SCIENCE AND INFORMATION TECHNOLOGY
SINTECH
JOURNAL

SINTECH JOURNAL | ISSN 2598-7305 | E-ISSN 2598-9642
Vol. 8 No 1 – April 2025 | https://s.id/sintechjournal
DOI: https://doi.org/10.31598
Publishing: Prahasta Publisher

Perbandingan *IndoBERT* dan *Bi-LSTM* Dalam Mendeteksi Pelanggaran Undang-Undang ITE

Muhammad Dhafa Maulana[1], Christian Sri Kusuma Aditya[2]

[1,2]Departemen Informatika, Teknik, Universitas Muhammadiyah Malang
Jl. Raya Tlogomas No.246 Malang, Jawa Timur, Indonesia

e-mail: dhafamaulana@webmail.umm.ac.id[1], christianskaditya@umm.ac.id[2]

Received : March, 2025	Accepted : April, 2025	Published : April, 2025

Abstract

Social media has become a widely used platform in Indonesia, facilitating daily information exchange. However, it also serves as a medium for negative content, including hate speech, cyberbullying, and the promotion of illegal activities such as online gambling. This study aims to develop an automatic classification system to detect ITE Law violations using deep learning approaches. Two models compared are IndoBERT and Bi-LSTM. The dataset used consists of labeled Indonesian-language comments collected from social media and public sources such as Kaggle. The types of ITE violations classified include cyberbullying, hate speech, and online gambling. Experimental results show that both IndoBERT and Bi-LSTM achieved an accuracy of 97%, with IndoBERT performing slightly better in detecting cyberbullying and hate speech. This research is expected to contribute to efforts in automatically preventing ITE Law violations through natural language processing technology.

Keywords: *Bi-LSTM, IndoBERT, UU ITE*

Abstrak

Media sosial telah menjadi platform yang banyak digunakan di Indonesia untuk bertukar informasi setiap hari. Namun, media ini juga menjadi sarana penyebaran konten negatif seperti ujaran kebencian, perundungan siber, dan promosi aktivitas ilegal seperti judi online. Penelitian ini bertujuan untuk mengembangkan sistem klasifikasi otomatis guna mendeteksi pelanggaran terhadap UU ITE menggunakan pendekatan deep learning. Dua model yang dibandingkan adalah IndoBERT dan Bi-LSTM. Dataset yang digunakan berupa komentar-komentar berbahasa Indonesia yang telah diberi label, dikumpulkan dari media sosial dan sumber publik seperti Kaggle. Label pelanggaran yang diklasifikasikan mencakup cyber bullying, hate speech, dan judi online. Hasil pengujian menunjukkan bahwa model IndoBERT dan Bi-LSTM mencapai akurasi sebesar 97%, dengan IndoBERT menunjukkan performa yang lebih baik dalam mendeteksi cyberbullying dan hatespeech. Penelitian ini diharapkan dapat memberikan kontribusi dalam upaya pencegahan pelanggaran UU ITE secara otomatis melalui teknologi pemrosesan bahasa alami.

Kata Kunci: *Bi-LSTM, IndoBERT, UU ITE*

1. PENDAHULUAN

Media sosial merupakan salah satu platform yang paling banyak digunakan oleh masyarakat Indonesia untuk mencari, menerima, dan memberikan informasi setiap harinya. Seiring dengan perkembangan teknologi, penggunaan media sosial di Indonesia terus mengalami peningkatan yang signifikan dari tahun ke tahun[1]. Berdasarkan laporan Digital 2024 Indonesia, sebanyak 66,5% dari total populasi Indonesia kini telah terhubung dengan internet, dan 49,9% dari populasi memiliki identitas media sosial [2].Peningkatan ini terjadi karena beberapa faktor salah satunya yaitu dampak dari covid 19, yang meningkatkan waktu online pada pengguna social media [3].

