

---

## Deteksi Cyberbullying Menggunakan BERT dan Bi-LSTM

Fidya Farasalsabila<sup>1</sup>, Ema Utami<sup>2</sup>, Hanafi<sup>3</sup>

<sup>123</sup>Magister Informatika, Universitas Amikom Yogyakarta

Email: [fidya@students.amikom.ac.id](mailto:fidya@students.amikom.ac.id)<sup>1</sup>

---

### ABSTRACT

*Cyberbullying is a digital problem that is not a new phenomenon. This existed before the advent of social networks, and cyberbullying has a wide impact, including a person's mental and physiological conditions such as sadness, anxiety and depression. The main objective of this research is to develop an effective cyberbullying detection system using natural language processing techniques. The method used in this research includes the application of the BERT (Bi-Directional Encoder Representations from Transformers) and Bi-LSTM (Bi-Directional Long Short-Term Memory) models as a deep learning approach to analyze text and detect cyberbullying behavior. This approach allows the system to understand complex language contexts and capture patterns that traditional methods may find difficult to identify. Testing was carried out using a dataset that included various types of Indonesian language texts containing cyber bullying acts. The research results show that the combination of BERT and Bi-LSTM is able to provide superior detection performance with a high accuracy rate of 90% and the ability to identify variations of cyber bullying. This research makes a significant contribution to efforts to protect individuals from the negative impacts of cyber bullying through the development of a sophisticated and adaptive detection system.*

**Keywords:** BERT, Bi-LSTM, Cyberbullying, Deep Learning.

### INTISARI

Perundungan dunia maya adalah permasalahan digital yang bukan merupakan fenomena baru. Hal ini sudah ada sebelum munculnya jaringan sosial, dan *cyberbullying* memiliki dampak yang luas, termasuk kondisi mental dan fisiologis seseorang seperti kesedihan, kecemasan hingga depresi. Tujuan utama penelitian ini adalah mengembangkan sebuah sistem deteksi *cyberbullying* yang efektif menggunakan teknik-teknik pemrosesan bahasa alami. Metode yang digunakan dalam penelitian ini mencakup penerapan model BERT (*Bidirectional Encoder Representations from Transformers*) dan Bi-LSTM (*Bidirectional Long Short-Term Memory*) sebagai pendekatan *deep learning* untuk menganalisis teks dan mendeteksi perilaku cyber bullying. Pendekatan ini memungkinkan sistem untuk memahami konteks bahasa yang kompleks dan menangkap pola-pola yang mungkin sulit diidentifikasi oleh metode tradisional. Pengujian dilakukan menggunakan dataset yang mencakup berbagai jenis teks berbahasa Indonesia yang mengandung tindakan *cyberbullying*. Hasil penelitian menunjukkan bahwa kombinasi BERT dan Bi-LSTM mampu memberikan performa deteksi yang unggul dengan tingkat akurasi yang tinggi sebesar 90% dan kemampuan mengidentifikasi variasi dari cyber bullying. Penelitian ini memberikan kontribusi signifikan dalam upaya melindungi individu dari dampak negatif *cyberbullying* melalui pengembangan sistem deteksi yang canggih dan adaptif.

**Kata kunci:** BERT, Bi-LSTM, *Cyberbullying*, *Deep Learning*.

---

### PENDAHULUAN

Di era digital yang semakin berkembang, permasalahan perundungan dunia maya (*cyberbullying*) menjadi sorotan utama dalam kehidupan sekarang. Ada beberapa penelitian dalam literatur untuk mencegah *cyberbullying* menggunakan teknik intervensi dan pencegahan. Ide-ide tersebut berasal dari domain psikologi dan pendidikan. Namun, teknik-teknik ini jarang terjadi dalam skala dunia. Selain itu, korban *cyberbullying* sering kali menolak berkomunikasi dengan orang tua (Li dkk. 2019), guru (Jiang dkk. 2017), atau orang dewasa lainnya (Medhat, Hassan, dan Korashy 2014). Mereka menghabiskan banyak waktu online (Al-Garadi dkk. 2019), mencari bantuan anonim (Ke dkk. t.t.), dan mempublikasikan kebutuhan akan informasi dan bantuan di internet (Hosmer, Lemeshow, dan Sturdivant t.t.). Namun, internet adalah cara paling efektif untuk memberikan perawatan *cyberbullying*. Metode berbasis *web* juga dapat digunakan kapan saja dan kapan saja pasien menginginkannya (Chavan dan S 2015). Dengan beragamnya bentuk dan konteks pesan yang dianggap merugikan, analisis sentimen menjadi instrumen penting dalam mendeteksi dan mengkategorikan perilaku *cyberbullying*.

