DAFTAR PUSTAKA

Agal, S., Dhirubhai Odedra, N., Chowdhary, H., Singh Ruprah, T., Ankalu Vuyyuru, V., Yousef ABaker El-Ebiary, T., & professor, A. (2024). Elevating Offensive Language Detection: CNN-GRU and BERT for Enhanced Hate Speech Identification. In *IJACSA) International Journal of Advanced Computer Science and Applications* (Vol. 15, Issue 5). www.ijacsa.thesai.org

Akter, M. S., Shahriar, H., Ahmed, N., & Cuzzocrea, A. (2022). Deep Learning Approach for Classifying the Aggressive Comments on Social Media: Machine Translated Data Vs Real Life Data. *Proceedings - 2022 IEEE International Conference on Big Data, Big Data 2022*, 5646–5655. https://doi.org/10.1109/BigData55660.2022.10020249

Chowdhury, S. B. R., Ghosh, S., Li, Y., Oliva, J. B., Srivastava, S., & Chaturvedi, S. (2021). *Adversarial Scrubbing of Demographic Information for Text Classification*. http://arxiv.org/abs/2109.08613

Das, D., Nayak, D. R., Dash, R., Majhi, B., & Zhang, Y. D. (2020). H-WordNet: A holistic convolutional neural network approach for handwritten word recognition. *IET Image Processing*, *14*(9), 1794–1805. https://doi.org/10.1049/iet-ipr.2019.1398

Deviyani, A., & Widjaja, H. (n.d.). *IndoHateSpeech: A Deep Learning Approach to Automated Multi-class Multi-label Hate Speech Detection on Informal Indonesian Corpora*.

Dirting, B. D., Chukwudebe, G. A., Nwokorie, E. C., & Ayogu, I. I. (2022a). Multi-Label Classification of Hate Speech Severity on Social Media using BERT Model. *Proceedings of the 2022 IEEE Nigeria 4th International Conference on Disruptive Technologies for Sustainable Development, NIGERCON 2022.* https://doi.org/10.1109/NIGERCON54645.2022.9803164

Dirting, B. D., Chukwudebe, G. A., Nwokorie, E. C., & Ayogu, I. I. (2022b). Multi-Label Classification of Hate Speech Severity on Social Media using

BERT Model. *Proceedings of the 2022 IEEE Nigeria 4th International Conference on Disruptive Technologies for Sustainable Development, NIGERCON 2022.* https://doi.org/10.1109/NIGERCON54645.2022.9803164

García-Díaz, J. A., Jiménez-Zafra, S. M., García-Cumbreras, M. A., & Valencia-García, R. (2023). Evaluating feature combination strategies for hate-speech detection in Spanish using linguistic features and transformers. *Complex and Intelligent Systems*, *9*(3), 2893–2914. https://doi.org/10.1007/s40747-022-00693-x

Guo, X., Anjum, U., & Zhan, J. (2022). Cyberbully Detection Using BERT with Augmented Texts. *Proceedings - 2022 IEEE International Conference on Big Data, Big Data 2022*, 1246□1253. https://doi.org/10.1109/BigData55660.2022.10020581

Hashmi, E., & Yayilgan, S. Y. (2024). Multi-class hate speech detection in the Norwegian language using FAST-RNN and multilingual fine-tuned transformers. *Complex and Intelligent Systems*, *10*(3), 4535□4556. https://doi.org/10.1007/s40747-024-01392-5

Ibrohim, M. O., & Budi, I. (2019). *Multi-label Hate Speech and Abusive Language Detection in Indonesian Twitter*. https://www.komnasham.go.id/index.php/

Jahan, S., Oussalah, M., Romaissa Beddia, D., Kabir Mim, J., & Arhab, N. (n.d.). *A Comprehensive Study on NLP Data Augmentation for Hate Speech Detection: Legacy Methods, BERT, and LLMs ⋆*. https://pypi.org/project/nlp-augment/

Liu, H., Jin, W., Karimi, H., Liu, Z., & Tang, J. (2021). *The Authors Matter: Understanding and Mitigating Implicit Bias in Deep Text Classification*. http://arxiv.org/abs/2105.02778

Luo, J., Bouazizi, M., & Ohtsuki, T. (2021). Data Augmentation for Sentiment Analysis Using Sentence Compression-Based SeqGAN with Data Screening. *IEEE Access*, *9*, 99922□99931. https://doi.org/10.1109/ACCESS.2021.3094023

Matamoros-Fern⟨ndez, A., & Farkas, J. (2021). Racism, Hate Speech, and Social Media: A Systematic Review and Critique. *Television and New Media*, *22*(2), 205□224. https://doi.org/10.1177/1527476420982230

Mnassri, K., Rajapaksha, P., Farahbakhsh, R., & Crespi, N. (2022a). BERT-based Ensemble Approaches for Hate Speech Detection. *Proceedings - IEEE Global*

Communications Conference, GLOBECOM, 4649-4654. https://doi.org/10.1109/GLOBECOM48099.2022.10001325

Mnassri, K., Rajapaksha, P., Farahbakhsh, R., & Crespi, N. (2022b). BERT-based Ensemble Approaches for Hate Speech Detection. *Proceedings - IEEE Global Communications Conference, GLOBECOM, 4649-4654. https://doi.org/10.1109/GLOBECOM48099.2022.10001325*

