

---

## BIBLIOGRAPHY

- [1] Mulugeta Deribe Damota. The effect of social media on society. *New Media and Mass Communication*, 2019. URL <https://api.semanticscholar.org/CorpusID:151260332>.
- [2] Titis Anindyajati. Limitation of the right to freedom of speech on the Indonesian constitutional court consideration. *Indonesian Law Journal*, 14(1):19–36, 2021.
- [3] Dhamir Raniah Kiasati Desrul and Ade Romadhony. Abusive language detection on Indonesian online news comments. In *2019 International Seminar on Research of Information Technology and Intelligent Systems (ISRITI)*, pages 320–325, 2019. doi: 10.1109/ISRITI48646.2019.9034620.
- [4] Komnas HAM. *Buku Saku Penanganan Ujaran Kebencian (Hate Speech)*. 2015. URL <https://www.ylbhu.org/wp-content/uploads/2016/06/Panduan-Teoritik-Hate-Speech.pdf>. Accessed: May 22, 2023.
- [5] Yonatan Lupu, Richard Sear, Nicolas Velásquez, Rhys Leahy, Nicholas Johnson Restrepo, Beth Goldberg, and Neil F. Johnson. Offline events and online hate. *PLOS ONE*, 18(1):1–14, 01 2023. doi: 10.1371/journal.pone.0278511. URL <https://doi.org/10.1371/journal.pone.0278511>.
- [6] Magdalena Obermaier and Desirée Schmuck. Youths as targets: factors of online hate speech victimization among adolescents and young adults. *Journal of Computer-Mediated Communication*, 27(4):zmac012, 2022.
- [7] University College London. Calls for verbal abuse of children by adults to be formally recognised as a form of child maltreatment, October 2023. URL <https://www.ucl.ac.uk/news/2023/oct/calls-verbal-abuse-children-adults-be-formally-recognised-form-child-maltreatment>. Accessed: 2024-07-30.
- [8] I. M. Kardiyasa, A. A. S. L. Dewi, and N. M. S. Karma. Sanksi pidana terhadap ujaran kebencian (hate speech). *Jurnal Analogi Hukum*, 2(1):78–82, Mar. 2020. doi: 10.22225/ah.2.1.1627.78-82.
- [9] Indonesia. *Undang-Undang Nomor 19 Tahun 2016 tentang Perubahan atas Undang-Undang Nomor 11 Tahun 2008 Tentang Informasi dan Transaksi Elektronik*. Sekretariat Indonesia, Jakarta, 2016.
- [10] APJII. Indonesian internet profile 2022, 2022. URL <https://apjii.or.id/download/cf790057fdac70557a6655945479b5ab>.

- 
- [11] S.Kemp. Digital 2022: Indonesia., 2022. URL <https://datareportal.com/reports/digital-2022-indonesia>.
- [12] A. International UK. Twitter: Anti-lgbtq+ hate speech surging on platform under elon musk, 2023. URL <https://www.amnesty.org.uk/press-releases/twitter-anti-lgbtq-hate-speech-surging-platform-under-elon-musk>.
- [13] M. State University. Study finds hate speech on twitter increased quickly after elon musk takeover, 2022. URL <https://www.montclair.edu/newscenter/2022/10/31/study-finds-hate-speech-on-twitter-increased-following-elon-musk-takeover/>.
- [14] C. Indonesia. Csis national hate speech dashboard, 2021. URL <https://hatespeech.csis.or.id/>.
- [15] Jan Kocoń, Alicja Figas, Marcin Gruza, Daria Puchalska, Tomasz Kajdanowicz, and Przemysław Kazienko. Offensive, aggressive, and hate speech analysis: From data-centric to human-centered approach. *Information Processing Management*, 58(5): 102643, 2021. ISSN 0306-4573. doi: <https://doi.org/10.1016/j.ipm.2021.102643>. URL <https://www.sciencedirect.com/science/article/pii/S0306457321001333>.
- [16] Fabio Poletto, Valerio Basile, Manuela Sanguinetti, Cristina Bosco, and Viviana Patti. Resources and benchmark corpora for hate speech detection: a systematic review. *Lang. Resour. Eval.*, 55(2):477–523, jun 2021. ISSN 1574-020X. doi: 10.1007/s10579-020-09502-8. URL <https://doi.org/10.1007/s10579-020-09502-8>.
- [17] Paula Fortuna and Sérgio Nunes. A survey on automatic detection of hate speech in text. *ACM Comput. Surv.*, 51(4), jul 2018. ISSN 0360-0300. doi: 10.1145/3232676. URL <https://doi.org/10.1145/3232676>.
- [18] Sulistya Evingrum. Pendekatan dualistik terhadap kasus ujaran kebencian berdasarkan pancasila. *YUSTISIA MERDEKA : Jurnal Ilmiah Hukum*, 9(1):85–91, Aug. 2023. doi: 10.33319/yume.v9i1.227. URL <https://yustisia.unmermadiun.ac.id/index.php/yustisia/article/view/227>.
- [19] M Mowafy, A Rezk, and H El-Bakry. An efficient classification model for unstructured text document. *American Journal of Computer Science and Information Technology*, 6(1):16, 2018.
- [20] Santiago González-Carvajal and Eduardo C. Garrido-Merchán. Comparing BERT against traditional machine learning text classification. *CoRR*, abs/2005.13012, 2020. URL <https://arxiv.org/abs/2005.13012>.
- [21] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*, 2019.
-

