

## DAFTAR PUSTAKA

- Aldisa, R. T., Maulana, P., & Abdullah, M. A. (2022). Penerapan Big Data Analytic Terhadap Strategi Pemasaran Job Portal di Indonesia dengan Karakteristik Big Data 5V. *Jurnal Sistem Komputer dan Informatika (JSON)*, 3(3), 267. <https://doi.org/10.30865/json.v3i3.3905>
- Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM*, 55(4), 77–84. <https://doi.org/10.1145/2133806.2133826>
- Blei, D. M., Ng, A. Y., & Edu, J. B. (2003). *Latent Dirichlet Allocation* Michael I. Jordan. Dalam *Journal of Machine Learning Research* (Vol. 3).
- Bouabdallaoui, I., Guerouate, F., & Sbihi, M. (2023). Assessing *Topic Modeling* in Online Forums: A Comparative Study of Hierarchical and -Based *Clustering* Algorithms. *Proceedings - 10th International Conference on Wireless Networks and Mobile Communications, WINCOM 2023*. <https://doi.org/10.1109/WINCOM59760.2023.10322986>
- Castillo, C. (2015). PPT K-Means Algorithm. Course Algorithmic Methods of Data Science. Sapienza University of Rome
- Danayanti, P., Purwitasari, D., & Suciatri, N. (2021). Eliminasi Data Non-Topic Menggunakan Pemodelan Topik Untuk Peringkasan Otomatis Data Tweet Dengan Konteks Covid-19. *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIK)*. Vol. 8, No. 1. <https://doi.org/10.25126/jtiik.202184324>
- Dharma, E. M., Lumban Gaol, F., Leslie, H., Warnars, H. S., & Soewito, B. (2022). The Accuracy Comparison Among Word2Vec, GloVe, And FastText Towards Convolution Neural Network (CNN) Text Classification. *Journal of Theoretical and Applied Information Technology*, 31(2). [www.jatit.org](http://www.jatit.org)
- Dimitriadis, N. (2020). ‘*Applying Topic Modeling Algorithms on Twitter messages in Greek language*’, Ikee.Lib.Auth.Gr.
- Elgendy, N., & Elragal, A. (2014). Big data analytics: A literature review paper. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 8557 LNAI, 214–227. [https://doi.org/10.1007/978-3-319-08976-8\\_16](https://doi.org/10.1007/978-3-319-08976-8_16)
- Eligüzel, N., Çetinkaya, C., & Dereli, T. (2023). Comparative analysis with *topic modeling* and *word embedding methods* after the Aegean Sea earthquake on Twitter. *Evolving Systems*, 14(2), 245–261. <https://doi.org/10.1007/s12530-022-09450-4>
- Farzadnia, S., Vanani, I. R., & Hanafizadeh, P. (2024). An experimental study for identifying customer prominent viewpoints on different flight classes by *topic modeling* methods. *International Journal of Information Management Data Insights*, 4(1). <https://doi.org/10.1016/j.jjime.2024.100223>

- Fauzi, M. A. (2019). Word2Vec model for sentiment analysis of product reviews in Indonesian language. *International Journal of Electrical and Computer Engineering (IJECE)*, 9(1), 525. <https://doi.org/10.11591/ijece.v9i1.pp525-530>
- Fernandes, N., Gkolia, A., Pizzo, N., Davenport, J., & Nair, A. (2020). Unification of HDP and LDA Models for Optimal Topic *Clustering* of Subject Specific Question Banks. <http://arxiv.org/abs/2011.01035>
- Greau-Hamard, P.-S., Djoko-Kouam, M., Louët, Y. (2022). Performance Analysis and Comparison of Sequence Identification Algorithms in IoT Context. <https://hal.science/hal-03537524>
- Habibi, M. (2021). *Topic Modeling of Gernas Related Content on Instagram Using Latent Dirichlet Allocation (LDA)*. 34(Ahms 2020), pp. 260–264. doi: 10.2991/ahsr.k.210127.060.
- Hasan, M., Hossain, M. M., Ahmed, A., & Rahman, M. S. (2019). *Topic Modeling: A Comparison of the Performance of Latent Dirichlet Allocation and LDA2vec Model on Bangla Newspaper*. 2019 *International Conference on Bangla Speech and Language Processing, ICBSLP 2019*. <https://doi.org/10.1109/ICBSLP47725.2019.202047>
- Hayat, C. (2021). Enhanced K-Means *Clustering* Approach for Diagnosis Types of Acne. 2021 2nd International Conference on Innovative and Creative Information Technology, ICITech 2021, 39–43. <https://doi.org/10.1109/ICITech 50181. 2021. 9590 124>
- Hermanto, T. H., Setyanto, A., Luthfi, E. T. (2021). Algoritma LSTM-CNN untuk Sentimen Klasifikasi dengan Word2Vec pada Media Online. *Citec Journal*. 8(1): 64-77.
- Jelodar, H., & Wang, Y. (2017). *Latent Dirichlet Allocation (LDA) and Topic modeling: models, applications, a survey*
- Khairunnisa, S. O. (2018). Analisis Topik dan Author Publikasi pada Repositori Open Access Journal of Information Systems dengan Menggunakan Metode Author-Topic Models. Undergraduate Thesis, Institut Teknologi Sepuluh Nopember.
- Lezama-Sánchez, A. L., Vidal, M.T., & Reyes-Ortiz, J. A. (2023). Integrating Text Classification into Topic Discovery Using Semantic Embedding Models. *Applied Sciences (Switzerland)*, 13(17). <https://doi.org/10.3390/app13179857>
- Li, J., Huang, G., Fan, C., Sun, Z., & Zhu, H. (2019). Key word extraction for short text via Word2Vec, doc2vec, and textrank. *Turkish Journal of Electrical Engineering and Computer Sciences*, 27(3), 1794–1805. <https://doi.org/10.3906/elk-1806-38>
- Liddy, E. D. (2001). *Natural Language Processing*. <https://surface.syr.edu/istpub>
- Meddeb, A., & Romdhane, L. Ben. (2022). Using *Topic Modeling* and *Word Embedding* for Topic Extraction in Twitter. *Procedia Computer Science*, 207, 790–799. <https://doi.org/10.1016/j.procs.2022.09.134>
- Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient Estimation of Word Representations in Vector Space. <http://arxiv.org/abs/1301.3781>