Mutanga, R. T., Naicker, N., & Olugbara, O. O. (2020). Hate Speech Detection in Twitter using Transformer Methods. In *IJACSA) International Journal of Advanced Computer Science and Applications* (Vol. 11, Issue 9). www.ijacsa.thesai.org

Pamungkas, E. W., Galih, D., Putri, P., & Fatmawati, A. (n.d.). Hate Speech Detection in Bahasa Indonesia: Challenges and Opportunities. In *IJACSA) International Journal of Advanced Computer Science and Applications* (Vol. 14, Issue 6). https://www.statista.com/statistics/242606/

Park, H., & Kim, H. K. (2021). Verbal Abuse Classification Using Multiple Deep Neural Networks. *3rd International Conference on Artificial Intelligence in Information and Communication, ICAIIC2021*, 316□319. https://doi.org/10.1109/ICAIIC51459.2021.9415218

Paul, C., & Bora, P. (2021). Detecting Hate Speech using Deep Learning Techniques. In *IJACSA) International Journal of Advanced Computer Science and Applications* (Vol. 12, Issue 2). www.ijacsa.thesai.org

Perez, J. M., Luque, F. M., Zayat, D., Kondratzky, M., Moro, A., Serrati, P. S., Zajac, J., Miguel, P., Debandi, N., Gravano, A., & Cotik, V. (2023). Assessing the Impact of Contextual Information in Hate Speech Detection. *IEEE Access*, *11*, 30575□30590. https://doi.org/10.1109/ACCESS.2023.3258973

Pruksachatkun, Y., Krishna, S., Dhamala, J., Gupta, R., & Chang, K.-W. (2021). *Does Robustness Improve Fairness? Approaching Fairness with Word Substitution Robustness Methods for Text Classification*. http://arxiv.org/abs/2106.10826

Putu Widiarta Nandana Githa, I., 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 2024*, 61□65. https://doi.org/10.1109/GECOST60902.2024.10474781

Sabry, S. S., Adewumi, T., Abid, N., Kovacs, G., Liwicki, F., & Liwicki, M. (2022). HaT5: Hate Language Identification using Text-to-Text Transfer Transformer. *Proceedings of the International Joint Conference on Neural Networks*, *2022-July*. https://doi.org/10.1109/IJCNN55064.2022.9892696

Sharmila, P., Anbananthen, K. S. M., Chelliah, D., Parthasarathy, S., & Kannan, S. (2022). PDHS: Pattern-Based Deep Hate Speech Detection with Improved Tweet Representation. *IEEE Access*, *10*, 105366□105376. https://doi.org/10.1109/ACCESS.2022.3210177

Soni, P. K., & Rambola, R. K. (2021, June 25). Deep Learning, WordNet, and spaCy based Hybrid Method for Detection of Implicit Aspects for Sentiment Analysis. *2021 International Conference on Intelligent Technologies, CONIT2021*. https://doi.org/10.1109/CONIT51480.2021.9498372

Teimas, R., & Saias, J. (2023). Detecting Persuasion Attempts on Social Networks: Unearthing the Potential of Loss Functions and Text Pre-Processing in Imbalanced Data Settings. *Electronics (Switzerland)*, *12*(21). https://doi.org/10.3390/electronics12214447

Yu, S., Song, S., & Kim, Y. (2024). Rational Text Augmentation Method with Korean Misspellings. *Digest of Technical Papers - IEEE International Conference on Consumer Electronics*. https://doi.org/10.1109/ICCE59016.2024.10444190

Yuan, S., & Maronikolakis, A. (2022). *Separating Hate Speech and Offensive Language Classes via Adversarial Debiasing*. https://github.com/

Fortuna, P., & Nunes, S. (2018). A survey on automatic detection of hate speech in social networks. *ACM Computing Surveys*, 51(4), 85.

Yin, D., & Zubiaga, A. (2021). Challenges and opportunities in hate speech detection: A review. *Journal of Computational Social Science*, 4(2), 147□165

Gr)ndahl, E., Modha, S., Joshi, A., & Pretorius, A. (2018). Adversarial attacks on hate speech detection models: Are they a real threat? In *Proceedings of the First Workshop on Trolling, Aggression and Cyberbullying* (pp. 55□65). ACL

De Smedt, Q., Moreau, E., & Lepoutre, A. (2018). Early detection of jihadist hate speech on social media. In *Proceedings of the International AAAI Conference on Web and Social Media* (Vol. 12, pp. 278-287).

Waseem, Z., & Hovy, D. (2016). Hateful symbols or hateful people? Predictive features for hate speech detection on Twitter. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics* (pp. 88□97).

Nobata, C., Tetreault, J., Thomas, A., Mehdad, Y., & Chang, Y. (2016). Abusive language detection in online user content. In *Proceedings of the 25th International Conference on World Wide Web* (pp. 145□153).

Badjatiya, P., Gupta, S., Varma, V., & Ferm⎝ller, C. (2017). Deep learning for hate speech detection in tweets. In *Proceedings of the 26th International Conference on World Wide Web Companion* (pp. 759□760).

Albadi, N., Habash, N., & Rambow, O. (2018). Dialectal variation and its impact on hate speech detection for Arabic. In *IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining* (pp. 23□30).

Peli⎝on, L., Martinc, M., & Kralj Novak, P. (2019). Semeval-2019 Task 6: Multilingual detection of hate speech against immigrants and women in Twitter. In *Proceedings of the 13th International Workshop on Semantic Evaluation* (pp. 54□63).