Di balik adanya peningkatan tersebut, ada pula berbagai masalah yang muncul, seperti banyaknya penyebaran konten negatif seperti ujaran kebencian (*hate speech*), perundungan siber (*cyberbullying*), dan promosi aktivitas ilegal seperti judi online. Menurut data UNICEF, sebanyak 45% anak muda Indonesia berusia 14-24 tahun telah menjadi korban *cyberbullying*, dengan bentuk pelecehan yang paling umum terjadi melalui aplikasi chatting (45%) dan penyebaran foto atau video tanpa izin (41%) [4]. Dampak dari perilaku ini tidak hanya berpengaruh pada kondisi psikologis korban, tetapi juga berdampak negatif terhadap prestasi akademik mereka, di mana hingga 40% kasus bunuh diri anak di Indonesia terkait dengan bullying [4].

Sebagai upaya dalam mengatasi masalah ini, pemerintah Indonesia menerapkan Undang-Undang Informasi dan Transaksi Elektronik (UU ITE**)** yang bertujuan untuk melindungi masyarakat dari berbagai jenis kejahatan siber [5]. Meskipun begitu, proses deteksi manual terhadap pelanggaran UU ITE di media sosial sangat sulit dilakukan karena besarnya volume data yang harus diawasi. Oleh karena itu, teknologi kecerdasan buatan (AI), terutama melalui *Natural Language Processing (NLP)* dengan pendekatan *Deep Learning* [6], menjadi salah satu solusi yang sangat potensial untuk mendeteksi pelanggaran secara otomatis [7].

Salah satu model *AI* yang efektif untuk menangani komentar berbahasa Indonesia adalah *IndoBERT* [8], pemilihan model *IndoBERT* ini didasarkan pada relevansi dan kekuatan model dalam memahami konteks teks berbahasa Indonesia [9]. *IndoBERT*, yang dilatih khusus pada korpus Bahasa Indonesia, mampu menangkap makna kata-kata dalam bahasa lokal secara mendalam, menjadikannya sangat sesuai untuk tugas deteksi teks informal di media sosial. Sementara itu, *Bi-LSTM* menawarkan kemampuan menangkap dependensi urutan kata melalui arsitektur dua arah yang memproses teks dari dua sisi (maju dan mundur) [10], sehingga memungkinkan pemahaman yang lebih luas terhadap konteks dalam teks yang kompleks.

Berbagai penelitian sebelumnya telah memanfaatkan model *IndoBERT* dan *Bi-LSTM* untuk berbagai tugas pemrosesan bahasa alami (NLP) di Indonesia, seperti deteksi ujaran kebencian, analisis sentimen, dan klasifikasi hoaks. Pada penelitian yang berfokus pada dataset ini [11] menunjukkan bahwa kombinasi *IndoBERT*weet dan *Bi-LSTM* berhasil meningkatkan akurasi klasifikasi ujaran kebencian hingga 93,7% pada dataset publik Twitter. Sementara itu, [1] membandingkan metode *LSTM* dan *IndoBERT* dalam mendeteksi hoaks, dengan hasil menunjukkan keunggulan *IndoBERT* yang mencapai akurasi rata-rata 92,07% dibandingkan *LSTM* dengan akurasi 87,54%. Penelitian oleh [12] juga mendemonstrasikan efektivitas *IndoBERT* dalam analisis opini publik terhadap vaksin *COVID-19*, dengan akurasi mencapai 80%, lebih unggul dibandingkan *IndoBERT*weet (68%) dan *CNN-LSTM* (53%). Selain itu, pendekatan *IndoBERT*-RCNN pada analisis sentimen ulasan berbahasa Indonesia memberikan hasil yang sangat baik, mencapai akurasi 95,16% dan *F1-score* 93,27%, yang menunjukkan kinerja yang lebih baik dibandingkan dengan model dasar *IndoBERT* [13]. Penelitian-penelitian tersebut menyoroti kemampuan model berbasis transformer seperti *IndoBERT* dalam memahami konteks bahasa Indonesia, terutama pada data yang kompleks.

