

DAFTAR PUSTAKA

- Adhitya, R. R., Wina Witanti, & Rezki Yuniarti. (2023). PERBANDINGAN METODE CART DAN NAÏVE BAYES UNTUK KLASIFIKASI CUSTOMER CHURN. *INFOTECH Journal*, 9(2), 307–318. <https://doi.org/10.31949/infotech.v9i2.5641>
- Aini, L. R., & Yulianti, E. (2025). Expertise Retrieval Using Adjusted TF-IDF and Keyword Mapping to ACM Classification Terms. *Jurnal RESTI*, 9(3), 497–505. <https://doi.org/10.29207/resti.v9i3.6397>
- Alami, N., Meknassi, M., Ennahahi, N., Adlouni, Y., & Ammor, O. (2021). Unsupervised neural networks for automatic Arabic text summarization using document clustering and topic modeling. *Expert Syst. Appl.*, 172, 114652. <https://doi.org/10.1016/j.eswa.2021.114652>
- Albalawi, R., Yeap, T., & Benyoucef, M. (2020). Using Topic Modeling Methods for Short-Text Data: A Comparative Analysis. *Frontiers in Artificial Intelligence*, 3. <https://doi.org/10.3389/frai.2020.00042>
- Apidianaki, M. (2022). From Word Types to Tokens and Back: A Survey of Approaches to Word Meaning Representation and Interpretation. *Computational Linguistics*, 49, 465–523. https://doi.org/10.1162/coli_a_00474
- Asudani, D. S., Nagwani, N. K., & Singh, P. (2023). Impact of word embedding models on text analytics in deep learning environment: a review. *Artificial Intelligence Review*, 1–81. <https://doi.org/10.1007/s10462-023-10419-1>
- Aufar, A. F., Mochamad Alfian Rosid, Eviyanti, A., & Astutik, I. R. I. (2023). Optimizing Text Preprocessing for Accurate Sentiment Analysis on E-Wallet Reviews. *JICTE (Journal of Information and Computer Technology Education)*, 7(2), 42–50. <https://doi.org/10.21070/jicte.v7i2.1650>
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research*, 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/tac1_a_00051
- Camacho-Collados, J., & Pilehvar, M. T. (2018). From Word To Sense Embeddings: A Survey on Vector Representations of Meaning. *Journal of Artificial Intelligence Research*, 63, 743–788. <https://doi.org/10.1613/jair.1.11259>

- Chai, C. P. (2023). Comparison of text preprocessing methods. *Natural Language Engineering*, 29(3), 509–553. <https://doi.org/10.1017/S1351324922000213>
- Choi, J., & Lee, S.-W. (2020). Improving FastText with inverse document frequency of subwords. *Pattern Recognition Letters*, 133, 165–172. <https://doi.org/10.1016/j.patrec.2020.03.003>
- Clear, A., & Parrish, A. (2021). Computing Curricula 2020. Paradigms for global computing education. In *Computing Curricula 2020*. <https://dl.acm.org/doi/book/10.1145/3467967>
- Dale, R., Moisl, H., & Somers, H. (2001). Handbook of Natural Language Processing. *Computational Linguistics*, 27(4), 602–603. <https://doi.org/10.1162/coli.2000.27.4.602>
- Dzallas, S., & Fincher, S. (2015). ACM Curriculum Reports. *Proceedings of the Eleventh Annual International Conference on International Computing Education Research*, 81–89. <https://doi.org/10.1145/2787622.2787714>
- Farhadpour, S., Warner, T. A., & Maxwell, A. E. (2024). Selecting and Interpreting Multiclass Loss and Accuracy Assessment Metrics for Classifications with Class Imbalance: Guidance and Best Practices. *Remote Sensing*, 16(3), 533. <https://doi.org/10.3390/rs16030533>
- Galhotra, S., Golshan, B., & Tan, W.-C. (2021). Adaptive Rule Discovery for Labeling Text Data. *Proceedings of the 2021 International Conference on Management of Data*, 2217–2225. <https://doi.org/10.1145/3448016.3457334>
- Garcia, M. (2021). Embeddings in Natural Language Processing: Theory and Advances in Vector Representations of Meaning. *Computational Linguistics*, 47(3), 699–701. https://doi.org/10.1162/coli_r_00410
- Goyal, A., & Kashyap, I. (2022). Latent Dirichlet Allocation - An approach for topic discovery. *2022 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COM-IT-CON)*, 1, 97–102. <https://doi.org/10.1109/COM-IT-CON54601.2022.9850912>
- Graves, A. (2012). *Long Short-Term Memory*. 37–45. https://doi.org/10.1007/978-3-642-24797-2_4

