

DAFTAR PUSTAKA

- [1] S. Dermawan and A. T. Ayunda, "Sentiment Analysis of Coretax on Social Media X Using Naive Bayes, SVM, and LSTM for Service Improvement," *Journal of Applied Informatics and Computing (JAIC)*, vol. 9, no. 6, pp. 3177–3190, Dec. 2025.
- [2] M. S. Erstiawan, "Modernisasi Perpajakan Indonesia Dengan Aplikasi Coretax: Perspektif Content Analysis," *Majalah Ekonomi: Telaah Manajemen, Akuntansi dan Bisnis*, vol. 31, no. 1, pp. 1–17, Jun. 2025.
- [3] C. Korat and A. Munandar, "Penerapan Core Tax Administration System (CTAS) Langkah Meningkatkan Kepatuhan Perpajakan di Indonesia," *Jurnal Riset Akuntansi Politala*, vol. 8, no. 1, pp. 17–30, Mar. 2025.
- [4] N. Khan, B. Zaman, and S. Ali, "Evaluation of Social Media Influence on Public Administration," *International Journal of Contemporary Issues in Social Sciences*, pp. 1536–1545, 2024.
- [5] A. Putra, B. Santoso, dan R. Wijaya, "Analisis Sentimen Menggunakan IndoBERT pada Media Sosial," *Jurnal Teknologi Informasi*, vol. 5, no. 2, pp. 120–128, 2021.
- [6] M. R. A. Rambe, I. Zufria, and M. I. Rifki, "Analisis Sentimen Masyarakat pada Platform Media Sosial X (Twitter) terhadap Pelantikan Kabinet Merah Putih Menggunakan Bernoulli Naïve Bayes," *Device: Journal of Information System, Computer Science and Information Technology*, vol. 6, no. 1, pp. 83–102, Jun. 2025.
- [7] A. A. Rohman and I. Verawati, "Perbandingan Kinerja Model Transformer Monolingual Dalam Analisis Sentimen Terhadap Website Coretax Pada Platform X," *IJCSR: The Indonesian Journal of Computer Science Research*, vol. 5, no. 1, pp. 24–36, Jan. 2026, doi: 10.37905.
- [8] D. K. Sumartha, "Implementasi IndoBERT untuk Analisis Sentimen Opini Publik terhadap Kebijakan Kenaikan UKT di Era Pemerintahan," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 13, no. 3S1, 2025, doi:10.23960/jitet.v13i3S1.7880.
- [9] A. A. Azhari, A. M. Hanif, dan M. Y. Fadilah, "TikTok sebagai Media Sosial dalam Melakukan Kritik terhadap Pembangunan di Lampung," *Indonesian Journal of Social Development*, vol. 1, no. 2, pp. 1–13, 2024.
- [10] M. Rizal, N. Permana, and F. Qalbia, "Transformasi Sistem Perpajakan di Era Digital: Tantangan, Inovasi, dan Kebijakan Adaptif," *CITIZEN: Jurnal Ilmiah Multidisiplin Indonesia*, vol. 4, no. 4, pp. 340–348, 2024, doi: 10.53866/jimi.v4i4.648.
- [11] E. M. Arbiani, A. Juraidah, A. Almaidah, et al., "Peran Media Sosial sebagai Sarana Komunikasi Pendidikan Bagi Mahasiswa di Era Digital," *Jurnal Al-Qolamuna: Jurnal Komunikasi dan Penyiaran Islam*, vol. 2, no. 4, pp. 477–489, 2025.