- 
- [22] Ari Muzakir, Kusworo Adi, and Retno Kusumaningrum. Classification of hate speech language detection on social media: Preliminary study for improvement. In *International Conference on Networking, Intelligent Systems and Security*, pages 146–156. Springer, 2022.
- [23] Muhammad Okky Ibrohim and Indra Budi. Multi-label hate speech and abusive language detection in indonesian twitter. In *Proceedings of the third workshop on abusive language online*, pages 46–57, 2019.
- [24] Guntur Budi Herwanto, Annisa Maulida Ningtyas, I Gede Mujiyatna, Kurniawan Eka Nugraha, and I Nyoman Prayana Trisna. Hate speech detection in indonesian twitter using contextual embedding approach. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, 15(2):177–188, 2021.
- [25] Fajri Koto, Jey Han Lau, and Timothy Baldwin. Indobertweet: A pretrained language model for indonesian twitter with effective domain-specific vocabulary initialization. *arXiv preprint arXiv:2109.04607*, 2021.
- [26] Tri Widarmanti, Mutia Putri Widodo, Dian Puteri Ramadhani, and Muktar Danlami. Text emotion detection: Discover the meaning behind youtube comments using indo roberta. In *2022 International Conference on Advanced Creative Networks and Intelligent Systems (ICACNIS)*, pages 1–6. IEEE, 2022.
- [27] Ghinaa Zain Nabiilah, Simeon Yuda Prasetyo, Zahra Nabila Izdihar, and Abba Suganda Girsang. Bert base model for toxic comment analysis on indonesian social media. *Procedia Computer Science*, 216:714–721, 2023.
- [28] Rahmat Hendrawan, Adiwijaya, and Said Al Faraby. Multilabel classification of hate speech and abusive words on indonesian twitter social media. In *2020 International Conference on Data Science and Its Applications (ICoDSA)*, pages 1–7, 2020. doi: 10.1109/ICoDSA50139.2020.9212962.
- [29] Muhammad Amien Ibrahim, Samsul Arifin, I Gusti Agung Anom Yudistira, Rinda Nariswari, Abdul Azis Abdillah, Nerru Pranuta Murnaka, and Puguh Wahyu Prasetyo. An explainable ai model for hate speech detection on indonesian twitter. *CommIT (Communication and Information Technology) Journal*, 16(2):175–182, 2022.
- [30] Ritesh Kumar and Atul Kr Ojha. Kmi-panlingua at hasoc 2019: Svm vs bert for hate speech and offensive content detection. In *FIRE (Working Notes)*, pages 285–292, 2019.
- [31] Fajri Koto, Afshin Rahimi, Jey Han Lau, and Timothy Baldwin. Indolem and indobert: A benchmark dataset and pre-trained language model for indonesian nlp. *arXiv preprint arXiv:2011.00677*, 2020.
-

