DAFTAR PUSTAKA

Abdullah, A. O, Ali, M. A, Karabatak, M, dan Sengur, A, mar 2018, A comparative analysis of common YouTube comment spam filtering techniques. In *2018 6th International Symposium on Digital Forensic and Security (ISDFS)*, pages 1–5, Antalya, Turkey. IEEE. ISBN 978-1-5386-3449-3. doi: 10.1109/ISDFS.2018.8355315. URL https://ieeexplore.ieee.org/document/8355315/.

Abinaya, R, Bertilla Niveda, E, dan Naveen, P, 2020, *Spam detection on social media platforms. 2020 7th International Conference on Smart Structures and Systems, ICSSS 2020*, pages 31–33. doi: 10.1109/ICSSS49621.2020.9201948.

Agarwal, B, Ramampiaro, H, Langseth, H, dan Ruocco, M, 2018, *A deep network model for paraphrase detection in short text messages. Information Processing and Management*, 54(6):922–937. ISSN 03064573. doi: 10.1016/j.ipm.2018.06.005. URL https://doi.org/10.1016/j.ipm.2018.06.005.

Aiyar, S dan Shetty, N. P, 2018, *N-Gram Assisted Youtube Spam Comment Detection. Procedia Computer Science*, 132:174–182. ISSN 18770509. doi: 10.1016/j.procs.2018.05.181. URL https://linkinghub.elsevier.com/retrieve/pii/S1877050918309153.

Akilan, A, jun 2015, Text mining: Challenges and future directions. In *2nd International Conference on Electronics and Communication Systems, ICECS 2015*, pages 1679–1683. Institute of Electrical and Electronics Engineers Inc. ISBN 9781479972241. doi: 10.1109/ECS.2015.7124872.

AL-Smadi, M, Jaradat, Z, AL-Ayyoub, M, dan Jararweh, Y, 2017, *Paraphrase identification and semantic text similarity analysis in Arabic news tweets using lexical, syntactic, and semantic features. Information Processing and Management*, 53(3):640–652. ISSN 03064573. doi: 10.1016/j.ipm.2017.01.002. URL http://dx.doi.org/10.1016/j.ipm.2017.01.002.

Alam, T, Khan, A, dan Alam, F, 2020, Bangla text classification using transformers. In *arXiv*.

Alias, N, Foozy, C. F. M, dan Ramli, S. N, 2019, *Video spam comment features selection using machine learning techniques. Indonesian Journal of Electrical*

Engineering and Computer Science, 15(2):1046–1053. ISSN 25024760. doi: 10.11591/ijeecs.v15.i2.pp1046-1053.

Alita, D, Priyanta, S, Rokhman, N, Zainal Abidin Pagaralam, J. H, Lampung, B, dan Utara Bulaksumur Kabupaten Sleman, S, oct 2019, *Analysis of Emoticon and Sarcasm Effect on Sentiment Analysis of Indonesian Language on Twitter*. *Journal of Information Systems Engineering and Business Intelligence*, 5(2):100–109. ISSN 2443-2555. doi: 10.20473/JISEBI.5.2.100-109. URL https://e-journal.unair.ac.id/JISEBI/article/view/12101http://creativecommons.org/licenses/by/4.0/.

Allstars, 2021. *10 Artis Followers Terbanyak di Indonesia pada Instagram di 2021*. URL https://www.allstars.id/blog/2021/09/23/artis-indonesia-dengan-followers-instagram-terbanyak/. [Diakses pada: 2021-10-27].

Almeida, T. A, Hidalgo, J. M. G, dan Yamakami, A, 2011, *Contributions to the study of SMS spam filtering: New collection and results*. *DocEng 2011 - Proceedings of the 2011 ACM Symposium on Document Engineering*, pages 259–262. doi: 10.1145/2034691.2034742.

Arthana, R, 2019. *Mengenal Accuracy, Precision, Recall dan Specificity serta yang diprioritaskan dalam Machine Learning*. URL https://rey1024.medium.com/mengenal-accuracy-precission-recall-dan-specificity-serta-yang-diprioritaskan-b79ff4d77de8. [Diakses pada: 2021-09-01].

Asif, M, Akhtar, N, Husnain, M, Muhammad, M, Asmat, H, dan Asghar, M, 2016, *Hashtag the Tweets: Experimental Evaluation of Semantic Relatedness Measures*. *International Journal of Advanced Computer Science and Applications*, 7(6):474–482. ISSN 21565570. doi: 10.14569/IJACSA.2016.070662. URL http://thesai.org/Publications/ViewPaper?Volume=7{&}Issue=6{&}Code=ijacsa{&}SerialNo=62.

Aslam, S, 2018. *Twitter by the Numbers (2018): Stats, Demographics & Fun Facts*. URL https://www.omnicoreagency.com/twitter-statistics/. [Diakses pada: 2021-05-11].

Aslam, S, 2021. *Instagram by the Numbers (2021): Stats, Demographics & Fun Facts*. URL https://www.omnicoreagency.com/instagram-statistics/. [Diakses pada: 2021-05-11].

Bahdanau, D, Cho, K. H, dan Bengio, Y, 2015, *Neural machine translation by jointly learning to align and translate. 3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, pages 1–15.

Ban, X, Chen, C, Liu, S, Wang, Y, dan Zhang, J, 2018, *Deep-learnt features for Twitter spam detection. 2018 International Symposium on Security and Privacy in Social Networks and Big Data, SocialSec 2018*, pages 22–26. doi: 10.1109/SocialSec.2018.8760377.

Bansal, A, 2021, *Advanced Natural Language Processing with TensorFlow 2*. Packt Publishing, Mumbai, India, 1st edition. ISBN 9781800200937. URL https://www.packtpub.com/product/advanced-natural-language-processing-with-tensorflow-2/9781800200937.

Bao, W, Bao, W, Du, J, Yang, Y, dan Zhao, X, nov 2018, Attentive Siamese LSTM Network for Semantic Textual Similarity Measure. In *2018 International Conference on Asian Language Processing (IALP)*, pages 312–317. IEEE. ISBN 978-1-7281-1175-9. doi: 10.1109/IALP.2018.8629212. URL https://ieeexplore.ieee.org/document/8629212/.

Barbieri, F, Ronzano, F, dan Saggion, H, 2016, *What does this emoji mean? A vector space skip-gram model for twitter emojis. Proceedings of the 10th International Conference on Language Resources and Evaluation, LREC 2016*, pages 3967–3972.

Barik, A. M, Mahendra, R, dan Adriani, M, 2019, Normalization of Indonesian-English Code-Mixed Twitter Data. In *Proceedings of the 5th Workshop on Noisy User-generated Text (W-NUT 2019)*, pages 417–424, Hong Kong, China. Association for Computational Linguistics. doi: 10.18653/v1/d19-5554. URL https://www.aclweb.org/anthology/D19-5554/.

Besti, A, Ilyas, R, Kasyidi, F, dan Djamal, E. C, 2020, *Semantic classification of scientific sentence pair using recurrent neural network. International Conference on Electrical Engineering, Computer Science and Informatics*

(EECSI), 2020-Octob:150–155. ISSN 2407439X. doi: 10.23919/EECSI50503.2020.9251897.

Bogdanova, D, dos Santos, C, Barbosa, L, dan Zadrozny, B, 2015, Detecting Semantically Equivalent Questions in Online User Forums. In *Proceedings of the Nineteenth Conference on Computational Natural Language Learning*, pages 123–131, Beijing, China. Association for Computational Linguistics. ISBN 9781941643778. doi: 10.18653/v1/K15-1013. URL http://aclweb.org/anthology/K15-1013.

