Education (JASE)*, 4(2), 1–7. <https://doi.org/10.54060/a2zjournals.jase.70>
- Sun, N., Lin, G., Qiu, J., & Rimba, P. (2022). Near real-time twitter spam detection with machine learning techniques. *International Journal of Computers and Applications*, 44(4). <https://doi.org/10.1080/1206212X.2020.1751387>
- Washha, M., Qaroush, A., Mezghani, M., & Sedes, F. (2019). Unsupervised collective-based framework for dynamic retraining of supervised real-time spam tweets detection model. *Expert Systems with Applications*, 135. <https://doi.org/10.1016/j.eswa.2019.05.052>
- Wibisono, A. D., Dadi Rizkiono, S., & Wantoro, A. (2020). FILTERING SPAM EMAIL MENGGUNAKAN METODE NAIVE BAYES. *TELEFORTECH: Journal of Telematics and Information Technology*, 1(1). <https://doi.org/10.33365/tft.v1i1.685>
- Xi, X. (2024). The role of LightGBM model in management efficiency enhancement of listed agricultural companies. *Applied Mathematics and Nonlinear Sciences*, 9(1). <https://doi.org/10.2478/amns.2023.2.00386>

- Yang, J. S., Zhao, C. Y., Yu, H. T., & Chen, H. Y. (2020). Use GBDT to Predict the Stock Market. *Procedia Computer Science*, 174. <https://doi.org/10.1016/j.procs.2020.06.071>
- Yuan, H., Tang, Y., Sun, W., & Liu, L. (2020). A detection method for android application security based on TF-IDF and machine learning. *PLoS ONE*, 15(9 September). <https://doi.org/10.1371/journal.pone.0238694>
- Zhang, D., & Gong, Y. (2020). The Comparison of LightGBM and XGBoost Coupling Factor Analysis and Prediagnosis of Acute Liver Failure. *IEEE Access*, 8. <https://doi.org/10.1109/ACCESS.2020.3042848>
- Zhang, P., Jia, Y., & Shang, Y. (2022). Research and application of XGBoost in imbalanced data. *International Journal of Distributed Sensor Networks*, 18(6). <https://doi.org/10.1177/15501329221106935>
- Zhang, T., Wang, D., Chen, H., Zeng, Z., Guo, W., Miao, C., & Cui, L. (2020). BDANN: BERT-Based Domain Adaptation Neural Network for Multi-Modal Fake News Detection. *Proceedings of the International Joint Conference on Neural Networks*. <https://doi.org/10.1109/IJCNN48605.2020.9206973>