- Mohd Amram, N. A. L., Keikhosrokiani, P., & Asl, M. P. (2023). Artificial intelligence approach for detection and classification of depression among refugees in selected diasporic novels. *Social Sciences and Humanities Open*, 8(1). <https://doi.org/10.1016/j.ssaho.2023.100558>
- Nisha & Kaur, P., J. (2015). A survey of *clustering* techniques and algorithms. 2nd International Conference on Computing for Sustainable Global Development (INDIACom), New Delhi, India, 2015, pp. 304-307.
- 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.
- Nurlayli, A., & Nasichuddin, M. (2019). Topik Modeling Penelitian Dosen JPTEI UNY Pada Google Scholar Menggunakan *Latent Dirichlet Allocation*. *Elinvo* (Electronics, Informatics, and Vocational Education), 4(2), 154-161. [doi:https://doi.org/10.21831/elinvo.v4i2.28254](https://doi.org/10.21831/elinvo.v4i2.28254)
- Paisley, J., Wang, C., Blei, D. M., & Jordan, M. I. (2015). Nested *hierarchical dirichlet processes*. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(2), 256–270. <https://doi.org/10.1109/TPAMI.2014.2318728>
- Pavithra, C. B., & Savitha, D. J. (2024a). Advancements In Dynamic *Topic Modeling*: A Comparative Analysis of LDA, DTM, GibbsLDA++, HDP and Proposed Hybrid Model HDP With Ct-DTM for Real-Time And Evolving Textual Data. *Journal of Theoretical and Applied Information Technology*, 31(10). [www.jatit.org](http://www.jatit.org)
- Pavithra, & Savitha. (2024b). *Topic Modeling* for Evolving Textual Data Using LDA, HDP, NMF, BERTOPIC, and DTM With a Focus on Research Papers. *Journal of Technology and Informatics (JoTI)*, 5(2), 53–63. <https://doi.org/10.37802/joti.v5i2.618>
- Prabowo, Y. D., Marselino, D.T., & Suryawiguna, M. (2019). Pembentukan Vector Space Model Bahasa Indonesia Menggunakan Metode Word to Vector.
- Pratama, E. D. (2022). Implementasi Model Long-Short Term Memory (LSTM) pada Klasifikasi Teks Data SMS Spam Berbahasa Indonesia. *The Journal on Machine Learning and Computing Intelligence (JMLCI)* 1(2): 38-42
- Putri, W., & Hendrowati, R. (2018). Penggalan Teks Dengan Model *Bag Of Words* Terhadap Data Twitter. *Jurnal Muara Sains, Ilmu Kesehatan*, 2(1), 129–138.
- Rahman, M.D., Djunaidy, A., & Mahananto, F. (2021). Penerapan Weighted *Word Embedding* pada Pengklasifikasian Teks Berbasis Recurrent Neural Network untuk Layanan Pengaduan Perusahaan Transportasi. *Jurnal Sains dan Seni ITS*.
- Röder, M., Both, A., & Hinneburg, A. (2015). Exploring the space of *topic coherence measures*. *WSDM 2015 - Proceedings of the 8th ACM International Conference on Web Search and Data Mining*, 399–408. <https://doi.org/10.1145/2684822.2685324>
- Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of *cluster* analysis. *Journal of Computational and Applied Mathematics*, 20, 53-65.

