

DAFTAR PUSTAKA

- Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., ... Zheng, X. (2016). *TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems*.
<https://doi.org/10.48550/arxiv.1603.04467>
- Alencar, R. (2019). *Dealing with very small datasets*. Kaggle.
<https://www.kaggle.com/code/rafjaa/dealing-with-very-small-datasets/notebook>
- Ba, J. L., Kiros, J. R., & Hinton, G. E. (2016). *Layer Normalization*.
<http://arxiv.org/abs/1607.06450>
- Bahdanau, D., Cho, K. H., & Bengio, Y. (2015). Neural machine translation by jointly learning to align and translate. *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*.
<https://doi.org/10.48550/arxiv.1409.0473>
- Bañón, M., Chen, P., Haddow, B., Heafield, K., Hoang, H., Esplà-Gomis, M., Forcada, M. L., Kamran, A., Kirefu, F., Koehn, P., Ortiz Rojas, S., Pla Sempere, L., Ramírez-Sánchez, G., Sarrías, E., Strelec, M., Thompson, B., Waites, W., Wiggins, D., & Zaragoza, J. (2020). ParaCrawl: Web-Scale Acquisition of Parallel Corpora. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 4555–4567.
<https://doi.org/10.18653/v1/2020.acl-main.417>
- Bengio, Y., Courville, A., & Vincent, P. (2013). Representation learning: A review and new perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(8), 1798–1828.
<https://doi.org/10.1109/TPAMI.2013.50>
- Bengio, Y., Ducharme, R., Vincent, P., & Janvin, C. (2003). A Neural Probabilistic Language Model. *J. Mach. Learn. Res.*, 3(null), 1137–1155.
- Chen, B., & Cherry, C. (2014). A Systematic Comparison of Smoothing Techniques for Sentence-Level BLEU. *Proceedings of the Ninth Workshop on*

- Statistical Machine Translation*, 362–367.
<https://doi.org/10.3115/v1/W14-3346>
- Cho, K., van Merriënboer, B., Bahdanau, D., & Bengio, Y. (2014). On the Properties of Neural Machine Translation: Encoder-Decoder Approaches. *Proceedings of SSST 2014 - 8th Workshop on Syntax, Semantics and Structure in Statistical Translation*, 103–111.
<https://doi.org/10.48550/arxiv.1409.1259>
- Church, K. W., & Hanks, P. (1989). Word association norms, mutual information, and lexicography. *Proceedings of the 27th Annual Meeting on Association for Computational Linguistics* -, 76–83.
<https://doi.org/10.3115/981623.981633>
- Collins, J. T. (2014). Keragaman Bahasa Dan Kesepakatan Masyarakat: Pluralitas Dan Komunikasi. *Keragaman Bahasa Dan Kesepakatan Masyarakat: Pluralitas Dan Komunikasi*, 1(2), 149–180.
- Collobert, R., Weston, J., Bottou, L., Karlen, M., Kavukcuoglu, K., & Kuksa, P. (2011). Natural Language Processing (Almost) from Scratch. *J. Mach. Learn. Res.*, 12(null), 2493–2537.
- Dagan, I., Marcus, S., & Markovitch, S. (1993). Contextual word similarity and estimation from sparse data. *Proceedings of the 31st Annual Meeting on Association for Computational Linguistics* -, 164–171.
<https://doi.org/10.3115/981574.981596>
- Denkowski, M., & Lavie, A. (2011). Meteor 1.3: Automatic Metric for Reliable Optimization and Evaluation of Machine Translation Systems. *Proceedings of the Sixth Workshop on Statistical Machine Translation*, 85–91. <https://aclanthology.org/W11-2107>
- Elman, J. L. (1990). Finding Structure in Time. *Cognitive Science*, 14(2), 179–211.
<https://doi.org/https://doi.org/10.1207/s15516709cog1402%1>
- Folkman, T. (2019). *How To Use Deep Learning Even with Small*. Towards Data Science. <https://towardsdatascience.com/how-to-use-deep-learning-even-with-small-data-e7f34b673987>
- Gal, Y., & Ghahramani, Z. (2015). Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning. *33rd International*

