

# KLASIFIKASI SENTIMEN NETIZEN MEDIA SOSIAL X TERHADAP KANDIDAT CAWAPRES PADA PILPRES 2024 MENGGUNAKAN INDOBERT

Oleh:

**Alfath Thoriq Kumara, Mujib Ridwan, Anang Kunaefi**  
Fakultas Saintek UIN Sunan Ampel Surabaya

## ABSTRAK

Penelitian ini bertujuan untuk mengklasifikasi sentimen masyarakat pada masa kampanye pemilu 2024 melalui *tweet* berbahasa Indonesia pada media sosial X (Twitter). Penelitian ini diharapkan dapat menjadi telaah untuk masyarakat atau peneliti lain mengenai masa pemilu 2024. Penelitian ini juga diharapkan menjadikan masyarakat lebih mawas terhadap informasi yang beredar di media sosial. Data yang digunakan berjumlah  $\pm 4500$  *tweet* yang diambil pada waktu yang telah ditentukan. Analisis penelitian ini berbasis deep learning, terkhusus menggunakan metode IndoBERT dari *huggingface* yang berbasis pada BERT dengan arsitektur Transformer. Hasil penelitian menunjukkan bahwa model IndoBERT menghasilkan performa yang baik hingga mencapai akurasi 93,67%. Hasil analisa sentimen positif dan negatif, keduanya menunjukkan bahwa *tweet* berisi kata-kata dengan konotasi positif seperti “pilih”, “menang”, dan “jadi” dan berisi kata yang merujuk pada subjek seperti “pak”, “mas”, “gus”, dan “prof”.

**Kata kunci:** sentimen, Indobert, *tweet*, pemilu, cawapres.

## ABSTRACT

### SENTIMENTS CLASSIFICATION OF SOCIAL MEDIA NETIZENS X TOWARDS THE VICE PRESIDENT CANDIDATE IN THE 2024 PRESIDENTIAL ELECTION USING INDOBERT

By:

**Alfath Thoriq Kumara, Mujib Ridwan, Anang Kunaefi**  
Fakultas Saintek UIN Sunan Ampel

*This study aims to classify public sentiment during the 2024 election campaign through Indonesian-language tweets on social media X (Twitter). This research is expected to be a study for the community or other researchers regarding the 2024 election period. This research is also expected to make people more aware of the information circulating on social media. The data used amounted to  $\pm 4500$  tweets taken at a predetermined time. This research analysis is based on deep learning, specifically using the IndoBERT method from *huggingface* based on BERT with Transformer architecture. The results show that the IndoBERT model produces good performance to achieve 93.67% accuracy. The results of positive and negative sentiment analysis both show that the tweets contain words with positive connotations such as "pilih", "menang", and "jadi" and contain words that refer to subjects such as "pak", "mas", "gus", and "prof".*

**Keywords:** sentiment, IndoBERT, *tweet*, election, vice president candidate

## **Pendahuluan**

Perkembangan teknologi dan informasi telah mencapai tingkat yang pesat di era modern saat ini. Salah satu penemuan yang sangat mempengaruhi perkembangan tersebut adalah internet. Di Indonesia, pengguna internet terus meningkat dengan cepat. Pada bulan Januari 2023, jumlah pengguna aktif internet di Indonesia mencapai 212,9 juta orang (*Hootsuite (We are Social): Indonesian Digital Report 2023*, 2023). Media sosial menjadi salah satu tempat bagi redaksi untuk menyampaikan berita. Tersebarinya berita melalui media sosial dapat dengan mudah mendapat perhatian pengguna media sosial atau *netizen*. Berbagai cuitan dan berita yang tersebar di media sosial merupakan data yang mudah untuk didapatkan. Data-data tersebut merupakan informasi yang berharga mengenai sentimen dan polaritas suatu hal apabila diolah dengan benar.

Analisis sentimen merupakan pendekatan komputasional untuk mempelajari opini, sentimen, dan emosi yang terungkap dalam bentuk teks. Tujuannya adalah untuk mengidentifikasi atribut dan komponen yang dikomentari dalam setiap dokumen teks, serta menentukan apakah komentar tersebut bersifat positif, negatif, atau netral (Alita & Isnain, 2020). Memanfaatkan teknik analisis sentimen dan *text mining* dapat mengungkap pola opini dan sentimen dalam teks, yang dapat memberikan wawasan berharga dalam berbagai bidang seperti analisis pasar, pengambilan keputusan, dan pengembangan produk.

Intensitas pembicaraan di media sosial mengenai pemilu yang akan datang mencerminkan tingginya antusiasme dan perhatian masyarakat terhadap proses demokrasi. Warganet secara aktif membahas survei elektabilitas, program kerja, serta rekam jejak para kandidat, menciptakan lingkungan digital yang penuh dengan opini yang beragam mengenai visi, misi, dan karakter masing-masing calon. Selain itu, perdebatan dan pertarungan argumen di media sosial semakin sering terjadi, terutama menjelang debat calon wakil presiden (cawapres). Maraknya data yang tersebar dapat digunakan sebagai acuan untuk menilai seperti apa tanggapan masyarakat mengenai keberpihakan netizen terhadap cawapres pemilu pada Februari 2024.