Bojanowski, P, Grave, E, Joulin, A, dan Mikolov, T, 2017, Enriching Word Vectors with Subword Information. In *Transactions of the Association for Computational Linguistics*, volume 5, pages 135–146. doi: 10.1162/tacl_a_00051.

Borlander, T, 2016. *Bolandish/PHP-Instagram-Grabber*. URL https://github.com/Bolandish/PHP-Instagram-Grabber.

Bowman, S. R, Angeli, G, Potts Christopher, dan Manning, C. D, 2015, A large annotated corpus for learning natural language inference. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Association for Computational Linguistics.

Brownlee, J, 2020. *4 Types of Classification Tasks in Machine Learning*. URL https://machinelearningmastery.com/types-of-classification-in-machine-learning/. [Diakses pada: 2021-09-02].

Burns, R, 2019. *188 Spam Words to Avoid: How to Stay Out of Spam Filters*. URL https://www.activecampaign.com/blog/spam-words. [Diakses pada: 2021-03-06].

Campos, R, Mangaravite, V, Pasquali, A, Jorge, A, Nunes, C, dan Jatowt, A, 2020, *YAKE! Keyword extraction from single documents using multiple local features*. *Information Sciences*, 509:257–289. ISSN 00200255. doi: 10.1016/j.ins.2019.09.013.

Cao, J dan Lai, C, 2020, *A bilingual multi-type spam detection model based on M-BERT*. *2020 IEEE Global Communications Conference, GLOBECOM 2020 - Proceedings*, 2020-Janua. doi: 10.1109/GLOBECOM42002.2020.9347970.

Chandra, A dan Khatri, S. K, 2019, *Spam SMS Filtering using Recurrent Neural Network and Long Short Term Memory. 2019 4th International Conference on Information Systems and Computer Networks, ISCON 2019*, pages 118–122. doi: 10.1109/ISCON47742.2019.9036269.

Chen, C, Wang, Y, Zhang, J, Xiang, Y, Zhou, W, dan Min, G, apr 2017a, *Statistical Features-Based Real-Time Detection of Drifted Twitter Spam. IEEE Transactions on Information Forensics and Security*, 12(4):914–925. ISSN 1556-6013. doi: 10.1109/TIFS.2016.2621888. URL http://ieeexplore.ieee.org/document/7707341/.

Chen, G, Ye, D, Xing, Z, Chen, J, dan Cambria, E, may 2017b, Ensemble application of convolutional and recurrent neural networks for multi-label text categorization. In *2017 International Joint Conference on Neural Networks (IJCNN)*, volume 2017-May, pages 2377–2383. IEEE. ISBN 978-1-5090-6182-2. doi: 10.1109/IJCNN.2017.7966144. URL http://ieeexplore.ieee.org/document/7966144/.

Chrismanto, A. R dan Lukito, Y, 2017a, *Identifikasi Komentar Spam Pada Instagram. Lontar Komputer : Jurnal Ilmiah Teknologi Informasi*, 8(3):219. ISSN 2088-1541. doi: 10.24843/lkjiti.2017.v08.i03.p08.

Chrismanto, A. R dan Lukito, Y, 2017b, Klasifikasi Komentar Spam Pada Instagram Berbahasa Indonesia Menggunakan K-NN. In *Seminar Nasional Teknologi Informasi Kesehatan (SNATIK)*, pages 298–306, Yogyakarta. STIKES Surya Global.

Chrismanto, A. R, Lukito, Y, dan Susilo, A, 2020, *Implementasi Distance Weighted K-Nearest Neighbor Untuk Klasifikasi Spam & Non-Spam Pada Komentar Instagram. Jurnal Edukasi dan Penelitian Informatika (JEPIN)*, 6(2):236. ISSN 2460-0741. doi: 10.26418/jp.v6i2.39996.

Chrismanto, A. R, Raharjo, W. S, dan Lukito, Y, sep 2018a, *Design and Development of REST-Based Instagram Spam Detector for Indonesian Language. Proceedings - 2018 International Seminar on Application for Technology of Information and Communication: Creative Technology for Human Life, iSemantic 2018*, pages 345–350. doi: 10.1109/ISEMANTIC.2018.8549725. URL https://ieeexplore.ieee.org/document/8549725/.

Chrismanto, A. R, Raharjo, W. S, dan Lukito, Y, 2019a, *Firefox Extension untuk Klasifikasi Komentar Spam pada Instagram Berbasis REST Services. Jurnal Edukasi dan Penelitian Informatika (JEPIN)*, 5(2):146. ISSN 2460-0741. doi: 10.26418/jp.v5i2.33010.

Chrismanto, A. R, Sari, A. K, dan Suyanto, Y, 2022, *SPAMID-PAIR: A Novel Indonesian Post–Comment Pairs Dataset Containing Emoji. International Journal of Advanced Computer Science and Applications*, 13(11):92–100. ISSN 2156-5570. doi: 10.14569/IJACSA.2022.0131110. URL www.ijacsa.thesai.org.

Chrismanto, A. R, Sari, A. K, dan Suyanto, Y, 2023, *Enhancing Spam Comment Detection on Social Media With Emoji Feature and Post-Comment Pairs Approach Using Ensemble Methods of Machine Learning. IEEE Access*, 11: 80246–80265. ISSN 2169-3536. doi: 10.1109/ACCESS.2023.3299853. URL https://ieeexplore.ieee.org/document/10196430/.

Chrismanto, A. R, Sudiarto, W, dan Lukito, Y, 2018b, *Integration of REST-Based Web Service and Browser Extension for Instagram Spam Detection. International Journal of Advanced Computer Science and Applications*, 9(12). ISSN 21565570. doi: 10.14569/IJACSA.2018.091253. URL https://thesai.org/Publications/ViewPaper?Volume=9{&}Issue=12{&}Code=IJACSA{&}SerialNo=53.

Chrismanto, A, Delima, R, Santoso, H, Wibowo, A, dan Kristiawan, R, 2019b, *Developing agriculture land mapping using Rapid Application Development (RAD): A case study from Indonesia. International Journal of Advanced Computer Science and Applications*, 10(10). ISSN 21565570. doi: 10.14569/ijacsa.2019.0101033.

Chromium, 2022. *ChromeDriver - WebDriver for Chrome* . URL https://chromedriver.chromium.org/downloads. [Diakses pada: 2022-05-14].

Convey, E, 1996. *Porn sneaks way back on web.* URL https://scholar.google.com/scholar?q=Convey{%}2CE.{%}3APornsneakswaybackonweb.TheBostonHerald{%}2Cp.028{%}28May22{%}2C1996{%}29. [Diakses pada: 2021-02-26].

Dada, E. G, Bassi, J. S, Chiroma, H, Abdulhamid, S. M, Adetunmbi, A. O, dan Ajibuwa, O. E, 2019, *Machine learning for email spam filtering: review, approaches and open research problems*. *Heliyon*, 5(6). ISSN 24058440. doi: 10.1016/j.heliyon.2019.e01802.

Databooks, 2020. *Ini Media Sosial Paling Populer Sepanjang April 2020*. URL https://databoks.katadata.co.id/datapublish/2020/05/25/ini-media-sosial-paling-populer-sepanjang-april-2020. [Diakses pada: 2020-11-04].

Devlin, J, Chang, M. W, Lee, K, dan Toutanova, K, 2019, *BERT: Pre-training of deep bidirectional transformers for language understanding*. *NAACL HLT 2019 - 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference*, 1(Mlm):4171–4186.

