



## Model *Indo-BERT* untuk Identifikasi Sentimen Kekerasan Verbal di Twitter

Nurjoko<sup>1\*</sup>, Agus Rahardi<sup>2</sup>

<sup>1</sup> Program Studi Sains Data, Institut Informatika dan Bisnis Darmajaya, Bandar Lampung, Indonesia

<sup>2</sup> Program Studi Sistem Informasi, Institut Informatika dan Bisnis Darmajaya, Bandar Lampung, Indonesia

\*Email Penulis Korespondensi: [nurjoko@darmajaya.ac.id](mailto:nurjoko@darmajaya.ac.id)

### Abstrak

Munculnya kekerasan verbal di platform media sosial Twitter menjadi isu yang semakin memprihatinkan dalam beberapa tahun terakhir. Kekerasan verbal mencakup berbagai bentuk komunikasi yang merendahkan, menghina, dan mengancam, seringkali merugikan individu atau kelompok. Permasalahan utama yang dibahas dalam penelitian ini adalah identifikasi sentimen terkait perilaku kekerasan verbal di media sosial, khususnya Twitter. Penelitian ini bertujuan untuk mengembangkan model analisis sentimen Verbal Violence Behavior (VVB) di Twitter dengan menggunakan *Indo-BERT*. Keakuratan model akan dibandingkan dengan *BERT*. Penelitian ini diawali dengan pengumpulan data melalui web crawling dengan menggunakan kategori perilaku kekerasan verbal sebagai referensi. Pelabelan dataset dilakukan melalui kombinasi metode manual dan otomatis dengan pendekatan semi-supervised learning. Proses ini melibatkan pembelajaran mandiri, di mana data yang tidak berlabel secara otomatis diberi label menggunakan model yang telah dilatih sebelumnya. Dataset dikategorikan menjadi sentimen positif, sentimen negatif, dan sentimen netral. Model *Indo-BERT* digunakan sebagai kerangka analisis. Evaluasi hasil dilakukan dengan menerapkan matriks konfusi. Temuan dari percobaan ini menunjukkan bahwa model *Indo-BERT* memiliki kemampuan yang lebih unggul dengan tingkat akurasi mencapai 72% dalam memproses teks berbahasa Indonesia dibandingkan dengan model *BERT* dengan tingkat akurasi sebesar 69%.

**Kata kunci**—Sentiment Analysis, Verbal Violence Behaviour (VVB), *Indo\_Bert*, Twitter

### Abstract

The emergence of verbal violence on the social media platform Twitter has become an increasingly concerning issue in recent years. Verbal violence encompasses various forms of communication that are degrading, insulting, derogatory, and threatening, often harming individuals or groups. The primary issue addressed in this study is the identification related to verbal violence behaviour social media, particularly on Twitter. This research aims to develop sentiment analysis model of Verbal Violence Behaviour (VVB) on Twitter by using *Indo-BERT*. The accuracy of the model will be compared with those of *BERT*. This research commences with data collection through web crawling, utilizing the category of verbal violence behaviour as a

reference. Labelling datasets is carried out through a combination of manual and automated methods using a semi-supervised learning approach. This process involves self-learning, where unlabelled data is automatically labelled using a pre-trained model. Datasets are categorized into positive sentiment, negative sentiment, and neutral sentiment. The Indo-BERT Model is employed as the analytical framework. The evaluation of results is conducted by implementing a confusion matrix. The findings from this experiment indicate that the Indo-BERT model has superior capabilities with an accuracy rate of 72% in processing Indonesian text compared to the BERT model, which has an accuracy rate of 69%.

**Keywords**—Sentiment Analysis, Verbal Violence Behaviour (VVB), Indo\_Bert, Twitter

## 1. PENDAHULUAN

Media sosial menjadi salah satu sarana yang sangat efektif dalam menyebarkan informasi dan menyampaikan pesan kepada masyarakat. *Twitter*, sebagai salah satu *platform* media sosial populer, digunakan oleh banyak orang untuk berkomunikasi, berbagi pendapat, berdiskusi, dan mengungkapkan pandangan dan pemikiran tentang berbagai masalah sosial dan politik. Namun, penggunaan *Twitter* sering kali melibatkan konten negatif seperti pelecehan verbal. Pelecehan verbal dapat memicu emosi dan menimbulkan ketidaknyamanan pada pihak yang terlibat [1]. Oleh karena itu, penting untuk melakukan analisis sentimen terhadap konten *Twitter*, terutama terkait konten kekerasan verbal. Perilaku kekerasan verbal sering terjadi dalam bentuk komunikasi lisan yang melibatkan penghinaan, bahasa yang menghina, atau penghinaan [2]. Hal ini mencakup perilaku seperti kekerasan mental, sikap menyalahkan, pelabelan negatif, atau tuduhan yang tidak berdasar. Perilaku kekerasan verbal juga mencakup segala bentuk ujaran yang bersifat merendahkan, menyinggung, mengancam, dan tidak pantas, seperti mencoreng nama baik seseorang, menghina agama, memprovokasi, atau bahkan menyebarkan informasi palsu [3].