- Conference on Machine Learning, ICML 2016*, 3, 1651–1660.
<https://doi.org/10.48550/arxiv.1506.02142>
- Gehring, J., Auli, M., Grangier, D., Yarats, D., & Dauphin, Y. N. (2017). Convolutional sequence to sequence learning. *34th International Conference on Machine Learning, ICML 2017*, 3, 2029–2042.
<https://doi.org/10.48550/arxiv.1705.03122>
- Google Developer. (2022). *Validation Set: Another Partition*. Google Machine Learning Courses. <https://developers.google.com/machine-learning/crash-course/validation/another-partition>
- Goyal, N., Gao, C., Chaudhary, V., Chen, P.-J., Wenzek, G., Ju, D., Krishnan, S., Ranzato, M., Guzman, F., & Fan, A. (2021). *The FLORES-101 Evaluation Benchmark for Low-Resource and Multilingual Machine Translation*.
<http://arxiv.org/abs/2106.03193>
- Graves, A. (2013). *Generating Sequences With Recurrent Neural Networks*.
<https://doi.org/10.48550/arxiv.1308.0850>
- Gunawan, W., Sujaini, H., & Tursina, T. (2021). Analisis Perbandingan Nilai Akurasi Mekanisme Attention Bahdanau dan Luong pada Neural Machine Translation Bahasa Indonesia ke Bahasa Melayu Ketapang dengan Arsitektur Recurrent Neural Network. *Jurnal Edukasi Dan Penelitian Informatika (JEPIN)*, 7(3), 488. <https://doi.org/10.26418/jp.v7i3.50287>
- He, K., Zhang, X., Ren, S., & Sun, J. (2015). Deep Residual Learning for Image Recognition. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2016-Decem*, 770–778.
<https://doi.org/10.1109/CVPR.2016.90>
- Hihi, S. El, & Bengio, Y. (1995). Hierarchical Recurrent Neural Networks for Long-Term Dependencies. In D. Touretzky, M. C. Mozer, & M. Hasselmo (Eds.), *Nips* (Vol. 8, pp. 493–499). MIT Press.
<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.38.7574&rep=rep1&type=pdf>
- Hinton, G. E., Srivastava, N., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. R. (2012). *Improving neural networks by preventing co-adaptation of feature detectors*. <https://doi.org/10.48550/arxiv.1207.0580>

- Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, 9, 1735–1780.
- Inan, H., Khosravi, K., & Socher, R. (2016). Tying Word Vectors and Word Classifiers: A Loss Framework for Language Modeling. *5th International Conference on Learning Representations, ICLR 2017 - Conference Track Proceedings*. <https://doi.org/10.48550/arxiv.1611.01462>
- Ioffe, S., & Szegedy, C. (2015). Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. *32nd International Conference on Machine Learning, ICML 2015*, 1, 448–456. <http://arxiv.org/abs/1502.03167>
- Junczys-Dowmunt, M., Grundkiewicz, R., Dwojak, T., Hoang, H., Heafield, K., Neckermann, T., Seide, F., Germann, U., Aji, A. F., Bogoychev, N., Martins, A. F. T., & Birch, A. (2018). Marian: Fast Neural Machine Translation in C++. *ACL 2018 - 56th Annual Meeting of the Association for Computational Linguistics, Proceedings of System Demonstrations*, 116–121. <https://doi.org/10.18653/v1/p18-4020>
- Jurafsky, D., & Martin, J. (2022). Speech and Language Processing. In *Speech and Language Processing*. (3rd ed.). <https://web.stanford.edu/~jurafsky/slp3/>
- Kim, Y. J., Junczys-Dowmunt, M., Hassan, H., Fikri Aji, A., Heafield, K., Grundkiewicz, R., & Bogoychev, N. (2019). From Research to Production and Back: Ludicrously Fast Neural Machine Translation. *Proceedings of the 3rd Workshop on Neural Generation and Translation*, 280–288. <https://doi.org/10.18653/v1/D19-5632>
- Kingma, D. P., & Ba, J. (2014). Adam: A Method for Stochastic Optimization. *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*. <http://arxiv.org/abs/1412.6980>
- Kingma, D. P., & Ba, J. L. (2015). Adam: A method for stochastic optimization. *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, 1–15.
- Koehn, P. (2005). Europarl : A Parallel Corpus for Statistical Machine Translation. *MT Summit*, 11, 79--86. <https://aclanthology.org/2005.mtsummit-papers.11/>

