

## DAFTAR PUSTAKA

- Abiodun, O. I., Jantan, A., Esther Omolara, A., Victoria Dada, K., AbdElatif Mohamed, N., & Arshad, H. (2018). State-of-the-Art in Artificial Neural Network Applications: A Survey. *Heliyon*, 4, 938. <https://doi.org/10.1016/j.heliyon.2018>
- Afzal, S., Asim, M., Beg, M. O., Baker, T., Awad, A. I., & Shamim, N. (2024). Context-aware Embeddings for Robust Multiclass Fraudulent URL Detection in Online Social Platforms. *Computers and Electrical Engineering*, 119. <https://doi.org/10.1016/j.compeleceng.2024.109494>
- Ahmadi, W., Zahra, A., & Salsabila. (2024). Ragam Bahasa Gaul Generasi Z di Media Sosial Twitter. *Jurnal Cahaya Mandalika*, 5(1). <http://ojs.cahayamandalika.com/index.php/JCM> (Diakses pada 12 November 2025)
- Albab, M. U., P., Y. K., & Fawaiq, M. N. (2023). Optimization of the Stemming Technique on Text Preprocessing President 3 Periods Topic. *Jurnal Transformatika*, 20(2), 1–12. <https://doi.org/10.26623/transformatika.v20i2.5374>
- Alkhulaifi, A., Alsahli, F., & Ahmad, I. (2021). Knowledge Distillation in Deep Learning and its Applications. *PeerJ Computer Science*, 7, 1–24. <https://doi.org/10.7717/peerj-cs.474>
- Alzubaidi, L., Zhang, J., Humaidi, A. J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., Santamaría, J., Fadhel, M. A., Al-Amidie, M., & Farhan, L. (2021). Review of Deep Learning: Concepts, CNN Architectures, Challenges, Applications, Future Directions. *Journal of Big Data*, 8(1). <https://doi.org/10.1186/s40537-021-00444-8>
- Amien, M. (2023). Sejarah dan Perkembangan Teknik Natural Language Processing (NLP) Bahasa Indonesia: Tinjauan Tentang Sejarah, Perkembangan Teknologi, dan Aplikasi NLP dalam Bahasa Indonesia. *ELANG: Journal of Interdisciplinary Research*. (Diakses pada 13 November 2025)
- APJII. (2024). *APJII Jumlah Pengguna Internet Indonesia Tembus 221 Juta Orang*. Laporan Survei Penetrasi Internet Indonesia. <https://apjii.or.id/berita/d/apjii-jumlah-pengguna-internet-indonesia-tembus-221-juta-orang> (Diakses pada 13 November 2025)
- Ashari, S. A., Taufik Bau, R. R., & Suhada, S. (2023). Eksplorasi Intensitas Penggunaan Sosial Media (Studi Deskriptif pada Mahasiswa Teknik Informatika UNG). *INVERTED: Journal of Information Technology Education*, 3(2). <http://ejournal.ung.ac.id/index.php/inverted> (Diakses pada 17 November 2025)
- Ba, J. L., Kiros, J. R., & Hinton, G. E. (2016). *Layer Normalization*. <http://arxiv.org/abs/1607.06450> (Diakses pada 11 November 2025)

- Baranandita, F., & Asfari, N. A. B. (2022). Remaja, Media Sosial, dan Cyberbullying: Kajian Literatur. *Jurnal Flourishing*, 2(10), 650–655. <https://doi.org/10.17977/10.17977/um070v2i102022p650-655>
- Bergstra, J., & Bengio, Y. (2012). Random Search for Hyper-Parameter Optimization Yoshua Bengio. *Journal of Machine Learning Research*, 13, 281–305.