- [12] M. J. Siddiq, S. Jayasri, A. Suhendi, T. Hidayat, dan R. Rizky, "Analisis Sentimen Opini Masyarakat terhadap Pilkada 2024 di Media Sosial Twitter Menggunakan Algoritma Naïve Bayes," *JITET (Jurnal Informatika dan Teknik Elektro Terapan)*, vol. 13, no. 2, pp. 609–622, 2025.
- [13] V. Valiant and S. Paramita, "Peran Aplikasi TikTok sebagai Media Komunikasi Ekspresif Penyintas Covid-19," *Kiwari*, vol. 1, no. 3, pp. 558–565, 2022.
- [14] R. Bur, F. Ayuningtyas, and M. A. Muqsith, "Pemanfaatan TikTok sebagai Media Informasi Baru Generasi Z," *Jurnal Komunikasi Nusantara*, vol. 5, no. 2, pp. 189–198, 2023, doi: 10.33366/jkn.v5i2.260.
- [15] Z. Cheng and Y. Li, "Like, Comment, and Share on TikTok: Exploring the Effect of Sentiment and Second-Person View on the User Engagement with TikTok News Videos," *Social Science Computer Review*, vol. 42, no. 1, pp. 201–223, 2024, doi: 10.1177/08944393231178603.
- [16] A. Bustamin, A. A. Prayogi, D. Siswanto, M. Rafrin, and A. Nurdin, "Text Normalization for Indonesian Slang Words in Sentiment Analysis Development," *ICIC Express Letters, Part B: Applications*, vol. 16, no. 2, pp. 121–129, 2025.
- [17] C. Suhaeni, S. A. Kamila, F. Fahira, M. Yusran, and G. A. Dito, "Exploring a Large Language Model on the ChatGPT Platform for Indonesian Text Preprocessing Tasks," *Indonesian Journal of Statistics and Its Applications*, vol. 9, no. 1, pp. 100–116, 2025.
- [18] T. Kolajo, O. Daramola, A. Adebisi, and A. Seth, "A Framework for Pre-processing of Social Media Feeds Based on Integrated Local Knowledge Base," *Information Processing & Management*, vol. 57, no. 6, Art. no. 102348, 2020.
- [19] R. Budianoor, S. W. Saputro, F. Abadi, R. A. Nugroho, and A. Farmadi, "Quantifying the Impact of Text Preprocessing on IndoBERT Fine-Tuning for Indonesian Informal Culinary Sentiment Analysis," *Journal of Computing Theories and Applications*, 2025.
- [20] E. Saraswati and Muljono, "Deteksi Emosi pada Twitter Berbasis FastText: Evaluasi Performa Arsitektur CNN dan GRU," *Rabit: Jurnal Teknologi dan Sistem Informasi Univrab*, 2025.
- [21] V. S. Dhaka and S. K. Singh, "Web Crawler: A Review," *International Journal of Computer Applications*.
- [22] T. M. Fahrudin et al., "Building a Technology Recommender System Using Web Crawling and NLP," *Algorithms*, 2022.
- [23] A. Z. Rizquina, "Implementasi Web Scraping untuk Pengambilan Data Website," *Jurnal Teknologi dan Sistem Informasi*, 2023.
- [24] Apify, "Apify Platform Documentation," 2026. [Online]. Available: <https://docs.apify.com/platform>. [Accessed: 23-Jul-2026].

- [25] M. A. Khder, "Web Scraping or Web Crawling: State of Art, Techniques, Approaches and Application," *Int. J. Advance Soft Computing Applications*, vol. 13, no. 3, Nov. 2021, doi: 10.15849/IJASCA.211128.11.
- [26] J. Jünger, "Scraping Social Media Data as Platform Research," 2023. Z. Tan et al., "Large Language Models for Data Annotation and Synthesis," *Proceedings of EMNLP*, 2024.
- [27] T. Young, D. Hazarika, S. Poria, and E. Cambria, "Recent Trends in Deep Learning Based Natural Language Processing," *IEEE Computational Intelligence Magazine*, vol. 13, no. 3, pp. 55–75, Aug. 2018. doi: 10.1109/MCI.2018.2840738.
- [28] T. Wolf et al., "Transformers: State-of-the-Art Natural Language Processing," *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, 2021.
- [29] R. K. Das, S. Dash, and S. Mishra, "Comparative Analysis of Machine Learning and Hybrid Deep Learning Models for Sentiment Analysis," *PeerJ Computer Science*, vol. 9, 2023.
- [30] M. Sivakumar, "Trade-off between Training and Testing Ratio in Machine Learning for Medical Image Processing," *PeerJ Computer Science*, vol. 10, e2245, 2024. doi: 10.7717/peerj-cs.2245. [32] B. Liu, *Sentiment Analysis: Mining Opinions, Sentiments, and Emotions*, Cambridge University Press, 2022.
- [31] B. Liu, *Sentiment Analysis: Mining Opinions, Sentiments, and Emotions*, Cambridge University Press, 2022.
- [32] E. Cambria, D. Das, S. Bandyopadhyay, and A. Feraco, *A Practical Guide to Sentiment Analysis*. Springer, 2017. <https://link.springer.com/book/10.1007/978-3-319-55394-8> A.
- [33] Giachanou and F. Crestani, "Like It or Not: A Survey of Twitter Sentiment Analysis Methods," *ACM Computing Surveys*, vol. 49, no. 2, 2016. <https://dl.acm.org/doi/10.1145/2938640>
- [34] M. R. Manoppo, I. C. Kolang, D. N. Fiat, R. M. C. Mawara, A. D. P. Sumarno, A. Yusupa, dan V. Tarigan, "Analisis Sentimen Publik di Media Sosial terhadap Kenaikan PPN 12% di Indonesia Menggunakan IndoBERT," *Jurnal Kecerdasan Buatan dan Teknologi Informasi*, vol. 4, no. 2, pp. 152–163, Mei 2025, doi: 10.69916/jkbt.v4i2.322.
- [35] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics (NAACL-HLT)*, Minneapolis, USA, pp. 4171–4186, 2019.

