

PAPER NAME

Analisis Sentimen Terhadap Kemajuan Kecerdasan Buatan di Indonesia Menggunakan BERT dan RoBERTa

AUTHOR

**NISRINA AKBAR RIZKY PUTRI,
ARDIANSYAH**

WORD COUNT

2234 Words

CHARACTER COUNT

14503 Characters

PAGE COUNT

8 Pages

FILE SIZE

838.4KB

SUBMISSION DATE

May 17, 2023 10:58 AM GMT+7

REPORT DATE

May 17, 2023 10:59 AM GMT+7

● **12% Overall Similarity**

The combined total of all matches, including overlapping sources, for each database.

- 9% Internet database
- 2% Publications database
- Crossref database
- Crossref Posted Content database
- 9% Submitted Works database

Analisis Sentimen Terhadap Kemajuan Kecerdasan Buatan di Indonesia Menggunakan BERT dan RoBERTa

Nisrina Akbar Rizky Putri¹⁾, Ardiansyah²⁾

1)2) Program Studi Teknologi Informasi, Fakultas Kesehatan Dan Teknologi,
Universitas Muhammadiyah Klaten

Jl. Ir. Soekarno Km.1 Buntalan, Klaten, Jawa Tengah

¹⁾ nisrinaakbar@umkla.ac.id

²⁾ ardiansyah@umkla.ac.id

Abstrak

Kemajuan dalam bidang kecerdasan buatan (AI) dalam beberapa tahun belakang menjadi sesuatu yang sering dibahas dikalangan masyarakat di Indonesia. Berbagai prespektif muncul hampir disemua platform sosial media untuk menyampaikan tanggapan terhadap suatu topik. Pada penelitian yang dilakukan memanfaatkan analisis sentimen sehingga dapat menginterpretasi sentimen pada masyarakat dengan memanfaatkan komentar masyarakat di kolom komentar YouTube sebagai objek penelitian dikarekan relevansi dengan topik sebesar 60-80%. Selain itu, peneliti menggunakan *fine-tuning* BERT dan RoBERTa sebagai labeling sentimen negatif, netral ataupun positif pada setiap komentar yang telah dikumpulkan menggunakan Python serta *library BeautifulSoup* dan *selenium*. Dataset yang dikumpulkan sebanyak 5796. Namun, setelah dilakukan preprosesing dataset menjadi 5788 yang terbagi menjadi 3028, 1435, 1333 dari sentimen negatif, positif, dan netral untuk labeling menggunakan RoBERTa. Sedangkan, BERT yaitu 3106, 1422, 1268 dari sentimen negatif, positif, dan netral. Pada penelitian ini juga mendapatkan hasil bahwa model indobenchmark/indobert-base-p1 dapat digunakan untuk labeling dataset berbahasa Indonesia. Selain itu, training model dengan skema 1 persebaran dataset 70:30:30 dan skema 2 adalah 80:20:20, menjadi model yang memiliki kinerja sebesar 84% untuk akurasi validation dan 83% akurasi testing pada skema 1. 83% akurasi validasi 84% akurasi tesing pada skema 2. Selanjutnya, peneliti menyarankan untuk adanya keterlibatan stackholder terkait penggunaan AI di Indonesia.

Kata kunci: BERT, RoBERTa, Transformers, Fine-Tuning, Analisis Sentimen

Abstract

The rise of artificial intelligence (AI) in recent years has often been discussed among people in Indonesia. Various perspectives appear on almost all social media platforms to convey responses to a topic. The research conducted leverage sentiment analysis obtained to interpret sentiment in the community by utilizing public comments in the YouTube comment column as the object of research because of the relevance to the topic of 60-80%. In addition, researchers use fine-tuning BERT and RoBERTa to label negative, neutral, or positive sentiments on each comment that has been collected using Python and BeautifulSoup, and selenium libraries. The dataset gathered was 5796. However, after preprocessing the dataset becomes 5788 divided into 3028, 1435, and 1333 to negative, positive, and neutral sentiments for labeling using RoBERTa. Additionally, BERTs are 3106, 1422, and 1268 from negative, positive, and neutral sentiments. This research also found that the indobenchmark/indobert-base-p1 allows labeling Indonesian datasets. In addition, the training model with scheme 1 dataset distribution is 70:30:30 and Scheme 2 is 80:20:20, being a model that has a performance of 84% validation accuracy and 83% testing accuracy in Scheme 1. 83% validation accuracy and 84% testing accuracy in scheme 2. Furthermore, researchers suggest the involvement of stakeholders regarding the use of AI in Indonesia.

