

DAFTAR PUSTAKA

- Aji, A. F., Winata, G. I., Koto, F., Cahyawijaya, S., Romadhony, A., Mahendra, R., Kurniawan, K., Moeljadi, D., Prasajo, R. E., Baldwin, T., Lau, J. H., & Ruder, S. (2022). One country, 700+ languages: NLP challenges for underrepresented languages and dialects in Indonesia. *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 7226–7249.
- Alvin, F., & Winarsih, N. A. S. (2025). Perbandingan Kinerja Model IndoBERT, IndoBERTweet, dan Algoritma Klasik pada Analisis Sentimen Isu Indonesia Gelap. *Building of Informatics, Technology and Science (BITS)*, 7(3), 1601–1613. <https://doi.org/10.47065/bits.v7i3.8636>
- Apriansyah, F. M., Ramadhan, T. I., Hidayat, C. R., & Karta, A. (2025). Perbandingan IndoBERT dan IndoRoBERTa Untuk Analisis Sentimen Pada Film Dokumenter Dirty Vote. *Jurnal Pengembangan IT*, 10(3), 593–605. <https://doi.org/10.30591/jpit.v10i3.8607>
- Aras, S., Yusuf, M., Ruimassa, R., Agustinus, E., Wambrauw, B., & Palalangan, E. B. (2024). Sentiment Analysis on Shopee Product Reviews Using IndoBERT. *Journal of Information Systems and Informatics*, 6(3), 1616–1627. <https://doi.org/10.51519/journalisi.v6i3.814>
- Ariyani, P. W., Sunarya, I. M. G., & Gunadi, I. G. A. (2025). Analisis Sentimen Masyarakat Terhadap Virus Corona Berdasarkan Opini dari Twitter Menggunakan Metode Naïve Bayes dan K-Nearest Neighbor. *Jurnal Pendidikan Teknologi dan Kejuruan*, 22(2), 128–138. <https://doi.org/10.23887/jptk-undiksha.v22i2.103233>
- Arka, I. W. (2003). *Balinese Morphosyntax: a Lexical-Functional Approach*. Pacific Linguistics, Research School of Pacific and Asian Studies, The Australian National University. <https://doi.org/10.13140/RG.2.1.4828.1202/1>
- Artalia, N. K. A., Maysanjaya, I. M. D., & Suputra, P. H. (2026). Sentiment Analysis Of Bali's Rapid Tourism Growth Using IndoBERT With InSet Lexicon Automatic Labelling. *KARMAPATI (Kumpulan Artikel Mahasiswa Pendidikan Teknik Informatika)*, 15(1), 194–201. <https://doi.org/10.23887/karmapati.v15i1.111839>
- Asshiddiq, M. H., & Witanti, A. (2025). Analisis Sentimen tentang Penundaan Pengangkatan CPNS 2025 pada Platform X Menggunakan Metode IndoBERT. *Jurnal teknika (Jurnal Fakultas Teknik Universitas Islam Lamongan)*, 17(2), 97–108. <https://doi.org/10.30736/jt.v17i2.1448>
- Birjali, M., Kasri, M., & Beni-Hssane, A. (2021). A Comprehensive Survey On Sentiment Analysis: Approaches, Challenges and Trends. *Knowledge-Based Systems*, 226, 107134. <https://doi.org/10.1016/j.knosys.2021.107134>

