

DAFTAR PUSTAKA

- Abodayeh, A., Shihadeh, L., Hejazi, R., Latif, R., & Najjar, W. (2023). *Web Scraping for Data Analytics: A BeautifulSoup Implementation*. <https://doi.org/https://doi.org/10.1109/WiDS-PSU57071.2023.00025>
- Aliyah Salsabila, N., Ardhitto Winatmoko, Y., Akbar Septiandri, A., & Jamal, A. (2018). Colloquial Indonesian Lexicon. *2018 International Conference on Asian Language Processing (IALP)*, 226–229. <https://doi.org/https://doi.org/10.1109/IALP.2018.8629151>
- Alva Principe, R., Chiarini, N., & Viviani, M. (2025). Long Document Classification in the Transformer Era: A Survey on Challenges, Advances, and Open Issues. In *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* (Vol. 15, Number 2). John Wiley and Sons Inc. <https://doi.org/https://doi.org/10.1002/widm.70019>
- Carvalho, M., Pinho, A. J., & Brás, S. (2025). Resampling approaches to handle class imbalance: a review from a data perspective. *Journal of Big Data*, 12(1). <https://doi.org/https://doi.org/10.1186/s40537-025-01119-4>
- Clark, K., Khandelwal, U., Levy, O., & Manning, C. D. (2019). *What Does BERT Look At? An Analysis of BERT's Attention*. <https://doi.org/https://doi.org/10.48550/arXiv.1906.04341>
- Daniati, E., Wibawa, A. P., Irianto, W. S. G., & Hernandez, L. (n.d.). *Few-Shot-BERT-RNN Narrative Structure Analysis for Andersen's Stories*. <https://doi.org/http://dx.doi.org/10.62527/joiv.9.4.3932>**
- De Santis, E., Martino, A., Ronci, F., & Rizzi, A. (2025). From Bag-of-Words to Transformers: A Comparative Study for Text Classification in Healthcare Discussions in Social Media. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 9(1), 1063–1077. <https://doi.org/https://doi.org/10.1109/TETCI.2024.3423444>
- Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*. <https://doi.org/https://doi.org/10.48550/arXiv.1810.04805>

- Dhiman, P., Kaur, A., Gupta, D., Juneja, S., Nauman, A., & Muhammad, G. (2024). GBERT: A hybrid deep learning model based on GPT-BERT for fake news detection. *Heliyon*, 10(16). <https://doi.org/https://doi.org/10.1016/j.heliyon.2024.e35865>
- Hamdikatama, B. (2025). Beyond Algorithms: An Integrated Approach to Fake News Detection Using Machine Learning Techniques. *Jurnal Ilmu Pengetahuan Dan Teknologi Komputer (JITK)*, 10(3), 609–622. <https://doi.org/https://doi.org/10.33480/jitk.v10i3.6061>
- Howard, J., & Ruder, S. (2018). *Universal Language Model Fine-tuning for Text Classification*. <https://doi.org/https://doi.org/10.48550/arXiv.1801.06146>
- Hu, J., Ruder, S., Siddhant, A., Neubig, G., Firat, O., & Johnson, M. (2020). *XTREME: A Massively Multilingual Multi-task Benchmark for Evaluating Cross-lingual Generalization*. <https://doi.org/https://doi.org/10.48550/arXiv.2003.11080>
- Imron, A., Daniati, E., Firliana, R., Fauzan, M., & Mufid, A. (2025). *Analisis Data Penjualan Pada UMKM Konveksi Menggunakan Metode K-Means Clustering dengan Menerapkan CRISP-DM*. <https://doi.org/https://doi.org/10.33005/sitasi.v5i1.2542>
- Jurafsky, D., & Martin, J. H. (2025). *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition with Language Models* (3rd ed.). <https://web.stanford.edu/~jurafsky/slp3/>
- Karita, S., Chen, N., Hayashi, T., Hori, T., Inaguma, H., Jiang, Z., Someki, M., Soplin, N. E. Y., Yamamoto, R., Wang, X., Watanabe, S., Yoshimura, T., & Zhang, W. (2019). *A Comparative Study on Transformer vs RNN in Speech Applications*. <https://doi.org/https://doi.org/10.1109/ASRU46091.2019.9003750>
- Kementerian Komunikasi dan Digital. (2025). *Komdigi Identifikasi 1.923 Konten Hoaks Sepanjang Tahun 2024*. <https://www.komdigi.go.id/berita/siaran-pers/detail/komdigi-identifikasi-1923-konten-hoaks-sepanjang-tahun-2024>