Dalam penelitian ini, dilakukan perbandingan kinerja antara model *IndoBERT* dan *Bi-LSTM* pada tiga topik yang berbeda: ujaran kebencian, perundungan daring (*cyberbullying*), dan konten perjudian. Pendekatan ini bertujuan untuk memberikan pemahaman yang lebih menyeluruh tentang efektivitas kedua model dalam menangani data yang kompleks dan tidak terstruktur. Dataset yang digunakan dalam

penelitian ini mencakup gabungan dari sumber terbuka (*open-source*) dan data hasil *crawling*, sehingga mencerminkan kondisi sebenarnya dari konten digital yang seringkali tidak terstruktur. Hal ini diharapkan dapat memberikan wawasan yang lebih dalam mengenai model mana yang paling optimal dalam mendeteksi pelanggaran Undang-Undang ITE pada komentar berbahasa Indonesia di media sosial, serta memperkaya upaya pengembangan teknologi pemantauan konten berbahasa Indonesia secara otomatis.

2. METODE PENELITIAN

Penelitian ini dilakukan melalui beberapa tahapan, dimulai dari proses pengumpulan data, pelabelan data, pemprosesan data teks, pelatihan model menggunakan pendekatan IndoBERT dan Bi-LSTM, hingga evaluasi performa model menggunakan metrik evaluasi seperti akurasi, precision, recall, dan F1-score. Setiap tahapan dijelaskan lebih rinci pada subbab berikut ini.

2.1 Dataset

Penelitian ini menggunakan tiga jenis dataset terkait pelanggaran undang-undang ITE pada komentar media sosial. Dataset *hate speech* diambil dari [14], sedangkan dataset *cyber bullying* merupakan kombinasi dari hasil crawling dengan alat Tweet-Harvest dan dataset dari [15]. Dataset terkait promosi judi online sepenuhnya diperoleh melalui teknik crawling.

Tabel 1: Deskripsi Dataset

Fitur	Deskripsi
cleaned_text	Teks komentar yang telah melalui tahap preprocessing
Label	Klasifikasi pelanggaran

Tabel 2: Distribusi Dataset

Dataset	Jumlah	Persentase
Cyber Bullying	2106	75%
Hate Speech	522	19%
Judi Online	175	6%

Pada tabel 1 dapat dilihat bahwa distribusi dataset tidak seimbang, Distribusi yang tidak merata ini menjadi salah satu tantangan utama dalam penelitian. Hal ini dapat digunakan untuk mengevaluasi performa model, terutama bagaimana model menangani dataset dengan ketidakseimbangan label.

2.2 Labelling

Setiap data yang diambil telah dilabeli secara manual menjadi empat kategori: cyberbullying, hate speech, dan judi online

Tabel 3: Labelling Dataset

Label	Jenis Pelanggaran
0	Cyber Bullying
1	Hate Speech
2	Judi Online

Data yang telah diberikan label kemudian diproses lebih lanjut dalam tahap preprocessing untuk memastikan teks bersih dan siap digunakan dalam model deteksi.

2.3 Preprocessing Data

Pre-processing adalah sebuah teknik yang mengubah data mentah yang tidak lengkap dan tidak konsisten menjadi format yang dapat dipahami oleh mesin [16]. Langkah pre-processing di penelitian ini bertujuan untuk mempersiapkan data agar dapat digunakan dalam penelitian [17]. Proses ini terdiri dari beberapa tahap yang dapat dilihat pada tabel 4, yaitu:

Tabel 4: Preprocessing Data

Langkah Preprocessing	Penjelasan	Text Sebelum	Text Sesudah
Penghapusan URL	Menghilangkan semua tautan dalam teks.	RT @ddestevie: habis di kasih jajan sama situs gacor habis wd langsung55 cuss nyalon dong https://t.co/dYZ0PAcnSZ	RT @ddestevie: habis di kasih jajan sama situs gacor habis wd langsung55 cuss nyalon dong
Penghapusan Angka	Semua angka dihapus dari teks.	RT @ddestevie: habis di kasih jajan sama situs gacor habis wd langsung55 cuss nyalon dong	RT @ddestevie: habis di kasih jajan sama situs gacor habis wd langsung cuss nyalon dong