- Conneau, A., Khandelwal, K., Goyal, N., Chaudhary, V., Wenzek, G., Guzmán, F., Grave, E., Ott, M., Zettlemoyer, L., & Stoyanov, V. (2020). Unsupervised Cross-lingual Representation Learning at Scale. *Proceedings of the 58th annual meeting of the association for computational linguistics*, 8440–8451. <https://doi.org/10.18653/v1/2020.acl-main.747>
- Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). 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: human language technologies, volume 1 (long and short papers)*, 4171–4186. <https://doi.org/10.48550/arXiv.1810.04805>
- Diwali, A., Saeedi, K., Dashtipour, K., Gogate, M., Cambria, E., & Hussain, A. (2023). Sentiment Analysis Meets Explainable Artificial Intelligence: A Survey on Explainable Sentiment Analysis. *IEEE Transactions on Affective Computing*, 15(3), 837–846. <https://doi.org/10.1109/TAFFC.2023.3296373>
- Dodge, J., Ilharco, G., Schwartz, R., Farhadi, A., Hajishirzi, H., & Smith, N. (2020). Fine-tuning pretrained language models: Weight initializations, data orders, and early stopping. *arXiv preprint arXiv:2002.06305*.
- Eberhard, D. M., Simons, G. F., & Fennig, C. D. (2025). *Ethnologue: Languages of the World* (28 ed.). SIL International: <http://www.ethnologue.com>
- Fagerland, M. W., Lydersen, S., & Laake, P. (2013). The McNemar Test for Binary Matched-Pairs Data: Mid-p and Asymptotic are Better than Exact Conditional. *BMC Medical Research Methodology*, 13(1), 91. <https://doi.org/10.1186/1471-2288-13-91>
- Guest, G., Bunce, A., & Johnson, L. (2006). How Many Interviews Are Enough?: An Experiment with Data Saturation and Variability. *Field Methods*, 18(1), 59–82. <https://doi.org/0.1177/1525822X05279903>
- Hakim, G., Fatyanosa, T. N., & Widodo, A. W. (2024). Analisis Sentimen Masyarakat terhadap Kereta Cepat Whoosh pada Platform X menggunakan IndoBERT. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 8(10), 1–10.
- Jim, J. R., Talukder, M. A. R., Malakar, P., Kabir, M. M., Nur, K., & Mridha, M. F. (2024). Recent advancements and challenges of NLP-based sentiment analysis: A state-of-the-art review. *Natural Language Processing Journal*, 6, 100059. <https://doi.org/10.1016/j.nlp.2024.100059>
- Jurafsky, D., & Martin, J. H. (2023). *Speech and Language Processing. An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. Prentice Hall. <https://books.google.co.id/books?id=b3PuzwEACAAJ>
- Kokalj, E., Škrlić, B., Lavrač, N., Pollak, S., & Robnik-Šikonja, M. (2021). BERT meets Shapley: Extending SHAP Explanations to Transformer-based

- Classifiers. *Proceedings of the EACL hackashop on news media content analysis and automated report generation*, 16–21.
- Koto, F., Rahimi, A., Lau, J. H., & Baldwin, T. (2020). IndoLEM and IndoBERT: A Benchmark Dataset and Pre-trained Language Model for Indonesian NLP. *Proceedings of the 28th International Conference on Computational Linguistics*, 757–770. <https://doi.org/10.18653/v1/2020.coling-main.66>
- Kudo, T. (2018). Subword Regularization: Improving Neural Network Translation Models with Multiple Subword Candidates. In I. Gurevych & Y. Miyao (Ed.), *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (hal. 66–75). Association for Computational Linguistics. <https://doi.org/10.18653/v1/P18-1007>
- Kudo, T., & Richardson, J. (2018). SentencePiece: A simple and language independent subword tokenizer and detokenizer for Neural Text Processing. *Proceedings of the 2018 conference on empirical methods in natural language processing: System demonstrations*, 66–71. <https://doi.org/10.18653/v1/D18-2012>
- Liu, B. (2012). Sentiment Analysis and Opinion Mining. *Morgan & Claypool Publishers*.
- Lundberg, S. M., & Lee, S.-I. (2017). A Unified Approach to Interpreting Model Predictions. *Proceedings of the 31st International Conference on Neural Information Processing Systems*, 4768–4777. <https://doi.org/10.48550/arXiv.1705.07874>
- Lydersen, S., Fagerland, M. W., & Laake, P. (2009). Recommended Tests for Association in 2 x 2 Tables. *Statistics in Medicine*, 28(7), 1159–1175. <https://doi.org/10.1002/sim.3531>
- McNemar, Q. (1947). Note on the Sampling Error of the Difference Between Correlated Proportions or Percentages. *Psychometrika*, 12(2), 153–157. <https://doi.org/10.1007/BF02295996>
- Mosca, E., Szigeti, F., Tragianni, S., Gallagher, D., & Groh, G. (2022). SHAP-Based Explanation Methods: A Review for NLP Interpretability. *Proceedings of the 29th international conference on computational linguistics*, 4593–4603.
- P. Sosiawan, Martha, I. N., & Artika, I. W. (2021). Penggunaan Bahasa Bali pada Keluarga Muda di Kota Singaraja. *Jurnal Pendidikan dan Pembelajaran Bahasa Indonesia*, 10(1). https://doi.org/10.23887/jurnal_bahasa.v10i1.403
- Parawangsa, I. D. G. M., Mahendra, G. S., & Pinatih, R. N. P. (2026). Implementasi Metode XLM-RoBERTa Untuk Analisis Sentimen Ulasan Pengunjung Objek Wisata Kerta Gosa Klungkung. *RIGGS: Journal of Artificial Intelligence and Digital Business*, 4(4), 9893–9900. <https://doi.org/10.31004/riggs.v4i4.5368>
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos,

