

DAFTAR PUSTAKA

- Aburbeian, A. H. M., Owda, A. Y., & Owda, M. (2022). A Technology Acceptance Model Survey of the Metaverse Prospects. *AI (Switzerland)*, 3(2), 285–302.
- Aliyah Salsabila, N., 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.
- Arifiandy, R., & Fahmi, H. (2024). Perbandingan Vektorisasi Deteksi Spam Email Menggunakan Bag of Word, TF IDF, dan Word2Vec pada Multinomial Naïve Bayes. *Syntax: Jurnal Informatika*, 13(01), 1–14.
- Ariyanto, B. T., Ruldeviyani, Y., Dharmawan, G. B., & Bahar, M. I. (2024). Optimizing User Satisfaction: A Comprehensive Evaluation of the Info BMKG App Using UEQ+ and IPA Methods. *International Journal of Engineering Continuity*, 3(1), 139–156.
- Asnawi, M. H., Pravitasari, A. A., Herawan, T., & Hendrawati, T. (2023). The Combination of Contextualized Topic Model and MPNet for User Feedback Topic Modeling. *IEEE Access*, 11(November), 130272–130286.
- Azhar, N. A., Mohd Pozi, M. S., Din, A. M., & Jatowt, A. (2023). An Investigation of SMOTE based Methods for Imbalanced Datasets with Data Complexity Analysis. *IEEE Transactions on Knowledge and Data Engineering*, 35(7), 6651–6672.
- Bao, Y., & Yang, S. (2023). Two Novel SMOTE Methods for Solving Imbalanced Classification Problems. *IEEE Access*, 11(December 2022), 5816–5823.
- Bayu Baskoro, B., Susanto, I., Khomsah, S., Informatika, P., Sains Data, P., Teknologi Telkom Purwokerto Jl Panjaitan, I. DI, & Tengah, J. (2021). Analisis Sentimen Pelanggan Hotel di Purwokerto Menggunakan Metode

- Random Forest dan TF-IDF (Studi Kasus: Ulasan Pelanggan Pada Situs TRIPADVISOR). *Journal of Informatics Information System Software Engineering and Applications (INISTA)*, , Volume 3 N(2), 21–029.
- Cai, Y., Yuan, Y., & Zhou, A. (2024). Predictive Slope Stability Early Warning Model based on CatBoost. *Scientific Reports*, 14(1), 1–14.
- Chai, C. P. (2023). Comparison of Text Preprocessing Methods. *Natural Language Engineering*, 29(3), 509–553.
- Chowdhury, G. (2003). Natural Language Processing. *The Annual Review of Information Science and Technology*, 37, 51–89.
- Curiskis, S. A., Drake, B., Osborn, T. R., & Kennedy, P. J. (2020). An Evaluation of Document Clustering and Topic Modelling in Two Online Social Networks: Twitter and Reddit. *Information Processing and Management*, 57(2), 102034.
- Dieng, A. B., Ruiz, F. J. R., & Blei, D. M. (2020). Topic Modeling in Embedding Spaces. *Transactions of the Association for Computational Linguistics*, 8, 439–453.
- Ding, B., Qin, C., Liu, L., Chia, Y. K., Li, B., Joty, S., & Bing, L. (2023). Is GPT-3 a Good Data Annotator? *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, 1, 11173–11195.
- Duong, H. T., & Nguyen-Thi, T. A. (2021). A Review: Preprocessing Techniques and Data Augmentation for Sentiment Analysis. *Computational Social Networks*, 8(1), 1–16.
- Efat, A. A., Atiq, A., Abeed, A. S., Momin, A., & Alam, M. G. R. (2023). EMPOLITICON: NLP and ML Based Approach for Context and Emotion Classification of Political Speeches From Transcripts. *IEEE Access*, 11(June), 54808–54821.
- Egger, R., & Yu, J. (2022). A Topic Modeling Comparison Between LDA, NMF, Top2Vec, and BERTopic to Demystify Twitter Posts. *Frontiers in Sociology*, 7(May), 1–16.