- Berrar, D. (2018). Cross-validation. *Encyclopedia of Bioinformatics and Computational Biology: ABC of Bioinformatics* (Vol. 1–3, hlm. 542–545). Elsevier. <https://doi.org/10.1016/B978-0-12-809633-8.20349-X>
- Berutu, S. S., Budiati, H., Jatmika, J., & Gulo, F. (2023). Data Preprocessing Approach for Machine Learning-based Sentiment Classification. *JURNAL INFOTEL*, 15(4), 317–325. <https://doi.org/10.20895/infotel.v15i4.1030>
- Bohr, A., & Memarzadeh, K. (2020). The Rise of Artificial Intelligence in Healthcare Applications. *Artificial Intelligence in Healthcare* (hlm. 25–60). Elsevier. <https://doi.org/10.1016/B978-0-12-818438-7.00002-2>
- Brauers, G., & Frasinicar, F. (2023). A General Survey on Attention Mechanisms in Deep Learning. *IEEE Transactions on Knowledge and Data Engineering*, 35(4), 3279–3298. <https://doi.org/10.1109/TKDE.2021.3126456>
- Bre, F., Gimenez, J. M., & Fachinotti, V. D. (2018). Prediction of Wind Pressure Coefficients on Building Surfaces Using Artificial Neural Networks. *Energy and Buildings*, 158, 1429–1441. <https://doi.org/10.1016/j.enbuild.2017.11.045>
- Chandra, K., Prasetya, K. A., Saputra, R. D., & Hasani, M. F. (2024). Leveraging IndoBert for CyberBullying Classification on Social Media. *ICSINTESA 2024 - 2024 4th International Conference of Science and Information Technology in Smart Administration: The Collaboration of Smart Technology and Good Governance for Sustainable Development Goals*, 407–411. <https://doi.org/10.1109/ICSINTESA62455.2024.10747874>
- Chen, B., Deng, W., & Du, J. (2017). Noisy Softmax: Improving the Generalization Ability of DCNN Via Postponing the Early Softmax Saturation. *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, 2017-January*, 4021–4030. <https://doi.org/10.1109/CVPR.2017.428>
- Chi, S., Qiu, X., Xu, Y., Huang, X., & Sun, C. (2019). *How to Fine-Tune BERT for Text Classification?* <https://doi.org/10.48550/arXiv.1905.05583>
- Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)*, 4171–4186. <http://arxiv.org/abs/1810.04805> (Diakses pada 16 November 2025)

- Dubey, S. R., Singh, S. K., & Chaudhuri, B. B. (2022). Activation Functions in Deep Learning: A Comprehensive Survey and Benchmark. *Neurocomputing (Elsevier)*. <http://arxiv.org/abs/2109.14545> (Diakses pada 19 November 2025)
- Efianingrum, A., Irene Astuti Dwiningrum, S., & Nurhayati, R. (2020). Cyberbullying Pelajar SMA di Media Sosial: Prevalensi dan Rekomendasi. *Jurnal Pembangunan Pendidikan: Fondasi dan Aplikasi*, 8(2), 144–153. <https://journal.uny.ac.id/index.php/jppfa> (Diakses pada 21 November 2025)
- Fan, A., Grave, E., & Joulin, A. (2020). Reducing Transformer Depth on Demand With Structured Dropout. In *International Conference on Learning Representations (ICLR 2020)*.
- Fazry, L., & Apsari, N. (2021). Pengaruh Media Sosial terhadap Perilaku Cyberbullying di Kalangan Remaja. *Jurnal Pengabdian dan Penelitian Kepada Masyarakat (JPPM)*, 2(2), 272–278.
- Ganesh, P., Chen, Y., Lou, X., Khan, M. A., Yang, Y., Sajjad, H., Nakov, P., Chen, D., & Winslett, M. (2021). Compressing Large-Scale Transformer-Based Models: A Case Study on BERT. *Transactions of the Association for Computational Linguistics (TACL)*, 9, 1061–1080. <https://doi.org/10.1162/tacl>
- Garg, S. (2019). *A Study on the Structure of Neural Networks and the Mathematics behind Backpropagation*. <https://www.researchgate.net/publication/342412635> (Diakses pada 18 November 2025)
- Garrido-Merchan, E. C., Gozalo-Brizuela, R., & Gonzalez-Carvajal, S. (2023). Comparing BERT Against Traditional Machine Learning Models in Text Classification. *Journal of Computational and Cognitive Engineering*, 2(4), 352–356. <https://doi.org/10.47852/bonviewJCCE3202838>
- Gong, L., He, D., Li, Z., Qin, T., Wang, L., & Liu, T.-Y. (2019). Efficient Training of BERT by Progressively Stacking. In *Proceedings of the 36th International Conference on Machine Learning (ICML 2019)*.