Penghapusan Karakter Khusus	Menghapus mention (@), hashtag (#), & tanda baca.	RT @ddestevie: habis di kasih jajan sama situs gacor habis wd langsung cuss nyalon dong	RT ddestevie habis di kasih jajan sama situs gacor habis wd langsung cuss nyalon dong
Penghapusan Spasi Berlebih	Spasi yang berlebihan dihilangkan.	RT ddestevie habis di kasih jajan sama situs gacor habis wd langsung cuss nyalon dong	ddestevie habis di kasih jajan sama situs gacor habis wd langsung cuss nyalon dong
Case Folding	Mengubah semua teks menjadi huruf kecil.	ddestevie habis di kasih jajan sama situs gacor habis wd langsung cuss nyalon dong	ddestevie habis di kasih jajan sama situs gacor habis wd langsung cuss nyalon dong
Tokenisasi & Stopwords Removal	Memecah teks menjadi kata-kata dan menghapus kata tidak penting (stopwords).	ddestevie habis di kasih jajan sama situs gacor habis wd langsung cuss nyalon dong	ddestevie kasih jajan situs gacor wd langsung cuss nyalon

2.4 IndoBERT

IndoBERT merupakan model berbasis arsitektur *BERT (Bidirectional Encoder Representations from Transformers)* yang dirancang khusus untuk bahasa Indonesia [18]. *BERT* [19] dikenal sebagai model transformer yang mampu memahami konteks kata dalam kalimat secara mendalam melalui pendekatan *bidirectional*, yaitu membaca teks dari dua arah, baik dari kiri ke kanan maupun sebaliknya [20]. Hal ini memberikan *BERT* kemampuan untuk menangkap makna kontekstual kata secara lebih akurat dibandingkan model yang hanya memproses satu arah. *IndoBERT* memiliki 12 lapisan tersembunyi (*hidden layers*) dan telah dilatih menggunakan lebih dari 220 juta kata [21].

Tabel 5: Model IndoBERT

Layer	Output Shape	Parameter Count	Connected To
Input IDs	(None, None)	0	-
Attention Mask	(None, None)	0	-
lambda	(None, None, 768)	0	Input_ids, attention_mask
lambda_1	(None, 768)	0	Lambda
Dense	(None, 3)	2,307	Lambda_1
Total		2,037	

Penelitian ini menggunakan tokenizer berbasis model indobenchmark/*IndoBERT*-base-p2. model dilatih dengan membekukan parameter *IndoBERT*, sementara lapisan akhir berupa dense layer dengan aktivasi *softmax*. Proses pelatihan dilakukan selama 15 *epoch* menggunakan optimizer Adam dan fungsi loss categorical crossentropy.

2.5 Bi-LSTM

Bi-LSTM (Bidirectional Long Short-Term Memory) adalah model yang dirancang untuk menangani urutan data, seperti teks, dengan mempertimbangkan dependensi jangka panjang dalam data. *Bi-LSTM* adalah versi modifikasi dari *LSTM*, di mana jaringan saraf ini memproses data dalam dua arah yaitu maju dan mundur [22][23]. *LSTM* dikembangkan sebagai solusi atas keterbatasan RNN dalam mempertahankan informasi jangka panjang dari teks masukan, yang sering kali memunculkan masalah *vanishing gradient [24].* Kemampuan ini memberikan model *Bi-LSTM* keunggulan dalam memahami konteks yang lebih luas dari teks, terutama dalam tugas klasifikasi teks seperti deteksi pelanggaran pada dataset yang berisi komentar yang melanggar UU ITE.

Tabel 6: Model Bi-LSTM

Layer	Output Shape	Parameter Count
Embedding	2803, 63, 128	640,000
SpatialDropout1D	2803, 63, 128	0
Bidirectional(*LSTM*)	2803, 120	90,720
Dense	2803, 3	363
Total	731,083	

Dapat dilihat pada tabel 6, model *Bi-LSTM* didalam penelitian ini menggunakan lapisan embedding dengan dimensi 128, lapisan *Bi-LSTM* dengan 60 unit, dan lapisan keluaran dengan aktivasi softmax. Pelatihan dilakukan selama 15 epoch menggunakan optimizer Adam dan fungsi loss categorical crossentropy.