## **TINJAUAN PUSTAKA**

Text mining adalah proses ekstraksi informasi berharga dari kumpulan dokumen teks yang tidak terstruktur atau semi terstruktur menggunakan alat analisis, yang merupakan bagian dari data mining. Tujuan utamanya adalah mendapatkan informasi yang berarti melalui teknik seperti kategorisasi, pengelompokan teks, dan identifikasi pola. Sumber data bisa berupa berbagai jenis dokumen seperti artikel, laporan, dan media sosial. Untuk mencapai tujuannya, text mining mengadopsi teknik dari berbagai bidang seperti Data Mining, Machine Learning, dan Natural Language Processing (NLP).

Dalam informasi tekstual, terdapat dua jenis informasi, yaitu fakta dan opini. Fakta bersifat objektif, sedangkan opini bersifat subyektif dan mencerminkan sentimen individu. Analisis sentimen adalah bidang yang fokus pada pengklasifikasian polaritas teks untuk menentukan apakah opini dalam dokumen bersifat positif, negatif, atau netral. Bidang ini memanfaatkan teknik komputasional untuk memahami aspek emosional dalam teks. Analisis sentimen semakin penting, terutama dalam mengolah data dari media sosial, ulasan produk, dan opini pengguna, guna memahami pendapat publik dan mendukung pengambilan keputusan. Fokus utama dari analisis sentimen adalah melakukan

klasifikasi terhadap *review* berdasarkan polaritasnya. Melalui analisis sentimen, informasi yang tersembunyi dalam teks dapat diungkapkan, dan hasilnya dapat digunakan untuk berbagai keperluan seperti pemahaman umum publik terhadap suatu produk, merek, atau topik tertentu, serta pengambilan keputusan berdasarkan sentimen yang terkandung dalam ulasan dan komentar pengguna.

Analisis sentimen melibatkan pengolahan data teks yang seringkali tidak terstruktur. Untuk memudahkan proses klasifikasi, data mentah diolah menjadi lebih terstruktur melalui beberapa tahapan. Penghapusan Duplikasi Data: Data yang memiliki duplikat dihapus untuk meningkatkan efisiensi klasifikasi. Preprocessing: proses pengolahan data mentah menjadi lebih bersih dan terstruktur, meliputi: Cleansing: Penghapusan simbol-simbol khusus, tanda baca, dan spasi yang tidak perlu. Case Folding: Mengubah semua huruf dalam teks menjadi huruf kecil untuk keseragaman. Tokenizing: Memisahkan teks menjadi unit-unit kecil atau kata-kata (token) untuk analisis lebih lanjut. Normalization: Menyamakan variasi penulisan kata agar sesuai dengan bentuk standar. Filtering: Menghilangkan kata-kata yang tidak relevan seperti kata hubung untuk fokus pada kata-kata yang lebih bermakna. Proses-proses ini menghasilkan data teks yang lebih terstruktur, siap untuk dianalisis dan diklasifikasikan. Stemming: Menghapus imbuhan pada kata-kata untuk mendapatkan kata dasar yang sesuai. Setelah proses preprocessing selesai, data yang sudah bersih dan terstruktur ini kemudian diklasifikasikan sentimennya menggunakan metode yang telah dipilih.

BERT (Bidirectional Encoder Representations from Transformers) adalah model canggih dalam bidang Natural Language Processing (NLP) yang dikembangkan oleh Google pada 11 Oktober 2018. BERT menggunakan pendekatan bidirectional dalam pemrosesan teks, yang berarti model ini dapat memahami konteks dengan melihat kata dari kedua arah, yakni dari kiri ke kanan dan dari kanan ke kiri. Pendekatan ini memungkinkan BERT untuk melakukan fine-tuning dengan hanya menambahkan satu lapisan tambahan, yang membuatnya mampu menangkap makna kata dalam konteks yang lebih luas.

BERT dibangun berdasarkan arsitektur Transformer encoder, yang terdiri dari beberapa lapisan. Setiap lapisan encoder memiliki dua sub-lapisan: mekanisme multi-head self-attention dan jaringan feedforward terhubung sepenuhnya. Dalam setiap lapisan, input melewati self-attention untuk memahami konteks secara lebih luas sebelum diteruskan ke jaringan feedforward untuk langkah berikutnya.

WordPiece Embeddings adalah metode yang digunakan dalam BERT untuk merepresentasikan kata dalam bentuk vektor. Dengan 30.000 token dalam kosakata, WordPiece dirancang untuk mengatasi masalah kata-kata yang tidak ada dalam kamus (Out of Vocabulary/OOV). Kata-kata yang sering digunakan disimpan utuh, sementara kata-kata yang jarang digunakan dibagi menjadi sub-kata hingga mencapai bentuk dasar.

