

DAFTAR PUSTAKA

- Akiba, T., Sano, S., Yanase, T., Ohta, T., & Koyama, M. (2019). *Optuna: A Next-generation Hyperparameter Optimization Framework* (Versi 1). arXiv. <https://doi.org/10.48550/ARXIV.1907.10902>
- Alammar, J., & Grootendorst, M. (2024). *Hands-on large language models: Language understanding and generation*. O'Reilly Media, Inc.
- Amalina, R. (2024). Optimization of clustering parameters in HDBSCAN for textual embeddings. *International Journal of Data Science and Analytics*, 18(3), 455–469. <https://doi.org/10.1007/s41060-024-00478-7>
- Amidei, J., Piwek, P., & Willis, A. (2018). Rethinking the Agreement in Human Evaluation Tasks. Dalam E. M. Bender, L. Derczynski, & P. Isabelle (Ed.), *Proceedings of the 27th International Conference on Computational Linguistics* (hlm. 3318–3329). Association for Computational Linguistics. <https://aclanthology.org/C18-1281/>
- Ananta, M., Utari, R., Akhyar, A., & Saptawati, G. A. P. (2024). Integrating bertopic and large language models for thematic identification of indonesian legal documents. *2024 11th International Conference on Advanced Informatics: Concept, Theory and Application (ICAICTA)*, 1–6.
- Artstein, R., & Poesio, M. (2008). Inter-Coder Agreement for Computational Linguistics. *Computational Linguistics*, 34(4), 555–596. <https://doi.org/10.1162/coli.07-034-R2>
- Azzahra, N. F. (2024). Analisis Sentimen Pada Postingan Media Sosial Twitter Dan Instagram Terkait Mangrove Surabaya. *JURNAL PILAR TEKNOLOGI Jurnal Ilmiah Ilmu Ilmu Teknik*, 9(1), 1–7. <https://doi.org/10.33319/piltek.v9i1.172>
- Bertsekas, D. P. (1997). Nonlinear Programming. *Journal of the Operational Research Society*, 48(3), 334–334. <https://doi.org/10.1057/palgrave.jors.2600425>
- Bishop, C. M., & Nasrabadi, N. M. (2006). *Pattern recognition and machine learning* (Vol. 4). Springer.
- Blei, D. M. (2012). Probabilistic topic models. *Commun. ACM*, 55(4), 77–84. <https://doi.org/10.1145/2133806.2133826>
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research*, 3, 993–1022.
- Boone, H., & Boone, D. (2012). Analyzing Likert Data. *Journal of Extension*, 50(2). <https://doi.org/10.34068/joe.50.02.48>
- Burger, C. (2023). *Adversarial Explanations for Text-Based Local Surrogate Models* [Master's Thesis]. The University of Mississippi.

- Campello, R. J. G. B., Moulavi, D., & Sander, J. (2013). Density-based clustering based on hierarchical density estimates. *Advances in Knowledge Discovery and Data Mining (PAKDD 2013)*, 160–172. https://doi.org/10.1007/978-3-642-37456-2_14
- Chang, J., Gerrish, S., Wang, C., Boyd-graber, J., & Blei, D. (2009). Reading Tea Leaves: How Humans Interpret Topic Models. Dalam Y. Bengio, D. Schuurmans, J. Lafferty, C. Williams, & A. Culotta (Ed.), *Advances in Neural Information Processing Systems* (Vol. 22). Curran Associates, Inc. https://proceedings.neurips.cc/paper_files/paper/2009/file/f92586a25bb3145facd64ab20fd554ff-Paper.pdf
- Coello Coello, C. A. (2000). Use of penalti functions for solving constrained optimization problems: A survey. *Computer Methods in Applied Mechanics and Engineering*, 191(11–12), 1245–1287. [https://doi.org/10.1016/S0045-7825\(01\)00323-1](https://doi.org/10.1016/S0045-7825(01)00323-1)
- Cohen, J. (1960). A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, 20(1), 37–46. <https://doi.org/10.1177/001316446002000104>
- Cohen, J. (1968). Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. *Psychological Bulletin*, 70(4), 213–220. <https://doi.org/10.1037/h0026256>
- 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*, 4171–4186. <https://doi.org/10.18653/v1/N19-1423>
- Dieng, A., Ruiz, F., & Blei, D. (2020). Topic Modeling in Embedding Spaces. *Transactions of the Association for Computational Linguistics*, 8, 439–453. https://doi.org/10.1162/tacl_a_00325
- Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. *arXiv preprint arXiv:1702.08608*. <https://arxiv.org/abs/1702.08608>
- Ehrgott, M. (2005). Multiobjective Versions of Polynomially Solvable Problems. Dalam *Multicriteria Optimization* (hlm. 221–270). Springer-Verlag. https://doi.org/10.1007/3-540-27659-9_9
- Gramegna, A., & Giudici, P. (2021). SHAP and LIME: An Evaluation of Discriminative Power in Credit Risk. *Frontiers in Artificial Intelligence*, 4, 752558. <https://doi.org/10.3389/frai.2021.752558>
- Grootendorst, M. (2022). BERTopic: Neural topic modeling with a class-based TF-IDF procedure. *arXiv preprint arXiv:2203.05794*.

- Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., & Pedreschi, D. (2018). A survey of methods for explaining black box models. *ACM Computing Surveys*, 51(5), 93:1-93:42. <https://doi.org/10.1145/3236009>
- Gunantara, N. (2018). A review of multi-objective optimization: Methods and its applications. *Cogent Engineering*, 5(1), 1502242. <https://doi.org/10.1080/23311916.2018.1502242>
- Jain, A. K., Murty, M. N., & Flynn, P. J. (1999). Data clustering: A review. *ACM Comput. Surv.*, 31(3), 264–323. <https://doi.org/10.1145/331499.331504>
- Jurafsky, D., & Martin, J. H. (2026). *Speech and Language Processing* (3rd Draft). Pearson.
- Keeney, R. L., & Raiffa, H. (1976). *Decisions with Multiple Objectives: Preferences and Value Trade-Offs*. John Wiley & Sons.
- Khandelwal, T. (2025). Using llm-based approaches to enhance and automate topic labeling. *arXiv preprint arXiv:2502.18469*.
- Koto, F., Lau, J. H., & Baldwin, T. (2021). Discourse Probing of Pretrained Language Models. Dalam K. Toutanova, A. Rumshisky, L. Zettlemoyer, D. Hakkani-Tur, I. Beltagy, S. Bethard, R. Cotterell, T. Chakraborty, & Y. Zhou (Ed.), *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (hlm. 3849–3864). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2021.naacl-main.301>
- Kozłowski, D., Pradier, C., & Benz, P. (2024). Generative AI for automatic topic labelling. *arXiv preprint arXiv:2408.07003*.
- Kusumaningrum, R., Khoerunnisa, S. F., Khadijah, K., & Syafrudin, M. (2024). Exploring Community Awareness of Mangrove Ecosystem Preservation through Sentence-BERT and K-Means Clustering. *Information*, 15(3), 165. <https://doi.org/10.3390/info15030165>
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *biometrics*, 159–174.
- Lau, J. H., Newman, D., & Baldwin, T. (2014). Machine Reading Tea Leaves: Automatically Evaluating Topic Coherence and Topic Model Quality. Dalam S. Wintner, S. Goldwater, & S. Riezler (Ed.), *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics* (hlm. 530–539). Association for Computational Linguistics. <https://doi.org/10.3115/v1/E14-1056>
- Laureate, C., Buntine, W., & Linger, H. (2023). A systematic review of the use of topic models for short text social media analysis. *Artificial Intelligence Review*, 56, 1–33. <https://doi.org/10.1007/s10462-023-10471-x>
- Liu, Y. (2024). Comparison of LDA and BERTopic in news topic modeling: A case study of The New York Times reports on China. *Pacific International Journal*, 7(2), 34–44.

- Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems (NeurIPS 2017)*, 4765–4774.
- M, N. R., & K.P, M. B. (2025). A hybrid explainability framework for recommender systems using SHAP with natural language interpretations. *International Journal of Innovative Research and Scientific Studies*, 8(6), 687–703. <https://doi.org/10.53894/ijirss.v8i6.9669>
- Marler, R., & Arora, J. (2004). Survey of Multi-Objective Optimization Methods for Engineering. *Structural and Multidisciplinary Optimization*, 26, 369–395. <https://doi.org/10.1007/s00158-003-0368-6>
- Mbake Woka, A. Y. P., Mahendra Dwifabri; Utama, Dody Qori. (2025). Sentiment Classification and Interpretation of Tokopedia Reviews: A Machine Learning, IndoBERT, and LIME Approach. *Building of Informatics, Technology and Science (BITS)*, (Vol 7 No 2 (2025): September 2025), 1164–1173.
- McInnes, L., Healy, J., & Melville, J. (2018). UMAP: Uniform Manifold Approximation and Projection for dimension reduction. *arXiv preprint arXiv:1802.03426*. <https://arxiv.org/abs/1802.03426>
- Medvecki, D., Bašaragin, B., Ljajić, A., & Milošević, N. (2024). Multilingual transformer and BERTopic for short text topic modeling: The case of Serbian. *arXiv preprint arXiv:2402.03067*. <https://arxiv.org/abs/2402.03067>
- Meng, Z. (2011). An Objective Penalty Functions Algorithm for Multiobjective Optimization Problem. *American Journal of Operations Research*, 01, 229–235. <https://doi.org/10.4236/ajor.2011.14026>
- Miettinen, K. (1999). *Nonlinear Multiobjective Optimization*. Kluwer Academic Publishers. <https://doi.org/10.1007/978-1-4615-5563-6>
- Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). *Efficient Estimation of Word Representations in Vector Space* (Versi 3). arXiv. <https://doi.org/10.48550/ARXIV.1301.3781>
- Molnar, C. (2022). *Interpretable Machine Learning*. Self-Published. <https://christophm.github.io/interpretable-ml-book/>
- Nguyen, D., Luu, A. T., & Le, T. (2023). Hashformers: Transformer-based contextual hashtag segmentation. *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (ACL)*.
- Nguyen, T., Wu, X., Dong, X., Nguyen, C.-D. T., Ng, S.-K., & Luu, A. T. (2024). *Topic Modeling as Multi-Objective Contrastive Optimization* (Versi 3). arXiv. <https://doi.org/10.48550/ARXIV.2402.07577>