2.6 Perbandingan Kinerja Model

Setelah kedua model (*IndoBERT* dan *Bi-LSTM*) dilatih, hasil kinerjanya dibandingkan menggunakan metrik performa seperti *accuracy*, *precision*, *recall*, dan *F1-score*. Evaluasi kinerja menggunakan metrik tersebut penting karena metrik seperti *precision, recall*, dan *F1-score* memberikan gambaran terperinci tentang kemampuan model dalam mengklasifikasikan data dengan benar [25]. Hasil ini dianalisis untuk menentukan model mana yang lebih unggul dalam mendeteksi pelanggaran Undang-Undang ITE pada komentar berbahasa Indonesia.

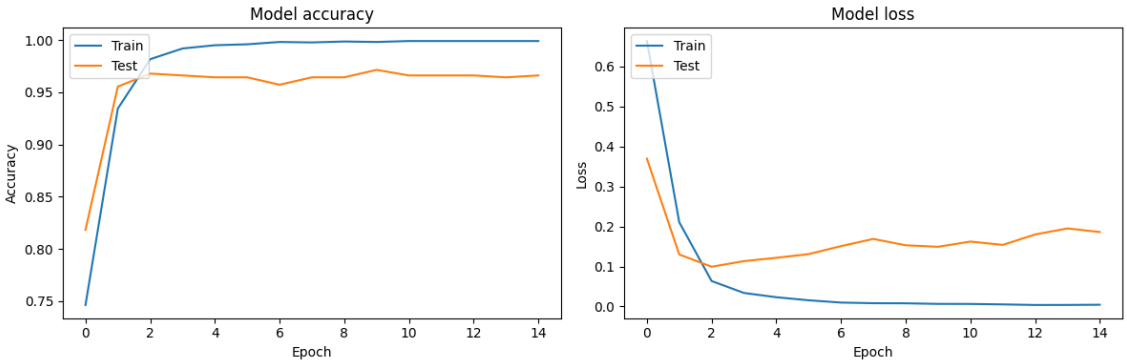
3. HASIL DAN PEMBAHASAN

Pada bab ini akan dipaparkan hasil dari penelitian yang telah dilakukan menggunakan model pretrained *IndoBERT* dan model *Bi-LSTM*, kemudian hasil tersebut akan di analisis kinerjanya dalam melatih dataset yang berisi komentar yang melanggar undang-undang ITE pada media sosial.