Perilaku kekerasan verbal di media sosial dapat menimbulkan berbagai permasalahan serius bagi individu dan masyarakat. Dampaknya meliputi masalah psikologis, ketidaknyamanan, *cyberbullying*, masalah kesehatan mental dan kesejahteraan, pelanggaran hukum, perpecahan sosial, dan pelanggaran privasi [4]. Salah satu permasalahan umum yang muncul dari perilaku kekerasan verbal di media sosial adalah *cyberbullying* [5]. Hal ini lebih mudah dilakukan karena pelaku tidak perlu berhadapan langsung dengan korbannya. Mereka dapat dengan mudah melakukan tindakan intimidasi melalui internet atau ponsel pintar tanpa melihat dampaknya terhadap orang lain. *Cyberbullying* sering terjadi di *platform* jejaring sosial seperti *Facebook*, *Twitter*, dan *Instagram* [6]. Hal ini memberikan dampak psikologis, khususnya pada anak-anak, karena korban *cyberbullying* dapat dijangkau oleh siapa saja dan kapan saja.

Salah satu metode yang efektif untuk analisis sentimen terhadap perilaku kekerasan verbal adalah model *BERT* (*Bidirect Encoder Representations from Transformers*), yang merupakan salah satu teknik pemrosesan bahasa alami (NLP) terbaru dan telah terbukti dapat melakukan tugas pemrosesan bahasa secara akurat dan efisien [7]. Di Indonesia, telah dikembangkan model *INDO-BERT* yang dirancang khusus untuk bahasa Indonesia. Model ini dapat secara efektif mengolah bahasa Indonesia dan menganalisis sentimen pada konten *Twitter* yang ditulis dalam bahasa Indonesia [8]. Untuk memastikan keakuratan dan efektivitas model *Indo-BERT* dalam menganalisis sentimen pada konten yang mengandung pelecehan verbal di *Twitter* dilakukan evaluasi menggunakan metode *Confusion Matrix*.

Penelitian yang pernah dilakukan oleh [9] menggunakan metode *Convolutional Neural Network* (*CNN*) untuk mengklasifikasikan sentimen pada data *Twitter* Indonesia dan mencapai akurasi klasifikasi sebesar 81,74%. Penelitian sebelumnya mengenai analisis sentimen dalam konteks media sosial di Indonesia masih terbatas, dan hanya sedikit penelitian yang berkonsentrasi pada pelecehan verbal. Misalnya penelitian [10] menggunakan model *BERT* untuk menganalisis sentimen terkait intimidasi di *Twitter* dan mencapai akurasi tinggi sebesar 81%. Namun penelitian tersebut tidak secara spesifik membahas pelecehan verbal. Oleh karena itu, penelitian ini bertujuan untuk melakukan analisis sentimen terhadap konten berbahasa Indonesia

yang mengandung pelecehan verbal di Twitter dengan menggunakan model *Indo-BERT*. Kinerja model *Indo-BERT* akan diukur dan dievaluasi dengan matriks konfusi. Penelitian ini diharapkan dapat memberikan wawasan baru mengenai penggunaan model *Indo-BERT* untuk analisis sentimen terhadap konten yang mengandung perilaku kekerasan verbal di Twitter.

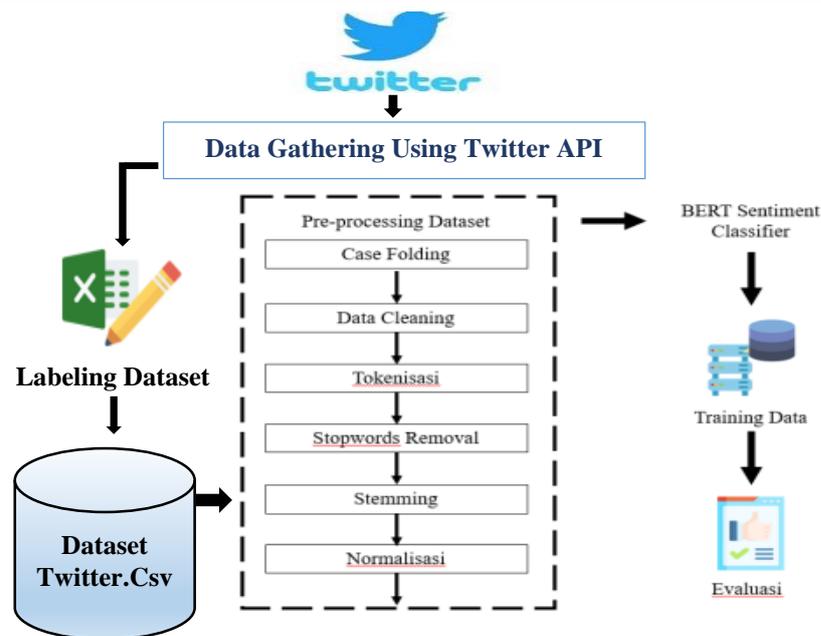
Penelitian di bidang NLP atau *Natural Language Processing* dengan menggunakan bahasa Indonesia belum banyak tersedia karena kurangnya sumber data yang dapat diakses, padahal bahasa Indonesia digunakan oleh sekitar 199 juta orang dan menduduki peringkat ke-11 sebagai salah satu bahasa yang paling umum digunakan bahasa di dunia pada tahun 2017 [11]. Oleh karena itu, model terlatih yang disebut *Indo-BERT* dikembangkan. *Indo-BERT* adalah variasi *BERT* (*Bidirect Encoder Representations from Transformers*) yang mengikuti model *BERT-Base* dengan 12 tugas, memiliki korpus lebih dari 220 juta kata yang sebagian besar bersumber dari Wikipedia bahasa Indonesia (74 juta kata), Kompas, Tempo, Artikel Liputan6 (total 55 juta kata), dan Indonesia *Web Corpus* (90 juta kata) [11]. *Indo-BERT* pertama kali diperkenalkan dalam sebuah makalah pada bulan September 2020, dengan pengumpulan data dari *Indo4B*, yang mencakup pra-pelatihan kontekstual tentang teks, blog, berita, dan situs web yang tersedia untuk umum [12]. [13] Eksperimen menunjukkan bahwa *Indo-BERT* mencapai kinerja optimal pada sebagian besar tugas di *INDOLEM*. Juga mencapai akurasi tertinggi dibandingkan penggunaan metode *KNN*, *SVM*, dan *Naïve Bayes*. Pada percobaan pertama terungkap bahwa *KNN* dan *SVM* hanya dapat mengklasifikasikan data ke dalam kelas mayoritas, dan *Naïve Bayes* dapat mengklasifikasikan kelas minoritas. Namun memiliki akurasi yang lebih rendah dibandingkan *Indo-BERT*, yaitu 0,3391 di kelas 1, 0,0863 di kelas 2, dan 0,2368 di kelas 3 [14].