- A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830. <https://doi.org/10.48550/arXiv.1201.0490>
- Petrov, A., La Malfa, E., Torr, P., & Bibi, A. (2023). Language Model Tokenizers Introduce Unfairness Between Languages. *Advances in neural information processing systems*, 36, 36963–36990. <https://doi.org/10.48550/arXiv.2305.15425>
- Pires, T., Schlinger, E., & Garrette, D. (2019). How Multilingual is Multilingual BERT? *Proceedings of the 57th annual meeting of the association for computational linguistics*, 4996–5001. <https://doi.org/10.18653/v1/P19-1493>
- Prabowo, A., & Sanjaya, F. I. (2024). PENERAPAN METODE TRANSFER LEARNING PADA INDOBERT UNTUK ANALISIS SENTIMEN TEKS BAHASA JAWA NGOKO LUGU. *Jurnal Sistem Informasi dan Sistem Komputer*, 9(2), 205–217. <https://doi.org/10.51717/simkom.v9i2.478>
- Pramayasa, K., Maysanjaya, I. M. D., & Indradewi, I. G. A. A. D. (2023). Analisis Sentimen Program Mbkm pada Media Sosial Twitter Menggunakan KNN dan SMOTE. *SINTECH (Science and Information Technology) Journal*, 6(2), 89–98. <https://doi.org/10.31598/sintechjournal.v6i2.1372>
- Purwarianti, A., & Crisdayanti, I. A. P. A. (2019). Improving bi-lstm performance for indonesian sentiment analysis using paragraph vector. *2019 International Conference of Advanced Informatics: Concepts, Theory and Applications (ICAICTA)*, 1–5.
- Qolbu, A. A., Fitriyati, N., & Inayah, N. (2025). Performa Naïve Bayes, SVM, dan IndoBERT pada Analisis Sentimen Twitter IndiHome dengan Strategi Penanganan Data Tidak Seimbang. *Jurnal Fourier*, 14(1), 29–44. <https://doi.org/10.14421/fourier.2025.141.29-44>
- Rainita, N. P. A., Maysanjaya, I. M. D., & Mahendra, G. S. (2025). Perbandingan Performa Algoritma Naive Bayes dan Support Vector Machine pada Analisis Sentimen Implementasi Program KIP-Kuliah. *Jurnal TEKNOSIA*, 19(2), 76–87.
- Rust, P., Pfeiffer, J., Vulić, I., Ruder, S., & Gurevych, I. (2021). How Good is Your Tokenizer? On the Monolingual Performance of Multilingual Language Models. *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, 3118–3135. <https://doi.org/10.18653/v1/2021.acl-long.243>
- Santhi, T., Maysanjaya, I. M. D., & Wijaya, I. N. S. W. (2026). Perancangan Aplikasi Analisis Sentimen Mengenai Utang Negara Indonesia Berbasis LSTM. *KARMAPATI (Kumpulan Artikel Mahasiswa Pendidikan Teknik Informatika)*, 15(1), 30–40. <https://doi.org/10.23887/karmapati.v15i1.108664>