- [32] Muhammad Ikram Kaer Sinapoy, Yuliant Sibaroni, and Sri Suryani Prasetyowati. Comparison of lstm and indobert method in identifying hoax on twitter. *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, 7(3):657–662, Jun. 2023. doi: 10.29207/resti.v7i3.4830.
- [33] Yoga Sagama and Andry Alamsyah. Multi-label classification of indonesian online toxicity using bert and roberta. In *2023 IEEE International Conference on Industry 4.0, Artificial Intelligence, and Communications Technology (IAICT)*, pages 143–149, 2023. doi: 10.1109/IAICT59002.2023.10205892.
- [34] Swapnanil Mukherjee and Sujit Das. Application of transformer-based language models to detect hate speech in social media. *Journal of Computational and Cognitive Engineering*, 2(4):278–286, 2023.
- [35] Azal Ahmad Khan, Omkar Chaudhari, and Rohitash Chandra. A review of ensemble learning and data augmentation models for class imbalanced problems: combination, implementation and evaluation, 2023. URL <https://arxiv.org/abs/2304.02858>.
- [36] Leno Dwi Cahya, Ardytha Luthfiarta, Julius Immanuel Theo Krisna, Sri Winarno, and Adhitya Nugraha. Improving multi-label classification performance on imbalanced datasets through smote technique and data augmentation using indobert model. *Jurnal Nasional Teknologi dan Sistem Informasi*, 2024. URL <https://api.semanticscholar.org/CorpusID:267187378>.
- [37] Huailong Dong, Bowen Zhu, and Jing Zhang. A cost-sensitive active learning for imbalance data with uncertainty and diversity combination. In *Proceedings of the 2020 12th international conference on machine learning and computing*, pages 218–224, 2020.
- [38] M. O. Ibrohim and I. Budi. A dataset and preliminaries study for abusive language detection in indonesian social media. *Procedia Computer Science*, 135:222–229, 2018. doi: 10.1016/j.procs.2018.08.169.
- [39] Fredy Rodríguez-Torres, José F. Martínez-Trinidad, and Jesús A. Carrasco-Ochoa. An oversampling method for class imbalance problems on large datasets. *Applied Sciences*, 12(7), 2022. ISSN 2076-3417. doi: 10.3390/app12073424. URL <https://www.mdpi.com/2076-3417/12/7/3424>.
- [40] Nurulfitriah Noorhalim, Aida Ali, and Siti Mariyam Shamsuddin. Handling imbalanced ratio for class imbalance problem using smote. In Liew-Kee Kor, Abd-Razak Ahmad, Zanariah Idrus, and Kamarul Ariffin Mansor, editors, *Proceedings of the Third International Conference on Computing, Mathematics and Statistics (iCMS2017)*, pages 19–30, Singapore, 2019. Springer Singapore. ISBN 978-981-13-7279-7.

- 
- [41] Kevin W. Bowyer, Nitesh V. Chawla, Lawrence O. Hall, and W. Philip Kegelmeyer. SMOTE: synthetic minority over-sampling technique. *CoRR*, abs/1106.1813, 2011. URL <http://arxiv.org/abs/1106.1813>.
- [42] S. T. Jishan, R. I. Rashu, N. Haque, and R. M. Rahman. Improving accuracy of students' final grade prediction model using optimal equal width binning and synthetic minority over-sampling technique. *Decision Analytics*, 2(1):1, Dec. 2015. doi: 10.1186/s40165-014-0010-2.
- [43] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. *CoRR*, abs/1706.03762, 2017. URL <http://arxiv.org/abs/1706.03762>.
- [44] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.
- [45] Kuncahyo Setyo Nugroho, Anantha Yullian Sukmadewa, Haftittah Wuswilahaken DW, Fitra A Bachtiar, and Novanto Yudistira. Bert fine-tuning for sentiment analysis on indonesian mobile apps reviews. In *Proceedings of the 6th International Conference on Sustainable Information Engineering and Technology*, pages 258–264, 2021.
- [46] Flax Community. Indonesian roberta base. hugging face, 2021. URL <https://huggingface.co/flax-community/indonesian-roberta-base>.
- [47] Wilson Wongso. Indo roberta small. hugging face., 2021. URL <https://huggingface.co/w11wo/indo-roberta-small>.
- [48] Alaa Tharwat. Classification assessment methods. *Applied computing and informatics*, 17(1):168–192, 2021.