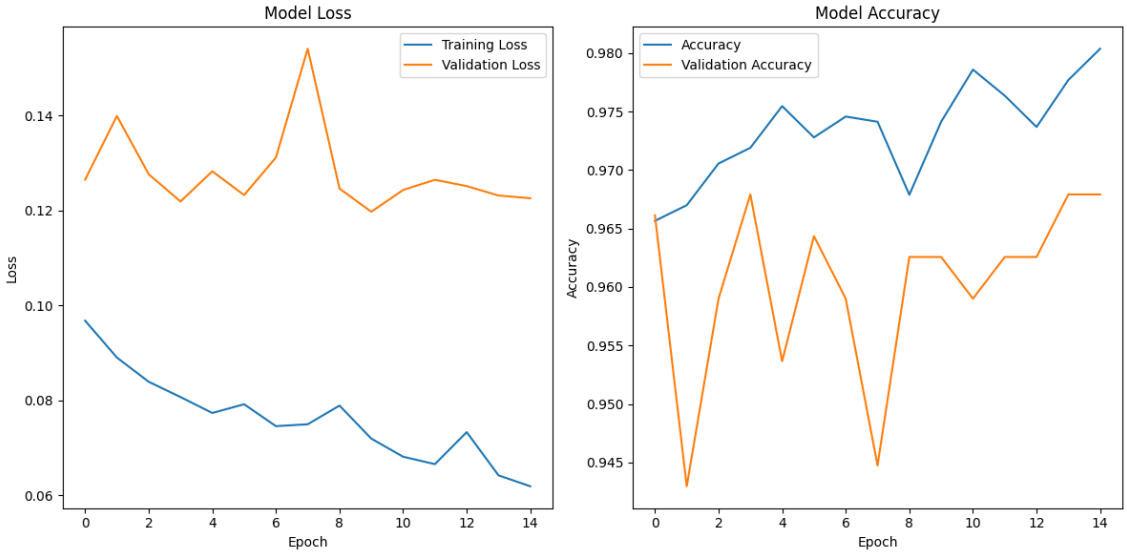
Tabel 7: Kinerja Model

Model	Label	*precision*	*recall*	*F1-score*	*accuracy*
Bi-LSTM	0 (Hatespeech)	0.92	0.90	0.91	0.97
	1 (Cyberbullying)	0.97	0.98	0.98	
	2 (Judi Online)	1.00	0.97	0.98	
IndoBERT	0 (Hatespeech)	0.90	0.94	0.92	0.97
	1 (Cyberbullying)	0.98	0.98	0.98	
	2 (Judi Online)	1.00	0.92	0.96	

Hasil pengujian pada tabel 7 menunjukkan bahwa kedua model memiliki kinerja yang baik secara keseluruhan, namun terdapat perbedaan pada kemampuan masing-masing model dalam menangani setiap kategori data. Berdasarkan hasil pada kategori Hate Speech, *Bi-LSTM* memiliki precision sebesar 0.92, sementara *IndoBERT* memiliki recall lebih tinggi yaitu 0.94, sehingga menghasilkan F1-score masing-masing sebesar 0.91 dan 0.92. Untuk kategori Cyberbullying, kedua model memberikan performa yang hampir sama dengan F1-score sebesar 0.98. Namun, pada kategori Judi Online, *Bi-LSTM* mencapai precision sempurna (1.00) dengan recall 0.97, sedangkan IndoBERT memiliki recall yang lebih rendah (0.92) meskipun precision-nya juga sempurna.



Gambar 1. Grafik *accuracy* dan *loss* model *Bi-LSTM*



Gambar 2. Grafik *accuracy* dan *loss* model *IndoBERT*

Performa kedua model juga dibandingkan berdasarkan rata-rata akurasi validasi pada 15 epoch pelatihan. *IndoBERT* menunjukkan stabilitas akurasi yang lebih baik dengan rata-rata akurasi validasi lebih tinggi dibandingkan *Bi-LSTM*, yang cenderung mengalami overfitting pada epoch akhir. Hal ini terlihat pada gambar 1, grafik akurasi dan loss di mana *Bi-LSTM* menunjukkan fluktuasi yang signifikan setelah epoch ke-10, sementara *IndoBERT* tetap konsisten.

Dalam hal waktu pelatihan, *IndoBERT* lebih unggul karena memanfaatkan transfer learning dengan pre-trained weights. Waktu pelatihan yang lebih cepat tidak mengurangi performanya secara signifikan, bahkan memberikan stabilitas yang lebih baik dibandingkan *Bi-LSTM*.

Hasil penelitian ini sejalan dengan beberapa studi terdahulu yang menunjukkan bahwa model berbasis transfer learning, seperti *IndoBERT*, lebih efektif dalam menangani data dengan struktur bahasa kompleks. Namun, Bi-LSTM tetap menjadi pilihan yang baik untuk dataset dengan distribusi kecil atau spesifik, terutama pada kelas dengan minoritas data, seperti pada kategori Hate Speech.

4. KESIMPULAN

Penelitian ini berhasil mengidentifikasi model yang lebih efektif untuk mendeteksi pelanggaran hukum digital berbasis teks berbahasa Indonesia. Model berbasis *transformer* menunjukkan keunggulan dalam menangani data dengan struktur kompleks, sementara model berbasis jaringan saraf tradisional memberikan performa kompetitif pada kondisi tertentu. Hasil ini menunjukkan bahwa pendekatan yang tepat dalam pemrosesan teks sangat bergantung pada sifat dataset dan kebutuhan aplikasi.

Temuan ini memberikan panduan bagi pengembang untuk memilih model yang sesuai dalam membangun sistem deteksi pelanggaran daring. Dengan hasil ini, penelitian dapat menjadi landasan pengembangan lebih lanjut di bidang analisis teks legal, termasuk eksplorasi integrasi model atau peningkatan metode klasifikasi untuk aplikasi praktis di Indonesia.

