

DAFTAR PUSTAKA

- AJI. 2024. 2024 Indonesian general election hate speech monitoring dashboard. Available at: <https://aji.or.id/> [Accessed 20 September 2024].
- Alexandra, L. A. & Satria, A., 2023. Identifying hate speech trends and prevention in Indonesia: A cross-case comparison. *Global Responsibility to Protect*, 15(2-3), pp.135–176. <https://doi.org/10.1163/1875984X-01502002>.
- Andriyani, A. D., Masyhur, A. M. & Riyadi, S., 2024. Improving hate speech detection accuracy using hybrid CNN-RNN and random oversampling techniques. *2024 IEEE Symposium on Industrial Electronics & Applications (ISIEA)*, pp.1-5. <https://doi.org/10.1109/ISIEA61920.2024.10607232>.
- Beddiar, D. R., Jahan, M. S. & Oussalah, M., 2021. Data expansion using back translation and paraphrasing for hate speech detection. *Online Social Networks and Media*, 24, p.100153. <https://doi.org/10.1016/j.osnem.2021.100153>.
- Chicco, D. & Jurman, G., 2020. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics*, 21, pp.1-13. <https://doi.org/10.1186/s12864-019-6413-7>.
- Citron, D. K. & Franks, M. A., 2014. Criminalizing revenge porn. *Wake Forest Law Review*, 49(2), pp.345-391.
- CSIS. 2022. Hate speech dashboard. Available at: <https://hatespeech.csis.or.id/> [Accessed 20 September 2024].
- Grave, E., Bojanowski, P., Gupta, P., Joulin, A. & Mikolov, T., 2018. Learning word vectors for 157 languages. *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, European Language Resources Association (ELRA), Miyazaki, Japan, pp.1-8. [Online] Available at: <https://aclanthology.org/L18-1550>.
- Hochreiter, S. & Schmidhuber, J., 1997. Long short-term memory. *Neural Computation*, 9(8), pp.1735-1780.

- Ibrohim, M. O. & Budi, I., 2019. Multi-label hate speech and abusive language detection in Indonesian Twitter. *Proceedings of the Third Workshop on Abusive Language Online*, pp.46-57. <https://doi.org/10.18653/v1/W19-3506>.
- Ingolfsson, T. M., 2021. Insights into LSTM architecture. Available at: https://thorirmar.com/post/insight_into_lstm/ [Accessed 27 September 2024].
- Kobayashi, S., 2018. Contextual augmentation: Data augmentation by words with paradigmatic relations. *Proceedings of NAACL-HLT*, pp.452-457.
- Komnas HAM, 2015. *Buku saku penanganan ujaran kebencian (Hate Speech)*. Komisi Nasional Hak Asasi Manusia, Jakarta.
- Kowsari, K., Heidarysafa, M., Brown, D. E., Meimandi, K. J. & Barnes, L. E., 2019. Text classification algorithms: A survey. *Information*, 10(4), p.150. <https://doi.org/10.3390/info10040150>.
- Kusuma, J. F. & Chowanda, A., 2023. Indonesian hate speech detection using IndoBERTweet and BiLSTM on Twitter. *International Journal on Informatics Visualization*.
- Li, Y. & Yang, T., 2017. Word embedding for understanding natural language: A survey. In: *Proceedings of the 8th International Conference on Learning Representations (ICLR)*.
- Manning, C. D., Raghavan, P. & Schütze, H., 2008. *Introduction to information retrieval*. Cambridge University Press.
- Mendez, J. R., Iglesias, E. L., Fdez-Riverola, F., Díaz, F. & Corchado, J. M., 2006. Tokenising, stemming and stopword removal on anti-spam filtering domain. In: R. Marín, E. Onaindía, A. Bugarín & J. Santos, eds. *Current Topics in Artificial Intelligence. Lecture Notes in Computer Science*, vol 4177. Springer, Berlin, Heidelberg. https://doi.org/10.1007/11881216_47.
- Mikhailov, S. & Kashevnik, A., 2021. Car tourist trajectory prediction based on bidirectional LSTM neural network. *Electronics*, 10(1390), pp.1-15. <https://doi.org/10.3390/electronics10121390>.
- Mikolov, T., Le, Q. V. & Sutskever, I., 2013. Exploiting similarities among languages for machine translation. *arXiv preprint arXiv:1309.4168*.