Penelitian ini akan menggunakan metode *Indo-BERT* untuk melakukan analisis klasifikasi sentimen terhadap perilaku kekerasan verbal di media sosial Twitter. Dengan mempertimbangkan fitur-fitur seperti *tweet* dan label data, diharapkan dapat diperoleh klasifikasi sentimen perilaku kekerasan verbal yang tepat dan akurat. Data akan dikumpulkan secara selektif dari konten pengguna Twitter. Twitter dipilih sebagai sumber data karena statusnya sebagai platform media sosial terbesar kedua di dunia setelah *Facebook*, dan Indonesia menempati peringkat kelima basis pengguna Twitter terbesar secara global, setelah Brasil [15]. Kinerja model *Indo-BERT* akan diukur dan dievaluasi menggunakan matriks konfusi. Hasil penelitian ini akan memberikan wawasan baru mengenai analisis sentimen perilaku kekerasan verbal berbahasa Indonesia di Twitter. Penelitian ini diharapkan dapat memberikan pemahaman yang lebih baik tentang efektivitas penggunaan model *Indo-BERT* untuk analisis sentimen konten yang mengandung pelecehan verbal di media sosial.

## 2. METODE PENELITIAN

### 2.1 Alur Penelitian

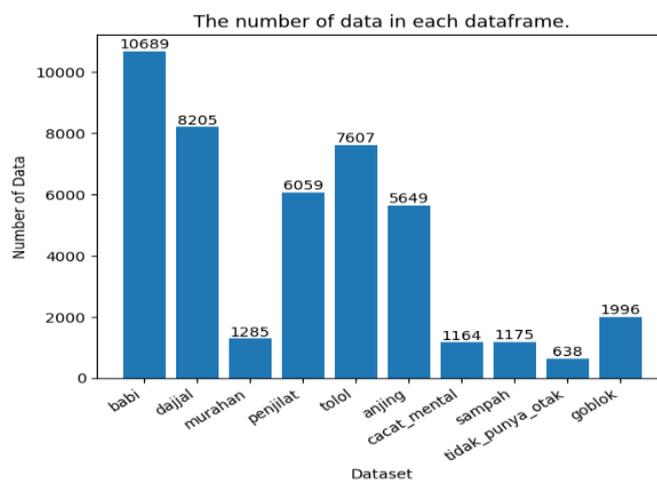
Tahapan penelitian ini terdiri dari empat langkah utama, dimulai dengan pengumpulan data dari *Twitter*, pelabelan data, *preprocessing* data, klasifikasi model, dan evaluasi. Gambar 1 merupakan diagram perancangan sistem yang digunakan dalam penelitian.



Gambar 1 Desain Alur Penelitian

## 2.2 Pengumpulan Data

Pengumpulan data dari *Twitter* dilakukan dengan metode *crawling* dengan menggunakan *library Python*. *Crawling* dilakukan dengan kategori kata kunci yang mengandung makna cacian, dalam rentang waktu 01 Maret 2024 hingga 01 Juni 2024. Total data yang diperoleh berjumlah 44.467 entri, didistribusikan berdasarkan kata kunci pencarian yang ditentukan seperti pada Gambar 2. Dokumentasi hasil *crawling* disimpan dalam format file CSV.