DAFTAR PUSTAKA

[1] M. Sinapoy, Y. Sibaroni, and S. S. Prasetyowati, "Comparison of LSTM and IndoBERT Method in Identifying Hoax on Twitter," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 7, no. 3, pp. 657–662, Jun. 2023, doi: 10.29207/resti.v7i3.4830.

[2] "DIGITAL 2024: THE ESSENTIAL GUIDE TO THE LATEST CONNECTED BEHAVIOURS - We Are Social Indonesia." Accessed: Jun. 10, 2024. [Online]. Available: https://wearesocial.com/id/blog/2024/01/digital-2024/

[3] I. S. Borualogo, H. Wahyudi, and S. Kusdiyati, "Prevalence and Predictors of Cyberbullying in Middle and High School Students During the COVID-19 Pandemic," *Jurnal Psikologi*, vol. 50, no. 2, p. 206, Aug. 2023, doi: 10.22146/jpsi.76494.

[4] "BULLYING IN INDONESIA: Key Facts, Solutions, and Recommendations." Accessed: Mar. 09, 2025. [Online]. Available: https://www.unicef.org/indonesia/media/5606/file/Bullying.in.Indonesia.pdf

[5] E. N. Putra, "LAW'S SILENCE ON CYBERBULLYING TO CHILDREN IN INDONESIA," *Brawijaya Law Journal*, vol. 11, no. 1, pp. 135–163, Mar. 2024, doi: 10.21776/ub.blj.2024.011.01.07.

[6] C. I. Garcia, F. Grasso, A. Luchetta, M. C. Piccirilli, L. Paolucci, and G. Talluri, "A comparison of power quality disturbance detection and classification methods using CNN, LSTM and CNN-LSTM," *Applied Sciences (Switzerland)*, vol. 10, no. 19, pp. 1–22, Oct. 2020, doi: 10.3390/app10196755.

[7] Y. Wen and P. Ti, "A Study of Legal Judgment Prediction Based on Deep Learning Multi-Fusion Models—Data from China," *Sage Open*, vol. 14, no. 3, Jul. 2024, doi: 10.1177/21582440241257682.

[8] S. M. Isa, G. Nico, and M. Permana, "INDOBERT FOR INDONESIAN FAKE NEWS DETECTION," *ICIC Express Letters*, vol. 16, no. 3, pp. 289–297, Mar. 2022, doi: 10.24507/icicel.16.03.289.

[9] L. Geni, E. Yulianti, and D. I. Sensuse, "Sentiment Analysis of Tweets Before the 2024 Elections in Indonesia Using IndoBERT Language Models," *Jurnal Ilmiah Teknik Elektro Komputer dan Informatika (JITEKI)*, vol. 9, no. 3, pp. 746–757, 2023, doi: 10.26555/jiteki.v9i3.26490.

[10] A. D. Safira and E. B. Setiawan, "Hoax Detection in Social Media using Bidirectional Long Short-Term Memory (Bi-LSTM) and 1 Dimensional-Convolutional Neural Network (1D-CNN) Methods," in *2023 11th International Conference on Information and Communication Technology, ICoICT 2023*, 2023, pp. 355–360. doi: 10.1109/ICoICT58202.2023.10262528.

[11] J. Kusuma and A. Chowanda, "Indonesian Hate Speech Detection Using IndoBERTweet and BiLSTM on Twitter," *JOIV : International Journal on Informatics Visualization*, vol. 7, pp. 773–780, 2023, doi: https://dx.doi.org/10.30630/joiv.7.3.1035.

[12] S. Saadah, K. Auditama, A. Fattahila, F. Amorokhman, A. Aditsania, and 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.

[13] H. Jayadianti, W. Kaswidjanti, A. T. Utomo, S. Saifullah, F. A. Dwiyanto, and R. Drezewski, "Sentiment analysis of Indonesian reviews using fine-tuning IndoBERT and R-CNN," *ILKOM Jurnal Ilmiah*, vol. 14, no. 3, pp. 348–354, Dec. 2022, doi: 10.33096/ilkom.v14i3.1505.348-354.