Keywords: BERT, RoBERTa, Transformers, Fine-Tuning, Sentiment Analysis

1. PENDAHULUAN

Kecerdasaan buatan (*AI*) merupakan sistem *Machine-Based* yang mampu membuat sebuah rekomendasi, prediksi, diagnosa yang sama seperti manusia [1], [2]. Beberapa tahun belakangan, dampak dari kemajuan yang terjadi pada *AI* telah memasuki banyak bidang seperti Bisnis [3], Agrikultur [4], Kesehatan [5], Manufaktur, Pendidikan [6], dan lainnya [7]. Meskipun, *AI* telah memberikan banyak manfaat pada berbagai aspek kehidupan masyarakat tetapi *AI* masih menjadi pro dan kontra terkait kemajuan *AI*. Perbedaan pandangan terhadap kecepatan inovasi *AI* seringkali ditemukan diberbagai kolom komentar platform media sosial seperti Twitter, Facebook, Tiktok bahkan YouTube [8]. Sehingga, media sosial menjadi wadah masyarakat menanggapi terkait situasi yang sedang terjadi [9]. Platform YouTube menjadi wadah yang cukup baik untuk masyarakat dalam berpendapat terhadap suatu masalah, 60-80% komentar berkaitan dengan topik yang sedang dibahas [10].

Dengan adanya perbedaan pandangan terhadap suatu masalah yang muncul, sentimen analisis seringkali digunakan [11] untuk menentukan interpretasi masyarakat terkait masalah yang sedang hangat dibahas [12]. Analisis sentimen merupakan bidang Natural Language Processing (NLP) yang mampu mempelajari komentar atau ulasan dari masyarakat yang berupa teks [13]. Pembelajaran yang dilakukan analisis sentimen akan memberikan polaritas teks yang dilatih menjadi positif, netral dan negatif [14]. Inovasi yang terjadi pada NLP tanpa terkecuali pada bagian sentimen analisis yaitu model pre-trained BERT [15]. Selain itu, BERT sangat kompetibel untuk melakukan labeling pada dataset [16].

Penelitian sentimen analisis yang memanfaatkan komentar pada YouTube yang dilakukan [17] memperoleh hasil sentimen negatif lebih banyak dibandingkan sentimen positif maupun netral. Penelitian lainnya yang telah dilakukan [9] terkait sentimen masyarakat terhadap *METAVVERSE* menggunakan Support Vector Machine (SVM). *METAVVERSE* merupakan hasil inovasi dari AI yang menawarkan kemudahan dalam berbagai aspek kehidupan. Pada penelitian tersebut mendapatkan hasil 1772, 397, 337 dari sentimen positif, netral dan negatif. Namun, pada penelitian tersebut tidak menjelaskan bagaimana peneliti melakukan labeling dataset yang digunakan. Selain itu, penelitian mendapatkan hasil akurasi sebesar 81% setelah dataset yang digunakan melalui proses preprocessing.

Penelitian [18] membandingkan antara SVM dan BERT yang menggunakan dataset dari hasil crawling di twitter berbahasa Indonesia. Dataset yang dikumpulkan sebanyak 2624 sedangkan labeling yang dibagi menjadi 2 kategori 0 dan 1. Pada penelitian tersebut memperoleh hasil akurasi 94% untuk BERT, sedangkan SVM mendapatkan hasil 93%. Penelitian [19] bertujuan mengklasifikasi judul berita clickbait atau bukan menggunakan RoBERTa dan BERT dengan dataset CLICK-ID memperoleh hasil Indobenchmark/IndoBERT-p1 sebagai model dengan kinerja yang terbaik untuk judul berita berbahasa Indonesia.

Penelitian sebelumnya telah memberikan gambaran terkait penggunaan analisis sentimen guna memberikan interpretasi tanggapan masyarakat. Sehingga penelitian yang akan dilakukan berfokus pada bagaimana sentimen masyarakat menanggapi kemajuan AI di Indonesia menggunakan BERT dan RoBERTa model serta membandingkan model yang terbaik untuk penggunaan teks klasifikasi pada bahasa Indonesia. Penelitian ini juga memanfaatkan media YouTube sebagai media pengumpulan data dikarenakan relevansi komentar yang diberikan masyarakat memiliki presentasi yang besar dibandingkan dengan platform lainnya.