- Sinaga, K. P., & Yang, M. S. (2020). Unsupervised K-Means *clustering* algorithm. IEEE Access, 8, 80716–80727. <https://doi.org/10.1109/ACCESS.2020.2988796>
- Soyusiawaty, D., Zakaria, Y. (2018). Book Data Content *Similarity Detector* With *Cosine Similarity* (Case study on [digilib.uad.ac.id](http://digilib.uad.ac.id)). 12th International Conference on Telecommunication Systems, Services, and Applications (TSSA), Yogyakarta, Indonesia, 2018, pp. 1-6, doi: 10.1109/TSSA.2018.8708758.
- Stevens, K., Andrzejewski, D., Buttler, D., & Kegelmeyer, P. (2012). Exploring *Topic Coherence* over many models and many topics. Association for Computational Linguistics. <http://mallet.cs.umass.edu/>
- Syed, S., & Spruit, M. (2017). Full-Text or abstract? Examining *topic coherence* scores using *latent dirichlet allocation*. *Proceedings - 2017 International Conference on Data Science and Advanced Analytics, DSAA 2017*, 2018-January, 165–174. <https://doi.org/10.1109/DSAA.2017.61>
- Tambunan, H. B., Barus, D. H., Hartono, J., Alam, A. S., Nugraha, D. A., & Usman, H. (2020). Electrical peak load *clustering* analysis using K-Means algorithm and silhouette coefficient. *Proceeding-2nd International Conference on Technology and Policy in Electric Power and Energy, ICT-PEP 2020*, 258–262. <https://doi.org/10.1109/ICT-PEP50916.2020.9249773>
- Teh, Y. W., Jordan, M. I., Beal, M. J., & Blei, D. M. (2006). *Hierarchical Dirichlet Processes*. *Journal of the American Statistical Association*, 101(476), 1566-1581.
- Thoomkuzhy, A. M. (2020). Drug Reviews: Cross-condition and Cross-source Analysis by Drug Reviews: Cross-condition and Cross-source Analysis by Review Quantification Using Regional CNN-LSTM Models Review Quantification Using Regional CNN-LSTM Models. <https://doi.org/10.21427/fcve-2g86>
- Vahidnia, S., Abbasi, A., & Abbass, H. A. (2021). Embedding-based Detection and Extraction of Research Topics from Academic Documents Using Deep *Clustering*. *Journal of Data and Information Science*, 6(3), 99–122. <https://doi.org/10.2478/jdis-2021-0024>
- Wakhidah, N. (2010). *Clustering Menggunakan K-Means Algorithm (K-Means Algorithm Clustering)*. *Jurnal Transformatika*.
- Wei, W., Nan, D., Zhang, L., Zhou, J., Wang, L., & Tang, X. (2020). Short text data model of secondary equipment faults in power grids based on LDA topic model and convolutional neural network. *Proceedings - 2020 35th Youth Academic Annual Conference of Chinese Association of Automation, YAC 2020*, 156–160. <https://doi.org/10.1109/YAC51587.2020.9337597>
- Wibowo, D. A., & Musdholifah, A. (2021). Sentiments Analysis of Indonesian Tweet about Covid-19 Vaccine Using Support Vector Machine and FastText Embedding. *2021 4th International Seminar on Research of Information Technology and Intelligent Systems, ISRITI 2021*, 184–188. <https://doi.org/10.1109/ISRITI54043.2021.9702871>

- Wu, X., & Li, C. (2019). "Short Text Topic modeling with Flexible Word Patterns," 2019 *International Joint Conference on Neural Networks (IJCNN)*, Budapest, Hungary, 2019, pp. 1-7, doi: 10.1109/IJCNN.2019.8852366.
- Xue, M. (2019). A Text Retrieval Algorithm Based on the Hybrid LDA and Word2Vec Model. *Proceedings - 2019 International Conference on Intelligent Transportation, Big Data and Smart City, ICITBS 2019*, 373–376. <https://doi.org/10.1109/ICITBS.2019.00098>
- Zhang, A., Lipton, Z.C., Li, M. and Smola, A. J. (2021). "Dive Into Deep Learning"
- Zhang, H., Huating, S., & Wu, X. (2020). Topic model for graph mining based on *hierarchical Dirichlet process*. *Statistical Theory and Related Fields*, 4(1), 66–77. <https://doi.org/10.1080/24754269.2019.1593098>
- Zhang, T., Chen, J., Zhan, X., Luo, X., Lo, D., & Jiang, H. (2021). Where2Change: Change Request Localization for App Reviews. *IEEE Transactions on Software Engineering*, 47(11), 2590–2616. <https://doi.org/10.1109/TSE.2019.2956941>