- Nurwenda, I., & Hidayat, R. (2024). Analisis percakapan publik Indonesia tentang kebijakan lingkungan menggunakan BERTopic berbasis IndoBERTweet. *Jurnal Teknologi Informasi dan Sains Data Indonesia*, 9(1), 55–66.
- Rahman, M. A., & Purwarianti, A. (2021). IndoBERTweet: A pretrained language model for Indonesian Twitter text. *arXiv preprint arXiv:2109.03663*. <https://arxiv.org/abs/2109.03663>
- Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). “Why should I trust you?”: Explaining the predictions of any classifier. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1135–1144. <https://doi.org/10.1145/2939672.2939778>
- Roberts, C. V., Elahi, E., & Chandrashekar, A. (2022). *On the Bias-Variance Characteristics of LIME and SHAP in High Sparsity Movie Recommendation Explanation Tasks*. <https://arxiv.org/abs/2206.04784>
- Röder, M., Both, A., & Hinneburg, A. (2015). Exploring the space of topic coherence measures. *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining (WSDM '15)*, 399–408. <https://doi.org/10.1145/2684822.2685324>
- Samsir, M., Rahman, F., & Lestari, D. (2023). BERTopic modeling of natural language processing abstracts: Thematic structure and trajectory. *Machine Intelligence and Business Journal*, 12(4), 88–97. <https://doi.org/10.5281/zenodo.7775619>
- Saputra, K., Mahendra, R., & Purnama, D. (2022). Evaluating Indonesian transformer models for social media text classification. *Indonesian Journal of Computational Linguistics*, 7(2), 55–68.
- Saracevic, T. (2007). Relevance: A review of the literature and a framework for thinking on the notion in information science. Part II: nature and manifestations of relevance. *Journal of the American Society for Information Science and Technology*, 58(13), 1915–1933. <https://doi.org/10.1002/asi.20682>
- Sullivan, G. M., & Artino Jr, A. R. (2013). Analyzing and interpreting data from Likert-type scales. *Journal of graduate medical education*, 5(4), 541–542.
- Sunan Muhammad, D., Kusumaningrum, R., & Tantyoko, H. (2026). *Analisis Sentimen Unggahan di Platform X Terkait Mangrove Menggunakan IndoBERTweet-BIGRU dengan Interpretasi Lime dan Shap*. Universitas Diponegoro.
- Terragni, S., Harrando, I., Lisena, P., Troncy, R., & Fersini, E. (2022). *One Configuration to Rule Them All? Towards Hyperparameter Transfer in Topic Models using Multi-Objective Bayesian Optimization* (Versi 1). *arXiv*. <https://doi.org/10.48550/ARXIV.2202.07631>
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2023). *Attention Is All You Need*. <https://arxiv.org/abs/1706.03762>

- Wanyonyi, M., Morris, Z. N., Musyoka, F. M., Makaa, D., & Kitavi, D. M. (2025). Balancing interpretability and accuracy: A comparative evaluation of logistic regression and machine learning approaches for coronary heart disease risk prediction in low-resource settings. *Authorea Preprints*.
- Wibowo, A., Pratama, T., & Sari, D. (2025). Pemodelan topik media sosial tentang pariwisata hijau menggunakan IndoBERTweet dan BERTopic. *Jurnal Data Mining dan Kecerdasan Buatan Indonesia*, 7(1), 12–25.
- Yang, E., Lewis, D. D., & Frieder, O. (2019). A Regularization Approach to Combining Keywords and Training Data in Technology-Assisted Review. *Proceedings of the Seventeenth International Conference on Artificial Intelligence and Law, ICAIL '19*, 153–162. <https://doi.org/10.1145/3322640.3326713>
- Zhou, Z. (2021). *Statistical inference for machine learning: Feature importance, uncertainty quantification and interpretation stability*. Cornell University.