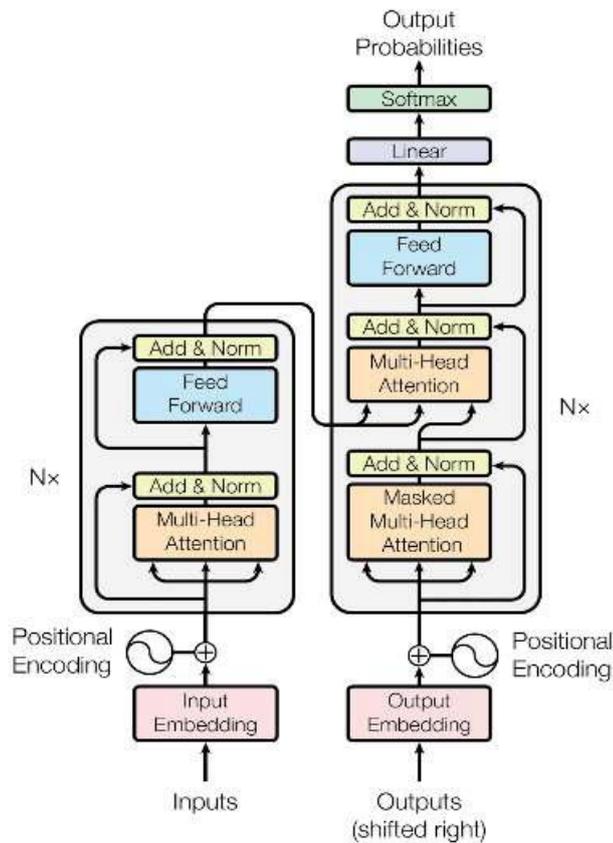


fig 1 Arsitektur Transformer

Setiap kalimat dalam BERT diawali dengan token khusus [CLS] untuk klasifikasi, dan diakhiri dengan token [SEP] untuk pemisahan antar kalimat. Selain itu, embedding tambahan digunakan untuk membedakan antara dua kalimat, yang terdiri dari token embedding, segment embedding, dan position embedding.

Input	[CLS]	my	dog	is	cute	[SEP]	he	likes	play	#ing	[SEP]
Token Embeddings	$E_{[CLS]}$	$E_{my}$	$E_{dog}$	$E_{is}$	$E_{cute}$	$E_{[SEP]}$	$E_{he}$	$E_{likes}$	$E_{play}$	$E_{#ing}$	$E_{[SEP]}$
	+	+	+	+	+	+	+	+	+	+	+
Segment Embeddings	$E_A$	$E_A$	$E_A$	$E_A$	$E_A$	$E_A$	$E_B$	$E_B$	$E_B$	$E_B$	$E_B$
	+	+	+	+	+	+	+	+	+	+	+
Position Embeddings	$E_0$	$E_1$	$E_2$	$E_3$	$E_4$	$E_5$	$E_6$	$E_7$	$E_8$	$E_9$	$E_{10}$

fig 2 Representasi Input BERT

BERT memiliki dua fase utama: pre-training dan fine-tuning. Pre-training bertujuan untuk menghasilkan representasi kata dan konteks dalam teks melalui dua proses utama. Masked Language Modeling (MLM): Dalam proses ini, sekitar 15% dari kata dalam kalimat dipilih secara acak untuk diganti dengan token [MASK]. Model kemudian berusaha memprediksi kata yang hilang berdasarkan konteks kalimat.

Next Sentence Prediction (NSP): Proses ini membantu model memahami hubungan antara dua kalimat dengan memprediksi apakah kalimat kedua mengikuti kalimat pertama atau tidak.

Fine-tuning memungkinkan BERT untuk diadaptasi ke berbagai tugas NLP dengan melakukan modifikasi pada model yang telah dilatih sebelumnya. Arsitektur Transformer dalam BERT memudahkan proses ini karena mekanisme self-attention yang dimilikinya.

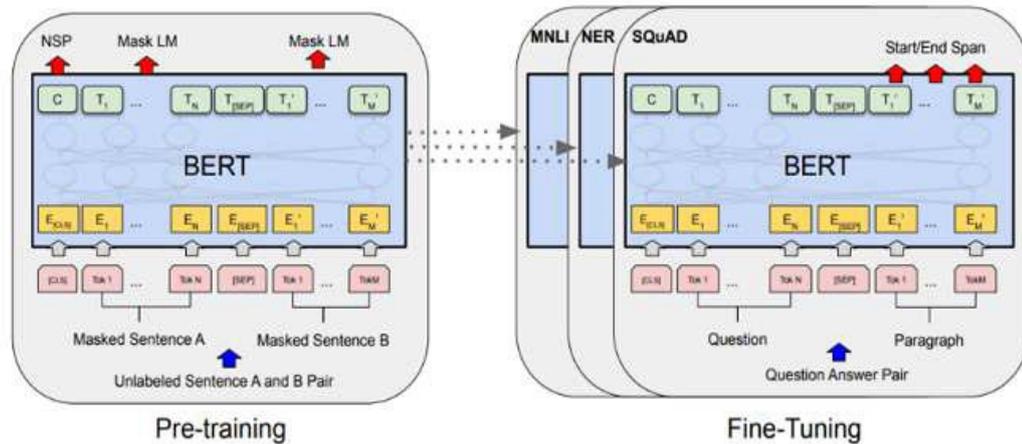


fig 3 Ilustrasi Alur Proses Pre-training dan Fine-tuning.

