

## DAFTAR PUSTAKA

- Abdelmotaleb, H., Wojtyś, M., & McNeile, C., (2023), GSDMM Model Evaluation Techniques with Application to British Telecom Data, *Proceedings of the 5th International Conference on Statistics: Theory and Applications*, 116, 1–8. <https://doi.org/10.11159/icsta23.116>
- Abdelrazek, A., Eid, Y., Gawish, E., Medhat, W., & Hassan, A., (2023), Topic modeling algorithms and applications: A survey, *Information Systems*, 112, 102131. <https://doi.org/10.1016/j.is.2022.102131>
- Agarwal, N., Sikka, G., & Awasthi, L. K., (2020), Evaluation of web service clustering using Dirichlet Multinomial Mixture model based approach for Dimensionality Reduction in service representation, *Information Processing and Management*, 57(4), 102238. <https://doi.org/10.1016/j.ipm.2020.102238>
- Aksoy, M., Yanık, S., & Amasyali, M. F., (2023), A comparative analysis of text representation, classification and clustering methods over real project proposals, In *International Journal of Intelligent Computing and Cybernetics* (Vol. 16, Issue 3). <https://doi.org/10.1108/IJICC-11-2022-0289>
- Alash, H. M., & Al-Sultany, G. A., (2020), Improve topic modeling algorithms based on Twitter hashtags, *Journal of Physics: Conference Series*, 1660(1). <https://doi.org/10.1088/1742-6596/1660/1/012100>
- Alipour, S., Galeazzi, A., Sangiorgio, E., Avalle, M., Bojic, L., Cinelli, M., & Quattrociochi, W., (2023), *Cross-Platform Social Dynamics: An Analysis of ChatGPT and COVID-19 Vaccine Conversations*, 1–25.
- Alsemaree, O., Alam, A. S., Gill, S., & Uhlig, S., (2024), Sentiment analysis of Arabic social media texts: A machine learning approach to deciphering customer perceptions, *Heliyon*, e27863. <https://doi.org/10.1016/j.heliyon.2024.e27863>
- Alsmadi, M. K., Alzaqebah, M., Jawarneh, S., Almarashdeh, I., Al-Betar, M. A., Alwohaibi, M., Al-Mulla, N. A., Ahmed, E. A. E., & Al Smadi, A., (2024), Hybrid topic modeling method based on dirichlet multinomial mixture and fuzzy match algorithm for short text clustering, *Journal of Big Data*, 11(1).

<https://doi.org/10.1186/s40537-024-00930-9>

Amaro, A., & Bacao, F., (2024), Topic Modeling: A Consistent Framework for Comparative Studies, *Emerging Science Journal*, 8(1), 125–139.

<https://doi.org/10.28991/ESJ-2024-08-01-09>

Aznag, M., Quafafou, M., & Jarir, Z., (2013), Correlated Topic Model for Web Services Ranking, *International Journal of Advanced Computer Science and Applications*, 4(6). <https://doi.org/10.14569/IJACSA.2013.040637>

Bastani, K., Namavari, H., & Shaffer, J., (2019), Latent Dirichlet allocation (LDA) for topic modeling of the CFPB consumer complaints, *Expert Systems with Applications*, 127, 256–271. <https://doi.org/10.1016/j.eswa.2019.03.001>

Blei, D. M., & Lafferty, J. D., (2006), Correlated topic models, *Advances in Neural Information Processing Systems*, 147–154.

Blei, D. M., Ng, A. Y., & Jordan, M. T., (2002), Latent dirichlet allocation, *Advances in Neural Information Processing Systems*, 3, 993–1022.

Chang, J., Boyd-Graber, J., Gerrish, S., Wang, C., & Blei, D. M., (2009), Reading tea leaves: How humans interpret topic models, *Advances in Neural Information Processing Systems 22 - Proceedings of the 2009 Conference*, 288–296.

De Santis, E., Martino, A., Ronci, F., & Rizzi, A., (2024), 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), 1–15.

<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, *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.

Egger, R., (2022), A Topic Modeling Comparison Between LDA, NMF, Top2Vec, and BERTopic to Demystify Twitter Posts, *Frontiers in Sociology*, 7.