- 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), 1–21.
- Eren, A. O., & Sert, M. (2023). Automated Audio Captioning With Topic Modeling. *IEEE Access*, 11(December 2022), 4983–4991.
- Gokcimen, T., & Das, B. (2024). Exploring climate change discourse on social media and blogs using a topic modeling analysis. *Heliyon*, 10(11), e32464.
- Gupta, A., Dengre, V., Kheruwala, H. A., & Shah, M. (2020). Comprehensive Review of Text-Mining Applications in Finance. *Financial Innovation*, 6(1), 1–25.
- Han, J., Pei, J., & Tong, H. (2023). Data Mining Concepts and Techniques Fourth Edition. In *Sustainability (Switzerland)* (Vol. 15, Issue 1).
- Hancock, J. T., & Khoshgoftaar, T. M. (2020). CatBoost for Big Data: an interdisciplinary review. *Journal of Big Data*, 7(1).
- Hassani, H., Beneki, C., Unger, S., Mazinani, M. T., & Yeganegi, M. R. (2020). Text Mining in Big Data Analytics. *Big Data and Cognitive Computing*, 4(1), 1–34.
- Hermans, T. D. G., Šakić Trogrlić, R., van den Homberg, M. J. C., Bailon, H., Sarku, R., & Mosurska, A. (2022). Exploring the Integration of Local and Scientific Knowledge in Early Warning Systems for Disaster Risk Reduction: a Review. *Natural Hazards*, 114(2), 1125–1152.
- Hoque, K. E., & Aljamaan, H. (2021). Impact of Hyperparameter Tuning on Machine Learning Models in Stock Price Forecasting. *IEEE Access*, 9, 163815–163830.
- Ibrahim, A. A., Ridwan, R. L., Muhammed, M. M., Abdulaziz, R. O., & Saheed, G. A. (2020). Comparison of the CatBoost Classifier with other Machine Learning Methods. *International Journal of Advanced Computer Science and Applications*, 11(11), 738–748.

- Jijo, B. T., & Abdulazeez, A. M. (2021). Classification Based on Decision Tree Algorithm for Machine Learning. *Journal of Applied Science and Technology Trends*, 2(1), 20–28.
- Karanikola, A., Davrazos, G., Liapis, C. M., & Kotsiantis, S. (2023). Financial Sentiment Analysis: Classic Methods vs. Deep Learning Models. *Intelligent Decision Technologies*, 17(4), 893–915.
- Kastrati, Z., Imran, A. S., Daudpota, S. M., Memon, M. A., & Kastrati, M. (2023). Soaring Energy Prices: Understanding Public Engagement on Twitter Using Sentiment Analysis and Topic Modeling with Transformers. *IEEE Access*, 11(March), 26541–26553.
- 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), 3713–3744.
- Korencic, D., Ristov, S., Repar, J., & Snajder, J. (2021). A Topic Coverage Approach to Evaluation of Topic Models. *IEEE Access*, 9, 123280–123312.
- Kumar, A., Karamchandani, A., & Singh, S. (2024). Topic Modeling of Neuropsychiatric Diseases Related to Gut Microbiota and Gut Brain Axis Using Artificial Intelligence Based BERTopic Model on PubMed Abstracts. *Neuroscience Informatics*, 4(4), 100175.
- Kumar, S., Kar, A. K., & Ilavarasan, P. V. (2021). Applications of Text Mining in Services Management: A Systematic Literature Review. *International Journal of Information Management Data Insights*, 1(1).
- Li, Q., Wu, Z., Wen, Z., & He, B. (2020). Privacy-Preserving Gradient Boosting Decision Trees. *AAAI 2020 - 34th AAAI Conference on Artificial Intelligence*, 784–791.
- Li, T., Cui, L., Wu, Y., Pandey, R., Liu, H., Dong, J., Wang, W., Song, X., Hao, Y., Cui, X., Du, J., Zhang, X., & Wang, Y. (2024). Unveiling and Advancing Grassland Degradation Research Using a BERTopic Modelling Approach. *Journal of Integrative Agriculture*.

