

BIBLIOGRAPHY

- 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
- Chung, H., & Shin, K. S. (2018). Genetic algorithm-optimized long short-term memory network for stock market prediction. *Sustainability (Switzerland)*, 10(10). <https://doi.org/10.3390/su10103765>
- Collobert, R., Weston, J., Bottou, L., Karlen, M., Kavukcuoglu, K., & Kuksa, P. (2011). Natural language processing (almost) from scratch. *Journal of Machine Learning Research*, 12, 2493–2537.
- D'Sa, A. G., Illina, I., & Fohr, tD. (2020). BERT and fastText Embeddings for Automatic Detection of Toxic Speech. *Proceedings of 2020 International Multi-Conference on: Organization of Knowledge and Advanced Technologies, OCTA 2020*. <https://doi.org/10.1109/OCTA49274.2020.9151853>
- Deerwester, S. C., Dumais, S. T., Landauer, T. K., Furnas, G. W., & Harshman, R. A. (1990). Indexing by Latent Semantic Analysis. *J. Am. Soc. Inf. Sci.*, 41, 391–407.
- Deniz, Erbay, Cosar (2022). Multi-Label Classification of E-Commerce Customer Reviews via Machine Learning via Machine Learning. *Axioms* 2022, 11, 436. <https://doi.org/10.3390/axioms11090436>
- Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. *NAACL HLT 2019 - 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference*, 1(Mlm), 4171–4186.
- Duque, A. B., Santos, L. L. J., Macêdo, D., & Zanchettin, C. (2019). Squeezed Very Deep Convolutional Neural Networks for Text Classification. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 11727 LNCS(2001), 193–207. https://doi.org/10.1007/978-3-030-30487-4_16
- Gerth, T. (2021). A Comparison of Word Embedding Techniques for Similarity Analysis. *Computer Science*. <https://scholarworks.uark.edu/csceuht>
- Grave, E., Bojanowski, P., Gupta, P., Joulin, A., & Mikolov, T. (2019). Learning word vectors for 157 languages. *LREC 2018 - 11th International Conference on Language Resources and Evaluation*, 3483–3487.
- Graves, A., & Schmidhuber, J. (2005). Framewise phoneme classification with bidirectional LSTM and other neural network architectures. *Neural Networks*, 18(5), 602–610. <https://doi.org/https://doi.org/10.1016/j.neunet.2005.06.042>

- Hermawan, L., & Bellanar Ismiati, M. (2020). Pembelajaran Text Preprocessing berbasis Simulator Untuk Mata Kuliah Information Retrieval. *Jurnal Transformatika*, 17(2), 188. <https://doi.org/10.26623/transformatika.v17i2.1705>
- Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, 9(8), 1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
- Imaduddin, H., Widyawan, & Fauziati, S. (2019). Word embedding comparison for Indonesian language sentiment analysis. *Proceeding - 2019 International Conference of Artificial Intelligence and Information Technology, ICAIIT 2019*, 426–430. <https://doi.org/10.1109/ICAIIIT.2019.8834536>
- Indrapurasih, R. D., Bijaksana, M. A., Sardi, I. L., & Belakang, L. (2018). Implementasi dan Analisis Kesamaan Semantik Antar Kata Bahasa Indonesia Menggunakan Metode GloVe Pendahuluan Studi Terkait Semantic Similarity. *EProceedings of Engineering*, 5(3), 7699–7706.
- Jang, B., Kim, I., & Kim, J. W. (2019). Word2vec convolutional neural networks for classification of news articles and tweets. *PLoS ONE*, 14(8), 1–20. <https://doi.org/10.1371/journal.pone.0220976>
- Joulin, A., Grave, E., Bojanowski, P., & Mikolov, T. (2017). Bag of tricks for efficient text classification. *15th Conference of the European Chapter of the Association for Computational Linguistics, EACL 2017 - Proceedings of Conference*, 2, 427–431. <https://doi.org/10.18653/v1/e17-2068>
- Lai, S., Liu, K., He, S., & Zhao, J. (2016). How to generate a good word embedding. *IEEE Intelligent Systems*, 31(6), 5–14. <https://doi.org/10.1109/MIS.2016.45>
- Li, H., Li, X., Caragea, D., & Caragea, C. (2018). Comparison of Word Embeddings and Sentence Encodings as Generalized Representations for Crisis Tweet Classification Tasks. *Proceedings of the ISCRAM Asian Pacific 2018 Conference, November*, 1–13.
- Li, S. and Ou, J. (2021) “Multi-label classification of research papers using multi-label K-nearest neighbour algorithm,” *Journal of Physics: Conference Series*, 1994(1), p. 012031. Available at: <https://doi.org/10.1088/1742-6596/1994/1/012031>.
- Manalu, L. N. T., Bijaksana, M. A., & Suryani, A. A. (2019). Analysis of the Word2Vec model for semantic similarities in Indonesian words. *2019 7th International Conference on Information and Communication Technology, ICoICT 2019*. <https://doi.org/10.1109/ICoICT.2019.8835330>
- Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. *ArXiv Preprint ArXiv:1301.3781*.
- Mikolov, T., Yih, W., & Zweig, G. (2013). Linguistic Regularities in Continuous Space Word Representations. *Proceedings of the 2013 Conference of the*

