

DAFTAR PUSTAKA

- Aggarwal, C.C., 2018. *Machine Learning for Text*. [online] Cham: Springer International Publishing. <https://doi.org/10.1007/978-3-319-73531-3>.
- Arras, L., Horn, F., Montavon, G., Müller, K.-R. and Samek, W., 2017. “What is relevant in a text document?”: An interpretable machine learning approach. *PLOS ONE*, 12(8), p.e0181142. <https://doi.org/10.1371/journal.pone.0181142>.
- Arrieta, A.B., Díaz-Rodríguez, N., Ser, J.D., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R. and Herrera, F., 2020. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible *AI*. *Information Fusion*, 58, pp.82–115. <https://doi.org/10.1016/j.inffus.2019.12.012>.
- Brown, A., 2017. What is hate speech? Part 1: The Myth of Hate. *Law and Philosophy*, 36(4), pp.419–468. <https://doi.org/10.1007/s10982-017-9297-1>.
- Buhrmester, V., Münch, D. and Arens, M., 2019. Analysis of Explainers of Black Box Deep Neural Networks for Computer Vision: A Survey. <https://doi.org/10.48550/arXiv.1911.12116>.
- Camacho-Collados, J. and Pilehvar, M.T., 2018. On the Role of Text Preprocessing in Neural Network Architectures: An Evaluation Study on Text Categorization and Sentiment Analysis. In: T. Linzen, G. Chrupala and A. Alishahi, eds. *Proceedings of the 2018 EMNLP Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP*. [online] Brussels, Belgium: Association for Computational Linguistics. pp.40–46. <https://doi.org/10.18653/v1/W18-5406>.
- Caselli, T., Basile, V., Mitrović, J. and Granitzer, M., 2021. HateBERT: Retraining BERT for Abusive Language Detection in English. <https://doi.org/10.48550/arXiv.2010.12472>.
- Davidson, T., Warmusley, D., Macy, M. and Weber, I., 2017. Automated Hate Speech Detection and the Problem of Offensive Language. *Proceedings of the International AAAI Conference on Web and Social Media*, 11(1), pp.512–515. <https://doi.org/10.1609/icwsm.v11i1.14955>.
- Devlin, J., Chang, M.-W., Lee, K. and Toutanova, K., 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In: J. Burstein, C. Doran and T. Solorio, eds. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*. [online] Minneapolis, Minnesota: Association for Computational Linguistics. pp.4171–4186. <https://doi.org/10.18653/v1/N19-1423>.
- Feldman, R. and Sanger, J., 2006. *The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data*. Cambridge University Press.

- Fortuna, P. and Nunes, S., 2018. A Survey on Automatic Detection of Hate Speech in Text. *ACM Comput. Surv.*, [online] 51(4). <https://doi.org/10.1145/3232676>.
- Hare, I. and Weinstein, J. eds., 2009. *Extreme Speech and Democracy*. [online] Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199548781.001.0001>.
- Hugging Face, 2023. Tokenizer. [online] Available at: https://huggingface.co/docs/transformers/en/main_classes/tokenizer [Accessed 26 February 2025].
- Johannemann, J., Hadad, V., Athey, S. and Wager, S., 2021. Sufficient Representations for Categorical Variables. Available at: <https://arxiv.org/abs/1908.09874>.
- Joseph, V.R., 2022. Optimal ratio for data splitting. *Statistical Analysis and Data Mining: The ASA Data Science Journal*, 15(4), pp.531–538. <https://doi.org/10.1002/sam.11583>.
- Jurafsky, D. and Martin, J., 2008. *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*.
- Kepolisian Negara Republik Indonesia, 2015. Surat Edaran No. SE/6/X/2015 tentang Ujaran Kebencian. [online] *Jakarta: Kepolisian Negara Republik Indonesia*. Available at: <https://pro.hukumonline.com/a/lt564076166fdd6/penanganan-ujaran-kebencian-hate-speech>.
- Kibriya, H., Siddiqa, A., Khan, W.Z. and Khan, M.K., 2024. Towards safer online communities: Deep learning and explainable AI for hate speech detection and classification. *Computers and Electrical Engineering*, 116, p.109153. <https://doi.org/10.1016/j.compeleceng.2024.109153>.
- Li, J., 2024. Causes of Action Identification in Disputes over Construction Project Contracts Using Hierarchical Learning Based on BERT. *Applied Mathematics and Nonlinear Sciences*, [online] 9(1). <https://doi.org/10.2478/amns-2024-2254>.
- Liu, B., Hu, M. and Cheng, J., 2005. Opinion observer: analyzing and comparing opinions on the Web. In: *Proceedings of the 14th International Conference on World Wide Web, WWW '05*. [online] New York, NY, USA: Association for Computing Machinery. pp.342–351. <https://doi.org/10.1145/1060745.1060797>.
- Lundberg, S.M. and Lee, S.-I., 2017. A Unified Approach to Interpreting Model Predictions. In: I. Guyon, U.V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan and R. Garnett, eds. *Advances in Neural Information Processing Systems*. [online] Curran Associates, Inc. Available at: https://proceedings.neurips.cc/paper_files/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf.
- Manning, C.D. and Schütze, H., 1999. *Foundations of Statistical Natural Language Processing*. [online] Cambridge, Massachusetts: The MIT Press. Available at: <http://nlp.stanford.edu/fsnlp/>.