- Kudo, T. (2018). Subword Regularization: Improving Neural Network Translation Models with Multiple Subword Candidates. *ACL 2018 - 56th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference (Long Papers), 1*, 66–75. <https://doi.org/10.48550/arxiv.1804.10959>
- Kudo, T., & Richardson, J. (2018). SentencePiece: A simple and language independent subword tokenizer and detokenizer for neural text processing. *EMNLP 2018 - Conference on Empirical Methods in Natural Language Processing: System Demonstrations, Proceedings*, 66–71. <https://doi.org/10.18653/v1/d18-2012>
- Liddy, E. (2001). Natural Language Processing. In *Encyclopedia of Library and Information Science* (2nd ed.). Marcel Decker, Inc. <https://surface.syr.edu/istpub/63>
- Lison, P., & Tiedemann, J. (2016). OpenSubtitles2016: Extracting large parallel corpora from movie and TV subtitles. *Proceedings of the 10th International Conference on Language Resources and Evaluation, LREC 2016*, 923–929. <https://aclanthology.org/L16-1147>
- Luhn, H. P. (1957). A Statistical Approach to Mechanized Encoding and Searching of Literary Information. *IBM Journal of Research and Development*, 1(4), 309–317. <https://doi.org/10.1147/rd.14.0309>
- Maheswari, J. P. (2018). *Breaking the curse of small datasets in Machine Learning: Part 1*. Towards Data Science. <https://towardsdatascience.com/breaking-the-curse-of-small-datasets-in-machine-learning-part-1-36f28b0c044d>
- Marcin and Pouliquen, B. Z. M. and J.-D. (2016). The United Nations Parallel Corpus v1.0. *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, 3530–3534. <https://aclanthology.org/L16-1561>
- Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013a). Distributed Representations of Words and Phrases and their Compositionality arXiv : 1310 . 4546v1 [cs . CL] 16 Oct 2013. In C. J. C. Burges, L. Bottou, Z. Ghahramani, & K. Q. Weinberger (Eds.), *arXiv preprint arXiv:1310.4546: Vol. cs.CL (pp. 1–9)*.

<http://arxiv.org/abs/1310.4546v1%0Apapers2://publication/uuid/FB1742A3-202C-44CA-98F5-6EA51EC019D2>

- Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013b). Efficient estimation of word representations in vector space. In Y. Bengio & Y. LeCun (Eds.), *1st International Conference on Learning Representations, ICLR 2013 - Workshop Track Proceedings*. <http://arxiv.org/abs/1301.3781>
- Mikolov, T., Karafiát, M., Burget, L., Cernocký, J. H., & Khudanpur, S. (2010). Recurrent neural network based language model. *INTERSPEECH*.
- Nguyen, T., & Chiang, D. (2018). Improving Lexical Choice in Neural Machine Translation. *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, 334–343. <https://doi.org/10.18653/v1/N18-1031>
- Niwa, Y., & Nitta, Y. (1994). Co-occurrence vectors from corpora vs. distance vectors from dictionaries. *Proceedings of the 15th International Conference On Computational Linguistics*, 304–309. <https://aclanthology.org/C94-1049>
- Pascanu, R., Gulcehre, C., Cho, K., & Bengio, Y. (2013). How to Construct Deep Recurrent Neural Networks. *2nd International Conference on Learning Representations, ICLR 2014 - Conference Track Proceedings*. <https://doi.org/10.48550/arxiv.1312.6026>
- Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Köpf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., ... Chintala, S. (2019). PyTorch: An Imperative Style, High-Performance Deep Learning Library. *Advances in Neural Information Processing Systems*, 32. <https://doi.org/10.48550/arxiv.1912.01703>
- Post, M. (2018). A Call for Clarity in Reporting BLEU Scores. *WMT 2018 - 3rd Conference on Machine Translation, Proceedings of the Conference, 1*, 186–191. <https://doi.org/10.18653/v1/w18-6319>
- Press, O., & Wolf, L. (2016). Using the Output Embedding to Improve Language Models. *15th Conference of the European Chapter of the Association for*

- Computational Linguistics, EACL 2017 - Proceedings of Conference*, 2, 157–163. <https://doi.org/10.48550/arxiv.1608.05859>
- Ranzato, M., Chopra, S., Auli, M., & Zaremba, W. (2015). Sequence Level Training with Recurrent Neural Networks. *4th International Conference on Learning Representations, ICLR 2016 - Conference Track Proceedings*. <https://doi.org/10.48550/arxiv.1511.06732>
- Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, 323(6088), 533–536. <https://doi.org/10.1038/323533a0>
- Salton, G. (1971). *The SMART Retrieval System—Experiments in Automatic Document Processing*. Prentice-Hall, Inc.
- Schmidhuber, J. (1992). Learning Complex, Extended Sequences Using the Principle of History Compression. *Neural Computation*, 4(2), 234–242. <https://doi.org/10.1162/neco.1992.4.2.234>
- Schuster, M., & Nakajima, K. (2012). Japanese and Korean Voice Search. *International Conference on Acoustics, Speech and Signal Processing*, 5149–5152.
- Schuster, M., & Paliwal, K. K. (1997). Bidirectional recurrent neural networks. *IEEE Trans. Signal Process.*, 45, 2673–2681.
- Sennrich, R., Firat, O., Cho, K., Birch, A., Haddow, B., Hirschler, J., Junczys-Dowmunt, M., Läubli, S., Barone, A. V. M., Mokry, J., & Nădejde, M. (2017). *Nematus: a Toolkit for Neural Machine Translation*. <http://arxiv.org/abs/1703.04357>
- Sennrich, R., & Haddow, B. (2016). Linguistic Input Features Improve Neural Machine Translation. *Proceedings of the First Conference on Machine Translation: Volume 1, Research Papers*, 83–91. <https://doi.org/10.18653/v1/W16-2209>
- Sennrich, R., Haddow, B., & Birch, A. (2016a). Neural Machine Translation of Rare Words with Subword Units. *54th Annual Meeting of the Association for Computational Linguistics, ACL 2016 - Long Papers*, 3, 1715–1725. <https://doi.org/10.18653/V1/P16-1162>
- Sennrich, R., Haddow, B., & Birch, A. (2016b). Edinburgh Neural Machine