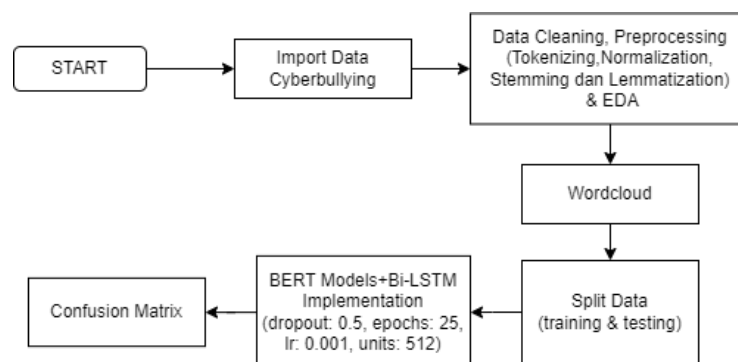
Banyak orang di seluruh dunia sekarang menggunakan platform media sosial seperti Twitter, Facebook, Instagram, dan TikTok sebagai sarana komunikasi rutin. Platform ini telah menjadi bagian integral dari kehidupan mereka sehari-hari (Anderson 2020). Platform ini telah menjadi bagian penting dari kehidupan kita sehari-hari (Akhter dkk. 2023), dan manfaatnya dalam meningkatkan komunikasi melalui koneksi, pemasaran, penyediaan informasi, hiburan, dan pendidikan tidak dapat dianggap remeh. Masalah *cyberbullying* ini merupakan tren yang relatif baru dan belakangan ini semakin populer sebagai topik permasalahan (Ogunleye dan Dharmaraj 2023). Banyak metode penambahan teks digunakan untuk melakukan analisis sentimen. Metode *Support Vector Machine* dipilih dari sekian banyak metode *text mining* karena memiliki nilai akurasi yang relatif tinggi jika dibandingkan dengan metode lainnya, seperti pada penelitian (Albayari dan Abdallah 2022) yang membandingkan Kinerja *Random Forest* (RF), *Multinomial Naïve Bayes* (MNB), Regresi Logistik (LR), dan *Support Vector Machine* (SVM). Penelitian lain (Eom, Yun, dan Byeon 2022) membandingkan kinerja prediktif dari lima model, yaitu SVM, RNN, LSTM, BERT, dan KoBERT. Hasilnya menegaskan bahwa model KoBERT menunjukkan kinerja terbaik (71%) di semua indikator kinerja prediktif (akurasi, presisi, dan skor F1). Meskipun masih banyak model lain yang perlu dieksplorasi, penelitian (Joshi dan Gupta 2022) membuktikan bahwa Jaringan Res-CNN-BiLSTM untuk klasifikasi *teks cyberbullying* merupakan pendekatan yang baik, namun tetap memiliki beberapa kekurangan yang perlu diatasi agar menghasilkan akurasi yang optimal.

Penelitian lain (Novianti dan Wardani 2023), (Rizki Aditya, Supriyati, dan Listyorini 2022) menggunakan model *Naive Bayes* dalam prosesnya. Hasil yang diperoleh masih rendah dengan akurasi masing-masing sebesar 75% dan 77,5%. Penelitian lainnya menggunakan kosakata bahasa Indonesia dari data postingan Twitter, Tiktok (Setiawan, M. Lhaksmana, dan Bunyamin 2023) dengan LSTM dan IndoBERTweet dengan tujuan mengidentifikasi pendekatan yang paling sesuai untuk mengembangkan model yang secara otomatis dapat mengkategorikan ulasan TikTok ke dalam sikap negatif, netral, dan positif. IndoBERTweet menunjukkan kinerja yang lebih unggul dibandingkan yang lain, mencapai akurasi 80%, sedangkan akurasi LSTM mencapai 78%.