- Kemp, S. (2025). *Digital 2026: Top Digital and Social Media Trends in Indonesia*.
<https://wearesocial.com/id/blog/2025/11/digital-2026-top-digital-and-social-media-trends-in-indonesia/>
- Koto, F., Rahimi, A., Lau, J. H., & Baldwin, T. (2020). *IndoLEM and IndoBERT: A Benchmark Dataset and Pre-trained Language Model for Indonesian NLP*.
<https://doi.org/https://doi.org/10.48550/arXiv.2011.00677>
- Loshchilov, I., & Hutter, F. (2019). *Decoupled Weight Decay Regularization*.
<https://doi.org/https://doi.org/10.48550/arXiv.1711.05101>
- Lu, H., Xu, Y., Ye, M., Yan, K., Gao, Z., & Jin, Q. (2019). Learning misclassification costs for imbalanced classification on gene expression data. *BMC Bioinformatics*, 20. <https://doi.org/https://doi.org/10.1186/s12859-019-3255-x>
- Masood, M. A., Abbasi, R. A., & Keong, N. W. (2020). Context-Aware Sliding Window for Sentiment Classification. *IEEE Access*, 8, 4870–4884.
<https://doi.org/https://doi.org/10.1109/ACCESS.2019.2963586>
- Nugroho, A., & Harini, D. (2024). Teknik Random Forest untuk Meningkatkan Akurasi Data Tidak Seimbang. *JSITIK*, 2(2).
<https://doi.org/https://doi.org/10.53624/jsitik.v2i2.XX>**
- Pappagari, R., Zelasko, P. , Villalba, J., Carmiel, Y., & Dehak, N. (2019). *HIERARCHICAL TRANSFORMERS FOR LONG DOCUMENT CLASSIFICATION*.
<https://doi.org/https://doi.org/10.48550/arXiv.1910.10781>
- Pekandi, L. A., Widjaja, R. G., Ananta, A., Harefa, J., & Jingga, K. (2025). Evaluating IndoBERT for Indonesian Hoax News Detection: A Comparative Study with Ensemble and CNN-LSTM Models. *Procedia Computer Science*, 269, 1625–1633. <https://doi.org/https://doi.org/10.1016/j.procs.2025.09.105>
- Peltarion. (2020). *Structure of BERT*. <https://peltarion.com/knowledge-center/documentation/modeling-view/build-an-ai-model/blocks/bert-encoder>
- Pes, B., & Lai, G. (2021). Cost-sensitive learning strategies for high-dimensional and imbalanced data: a comparative study. *PeerJ Computer Science*, 7. <https://doi.org/https://doi.org/10.7717/PEERJ-CS.832>

