Indonesian Headline Detection on Twitter Social Media

Kaenova Mahendra Auditama
School of Computing
Telkom Univeristy
Bandung, Indonesia
kaenova@student.telkomuniversity.ac.id

Mahendra Dwifebri Purbolaksono
School of Computing
Telkom University
Bandung, Indonesia
mahendradp@telkomuniversity.ac.id

Ade Romadhony
School of Computing
Telkom University
Bandung, Indonesia
aderomadhony@telkomuniversity.ac.id

REFERENCES

[1] S. A. Salloum, M. Al-Emran, A. A. Monem, and K. Shaalan, "A Survey of Text Mining in Social Media: Facebook and Twitter Perspectives", doi: 10.25046/aj020115.

[2] I. Dilrukshi, K. De Zoysa, and A. Caldera, "Twitter news classification using SVM," *Proceedings of the 8th International Conference on Computer Science and Education, ICCSE 2013*, pp. 287–291, 2013, doi: 10.1109/ICCSE.2013.6553926.

[3] T. A. S. Rohmah and W. Maharani, "Personality Detection on Twitter Social Media Using IndoBERT Method," *Building of Informatics, Technology and Science (BITS)*, vol. 4, no. 2, pp. 448–453, Sep. 2022, doi: 10.47065/BITS.V4I2.1895.

[4] A. Agarwal, B. Xie, I. Vovsha, O. Rambow, and R. Passonneau, "Sentiment Analysis of Twitter Data," pp. 30–38, 2011, Accessed: Oct. 17, 2022. [Online]. Available: http://www.webconfs.com/stop-words.php

[5] S. Wakade, C. Shekar, K. J. Liszka, and C.-C. Chan, "Text Mining for Sentiment Analysis of Twitter Data", Accessed: Oct. 17, 2022. [Online]. Available: http://jaiku.com

[6] S. Saadah, K. M. Auditama, A. A. Fattahila, F. I. Amorokhman, A. Aditsania, and A. A. Rohmawati, "Implementation of BERT, IndoBERT, and CNN-LSTM in Classifying Public Opinion about COVID-19 Vaccine in Indonesia," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 6, no. 4, pp. 648–655, Aug. 2022, doi: 10.29207/RESTI.V6I4.4215.

[7] A. R. T. Lestari, R. S. Perdana, and M. A. Fauzi, "Analisis Sentimen Tentang Opini Pilkada DKI 2017 Pada Dokumen Twitter Berbahasa Indonesia Menggunakan Naïve Bayes dan Pembobotan Emoji," *Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 1, no. 12, 2017.

[8] W. F. Abdillah, A. Premana, and R. M. H. Bhakti, "Analisis Sentimen Penanganan Covid-19 dengan Support Vector Machine: Evaluasi Leksikon dan Metode Ekstraksi Fitur," *Jurnal Ilmiah Intech : Information Technology Journal of UMUS*, vol. 3, no. 02, 2021, doi: 10.46772/intech.v3i02.556.

[9] T. Wu, S. Liu, J. Zhang, and Y. Xiang, "Twitter spam detection based on deep learning," in *ACM International Conference Proceeding Series*, 2017. doi: 10.1145/3014812.3014815.

[10] A. G. D'Sa, I. Illina, and D. Fohr, "BERT and fastText Embeddings for Automatic Detection of Toxic Speech," in *Proceedings of 2020 International Multi-Conference on: Organization of Knowledge and Advanced Technologies, OCTA 2020*, 2020. doi: 10.1109/OCTA49274.2020.9151853.

[11] Merriam-Webster, "headline," *Merriam-Webster.com*, 2011. https://www.merriam-webster.com/dictionary/headline (accessed Nov. 01, 2022).

[12] M. I. Rana, S. Khalid, and M. U. Akbar, "News classification based on their headlines: A review," in *17th IEEE International Multi Topic Conference: Collaborative and Sustainable Development of Technologies, IEEE INMIC 2014 - Proceedings*, 2015. doi: 10.1109/INMIC.2014.7097339.

[13] C. Naury, D. H. Fudholi, and A. F. Hidayatullah, "Topic Modelling pada Sentimen Terhadap Headline Berita Online Berbahasa Indonesia Menggunakan LDA dan LSTM," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 5, no. 1, 2021, doi: 10.30865/mib.v5i1.2556.

[14] N. Napu, "English and Indonesian Newspaper Headlines: A Comparative Study of Lexical Features," *European Journal of Literature, Language and ...*, no. 1998, 2018.

[15] A. Piotrkowicz, V. Dimitrova, J. Otterbacher, and K. Markert, "The Impact of News Values and Linguistic Style on the Popularity of Headlines on Twitter and Facebook," in *Eleventh International AAAI Conference on Web and Social Media*, 2017.

[16] A. William and Y. Sari, "CLICK-ID: A novel dataset for Indonesian clickbait headlines," *Data Brief*, vol. 32, 2020, doi: 10.1016/j.dib.2020.106231.