2. TINJAUAN PUSTAKA

2.1 Sentimen Analisis

Analisis sentiment merupakan *task* dari NLP yang memiliki tujuan untuk menganalisa opini, sentiment, tingkah laku, atau pandangan masyarakat luas yang berupa teks sehingga, sering disebut dengan penggalian opini [20]. Opini masyarakat menjadi informasi yang sangat penting dan mudah ditemukan dikarenakan seringkali masyarakat meninggalkan pandangannya diberbagai platform ataupun forum. Pemanfaatan analisis sentiment tidak hanya dilakukan guna penelitian saja sehingga analisis sentimen mengalami peningkatan semenjak tahun 2004. Tetapi perusahaan atau pemerintahan menggunakan analisis sentimen sebagai pertimbangan

pengambilan keputusan ataupun mendapatkan opini masyarakat terkait produk dari perusahaan [20].

2.2 BERT dan RoBERTa

Bidirectional Encoder Representation from Transformers (BERT) merupakan bagian *encoder-decoder* transformer model [21], [22] yang telah dilatih untuk mengerti suatu bahasa dengan kumpulan data teks yang sangat besar atau yang sering disebut korpus [15]. *Pre-train* BERT menggunakan 2 *tasks unsupervised* yaitu *Masked Language Modeling* (MLM) dan *Next Sentence Prediction* (NSP) [23]. Selanjutnya, tahapan *Fine-Tuning* yang bertujuan agar BERT dapat menyelesaikan masalah yang lebih spesifik [15]. *Robustly Optimized BERT Pretraining Approach* (RoBERTa) merupakan pengembangan dari BERT yang pada *Pre-train* menggunakan *Dynamic Masking*, menghilangkan NPS, melatih data dengan kalimat yang Panjang, menggantikan pola *Mask* secara dinamis saat pelatihan [24].

2.3 Evaluasi

Pada penelitian yang dilakukan pada tahapan evaluasi yaitu *Confusion Matrix* untuk mengukur akurasi klasifikasi atau model [25] seperti yang terlihat pada gambar 1.

		True Class	
		Positive	Negative
Predictive Class	Positive	True Positive (TP)	False Positive (FP)
	Negative	False Negative (FN)	True Negative (TN)

Gambar 1. *Confusion Matrix*

Selain itu, penggunaan perhitungan akurasi, *Precision*, *Recall*, dan *F1-Score* sebagai metode evaluasi model berdasarkan *Confusion Matrix* yang telah didapatkan. Berikut persamaan (1) akurasi, persamaan (2) *Precision*, persamaan (3) *Recall*, persamaan (4) *F1-Score*.

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

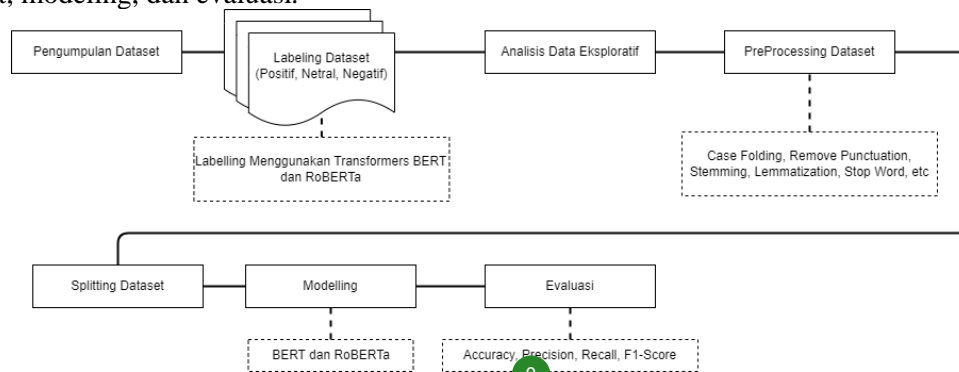
$$\text{Precision} = \frac{TP}{TP+FP} \quad (2)$$

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

$$f1 - score = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} = 2 * \frac{2TP}{2TP+FP+FN} \quad (4)$$

3. METODE PENELITIAN

Metode penelitian yang dilakukan terlihat pada gambar 2, yang dimana ada beberapa tahapan yaitu pengumpulan data, labeling dataset, Analisis data eksploratif, preprosesing dataset, splitting dataset, modeling, dan evaluasi.