- Saputri, N. K. T. A., Gunadi, I. G. A., & Sunarya, I. M. G. (2024). Perbandingan Metode Analisis Sentimen Pelayanan Daring di Fakultas Teknik dan Kejuruan Universitas Pendidikan Ganesha Menggunakan Algoritma Naïve Bayes dan LSTM: Sentiment Analysis of Online Services at the Engineering and Vocational Faculty of Ganesha Education University Using Naïve Bayes and LSTM Algorithms. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 4(3), 1120–1129. <https://doi.org/10.57152/malcom.v4i3.1336>
- Sari, A. M., Sunarya, I. M. G., & Maysanjaya, I. M. D. (2026). Perancangan Aplikasi Analisis Sentimen Mengenai Program Petani Milenial Berbasis IndoBERT. *KARMAPATI (Kumpulan Artikel Mahasiswa Pendidikan Teknik Informatika)*, 15(1), 118–126. <https://doi.org/10.23887/karmapati.v15i1.108665>
- Sennrich, R., Haddow, B., & Birch, A. (2016). Neural Machine Translation of Rare Words with Subword Units. *Proceedings of the 54th annual meeting of the association for computational linguistics (volume 1: long papers)*, 1715–1725. <https://doi.org/10.18653/v1/P16-1162>
- Setiawan, V. D., Iswavigra, D. U., & Anggiratih, E. (2025). Implementation of IndoBERT for Sentiment Analysis of the Constitutional Court's Decision Regarding the Minimum Age of Vice Presidential Candidates. *Scientific Journal of Informatics*, 12(3), 397–406. <https://doi.org/10.15294/sji.v12i3.26360>
- Sidik, P., Sunarya, I. M. G., & Gunadi, I. G. A. (2025). Comparison of Random Forest and Support Vector Machine Methods in Sentiment Analysis of Student Satisfaction Questionnaire Comments at ITB STIKOM Bali. *Journal of Applied Informatics and Computing*, 9(3), 794–802. <https://doi.org/10.30871/jaic.v9i3.9617>
- Tarwoto, Nugroho, R., Azka, N., & Graha, W. S. R. (2025). Analisis Sentimen Ulasan Aplikasi Mobile JKN di Google PlayStore Menggunakan IndoBERT. *Jurnal JTik (Jurnal Teknologi Informasi dan Komunikasi)*, 9(2), 495–505. <https://doi.org/10.35870/jtik.v9i2.3340>
- UNESCO. (2022). *Global action plan of the International Decade of Indigenous Languages (IDIL2022–2032)*. UNESCO.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention Is All You Need. *Advances in neural information processing systems*, 30. <https://doi.org/10.48550/arXiv.1706.03762>
- Wibawa, P. W. A., & Pramarta, C. (2023). Analisis Sentimen pada Teks Berbahasa Bali Menggunakan Metode Multinomial Naive Bayes dengan TF-IDF dan BoW. *Jurnal Nasional Teknologi Informasi Dan Aplikasinya*, 2(1),

- 37–46. <https://doi.org/10.24843/JNATIA.2023.v02.i01.p05>
- Wijaya, I. N. S. W., Seputra, K. A., & Dewi, N. P. N. P. (2025). Fine Tuning Model IndoBERT untuk Analisis Sentimen Berita Pariwisata Indonesia. *Jurnal Pendidikan Teknologi dan Kejuruan*, 22(2), 195–204. <https://doi.org/10.23887/jptk-undiksha.v22i2.104056>
- Wilie, B., Vincentio, K., Winata, G. I., Cahyawijaya, S., Li, X., Lim, Z. Y., Soleman, S., Mahendra, R., Fung, P., Bahar, S., & Purwarianti, A. (2020). IndoNLU: Benchmark and Resources for Evaluating Indonesian Natural Language Understanding. *Proceedings of the 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 10th International Joint Conference on Natural Language Processing*, 843–857. <https://doi.org/10.18653/v1/2020.aacl-main.85>
- Winata, G. I., Aji, A. F., Cahyawijaya, S., Mahendra, R., Koto, F., Romadhony, A., Kurniawan, K., Moeljadi, D., Prasajo, R. E., Fung, P., Baldwin, T., Lau, J. H., Sennrich, R., & Ruder, S. (2023). NusaX: Multilingual Parallel Sentiment Dataset for 10 Indonesian Local Languages. *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics*, 815–834. <https://doi.org/10.18653/v1/2023.eacl-main.57>
- Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Louf, R., Funtowicz, M., Davison, J., Shleifer, S., Platen, P. von, Ma, C., Jernite, Y., Plu, J., Xu, C., Scao, T. Le, Gugger, S., ... Rush, A. (2020). Transformers: State-of-the-Art Natural Language Processing. *Proceedings of the 2020 conference on empirical methods in natural language processing: system demonstrations*, 38–45. <https://doi.org/10.18653/v1/2020.emnlp-demos.6>