Karena algoritma BERT dan Bi-LSTM dinilai memiliki tingkat akurasi yang lebih tinggi dibandingkan pendekatan lainnya, penulis penelitian ini memutuskan untuk menggunakannya. Keputusan ini didasarkan pada penelitian sebelumnya (Anderson 2020), (Chintalapudi, Battineni, dan Amenta 2021), (Jason Wang, Kaiqun Fu, dan Chang-Tien Lu 2020). Pemanfaatan *library machine learning* yang *open source* inilah yang membedakan penelitian ini dengan penelitian serupa yang pernah dilakukan sebelumnya. Pytorch, yang sebelumnya dikembangkan oleh Facebook, dimanfaatkan oleh para peneliti dalam pekerjaan ini. Penelitian ini akan memberikan lebih banyak informasi untuk membantu upaya pencegahan dan penghapusan *cyberbullying* dengan menggunakan sarana teknologi kreatif.

## METODE

Model yang diusulkan pada penelitian ini adalah BERT dan Bi-LSTM. Ada beberapa langkah yang perlu dilakukan sebelum menerapkan model ini dalam penelitian untuk melakukan klasifikasi *cyberbullying*.



**Gambar 1.** Alur Penelitian Deteksi *Cyberbullying* menggunakan BERT dan Bi-LSTM

Gambar 1 menggambarkan alur dari penelitian yang diusulkan. Langkah awal, mengimport dan membersihkan dataset *cyberbullying* agar data dapat dilanjutkan ke proses implementasi model. Selanjutnya, data akan di bagi menjadi beberapa subset baru akan di implementasikan ke model BERT dan Bi-LSTM. Langkah akhir adalah mengevaluasi kinerja model BERT dan Bi-LSTM terhadap data *cyber bullying*.

Kumpulan data mentah dalam penelitian ini diperoleh dari (Yudi Setiawan 2023) dan disimpan dalam “*comma separated values*” (CSV) dan dapat diakses oleh publik melalui *Mendeley Data*. Sebanyak 15.648 dataset dipilih untuk digunakan dan akan dibagi menjadi 2 jenis data untuk keperluan penelitian ini, yakni data latih dan data uji. Data Latih digunakan untuk melatih algoritma dalam mencari model yang sesuai. Sedangkan data uji merupakan data baru yang belum memiliki kelas sehingga diperlukan proses klasifikasi untuk menentukan kelas yang cocok. Dataset memiliki 2 label, yakni label 0 (*non-cyberbullying*) dan label 1 (*cyberbullying*). Langkah penting dalam menyiapkan data untuk analisis lebih lanjut diperlukan sehingga data *cleaning* dan *preprocessing* dilakukan dalam penelitian ini, sehingga keputusan yang diambil berdasarkan data tersebut dapat lebih akurat dan relevan. Proses data *cleaning* dan *preprocessing* dilakukan menggunakan bahasa pemrograman python dan librarynya. Langkah *preprocessing* yang dipilih adalah 4 langkah (Ignatow dan Mihalcea 2018), yakni Tokenization, Stopword Removal, Stemming dan Lemmatization.