- Mnassri, K., Rajapaksha, P., Farahbakhsh, R. and Crespi, N., 2022. BERT-based Ensemble Approaches for Hate Speech Detection. In: *GLOBECOM 2022 - 2022 IEEE Global Communications Conference*. pp.4649–4654. <https://doi.org/10.1109/GLOBECOM48099.2022.10001325>.
- Molnar, C., 2022. *Interpretable Machine Learning: A Guide for Making Black Box Models Explainable*. 2nd ed. [online] Available at: <<https://christophm.github.io/interpretable-ml-book>>.
- Mudde, C. and Rovira Kaltwasser, C., 2017. *Populism: A Very Short Introduction*. [online] Oxford University Press. <https://doi.org/10.1093/actrade/9780190234874.001.0001>.
- Müller, A.C. and Guido, S., 2016. *Introduction to Machine Learning with Python: A Guide for Data Scientists*. O'Reilly Media, Inc.
- Muraina, I., 2022. IDEAL DATASET SPLITTING RATIOS IN MACHINE LEARNING ALGORITHMS: GENERAL CONCERNS FOR DATA SCIENTISTS AND DATA ANALYSTS.
- Murel, J. and Kavlakoglu, E., 2024. What is transfer learning? | IBM. [online] Available at: <<https://www.ibm.com/topics/transfer-learning>> [Accessed 3 December 2024].
- Mutasodirin, M.A. and Prasojo, R.E., 2021. Investigating Text Shortening Strategy in BERT: Truncation vs Summarization. In: *2021 International Conference on Advanced Computer Science and Information Systems (ICACSIS)*. [online] IEEE. pp.1–5. <https://doi.org/10.1109/icacsis53237.2021.9631364>.
- Pan, S.J. and Yang, Q., 2010. A Survey on Transfer Learning. *IEEE Transactions on Knowledge and Data Engineering*, 22(10), pp.1345–1359. <https://doi.org/10.1109/TKDE.2009.191>.
- Powers, D., 2011. Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness & Correlation. *Journal of Machine Learning Technologies*, 2(1), pp.37–63.
- Reiter, E. and Dale, R., 2000. *Building Natural Language Generation Systems*. *Studies in Natural Language Processing*. Cambridge University Press.
- Ribeiro, M.T., Singh, S. and Guestrin, C., 2016. ‘Why Should I Trust You?’: Explaining the Predictions of Any Classifier. Available at: <<https://arxiv.org/abs/1602.04938>>.
- Ribeiro, M.T., Singh, S. and Guestrin, C., 2018. Anchors: High-Precision Model-Agnostic Explanations. *Proceedings of the AAAI Conference on Artificial Intelligence*, [online] 32(1). <https://doi.org/10.1609/aaai.v32i1.11491>.
- Rodríguez, P., Bautista, M.A., González, J. and Escalera, S., 2018. Beyond one-hot encoding: Lower dimensional target embedding. *Image and Vision Computing*, 75, pp.21–31. <https://doi.org/10.1016/j.imavis.2018.04.004>.
- Ruder, S., Peters, M.E., Swayamdipta, S. and Wolf, T., 2019. Transfer Learning in Natural Language Processing. In: A. Sarkar and M. Strube, eds. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Tutorials*. [online] Minneapolis, Minnesota: Association for Computational Linguistics. pp.15–18. <https://doi.org/10.18653/v1/N19-5004>.

