

DAFTAR PUSTAKA

- [1] Kompas Gramedia, “Retail & publishing - kompas gramedia,” accessed: 2025-05-03. [Online]. Available: <https://kompasgramedia.com/business/retail-publishing>
- [2] Wikipedia contributors, “Kompas gramedia group,” Wikipedia, The Free Encyclopedia, 2025, accessed: 2025-05-03. [Online]. Available: https://en.wikipedia.org/wiki/Kompas_Gramedia_Group
- [3] Umnadmin, “Big data: Pengertian, contoh, dan fungsinya,” Online; Universitas Multimedia Nusantara, Aug. 2023, accessed: 2025-05-03. [Online]. Available: <https://www.umn.ac.id/big-data-pengertian-contoh-dan-fungsinya/>
- [4] Google Cloud, “Big data defined: Examples and benefits,” Online; Google Cloud, n.d., accessed: 2025-05-03. [Online]. Available: <https://cloud.google.com/learn/what-is-big-data>
- [5] Insight Editor, “Unpacking gen ai: Demystifying the current artificial intelligence landscape,” Online; Insight, Jan. 2024, accessed: 2025-05-03. [Online]. Available: https://www.insight.com/en_US/content-and-resources/blog/unpacking-gen-ai--demystifying-the-current-artificial-intelligence-landscape.html
- [6] e. a. Duong, “Customer gender prediction based on e-commerce data,” in *ResearchGate*, 2015. [Online]. Available: https://www.researchgate.net/publication/311316135_Customer_gender_prediction_based_on_E-commerce_data
- [7] M. M. Islam and J.-H. Baek, “Deep learning based real age and gender estimation from unconstrained face image towards smart store customer relationship management,” *Applied Sciences*, vol. 11, no. 10, p. 4549, 2021.
- [8] P. Kotler, H. Kartajaya, and I. Setiawan, *Marketing 5.0: Technology for Humanity*. Hoboken, NJ: John Wiley & Sons, 2021.
- [9] TechTarget, “What are the 5 v’s of big data?” accessed: 2025-05-03. [Online]. Available: <https://www.techtarget.com/searchdatamanagement/definition/5-Vs-of-big-data>
- [10] IBM, “What is big data analytics?” accessed: 2025-05-03. [Online]. Available: <https://www.ibm.com/think/topics/big-data-analytics>
- [11] Express Analytics, “The progress of customer profiling using machine learning,” *Express Analytics Blog*, 2024. [Online]. Available: <https://www.expressanalytics.com/blog/customer-profiling-with-machine-learning/>

- [12] Sonora ID, "Sejarah lahirnya Kompas Gramedia hingga dikenal seperti sekarang," *Sonora ID*, 2020. [Online]. Available: <https://bangka.sonora.id/read/502329664/sejarah-lahirnya-kompas-gramedia-hingga-dikenal-seperti-sekarang?page=all>
- [13] Kompas Gramedia. (2025) Kg media. <https://www.kompasgramedia.com/business/media>. Accessed: May 11, 2025.
- [14] —. (2025) Kg hospitality. <https://www.kompasgramedia.com/business/hospitality>. Accessed: May 11, 2025.
- [15] —. (2025) Kg education. <https://www.kompasgramedia.com/business/education>. Accessed: May 11, 2025.
- [16] —. (2025) Kg retail & publishing. <https://www.kompasgramedia.com/business/retail-publishing>. Accessed: May 11, 2025.
- [17] —, "Kg manufacture," <https://www.kompasgramedia.com/business/manufacture>, 2025, accessed: May 11, 2025.
- [18] —, "Kg event & venue," <https://www.kompasgramedia.com/business/event-venue>, 2025, accessed: May 11, 2025.
- [19] —, "Kg property," <https://www.kompasgramedia.com/business/property>, 2025, accessed: May 11, 2025.
- [20] —, "Kg digital," <https://www.kompasgramedia.com/business/digital>, 2025, accessed: May 11, 2025.
- [21] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," *arXiv preprint arXiv:1810.04805*, 2018. [Online]. Available: <https://doi.org/10.48550/arxiv.1810.04805>
- [22] S. Metzger, "A beginner's guide to tokens, vectors, and embeddings in nlp," December 2022, medium. [Online]. Available: <https://medium.com/@saschametzger/what-are-tokens-vectors-and-embeddings-how-do-you-create-them-e2a3e698e037>
- [23] B. Wilie, K. Vincentio, G. I. Winata, S. Cahyawijaya, X. Li, Z. Y. Lim, S. Soleman, R. Mahendra, P. Fung, S. Bahar, and A. Purwarianti, "Indonlu: Benchmark and resources for evaluating Indonesian natural language understanding," *arXiv preprint arXiv:2009.05387*, 2020. [Online]. Available: <https://doi.org/10.48550/arxiv.2009.05387>
- [24] F. Koto, A. Rahimi, J. H. Lau, and T. Baldwin, "Indolem and indobert: a benchmark dataset and pre-trained language model for Indonesian NLP," *arXiv preprint arXiv:2011.00677*, 2020. [Online]. Available: <https://doi.org/10.48550/arxiv.2011.00677>