Gambar 2. Alur penelitian

3.1 Pengumpulan Data

Tahapan pengumpulan data pada penelitian menggunakan metode *crawling* komentar pada video dan short YouTube yang membahas terkait AI. Proses *crawling* dilakukan menggunakan bahasa pemrograman Python serta *library BeautifulSoup* dan *Selenium*. Proses pengumpulan dataset yang berhasil didapatkan sebanyak 5796 data terlihat pada tabel 1. Namun, peneliti tidak mengambil secara keseluruhan komentar. Komentar yang berupa balasan terhadap komentar lain tidak diambil sebagai dataset untuk menjaga relevansi komentar dengan topik [10] seperti yang terlihat pada tabel 2.

Tabel 1. Contoh dataset hasil crawling

Index	Komentar
1	Sesuai motto terkenal "programmer adalah pekerjaan yang menghapus pekerjaan orang lain" Jalan tol sudah pake, kartu, makan bisa pesen di hp, dll
2	ChatGPT aja udah serem bang. Di satu sisi senang dengan perkembangan teknologi udah se hebat ini, tapi disisi lain rada khawatir juga.
3	Justru adanya AI membantu bgt, apalagi udah integrasi antara Microsoft office dg Bing + GPT-4. Semua kerjaanku bikin laporan administrasi udah bisa otomatis ga perlu capek-capek mikir, tinggal ngoreksi doang. Trus aku yg dulu coding butuh waktu 2-3 hari, sekarang pake chatGPT cuman 2 jam kelar. Intinya aku pribadi merasa kebantu banget dg AI karna ningkatin kinerja dan produktivitas. Semoga semua industri bisa memanfaatkan AI dg maksimal dan mendapatkan lebih banyak keuntungan.
4	Lantas para pelayan restoran udah gada dong berarti ko?
5	Akan banyak orang-orang kehilangan pekerjaan kecuali pekerjaan dengan hati dan intuisi

Tabel 2. Contoh komentar balasan pada youtube

Komentar	Komentar Balasan
semoga anggota DPR kedepannya bisa di kerjakan chatGPT	Hemat anggaran dan mengurangi atu menghilangkan angka koruptor . 🙏 Sangat cerdas semoga dan semoga Setuju! Jadi menghapus Korupsi/Kolusi/Non Produktif. Mari dukung Pencalonan AI jadi Ketua & Anggota DPR/MPR RI!!!

3.2 Labeling Dataset

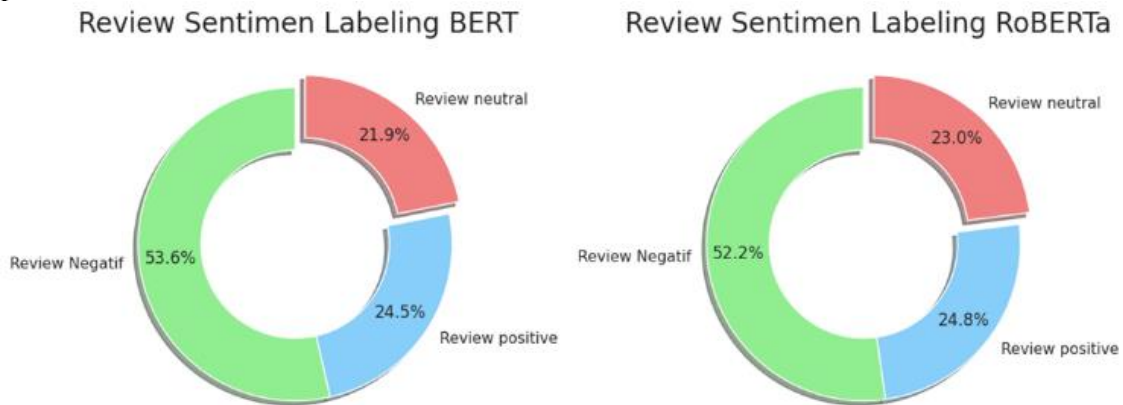
Labeling dataset pada penelitian yang dilakukan memanfaatkan *Fine-Tuning* BERT dan RoBERTa model berbahasa Indonesia seperti yang terlihat pada tabel 3. Hal ini dilakukan dikarenakan dataset yang telah berlabel sulit didapatkan serta membutuhkan pembiayaan yang lebih tinggi [26]. Selain itu, terdapat beberapa perbedaan hasil klasifikasi seperti contoh yang terlihat pada tabel 3 nomor 4 sampai 6.