Eisner, B, Augenstein, I, Rockt, T, dan Riedel, S, 2016, emoji2vec : Learning Emoji Representations from their Description. In *Conference: Proceedings of The Fourth International Workshop on Natural Language Processing for Social Media*, Austin, TX. Association for Computational Linguistics. ISBN 7653904726.

Etaiwi, W dan Naymat, G, 2017, *The Impact of applying Different Preprocessing Steps on Review Spam Detection*. *Procedia Computer Science*, 113:273–279. ISSN 18770509. doi: 10.1016/j.procs.2017.08.368. URL http://dx.doi.org/10.1016/j.procs.2017.08.368.

Ezzikouri, H, Erritali, M, dan Oukessou, M, mar 2016, Semantic Similarity/Relatedness for Cross Language Plagiarism Detection. In *2016 13th International Conference on Computer Graphics, Imaging and Visualization (CGiV)*, pages 372–374. IEEE. ISBN 978-1-5090-0811-7. doi: 10.1109/CGiV.2016.78. URL https://ieeexplore.ieee.org/document/7467740/.

Fatichah, C, Lazuardi, W. F, Navastara, D. A, Suciati, N, dan Munif, A, 2021, *A Content Filtering from Spam Posts on Social Media using Weighted Multimodal Approach*. *Journal of Computer Science*, 17(1):55–66. ISSN 15526607. doi: 10.3844/jcssp.2021.55.66.

Feldman, R dan Sanger, J, 2006, *The Text Mining Handbook*. Cambridge University Press, Cambridge, 1 edition. ISBN 9780511546914. doi: 10.1017/CBO9780511546914. URL https://www.cambridge.org/core/product/identifier/9780511546914/type/book.

Fonseca, E dan Alvarenga, J. P. R, 2020, *Wide and deep transformers applied to semantic relatedness and textual entailment*. *CEUR Workshop Proceedings*, 2583(1):68–76. ISSN 16130073.

Freischlag, C, 2020. *Combining numerical and text features in deep neural networks*. URL https://towardsdatascience.com/combining-numerical-and-text-features-in-deep-neural-networks-e91f0237eea4. [Diakses pada: 2022-01-09].

Géron, A, 2022, *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*. O'Reilly Media, Inc, Sebastopol, 3rd edition. ISBN 978-1-098-12597-4. URL https://oreilly.com/catalog/errata.csp?isbn=9781492032649.

Ghanem, R dan Erbay, H, 2020, *Context-dependent model for spam detection on social networks*. *SN Applied Sciences*, 2(9):1–8. ISSN 2523-3963. doi: 10.1007/s42452-020-03374-x. URL https://doi.org/10.1007/s42452-020-03374-x.

Godbole, A, Dalmia, A, dan Sahu, S. K, jan 2018, Siamese Neural Networks with Random Forest for detecting duplicate question pairs. In *arXiv*, pages 1–5. URL http://arxiv.org/abs/1801.07288.

Goodfellow, I, Bengio, Y, dan Courville, A, 2016, *Deep Learning*. MIT Press, Massachusetts. ISBN 9780262035613.

Gunawan, D, Saniyah, Z, dan Hizriadi, A, 2019, *Normalization of abbreviation and acronym on microtext in Bahasa Indonesia by using dictionary-based and longest common subsequence (LCS)*. *Procedia Computer Science*, 161:553–559. ISSN 18770509. doi: 10.1016/j.procs.2019.11.155. URL https://doi.org/10.1016/j.procs.2019.11.155.

Hadj Taieb, M. A, Zesch, T, dan Ben Aouicha, M, dec 2019, *A survey of semantic relatedness evaluation datasets and procedures*. *Artificial Intelligence Review*,

53(6):4407–4448. ISSN 0269-2821. doi: 10.1007/s10462-019-09796-3. URL http://link.springer.com/10.1007/s10462-019-09796-3.

Hammad, M, Al-Smadi, M, Baker, Q. B, dan Al-Zboon, S. A, aug 2021, *Using deep learning models for learning semantic text similarity of Arabic questions. International Journal of Electrical and Computer Engineering (IJECE)*, 11(4): 3519. ISSN 2722-2578. doi: 10.11591/ijece.v11i4.pp3519-3528. URL http://ijece.iaescore.com/index.php/IJECE/article/view/24008.

Han, J, Kamber, M, dan Pei, J, 2011, *Data Mining : Concepts and Techniques*. Morgan Kaufmann, 3rd edition. ISBN 978-9380931913. URL https://www.amazon.com/Data-Mining-Concepts-Techniques-Management/dp/0123814790.

Haqimi, N. A, Rokhman, N, dan Priyanta, S, jul 2019, *Detection Of Spam Comments On Instagram Using Complementary Naïve Bayes. IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, 13(3):263. ISSN 2460-7258. doi: 10.22146/ijccs.47046. URL https://jurnal.ugm.ac.id/ijccs/article/view/47046.

He, P, Liu, X, Gao, J, dan Chen, W, 2020, DeBERTa: Decoding-enhanced BERT with Disentangled Attention. In *arXiv*, pages 1–21.

Hines, K, 2021. *How to Identify and Control Blog Comment Spam on WordPress*. URL https://neilpatel.com/blog/control-blog-comment-spam/. [Diakses pada: 2021-02-27].

Hinton, S dan Hjorth, L, 2013, *Understanding Social Media*. SAGE Publications Ltd, 1 Oliver's Yard, 55 City Road, London EC1Y 1SP United Kingdom. ISBN 9781446201213. doi: 10.4135/9781446270189. URL http://sk.sagepub.com/books/understanding-social-media.

Hochreiter, S dan Schmidhuber, J, 1997, *Long Short-Term Memory. Neural Computation*, 9(8):1735–1780. ISSN 08997667. doi: 10.1162/neco.1997.9.8.1735.

Homma, Y, Sy, S, dan Yeh, C, 2016, Detecting Duplicate Questions with Deep Learning. In *30th Conference on Neural Information Processing Systems (NIPS 2016)*, pages 1–8, Barce-

lona, Spain. URL https://pdfs.semanticscholar.org/6ffd/e80e503fe6125237476494e777f4fe6d62c4.pdf.

Huang, J, Yao, S, Lyu, C, dan Ji, D, 2017, Multi-Granularity Neural Sentence Model for Measuring Short Text Similarity. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 10177 LNCS, pages 439–455. Springer Verlag. doi: 10.1007/978-3-319-55753-3_28. URL http://link.springer.com/10.1007/978-3-319-55753-3{_}28.

Instagram, 2020. *How do I filter out and hide comments I don't want to appear on my posts on Instagram? | Instagram Help Centre*. URL https://www.facebook.com/help/instagram/700284123459336. [Diakses pada: 2021-03-06].

Instagram, 2022. *Community Guidelines | Instagram Help Centre*. URL https://help.instagram.com/477434105621119. [Diakses pada: 2022-01-06].

InternetSociety, 2014. *What Is Spam | Internet Society*. URL https://www.internetsociety.org/resources/doc/2014/what-is-spam/. [Diakses pada: 2021-02-26].

Inuwa-Dutse, I, Liptrott, M, dan Korkontzelos, I, nov 2018, *Detection of spam-posting accounts on Twitter*. *Neurocomputing*, 315:496–511. ISSN 09252312. doi: 10.1016/j.neucom.2018.07.044. URL https://linkinghub.elsevier.com/retrieve/pii/S0925231218308798.

Jain, G, Sharma, M, dan Agarwal, B, 2019, *Spam detection in social media using convolutional and long short term memory neural network*. *Annals of Mathematics and Artificial Intelligence*, 85(1):21–44. ISSN 15737470. doi: 10.1007/s10472-018-9612-z. URL https://link.springer.com/article/10.1007/s10472-018-9612-z.