<https://doi.org/10.3389/fsoc.2022.886498>

- Egger, R., & Yu, J., (2022a), A Topic Modeling Comparison Between LDA, NMF, Top2Vec, and BERTopic to Demystify Twitter Posts, *Frontiers in Sociology*, 7(May), 1–16. <https://doi.org/10.3389/fsoc.2022.886498>
- Egger, R., & Yu, J., (2022b), A Topic Modeling Comparison Between LDA, NMF, Top2Vec, and BERTopic to Demystify Twitter Posts, *Frontiers in Sociology*, 7(May), 1–16. <https://doi.org/10.3389/fsoc.2022.886498>
- Hajek, P., Barushka, A., & Munk, M., (2021), Neural Networks with Emotion Associations, Topic Modeling and Supervised Term Weighting for Sentiment Analysis, *International Journal of Neural Systems*, 31(10), 2150013. <https://doi.org/10.1142/S0129065721500131>
- Hajjem, M., & Latiri, C., (2017), Combining IR and LDA Topic Modeling for Filtering Microblogs, *Procedia Computer Science*, 112, 761–770. <https://doi.org/10.1016/j.procs.2017.08.166>
- Hofmann, T., (1999), Probabilistic latent semantic indexing, *Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 50–57. <https://doi.org/10.1145/312624.312649>
- Jang, H., Rempel, E., Roth, D., Carenini, G., & Janjua, N. Z., (2021), Tracking COVID-19 discourse on twitter in north america: Infodemiology study using topic modeling and aspect-based sentiment analysis, *Journal of Medical Internet Research*, 23(2). <https://doi.org/10.2196/25431>
- Jelodar, H., Wang, Y., Yuan, C., Feng, X., Jiang, X., Li, Y., & Zhao, L., (2019), Latent Dirichlet allocation (LDA) and topic modeling: models, applications, a survey, *Multimedia Tools and Applications*, 78(11), 15169–15211. <https://doi.org/10.1007/s11042-018-6894-4>
- Jha, B., (2019), The Role of Social Media Communication: Empirical Study of Online Purchase Intention of Financial Products, *Global Business Review*, 20(6), 1445–1461. <https://doi.org/10.1177/0972150919848912>
- 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/>

- Kalepalli, Y., Tasneem, S., Teja, P. D. P., & Manne, S., (2020), Effective Comparison of LDA with LSA for Topic Modelling, *Proceedings of the International Conference on Intelligent Computing and Control Systems, ICICCS 2020, Iciccs*, 1245–1250. <https://doi.org/10.1109/ICICCS48265.2020.9120888>
- Khadija, M. A., & Nurharjadmo, W., (2024), Enhancing Indonesian customer complaint analysis: LDA topic modelling with BERT embeddings, *Sinergi (Indonesia)*, 28(1), 153–162. <https://doi.org/10.22441/sinergi.2024.1.015>
- Kherwa, P., & Bansal, P., (2020), Topic Modeling: A Comprehensive Review, *EAI Endorsed Transactions on Scalable Information Systems*, 7(24), 1–16. <https://doi.org/10.4108/eai.13-7-2018.159623>
- Koto, F., Lau, J. H., & Baldwin, T., (2021a), IndoBERTweet: A Pretrained Language Model for Indonesian Twitter with Effective Domain-Specific Vocabulary Initialization, *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, 10660–10668. <https://doi.org/10.18653/v1/2021.emnlp-main.833>
- Koto, F., Lau, J. H., & Baldwin, T., (2021b), IndoBERTweet: A Pretrained Language Model for Indonesian Twitter with Effective Domain-Specific Vocabulary Initialization, *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, 10660–10668. <https://doi.org/10.18653/v1/2021.emnlp-main.833>
- Lafferty, J., & Blei, D., (2005), Correlated Topic Models, In Y. Weiss, B. Schölkopf, & J. Platt (Eds.), *Advances in Neural Information Processing Systems* (Vol. 18), MIT Press. [https://proceedings.neurips.cc/paper\\_files/paper/2005/file/9e82757e9a1c12cb710ad680db11f6f1-Paper.pdf](https://proceedings.neurips.cc/paper_files/paper/2005/file/9e82757e9a1c12cb710ad680db11f6f1-Paper.pdf)
- Landauer Thomas, K., & Foltz Peter, W., (1998), Laham Darrell. An Introduction to Latent Semantic Analysis, *Discourse Processes*. Springer-Verlag, June 2012, 259–284.
- Lossio-Ventura, J. A., Gonzales, S., Morzan, J., Alatrística-Salas, H., Hernandez-