- Translation Systems for WMT 16. *Proceedings of the First Conference on Machine Translation: Volume 2, Shared Task Papers*, 371–376. <https://doi.org/10.18653/v1/w16-2323>
- Sennrich, R., & Zhang, B. (2019). Revisiting Low-Resource Neural Machine Translation: A Case Study. *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 211–221. <https://doi.org/10.18653/v1/P19-1021>
- Shen, S., Cheng, Y., He, Z., He, W., Wu, H., Sun, M., & Liu, Y. (2016). Minimum Risk Training for Neural Machine Translation. *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 1683–1692. <https://doi.org/10.18653/v1/P16-1159>
- SPARCK JONES, K. (1972). A STATISTICAL INTERPRETATION OF TERM SPECIFICITY AND ITS APPLICATION IN RETRIEVAL. *Journal of Documentation*, 28(1), 11–21. <https://doi.org/10.1108/eb026526>
- Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *Journal of Machine Learning Research*, 15(56), 1929–1958. <http://jmlr.org/papers/v15/srivastava14a.html>
- Stahlberg, F. (2019). *Neural Machine Translation: A Review and Survey*. <https://scholar.google.com/scholar?q=%22neural+machine+>
- Stahlberg, F. (2020). Neural machine translation: A review. *Journal of Artificial Intelligence Research*, 69, 343–418. <https://doi.org/10.1613/JAIR.1.12007>
- Stanojevic, M., & Sima'an, K. (2014). BEER: BEtter Evaluation as Ranking. *Proceedings of the Ninth Workshop on Statistical Machine Translation*, 414–419. <https://doi.org/10.3115/v1/W14-3354>
- Sutskever, I., Vinyals, O., & Le, Q. V. (2014). Sequence to Sequence Learning with Neural Networks. *Advances in Neural Information Processing Systems*, 4(January), 3104–3112. <https://doi.org/10.48550/arxiv.1409.3215>
- Tan, Z., Wang, S., Yang, Z., Chen, G., Huang, X., Sun, M., & Liu, Y. (2020). Neural machine translation: A review of methods, resources, and tools. *AI Open*, 1, 5–21. <https://doi.org/10.1016/j.aiopen.2020.11.001>
- Tiedemann, J. (2012). Parallel data, tools and interfaces in OPUS. *Proceedings of*

- the 8th International Conference on Language Resources and Evaluation, LREC* 2012, 2214–2218.
<http://opus.lingfil.uu.se/OpenSubtitles2011/wordalign/>
- Tieleman, T., Hinton, G., & others. (2012). Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude. *COURSERA: Neural Networks for Machine Learning*, 4(2), 26–31.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention Is All You Need. *Advances in Neural Information Processing Systems, 2017-Decem*, 5999–6009.
<http://arxiv.org/abs/1706.03762>
- Wdowiak, E. (2021). *Sicilian Translator: A Recipe for Low-Resource NMT Arba Sicula*.
- Wu, Y., Schuster, M., Chen, Z., Le, Q. V., Norouzi, M., Macherey, W., Krikun, M., Cao, Y., Gao, Q., Macherey, K., Klingner, J., Shah, A., Johnson, M., Liu, X., Kaiser, L., Gouws, S., Kato, Y., Kudo, T., Kazawa, H., ... Dean, J. (2016). *Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation*.
<http://arxiv.org/abs/1609.08144>
- Zeiler, M. D. (2012). *ADADELTA: An Adaptive Learning Rate Method*.
<https://doi.org/10.48550/arxiv.1212.5701>
- Zilly, J. G., Srivastava, R. K., Koutník, J., & Schmidhuber, J. (2016). Recurrent Highway Networks. *34th International Conference on Machine Learning, ICML 2017*, 8, 6346–6357. <https://doi.org/10.48550/arxiv.1607.03474>