

## DAFTAR PUSTAKA

- Bhattacharya, M., Jurkowitz, C., & Shatkay, H. (2017). Identifying patterns of associated-conditions through topic models of Electronic Medical Records. *Proceedings - 2016 IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2016*, 466–469. <https://doi.org/10.1109/BIBM.2016.7822561>
- Bíró, I., & Szabó, J. (2010). *Large scale link based latent Dirichlet allocation for web document classification*. <http://arxiv.org/abs/1006.4953>
- Blei, D. M. (2012a). Probabilistic topic models. *Communications of the ACM*, 55(4), 77–84. <https://doi.org/10.1145/2133806.2133826>
- Blei, D. M. (2012b). *Probabilistic Topic Models*. 77–84.
- Blei, D. M., & Lafferty, J. D. (2006). *Dynamic Topic Models*.
- Blei, D. M., Ng, A. Y., & Edu, J. B. (2003). Latent Dirichlet Allocation Michael I. Jordan. In *Journal of Machine Learning Research* (Vol. 3).
- Bojanowski, P., Grave, E., Joulin, A., & Mikolov, T. (2017). Enriching Word Vectors with Subword Information. *Transactions of the Association for Computational Linguistics*, 5, 135–146. [https://doi.org/10.1162/tacl\\_a\\_00051](https://doi.org/10.1162/tacl_a_00051)
- Brill, E. (1992). A Simple Rule-Based Part of Speech Tagger. *HLT*.
- Carnot, M. L., Bernardino, J., Laranjeiro, N., & Oliveira, H. G. (2020). Applying text analytics for studying research trends in dependability. *Entropy*, 22(11), 1–20. <https://doi.org/10.3390/e22111303>
- Chen, L.-C. (2017). An effective LDA-based time topic model to improve blog search performance. *Information Processing & Management*, 53(6), 1299–1319. <https://doi.org/https://doi.org/10.1016/j.ipm.2017.08.001>
- Corpling@GU. (n.d.). *Extended Penn Treebank Tag Set*. Computational Corpus Linguistics Lab, Department of Linguistics, Georgetown University.

- Devlin, J., Chang, M.-W., Lee, K., Google, K. T., & Language, A. I. (2018). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *ArXiv Preprint ArXiv:1810.04805*. <https://github.com/tensorflow/tensor2tensor>
- Fahlevvi, M. R. (2019). PEMODELAN TOPIK PADA PORTAL BERITA ONLINE MENGGUNAKAN LATENT DIRICHLET ALLOCATION (LDA). *Universitas Gadjah Mada*.
- Farizki Wicaksono, A., & Purwarianti, A. (2010, June). *HMM Based Part-of-Speech Tagger for Bahasa Indonesia*.
- Garcia, S. (2007). *Search Engine Optimisation Using Past Queries*. RMIT University.
- Greene, B. B., & Rubin, G. M. (1971). *Automatic Grammatical Tagging of English*. Department of Linguistics, Brown University.
- Guo, C., Lu, M., & Wei, W. (2021). An Improved LDA Topic Modeling Method Based on Partition for Medium and Long Texts. *Annals of Data Science*, 8(2), 331–344. <https://doi.org/10.1007/s40745-019-00218-3>
- Harzing, A.-W. (2007). *Publish or Perish*. <https://harzing.com/resources/publish-or-perish>
- Jurafsky, D. (2000). *Speech and Language Processing “An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition.”* Prentice-Hall, Inc.
- Kavvadias, S., Drosatos, G., & Kaldoudi, E. (2020). Supporting topic modeling and trends analysis in biomedical literature. *Journal of Biomedical Informatics*, 110, 103574. <https://doi.org/https://doi.org/10.1016/j.jbi.2020.103574>
- Kengken, R. I. (2014). *Pemodelan Topik untuk Media Sosial Menggunakan Latent Dirichlet Allocation*. Universitas Gadjah Mada.
- Krasnashchok, K., & Jouili, S. (2018). *Improving Topic Quality by Promoting Named Entities in Topic Modeling*. 247–253. <https://doi.org/10.18653/v1/P18-2040>