- He, B., Yang, Y., Wang, L., & Zhou, J. (2024). The Text Classification Method Based on BiLSTM and Multi-Scale CNN. *Computer Life*, 12(2), 43–49. <https://doi.org/10.54097/ypxxse31>
- Hingmire, S., & Chakraborti, S. (2014). Topic labeled text classification. *Proceedings of the 37th International ACM SIGIR Conference on Research & Development in Information Retrieval*, 385–394. <https://doi.org/10.1145/2600428.2609565>
- Jasmir, J., Riyadi, W., Agustini, S. R., Arvita, Y., Meisak, D., & Aryani, L. (2022). Bidirectional Long Short-Term Memory and Word Embedding Feature for Improvement Classification of Cancer Clinical Trial Document. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 6(4), 505–510. <https://doi.org/10.29207/resti.v6i4.4005>
- 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>
- Johnson, S. J., Murty, M. R., & Navakanth, I. (2023). A detailed review on word embedding techniques with emphasis on word2vec. *Multimedia Tools and Applications*, 83(13), 37979–38007. <https://doi.org/10.1007/s11042-023-17007-z>
- Jude Chukwura Obi. (2023). A comparative study of several classification metrics and their performances on data. *World Journal of Advanced Engineering Technology and Sciences*, 8(1), 308–314. <https://doi.org/10.30574/wjaets.2023.8.1.0054>
- Juna, M. F., & Hayaty, M. (2023). The observed preprocessing strategies for doing automatic text summarizing. *Computer Science and Information Technologies*, 4(2), 119–126. <https://doi.org/10.11591/csit.v4i2.p119-126>
- Koltcov, S., Surkov, A., Filippov, V., & Ignatenko, V. (2024). Topic models with elements of neural networks: investigation of stability, coherence, and determining the optimal number of topics. *PeerJ Computer Science*, 10, 1–41. <https://doi.org/10.7717/peerj-cs.1758>
- Laba, I. N., Supinganto, A., Martini, M., Lestari, R., Aribowo, E. K., Badi'ah, A., & Pramana, C. (2020). *Tips dan Trik publikasi jurnal bereputasi*. Bandung:... - Google Scholar (Number December).

- Lauriola, I., Lavelli, A., & Aiolli, F. (2021). An introduction to Deep Learning in Natural Language Processing: Models, techniques, and tools. *Neurocomputing*, 470, 443–456. <https://doi.org/10.1016/j.neucom.2021.05.103>
- Malashin, I., Tynchenko, V., Gantimurov, A., Nelyub, V., & Borodulin, A. (2024). Applications of Long Short-Term Memory (LSTM) Networks in Polymeric Sciences: A Review. *Polymers*, 16. <https://doi.org/10.3390/polym16182607>
- Mashtalir, S. V., & Nikolenko, O. V. (2023). Data preprocessing and tokenization techniques for technical Ukrainian texts. *Applied Aspects of Information Technology*, 6(3), 318–326. <https://doi.org/10.15276/aait.06.2023.22>
- Melamud, O., Goldberger, J., & Dagan, I. (2016). context2vec: Learning Generic Context Embedding with Bidirectional LSTM. *Proceedings of The 20th SIGNLL Conference on Computational Natural Language Learning*, 51–61. <https://doi.org/10.18653/v1/K16-1006>
- 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. <http://arxiv.org/abs/1301.3781>
- Mimno, D., Wallach, H. M., Talley, E., Leenders, M., & McCallum, A. (2011). Optimizing semantic coherence in topic models. *EMNLP 2011 - Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference*, (2), 262–272.
- Minaee, S., Kalchbrenner, N., Cambria, E., Nikzad, N., Chenaghlu, M., & Gao, J. (2022). Deep Learning--based Text Classification. *ACM Computing Surveys*, 54(3), 1–40. <https://doi.org/10.1145/3439726>
- MR ADEPU RAJESH, & DR TRYAMBAK HIWARKAR. (2023). Exploring Preprocessing Techniques for Natural LanguageText: A Comprehensive Study Using Python Code. *International Journal of Engineering Technology and Management Sciences*, 7(5), 390–399. <https://doi.org/10.46647/ijetms.2023.v07i05.047>
- Mursalin, Rohantizani, Fajriana, Azhari, T., & Satria, D. M. (2023). Sharing Session Publikasi Ilmiah dan Sitasi dengan Aplikasi Mendeley: Workshop Pengabdian Kolaborasi. *Jurnal Solusi Masyarakat Dikara*, 3(3), 155–161.