- Gou, J., Yu, B., Maybank, S. J., & Tao, D. (2021). *Knowledge Distillation: A Survey*. <https://doi.org/10.1007/s11263-021-01453-z>
- Hadiprakoso, R. B., Setiawan, H., Yasa, R. N., & Girinoto. (2021). Text Preprocessing for Optimal Accuracy in Indonesian Sentiment Analysis Using a Deep Learning Model with Word Embedding. *AIP Conference Proceedings*. <https://www.researchgate.net/publication/365799401> (Diakses pada 16 November 2025)
- Hanni, C. T. (2021). *Cyberbullying Bahasa Indonesia*. <https://www.kaggle.com/datasets/cttrhnn/cyberbullying-bahasa-indonesia?select=stopwordsID.csv> (Diakses pada 10 April 2025)

- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *Springer Series in Statistics The Elements of Statistical Learning Data Mining, Inference, and Prediction* (Second Edition). Springer.
- He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 770–778. <https://doi.org/10.1109/CVPR.2016.90>
- Heaton, J. (2018). Ian Goodfellow, Yoshua Bengio, and Aaron Courville: Deep learning. *Genetic Programming and Evolvable Machines*, 19(1–2), 305–307. <https://doi.org/10.1007/s10710-017-9314-z>
- Hinton, G., Vinyals, O., & Dean, J. (2015). *Distilling the Knowledge in a Neural Network*. <http://arxiv.org/abs/1503.02531> (Diakses pada 16 November 2025)
- Huang, J., Ding, R., Zheng, Y., Wu, X., Chen, S., & Jin, X. (2024). Does Part of Speech Have an Influence on Cyberbullying Detection? *Analytics*, 3(1), 1–13. <https://doi.org/10.3390/analytics3010001>
- Hussain, S. A. (2023). Exploring the Effectiveness of Transfer Learning for Text Classification in Low-Resource Languages. *International Journal Of Novel Research And Development*, 8(5).
- Hutabarat, R. K. (2023). Interaksi Sosial di Era Digital: Dampak Media Sosial Terhadap Perubahan Budaya. *Pragmatik : Jurnal Rumpun Ilmu Bahasa dan Pendidikan*, 2(1), 106–110. <https://doi.org/10.61132/pragmatik.v2i1.167>
- Ijomah, W. L., Nwankpa, C. E., Marshall, S., Ijomah, W. L., & Gachagan, A. (2020). Activation Functions: Comparison of trends in Practice and Research for Deep Learning. In *Proceedings of the 2nd International Conference on Computational Sciences and Technologies (INCCST 2020)*, 124–135. <https://doi.org/10.48550/arXiv.1811.03378>
- Ingole, K., & Patil, N. (2020). Performance Analysis of Various Activation Function on a Shallow Neural Network. *Journal of Emerging Technologies and Innovative Research (JETIR)*, 7.
- Jain, A. K., Murty, M. N., & Flynn, P. J. (1999). Data Clustering: A Review. *ACM Computing Surveys*, 31(3), 264–323. <https://doi.org/https://doi.org/10.1145/331499.331504>
- Janocha, K., & Czarnecki, W. M. (2017). On Loss Functions for Deep Neural Networks in Classification. *Computer Vision and Image Understanding (CVIU)*, 150, 66–87. <http://arxiv.org/abs/1702.05659> (Diakses pada 19 November 2025)
- Jawahar, G., Sagot, B., & Djame, S. (2019). What does BERT learn about the structure of language? In *Proceedings of the 57th Annual Meeting of the Association for*

- Computational Linguistics (ACL 2019)*, 3651–3657. <https://aclanthology.org/P19-1356/> (Diakses pada 18 November 2025)
- Jiao, X., Yin, Y., Shang, L., Jiang, X., Chen, X., Li, L., Wang, F., & Liu, Q. (2020). *TinyBERT: Distilling BERT for Natural Language Understanding*. <http://arxiv.org/abs/1909.10351> (Diakses pada 20 November 2025)
- Kalogirou, S. A. (2001). Artificial Neural Networks in Renewable Energy Systems Applications: A Review. *Renewable and Sustainable Energy Reviews* (Vol. 5).