Jang, B, Kim, M, Harerimana, G, Kang, S. U, dan Kim, J. W, 2020, *Bi-LSTM model to increase accuracy in text classification: Combining word2vec CNN and attention mechanism*. *Applied Sciences (Switzerland)*, 10(17). ISSN 20763417. doi: 10.3390/app10175841.

Joe, S, may 2019. *Instagram Scrapper*. URL https://github.com/realsirjoe/instagram-scraper. [Diakses pada: 2022-05-14].

Kacmajor, M, Kelleher, J. D, Klubicka, F, dan Maldonado, A, feb 2020, Semantic Relatedness and Taxonomic Word Embeddings. In *arXiv*. URL http://arxiv.org/abs/2002.06235.

Kanodia, S, Sasheendran, R, dan Pathari, V, 2018, *A Novel Approach for Youtube Video Spam Detection using Markov Decision Process. 2018 International Conference on Advances in Computing, Communications and Informatics, ICACCI 2018*, pages 60–66. doi: 10.1109/ICACCI.2018.8554405.

Kaplan, A. M dan Haenlein, M, jan 2010, *Users of the world, unite! The challenges and opportunities of Social Media. Business Horizons*, 53(1):59–68. ISSN 00076813. doi: 10.1016/j.bushor.2009.09.003. URL https://www.sciencedirect.com/science/article/pii/S0007681309001232https://linkinghub.elsevier.com/retrieve/pii/S0007681309001232.

Kapoor, A, Gulli, A, dan Pal, S, 2022, *Deep Learning with TensorFlow and Keras*. Packt Publishing, Birmingham, 3rd edition. ISBN 978-1-80323-291-1.

Kemendikbud, 2020. *Hasil Pencarian - KBBI Daring*. URL https://kbbi.kemdikbud.go.id/entri/spam. [Diakses pada: 2021-02-26].

Khatun, A, Matin, M. H, Miah, A.-a, dan Miah, R, 2020, *Comparative Study on Text Classification. International Journal of Engineering Science Invention (IJESI)*, 9(9):21–33. doi: 10.35629/6734-0909012133.

Khawandi, S, Abdallah, F, dan Ismail, A, 2019, A Survey On Image Spam Detection Techniques. In *3rd International Conference on Computer Science and Information Technology (COMIT 2019)*, pages 13–27. doi: 10.5121/csit.2019.90102.

Khobragade, R dan Bhattacharyya, P, 2019. *Survey of Textual Entailment Approaches*. URL https://www.cfilt.iitb.ac.in/resources/surveys/rakesh{_}heaven{_}survey{_}2019.pdf. [Diakses pada: 2021-05-05].

Khodake, P. S, 2020, *Analysis and Detection of Spam Comments on Social Networking Platforms like YouTube using Machine Learning. International*

Journal for Research in Applied Science and Engineering Technology, 8(9): 1280–1282. doi: 10.22214/ijraset.2020.31677.

Khot, T, Sabharwal, A, dan Clark, P, apr 2018, *SciTaiL: A Textual Entailment Dataset from Science Question Answering. Proceedings of the AAAI Conference on Artificial Intelligence*, 32(1):5189–5197. ISSN 2374-3468. doi: 10.1609/aaai.v32i1.12022. URL https://ojs.aaai.org/index.php/AAAI/article/view/12022.

Kilgarriff, A dan Fellbaum, C, 2000, *WordNet: An Electronic Lexical Database. Language*, 76(3):706. ISSN 00978507. doi: 10.2307/417141.

Kim, J, Seo, D, Kim, H, dan Kang, P, 2017, *Facebook Spam Post Filtering based on Instagram-based Transfer Learning and Meta Information of Posts. Journal of the Korean Institute of Industrial Engineers*, 43(3):192–202. ISSN 1225-0988. doi: 10.7232/jkiie.2017.43.3.192.

Kingma, D. P dan Ba, J, 2017. *Adam: A Method for Stochastic Optimization.*

Krithiga, R dan Ilavarasan, D. R, 2020, *Machine learning techniques for spammer identification: State of the art and analysis. Journal of Critical Reviews*, 7(1):446–448. ISSN 23945125. doi: 10.31838/jcr.07.01.87.

Kusumawardani, R. P, Priansya, S, dan Atletiko, F. J, 2018, *Context-sensitive normalization of social media text in bahasa Indonesia based on neural word embeddings. Procedia Computer Science*, 144:105–117. ISSN 18770509. doi: 10.1016/j.procs.2018.10.510. URL https://doi.org/10.1016/j.procs.2018.10.510.

Ladha, L dan Deepa, T, 2011, *Feature Selection Methods And Algorithms. International Journal on Computer Science and Engineering*, 3(5):1787–1797. URL http://journals.indexcopernicus.com/abstract.php?icid=945099.

Lan, Z, Chen, M, Goodman, S, Gimpel, K, Sharma, P, dan Soricut, R, 2019, *Albert: A lite bert for self-supervised learning of language representations. arXiv*, pages 1–17. ISSN 23318422.

Lee, J dan Cheah, Y.-N, 2015, Semantic Relatedness Measure for Identifying Relevant Answers in Online Community Question Answering Services. In

9th International Conference on Information Technology in Asia, Kuching, Malaysia. Faculty of Computer Science and IT, Universiti Malaysia Sarawak (UNIMAS). URL https://ir.unimas.my/id/eprint/12080/.

Li, D, Wang, Y, Rzepka, R, dan Araki, K, 2020, Preliminary Analysis of Weibo Emoji Usage in Spam Analysis of Weibo Comments. In *LAU Spring 2020*, pages 24–26. URL http://arakilab.media.eng.hokudai.ac.jp/{~}araki/2019/2019-D-6.pdf.

Li, J, Chen, W, Gu, B, Fang, J, Li, Z, dan Zhao, L, 2019a, Measuring Semantic Relatedness with Knowledge Association Network. In Li G., Yang J., Gama J., Natwichai J., T. Y, editor, *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 11446 LNCS, pages 676–691. Springer International Publishing, Switcherland. ISBN 9783030185756. doi: 10.1007/978-3-030-18576-3_40. URL http://dx.doi.org/10.1007/978-3-030-18576-3{_}40http://link.springer.com/10.1007/978-3-030-18576-3{_}40.

Li, M, Wu, B, dan Wang, Y, 2019b, Comment Spam Detection via Effective Features Combination. In *IEEE International Conference on Communications*, Shanghai, China. IEEE. ISBN 9781538680889. doi: 10.1109/ICC.2019.8761340. URL https://ieeexplore.ieee.org/abstract/document/8761340.

Liang, H, Sun, X, Sun, Y, dan Gao, Y, 2017, *Text feature extraction based on deep learning: a review. Eurasip Journal on Wireless Communications and Networking*, 2017(1):1–12. ISSN 16871499. doi: 10.1186/s13638-017-0993-1.

Liu, C, Fang, F, Lin, X, Cai, T, Tan, X, Liu, J, dan Lu, X, dec 2021, *Improving sentiment analysis accuracy with emoji embedding. Journal of Safety Science and Resilience*, 2(4):246–252. ISSN 26664496. doi: 10.1016/j.jnlssr.2021.10.003. URL https://linkinghub.elsevier.com/retrieve/pii/S2666449621000529.

Liu, G dan Guo, J, 2019, *Bidirectional LSTM with attention mechanism and convolutional layer for text classification. Neurocomputing*, 337:325–338. ISSN 18728286. doi: 10.1016/j.neucom.2019.01.078. URL https://doi.org/10.1016/j.neucom.2019.01.078.

