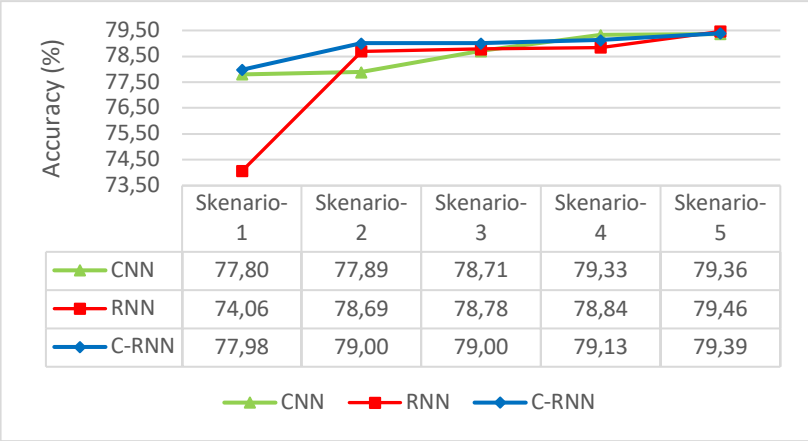

	Skenario-1	Skenario-2	Skenario-3	Skenario-4	Skenario-5
CNN	77,80	77,89	78,71	79,33	79,36
RNN	74,06	78,69	78,78	78,84	79,46
C-RNN	77,98	79,00	79,00	79,13	79,39

Gambar 6. Perkembangan akurasi terhadap setiap skenario.

Berdasarkan hasil pengujian yang dilakukan pada Gambar 6, pengujian yang dilakukan menggunakan metode CNN, RNN dan C-RNN dengan menerapkan kombinasi fitur perluasan FastText dan ekstraksi fitur TF-IDF, terdapat peningkatan akurasi dengan menggabungkan ekstraksi fitur TF-IDF dengan fitur perluasan FastText. Akurasi tertinggi dicapai oleh model RNN pada fitur 1 teratas dengan menggunakan korpus IndoNews yaitu sebesar 5.39% dibandingkan dengan baseline, kemudian model CNN mengalami peningkatan dengan menggunakan korpus IndoNews sebesar 1.56% pada fitur 5 teratas, dan model C-RNN mengalami peningkatan sebesar 1.41% pada korpus IndoNews pada fitur 1 teratas.

4. Kesimpulan

Penelitian ini menggunakan 29.085 tweet dalam Bahasa Indonesia dengan pola linguistik terbuka untuk mendeteksi *cyberbullying*. Dalam klasifikasi, CNN, RNN, dan C-RNN digunakan. Data dikumpulkan menggunakan perayapan data berbasis bahasa pemrograman Python dan diberi label secara manual. TF-IDF digunakan sebagai ekstraksi fitur yang mengekstraksi informasi dan mengubahnya menjadi format vector sehingga dapat diproses oleh model *deep learning*. Kombinasi N-Gram digunakan untuk ekstraksi TF-IDF, dengan memberikan ketepatan klasifikasi yang tinggi. FastText digunakan untuk membangun korpus kemiripan untuk perluasan fitur.

Hasil akurasi dari model hybrid deep learning C-RNN belum berhasil melampaui nilai akurasi dari model RNN, model deep learning RNN lebih unggul dalam hal akurasi dibandingan dengan model lainnya. Jika Dilihat dari jenis N-gram dan jumlah maksimum fitur yang digunakan sebagai parameter TF-IDF, maka Allgram (*Unigram*, *Bigram*, *Trigram*) dan jumlah maksimum fitur sebesar 1000, berpengaruh baik terhadap sistem deteksi *cyberbullying*.

Penggunaan fitur ekspansi telah terbukti meningkatkan akurasi pada setiap model yang digunakan pada penelitian ini. Hasil dari akurasi CNN adalah 79,36% meningkat 1,56% dari baseline, sedangkan akurasi RNN adalah 79,46% meningkat 5,40% dari baseline. Kemudian, untuk model C-RNN mendapatkan peningkatan akurasi dari baseline sebesar 1,41% hingga mencapai akurasi 79,39%.