Tabel 3. Sample hasil labeling dataset

Nomor	Komentar	RoBERTa	BERT
1	ChatGPT aja udah serem bang.	negative	negative
2	Belajar pakai chat gpt membantu banget, terutama sebagai anak kuliah yg bingung mau nyari bahan belajar bagus, pertama cari di gpt, setelah itu baru pakai Google untuk cari lebih detail	positive	positive
3	informasi adalah kekuatan	neutral	neutral
4	Gue dah pake model GPT-4.	neutral	positive
5	Jarvis, Friday, Karen, Edith in real life. Serem sih sebenarnya yang kaya Elon Musk bilang. Tapi sisi positifnya efisien.	positive	negative
6	menurut maa Raymond tentang Timothy sekarang gimana?	negative	neutral

3.3 Analisis Eksplorasi Dataset

Analisis Eksplorasi dataset dilakukan untuk melihat proporsi setiap kelas pada dataset yang telah melalui proses labeling dataset. Pada penelitian didapatkan kelas negatif, positif, dan netral sebesar 3106, 1422, 1268 dari labelling menggunakan BERT. Sedangkan, labelling menggunakan RoBERTa 3028 negatif, 1435 positif, 1333 netral berikut gambar 3 grafik persebaran setiap kelas pada dataset.



Gambar 3. Perbandingan kelas pada dataset

3.4 Preprocessing Dataset

Tahapan preprocessing dataset bertujuan untuk menghilangkan *noise* pada teks dataset yang dimiliki [27]. Selain itu tahapan Preprocessing merupakan tahapan yang penting untuk mendapatkan hasil yang lebih baik meskipun penggunaan bahasa Inggris perlu melalui tahap ini [28]. Pada tahapan ini peneliti melakukan perubahan teks secara keseluruhan menjadi huruf kecil, menggantikan enter menjadi spasi, penghapusan emotikon, penghapusan symbol, serta penghapusan karakter yang berulang. Sehingga pada penelitian dilakukan dataset menjadi 5788 dari 5796.

3.5 Split Dataset

Pada penelitian split dataset dilakukan menjadi 3 bagian yaitu data latih, data validasi, dan data testing. Ratio yang digunakan 70:30:30 adalah 70% dari dataset menjadi data latih, 30% dari dataset akan dibagikan lagi menjadi 70% data validation dan 30% data testing seperti yang terlihat pada tabel 4.

Tabel 4. Split dataset

	Data Latih	Data Validation	Data Testing	Ratio
Skema-dataset-1	4051	1215	522	70:30:30
Skema-dataset-2	4630	926	232	80:20:20

