

REFERENCES

- Aghdam, M. H. (2023). Automatic extractive and generic document summarization based on nmf. *Journal of Artificial Intelligence and Soft Computing Research*, 13(1):37–49.
- Angelov, D. (2020). Top2vec: Distributed representations of topics. *arXiv preprint arXiv:2008.09470*.
- Babalola, O., Ojokoh, B., and Boyinbode, O. (2024). Comprehensive evaluation of lda, nmf, and bertopic's performance on news headline topic modeling. *Journal of Computing Theories and Applications*, 2(2):268–289.
- Blei, D. M., Ng, A. Y., and Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan):993–1022.
- Campello, R. J. G. B., Moulavi, D., and Sander, J. (2013). Density-based clustering based on hierarchical density estimates. In Pei, J., Tseng, V. S., Cao, L., Motoda, H., and Xu, G., editors, *Advances in Knowledge Discovery and Data Mining*, pages 160–172, Berlin, Heidelberg. Springer Berlin Heidelberg.
- Davies, D. L. and Bouldin, D. W. (2009). A cluster separation measure. *IEEE transactions on pattern analysis and machine intelligence*, (2):224–227.
- de Groot, M., Aliannejadi, M., and Haas, M. R. (2022). Experiments on generalizability of bertopic on multi-domain short text. *arXiv preprint arXiv:2212.08459*.
- Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2019). Bert: Pre-training of deep bidirectional transformers for language understanding. In *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), pages 4171–4186.

Diana, K. D. R. and Khodra, M. L. (2023). Indosbert: Enhancing indonesian sentence embeddings with siamese networks fine-tuning. In *2023 10th International Conference on Advanced Informatics: Concept, Theory and Application (ICAICTA)*, pages 1–6. IEEE.

Egger, R. and Yu, J. (2022). A topic modeling comparison between lda, nmf, top2vec, and bertopic to demystify twitter posts. *Frontiers in sociology*, 7:886498.

Ghasiya, P. and Okamura, K. (2021). Investigating covid-19 news across four nations: A topic modeling and sentiment analysis approach. *Ieee Access*, 9:36645–36656.

Grootendorst, M. (2022). Bertopic: Neural topic modeling with a class-based tf-idf procedure. *arXiv preprint arXiv:2203.05794*.

Hutama, L. B. and Suhartono, D. (2022). Indonesian hoax news classification with multilingual transformer model and bertopic. *Informatica*, 46(8).

Joachims, T. et al. (1997). A probabilistic analysis of the rocchio algorithm with tfidf for text categorization. In *ICML*, volume 97, pages 143–151. Citeseer.

Karami, A., Shah, V., Vaezi, R., and Bansal, A. (2020). Twitter speaks: A case of national disaster situational awareness. *Journal of Information Science*, 46(3):313–324.

Koto, F., Rahimi, A., Lau, J. H., and Baldwin, T. (2020). Indolem and indobert: A benchmark dataset and pre-trained language model for indonesian nlp. *arXiv preprint arXiv:2011.00677*.

- Kumar, A. and Sharma, D. (2021). Role and impact of digital media in spreading news and information during pandemic covid-19 in india. *Journal of Contemporary Issues in Business and Government*, 27(1):2538–2545.
- Kurniat, P. Y., Ayu Indrayathi, P., Erma Pradnyan, P., Ulandar, P. S., Putu Januraga, P., Sri Yuliarti, M., Samaria Santosa, K., et al. (2023). Infodemic, health promotion efforts, and preventive behavior during the covid-19 pandemic in indonesia: A quantitative analysis study.
- Lukman, E. (2013). Indonesia’s kapanlagi wants to surpass Kompas and detik in 2 years. Accessed: July 2025.
- McInnes, L., Healy, J., and Melville, J. (2018). Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426*.
- Reimers, N. and Gurevych, I. (2019). Sentence-bert: Sentence embeddings using siamese bert-networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics.
- Reporters Without Borders (2025). Indonesia. Accessed: July 2025.
- Röder, M., Both, A., and Hinneburg, A. (2015). Exploring the space of topic coherence measures. In *Proceedings of the eighth ACM international conference on Web search and data mining*, pages 399–408.
- Schäfer, K., Choi, J.-E., Vogel, I., and Steinebach, M. (2024). Unveiling the potential of bertopic for multilingual fake news analysis—use case: Covid-19. *arXiv preprint arXiv:2407.08417*.
- Taghandiki, K. and Mohammadi, M. (2023). Topic modeling: Exploring the processes, tools, challenges and applications. *Authorea Preprints*.

- Wahyudi, M. D. R., Fatwanto, A., Kiftiyani, U., and Wonoseto, M. G. (2021). Topic modeling of online media news titles during covid-19 emergency response in indonesia using the latent dirichlet allocation (lda) algorithm. *Telematika*, 14(2):101–111.
- Wang, Q. and Ma, B. (2024). Enhancing bertopic with pre-clustered knowledge: Reducing feature sparsity in short text topic modeling. *Journal of Data Analysis and Information Processing*, 12(4):597–611.
- Yadav, V. and Shakya, S. (2022). Sentiment analysis and topic modeling on news headlines. *Journal of Ubiquitous Computing and Communication Technologies*, 4(3):204–218.
- Yin, J. and Wang, J. (2014). A dirichlet multinomial mixture model-based approach for short text clustering. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 233–242.
- Zhang, X. et al. (2023). Mptopic: Improving topic modeling via masked permuted pre-training. *arXiv preprint arXiv:2309.01015*.