## HASIL DAN PEMBAHASAN

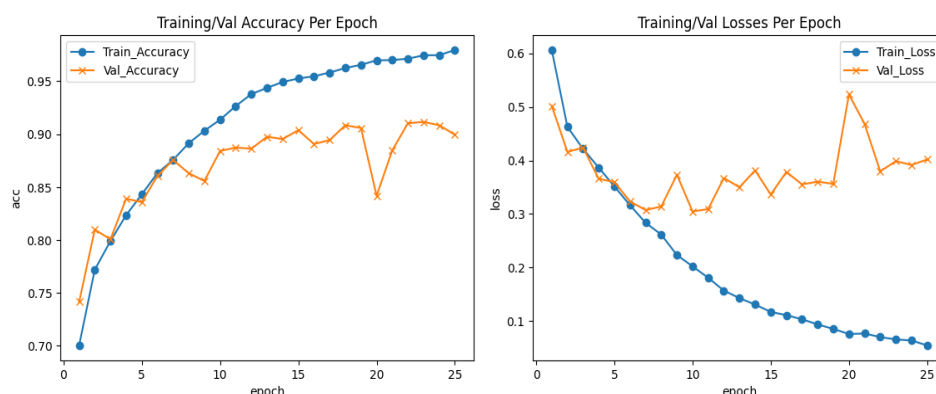
### Hasil

Studi kasus yang diambil merupakan studi kasus *cyberbullying* yang dataset diambil sebanyak 15.648 didapat melalui *Mendeley Data* (Yudi Setiawan 2023). Dataset memiliki 2 label, yakni label 0 (*non-cyberbullying*) dan label 1 (*cyberbullying*). Setelah proses *preprocessing*, sebanyak 10953 data latih dan 4695 data uji digunakan dalam implementasi ke dalam model BERT dan Bi-LSTM. Hasil dari keseluruhan *preprocessing* dapat dilihat pada gambar 2. Data hasil *preprocessing* akan digunakan untuk menganalisis hasil prediksi model. Hasil dari *preprocessing* ini akan meminimalkan potensi bias dalam model, meningkatkan keberlanjutan, dan memberikan pemahaman yang lebih mendalam terhadap permasalahan *cyberbullying*.

	Text	casefolding	cleansing	remove_emoticon	tokenizing	normalization	stopwords	stemming	lemmatization
0	Oooo statusnya masih darurat tho?... Daruratny...	o000 statusnya masih darurat tho?... daruratny...	o000 statusnya masih darurat tho daruratnya ...	o000 statusnya masih darurat tho daruratnya ...	[o000, statusnya, masih, darurat, tho, darurat...	[o000, statusnya, masih, darurat, though, daru...	[o000, statusnya, darurat, though, daruratnya,...	[o000, status, darurat, though, darurat, orang...	o000 status darurat though darurat mana orang ...
1	Pembicaraan antar profesor pas mengemukakan pe...	pembicaraan antar profesor pas mengemukakan pe...	pembicaraan antar profesor pas mengemukakan pe...	pembicaraan antar profesor pas mengemukakan pe...	[pembicaraan, antar, profesor, pas, mengemukak...	[pembicaraan, antar, profesor, pas, mengemukak...	[pembicaraan, profesor, pas, mengemukakan, pen...	[bicara, profesor, pas, muka, tua]	bicara profesor pas muka tua
2	Pasti bagus kalau ad acara talkshow . Host nya...	pasti bagus kalau ad acara talkshow . host nya...	pasti bagus kalau ad acara talkshow host nya ...	pasti bagus kalau ad acara talkshow host nya ...	[pasti, bagus, kalau, ad, acara, talkshow, hos...	[pasti, bagus, kalau, ada, acara, talkshow, ho...	[bagus, acara, talkshow, host, bertiga]	[bagus, acara, talkshow, host]	bagus acara talkshow host tiga
3	Uppppp	uppppp	uppppp	uppppp	[uppppp]	[uppppp]	[uppppp]	[uppppp]	uppppp
4	nonton cuma liat timecode berjalan...soalnya k...	nonton cuma liat timecode berjalan...soalnya k...	nonton cuma liat timecode berjalan soalnya k...	nonton cuma liat timecode berjalan soalnya k...	[nonton, cuma, liat, timecode, berjalan, soaln...	[menonton, cuma, lihat, timecode, berjalan, so...	[menonton, lihat, timecode, berjalan, tidak, m...	[tonton, lihat, timecode, jalan, tidak, bicaraj]	tonton lihat timecode jalan tidak erti bicara
...	...	...	...	...	...	...	...	...	...