IndoBERT adalah varian dari BERT yang dirancang khusus untuk bahasa Indonesia. Walaupun mengikuti arsitektur BERT, IndoBERT dilatih menggunakan korpus teks berbahasa Indonesia yang luas, sehingga mampu memberikan representasi yang lebih baik untuk konteks bahasa Indonesia dibandingkan BERT standar.

IndoBERT menggunakan SentencePiece dengan Byte Pair Encoding (BPE) sebagai metode penyusunan vocabulary, berbeda dengan BERT yang menggunakan WordPiece. Dengan dataset Indo4B, IndoBERT dilatih menggunakan TPU v3-8 dalam dua tahap, menghasilkan beberapa varian model seperti IndoBERTBASE, IndoBERTLARGE, IndoBERT-liteBASE, dan IndoBERT-liteLARGE, masing-masing dengan kelebihan dan kekurangannya sendiri.

Confusion Matrix adalah metode yang digunakan untuk menampilkan dan memvisualisasikan hasil prediksi sebuah model dalam kaitannya dengan kondisi faktual. Confusion Matrix membantu dalam mengevaluasi kinerja model dengan mengukur beberapa metrik penting seperti accuracy, precision, recall, dan F-measure. Dalam sebuah Confusion Matrix, biasanya terdapat empat komponen utama. True Positive (TP): Jumlah data yang diprediksi benar oleh model sebagai positif dan memang benar positif. True Negative (TN): Jumlah data yang diprediksi benar oleh model sebagai negatif dan memang benar negatif. False Positive (FP): Jumlah data yang diprediksi salah oleh model sebagai positif, padahal sebenarnya negatif. Ini dikenal juga sebagai Type I Error. False Negative (FN): Jumlah data yang diprediksi salah oleh model sebagai negatif, padahal sebenarnya positif. Ini dikenal juga sebagai Type II Error.

table 1 Tabel Confussion Matrix

	Aktual Positif	Aktual Negatif
Prediksi Positif	True Positif (TP)	False Positif (FP)
Prediksi Negatif	False Negative (FN)	True Negative (TN)

Merujuk pada Tabel 2.2 di atas dan tujuan penggunaannya yang mencakup penghitungan *accuracy*, *recall*, *F-Measure*, dan *precision*, informasinya dapat diuraikan sebagai berikut:

1. *Accuracy*

Proporsi prediksi yang benar dari keseluruhan prediksi.

$$Accuracy = (True Positives + True Negatives) / (True Positives + True Negatives + False Positives + False Negatives)$$

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}$$

2. *Recall*

Proporsi prediksi yang benar dari semua data yang sebenarnya.

$$Recall(sensitivity) = True Positives / (True Positives + False Negatives)$$

$$Recall(specificity) = True Negatives / (True Negatives + False Positives)$$

$$Sensitivity = \frac{TP}{TP + FN}$$

$$Specificity = \frac{TN}{TN + FP}$$

3. *Precision*

Proporsi prediksi positif yang benar dari semua prediksi positif.

$$Precision Positives = True Positives / (True Positives + False Positives)$$

$$Precision Negatives = True Negatives / (True Negatives + False Negatives)$$

$$precision\ positif = \frac{TP}{TP + FP}$$

$$precision\ negatif = \frac{TN}{TN + FN}$$

4. *F-Measure*

Harmonic mean dari precision dan recall, memberikan gambaran keseimbangan antara keduanya.

$$F\text{-measure Positive} = 2 * (Precision Positive * Recall Positive) / (Precision Positive + Recall Positive)$$

$$F\text{-measure Negative} = 2 * (Precision Negative * Recall Negative) / (Precision Negative + Recall Negative)$$

$$F - \text{measure positif} = 2 \frac{\text{precision positif} * \text{recall positif}}{\text{precision positif} + \text{recall positif}}$$

$$F - \text{measure negatif} = 2 \frac{\text{precision negatif} * \text{recall negatif}}{\text{precision negatif} + \text{recall negatif}}$$

Pseudo labelling adalah metode semi-supervised learning yang melabeli data yang belum terlabel dengan bantuan data yang sudah terlabel untuk meningkatkan kinerja model klasifikasi. Prosesnya melibatkan pelatihan model menggunakan data terlabel, memprediksi label pada data belum terlabel, lalu melatih kembali model dengan gabungan data terlabel dan pseudo label. Meskipun bermanfaat, pseudo labelling harus digunakan dengan hati-hati karena prediksinya mungkin tidak selalu akurat. Ada beberapa jenis pseudo labelling, termasuk berbasis confidence threshold, probabilitas kelas, konsensus, dan bobot, yang dipilih berdasarkan karakteristik data dan kebutuhan klasifikasi.