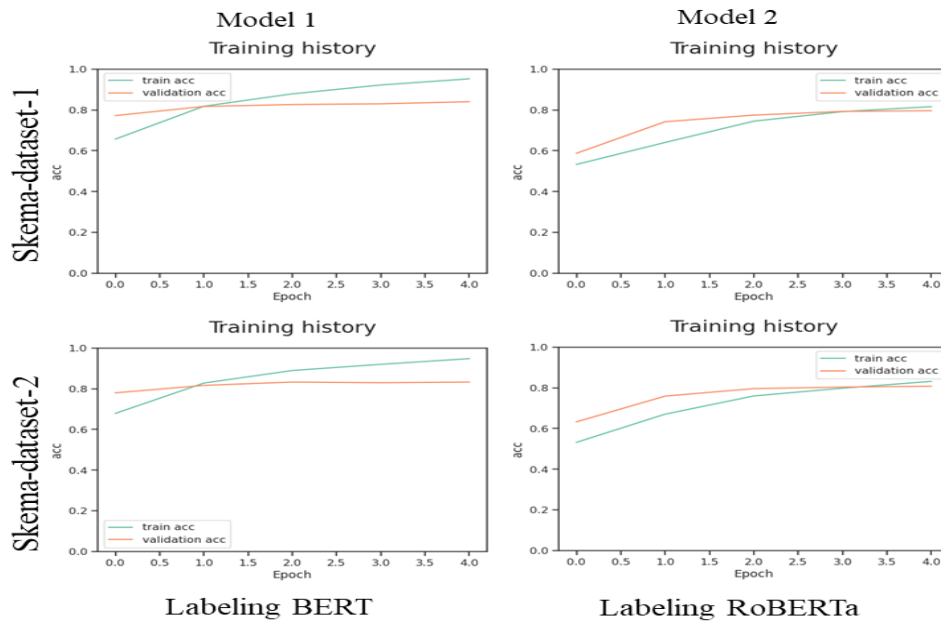
3.6 Modeling

Pada tahapan modeling peneliti menggunakan transformers model *pre-trained* huggingface yaitu: `flax-community/indonesian-roberta-base` menggunakan dataset OSCAR, dan `indobenchmark/indobert-base-p1` dengan dataset indonlu. Saat penelitian dilakukan menggunakan Google Colab GPU tesla t4. Penelitian yang dilakukan menggunakan 5 epochs, learning rate $3e-6$, batch size 32, workers 16, max token 512, num_labels 3 karena sesuai dengan kelas kategori sentimen pada dataset, optimizer Adam.

4. PEMBAHASAN

4.1 Modeling

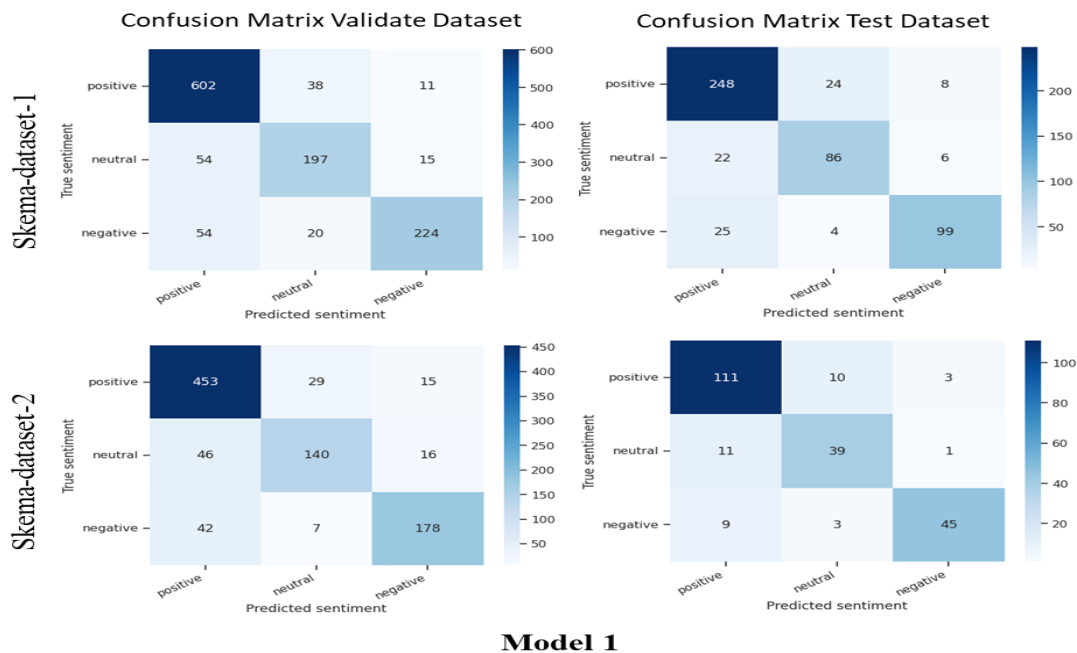
Tahapan modeling yaitu Model 1 indobenchmark/indobert-base-p1, model 2 flax-community/indonesian-roberta-base. Gambar 4 memperlihatkan grafik hasil training dari model yang digunakan pada saat training dataset dengan skema seperti pada tabel 4 serta yang telah dijelaskan pada bagian 3.6.