- [36] J. Ohme, "Digital Trace Data Collection for Social Media Effects Research," *Communication Methods and Measures*, vol. 17, no. 2, pp. 89–105, 2023.
- [37] K. Kowsari, K. J. Meimandi, M. Heidarysafa, S. Mendu, L. Barnes, and D. Brown, "Text Classification Algorithms: A Survey," *Information*, vol. 10, no. 4, pp. 1–68, 2021.
- [38] T. Kolajo, O. Daramola, and A. Adebiyi, "A Framework for Pre-processing of Social Media Data Using Integrated Local Knowledge Base," *Information Processing & Management*, vol. 57, no. 6, Art. no. 102348, 2020, doi: 10.1016/j.ipm.2020.102348.
- [39] A. Sharma and P. Gupta, "Web Data Preprocessing Techniques for Text Mining Applications," *Journal of Data Science and Artificial Intelligence*, vol. 3, no. 2, pp. 45–53, 2022.
- [40] D. Thakur, J. Singh, G. Dhiman, M. Shabaz, and T. Gera, "Identifying Major Research Areas and Minor Research Themes of Android Malware Analysis and Detection Field Using LSA," *Complexity*, vol. 2021, Article ID 4551067, pp. 1–28, Sep. 2021, doi: 10.1155/2021/4551067.
- [41] M. M. Karim et al., "Transforming Data Annotation with AI Agents: A Review of Architectures, Reasoning, Applications, and Impact," *Future Internet*, vol. 17, no. 8, 2025. [Online]. Available: <https://www.mdpi.com/1999-5903/17/8/353>
- [42] A. Romadhony et al., "Sentiment Analysis on a Large Indonesian Product Review Dataset," *Journal of Information Systems Engineering and Business Intelligence*, vol. 10, no. 1, 2024.
- [43] M. Barnes, P. Oberländer, dan S. Rosenthal, "SemEval-2022 Task 10: Structured Sentiment Analysis," *Proceedings of the 16th International Workshop on Semantic Evaluation (SemEval-2022)*, 2022.
- [44] W. van Atteveldt, M. A. C. G. van der Velden, and M. Boukes, "The Validity of Sentiment Analysis: Comparing Manual Annotation, Crowd-Coding, Dictionary Approaches, and Machine Learning Algorithms," *Communication Methods and Measures*, vol. 15, no. 2, pp. 121–140, 2021, doi: 10.1080/19312458.2020.1869198.
- [45] A. Baledent, Y. Mathet, A. Widlöcher, C. Couronne, and J.-L. Manguin, "Validity, Agreement, Consensuality and Annotated Data Quality," in *Proceedings of the Language Resources and Evaluation Conference (LREC)*, Marseille, France, 2022, pp. 2940–2948.
- [46] Y. Guo, S. Das, S. Lakamana, and A. Sarker, "An Aspect-Level Sentiment Analysis Dataset for Therapies on Twitter," *Data in Brief*, vol. 50, Art. no. 109618, 2023, doi: 10.1016/j.dib.2023.109618.
- [47] B. Więckowska, K. B. Kubiak, P. Józwiak, W. Moryson, and B. Stawińska-Witoszyńska, "Cohen's Kappa Coefficient as a Measure to Assess Classification Improvement following the Addition of a New Marker to a Regression Model," *Int.*