K-fold cross validation adalah metode pengujian yang menilai kinerja algoritme dengan membagi data secara acak menjadi k kelompok. Satu kelompok digunakan sebagai data uji, sementara sisanya sebagai data latih.

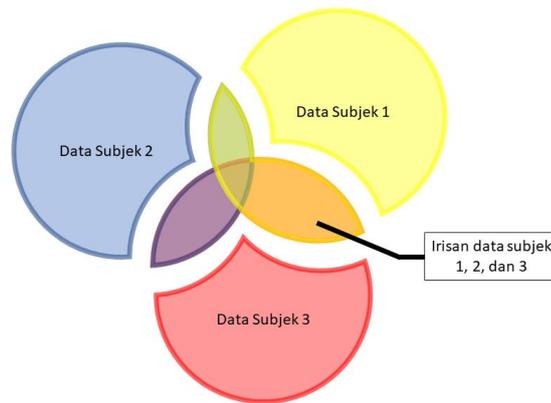
## **ARSITEKTUR PENELITIAN**

### **Jenis dan Sumber Data**

Data penelitian ini adalah komentar masyarakat tentang cawapres pemilu 2024 yang dikumpulkan melalui scraping di media sosial X (sebelumnya Twitter) menggunakan ekstensi Web Scraper di Nitter. Web Scraper memungkinkan pengambilan data teks, link, dan gambar dengan mudah. Data dikumpulkan menggunakan kata kunci seperti “@cakimiNOW”, “@gibran\_tweet”, dan “@mohmahfudmd”, mencakup cuitan tanpa reply, link, retweet, atau retweet quotes, dari pendaftaran cawapres ke KPU hingga 30 hari setelahnya, termasuk 23-25 Desember 2023 setelah debat cawapres pertama. Jenis tweet yang diambil adalah unggahan asli, bukan reply, yang mengandung link, retweet, atau retweet quotes. Hasil scraping adalah tiga set data mentah untuk masing-masing cawapres: 1578 tweet untuk subjek pertama, 2573 tweet untuk subjek kedua, dan 1305 tweet untuk subjek ketiga, yang akan digunakan dalam tahap pemrosesan berikutnya.

### **Penghapusan Duplikasi Data**

Proses penghapusan data terdiri dari dua langkah. Langkah pertama adalah menghapus data yang muncul di lebih dari satu set data mentah (misalnya, di data subjek 1 dan 2) karena tidak dapat dipastikan subjeknya. Setelah proses ini, data yang tersisa di tiap subjek hanya mencakup tweet yang khusus untuk satu subjek. Hasilnya adalah: dataset subjek pertama dengan 1191 tweet, subjek kedua dengan 2231 tweet, dan subjek ketiga dengan 885 tweet.



*fig 4 Irisan Data Tiap Subjek*

Proses penghapusan duplikasi kedua melibatkan penghapusan tweet yang sama dalam dataset, dilakukan pada awal dan akhir tahap preprocessing. Tujuannya adalah untuk mengurangi data duplikat dan mempermudah proses klasifikasi. Hasilnya adalah dataset yang bebas dari duplikasi.

### **Preprocessing**

Preprocessing adalah tahap pengolahan data setelah penghapusan duplikasi. Data yang sudah terkelompok menurut subjek diproses melalui beberapa langkah. Langkah pertama adalah cleansing, yaitu membersihkan data tweet dari tautan, emoji, tanda baca, simbol, dan karakter non-ASCII menggunakan library regex dan fungsi compile(). Selain itu, spasi berlebihan juga diubah menjadi spasi tunggal.

### **Pelabelan**

Data yang telah diproses akan diberi label dalam dua cara: menggunakan algoritma dan label dari pakar.

1. Pelabelan Algoritma: Menggunakan kamus leksikon dari jurnal yang berisi bobot kata-kata dalam Bahasa Indonesia. Kalimat diberi label 1 jika bobotnya positif dan 0 jika bobotnya negatif.
2. Pelabelan Pakar: Melibatkan 3 pakar yang memberikan label 0 atau 1 pada 10% data acak dari masing-masing subjek. Pakar menilai data mentah dan memberikan label berdasarkan kata-kata positif atau negatif.

Hasil akhir adalah dataset dengan label 0 (negatif) atau 1 (positif) dari kedua metode tersebut.