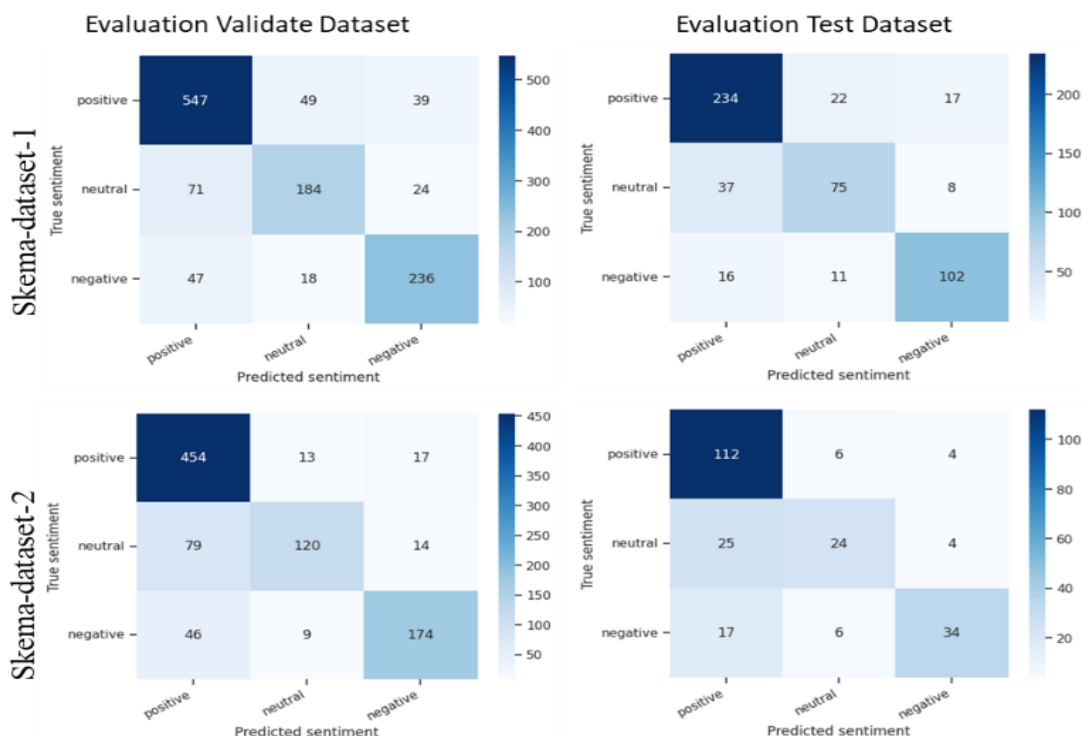
Gambar 4. Grafik training dataset

4.2 Evaluasi

Hasil Confusion matrix dari data latih yang dilakukan dengan beberapa parameter yang mengacu pada bagian 3.6 dengan membagi berdasarkan type labeling BERT dan RoBERTa terlihat pada gambar 5 serta gambar 6. Pada evaluasi peneliti melakukan 2 tahapan yaitu validasi data latih dan juga test dataset.



Gambar 5. Evaluasi confusion matrix model 1 terhadap penggunaan labeling BERT



Model 2

Gambar 6. Evaluasi *confusion matrix* model 2 terhadap penggunaan labeling RoBERTa

Setelah mendapatkan confusion matrix, peneliti melakukan perhitungan *precision*, *recall*, *f1-score*, serta akurasi dari model yang digunakan mendapatkan hasil seperti yang terlihat pada gambar 7 dan gambar 8. Selain itu gambar 9 merupakan hasil akurasi terbaik dari proses train dataset yang dimana kedua model memiliki akurasi diatas 80% saat proses validation. Namun, ketika proses test dataset model 2 mengalami penurunan akurasi menjadi 79% pada skema dataset 1 dan 73% pada skema dataset 2.

	Evaluation Validate Dataset				Evaluation Test Dataset					
	precision	recall	f1-score	support	precision	recall	f1-score	support		
Skema-dataset-1	positive	0.85	0.92	0.88	651	positive	0.84	0.89	0.86	280
	neutral	0.77	0.74	0.76	266	neutral	0.75	0.75	0.75	114
	negative	0.90	0.75	0.82	298	negative	0.88	0.77	0.82	128
	accuracy			0.84	1215	accuracy			0.83	522
	macro avg	0.84	0.81	0.82	1215	macro avg	0.82	0.80	0.81	522
	weighted avg	0.84	0.84	0.84	1215	weighted avg	0.83	0.83	0.83	522
Skema-dataset-2	positive	0.84	0.91	0.87	497	positive	0.85	0.90	0.87	124
	neutral	0.80	0.69	0.74	202	neutral	0.75	0.76	0.76	51
	negative	0.85	0.78	0.82	227	negative	0.92	0.79	0.85	57
	accuracy			0.83	926	accuracy			0.84	232
	macro avg	0.83	0.80	0.81	926	macro avg	0.84	0.82	0.83	232
	weighted avg	0.83	0.83	0.83	926	weighted avg	0.84	0.84	0.84	232