Gambar 2 Distribusi Jumlah dataset

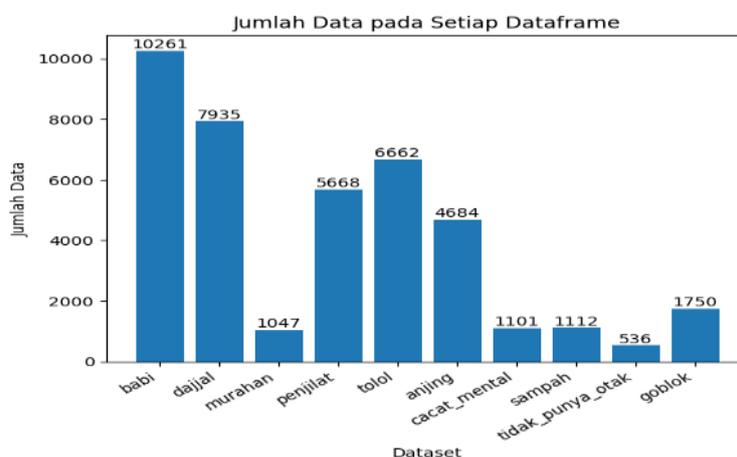
## 2.3 Pelabelan Data

Proses pelabelan data dilakukan secara manual dan otomatis. Pelabelan manual dilakukan pada data sampel dari seluruh kumpulan data. Data tersebut kemudian diberi label oleh tiga *anotator* berbeda sesuai dengan kriteria berikut: 1) Data dengan sentimen negatif diberi label "2", 2) Data dengan sentimen positif diberi label "1", dan 3) Data dengan sentimen netral diberi label sebagai "0". Proses pelabelan data otomatis dilakukan dengan menggunakan salah satu pendekatan dalam proses *Semisupervised Learning* untuk analisis sentimen. Pada metode ini, data yang tidak berlabel akan otomatis diberi label menggunakan model yang telah dilatih sebelumnya.

Proses ini memanfaatkan prediksi sentimen yang dipelajari model dari data berlabel. Penerapan teknik ini bertujuan untuk menggunakan data secara efisien dan meningkatkan jumlah data berlabel untuk melatih model analisis sentimen.

### 2.3 Data Preprocessing

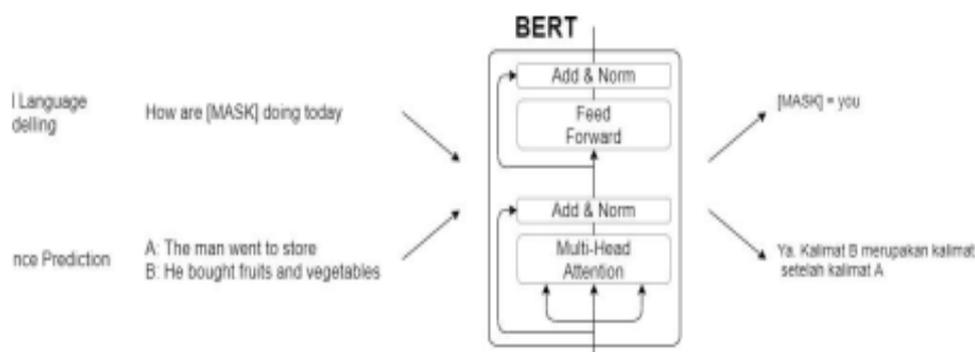
Sebelum data digunakan untuk melatih model *Supervised Learning*, penting untuk menjalani langkah-langkah pra-pemrosesan data untuk memastikan kualitas dan konsistensi data. Pada tahap ini dilakukan penghapusan *stop word* dan penggantian kata *slang* menggunakan *NLP Library* untuk menghilangkan kata-kata yang tidak relevan dan mengganti kata-kata *slang* dengan bentuk baku yang sesuai. Selain itu, *stemming* dilakukan dengan menggunakan *stemmer Sastrawi* untuk mentransformasikan kata-kata dalam *dataset* menjadi bentuk dasar atau akar kata, mengatasi variasi kata yang mirip, sehingga kata yang memiliki akar kata yang sama dianggap sebagai satu kesatuan. Hasil *pre-processing* data dapat dilihat pada Gambar 3. Jumlah *dataset* mengalami penurunan menjadi 40.756 dari sebelumnya berjumlah 44.467 data, hal ini disebabkan adanya proses pembersihan data pada tahap *preprocessing* data ini.



Gambar 3 Distribusi *dataset* setelah data *preprocessing*

### 2.4 Model Klasifikasi

Peneliti mengadopsi model *Indo-BERT* sebagai kerangka analisis. Keputusan ini didasarkan pada fakta bahwa *Indo-BERT* telah dilatih lebih dari 220 juta kata bahasa Indonesia. Setelah tahap pra-pemrosesan data awal, kumpulan data akan diubah ke dalam format yang kompatibel dengan *Indo-BERT*, yang melibatkan representasi kata dalam format vektor menggunakan alat *Indo-BERT Tokenizer*. Langkah berikutnya melibatkan penyempurnaan *hyperparameter*, dimana model *Indo-BERT* yang telah dilatih sebelumnya disesuaikan untuk melakukan tugas klasifikasi sentimen. Dengan metode *BERT*, suatu model bahasa dapat membaca dua arah secara bersamaan. *BERT* dirancang untuk membantu komputer memahami arti bahasa ambigu dalam teks dengan menggunakan teks di sekitarnya untuk menetapkan konteks melalui dua tugas *BERT* yang telah dilatih sebelumnya, *Masked Language Model (MLM)* dan *Next Sentence Prediction (NSP)*

Gambar 4 Proses *Pre-training* pada BERT

Gambar 4 memperlihatkan proses *training BERT*, model dapat menerima pasangan kalimat sebagai input dan dilatih untuk memprediksi jika kalimat kedua pada pasangan tersebut adalah kalimat berikutnya pada dokumen aslinya atau hanya satu kalimat saja. Selama *training*, 50% dari input adalah pasangan kalimat dimana kalimat kedua adalah kalimat berikutnya pada dokumen asli. Sedangkan 50% lainnya adalah kalimat yang diambil secara acak dari *corpus* sebagai kalimat kedua.