### **Validasi Label**

Proses validasi label adalah proses membandingkan label dari algoritma dan label dari pakar. Pada dataset pertama terdapat 135 data berlabel dari pakar, sebanyak 128 data dari data tersebut valid dan dapat digunakan. Pada dataset kedua terdapat 250 data berlabel dari pakar, sebanyak 226 data dari data tersebut valid dan dapat digunakan. Pada dataset ketiga terdapat 129 data berlabel dari pakar, sebanyak 113 data dari data tersebut valid dan dapat digunakan. Data (berlabel dari pakar) tersebut akan digunakan sebagai pembandingan bagi data berlabel dari algoritma. Tingkat kecocokan label dataset 1 adalah 61,5%, dataset 2 adalah 56,5%, dan dataset 3 adalah 60,6%.

### **Modelling**

Proses modeling melibatkan beberapa tahap: membangun model, fine-tuning, dan tokenisasi. Sebelum modeling, data dibagi menjadi 90% data training

dan 10% data testing. Data training juga digunakan untuk validasi model dengan k-fold cross-validation. Jika ada ketidakseimbangan label, data akan dihapus secara acak untuk menyeimbangkan jumlah label.

Model dibangun menggunakan "indolem/indobert-base-uncased" dengan kelas `TFAutoModelForSequenceClassification` untuk klasifikasi urutan. Tokenisasi dilakukan dengan `AutoTokenizer` dari transformer, yang mengatur padding, panjang maksimal token (128 karakter), dan token spesial [CLS] dan [SEP]. Hasil akhir adalah data yang telah di-tokenisasi dan model siap dilatih.

table 2 Nilai hyperparameter yang digunakan pada tahap fine-tuning

No.	Hyperparameter	Nilai
1.	Learning rate	5e-5
2.	Epoch	5 dan 10
3.	Batch size	16 dan 32

### Confussion Matrix

table 3 Hasil Confusion Matrix Dataset 1

	Model 1 Batch Size 16, 5 Epochs	Model 2 Batch Size 16, 10 Epochs	Model 3 Batch Size 32, 5 Epochs	Model 4 Batch Size 32, 10 Epochs
Overall Accuracy	88,90%	90,83%	93,67%	91,35%
Overall Precision	87,37%	90,31%	93,51%	90,18%
Overall Recall	92%	92,02%	94,29%	93,46%
Overall F1-Score	89,53%	91,12%	93,9%	91,73

Pada tabel 3 terlihat nilai overall metrik yang dihasilkan model 3 pada proses klasifikasi pada dataset 1 lebih tinggi daripada model lain. Nilai overall accuracy adalah 93,67%, nilai overall precision adalah 93,67%, nilai overall recall adalah 94,29%, dan nilai overall f1-score adalah 93,9%. Hasil ini berselisih sedikit dengan model yang lain sekitar 2-6%.

table 4 perbandingan predicted positive dan predicted negative dataset 1

	Predicted positive	Predictied negative
Model 1	53,1%	46,9%
Model 2	51,16%	49,83%
Model 3	50,38%	49,61%
Model 4	51,94%	48,05%

Pada tabel 4 perbandingan Predicted Negative dan Predicted Positif setiap model pada dataset 1, dapat dilihat bahwa model 3 adalah model dengan selisih paling sedikit dengan nilai Predicted Positive sebesar 50,38% dan Predicted Negative sebesar 49,61%. Maka dapat diambil kesimpulan bahwa model 3 adalah model yang menghasilkan hasil prediksi paling balance pada dataset 1.

table 5 Hasil Confusion Matrix Dataset 2

	Model 1 Batch Size 16, 5 Epochs	Model 2 Batch Size 16, 10 Epochs	Model 3 Batch Size 32, 5 Epochs	Model 4 Batch Size 32, 10 Epochs
Overall Accuracy	87,75%	87,75%	91,1%	89,63%
Overall Precision	87,67%	88,23%	91,1%	89,6%
Overall Recall	89,47%	90,61%	93,18%	89,69%
Overall F1-Score	88,51%	88,86%	91,99%	89,52%

Pada Tabel 5 dapat dilihat bahwa pada dataset 2, model 3 mendapat nilai overall metrik tertinggi daripada model lain. Nilai overall accuracy adalah 91,1%, nilai overall precision adalah 91,1%, nilai overall recall adalah 93,18%, dan nilai overall f1-score adalah 91,99%. Nilai metrik ini adalah yang tertinggi daripada model lainnya yang berselisih sekitar 2-4%.

table 6 perbandingan predicted positive dan predicted negative dataset 2

	Predicted Positive	Predictied Negative
Model 1	51,45%	48,55%
Model 2	52,97%	47,03%
Model 3	51,9%	48,1%
Model 4	49,7%	50,3%

Pada tabel 6 perbandingan Predicted Negative dan Predicted Positif setiap model pada dataset 2, dapat dilihat bahwa model 4 adalah model dengan selisih paling sedikit dengan nilai Predicted Positive sebesar 49,7% dan Predicted Negative sebesar 50,3%. Maka dapat diambil kesimpulan bahwa model 4 adalah model yang menghasilkan hasil prediksi paling balance pada dataset 2.