Gambar 7. Evaluasi model 1 terhadap penggunaan labeling BERT

		Evaluation Validate Dataset				Evaluation Test Dataset				
		precision	recall	f1-score	support	precision	recall	f1-score	support	
Skema-dataset-1	positive	0.82	0.86	0.84	635	positive	0.82	0.86	0.84	273
	neutral	0.73	0.66	0.69	279	neutral	0.69	0.62	0.66	120
	negative	0.79	0.78	0.79	301	negative	0.80	0.79	0.80	129
	accuracy			0.80	1215	accuracy			0.79	522
	macro avg	0.78	0.77	0.77	1215	macro avg	0.77	0.76	0.76	522
	weighted avg	0.79	0.80	0.79	1215	weighted avg	0.78	0.79	0.79	522
Skema-dataset-2	positive	0.78	0.94	0.85	484	positive	0.73	0.92	0.81	122
	neutral	0.85	0.56	0.68	213	neutral	0.67	0.45	0.54	53
	negative	0.85	0.76	0.80	229	negative	0.81	0.60	0.69	57
	accuracy			0.81	926	accuracy			0.73	232
	macro avg	0.83	0.75	0.78	926	macro avg	0.73	0.66	0.68	232
	weighted avg	0.81	0.81	0.80	926	weighted avg	0.73	0.73	0.72	232

Gambar 8. Evaluasi model 2 terhadap penggunaan labeling RoBERTa

Labelling	Model	skema dataset	val accuracy	test accuracy
BERT	indobenchmark/indobert-base-p1	1	0.84	0.83
		2	0.83	0.84
RoBERTa	flax-community/indonesian-roberta-base	1	0.8	0.79
		2	0.81	0.73

Gambar 9. Validation accuracy dan test accuracy

5. KESIMPULAN

Berdasarkan penelitian yang dilakukan dengan beberapa skema peneliti dapat menyimpulkan model indobenchmark/indobert-base-p1 menjadi akurasi yang paling baik disemua skema split dataset. Sehingga, model tersebut dapat digunakan untuk melabeling dataset berbahasa Indonesia hal ini dikarenakan data latih IndoNLU berukuran jauh lebih besar sehingga model dapat belajar lebih banyak daripada model yang lain. Namun, labeling dataset disarankan dilakukan pendampingan ahli Bahasa sehingga keakuratan terhadap teks dapat dipertanggung jawabkan. Selain itu, penelitian ini yang menggunakan analisis eksplorasi data mendapatkan sentimen negative menjadi sentiment tertinggi dengan 53.6% dibandingkan dengan sentimen netral maupun positif. Sehingga perlunya keterlibatan stackholder dalam mengatur penggunaan AI di Indonesia agar tidak disalahgunakan sebagai alat tindak kejahatan.

● **12% Overall Similarity**

Top sources found in the following databases:

- 9% Internet database
- 2% Publications database
- Crossref database
- Crossref Posted Content database
- 9% Submitted Works database

TOP SOURCES

The sources with the highest number of matches within the submission. Overlapping sources will not be displayed.

1	Universitas Pendidikan Indonesia on 2023-01-06 Submitted works	7%
2	Volkan KILIC , O yku B. MERCAN, Mehmet TETIK, O zlem KAP, Nesri Crossref	<1%
3	slideshare.net Internet	<1%
4	prosiding.unimus.ac.id Internet	<1%
5	huggingface.co Internet	<1%
6	Muhammad Muharrom Al Haromainy, Dimas Ari Setyawan, Onny Kartik... Crossref	<1%
7	core.ac.uk Internet	<1%
8	ejurnal.dipanegara.ac.id Internet	<1%

-
- 9 Arief Gilang Ramadhan, Teguh Susyanto, Iwan Ady Prabowo. "Sistem P... <1%
Crossref
-
- 10 text-id.123dok.com <1%
Internet
-
- 11 ontolux.de <1%
Internet
-
- 12 scribd.com <1%
Internet
-
- 13 Dihin Muriyatmoko, Triana Harmini, Maulana Kemal Ardiansyah. "Senti... <1%
Crossref
-
- 14 Sriwijaya University on 2022-01-21 <1%
Submitted works