- North {A}merican Chapter of the Association for Computational Linguistics: Human Language Technologies*, 746–751. <https://aclanthology.org/N13-1090>
- Naili, M., Chaibi, A. H., & Ben Ghezala, H. H. (2017). Comparative study of word embedding methods in topic segmentation. *Procedia Computer Science*, 112, 340–349. <https://doi.org/10.1016/j.procs.2017.08.009>
- Nurdin, A., Anggo Seno Aji, B., Bustamin, A., & Abidin, Z. (2020). Perbandingan Kinerja Word Embedding Word2Vec, Glove, Dan Fasttext Pada Klasifikasi Teks. *Jurnal Tekno Kompak*, 14(2), 74. <https://doi.org/10.33365/jtk.v14i2.732>
- Pennington, J., Socher, R., & Manning, C. D. (2014). GloVe: Global vectors for word representation. *EMNLP 2014 - 2014 Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference, June 2018*, 1532–1543. <https://doi.org/10.3115/v1/d14-1162>
- Reimers, N., & Gurevych, I. (2020). Sentence-BERT: Sentence embeddings using siamese BERT-networks. *EMNLP-IJCNLP 2019 - 2019 Conference on Empirical Methods in Natural Language Processing and 9th International Joint Conference on Natural Language Processing, Proceedings of the Conference*, 3982–3992. <https://doi.org/10.18653/v1/d19-1410>
- Roy, A. (2021). *Recent Trends in Named Entity Recognition (NER)*. 1–27. <http://arxiv.org/abs/2101.11420>
- Susanty, M., & Sukardi, S. (2021). Perbandingan Pre-trained Word Embedding dan Embedding Layer untuk Named-Entity Recognition Bahasa Indonesia. *Petir*, 14(2), 247–257. <https://doi.org/10.33322/petir.v14i2.1164>
- Symeonidis & Duong, Huu-Thanh (2021). A review: preprocessing techniques and data augmentation for sentiment analysis. <https://doi.org/10.1186/s40649-020-00080-x>
- Azeema Sadia, Fariha Khan, Fatima Bashir (2018) An Overview of Lexicon-Based Approach For Sentiment Analysis https://ieec.neduet.edu.pk/2018/Papers_2018/15.pdf
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems, 2017-Decem(Nips)*, 5999–6009.
- Wang, B., Wang, A., Chen, F., Wang, Y., & Kuo, C. C. J. (2019). Evaluating word embedding models: Methods and experimental results. *APSIPA Transactions on Signal and Information Processing*, 8(2019). <https://doi.org/10.1017/ATSIP.2019.12>
- Wilie, B., Vincentio, K., Winata, G. I., Cahyawijaya, S., Li, X., Lim, Z. Y., Soleman, S., Mahendra, R., Fung, P., Bahar, S., & Purwarianti, A. (2020). *IndoNLU: Benchmark and Resources for Evaluating Indonesian Natural Language Understanding*. 843–857. <http://arxiv.org/abs/2009.05387>

- Xia M, Kong X, Anastasopoulos A, Neubig G. Generalized Data Augmentation for Low-Resource Translation, in Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, Florence. 2019; pp. 5786–5796. <https://doi.org/10.18653/v1/P19-1579>.
- Xiao, Y., & Cho, K. (2016). *Efficient Character-level Document Classification by Combining Convolution and Recurrent Layers*. <http://arxiv.org/abs/1602.00367>
- Zhang, X., Zhao, J., & LeCun, Y. (2015). Character-level Convolutional Networks for Text Classification. In C. Cortes, N. Lawrence, D. Lee, M. Sugiyama, & R. Garnett (Eds.), *Advances in Neural Information Processing Systems* (Vol. 28). Curran Associates, Inc. <https://proceedings.neurips.cc/paper/2015/file/250cf8b51c773f3f8dc8b4be867a9a02-Paper.pdf>
- Zhao, Z., Chen, W., Wu, X., Chen, P. C. Y., & Liu, J. (2017). LSTM network: a deep learning approach for short-term traffic forecast. *IET Intelligent Transport Systems*, 11(2), 68–75. <https://doi.org/https://doi.org/10.1049/iet-its.2016.0208>