### 2.5 Evaluasi

Tahap evaluasi dirancang untuk menilai hasil pelabelan berbasis pembelajaran yang diawasi dan analisis sentimen terhadap kalimat-kalimat dalam kumpulan data. Saat menganalisis sentimen pelecehan verbal menggunakan model *Indo-BERT*, matriks konfusi dapat digunakan untuk menilai kinerja model dalam mengkategorikan komentar sebagai positif, netral, atau negatif. Nilai akurasi tertinggi yang diperoleh pada proses pelatihan sebelumnya akan dijadikan sebagai akurasi model. Untuk memperoleh prediksi dari model digunakan matriks konfusi seperti yang digambarkan pada Tabel 1.

Tabel 1 *Confusion Matrix*

		<i>True Class</i>		
		<i>Positive</i>	<i>Neutral</i>	<i>Negative</i>
<i>Predicted Class</i>	<i>Positive</i>	<i>True Positive (TP)</i>	<i>False Positive (FP)</i>	<i>False Positive (FP)</i>
	<i>Neutral</i>	<i>False Neutral (FNt)</i>	<i>True Neutral (TNt)</i>	<i>False Neutral (FNt)</i>
	<i>Negative</i>	<i>False Negative (FN)</i>	<i>False Negative (FN)</i>	<i>True Negative (TN)</i>

Empat indikator yang dihasilkan oleh *Confusion Matrix* yaitu *Accuracy*, *Precision*, *Recall*, dan *F1 Score* atau *F-Measure*. Dalam perhitungan evaluasi tetap memperhitungkan nilai-nilai seperti yang telah dijelaskan sebelumnya. Persamaan matriks konfusi dapat dilihat pada rumus (1) sampai dengan rumus (4).

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \quad (1)$$

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

$$F1\ Score = \frac{2 \times Recall \times Precision}{Recall + Precision} \quad (4)$$

### 3. HASIL DAN PEMBAHASAN

Penelitian ini dilakukan dengan menggunakan 40.756 *dataset* yang diperoleh dari *Crawling konten data Twitter*. Untuk memastikan bahwa seluruh *dataset* memiliki struktur yang

seragam dan memudahkan proses pelatihan model *Indo-BERT* dilakukan proses *shaping dataset* dengan menggunakan fungsi *shape\_dataframe*, *Code*. Data yang telah di *shape* mempunyai struktur kolom *Tweet* dan label, dimana pada kolom label nantinya akan digunakan untuk menyimpan data sentimen yang di berikan secara manual maupun otomatis. Pada proses *labeling dataset* menggunakan model *Supervised-Learning*, karena porsi *dataset* sangat tidak berimbang, dan kategori *dataset* terlalu banyak, diputuskan untuk menambahkan data dengan *sentimen* positif ke *dataset* karena *sentiment* positif memiliki jumlah data yang lebih sedikit. Selain itu, juga dilakukan *fine-tuning* pada *hyperparameter* model untuk memastikan bahwa model memiliki performa yang optimal dalam melakukan analisis sentimen pada *dataset* yang sangat beragam ini. *Dataset* dibagi menjadi 2 data yaitu 80 % data latih dan 20% *data tesing* yang bertujuan memberikan data latih yang cukup untuk klasifikasi multi kelas. Langkah selanjutnya adalah membangun model untuk mengklasifikasikan perilaku kekerasan verbal menggunakan metode *Indo-BERT*.

Metode *Indo-BERT* sendiri mempunyai dua mekanisme yaitu *encoder* dan *decoder*. *Encoder* bertanggung jawab untuk membaca masukan, sedangkan *decoder* menghasilkan prediksi. Selain itu, karena *Indo-BERT* dapat membaca teks dua arah, model ini dapat belajar dari konteks sekitarnya. Model *Indo-BERT* membutuhkan *resource* yang tinggi, perlu dibuat beberapa optimalisasi seperti penggunaan *data loader* agar menghemat penggunaan memori saat pelatihan, sehingga seluruh *dataset* tidak dimuat secara bersamaan dalam memori. Untuk tujuan ini, digunakan juga *\*blocks\** untuk membuat *data loader* yang akan menghasilkan komentar komentar yang sudah di *tokenisasi*. Komentar dan sentimen memiliki panjang maksimal 130 kata. Untuk menganalisis sentimen, digunakan *\*layer\** tambahan, termasuk *\*dropout layer\** dengan probabilitas 0.1. Pada penelitian ini dilakukan *\*fine-tuning\** dengan mengatur *hyperparameter* seperti berikut:

- 1 *Epoch* : 10
- 2 *Learning rate* :  $1e-3$
- 3 *Batch size* : 32

Pemilihan *hyperparameter* dengan *Epoch* sebesar 10 menunjukkan bahwa seluruh *dataset* pelatihan akan dilewati oleh model. Ini berarti model akan memperbarui bobotnya sebanyak 10 kali iterasi. Selanjutnya, *learning rate* sebesar  $1e-3$  mengindikasikan seberapa besar langkah perubahan bobot yang diambil oleh model saat melakukan pembaruan berdasarkan *gradient* dari fungsi *loss*. Semakin kecil *learning rate*, pembaruan menjadi lebih halus dan stabil, tetapi bisa memerlukan lebih banyak iterasi untuk mencapai konvergensi. Terakhir, *batch size* sebesar 32 menentukan berapa banyak sampel data yang digunakan dalam setiap iterasi saat melatih model. *Batch size* yang lebih besar dapat mempercepat proses pelatihan karena lebih banyak sampel yang dievaluasi sekaligus, tetapi juga dapat memakan lebih banyak memori