Dari hasil penelitian ini, pendekatan ekspansi fitur menunjukkan efektivitas yang signifikan dalam mendeteksi cyberbullying pada media social Twitter. Meskipun pendekatan ini telah menunjukkan hasil yang menjanjikan, masih ada ruang untuk peningkatan dan optimasi.

Daftar Pustaka

[1] Asosiasi Penyelenggara Jasa Internet Indonesia, “Profil Internet Indonesia 2022,” pp. 1–104, 2022. Accessed: Nov. 27, 2022. [Online]. Available: https://apjii.or.id/survei

[2] S. Kemp, “DIGITAL 2022: INDONESIA.” https://datareportal.com/reports/digital-2022-indonesia (accessed Nov. 27, 2022).

[3] “Explained: What is Twitter?” https://www.webwise.ie/parents/explained-what-is-twitter-2/ (accessed Nov. 28, 2022).

[4] Unicef, “Cyberbullying: What is it and how to stop it.” https://www.unicef.org/end-violence/how-to-stop-cyberbullying (accessed Nov. 28, 2022).

[5] Z. Jianqiang, G. Xiaolin, and Z. Xuejun, “Deep Convolution Neural Networks for Twitter Sentiment Analysis,” IEEE Access, vol. 6, pp. 23253–23260, 2018, doi: 10.1109/ACCESS.2017.2776930.

[6] D. H. Jayani, “Survei APJII: 49% Pengguna Internet Pernah Dirisak di Medsos.” https://databoks.katadata.co.id/datapublish/2019/05/16/survei-apjii-49-pengguna-internet-pernah-dirisak-di-medsos/ (accessed Nov. 28, 2022).

[7] T. A. Buan and R. Ramachandra, "Automated Cyberbullying Detection in Social Media Using an SVM Activated Stacked Convolution LSTM Network," in *ACM International Conference Proceeding Series*, Association for Computing Machinery, Mar. 2020, pp. 170–174. doi: 10.1145/3388142.3388147.

[8] M. Zulqarnain, R. Ghazali, M. Ghulam Ghouse, and M. Faheem Mushtaq, "Efficient Processing of GRU Based on Word Embedding for Text Classification," 2019.

[9] W. K. Sari, D. P. Rini, and R. F. Malik, "Text Classification Using Long Short-Term Memory With GloVe Features," *Jurnal Ilmiah Teknik Elektro Komputer dan Informatika*, vol. 5, no. 2, p. 85, Feb. 2020, doi: 10.26555/jiteki.v5i2.15021.

[10] Y. Q. Yuan, "Research on Text Classification Algorithm Based on BiLSTM-WSAttention," in *IEEE Advanced Information Technology, Electronic and Automation Control Conference (IAEAC)*, 2021, pp. 2235–2239. doi: 10.1109/IAEAC50856.2021.9390982.

[11] H. Wang, K. Tian, Z. Wu, and L. Wang, "A short text classification method based on convolutional neural network and semantic extension," *International Journal of Computational Intelligence Systems*, vol. 14, no. 1, pp. 367–375, 2021, doi: 10.2991/ijcis.d.201207.001.

[12] S. T. Laxmi, R. Rismala, and H. Nurrahmi, "Cyberbullying Detection on Indonesian Twitterusing Doc2Vec and Convolutional Neural Network," in *9th International Conference on Information and Communication Technology (ICoICT)*, 2021, pp. 82–86. doi: 10.1109/ICoICT52021.2021.9527420.

[13] H. Nurrahmi and D. Nurjanah, "Indonesian Twitter Cyberbullying Detection using Text Classification and User Credibility," in *International Conference on Information and Communications Technology (ICOIACT)978-1-5386-0954-5/18/$31.00 ©2018 IEEE543*, 2018, pp. 543–548. doi: 10.1109/ICOIACT.2018.8350758.