[14] I. Alfina, R. Mulia, M. I. Fanany, and Y. Ekanata, "Hate speech detection in the Indonesian language: A dataset and preliminary study," in *2017 International Conference on Advanced*

Computer Science and Information Systems, ICACSIS 2017, Institute of Electrical and Electronics Engineers Inc., Jul. 2017, pp. 233–237. doi: 10.1109/ICACSIS.2017.8355039.

[15] W. Athira Luqyana, I. Cholissodin, and R. S. Perdana, "Analisis Sentimen Cyberbullying pada Komentar Instagram dengan Metode Klasifikasi Support Vector Machine," vol. 2, no. 11, pp. 4704–4713, 2018, [Online]. Available: http://j-ptiik.ub.ac.id

[16] M. Umer, Z. Imtiaz, S. Ullah, A. Mehmood, G. S. Choi, and B. W. On, "Fake news stance detection using deep learning architecture (CNN-LSTM)," *IEEE Access*, vol. 8, pp. 156695–156706, 2020, doi: 10.1109/ACCESS.2020.3019735.

[17] S. Pradha, M. N. Halgamuge, and N. Tran Quoc Vinh, "Effective text data preprocessing technique for sentiment analysis in social media data," in *Proceedings of 2019 11th International Conference on Knowledge and Systems Engineering, KSE 2019*, Institute of Electrical and Electronics Engineers Inc., Oct. 2019. doi: 10.1109/KSE.2019.8919368.

[18] B. Wilie *et al.*, "IndoNLU: Benchmark and Resources for Evaluating Indonesian Natural Language Understanding," in *AACL-IJCNLP*, Sep. 2020. [Online]. Available: http://arxiv.org/abs/2009.05387

[19] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," in *Proc. 2019 Conf. North American Chapter Assoc. Comput. Linguist.: Human Language Technology*, 2019, pp. 4171–4186. doi: 10.18653/v1/N19-1423.

[20] N. Rai, D. Kumar, N. Kaushik, C. Raj, and A. Ali, "Fake News Classification using transformer based enhanced LSTM and BERT," *International Journal of Cognitive Computing in Engineering*, vol. 3, pp. 98–105, Jun. 2022, doi: 10.1016/j.ijcce.2022.03.003.

[21] F. Koto, A. Rahimi, J. H. Lau, and T. Baldwin, "IndoLEM and IndoBERT: A Benchmark Dataset and Pre-trained Language Model for Indonesian NLP," in *Proceedings of the 28th International Conference on Computational Linguistics*, Online, 2020, pp. 757–770. doi: 10.18653/v1/2020.coling-main.66.

[22] F. Shahid, A. Zameer, and M. Muneeb, "Predictions for COVID-19 with deep learning models of LSTM, GRU and Bi-LSTM," *Chaos Solitons Fractals*, vol. 140, Nov. 2020, doi: 10.1016/j.chaos.2020.110212.

[23] A. Wani, I. Joshi, S. Khandve, V. Wagh, and R. Joshi, "Evaluating Deep Learning Approaches for Covid19 Fake News Detection," in *Combating Online Hostile Posts in Regional Languages during Emergency Situation*, Jan. 2021, pp. 153–163. doi: 10.1007/978-3-030-73696-5_15.

[24] R. K. Kaliyar, K. Fitwe, P. Rajarajeswari, and A. Goswami, "Classification of Hoax/Non-Hoax News Articles on Social Media using an Effective Deep Neural Network," in *Proceedings - 5th International Conference on Computing Methodologies and Communication, ICCMC 2021*, Institute of Electrical and Electronics Engineers Inc., Apr. 2021, pp. 935–941. doi: 10.1109/ICCMC51019.2021.9418282.

[25] R. Yacouby and D. Axman, "Probabilistic Extension of Precision, Recall, and F1 Score for More Thorough Evaluation of Classification Models," in *Proceedings of the First Workshop on Evaluation and Comparison of NLP Systems*, 2020, pp. 79–91. doi: 10.18653/v1/2020.eval4nlp-1.9.