Percobaan dilakukan sebanyak tiga kali. Untuk mencapai hasil optimal, jumlah *epoch* diatur menjadi 10 kali. Hasil akurasi dari ketiga percobaan menunjukkan tingkat akurasi masing-masing sebesar 71%, 72%, dan 72% seperti terlihat pada Gambar 5, Gambar 6, dan Gambar 7. Penurunan performa tersebut kemungkinan disebabkan oleh kecenderungan *overfitting* yang disebabkan oleh jumlah *epoch* yang tinggi. *Overfitting* terjadi ketika model menjadi terlalu fokus pada data pelatihan dan kehilangan kemampuan untuk menggeneralisasi data yang tidak terlihat. Fenomena ini mengakibatkan kemampuan prediksi menjadi kurang akurat dan kurang konsisten.

	precision	recall	f1-score	support
netral	0.55	0.49	0.52	1261
positif	0.65	0.63	0.64	676
negatif	0.78	0.82	0.80	2883
accuracy			0.71	4820
macro avg	0.66	0.65	0.65	4820
weighted avg	0.70	0.71	0.70	4820

Gambar 5 Tingkat akurasi percobaan 1

Pada percobaan pertama dilakukan dengan menerapkan nilai parameter *learning rate*  $2e-2$ , *dropout*: 0.002, dan *batch size*:8 menghasilkan tingkat akurasi sebesar 0.71 atau 71%.

	precision	recall	f1-score	support
Netral	0.57	0.53	0.55	1261
Positif	0.66	0.62	0.64	676
Negatif	0.79	0.83	0.81	2883
accuracy			0.72	4820
macro avg	0.68	0.66	0.67	4820
weighted avg	0.72	0.72	0.72	4820

Gambar 6 Tingkat akurasi percobaan 2

Pada percobaan pertama dilakukan dengan menerapkan nilai parameter *learning rate*  $2e-3$ , *dropout*: 0.003, dan *batch size*:16 menghasilkan tingkat akurasi sebesar 0.72 atau 72%.

	precision	recall	f1-score	support
netral	0.59	0.50	0.54	1261
positif	0.64	0.64	0.64	676
negatif	0.78	0.84	0.81	2883
accuracy			0.72	4820
macro avg	0.67	0.66	0.66	4820
weighted avg	0.71	0.72	0.71	4820

Gambar 7 Tingkat akurasi percobaan 3

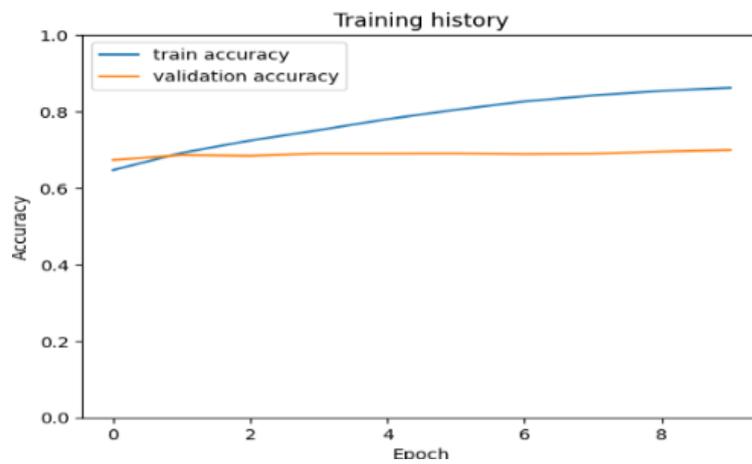
Pada percobaan pertama dilakukan dengan menerapkan nilai parameter *learning rate*  $2e-5$ , *dropout*: 0.004, dan *batch size*:32 menghasilkan tingkat akurasi sebesar 0.72 atau 72%.

Dalam penelitian ini dilakukan perbandingan antara model *BERT Uncased* yang merupakan model dasar *Indo-BERT* dan memiliki kinerja yang sangat baik dalam analisis dan klasifikasi sentimen. Sebagai perbandingan, percobaan dilakukan dengan menggunakan data dan parameter yang sama dengan model *BERT*. Gambar 8 menunjukkan hasil yang diperoleh dari model *BERT* dengan tingkat akurasi sebesar 69%, lebih rendah dibandingkan dengan akurasi yang diperoleh model *Indo-BERT* seperti terlihat pada Gambar 7, dengan tingkat akurasi sebesar 72%.