- [25] H. Q. To, K. V. Nguyen, N. L.-T. Nguyen, and A. G.-T. Nguyen, "Gender prediction based on vietnamese names with machine learning techniques," in *Proceedings of the 4th International Conference on Natural Language Processing and Information Retrieval*, 2020.
- [26] I. Novogroder. (2024) Data preprocessing in machine learning: Steps & best practices. Accessed: 2025-06-02. [Online]. Available: <https://lakefs.io/blog/data-preprocessing-in-machine-learning/>
- [27] M. Narang. (2024, Feb.) Top 4 techniques for handling missing values in machine learning. Paperspace by DigitalOcean Blog. [Online]. Available: <https://blog.paperspace.com/top-4-techniques-for-handling-the-missing-values-in-machine-learning/>
- [28] Hugging Face, "Tokenizer," 2025, accessed: 2025-06-02. [Online]. Available: https://huggingface.co/docs/transformers/main_classes/tokenizer
- [29] M. De Silva. (2023) Preprocessing steps for natural language processing (nlp): A beginner's guide. Accessed: 2025-06-02. [Online]. Available: <https://medium.com/@maleeshadesilva21/preprocessing-steps-for-natural-language-processing-nlp-a-beginners-guide-d6d9bf7689c9>
- [30] The pandas development team, *pandas: powerful Python data analysis toolkit*, pandas, 2024, <https://pandas.pydata.org/docs/>.
- [31] Python Software Foundation, *re — Regular expression operations*, 2025, <https://docs.python.org/3/library/re.html>.
- [32] Hugging Face, "Models," 2025, accessed: 2025-06-02. [Online]. Available: https://huggingface.co/docs/transformers/main/en/main_classes/model
- [33] S. Kumar. (2021) Data splitting technique to fit any machine learning model. Accessed: 2025-06-02. [Online]. Available: <https://medium.com/data-science/data-splitting-technique-to-fit-any-machine-learning-model-c0d7f3f1c790>
- [34] A. S. Gillis. (2024) Data splitting. Accessed: 2025-06-02. [Online]. Available: <https://www.techtarget.com/searchenterpriseai/definition/data-splitting#:~:text=In%20machine%20learning%2C%20data%20splitting,into%20three%20or%20four%20sets.>
- [35] GeeksforGeeks, "__init__ in python," December 2024. [Online]. Available: https://www.geeksforgeeks.org/__init__-in-python/
- [36] PyTorch, "Datasets & dataloaders — pytorch tutorials 2.7.0+cu126 documentation," 2025, accessed: 2025-06-09. [Online]. Available: https://pytorch.org/tutorials/beginner/basics/data_tutorial.html

- [37] Hugging Face, “Trainer,” 2025, accessed: 2025-05-20. [Online]. Available: https://huggingface.co/docs/transformers/en/main_classes/trainer
- [38] PyTorch, “torch.utils.data — pytorch 2.7 documentation,” 2025, accessed: 2025-05-20. [Online]. Available: <https://pytorch.org/docs/stable/data.html>
- [39] Fatmanurkutlu. (2024, Jan.) Model evaluation techniques in machine learning. Medium. [Online]. Available: <https://medium.com/@fatmanurkutlu1/model-evaluation-techniques-in-machine-learning-8cd88deb8655>
- [40] scikit-learn, “classification_report — scikit-learn documentation,” 2025, accessed: 2025-03-12. [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.metrics.classification_report.html
- [41] GeeksforGeeks. (2024, June) Essential metrics for model assessment: Tp, tn, fp, fn in machine learning. Accessed: 2025-06-04. [Online]. Available: <https://www.geeksforgeeks.org/essential-metrics-for-model-assessment-tp-tn-fp-fn-in-machine-learning/>
- [42] Google, “Classification: Accuracy, recall, precision, and related metrics,” <https://developers.google.com/machine-learning/crash-course/classification/accuracy-precision-recall>, n.d., google for Developers.
- [43] E. Software. (2024, November) Understanding training, validation, and testing data in ml. Medium. [Online]. Available: <https://medium.com/@eastgate/understanding-training-validation-and-testing-data-in-ml-c297e53d6096>
- [44] Ibm, “Overfitting,” <https://www.ibm.com/think/topics/overfitting>, Apr. 17 2025.
- [45] PyTorch, “torch.argmax — pytorch 2.7 documentation,” 2025, accessed: 2025-05-24. [Online]. Available: <https://docs.pytorch.org/docs/stable/generated/torch.argmax.html>
- [46] GeeksforGeeks. (2024, Jul.) Why deep learning need gpu. GeeksforGeeks. Accessed: 2025-06-08. [Online]. Available: <https://www.geeksforgeeks.org/why-deep-learning-need-gpu/>
- [47] NVIDIA Developer. (n.d.) Deep learning. NVIDIA. Accessed: 2025-06-08. [Online]. Available: <https://developer.nvidia.com/deep-learning>
- [48] A. Burkov, *The Hundred-Page Machine Learning Book*. Andriy Burkov, 2019.
- [49] G. James, D. Witten, T. Hastie, R. Tibshirani, and J. Taylor, *An Introduction to Statistical Learning: With Applications in Python*. Springer, 2023.