- Kalogirou, S. A., Mathioulakis, E., & Belessiotis, V. (2014). Artificial Neural Networks for the Performance Prediction of Large Solar Systems. *Renewable Energy*, 63, 90–97. <https://doi.org/10.1016/j.renene.2013.08.049>
- Kaplan, A. M., & Haenlein, M. (2010). Users of the World, Unite! The Challenges and Opportunities of Social Media. *Business Horizons*, 53(1), 59–68. <https://doi.org/10.1016/j.bushor.2009.09.003>
- Kemp, S. (2024). *Digital 2024 : Indonesia*. <https://datareportal.com/reports/digital-2024-indonesia> (Diakses pada 10 November 2025)
- Kingma, D. P., & Ba, J. (2015). Adam: A Method for Stochastic Optimization. *International Conference on Learning Representations (ICLR)* . <http://arxiv.org/abs/1412.6980> (Diakses pada 17 November 2025)
- Kobayashi, S., Akram, Y., & Von Oswald, J. (2024). Weight Decay Induces Low-Rank Attention Layers. In *Proceedings of the 38th Conference on Neural Information Processing Systems (NeurIPS 2024)*. <http://arxiv.org/abs/2410.23819> (Diakses pada 19 November 2025)
- Koto, F., Rahimi, A., Lau, J. H., & Baldwin, T. (2020). IndoLEM and IndoBERT: A Benchmark Dataset and Pre-trained Language Model for Indonesian NLP. In *Proceedings of the 28th International Conference on Computational Linguistics (COLING 2020)*, 757–770. <https://doi.org/10.18653/v1/2020.coling-main.66>
- Kotsiantis, S. B. (2007). Supervised Machine Learning: A Review of Classification Techniques. *Informatica*, 31, 249–268.
- Kovaleva, O., Romanov, A., Rogers, A., & Rumshisky, A. (2019). Revealing the Dark Secrets of BERT. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing (EMNLP 2019) and the 9th International Joint Conference on Natural Language Processing (IJCNLP 2019)*, 4365–4374.
- Kuss, D. J., & Griffiths, M. D. (2017). Social Networking Sites and Addiction: Ten Lessons Learned. *International Journal of Environmental Research and Public Health*, 14(3). <https://doi.org/10.3390/ijerph14030311>

- Kusuma, E. J., Lavindi, E. E., Jaya Kusuma, E., Tyas Pernanda, S., & Apriantoro, R. (2023). Cyber-bullying Detection based on Machine Learning Method (Case Study: Instagram Comment Section). *Journal of Applied Information and Communication Technologies*, 8(1). <https://doi.org/10.32497/jaict.v8i1>
- Lecun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436–444. <https://doi.org/10.1038/nature14539>
- Lei, T., Bai, J., Brahma, S., Ainslie, J., Lee, K., Zhou, Y., Du, N., Zhao, V. Y., Wu, Y., Li, B., Zhang, Y., & Google, M.-W. C. (2023). Conditional Adapters: Parameter-efficient Transfer Learning with Fast Inference. In *Proceedings of the 37th Conference on Neural Information Processing Systems (NeurIPS 2023)*.
- Li, C., Liu, K., & Liu, S. (2025). A Survey of Loss Functions in Deep Learning. *Mathematics*, 13(15). <https://doi.org/10.3390/math13152417>
- Li, X., Lu, R., Liu, P., & Zhu, Z. (2022). Graph Convolutional Networks With Hierarchical Multi-Head Attention for Aspect-Level Sentiment Classification. *Journal of Supercomputing*, 78(13), 14846–14865. <https://doi.org/10.1007/s11227-022-04480-w>
- Lin, T., Wang, Y., Liu, X., & Qiu, X. (2022). A Survey of Transformers. *AI Open*, 3, 111–132. <https://doi.org/10.1016/j.aiopen.2022.10.001>
- Lo, K. A., Briant Joe, C., Philip, S., & Hidayaturrehman. (2024). Comparison of Machine Learning and Deep Learning Models for Detecting Cyberbullying. *2024 International Visualization, Informatics and Technology Conference, IVIT 2024*, 138–144. <https://doi.org/10.1109/IVIT62102.2024.10692892>
- Loshchilov, I., & Hutter, F. (2019). Decoupled Weight Decay Regularization. *International Conference on Learning Representations (ICLR)*. <http://arxiv.org/abs/1711.05101> (Diakses pada 19 November 2025)
- Luqyana, W. A., Cholissodin, I., & Perdana, R. S. (2018). Analisis Sentimen Cyberbullying pada Komentar Instagram dengan Metode Klasifikasi Support Vector Machine. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 2(11), 4704–4713.