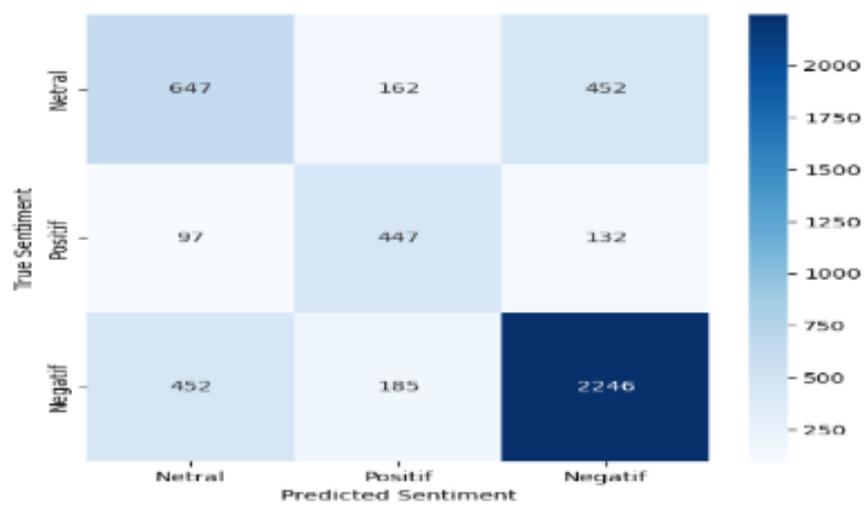
	precision	recall	f1-score	support
Netral	0.54	0.51	0.53	1261
Positif	0.56	0.66	0.61	676
Negatif	0.79	0.78	0.79	2883
accuracy			0.69	4820
macro avg	0.63	0.65	0.64	4820
weighted avg	0.70	0.69	0.69	4820

Gambar 8 Tingkat akurasi percobaan BERT

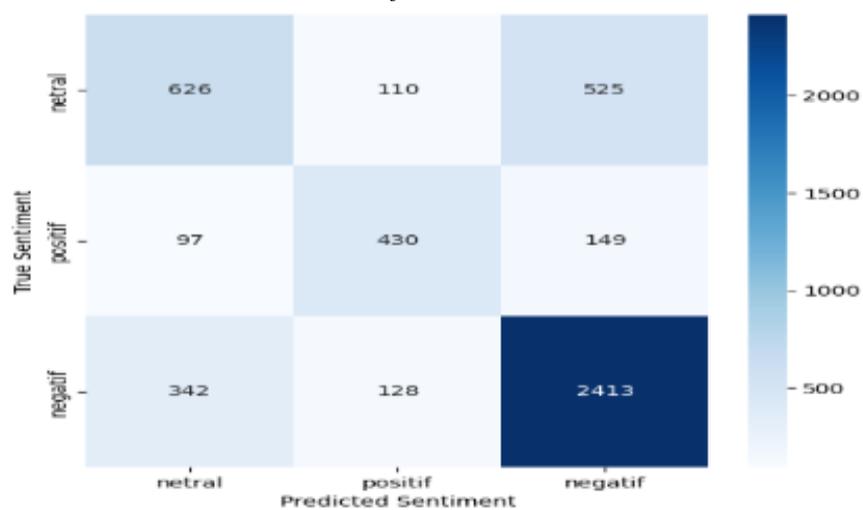
Hal ini menunjukkan bahwa *Indo-BERT* bekerja lebih baik dalam menganalisis sentimen dalam bahasa Indonesia. Selain itu, melihat matriks konfusi pada Gambar 10, model *BERT* dengan tepat memperkirakan 2.246 sentimen negatif, 447 sentimen positif, dan 647 sentimen netral. Sebaliknya, *Indo-BERT* memperoleh 2.413 sentimen negatif, 430 sentimen positif, dan 626 sentimen netral, seperti terlihat pada Gambar 11. Kurva kinerja pelatihan ditunjukkan pada Gambar 9.



Gambar 9 Kurva kinerja pelatihan



Gambar 10 Confusion matrix BERT



Gambar 11 Confusion matrix Indo-BERT

#### 4. KESIMPULAN

Berdasarkan hasil pengujian analisis sentimen terkait kekerasan verbal dalam bahasa Indonesia dengan menggunakan model *Indo-BERT*, terlihat bahwa proses pelabelan dengan metode *Supervised Learning* efektif untuk data yang sebarannya seimbang. Hal ini memungkinkan model mempelajari pola sentimen secara efektif di setiap kategori. Keberhasilan dalam analisis sentimen sangat bergantung pada langkah-langkah pemrosesan dan pengelolaan data yang cermat. Evaluasi model *Indo-BERT* menunjukkan kemampuannya yang kuat dalam analisis sentimen dalam bahasa Indonesia dengan tingkat akurasi 72% yang dihasilkan dari percobaan ke tiga dengan nilai parameter *learning rate*  $2e-5$ , *dropout*: 0.004, dan *batch size*:32, dibandingkan model *BERT* dengan tingkat akurasi 69%. Meskipun terdapat beberapa tantangan dalam kategori sentimen tertentu, model ini secara konsisten memberikan kinerja yang stabil. Keandalan model ini dalam memahami pola sentimen patut diperhatikan, karena *Indo-BERT* dapat membedakan pola sentimen bahasa sehari-hari secara efektif. Berdasarkan hasil evaluasi, model *Indo-BERT* memiliki potensi besar untuk mendukung berbagai aplikasi, seperti pemantauan opini publik, analisis risiko, dan pemahaman konten media sosial.

#### 5. SARAN

Untuk lebih meningkatkan akurasi pada model, penelitian kedepan dapat melakukan *tuning hyperparameter* untuk mengoptimalkan *hyperparameter model Indo-BERT* guna meningkatkan akurasi dan efisiensi. Menggunakan teknik *cross-validation* yang lebih komprehensif untuk memastikan model tidak *overfitting* dan performa yang konsisten.

#### UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Tim Redaksi Jurnal Teknik Politeknik Negeri Sriwijaya yang telah memberi kesempatan, sehingga artikel ilmiah ini dapat diterbitkan.

#### DAFTAR PUSTAKA

- [1] A. D. Vega, H. Hapidin, and K. Karnadi, "Pengaruh Pola Asuh dan Kekerasan Verbal terhadap Kepercayaan Diri (Self-Confidence)," *J. Obsesi J. Pendidik. Anak Usia Dini*, vol. 3, no. 2, p. 433, Jul. 2019, doi: 10.31004/obsesi.v3i2.227.
- [2] M. A. Ahdiyati, "Kekerasan Verbal di Konten Youtube Indonesia Dalam Perspektif Kultivasi," *J. Commun.*, vol. 5, no. 2, 2020.
- [3] F. Wibowo and R. B. Parancika, "Kekerasan Verbal (Verbal Abuse) di Era Digital Sebagai Faktor Penghambat Pembentukan Karakter," Apr. 2018, Accessed: Mar. 28, 2024. [Online]. Available: <http://publikasiilmiah.ums.ac.id/handle/11617/9908>
- [4] Desmana, "KESEHATAN MENTAL REMAJA KORBAN KEKERASAN VERBAL DI DESA LUBUK LADUNG KECAMATAN KEDURANG ILIR KABUPATEN BENGKULU SELATAN," diploma, UIN Fatmawati Sukarno Bengkulu, 2022. Accessed: Mar. 28, 2024. [Online]. Available: <http://repository.iainbengkulu.ac.id/8947/>
- [5] W. N. H. Wan Ali, M. Mohd, F. Fauzi, K. Shirai, and M. J. Mahamad Noor, "IMPLEMENTATION OF HYPERPARAMETER OPTIMISATION AND OVER-SAMPLING IN DETECTING CYBERBULLYING USING MACHINE LEARNING APPROACH," *Malays. J. Comput. Sci.*, pp. 78–100, Dec. 2021, doi: 10.22452/mjcs.sp2021no2.6.
- [6] C. M. Castorena, I. M. Abundez, R. Alejo, E. E. Granda-Gutiérrez, E. Rendón, and O. Villegas, "Deep Neural Network for Gender-Based Violence Detection on Twitter Messages," *Mathematics*, vol. 9, no. 8, p. 807, Apr. 2021, doi: 10.3390/math9080807.

- 
- [7] A. Syarif, R. Z. A. Aziz, and A. Ulvan, "Sentiment Analysis Model of Verbal Violence Behaviour on Twitter by Using *Indo-BERT*".
- [8] Steven Dharmawan, Viny Christanti Mawardi, and Novario Jaya Perdana, "Klasifikasi Ujaran Kebencian Menggunakan Metode FeedForward Neural Network (*Indo-BERT*)," *J. Ilmu Komput. Dan Sist. Inf.*, vol. 11, no. 1, Jun. 2023, doi: 10.24912/jiksi.v11i1.24066.
- [9] Alon Jacovi, A. Jacovi, Oren Sar Shalom, O. S. Shalom, Yoav Goldberg, and Y. Goldberg, "Understanding Convolutional Neural Networks for Text Classification," *ArXiv Comput. Lang.*, Sep. 2018, doi: 10.18653/v1/w18-5408.
- [10] T. A. S. Rohmah and W. Maharani, "Personality Detection on Twitter Social Media Using *Indo-BERT* Method," *Build. Inform. Technol. Sci. BITS*, vol. 4, no. 2, pp. 448–453, Sep. 2022, doi: 10.47065/bits.v4i2.1895.
- [11] H. Weissbart and A. E. Martin, "The Structure and Statistics of Language jointly shape Cross-frequency Dynamics during Spoken Language Comprehension," 2023.
- [12] E. Smyrnova-Trybulska and E. Smyrnova-Trybulska, *E-learning in the Transformation of Education in Digital Society*, 1st ed., vol. 14. in E-learning, vol. 14. STUDIO NOA, 2022. doi: 10.34916/el.2022.14.
- [13] F. Koto, A. Rahimi, J. H. Lau, and T. Baldwin, "IndoLEM and *Indo-BERT*: A Benchmark Dataset and Pre-trained Language Model for Indonesian NLP," in *Proceedings of the 28th International Conference on Computational Linguistics*, Barcelona, Spain (Online): International Committee on Computational Linguistics, 2020, pp. 757–770. doi: 10.18653/v1/2020.coling-main.66.
- [14] D. J. M. Pasaribu, K. Kusrini, and S. Sudarmawan, "Peningkatan Akurasi Klasifikasi Sentimen Ulasan Makanan Amazon dengan Bidirectional LSTM dan Bert Embedding," *Inspir. J. Teknol. Inf. Dan Komun.*, vol. 10, no. 1, Jun. 2020, doi: 10.35585/inspir.v10i1.2568.
- [15] S. Senbel, C. Seigel, and E. Bryan, "Religious Violence and Twitter: Networks of Knowledge, Empathy and Fascination," *Religions*, vol. 13, no. 3, p. 245, Mar. 2022, doi: 10.3390/rel13030245.