- Mahadevkar, S. V., Khemani, B., Patil, S., Kotecha, K., Vora, D. R., Abraham, A., & Gabralla, L. A. (2022). A Review on Machine Learning Styles in Computer Vision - Techniques and Future Directions. *IEEE Access*, 10(October), 107293–107329.
- Mehta, P., & Pandya, S. (2020). A Review on Sentiment Analysis Methodologies, Practices and Applications. *International Journal of Scientific and Technology Research*, 9(2), 601–609.
- Mishra, P., Biancolillo, A., Roger, J. M., Marini, F., & Rutledge, D. N. (2020). New Data Preprocessing Trends Based on Ensemble of Multiple Preprocessing Techniques. *TrAC - Trends in Analytical Chemistry*, 132.
- Molina-Coronado, B., Mori, U., Mendiburu, A., & Miguel-Alonso, J. (2020). Survey of Network Intrusion Detection Methods from the Perspective of the Knowledge Discovery in Databases Process. *IEEE Transactions on Network and Service Management*, 17(4), 2451–2479.
- Mukherjee, M., & Khushi, M. (2021). Smote-enc: A Novel Smote-Based Method to Generate Synthetic Data for Nominal and Continuous Features. *Applied System Innovation*, 4(1).
- Neelakandan, S., & Paulraj, D. (2020). A Gradient Boosted Decision Tree-Based Sentiment Classification of Twitter Data. *International Journal of Wavelets, Multiresolution and Information Processing*, 18(4), 1–21.
- Obiedat, R., Al-Darras, D., Alzaghoul, E., & Harfoushi, O. (2021). Arabic Aspect-Based Sentiment Analysis: A Systematic Literature Review. *IEEE Access*, 9, 152628–152645.
- Olagunju, T., Oyeboode, O., & Orji, R. (2020). Exploring Key Issues Affecting African Mobile eCommerce Applications Using Sentiment and Thematic Analysis. *IEEE Access*, 8, 114475–114486.
- Oyeboode, O., Alqahtani, F., & Orji, R. (2020). Using Machine Learning and Thematic Analysis Methods to Evaluate Mental Health Apps Based on User Reviews. *IEEE Access*, 8, 111141–111158.

- Permata Aulia, T. M., Arifin, N., & Mayasari, R. (2021). Perbandingan Kernel Support Vector Machine (Svm) Dalam Penerapan Analisis Sentimen Vaksinisasi Covid-19. *SINTECH (Science and Information Technology) Journal*, 4(2), 139–145.
- Pratama, K. D., Brata, D. W., & Purnomo, W. (2023). Analisis Sentimen Ulasan Pengguna Aplikasi Info BMKG pada Google Play Store di Indonesia. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 7(4), 1826–1834.
- Primanda, D., Nofrini Burga, M., Juliansyah, A., & Nurfauziah, N. (2022). Efektivitas Aplikasi Info Bmkg Dalam Memberikan Informasi Cuaca Dan Bencana Terhadap Masyarakat Kota Tangerang Selatan. *Neo Politea*, 3(2), 1–9.
- Qureshi, K. A., & Sabih, M. (2021). Un-Compromised Credibility: Social Media Based Multi-Class Hate Speech Classification for Text. *IEEE Access*, 9, 109465–109477.
- Rahman, M. M., Khan, N. I., Sarker, I. H., Ahmed, M., & Islam, M. N. (2023). Leveraging Machine Learning to Analyze Sentiment from COVID-19 Tweets: A Global Perspective. *Engineering Reports*, 5(3), 1–23.
- Rajput, A. (2019). Natural language processing, sentiment analysis, and clinical analytics. *Innovation in Health Informatics: A Smart Healthcare Primer*, 79–97.
- Röder, M., Both, A., & Hinneburg, A. (2015). Exploring the space of topic coherence measures. *WSDM 2015 - Proceedings of the 8th ACM International Conference on Web Search and Data Mining*, 399–408.
- Sabarmathi G, Chinnaiyan, R., & Head, &. (2021). Sentiment Analysis for Evaluating the Patient Medicine Satisfaction. *International Journal of Computational Intelligence in Control Copyrights @Muk Publications*, 13(2), 113–118.
- Sabeh, H. N., Husin, M. H., Kee, D. M. H., Baharudin, A. S., & Abdullah, R. (2021). A Systematic Review of the DeLone and McLean Model of