Liu, P, Qiu, X, dan Xuanjing, H, 2016a, *Recurrent neural network for text classification with multi-task learning. IJCAI International Joint Conference on Artificial Intelligence*, 2016-Janua:2873–2879. ISSN 10450823.

Liu, P, Qiu, X, Zhou, Y, Chen, J, dan Huang, X, 2016b, *Modelling interaction of sentence pair with coupled-LSTMs. EMNLP 2016 - Conference on Empirical Methods in Natural Language Processing, Proceedings*, pages 1703–1712. doi: 10.18653/v1/d16-1176.

Liu, Y, Ott, M, Goyal, N, Du, J, Joshi, M, Chen, D, Levy, O, Lewis, M, Zettlemoyer, L, dan Stoyanov, V, 2019, RoBERTa: A robustly optimized BERT pretraining approach. In *arXiv*. ArXiv. URL https://arxiv.org/abs/1907.11692.

Lukito, Y dan Chrismanto, A. R, nov 2017, Recurrent neural networks model for WiFi-based indoor positioning system. In *2017 International Conference on Smart Cities, Automation & Intelligent Computing Systems (ICON-SONICS)*, volume 2018-Janua, pages 121–125. IEEE. ISBN 978-1-5090-6278-2. doi: 10.1109/ICON-SONICS.2017.8267833. URL http://ieeexplore.ieee.org/document/8267833/.

Luong, T, Pham, H, dan Manning, C. D, 2015, Effective Approaches to Attention-based Neural Machine Translation. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 1412–1421, Stroudsburg, PA, USA. Association for Computational Linguistics. ISBN 9781941643327. doi: 10.18653/v1/D15-1166. URL http://aclweb.org/anthology/D15-1166.

Mahendra, R, Aji, A. F, Louvan, S, Rahman, F, dan Vania, C, 2021, *IndoNLI: A Natural Language Inference Dataset for Indonesian. EMNLP 2021 - 2021 Conference on Empirical Methods in Natural Language Processing, Proceedings*, pages 10511–10527. doi: 10.18653/v1/2021.emnlp-main.821.

Majumder, G, Pakray, P, Gelbukh, A, dan Pinto, D, dec 2016, *Semantic Textual Similarity Methods, Tools, and Applications: A Survey. Computación y Sistemas*, 20(4):647–665. ISSN 2007-9737. doi: 10.13053/cys-20-4-2506. URL http://www.cys.cic.ipn.mx/ojs/index.php/CyS/article/view/2506.

Mani, jan 2022. *NLP Profiler*. URL https://github.com/neomatrix369/nlp_profiler. [Diakses pada: 2022-05-15].

Manning, C. D, Raghavan, P, dan Schütze, H, 2008, *Introduction to Information Retrieval*. Cambridge University Press. ISBN 0521865719. URL https://nlp.stanford.edu/IR-book/.

Mansoor, M, ur Rehman, Z, Shaheen, M, Khan, M. A, dan Habib, M, dec 2020, *Deep Learning based Semantic Similarity Detection using Text Data*. *Information Technology And Control*, 49(4):495–510. ISSN 2335-884X. doi: 10.5755/j01.itc.49.4.27118. URL https://itc.ktu.lt/index.php/ITC/article/view/27118.

Mawardi, V, Susanto, N, dan Naga, D, apr 2018, *Spelling Correction for Text Documents in Bahasa Indonesia Using Finite State Automata and Levinshtein Distance Method*. *MATEC Web of Conferences*, 164:01047. ISSN 2261-236X. doi: 10.1051/matecconf/201816401047. URL https://www.matec-conferences.org/10.1051/matecconf/201816401047.

McCann, B, Bradbury, J, Xiong, C, dan Socher, R, jul 2017, *Learned in Translation: Contextualized Word Vectors*. *Advances in Neural Information Processing Systems*, 2017-Decem:6295–6306. ISSN 10495258. URL http://arxiv.org/abs/1708.00107.

McCord, M dan Chuah, M, 2011, 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)*, pages 175–186. Springer. ISBN 9783642234958. doi: 10.1007/978-3-642-23496-5_13. URL https://link.springer.com/chapter/10.1007{%}2F978-3-642-23496-5{_}13.

Mikolov, T, Chen, K, Corrado, G, dan Dean, J, 2013, *Distributed Representations of Words and Phrases and their Compositionality*, Technical report, Google.

Mishne, G, Carmel, D, dan Lempel, R, 2005, Blocking blog spam with language model disagreement. In *Proceedings of the 1st International Workshop on Adversarial Information Retrieval on the Web, AIRWeb 2005 - Held in*

Conjunction with the 14th International World Wide Web Conference, pages 1–6, Chiba, Japan. AIRWeb 2005.

Mishra, P, 2019, *Correlated Feature Selection for Tweet Spam Classification using Artificial Neural Networks. arXiv*. URL https://arxiv.org/abs/1911.05495.

Mishra, P, Kaushik, S, dan Dey, K, 2021, *Bi-ISCA: Bidirectional Inter-Sentence Contextual Attention Mechanism for Detecting Sarcasm in User Generated Noisy Short Text. CEUR Workshop Proceedings*, 2995:1–8. ISSN 16130073.

Mishra, U, 2021. *Binary and Multiclass Classification in Machine Learning*. URL https://www.analyticssteps.com/blogs/binary-and-multiclass-classification-machine-learning. [Diakses pada: 2021-09-02].

Mohsin, M, 2021. *10 Facebook Statistics Every Marketer Should Know in 2021 [Infographic]*. URL https://www.oberlo.com/blog/facebook-statistics. [Diakses pada: 2021-05-11].

Mostafavi, M, Varnosfaderani, M. P, Nikseresht, F, dan Mansouri, S. A, 2022. *emojiSpace: Spatial Representation of Emojis*.

Mozilla Developer Network, 2021. *Anatomy of an extension*. URL https://developer.mozilla.org/en-US/docs/Mozilla/Add-ons/WebExtensions/Anatomy{_}of{_}a{_}WebExtension. [Diakses pada: 2021-09-01].

Mus, C, 2015. *10+ Akun Instagram Dengan Followers Terbanyak Di Indonesia*. URL http://www.musdeoranje.net/2016/08/akun-instagram-dengan-followers-terbanyak-di-indonesia.html. [Diakses pada: 2021-10-13].

Musyarofah, R. L, Utami, E. U, dan Raharjo, S. R, 2020, *Analisis Komentar Potensial pada Social Commerce Instagram Menggunakan TF-IDF. Jurnal Eksplora Informatika*, 9(2):130–139. ISSN 2089-1814. doi: 10.30864/eksplora.v9i2.360.

Olah, C, 2015. *Understanding LSTM Networks*. URL https://colah.github.io/posts/2015-08-Understanding-LSTMs/. [Diakses pada: 2020-03-17].

O'Reilly, T, 2005. *What Is Web 2.0 - O'Reilly Media*. URL https://www.oreilly.com/pub/a/web2/archive/what-is-web-20.html. [Diakses pada: 2021-02-26].

Parikh, A, Täckström, O, Das, D, dan Uszkoreit, J, 2016, A Decomposable Attention Model for Natural Language Inference. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 2249–2255, Stroudsburg, PA, USA. Association for Computational Linguistics. ISBN 9781945626258. doi: 10.18653/v1/D16-1244. URL http://aclweb.org/anthology/D16-1244.

Patil, M. G, Galande, M. V, Kekan, V, dan Dange, M. K, 2014, *Sentiment Analysis Using Support Vector Machine. International Journal of Innovative Research in Computer and Communication Engineering*, 2(1):2607–2612. doi: 10.1109/I4CT.2014.6914200.