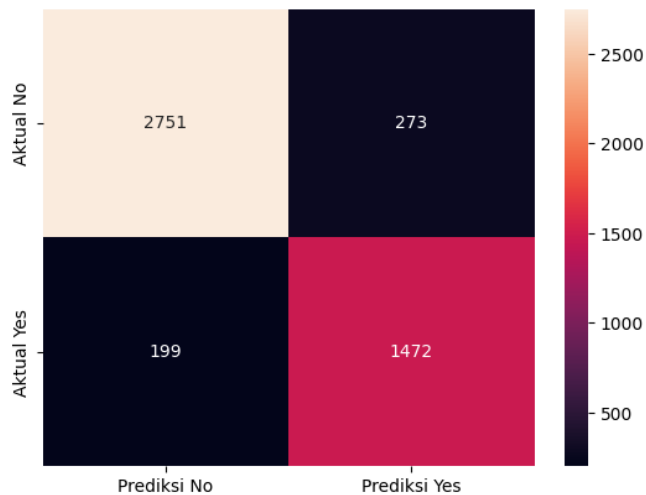
Gambar 2. Hasil Data *Preprocessing*

Dalam konteks penelitian ini, sebuah model yang menggabungkan BERT dengan arsitektur jaringan rekuren Bi-LSTM telah dikembangkan dan dievaluasi untuk tugas klasifikasi *cyberbullying*. Pencapaian terpenting adalah peningkatan signifikan dalam validasi kinerja model, yang naik dari 0.98% menjadi 0.99%. Ini menunjukkan peningkatan yang substansial dalam akurasi prediksi model, mengindikasikan kemampuannya untuk memahami dan mengklasifikasikan data secara lebih tepat.



Gambar 3. Grafik hasil akurasi dan loss pada setiap epoch

Gambar 3 di bagian akurasi (acc) menunjukkan pada ada awal proses pelatihan, akurasi pelatihan dan validasi dimulai pada sekitar 0,70. Seiring bertambahnya jumlah epoch, akurasi pelatihan juga meningkat dan mencapai akurasi puncak sekitar 0,95 pada epoch ke-20. Namun, akurasi validasi meningkat lebih lambat dan mencapai akurasi puncak sekitar 0,90 pada epoch ke-15. Setelah epoch ke-15, akurasi validasi mulai menurun sementara akurasi pelatihan terus meningkat. Ini merupakan tanda *overfitting*, yang terjadi ketika model menjadi terlalu kompleks dan mulai mengingat data pelatihan alih-alih mempelajari pola dasarnya. Untuk menghindari *overfitting*, penting untuk memantau akurasi validasi selama proses pelatihan dan menghentikan pelatihan ketika akurasi validasi mulai menurun. Gambar 3 bagian loss menunjukkan grafik kehilangan pelatihan dan validasi per era selama proses pelatihan model pembelajaran mesin yang menggunakan BERT dan Bi-LSTM. Pada awal proses pelatihan, kehilangan pelatihan dan validasi dimulai sekitar 0,6. Ketika jumlah epoch meningkat, baik pelatihan dan kerugian validasi menurun, menunjukkan bahwa model ini sedang membaik. Kerugian pelatihan menurun lebih cepat daripada kerugian validasi, menunjukkan bahwa model ini mengingat data pelatihan lebih baik daripada data validasi. Pada era ke-15, hilangnya pelatihan mencapai minimum sekitar 0,1, sementara hilangnya validasi mencapai minimal sekitar 0,2. Setelah era ke-15, baik pelatihan dan kerugian validasi mulai meningkat sedikit, menunjukkan bahwa model mulai berlebihan. *Overfitting* terjadi ketika model menjadi terlalu kompleks dan mulai menghafal data pelatihan daripada mempelajari pola yang mendasarinya. Secara ringkas, gambar 3 menunjukkan kerugian pelatihan dan validasi per era selama proses pelatihan model pembelajaran mesin yang menggunakan BERT dan Bi-LSTM. Model ini dimulai dengan kerugian tinggi dan secara bertahap ditingkatkan seiring dengan meningkatnya jumlah era. Kedua kehilangan pelatihan dan validasi menurun, tetapi setelah titik tertentu, kerugian mulai meningkat, menunjukkan *overfitting*.