[17] J. Sirusstara, N. Alexander, A. Alfarisy, S. Achmad, and R. Sutoyo, "Clickbait Headline Detection in Indonesian News Sites using Robustly Optimized BERT Pre-training Approach (RoBERTa)," in *2022 3rd International Conference on Artificial Intelligence and Data Sciences: Championing Innovations in Artificial Intelligence and Data Sciences for Sustainable Future, AiDAS 2022 - Proceedings*, 2022. doi: 10.1109/AiDAS56890.2022.9918678.

[18] Diyah Utami Kusumaning Putri and Dinar Nugroho Pratomo, "Clickbait Detection of Indonesian News Headlines using Fine-Tune Bidirectional Encoder Representations from Transformers (BERT)," *Inform : Jurnal Ilmiah Bidang Teknologi Informasi dan Komunikasi*, vol. 7, no. 2, 2022, doi: 10.25139/inform.v7i2.4686.

[19] M. Noor Fakhruzzaman, idah Zahrotul Jannah, R. Ardiati Ningrum, and I. Fahmiyah, "Clickbait Headline Detection in Indonesian News Sites using Multilingual Bidirectional Encoder Representations from Transformers (M-BERT)," Feb. 2021, Accessed: Apr. 12, 2023. [Online]. Available: https://arxiv.org/abs/2102.01497v1

[20] B. Siregar, I. Habibie, E. B. Nababan, and Fahmi, "Identification of Indonesian clickbait news headlines with long short-term memory recurrent neural network algorithm," *J Phys Conf Ser*, vol. 1882, no. 1, p. 012129, May 2021, doi: 10.1088/1742-6596/1882/1/012129.

[21] M. N. Fakhruzzaman and S. W. Gunawan, "Web-based Application for Detecting Indonesian Clickbait Headlines using IndoBERT," Feb. 2021, Accessed: Apr. 12, 2023. [Online]. Available: https://arxiv.org/abs/2102.10601v1

[22] M. McCord and M. Chuah, "Spam detection on twitter using traditional classifiers," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2011. doi: 10.1007/978-3-642-23496-5_13.

[23] M. Z. Amin and N. Nadeem, "Convolutional Neural Network: Text Classification Model for Open Domain Question Answering System," Sep. 2018, Accessed: Apr. 12, 2023. [Online]. Available: https://arxiv.org/abs/1809.02479v2

[24] P. Zhou, Z. Qi, S. Zheng, J. Xu, H. Bao, and B. Xu, "Text classification improved by integrating bidirectional LSTM with two-dimensional max pooling," in *COLING 2016 - 26th International Conference on Computational Linguistics, Proceedings of COLING 2016: Technical Papers*, 2016.

[25] A. A. Fattahila, F. Irfan Amorokhman, K. M. Auditama, K. Ahmad Wijaya, and A. Romadhony, "Indonesian Digital Wallet Sentiment Analysis Using CNN and LSTM Method," in *2021 International Conference on Artificial Intelligence and Big Data Analytics, ICAIBDA 2021*, 2021. doi: 10.1109/ICAIBDA53487.2021.9689712.

[26] A. Vaswani *et al.*, "Attention is all you need," in *Advances in Neural Information Processing Systems*, 2017.

[27] 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.

[28] F. Koto, J. H. Lau, and T. Baldwin, "INDOBERTWEET: A Pretrained Language Model for Indonesian Twitter with Effective Domain-Specific Vocabulary Initialization," in *EMNLP 2021 - 2021 Conference on Empirical Methods in Natural Language*

Processing, Proceedings, 2021. doi: 10.18653/v1/2021.emnlp-main.833.

[29] 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.

[30] R. Wongso, F. A. Luwinda, B. C. Trisnajaya, O. Rusli, and Rudy, "News Article Text Classification in Indonesian Language," in *Procedia Computer Science*, 2017. doi: 10.1016/j.procs.2017.10.039.

[31] A. Joulin, E. Grave, P. Bojanowski, and T. Mikolov, "Bag of tricks for efficient text classification," in *15th Conference of the European Chapter of the Association for Computational Linguistics, EACL 2017 - Proceedings of Conference*, 2017. doi: 10.18653/v1/e17-2068.

[32] P. Bojanowski, E. Grave, A. Joulin, and T. Mikolov, "Enriching Word Vectors with Subword Information," *Trans Assoc Comput Linguist*, vol. 5, pp. 135–146, Jul. 2016, doi: 10.1162/tacl_a_00051.

[33] Y. Kim, "Convolutional neural networks for sentence classification," in *EMNLP 2014 - 2014 Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference*, 2014. doi: 10.3115/v1/d14-1181.

[34] Anjali, G. Jivani, and M. Anjali, "A Comparative Study of Stemming Algorithms," *October*, vol. 2, no. 2004, 2007.

[35] Bunyamin, A. F. Huda, and A. A. Suryani, "Indonesian Stemmer for Ambiguous Word based on Context," in *2021 International Conference on Data Science and Its Applications, ICoDSA 2021*, 2021. doi: 10.1109/ICoDSA53588.2021.9617514.