- Mahdavinejad, M. S., Rezvan, M., Barekatin, M., Adibi, P., Barnaghi, P., & Sheth, A. P. (2018). Machine Learning for Internet of Things Data Analysis: a Survey. *Digital Communications and Networks*, 4(3), 161–175. <https://doi.org/10.1016/j.dcan.2017.10.002>
- Markchom, T., & Liang, H. (2021). UoR at SemEval-2021 Task 4: Using Pre-trained BERT Token Embeddings for Question Answering of Abstract Meaning. *15th International Workshop on Semantic Evaluation (SemEval-2021)*, 799–804. <https://aclanthology.org/2021.semeval-1.106/>

- Marlef, A., & Muda, Y. (2024). Mengenal dan Mencegah Cyberbullying: Tantangan Dunia Digital. *Journal of Education Research*, 5(3).
- Michel, P., Levy, O., & Neubig, G. (2019). Are Sixteen Heads Really Better than One? *Conference on Neural Information Processing Systems (NeurIPS)*. <http://arxiv.org/abs/1905.10650> (Diakses pada 21 November 2025)
- Mingote, V., Miguel, A., Ortega, A., & Lleida, E. (2023). Class Token and Knowledge Distillation for Multi-Head Self-Attention Speaker Verification Systems. *Digital Signal Processing: A Review Journal*, 133. <https://doi.org/10.1016/j.dsp.2022.103859>
- Montavon, G., Samek, W., & Müller, K. R. (2018). Methods for Interpreting and Understanding Deep Neural Networks. *Digital Signal Processing: A Review Journal*, 73, 1–15. <https://doi.org/10.1016/j.dsp.2017.10.011>
- Muraina, I. O. (2022). Ideal Dataset Splitting Ratios in Machine Learning Algorithms: General Concerns for Data Scientists and Data Analysts. *7th International Mardin Artuklu Scientific Researches Conference*, 496–504.
- Ningrum, F. S., & Amna, Z. (2020). Cyberbullying Victimization dan Kesehatan Mental pada Remaja. *INSAN Jurnal Psikologi dan Kesehatan Mental*, 5(1), 35. <https://doi.org/10.20473/jpkm.v5i12020.35-48>
- Novandian, Y. D., Luthfiarta, A., Assyifa, D. S., Setiawan, J., Cahyaningrum, L., Althoff, N., Rahayu, M., Nugraha, A., & Rismiyati. (2024). IndoBERT-based Indonesian Cyberbullying Detection with Multi-stage Labeling. In *Proceedings - 2024 International of Seminar on Application for Technology of Information and Communication: Smart And Emerging Technology for a Better Life, iSemantic 2024*, 515–521. <https://doi.org/10.1109/iSemantic63362.2024.10762553>
- Othman, R. Sh., & Ibrahim, I. M. (2025). Deep Learning for Natural Language Processing: A Review of Models and Applications. *International Journal of Scientific World*, 11(2), 71–78. <https://doi.org/10.14419/t1xnaq87>
- Putri, F. K., Manalu, S. R., & Gono, J. N. S. (2024). Pola Konsumsi Informasi Melalui Media di Kalangan Generasi Z (Studi terhadap SMAN 4, SMAN 9, SMA Mardisiswa, dan SMA Al-Azhar 14 di Kota Semarang). *Interaksi Online*, 12(4), 279–295. <https://ejournal3.undip.ac.id/index.php/interaksi-online/article/view/47223> (Diakses pada 20 November 2025)
- Putri, M., Edy Sutanto, T., & Inna, S. (2023). Studi Empiris Model BERT dan DistilBERT: Analisis Sentimen pada Pemilihan Presiden Indonesia. *Indonesian Journal of Computer Science Attribution*.