Pedregosa, F, Varoquaux, G, Gramfort, A, Michel, V, Thirion, B, Grisel, O, Blondel, M, Prettenhofer, P, Weiss, R, Dubourg, V, Vanderplas, J, Passos, A, Cournapeau, D, Brucher, M, Perrot, M, dan Duchesnay, E, 2011, *Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research*, 12:2825–2830.

Pei, J, Wu, Y, Qin, Z, Cong, Y, dan Guan, J, 2021, Attention-based model for predicting question relatedness on Stack Overflow. In *International Conference on Mining Software Repositories (MSR) 2021*, pages 97–107. IEEE. ISBN 9781728187105. doi: 10.1109/msr52588.2021.00023. URL http://arxiv.org/abs/2103.10763.

Pennington, J, Socher, R, dan Manning, C. D, 2014, GloVe: Global Vectors for Word Representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1532–1543, Doha, Qatar. Association for Computational Linguistics. doi: 10.3115/v1/D14-1162.

Perrin, A dan Anderson, M, 2019. *Social media usage in the U.S. in 2019 | Pew Research Center*. URL https://www.pewresearch.org/fact-tank/2019/04/10/share-of-u-s-adults-using-social-media-including-facebook-is-mostly-unchanged-since-2018/. [Diakses pada: 2021-04-26].

Peters, M. E, Neumann, M, Gardner, M, Clark, C, Lee, K, dan Zettlemoyer, L, 2018, Deep contextualized word representations. In *Proceedings ofNAACL-HLT 2018*, pages 2227–2237, New Orleans, Louisiana. Association for Computational Linguistics. URL http://allennlp.org/elmo.

Prabowo, F dan Purwarianti, A, 2018, Instagram online shop's comment classification using statistical approach. In *Proceedings - 2017 2nd International Conferences on Information Technology, Information Systems and Electrical Engineering, ICITISEE 2017*, pages 282–287, Yogyakarta. IEEE. ISBN 9781538606582. doi: 10.1109/ICITISEE.2017.8285512. URL https://ieeexplore.ieee.org/document/8285512.

Prince, S, 2023, *Understanding Deep Learning*. MIT Press, 1st edition. URL https://udlbook.github.io/udlbook/.

Priyoko, B dan Yaqin, A, 2019, *Implementation of naive bayes algorithm for spam comments classification on Instagram. 2019 International Conference on Information and Communications Technology, ICOIACT 2019*, pages 508–513. doi: 10.1109/ICOIACT46704.2019.8938575.

Qiao, C, Huang, B, Niu, G, Li, D, Dong, D, He, W, Yu, D, dan Wu, H, 2018, A new method of region embedding for text classification. In *6th International Conference on Learning Representations, ICLR 2018 - Conference Track Proceedings*, pages 1–12, Vancouver, CANADA. ICLR 2018.

Rachmat, A dan Lukito, Y, 2017, *Deteksi Komentar Spam Bahasa Indonesia Pada Instagram Menggunakan Naive Bayes. Jurnal ULTIMATICS*, 9(1):50–58. ISSN 2085-4552. doi: 10.31937/ti.v9i1.564.

Raffel, C, Shazeer, N, Roberts, A, Lee, K, Narang, S, Matena, M, Zhou, Y, Peter, W. L, dan Liu, J, 2019, *Exploring the limits of transfer learning with a unified text-to-text transformer. arXiv*, 21:1–67.

Ramaprabha, J, Das, S, dan Mukerjee, P, apr 2018, *Survey on Sentence Similarity Evaluation using Deep Learning. Journal of Physics: Conference Series*, 1000(1):012070. ISSN 1742-6588. doi: 10.1088/1742-6596/1000/1/012070. URL https://iopscience.iop.org/article/10.1088/1742-6596/1000/1/012070.

Reelfs, J. H, Hohlfeld, O, Strohmaier, M, dan Henckell, N, may 2020, Word-Emoji Embeddings from large scale Messaging Data reflect real-world Semantic Associations of Expressive Icons. In *3rd International Workshop on Emoji Understanding and Applications in Social Media*. ArXiv. URL http://arxiv.org/abs/2006.01207.

Reimers, N dan Gurevych, I, 2019, *Sentence-BERT: Sentence embeddings using Siamese BERT-networks*. *arXiv*, pages 3982–3992. ISSN 23318422.

Rennie, J. D, Shih, L, Teevan, J, dan Karger, D, 2003, *Tackling the Poor Assumptions of Naive Bayes Text Classifiers*. *Proceedings, Twentieth International Conference on Machine Learning*, 2(1973):616–623.

Rothman, D, 2021, *Transformers for Natural Language Processing Build innovative deep neural network architectures for NLP & Python, PyTorch, TensorFlow, BERT, RoBERTa & more*. Packt Publishing, 1st edition. ISBN 9781800565791. URL https://www.packtpub.com/product/transformers-for-natural-language-processing/9781800565791.

Rousset, D, 2017. *Creating One Browser Extension For All Browsers: Edge, Chrome, Firefox, Opera, Brave And Vivaldi*. URL https://www.smashingmagazine.com/2017/04/browser-extension-edge-chrome-firefox-opera-brave-vivaldi/. [Diakses pada: 2021-09-01].

Roy, P. K, Singh, J. P, dan Banerjee, S, 2020, *Deep learning to filter SMS Spam*. *Future Generation Computer Systems*, 102:524–533. ISSN 0167739X. doi: 10.1016/j.future.2019.09.001. URL https://doi.org/10.1016/j.future.2019.09.001.

Ruskanda, F. Z, 2019, *Study on the Effect of Preprocessing Methods for Spam Email Detection*. *Indonesian Journal on Computing (Indo-JC)*, 4(1):109. ISSN 2460-9234. doi: 10.21108/indojc.2019.4.1.284.

Sahoo, S. R dan Gupta, B. B, 2021, *Multiple features based approach for automatic fake news detection on social networks using deep learning*. *Applied Soft Computing*, 100:106983. ISSN 15684946. doi: 10.1016/j.asoc.2020.106983. URL https://doi.org/10.1016/j.asoc.2020.106983.

Saidani, N, Adi, K, dan Allili, M. S, 2020, *A semantic-based classification approach for an enhanced spam detection. Computers and Security*, 94: 101716. ISSN 01674048. doi: 10.1016/j.cose.2020.101716. URL https://doi.org/10.1016/j.cose.2020.101716.

Samsudin, N. M, Mohd Foozy, C. F. B, Alias, N, Shamala, P, Othman, N. F, dan Wan Din, W. I. S, 2019, *Youtube spam detection framework using naïve bayes and logistic regression. Indonesian Journal of Electrical Engineering and Computer Science*, 14(3):1508–1517. ISSN 25024760. doi: 10.11591/ijeecs.v14.i3.pp1508-1517.

Sanborn, A dan Skryzalin, J, 2015, *Deep Learning for Semantic Similarity. CS224d: Deep Learning for Natural Language Processing Stanford, CA, USA: Stanford University*, pages 1–7.

Sanh, V, Debut, L, Chaumond, J, dan Wolf, T, 2019, *DistilBERT, a distilled version of BERT: Smaller, faster, cheaper and lighter. arXiv*, pages 2–6. ISSN 23318422.

Saumya, S dan Singh, J. P, 2020, *Spam review detection using LSTM autoencoder: an unsupervised approach. Electronic Commerce Research*, 20 (0123456789). ISSN 15729362. doi: 10.1007/s10660-020-09413-4. URL https://doi.org/10.1007/s10660-020-09413-4.