**Gambar 4.** Evaluasi Confusion Matrix

Gambar 4 menunjukkan matriks evaluasi yang merupakan tabel *confusion matrix* yang digunakan untuk mengevaluasi kinerja model pembelajaran mesin yang mendeteksi *cyberbullying* menggunakan metode BERT dan Bi-LSTM. Confusion matrix menunjukkan jumlah prediksi *true positive* (TP), *false positive* (FP), *false negative* (FN), dan *true negative* (TN) yang dibuat oleh model BERT+Bi-LSTM. Dalam hal akurasi untuk kelas No (non-cyberbullying) adalah 0,9325, yang menunjukkan bahwa model tersebut secara benar memprediksi 93,25% dari *posting non-cyberbullying*. Dalam hal *recall* untuk kelas Ya (cyberbullying) adalah 0,8809, yang menunjukkan bahwa model tersebut benar memprediksi 88,09% dari postingan *cyberbullying*. Dalam hal F1-Score untuk kelas Ya (cyberbullying) adalah 0,8618. Akurasi model adalah 0,8995, yang menunjukkan bahwa model tersebut benar memprediksi sekitar 90% dari total posting. Hasil yang diperoleh merupakan kombinasi model BERT dengan Bi-LSTM berfungsi dengan baik dalam mendeteksi postingan *cyberbullying*, dengan akurasi tinggi, *recall*, dan F1-Score untuk kelas No (non-cyberbullying). Namun, kinerja model untuk kelas Ya (cyberbullying) tidak begitu tinggi, menunjukkan bahwa ada ruang untuk perbaikan dalam mendeteksi *posting cyberbullying*.

## KESIMPULAN

*Cyberbullying* mengacu pada tindakan terlibat dalam perilaku pelecehan online, yang melibatkan penggunaan bentuk komunikasi yang merendahkan, *forceful*, atau intimidatif. Tujuan dari studi ini adalah untuk mengembangkan dan meningkatkan sistem deteksi *cyberbullying* yang mampu menganalisis dan menghilangkan kasus pelecehan online yang dilakukan oleh individu di platform media sosial. Model BERT dan Bi-LSTM di gabungkan dan digunakan untuk melakukan analisis sentiment pada dataset cyber bullying dalam penelitian ini. Penelitian menunjukkan bahwa kombinasi model yang diusulkan mencapai akurasi sekitar 90%. Penelitian ini terhambat oleh fakta bahwa hanya dataset bahasa Indonesia yang digunakan. Potensi upaya penelitian masa depan mungkin melibatkan meneliti metodologi untuk mengklarifikasi proses pengambilan keputusan untuk memilih model yang tepat untuk kategorisasi multi-label. Meningkatkan kinerja BERT membutuhkan pemilihan parameter yang tepat, dengan tujuan memajukan pemahaman tentang bagaimana menentukan parameter paling ideal di masa depan.