- Boussard, T., & Bian, J., (2021), Evaluation of clustering and topic modeling methods over health-related tweets and emails, *Artificial Intelligence in Medicine*, 117(May), 102096. <https://doi.org/10.1016/j.artmed.2021.102096>
- Manning, C. D., (2009), *An introduction to information retrieval*.
- Mikolov, T., Chen, K., Corrado, G., & Dean, J., (2013), Efficient estimation of word representations in vector space, *1st International Conference on Learning Representations, ICLR 2013 - Workshop Track Proceedings*, 1–12.
- Mimno, D., Wallach, H., Talley, E., Leenders, M., & McCallum, A., (2011), Optimizing semantic coherence in topic models, *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, 262–272.
- Mohammed, S. H., & Al-Augby, S., (2020), LSA & LDA topic modeling classification: Comparison study on E-books, *Indonesian Journal of Electrical Engineering and Computer Science*, 19(1), 353–362. <https://doi.org/10.11591/ijeecs.v19.i1.pp353-362>
- Odden, T. O. B., Tyseng, H., Mjaaland, J. T., Kreutzer, M. F., & Malthe-Sørensen, A., (2024), Using text embeddings for deductive qualitative research at scale in physics education, *Physical Review Physics Education Research*, 20(2), 20151. <https://doi.org/10.1103/PhysRevPhysEducRes.20.020151>
- Pennington, J., Socher, R., & Manning, C. D., (2014), Glove: Global vectors for word representation, *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 1532–1543.
- Qiang, J., Qian, Z., Li, Y., Yuan, Y., & Wu, X., (2022a), Short Text Topic Modeling Techniques, Applications, and Performance: A Survey, *IEEE Transactions on Knowledge and Data Engineering*, 34(3), 1427–1445. <https://doi.org/10.1109/TKDE.2020.2992485>
- Qiang, J., Qian, Z., Li, Y., Yuan, Y., & Wu, X., (2022b), Short Text Topic Modeling Techniques, Applications, and Performance: A Survey, *IEEE Transactions on Knowledge and Data Engineering*, 34(3), 1427–1445. <https://doi.org/10.1109/TKDE.2020.2992485>
- Reimers, N., & Gurevych, I., (2019), *Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks*. <https://doi.org/10.48550/arXiv.1908.10084>

- 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*, 399–408.
- Romero, J. D., Feijoo-Garcia, M. A., Nanda, G., Newell, B., & Magana, A. J., (2024), Evaluating the Performance of Topic Modeling Techniques with Human Validation to Support Qualitative Analysis, *Big Data and Cognitive Computing*, 8(10). <https://doi.org/10.3390/bdcc8100132>
- Schivinski, B., & Dabrowski, D., (2016), The effect of social media communication on consumer perceptions of brands, *Journal of Marketing Communications*, 22(2), 189–214. <https://doi.org/10.1080/13527266.2013.871323>
- Si, Y., Wang, J., Xu, H., & Roberts, K., (2019), Enhancing clinical concept extraction with contextual embeddings, *Journal of the American Medical Association Informatics*, 26(11), 1297–1304. <https://doi.org/10.1093/jamia/ocz096>
- Tannia, L. N., & Monika, M., (2022), Intensitas Penggunaan Media Sosial dan Prokrastinasi Akademik pada Mahasiswa Saat sistem Pembelajaran Jarak Jauh, *Edukatif: Jurnal Ilmu Pendidikan*, 4(4), 5203–5212. <https://doi.org/10.31004/edukatif.v4i4.3135>
- Teh, Y. W., Jordan, M. I., Beal, M. J., & Blei, D. M., (2005), Sharing clusters among related groups: Hierarchical dirichlet processes, *Advances in Neural Information Processing Systems*, 1.
- Thakral, P., Sharma, D., & Ghosh, K., (2024), Evidence-based knowledge management: a topic modeling analysis of research on knowledge management and analytics, *VINE Journal of Information and Knowledge Management Systems*. <https://doi.org/10.1108/VJKMS-03-2023-0079>
- Walsh, J., Cave, J., & Griffiths, F., (2024), Combining Topic Modeling, Sentiment Analysis, and Corpus Linguistics to Analyze Unstructured Web-Based Patient Experience Data: Case Study of Modafinil Experiences, *Journal of Medical Internet Research*, 26, e54321. <https://doi.org/10.2196/54321>
- Wang, Y., Bashar, M. A., Chandramohan, M., & Nayak, R., (2023), Exploring topic models to discern cyber threats on Twitter: A case study on Log4Shell,

*Intelligent Systems with Applications*, 20(March), 200280.  
<https://doi.org/10.1016/j.iswa.2023.200280>

Weisser, C., Gerloff, C., Thielmann, A., Python, A., Reuter, A., Kneib, T., & Säfken, B., (2023), Pseudo-document simulation for comparing LDA, GSDMM and GPM topic models on short and sparse text using Twitter data, *Computational Statistics*, 38(2), 647–674. <https://doi.org/10.1007/s00180-022-01246-z>

Widiantoro, A. D., Mustafid, M., & Sanjaya, R., (2023), Analyzing User Comments on Fintech E-wallet Applications in Indonesia Using the LDA Algorithm, *2023 7th International Conference on Information Technology (InCIT)*, 169–174. <https://doi.org/10.1109/InCIT60207.2023.10412996>

Yadav, P., Kumar, A., Shivani, S., Hooda, R., Sudhir, S., & Pooja, P., (2023), Enhanced Spam Detection System for Twitter Social Networking Platform, *International Journal on Recent and Innovation Trends in Computing and Communication*, 11(11s), 195–201.  
<https://doi.org/10.17762/ijritcc.v11i11s.8086>

Yin, J., & Wang, J., (2014), A Dirichlet multinomial mixture model-based approach for short text clustering, *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 233–242.  
<https://doi.org/10.1145/2623330.2623715>

SEKOLAH PASCASARJANA