- <http://jsmd.dikara.org/jsmd/article/view/85%0Ahttps://jsmd.dikara.org/jsmd/article/download/85/95>
- Nasution, M. K. M. (2022). Publikasi Ilmiah-Artikel di Jurnal Bereputasi: Suatu Strategi. *Publikasi Artikel Di Jurnal Internasional*, (November). <https://doi.org/10.13140/RG.2.2.22282.54722>
- Pion-Tonachini, L., Makeig, S., & Kreutz-Delgado, K. (2017). Crowd labeling latent Dirichlet allocation. *Knowledge and Information Systems*, 53(3), 749–765. <https://doi.org/10.1007/s10115-017-1053-1>
- Prabhakar, S. K., & Won, D.-O. (2021). Medical Text Classification Using Hybrid Deep Learning Models with Multihead Attention. *Computational Intelligence and Neuroscience*, 2021(1). <https://doi.org/10.1155/2021/9425655>
- Raghuvanshi, K. P. (2024). A Systematic Literature Review on The Role of LSTM Networks in Capturing Temporal Dependencies in Data Mining Algorithms. *International Journal for Research in Applied Science and Engineering Technology*, 12(10), 1219–1224. <https://doi.org/10.22214/ijraset.2024.64761>
- Rahman, M. M., Shiplu, A. I., Watanobe, Y., & Alam, M. A. (2025). RoBERTa-BiLSTM: A Context-Aware Hybrid Model for Sentiment Analysis. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 9(6), 3788–3805. <https://doi.org/10.1109/TETCI.2025.3572150>
- Rehurek, R., & Sojka, P. (2010). Software Framework for Topic Modelling with Large Corpora. *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*, 45–50.
- Sadat, M., & Caragea, C. (2022). Hierarchical Multi-Label Classification of Scientific Documents. *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing, EMNLP 2022*, 8923–8937. <https://doi.org/10.18653/v1/2022.emnlp-main.610>
- Septianda, M. R., Widodo, & Nugraheni, M. (2025). Identification of Topic Areas by Scientific Publication Using K-Means on LDA. *The 2nd International Conference of Information System & Information Technology*.
- Siino, M., Tinnirello, I., & La Cascia, M. (2024). Is text preprocessing still worth the time? A comparative survey on the influence of popular preprocessing methods on

- Transformers and traditional classifiers. *Information Systems*, 121, 102342. <https://doi.org/10.1016/j.is.2023.102342>
- Siti Khomsah, Rima Dias Ramadhani, & Sena Wijaya. (2022). The Accuracy Comparison Between Word2Vec and FastText On Sentiment Analysis of Hotel Reviews. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 6(3), 352–358. <https://doi.org/10.29207/resti.v6i3.3711>
- Smit, A., Jain, S., Rajpurkar, P., Pareek, A., Ng, A., & Lungren, M. (2020). Combining Automatic Labelers and Expert Annotations for Accurate Radiology Report Labeling Using BERT. *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 1500–1519. <https://doi.org/10.18653/v1/2020.emnlp-main.117>
- Syahrial, S., & Afidh, R. P. F. (2024). Fine-Tuning Topic Modelling: A Coherence-Focused Analysis of Correlated Topic Models. *Infolitika Journal of Data Science*, 2(2), 82–87. <https://doi.org/10.60084/ijds.v2i2.236>
- Syed, S., & Spruit, M. (2017). Full-Text or Abstract? Examining Topic Coherence Scores Using Latent Dirichlet Allocation. *2017 IEEE International Conference on Data Science and Advanced Analytics (DSAA)*, 165–174. <https://doi.org/10.1109/DSAA.2017.61>
- Tulu, C. (2022). Experimental Comparison of Pre-Trained Word Embedding Vectors of Word2Vec, Glove, FastText for Word Level Semantic Text Similarity Measurement in Turkish. *Advances in Science and Technology Research Journal*, 16(4), 147–156. <https://doi.org/10.12913/22998624/152453>
- Wei, H., Gao, M., Zhou, A., Chen, F., Qu, W., Wang, C., & Lu, M. (2019). Named Entity Recognition From Biomedical Texts Using a Fusion Attention-Based BiLSTM-CRF. *IEEE Access*, 7, 73627–73636. <https://doi.org/10.1109/ACCESS.2019.2920734>
- Williams, S. B. (1954). The Association for Computing Machinery. *Journal of the ACM (JACM)*, 1(1), 1–3. <https://doi.org/10.1145/320764.320765>
- Yu, Y., Si, X., Hu, C., & Zhang, J. (2019). A Review of Recurrent Neural Networks: LSTM Cells and Network Architectures. *Neural Computation*, 31(7), 1235–1270. https://doi.org/10.1162/neco_a_01199

- Zhu, L., & Luo, D. (2023). A Novel Efficient and Effective Preprocessing Algorithm for Text Classification. *Journal of Computer and Communications*.
<https://doi.org/102342>
- Zulqarnain, M., Ghazali, R., Hassim, Y., & Rehan, M. (2020). A comparative review on deep learning models for text classification. *Indonesian Journal of Electrical Engineering and Computer Science*, 19, 325–335.
<https://doi.org/10.11591/ijeecs.v19.i1.pp325-335>