Sawers, P, 2010. *The origin of the word 'spam' | THE GOOD WORD*. URL http://www.thegoodword.co.uk/2010/09/20/the-origin-of-the-word-spam/. [Diakses pada: 2021-02-26].

Septiandri, A. A dan Wibisono, O, 2017, *Detecting spam comments on Indonesia's Instagram posts. Journal of Physics: Conference Series*, 801(012069): 1–7. ISSN 1742-6588. doi: 10.1088/1742-6596/755/1/011001. URL https://iopscience.iop.org/article/10.1088/1742-6596/755/1/011001.

Setya, K. N dan Mahendra, R, 2023a, Semi-supervised Textual Entailment on Indonesian Wikipedia Data. In Gelbukh, A, editor, *Computational Linguistics and Intelligent Text Processing - 19th International Conference, CICLing 2018, Revised Selected Paper*, pages 416–427. Springer Science and Business Media Deutschland GmbH, Hanoi, Viet Nam. doi: 10.1007/978-3-

031-23793-5_34. URL https://link.springer.com/10.1007/978-3-031-23793-5_34.

Setya, K. N dan Mahendra, R, 2023b, Semi-supervised Textual Entailment on Indonesian Wikipedia Data. In Gelbukh, A, editor, *Computational Linguistics and Intelligent Text Processing - 19th International Conference, CICLing 2018, Revised Selected Paper*, pages 416–427. Springer Science and Business Media Deutschland GmbH, Hanoi, Viet Nam. doi: 10.1007/978-3-031-23793-5_34. URL https://link.springer.com/10.1007/978-3-031-23793-5_34.

Severyn, A dan Moschitti, A, aug 2015, Learning to Rank Short Text Pairs with Convolutional Deep Neural Networks. In *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 373–382, New York, NY, USA. ACM. ISBN 9781450336215. doi: 10.1145/2766462.2767738. URL https://dl.acm.org/doi/10.1145/2766462.2767738.

Shahariar, G. M, Biswas, S, Omar, F, Shah, F. M, dan Binte Hassan, S, 2019, *Spam Review Detection Using Deep Learning. 2019 IEEE 10th Annual Information Technology, Electronics and Mobile Communication Conference, IEMCON 2019*, pages 27–33. doi: 10.1109/IEMCON.2019.8936148.

Singh, A, Blanco, E, dan Jin, W, 2019, Incorporating emoji descriptions improves tweet classification. In *NAACL HLT 2019 - 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference*, volume 1, pages 2096–2101. Association for Computational Linguistics. ISBN 9781950737130. doi: 10.18653/v1/n19-1214. URL https://aclanthology.org/N19-1214/.

Singh, P dan Manure, A, 2020, *Learn TensorFlow 2.0*. Apress, Berkeley, CA, 1st edition. ISBN 978-1-4842-5560-5. doi: 10.1007/978-1-4842-5558-2. URL https://www.apress.com/gp/book/9781484255605http://link.springer.com/10.1007/978-1-4842-5558-2.

Srinivasan, S, Ravi, V, Sowmya, V, Krichen, M, Ben Noureddine, D, Anivilla, S, dan Soman, K. P, 2020, *Deep Convolutional Neural Network Based Image Spam Classification. Proceedings - 2020 6th Conference on Data Science and*

Machine Learning Applications, CDMA 2020, pages 112–117. doi: 10.1109/CDMA47397.2020.00025.

StatCounter, 2020. *Social Media Stats Indonesia | StatCounter Global Stats*. URL https://gs.statcounter.com/social-media-stats/all/indonesia. [Diakses pada: 2021-04-26].

Statista, 2020. • *Indonesia: social media platforms used daily 2020 | Statista*. URL https://www.statista.com/statistics/1168819/indonesia-social-media-platforms-used-daily/. [Diakses pada: 2021-04-26].

Sun, X dan Lu, W, 2020, Understanding Attention for Text Classification. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 3418–3428, Stroudsburg, PA, USA. Association for Computational Linguistics. doi: 10.18653/v1/2020.acl-main.312. URL https://www.aclweb.org/anthology/2020.acl-main.312.

Tehlan, P, Madaan, R, dan Bhatia, K. K, 2019, *A spam detection mechamism in social media using soft computing. Proceedings of the 2019 6th International Conference on Computing for Sustainable Global Development, INDIACom 2019*, pages 950–955.

Tian, W, Li, J, dan Li, H, 2018, *A Method of Feature Selection Based on Word2Vec in Text Categorization. Chinese Control Conference, CCC*, 2018-July: 9452–9455. ISSN 21612927. doi: 10.23919/ChiCC.2018.8483345.

Vaswani, A, Shazeer, N, Parmar, N, Uszkoreit, J, Jones, L, Gomez, A. N, Kaiser, Ł, dan Polosukhin, I, 2017, *Attention is all you need. Advances in Neural Information Processing Systems*, 2017-Decem(Nips):5999–6009. ISSN 10495258.

Venkatesh, B dan Anuradha, J, 2019, *A review of Feature Selection and its methods. Cybernetics and Information Technologies*, 19(1):3–26. ISSN 13144081. doi: 10.2478/CAIT-2019-0001.

Vennila, G dan Manikandan, M. S, 2017, *Detection of Human and Computer Voice Spammers Using Hidden Markov Model in Voice over Internet Protocol Network. Procedia Computer Science*, 115:588–595. ISSN 18770509. doi: 10.1016/j.procs.2017.09.169. URL https://doi.org/10.1016/j.procs.2017.09.169.

Verma, V. K, Ranjan, M, dan Mishra, P, jan 2015, Text mining and information professionals: Role, issues and challenges. In *2015 4th International Symposium on Emerging Trends and Technologies in Libraries and Information Services*, pages 133–137, Noida, India. IEEE. ISBN 978-1-4799-7999-8. doi: 10.1109/ETTLIS.2015.7048186. URL http://ieeexplore.ieee.org/document/7048186/.

Wang, L, Zhang, L, dan Jiang, J, 2020a, *Duplicate Question Detection With Deep Learning in Stack Overflow. IEEE Access*, 8:25964–25975. ISSN 2169-3536. doi: 10.1109/ACCESS.2020.2968391. URL https://ieeexplore.ieee.org/document/8964380/.

Wang, L, Zhang, L, dan Jiang, J, 2020b, *Duplicate Question Detection With Deep Learning in Stack Overflow. IEEE Access*, 8:25964–25975. ISSN 2169-3536. doi: 10.1109/ACCESS.2020.2968391. URL https://ieeexplore.ieee.org/document/8964380/.

Wang, M, Yang, H, Qin, Y, Sun, S, dan Deng, Y, 2020c, *Unified Humor Detection Based on Sentence-pair Augmentation and Transfer Learning. Proceedings of the 22nd Annual Conference of the European Association for Machine Translation*, pages 53–59. URL https://www.aclweb.org/anthology/2020.eamt-1.7.

Wang, Y, Liu, B, Wu, H, Zhao, S, Cai, Z, Li, D, dan Fong, C. C, 2020d, *An opinion spam detection method based on multi-filters convolutional neural network. Computers, Materials and Continua*, 65(1):355–367. ISSN 15462226. doi: 10.32604/cmc.2020.09835.

Wang, Z, Mi, H, dan Ittycheriah, A, feb 2016, *Sentence Similarity Learning by Lexical Decomposition and Composition. COLING 2016 - 26th International Conference on Computational Linguistics, Proceedings of COLING 2016: Technical Papers*, pages 1340–1349. URL http://arxiv.org/abs/1602.07019.

Weiss, S. M, Indurkhya, N, Zhang, T, dan Damerau, F. J, 2005, *Text mining: Predictive methods for analyzing unstructured information.* Springer New York. ISBN 0387954333. doi: 10.1007/978-0-387-34555-0.