- J. Environ. Res. Public Health, vol. 19, no. 16, Art. no. 10213, 2022, doi: 10.3390/ijerph191610213.
- [48] I. T. Julianto et al., "Automatic Sentiment Annotation Using Grok AI for Opinion Classification," JISTICS Journal, 2025
- [49] M. Jafarlou et al., "Reducing Labeling Costs in Sentiment Analysis via Semi-Supervised Learning," ACM Digital Library, 2024.
- [50] V. R. Joseph, "Optimal ratio for data splitting," Statistical Analysis and Data Mining: The ASA Data Science Journal, vol. 15, no. 4, pp. 531–538, 2022.
- [51] M. Sadikin, R. Sari, and A. Nugroho, "Sentiment Analysis Using Machine Learning: A Comparative Study," European Journal of Electrical Engineering and Computer Science, vol. 7, no. 2, pp. 21–27, 2023.
- [52] R. K. Das, S. Dash, and S. Mishra, "Comparative analysis of machine learning and hybrid deep learning models for sentiment analysis," PeerJ Computer Science, vol. 9, 2023.
- [53] V. R. Joseph and A. Vakayil, "SPlit: An Optimal Method for Data Splitting," Technometrics, vol. 64, no. 2, pp. 166–176, 2022, doi: 10.1080/00401706.2021.1921037.
- [54] B. Bakirarar and A. H. Elhan, "Class Weighting Technique to Deal with Imbalanced Class Problem in Machine Learning: Methodological Research," Turkiye Klinikleri Journal of Biostatistics, vol. 15, no. 1, pp. 19–29, 2023, doi: 10.5336/biostatic.2022-93961
- [55] A. F. Al Farizi and Y. Sibaroni, "Implementation of BiLSTM and IndoBERT for Sentiment Analysis of TikTok Reviews," JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika), vol. 10, no. 1, pp. 96–106, Mar. 2025, doi: 10.29100/jipi.v10i1.5815.
- [56] A. Zhang, Z. C. Lipton, M. Li, and A. Smola, Dive into Deep Learning. Cambridge, MA, USA: MIT Press, 2021. Link: <https://d2l.ai>
- [57] F. Koto, A. Rahimi, J. H. Lau, and T. Baldwin, "IndoLEM and IndoBERT: A Benchmark Dataset and Pre-trained Language Model for Indonesian NLP," Proceedings of the 28th International Conference on Computational Linguistics, pp. 757–770, 2020. <https://aclanthology.org/2020.coling-main.66/>
- [58] M. F. Kono, I. N. Fajri, dan Y. Pristyanto, "Public Sentiment Analysis on Corruption Issues in Indonesia Using IndoBERT Fine-Tuning, Logistic Regression, and Linear SVM," *Journal of Applied Informatics and Computing (JAIC)*, vol. 9, no. 5, pp. 2616–2628, Oct. 2025.
- [59] A. Wicaksono and A. Purwarianti, "Indonesian Pre-trained Language Model for Sentiment Analysis," Proceedings of the Asian Conference on Machine Learning, 2021. <https://arxiv.org/abs/2109.07932>

- [60] B. Wilie et al., "IndoNLU: Benchmark and Resources for Evaluating Indonesian Natural Language Understanding," arXiv preprint arXiv:2009.05387, 2020.
- [61] N. P. D. Agustina, I. D. A. P. P. Tentriajaya, dan I. G. N. L. Wijayakusuma, "Public Sentiment Analysis on Demonstration Actions Using IndoBERT Based on Transfer Learning," *Journal of Applied Informatics and Computing (JAIC)*, vol. 9, no. 6, pp. 969–974, Dec. 2025.
- [62] B. M. R. E. Nilapaksi, T. N. Fatyanosa, and P. P. Adikara, "Analisis Sentimen Publik Terhadap Magang Berdampak 2025 di Platform X/Twitter Menggunakan Model IndoBERT," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 10, no. 4, pp. x–x, Apr. 2026.
- [63] T. Wolf, L. Debut, V. Sanh, et al., "Transformers: State-of-the-Art Natural Language Processing," in *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, 2020, pp. 38–45, doi: 10.18653/v1/2020.emnlp-demos.6.
- [64] A. Géron, *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*, 2nd ed. Sebastopol, CA: O'Reilly Media, 2019.
- [65] S. Kulkarni and V. R. Kshirsagar, "Confusion Matrix-Based Performance Evaluation Metrics," *International Journal of Advanced Computer Science and Applications*, vol. 15, no. 1, pp. 1–7, 2024.
- [66] M. R. Pratama, F. A. Gunawan S., R. R. Zhafari, Rendy, dan H. N. Imanda, "Sentiment Analysis of Beauty Product E-Commerce Using Support Vector Machine Method," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 6, no. 2, pp. 269–274, 2022.
- [67] C. Bao and Y. Wang, "A Survey of Word Cloud Visualization," *Journal of Computer-Aided Design & Computer Graphics*, vol. 33, no. 4, pp. 532–544, 2021.
- [68] U. Torres Parejo, J. R. Campaña, M. A. Vila, and M. Delgado, "A Survey of Tag Clouds as Tools for Information Retrieval and Content Representation," *Information Visualization*, vol. 20, no. 1, pp. 83–97, 2021.