- Samek, W., Wiegand, T. and Müller, K.-R., 2017. Explainable Artificial Intelligence: Understanding, Visualizing and Interpreting Deep Learning Models. Available at: <<https://arxiv.org/abs/1708.08296>>.
- Schmidt, A. and Wiegand, M., 2017. A Survey on Hate Speech Detection using Natural Language Processing. In: *Proceedings of the Fifth International Workshop on Natural Language Processing for Social Media*. [online] Proceedings of the Fifth International Workshop on Natural Language Processing for Social Media. Valencia, Spain: Association for Computational Linguistics. pp.1–10. <https://doi.org/10.18653/v1/W17-1101>.
- Schuster, M. and Nakajima, K., 2012. Japanese and Korean voice search. In: *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. pp.5149–5152. <https://doi.org/10.1109/ICASSP.2012.6289079>.
- Song, X., Salcianu, A., Song, Y., Dopson, D. and Zhou, D., 2021. Fast WordPiece Tokenization. Available at: <<https://arxiv.org/abs/2012.15524>>.
- Tjoa, E. and Guan, C., 2021. A Survey on Explainable Artificial Intelligence (XAI): Toward Medical XAI. *IEEE Transactions on Neural Networks and Learning Systems*, 32(11), pp.4793–4813. <https://doi.org/10.1109/tnnls.2020.3027314>.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł. ukasz and Polosukhin, I., 2017. Attention is All you Need. In: I. Guyon, U.V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan and R. Garnett, eds. *Advances in Neural Information Processing Systems*. [online] Curran Associates, Inc. Available at: <https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf>.
- Waldron, J., 2012. The harm in hate speech. Cambridge, MA, US: Harvard University Press. pp.vii, 292. <https://doi.org/10.4159/harvard.9780674065086>.
- Wang, R., Wang, Z., Xu, Z., Wang, C., Li, Q., Zhang, Y. and Li, H., 2021. A Real-Time Object Detector for Autonomous Vehicles Based on YOLOv4. *Computational Intelligence and Neuroscience*, 2021, pp.1–11. <https://doi.org/10.1155/2021/9218137>.
- Wiącek, M., Rybak, P., Pszeny, Ł. and Wróblewska, A., 2024. NLPre: A Revised Approach towards Language-centric Benchmarking of Natural Language Preprocessing Systems. In: N. Calzolari, M.-Y. Kan, V. Hoste, A. Lenci, S. Sakti and N. Xue, eds. *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*. [online] Torino, Italia: ELRA and ICCL. pp.12271–12287. Available at: <<https://aclanthology.org/2024.lrec-main.1073/>>.
- Young, T., Hazarika, D., Poria, S. and Cambria, E., 2018. *Recent Trends in Deep Learning Based Natural Language Processing*. Available at: <<https://arxiv.org/abs/1708.02709>>.

- Zampieri, M., Malmasi, S., Nakov, P., Rosenthal, S., Farra, N. and Kumar, R., 2019. Predicting the Type and Target of Offensive Posts in Social Media. In: J. Burstein, C. Doran and T. Solorio, eds. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*. [online] Minneapolis, Minnesota: Association for Computational Linguistics. pp.1415–1420. <https://doi.org/10.18653/v1/N19-1144>.
- Zhuang, F., Qi, Z., Duan, K., Xi, D., Zhu, Y., Zhu, H., Xiong, H. and He, Q., 2020. A Comprehensive Survey on Transfer Learning. *Proceedings of the IEEE*, PP, pp.1–34. <https://doi.org/10.1109/JPROC.2020.3004555>.