Wicaksono, I, 2020, *Sistem identifikasi komentar negatif pada instagram*, Technical report, Universitas Muhammadiyah Surakarta, Surakarta. URL http://eprints.ums.ac.id/82841/7/NaskahPublikasi-UploadPerpus.pdf.

Wilie, B, Vincentio, K, Winata, G. I, Cahyawijaya, S, Li, X, Lim, Z. Y, Soleman, S, Mahendra, R, Fung, P, Bahar, S, dan Purwarianti, A, 2020, *IndoNLU: Benchmark and resources for evaluating indonesian natural language understanding. arXiv*. ISSN 23318422.

Wongso, R, Luwinda, F. A, Trisnajaya, B. C, Rusli, O, dan Rudy, 2017, *News Article Text Classification in Indonesian Language. Procedia Computer Science*, 116:137–143. ISSN 18770509. doi: 10.1016/j.procs.2017.10.039. URL https://doi.org/10.1016/j.procs.2017.10.039.

Wu, T, Wen, S, Xiang, Y, dan Zhou, W, jul 2018, *Twitter spam detection: Survey of new approaches and comparative study. Computers & Security*, 76:265–284. ISSN 01674048. doi: 10.1016/j.cose.2017.11.013. URL https://linkinghub.elsevier.com/retrieve/pii/S016740481730250X.

Yang, R, Zhang, J, Gao, X, Ji, F, dan Chen, H, 2020, *Simple and effective text matching with richer alignment features. ACL 2019 - 57th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference*, pages 4699–4709. doi: 10.18653/v1/p19-1465.

Yang, Z, Dai, Z, Yang, Y, Carbonell, J, Salakhutdinov, R, dan Le, Q. V, 2019, XLNet: Generalized autoregressive pretraining for language understanding. In Wallach, H, Larochelle, H, Beygelzimer, A, D'Alché-Buc, F, Fox, E, dan Garnett, R, editors, *arXiv*, pages 1–18, Vancouver, Canada. NeurIPS Proceedings. URL https://papers.nips.cc/paper/2019/hash/dc6a7e655d7e5840e66733e9ee67cc69-Abstract.html.

Yin, W, Schütze, H, Xiang, B, dan Zhou, B, dec 2016, *ABCNN: Attention-Based Convolutional Neural Network for Modeling Sentence Pairs. Transactions of the Association for Computational Linguistics*, 4:566–567. ISSN 2307-387X. doi: 10.1162/tacl_a_00244. URL https://direct.mit.edu/tacl/article/43468.

YouTube, 2022. *Kebijakan mengenai spam, praktik penipuan, dan scam - Bantuan YouTube*. URL https://support.google.com/youtube/answer/2801973?hl=id{#}zippy={%}2Cspam-komentar. [Diakses pada: 2022-01-06].

Yu, S, Su, J, dan Luo, D, 2019, *Improving BERT-Based Text Classification with Auxiliary Sentence and Domain Knowledge*. *IEEE Access*, 7:176600–176612. ISSN 21693536. doi: 10.1109/ACCESS.2019.2953990.

Yuan, Z dan Jun, S, oct 2020, Siamese Network cooperating with Multi-head Attention for semantic sentence matching. In *2020 19th International Symposium on Distributed Computing and Applications for Business Engineering and Science (DCABES)*, pages 215–218. IEEE. ISBN 978-1-7281-9724-1. doi: 10.1109/DCABES50732.2020.00068. URL https://ieeexplore.ieee.org/document/9277782/.

Yüksel, A. E, Türkmen, Y. A, Özgür, A, dan Altınel, A. B, 2019, *Turkish tweet classification with transformer encoder*. *International Conference Recent Advances in Natural Language Processing, RANLP*, 2019-Septe:1380–1387. ISSN 13138502. doi: 10.26615/978-954-452-056-4_158.

Yusof, Y dan Sadoon, O. H, 2017, Detecting Video Spammers in Youtube Social Media. In *ICOCI Kuala Lumpur. Universiti Utara Malaysia*, pages 228–234, Kuala Lumpur. Universiti Utara Malaysia. URL http://www.uum.edu.my.

Zaheer, M, Guruganesh, G, Dubey, A, Ainslie, J, Alberti, C, Ontanon, S, Pham, P, Ravula, A, Wang, Q, Yang, L, dan Ahmed, A, 2020, Big Bird: Transformers for Longer Sequences. In *arXiv*, Vancouver, CANADA. NeurIPS Proceedings. URL https://proceedings.neurips.cc/paper/2020/file/c8512d142a2d849725f31a9a7a361ab9-Paper.pdf.

Zainab, A, Syed, D, dan Al-Thani, D, 2019, Deployment of deep learning models to mobile devices for spam classification. In *Proceedings - 2019 IEEE 1st International Conference on Cognitive Machine Intelligence, CogMI 2019*, pages 112–117, Los Angeles, CA, USA. IEEE. ISBN 9781728167374. doi: 10.1109/CogMI48466.2019.00024. URL https://ieeexplore.ieee.org/document/8998978.

Zhang, C, Liu, C, Zhang, X, dan Almpanidis, G, 2017, *An up-to-date comparison of state-of-the-art classification algorithms*. *Expert Systems with Applications*,

82:128–150. ISSN 09574174. doi: 10.1016/j.eswa.2017.04.003. URL http://dx.doi.org/10.1016/j.eswa.2017.04.003.

Zhang, L dan Moldovan, D, 2018, *Rule-based vs. Neural Net Approaches to Semantic Textual Similarity. Proceedings of the First Workshop on Linguistic Resources for Natural Language Processing*, pages 12–17. URL https://www.aclweb.org/anthology/W18-3803.

Zhang, W dan Sun, H.-M, jan 2017, Instagram Spam Detection. In *2017 IEEE 22nd Pacific Rim International Symposium on Dependable Computing (PRDC)*, pages 227–228. IEEE. ISBN 978-1-5090-5652-1. doi: 10.1109/PRDC.2017.43. URL http://ieeexplore.ieee.org/document/7920623/.

Zhang, X, Zhao, J. Z, dan LeCun, Y, 2015a, Character-level Convolutional Networks for TextClassification. In *NIPS'15: Proceedings of the 28th International Conference on Neural Information Processing Systems*, pages 649–657. URL https://dl.acm.org/doi/10.5555/2969239.2969312.

Zhang, Y, Chen, M, dan Liu, L, 2015b, A review on text mining. In *Proceedings of the IEEE International Conference on Software Engineering and Service Sciences, ICSESS*, volume 2015-November, pages 681–685, Beijing, China. IEEE. ISBN 9781479983520. doi: 10.1109/ICSESS.2015.7339149. URL https://ieeexplore.ieee.org/document/7339149.

Zhao, J, Lan, M, Niu, Z. Y, dan Lu, Y, 2015, Integrating word embeddings and traditional NLP features to measure textual entailment and semantic relatedness of sentence pairs. In *Proceedings of the International Joint Conference on Neural Networks*, Killarney, Ireland. IEEE. ISBN 9781479919604. doi: 10.1109/IJCNN.2015.7280462. URL https://ieeexplore.ieee.org/document/7280462.

Zhao, J, Zhu, T, dan Lan, M, 2014, ECNU: One Stone Two Birds: Ensemble of Heterogenous Measures for Semantic Relatedness and Textual Entailment. In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, pages 271–277, Stroudsburg, PA, USA. Association for Computational Linguistics. doi: 10.3115/v1/S14-2044. URL http://aclweb.org/anthology/S14-2044.