- Rahma, I. A., & Suadaa, L. H. (2023). Penerapan Text Augmentation untuk Mengatasi Data yang Tidak Seimbang pada Klasifikasi Teks Berbahasa Indonesia. *Jurnal Teknologi*

- Informasi dan Ilmu Komputer*, 10(6), 1329–1340.  
<https://doi.org/10.25126/jtiik.2023107325>
- Ramadhan, A. D. (2024). *CyberBullyingDetector*.  
<https://huggingface.co/aditdwi123/CyberBullyingDetector> (Diakses pada 5 Juni 2025)
- Reyad, M., Sarhan, A. M., & Arafa, M. (2023). A Modified Adam Algorithm for Deep Neural Network Optimization. *Neural Computing and Applications*, 35(23), 17095–17112. <https://doi.org/10.1007/s00521-023-08568-z>
- Riyaddulloh, R., & Romadhony, A. (2021). Normalisasi Teks Bahasa Indonesia Berbasis Kamus Slang Studi Kasus: Tweet Produk Gadget Pada Twitter. *e-Proceeding of Engineering*, 8(4), 4216.
- Rogers, A., Kovaleva, O., & Rumshisky, A. (2020). A Primer in BERTology: What We Know About How BERT Works. *Transactions of the Association for Computational Linguistics*, 8, 842–866. [https://doi.org/10.1162/tacl\\_a\\_00349](https://doi.org/10.1162/tacl_a_00349)
- Ruder, S. (2017). *An overview of gradient descent optimization algorithms*.  
<http://arxiv.org/abs/1609.04747> (Diakses pada 22 November 2025)
- Sanh, V., Debut, L., Chaumond, J., & Wolf, T. (2020). *DistilBERT, a Distilled Version of BERT: Smaller, Faster, Cheaper, and Lighter*. <http://arxiv.org/abs/1910.01108> (Diakses pada 22 November 2025)
- Santos, F. de N., D'Antuono, P., Robbelein, K., Noppe, N., Weijtjens, W., & Devriendt, C. (2023). Long-Term Fatigue Estimation on Offshore Wind Turbines Interface Loads Through Loss Function Physics-Guided Learning of Neural Networks. *Renewable Energy*, 205, 461–474. <https://doi.org/10.1016/j.renene.2023.01.093>
- Schmidhuber, J. (2015). Deep Learning in Neural Networks: An Overview. *Neural Networks*, 61, 85–117. <https://doi.org/10.1016/j.neunet.2014.09.003>
- Schwartz, R., Dodge, J., Smith, N. A., & Etzioni, O. (2020). Green AI. *Communications of the ACM*, 63(12), 54–63. <https://doi.org/10.1145/3381831>
- Sebayang, R. R., Purba, E., Damanik, S. P., & Surip, M. (2024). Dinamika Bahasa Gaul dan Serapan Asing di Era Digital: Dampaknya terhadap Kemampuan Berbahasa Indonesia Baku. *BAHTRA: Jurnal Pendidikan Bahasa dan Sastra*, 5(2).
- Setiabudi, R., Iswari, N. M. S., & Rusli, A. (2021). Enhancing Text Classification Performance by Preprocessing Misspelled Words in Indonesian Language. *Telkonnika (Telecommunication Computing Electronics and Control)*, 19(4), 1234–1241. <https://doi.org/10.12928/TELKOMNIKA.v19i4.20369>
- Setiawan, Y., Ulva Maulidevi, N., & Surendro, K. (2022). Deteksi Cyberbullying dengan Mesin Pembelajaran Klasifikasi (Supervised Learning): Peluang dan Tantangan. *Jurnal*

- Teknologi Informasi dan Ilmu Komputer (JTIK)*, 9(7), 1577–1582.  
<https://doi.org/10.25126/jtiik.202296747>
- Shabir, A., Ahmed, K. T., Kanwal, K., Almas, A., Raza, S., Fatima, M., & Abbas, T. (2024). A Systematic Review of Attention Models in Natural Language Processing. *Statistics, Computing and Interdisciplinary Research*, 6(1), 33–56.  
<https://doi.org/10.52700/scir.v6i1.157>
- Shrestha, A., & Mahmood, A. (2019). Review of Deep Learning Algorithms and Architectures. *IEEE Access*, 7, 53040–53065.  