table 7 Hasil Confusion Matrix Dataset 3

	Model 1 Batch Size 16, 5 Epochs	Model 2 Batch Size 16, 10 Epochs	Model 3 Batch Size 32, 5 Epochs	Model 4 Batch Size 32, 10 Epochs
Overall Accuracy	88,5%	79,35%	85,75%	91,06%
Overall Precision	87,99%	79,38%	83,99%	92,41%
Overall Recall	87,77%	84,79%	85,26%	86,49%
Overall F1-Score	87,78%	80,66%	84,24%	89,11%

Pada proses klasifikasi dataset 3 yang terlihat pada Tabel 7, nilai-nilai metrik terbaik dihasilkan oleh model 4, dengan nilai overall accuracy adalah 91,06%, nilai overall precision adalah 92,41%, dan nilai overall f1-score adalah 89,11%. Pada proses klasifikasi dataset 3 nilai overall recall terbaik dihasilkan oleh model 1 dengan nilai overall recall sebesar 87,77%. Berdasarkan angka-angka yang terlihat, model 4 terlihat lebih unggul dari metrik lain. Maka dari itu model 4 adalah yang terbaik dibanding model lain pada proses klasifikasi dataset 3.

table 8 perbandingan predicted positive dan predicted negative dataset 3

	Predicted Positive	Predictied Negative
Model 1	50,2%	49,8%
Model 2	56,04%	43,96%
Model 3	50,4%	49,6%
Model 4	46,52%	53,48%

Pada tabel 8, perbandingan Predicted Negative dan Predicted Positif setiap model pada dataset 3, dapat dilihat bahwa model 1 dan model 3 adalah model dengan selisih paling sedikit dengan nilai Predicted Positive dari model 1 sebesar 50,2% dan 50,4% pada model 3. Predicted Negative sebesar 49,8% pada model 1

dan 49,6% pada model 3. Dapat diambil kesimpulan bahwa model 1 dan 3 adalah model yang menghasilkan hasil prediksi yang balance pada dataset 3.

Berdasarkan hasil pengamatan Tabel 3, 5, dan 7 model terbaik adalah model 3 dan 4. Model 3 menghasilkan nilai metrik yang tinggi pada dataset 1 dan 2. Model 4 menghasilkan nilai metrik tertinggi pada dataset 3. Model 4 mendapatkan nilai metrik yang kurang pada dataset 1 dan 2, hal ini menandakan bahwa model 4 tidak bisa digunakan pada beragam kondisi dataset. Jadi kesimpulannya adalah model 3 dengan batch size 32 dan 5 epochs adalah model dengan nilai metrik terbaik. Model 3 memiliki performa sangat baik yang ditunjukkan dengan hasil *overall* metrik yang baik pada dataset 1 dan dataset 2. Pada dataset 3, model 3 menunjukkan performa yang baik dan stabil. Hasil yang baik pada dataset 1 dan dataset 2 dapat menjadi pertimbangan bahwa model 3 dapat digunakan untuk proses klasifikasi pada data lain (data nyata).

### **KESIMPULAN**

Berdasarkan rangkaian proses yang telah dilakukan oleh peneliti dan hasil yang diperoleh pada bab sebelumnya, dapat disimpulkan bahwa:

1. Model Indobert menghasilkan performa klasifikasi yang sangat baik. Performa model Indobert dengan batch size 32 dan 5 epoch memberikan hasil klasifikasi tertinggi. Nilai akurasi mencapai 93,67% pada dataset 1 dan 91,1% pada dataset 2.
2. Opini masyarakat terhadap pemilu 2024 dapat dilihat dari sentimen-sentimen positif dan negatif yang terlihat. Masyarakat cenderung mendukung pilihannya dan (tidak mendukung) mengkritik yang bukan pilihannya.
3. Sentimen positif dan negatif pada setiap dataset tidak memiliki selisih yang besar atau dengan kata lain balance. Yaitu berkisar pada angka 49% hingga 50% untuk predicted positif dan negatif.