## DAFTAR PUSTAKA

- Akhter, Arnisha, Uzzal Kumar Acharjee, Md. Alamin Talukder, Md. Manowarul Islam, dan Md Ashraf Uddin. 2023. "A robust hybrid machine learning model for Bengali cyber bullying detection in social media." *Natural Language Processing Journal* 4:100027. doi: 10.1016/j.nlp.2023.100027.
- Albayari, Reem, dan Sherief Abdallah. 2022. "Instagram-Based Benchmark Dataset for Cyberbullying Detection in Arabic Text." *Data* 7(7). doi: 10.3390/data7070083.
- Al-Garadi, Mohammed Ali, Mohammad Rashid Hussain, Nawsher Khan, Ghulam Murtaza, Henry Friday Nweke, Ihsan Ali, Ghulam Mujtaba, Haruna Chiroma, Hasan Ali Khattak, dan Abdullah Gani. 2019. "Predicting Cyberbullying on Social Media in the Big Data Era Using Machine Learning Algorithms: Review of Literature and Open Challenges." *IEEE Access* 7:70701–18. doi: 10.1109/ACCESS.2019.2918354.
- Anderson, Katie Elson. 2020. "Getting acquainted with social networks and apps: it is time to talk about TikTok." *Library Hi Tech News* 37(4):7–12. doi: 10.1108/LHTN-01-2020-0001.
- Chavan, Vikas S., dan Shylaja S. S. 2015. *Machine Learning Approach for Detection of Cyber-Aggressive Comments by Peers on Social Media Network*.
- Chintalapudi, Nalini, Gopi Battineni, dan Francesco Amenta. 2021. "Sentimental analysis of COVID-19 tweets using deep learning models." *Infectious Disease Reports* 13(2). doi: 10.3390/IDR13020032.
- Eom, Gayeong, Sanghyun Yun, dan Haewon Byeon. 2022. "Predicting the sentiment of South Korean Twitter users toward vaccination after the emergence of COVID-19 Omicron variant using deep learning-based natural language processing." *Frontiers in Medicine* 9. doi: 10.3389/fmed.2022.948917.
- Hosmer, David W., Stanley Lemeshow, dan Rodney X. Sturdivant. t.t. *Applied logistic regression*.
- Ignatow, Gabe, dan Rada Mihalcea. 2018. *An Introduction to Text Mining*.
- Jason Wang, Kaiqun Fu, dan Chang-Tien Lu. 2020. "Fine-Grained Balanced Cyberbullying Dataset."
- Jiang, Chunxiao, Haijun Zhang, Yong Ren, Zhu Han, Kwang Cheng Chen, dan Lajos Hanzo. 2017. "Machine Learning Paradigms for Next-Generation Wireless Networks." *IEEE Wireless Communications* 24(2):98–105. doi: 10.1109/MWC.2016.1500356WC.
- Joshi, Raunak, dan Abhishek Gupta. 2022. "Performance Comparison of Simple Transformer and Res-CNN-BiLSTM for Cyberbullying Classification."
- Ke, Guolin, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, dan Tie-Yan Liu. t.t. *LightGBM: A Highly Efficient Gradient Boosting Decision Tree*.
- Li, Jun, Guimin Huang, Chunli Fan, Zhenglin Sun, dan Hongtao Zhu. 2019. "Key word extraction for short text via word2vec, doc2vec, and textrank." *Turkish Journal of Electrical Engineering and Computer Sciences* 27(3):1794–1805. doi: 10.3906/elk-1806-38.
- Medhat, Walaa, Ahmed Hassan, dan Hoda Korashy. 2014. "Sentiment analysis algorithms and applications: A survey." *Ain Shams Engineering Journal* 5(4):1093–1113. doi: 10.1016/j.asej.2014.04.011.
- Novianti, Fenny, dan Kiky Rizky Nova Wardani. 2023. "Analisis Sentimen Masyarakat Terhadap Data Tweet Traveloka Selama Rapid Test Antigen Menggunakan Algoritma Naïve Bayes." *JIPi (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)* 8(3):922–33. doi: 10.29100/jipi.v8i3.3973.
- Ogunleye, Bayode, dan Babitha Dharmaraj. 2023. "The Use of a Large Language Model for Cyberbullying Detection." *Analytics* 2(3):694–707. doi: 10.3390/analytics2030038.

- Rizki Aditya, Dio, Endang Supriyati, dan Tri Listyorini. 2022. *Analisis Sentimen Pengguna Twitter Terhadap Rokok Elektrik (VAPE) di Indonesia Menggunakan Metode Naïve Bayes*.
- Setiawan, Jerry Cahyo, Kemas M. Lhaksmana, dan Bunyamin Bunyamin. 2023. "Sentiment Analysis of Indonesian TikTok Review Using LSTM and IndoBERTweet Algorithm." *JIPi (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)* 8(3):774–80. doi: 10.29100/jipi.v8i3.3911.
- Setiawan, Yudi, Nur Ulfa Maulidevi, dan Kridanto Surendro. 2023. "The Use of Dynamic n-Gram to Enhance TF-IDF Features Extraction for Bahasa Indonesia Cyberbullying Classification." Hlm. 200–205 dalam *ACM International Conference Proceeding Series*. Association for Computing Machinery.
- Yudi Setiawan. 2023. "Bahasa Cyberbullying Dataset (Source Data: Instagram, Twitter, and Youtube)." *Mendeley Data*.