<https://doi.org/10.1109/ACCESS.2019.2912200>
- Smelyanskiy, M., Shirish Keskar, N., Mudigere, D., Nocedal, J., Tak, P., & Tang, P. (2017). On Large-Batch Training for Deep Learning: Generalization Gap and Sharp Minima. *International Conference on Learning Representations (ICLR)*.  
<https://doi.org/10.48550/arXiv.1609.04836>
- Smith, L. N. (2018). *A Disciplined Approach to Neural Network Hyper-Parameters: Part 1 — Learning Rate, Batch Size, Momentum, and Weight Decay*.  
<http://arxiv.org/abs/1803.09820> (Diakses pada 23 November 2025)
- Sokolova, M., & Lapalme, G. (2009). A Systematic Analysis of Performance Measures for Classification Tasks. *Information Processing and Management*, 45(4), 427–437.  
<https://doi.org/10.1016/j.ipm.2009.03.002>
- Srivastava, N., Hinton, G., Krizhevsky, A., & Salakhutdinov, R. (2014). Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *Journal of Machine Learning Research*, 15, 1929–1958.
- Sutton, R. S., & Barto, A. G. (2018). *Reinforcement Learning: An Introduction* (2nd edition). MIT Press.
- Szandała, T. (2020). *Review and Comparison of Commonly Used Activation Functions for Deep Neural Networks*. <https://doi.org/10.48550/arXiv.2010.09458>
- Tharwat, A. (2018). Classification Assessment Methods. *Applied Computing and Informatics*, 17(1), 168–192. <https://doi.org/10.1016/j.aci.2018.08.003>
- Tian, Y., & Zhang, Y. (2022). A Comprehensive Survey on Regularization Strategies in Machine Learning. *Information Fusion*, 80, 146–166.  
<https://doi.org/10.1016/j.inffus.2021.11.005>
- UNICEF Indonesia. (2020). Bullying in Indonesia: Key facts, solutions, and recommendations [Fact sheet]. <https://www.unicef.org/indonesia/media/5606/file/Bullying.in.Indonesia.pdf> (Diakses pada 10 November 2025)

- Vaswani, A., Brain, G., 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 (NIPS 2017)*, 5998–6008.
- Voita, E., Talbot, D., Moiseev, F., Sennrich, R., & Titov, I. (2019). Analyzing Multi-Head Self-Attention: Specialized Heads Do the Heavy Lifting, the Rest Can Be Pruned. *In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (ACL 2019)*, 5797–5808. <https://aclanthology.org/P19-1580/>
- Wilie, B., Vincentio, K., Indra Winata, G., Cahyawijaya, S., Li, X., Lim, Z. Y., Soleman, S., Mahendra, R., Fung, P., Bahar, S., Purwarianti, A., & Bandung, I. T. (2020). IndoNLU: Benchmark and Resources for Evaluating Indonesian Natural Language Understanding. *In Proceedings of the 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics & the 10th International Joint Conference on Natural Language Processing (AAACL-IJCNLP 2020)*, 843–857. <https://doi.org/10.18653/v1/2020.aacl-main.85>
- Wirawan, C. (2024). *distilbert-base-indonesian*. <https://huggingface.co/cahya/distilbert-base-indonesian> (Diakses pada 5 Mei 2025)
- Ying, X. (2019). An Overview of Overfitting and its Solutions. *Journal of Physics: Conference Series*, 1168(2). <https://doi.org/10.1088/1742-6596/1168/2/022022>
- Yunior, K. H., Vitianingsih, A. V., Kacung, S., Maukar, A. L., & Arumsari, A. D. (2024). Sentiment Analysis of Cyberbullying Detection on Social Networks using the Sentistrength Method. *Sistemasi: Jurnal Sistem Informasi*, 13(4), 1587–1596.
- Zhang, L., Wang, S., & Liu, B. (2018). mixup: Beyond Empirical Risk Minimization. *In International Conference on Learning Representations (ICLR 2018)*.
- Zhuang, F., Qi, Z., Duan, K., Xi, D., Zhu, Y., Zhu, H., Xiong, H., & He, Q. (2021). A Comprehensive Survey on Transfer Learning. *Proceedings of the IEEE*, 109(1), 43–76. <https://doi.org/10.1109/JPROC.2020.3004555>