- Rahmat, M. A., Indrabayu, & Areni, I. S. (2019). *Hoax Web Detection For News in Bahasa Using Support Vector Machine*.
<https://doi.org/https://doi.org/10.1109/ICOIACT46704.2019.8938425>
- Ridho, M. Y., & Yulianti, E. (2024). From Text to Truth: Leveraging IndoBERT and Machine Learning Models for Hoax Detection in Indonesian News. *Jurnal Ilmiah Teknik Elektro Komputer Dan Informatika*, 10(3), 544–555.
<https://doi.org/https://doi.org/10.26555/jiteki.v10i3.29450>
- Ripa'i, A., Santoso, F., & Lazim, F. (2024). Deteksi Berita Hoax dengan Perbandingan Website Menggunakan Pendekatan Deep Learning Algoritma BERT. *G-Tech: Jurnal Teknologi Terapan*, 8(3), 1749–1758.
<https://doi.org/https://doi.org/10.33379/gtech.v8i3.4541>
- Ristyawan, A., Nugroho, A., & Amarya, T. K. (2025). Optimasi Preprocessing Model Random Forest Untuk Prediksi Stroke. *Jurnal Teknik Informatika Dan Sistem Informasi (JATISI)*, 12(1), 29–44.**
<https://doi.org/10.25008/jatisi.v12i1.9587>
- Sabir, M., Farooq Khan, T., & Azam, M. (2025). A Comparative Study of Traditional and Hybrid Models for Text Classification. In *Journal of Computers and Intelligent Systems* (Vol. 3, Number 1).
<https://journals.iub.edu.pk/index.php/JCIS/>
- Shimaoka, A. M., Ferreira, R. C., & Goldman, A. (2024). The evolution of CRISP-DM for Data Science: Methods, Processes and Frameworks. *SBC Reviews on Computer Science*, 4(1), 28–43.
<https://doi.org/https://doi.org/10.5753/reviews.2024.3757>
- Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. (2017). Fake News Detection on Social Media: A Data Mining Perspective. *SIGKDD Explor. Newsl.*, 19(1), 22–36. <https://doi.org/https://doi.org/10.1145/3137597.3137600>
- Sucipto, S., Dwi Prasetya, D., & Widiyaningtyas, T. (2024). Educational Data Mining: Multiple Choice Question Classification in Vocational School. *MATRIK: Jurnal Manajemen, Teknik Informatika Dan Rekayasa Komputer*, 23(2), 379–388.**
<https://doi.org/https://doi.org/10.30812/matrik.v23i2.3499>

- Sun, C., Qiu, X., Xu, Y., & Huang, X. (2020). *How to Fine-Tune BERT for Text Classification?* <https://doi.org/https://doi.org/10.48550/arXiv.1905.05583>
- Tammam, A., Sucipto, & Indriati, R. (2018). Hoax Detection At Social Media With Text Mining Clarification System-Based . *Jurnal Ilmiah Penelitian Dan Pembelajaran Informatika*, 3. <https://doi.org/https://doi.org/10.29100/jipi.v3i2.837>**
- Tandoc, E. C., Lim, Z. W., & Ling, R. (2018). Defining “Fake News”: A typology of scholarly definitions. *Digital Journalism*, 6(2), 137–153. <https://doi.org/https://doi.org/10.1080/21670811.2017.1360143>
- Tenney, I., Das, D., & Pavlick, E. (2019). *BERT Rediscovered the Classical NLP Pipeline*. 4593–4601. <https://doi.org/https://doi.org/10.18653/v1/P19-1452>
- Thölke, P., Mantilla-Ramos, Y. J., Abdelhedi, H., Maschke, C., Dehgan, A., Harel, Y., Kemtur, A., Mekki Berrada, L., Sahraoui, M., Young, T., Bellemare Pépin, A., El Khantour, C., Landry, M., Pascarella, A., Hadid, V., Combrisson, E., O’Byrne, J., & Jerbi, K. (2023). Class imbalance should not throw you off balance: Choosing the right classifiers and performance metrics for brain decoding with imbalanced data. *NeuroImage*, 277. <https://doi.org/https://doi.org/10.1016/j.neuroimage.2023.120253>
- Tobing, C. J. L., Wijayakusuma, I. L., & Harini, L. P. I. (2025a). Detection of Political Hoax News Using Fine-Tuning IndoBERT. *Journal of Applied Informatics and Computing*, 9(2), 354–360. <https://doi.org/https://doi.org/10.30871/jaic.v9i2.8989>
- Tobing, C. J. L., Wijayakusuma, I. L., & Harini, L. P. I. (2025b). Perbandingan Kinerja IndoBERT dan MBERT Untuk Deteksi Berita Hoaks Politik dalam Bahasa Indonesia. *JST (Jurnal Sains Dan Teknologi)*, 14(1), 114–123. <https://doi.org/https://doi.org/10.23887/jstundiksha.v14i1.92126>
- Tobing, E. (2020). *NormalisasiKata: Kamus Normalisasi Kata Tidak Baku Bahasa Indonesia*. Github. <https://github.com/evrintobing17/NormalisasiKata>
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2023). *Attention Is All You Need*. <https://doi.org/https://doi.org/10.48550/arXiv.1706.03762>