- Information Systems Success in an E-Learning Context (2010-2020). *IEEE Access*, 9, 81210–81235.
- Setiawan, I. N., Krismawati, D., Pramana, S., & Tanur, E. (2022). Klasterisasi Wilayah Rentan Bencana Alam Berupa Gerakan Tanah Dan Gempa Bumi Di Indonesia. *Seminar Nasional Official Statistics*, 2022(1), 669–676.
- Shahraki, A., Abbasi, M., & 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(June), 103770.
- Sibitenda, H., Member, G. S., Diattara, A. W. A., & Ba, C. (2024). Extracting Semantic Topics About Development in Africa From Social Media. *IEEE Access*, 12(August), 142343–142359.
- Slamet Rusydiana, A., & Marlina, L. (2020). Analisis Sentimen terkait Sertifikasi Halal. *JEBA (Journal of Economics and Business Aseanomics)*, 5(1), 69–85.
- Stewart, G., & Al-Khassaweneh, M. (2022). An Implementation of the HDBSCAN Clustering Algorithm. *Applied Sciences (Switzerland)*, 12(5), 1–21.
- Sun, R., Wang, G., Zhang, W., Hsu, L. T., & Ochieng, W. Y. (2020). A Gradient Boosting Decision Tree Based GPS Signal Reception Classification Algorithm. *Applied Soft Computing Journal*, 86(November).
- Sun, Y., Lank, E., & Terry, M. (2017). Label-and-Learn: Visualizing the Likelihood of Machine Learning Classifier's Success During Data Labeling. *International Conference on Intelligent User Interfaces, Proceedings IUI*, 523–534.
- Syahri, R., & Sahrul, S. (2019). Analisis Kesiapsiagaan Masyarakat Pada Wilayah Rentan Bencana Banjir (Studi Di Kelurahan Paruga Kecamatan Rasana`E Barat Kota Bima). *Jurnal Ilmu Administrasi Negara*, 16(1), 71–81.
- Syahrial, S., Perucha, R., & Afidh, F. (2024). *Fine-Tuning Topic Modelling : A Coherence-Focused Analysis of Correlated Topic Models*. 2(2), 82–87.

- Syarif, A., Alimuddin, A., & Asrul, L. (2014). Pentingnya Komunikasi dan Informasi Pada Implementasi Kebijakan Penyelenggaraan Penanggulangan Bencana di Kota Makassar. *Jurnal Komunikasi KAREBA*, 3(3), 142–152.
- Tan, M. L., Prasanna, R., Stock, K., Doyle, E. E. H., Leonard, G., & Johnston, D. (2020). Modified Usability Framework for Disaster Apps: A Qualitative Thematic Analysis of User Reviews. *International Journal of Disaster Risk Science*, 11(5), 615–629.
- Tiwari, D., Nagpal, B., Bhati, B. S., Mishra, A., & Kumar, M. (2023). A Systematic Review of Social Network Sentiment Analysis with Comparative Study of Ensemble-Based Techniques. In *Artificial Intelligence Review* (Vol. 56, Issue 11). Springer Netherlands.
- Uwiragiye, E., & Rhinehardt, K. L. (2022). TFIDF-Random Forest: Prediction of Aptamer-Protein Interacting Pairs. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 19(5), 3032–3037.
- Vermeulen, M., Smith, K., Eremin, K., Rayner, G., & Walton, M. (2021). Application of Uniform Manifold Approximation and Projection (UMAP) in Spectral Imaging of Artworks. *Spectrochimica Acta - Part A: Molecular and Biomolecular Spectroscopy*, 252(February).
- Wong, T. T., & Yeh, P. Y. (2020). Reliable Accuracy Estimates from k-Fold Cross Validation. *IEEE Transactions on Knowledge and Data Engineering*, 32(8), 1586–1594.
- Wu, W. T., Li, Y. J., Feng, A. Z., Li, L., Huang, T., Xu, A. D., & Lyu, J. (2021). Data Mining in Clinical Big Data: the Frequently Used Databases, Steps, and Methodological Models. *Military Medical Research*, 8(1), 1–12.
- Yang, Z. (2020). Sentiment Analysis of Movie Reviews Based on Machine Learning. *ACM International Conference Proceeding Series*, 1–4.
- Zamzam, Y. F., Saragih, T. H., Herteno, R., Muliadi, Nugrahadi, D. T., & Huynh, P. H. (2024). Comparison of CatBoost and Random Forest Methods for Lung Cancer Classification using Hyperparameter Tuning Bayesian Optimization-

based. *Journal of Electronics, Electromedical Engineering, and Medical Informatics*, 6(2), 125–136.