[14] M. Andriansyah *et al.*, "Cyberbullying Comment Classification on Indonesian Selebgram Using Support Vector Machine Method," in *Second International Conference on Informatics and Computing (ICIC)*, 2017, pp. 1–5. doi: 10.1109/IAC.2017.8280617.

[15] G. Sarna and M. P. S. Bhatia, "Content based approach to find the credibility of user in social networks: an application of cyberbullying," *International Journal of Machine Learning and Cybernetics*, vol. 8, no. 2, pp. 677–689, Apr. 2017, doi: 10.1007/s13042-015-0463-1.

[16] E. B. Setiawan, D. H. Widyantoro, and K. Surendro, "Feature expansion using word embedding for tweet topic classification," in *Proceeding of 2016 10th International Conference on Telecommunication Systems Services and Applications, TSSA 2016: Special Issue in Radar Technology*, Institute of Electrical and Electronics Engineers Inc., Mar. 2017. doi: 10.1109/TSSA.2016.7871085.

[17] I. Kaibi, E. H. Nfaoui, and H. Satori, "A Comparative Evaluation of Word Embeddings Techniques for Twitter Sentiment Analysis," in *International Conference on Wireless Technologies, Embedded and Intelligent Systems (WITS)*, IEEE, 2019, pp. 1–4. doi: 10.1109/WITS.2019.8723864.

[18] W. Bourequat and H. Mourad, "Sentiment Analysis Approach for Analyzing iPhone Release using Support Vector Machine," *International Journal of Advances in Data and Information Systems*, vol. 2, no. 1, pp. 36–44, Apr. 2021, doi: 10.25008/ijadis.v2i1.1216.

[19] S. Gharatkar, A. Ingle, T. Naik, and A. Save, "Review Preprocessing Using Data Cleaning And Stemming Technique," in *International Conference on Innovations in Information,Embedded and Communication Systems (ICIIECS)*, 2017, pp. 1–4.

[20] S. Akuma, T. Lubem, and I. T. Adom, "Comparing Bag of Words and TF-IDF with different models for hate speech detection from live tweets," *International Journal of Information Technology (Singapore)*, 2022, doi: 10.1007/s41870-022-01096-4.

[21] S. W. Kim and J. M. Gil, "Research paper classification systems based on TF-IDF and LDA schemes," *Human-centric Computing and Information Sciences*, vol. 9, no. 1, Dec. 2019, doi: 10.1186/s13673-019-0192-7.

[22] P. Bojanowski, E. Grave, A. Joulin, and T. Mikolov, "Enriching Word Vectors with Subword Information," Jul. 2016, [Online]. Available: http://arxiv.org/abs/1607.04606

[23] O. Bayat, S. Aljawarneh, H. F. Carlak, International Association of Researchers, Institute of Electrical and Electronics Engineers, and Akdeniz Üniversitesi, *Proceedings of 2017 International Conference on Engineering & Technology (ICET'2017) : Akdeniz University, Antalya, Turkey, 21-23 August, 2017*.

[24] H. Khotimah, F. #1, E. Budi, S. #2, and I. Kurniawan, "Implementation Information Gain Feature Selection for Hoax News Detection on Twitter using Convolutional Neural Network (CNN)", doi: 10.34818/indojc.2021.5.3.506.

[25] T. Tom´ and T. Mikolov, "Overview Introduction Model description ASR Results Extensions MT Results Comparison Main outcomes Future work Recurrent neural network based language model," 2010.

[26] H. A. A. El Aal, S. A. Taie, and N. El-Bendary, "An optimized rnn-lstm approach for parkinson's disease early detection using speech features," *Bulletin of Electrical Engineering and Informatics*, vol. 10, no. 5, pp. 2503–2512, Oct. 2021, doi: 10.11591/eei.v10i5.3128.

[27] S. Afaq and S. Rao, “Significance Of Epochs On Training A Neural Network,” *International Journal of Scientific & Technology Research*, vol. 9, no. 6, 2020, [Online]. Available: www.ijstr.org