- Nandana Githa, I. P. W., Syananda, A., Faustine, R., Edbert, I. S. & Suhartono, D., 2024. Hate speech classification in Indonesian tweets using TF-IDF and data augmentation. *2024 International Conference on Green Energy, Computing and Sustainable Technology (GECOST)*, pp.61-65. <https://doi.org/10.1109/GECOST60902.2024.10474781>.
- Opitz, J. & Burst, S., 2019. Macro F1 and Macro F1. *arXiv preprint arXiv:1911.03347*.
- Pasupa, K., Karnbanjob, W. & Aksornsiri, M., 2022. Hate speech detection in Thai social media with ordinal-imbalanced text classification. *2022 19th International Joint Conference on Computer Science and Software Engineering (JCSSE)*, pp.1-6. <https://doi.org/10.1109/JCSSE54890.2022.9836312>.
- Powers, D. M. W., 2020. Evaluation: From precision, recall, and F-measure to ROC, informedness, markedness & correlation. *Journal of Machine Learning Technologies*, 2(1), pp.37–63.
- Pustejovsky, J. & Stubbs, A., 2013. *Natural language annotation for machine learning: A guide to corpus-building for applications*. O'Reilly Media, Inc.
- Sanya, A. D. & Suadaa, L. H., 2022. Handling imbalanced dataset on hate speech detection in Indonesian online news comments. *2022 10th International Conference on Information and Communication Technology (ICoICT)*, pp.380-385. <https://doi.org/10.1109/ICoICT55009.2022.9914883>.
- Sarwar, S. M. & Murdock, V., 2021. Unsupervised domain adaptation for hate speech detection using a data augmentation approach. *arXiv preprint arXiv:2107.12866*.
- Schmidt, A. & Wiegand, M., 2017. A survey on hate speech detection using natural language processing. *Proceedings of the Fifth International Workshop on Natural Language Processing for Social Media*, pp.1-10.
- Sokolova, M. & Lapalme, G., 2009. A systematic analysis of performance measures for classification tasks. *Information Processing & Management*, 45(4), pp.427-437.
- Susanto, L., Wijanarko, M. I., Pratama, P. A., Hong, T., Idris, I., Aji, A. F. & Wijaya, D., 2024. IndoToxic2024: A demographically-enriched dataset of

hate speech and toxicity types for Indonesian language. *arXiv preprint arXiv:2406.19349*.

Tala, F. Z., 2003. A study of stemming effects on information retrieval in Bahasa Indonesia (M.Sc. Thesis). Master of Logic Project, Institute for Logic, Language and Computation, Universiteit van Amsterdam, The Netherlands.

Undang-Undang Republik Indonesia No. 11 Tahun 2008 tentang Informasi dan Transaksi Elektronik (UU ITE).

United Nations, 1965. *International Convention on the Elimination of All Forms of Racial Discrimination (ICERD)*.

Unlu, A. & Kotonen, T., 2024. Online polarization and identity politics: An analysis of Facebook discourse on Muslim and LGBTQ+ communities in Finland. *Scandinavian Political Studies*. <https://doi.org/10.1111/1467-9477.12245>.

Wei, J. & Zou, K., 2019. EDA: Easy data augmentation techniques for boosting performance on text classification tasks. *arXiv preprint arXiv:1901.11196*.

Williams, M. L., Burnap, P., Javed, A., Liu, H. & Ozalp, S., 2019. Hate in the machine: Anti-Black and anti-Muslim social media posts as predictors of offline racially and religiously aggravated crime. *The British Journal of Criminology*, 60(1), pp.93-117. <https://doi.org/10.1093/bjc/azy033>.

Wijaya, K. U. & Setiawan, E. B., 2023. Hate speech detection using convolutional neural network and gated recurrent unit with FastText feature expansion on Twitter. *Jurnal Ilmiah Teknik Elektro Komputer dan Informatika (JITEKI)*, 9(3), pp.619-631. <https://doi.org/10.26555/jiteki.v9i3.26532>.

Yanaputeranto, M. R. & Rifai, S., 2021. Hate speech in the digital age: Analysis of government policies and societal responses in Indonesia. *Journal of Asian Social Policy*, 15(2), pp.241-260. <https://doi.org/10.21670/jasp-15147>.

Young, J. & Rusli, A., 2019. Review and visualization of Facebook's FastText pretrained word vector model. *Proceedings of the 2019 International Conference on Electrical, Electronics and System Engineering (ICEESE)*, pp.1-6. <https://doi.org/10.1109/ICESI.2019.8863015>.

Young, T., Hazarika, D., Poria, S. & Cambria, E., 2018. Recent trends in deep learning based natural language processing. *IEEE Computational Intelligence Magazine*, 13(3), pp.55-75. <https://doi.org/10.1109/MCI.2018.2840738>.

Zolotov, V. & Kung, D., 2017. Analysis and optimization of fastText linear text classifier. *arXiv preprint arXiv:1702.05531*.

Zhang, Z., Robinson, D. & Tepper, J., 2018. Detecting hate speech on Twitter using a convolution-GRU based deep neural network. *European Semantic Web Conference*, Springer, pp.745-760.