- Martin, F., & Johnson, M. (2015). *More Efficient Topic Modelling Through a Noun Only Approach*. <http://www.nltk.org/howto/wordnet.html>
- Mehrotra, R., Sanner, S., Buntine, W., & Xie, L. (2013). *Improving LDA Topic Models for Microblogs via Tweet Pooling and Automatic Labeling*. 889–892. <https://doi.org/10.1145/2484028.2484166>
- Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). *Efficient Estimation of Word Representations in Vector Space*. <http://arxiv.org/abs/1301.3781>
- Mitchell, M. P., Santorini, B., & Marcinkiewicz, M. A. (2003). *Alphabetical list of part-of-speech tags used in the Penn Treebank Project*. University of Pennsylvania.
- Newman, D., Bonilla, E. v, & Buntine, W. (2011). Improving Topic Coherence with Regularized Topic Models. In J. Shawe-Taylor, R. Zemel, P. Bartlett, F. Pereira, & K. Q. Weinberger (Eds.), *Advances in Neural Information Processing Systems* (Vol. 24). Curran Associates, Inc.  
<https://proceedings.neurips.cc/paper/2011/file/5ef698cd9fe650923ea331c15af3b160-Paper.pdf>
- Newman, D., Lau, J. H., Grieser, K., & Baldwin, T. (2010). *Automatic Evaluation of Topic Coherence*. June, 100–108.
- Nuridin, A., Anggo, B., Aji, S., Bustamin, A., & Abidin, Z. (2020). Perbandingan Kinerja Word Embedding Word2Vec, GloVe, dan fastText pada Klasifikasi Teks. *Jurnal TEKNOKOMPAK*, 14(2), 74.
- Paul, M., & Girju, R. (2009). *Topic Modeling of Research Fields: An Interdisciplinary Perspective*. [www.scopus.com](http://www.scopus.com)
- Pisceldo, F. (2009). *Probabilistic Part Of Speech Tagging for Bahasa Indonesia*.
- Robertson, S. E. (2004). Understanding inverse document frequency: on theoretical arguments for IDF. *J. Documentation*, 60, 503–520.
- Robinson, L. (2016). *Implementasi Metode Generalized Vector Space Model Pada Aplikasi Information Retrieval Untuk Pencarian Informasi Pada*

- Kumpulan Dokumen Teknik Elektro di UPT BPI LIPI*. Universitas Komputer Indonesia.
- Septian, J. A., Fachrudin, T. M., & Nugroho, A. (2019). Analisis Sentimen Pengguna Twitter Terhadap Polemik Persepakbolaan Indonesia Menggunakan Pembobotan TF-IDF dan K-Nearest Neighbor. *INSYST: Journal of Intelligent System and Computation*, 1.
- Socrates, I. G. A., Akbar, A. L., Akbar, M. S., Arifin, A. Z., & Herumurti, D. (2017). Optimasi Naive Bayes Dengan Pemilihan Fitur Dan Pembobotan Gain Ratio. *Lontar Komputer : Jurnal Ilmiah Teknologi Informasi*, 7(1), 22. <https://doi.org/10.24843/lkjiti.2016.v07.i01.p03>
- Stevens, K., Kegelmeyer, W. P., Andrzejewski, D., & Buttlar, D. J. (2012). Exploring Topic Coherence over Many Models and Many Topics. *EMNLP*.
- Tran Bach Xuan and Nghiem, S. and S. O. and V. T. M. and H. G. H. and V. G. T. and P. H. Q. and D. H. T. and L. C. A. and T. W. and H. C. S. H. and H. R. C. M. (2019). Modeling Research Topics for Artificial Intelligence Applications in Medicine: Latent Dirichlet Allocation Application Study. *J Med Internet Res*, 21(11), e15511. <https://doi.org/10.2196/15511>
- Utami, K. P. (2017). *ANALISIS TOPIK DATA MEDIA SOSIAL TWITTER MENGGUNAKAN MODEL TOPIK LATENT DIRICHLET ALLOCATION*. Institut Pertanian Bogor.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł. ukasz, & Polosukhin, I. (2017). Attention is All you Need. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, & R. Garnett (Eds.), *Advances in Neural Information Processing Systems* (Vol. 30). Curran Associates, Inc. [https://proceedings.neurips.cc/paper\\_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf](https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf)
- Wahyuni, R. T., Prastiyanto, D., & Suprpto, D. E. (2017). *Penerapan Algoritma Cosine Similarity dan Pembobotan TF-IDF pada Sistem*

*Klasifikasi*

*Dokumen*

*Skripsi.*

<https://doi.org/https://doi.org/10.15294/jte.v9i1.10955>

Williams, E. M., Levin, D., & McCulloh, I. (2020). Improving LDA Topic Modeling with Gamma and Simmelian Filtration. *2020 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, 692–696.

<https://doi.org/10.1109/ASONAM49781.2020.9381330>

Yang, H.-L., Chang, T.-W., & Choi, Y. (2018). Exploring the Research Trend of Smart Factory with Topic Modeling. *Sustainability*, 10(8).  
<https://doi.org/10.3390/su10082779>

Yang, T.-I., Torget, A., & Mihalcea, R. (2011). *Topic modeling on historical newspapers*. 96–104.