## DAFTAR PUSTAKA

- Abdillah, W. F., Premana, A., & Bhakti, R. M. H. (n.d.). Analisis Sentimen Penanganan Covid-19 dengan Support Vector Machine: Evaluasi Leksikon dan Metode Ekstraksi Fitur. *2021/11/30*, 3(02).  
<https://doi.org/10.46772/intech.v3i02.556>
- Aldayel, A., & Magdy, W. (2019). Your Stance is Exposed! Analysing Possible Factors for Stance Detection on Social Media. *Proc. ACM Hum.-Comput. Interact.*, 3(CSCW). <https://doi.org/10.1145/3359307>
- Alita, D., & Isnain, A. R. (2020). Pendeteksian Sarkasme pada Proses Analisis Sentimen Menggunakan Random Forest Classifier. *jurnal komputasi*, 8(2).  
<https://doi.org/10.23960/komputasi.v8i2.2615>
- Biber, D., & Finegan, E. (1988). Adverbial stance types in English. *Discourse Processes*, 11(1), 1–34. <https://doi.org/10.1080/01638538809544689>
- Darwish, K., Magdy, W., Rahimi, A., Baldwin, T., & Abokhodair, N. (2018). Predicting Online Islamophobic Behavior after #ParisAttacks. *The Journal of Web Science*, 4(3), 34–52. <https://doi.org/10.1561/106.00000013>
- Dave, K., Lawrence, S., & Pennock, D. M. (2003). Mining the peanut gallery: Opinion extraction and semantic classification of product reviews. *Proceedings of the Twelfth International Conference on World Wide Web - WWW '03*, 519. <https://doi.org/10.1145/775152.775226>
- Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In J. Burstein, C. Doran, & T. Solorio (Eds.), *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)* (pp. 4171–4186). Association for Computational Linguistics. <https://doi.org/10.18653/v1/N19-1423>
- Feldman, R., & Sanger, J. (2007). *The text mining handbook: Advanced approaches in analyzing unstructured data*. Cambridge university press.
- Hootsuite (*We are Social*): *Indonesian Digital Report 2023*. (2023).  
<https://andi.link/hootsuite-we-are-social-indonesian-digital-report-2023/>
- Kementerian Komunikasi dan Informatika. (2022).  
[https://kominfo.go.id/content/detail/2366/indonesia-peringkat-lima-pengguna-twitter/0/sorotan\\_media](https://kominfo.go.id/content/detail/2366/indonesia-peringkat-lima-pengguna-twitter/0/sorotan_media)
- Koto, F., Rahimi, A., Lau, J. H., & Baldwin, T. (2020). *IndoLEM and IndoBERT: A Benchmark Dataset and Pre-trained Language Model for Indonesian NLP* (arXiv:2011.00677). arXiv. <https://doi.org/10.48550/arXiv.2011.00677>
- Küçük, D., & Can, F. (2021). Stance Detection: Concepts, Approaches, Resources, and Outstanding Issues. *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2673–2676. <https://doi.org/10.1145/3404835.3462815>
- Lahoti, P., Garimella, K., & Gionis, A. (2018). Joint Non-Negative Matrix Factorization for Learning Ideological Leaning on Twitter. *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*, 351–359. <https://doi.org/10.1145/3159652.3159669>
- Liu, B. (2010). Sentiment analysis and subjectivity. *Handbook of Natural Language Processing*, 2(2010), 627–666.

- Markoulidakis, I., Rallis, I., Georgoulas, I., Kopsiaftis, G., Doulamis, A., & Doulamis, N. (2021). Multiclass Confusion Matrix Reduction Method and Its Application on Net Promoter Score Classification Problem. *Technologies*, 9(4), Art. 4. <https://doi.org/10.3390/technologies9040081>
- Nurhuda, F., Widya Sihwi, S., & Doewes, A. (2016). Analisis Sentimen Masyarakat terhadap Calon Presiden Indonesia 2014 berdasarkan Opini dari Twitter Menggunakan Metode Naive Bayes Classifier. *Jurnal Teknologi & Informasi ITSmart*, 2(2), 35. <https://doi.org/10.20961/its.v2i2.630>
- Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends® in Information Retrieval*, 2(1–2), 1–135.
- Reveilhac, M., & Schneider, G. (2023). Replicable semi-supervised approaches to state-of-the-art stance detection of tweets. *Information Processing & Management*, 60(2), 103199. <https://doi.org/10.1016/j.ipm.2022.103199>
- Triawati, C., Bijaksana, M. A., Indrawati, N., & Saputro, W. A. (2009). Pemodelan Berbasis Konsep Untuk Kategorisasi Artikel Berita Berbahasa Indonesia. *Seminar Nasional Aplikasi Teknologi Informasi (SNATI)*.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (n.d.). *Attention is All you Need*.
- Wilie, B., Vincentio, K., Winata, G. I., Cahyawijaya, S., Li, X., Lim, Z. Y., Soleman, S., Mahendra, R., Fung, P., Bahar, S., & Purwarianti, A. (2020). *IndoNLU: Benchmark and Resources for Evaluating Indonesian Natural Language Understanding* (arXiv:2009.05387). arXiv. <https://doi.org/10.48550/arXiv.2009.05387>
- Wu, Y., Schuster, M., Chen, Z., Le, Q. V., Norouzi, M., Macherey, W., Krikun, M., Cao, Y., Gao, Q., Macherey, K., Klingner, J., Shah, A., Johnson, M., Liu, X., Kaiser, Ł., Gouws, S., Kato, Y., Kudo, T., Kazawa, H., ... Dean, J. (2016). *Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation* (arXiv:1609.08144). arXiv. <https://doi.org/10.48550/arXiv.1609.08144>
- Yuill, S., & Harpin, H. (2006). *Python*.