# ANALISIS SENTIMEN PENGGUNA PLATFORM MEDIA SOSIAL X PADA TOPIK PEMILIHAN PRESIDEN 2024 MENGGUNAKAN PERBANDINGAN MODEL MONOLINGUAL DAN MULTILINGUAL BERT

**Nurhasiyah, Ramaditia Dwiyansaputra, Santi Ika Murpratiwi, Arik Aranta**
Teknik Informatika, Universitas Mataram
Jalan Majapahit 62, Mataram, Lombok NTB, Indonesia
*nurhasiyah685@gmail.com*

**ABSTRAK**

Pemilihan Presiden 2024 di Indonesia merupakan topik penting yang banyak dibahas di media sosial, terutama platform X (sebelumnya Twitter). Media sosial ini menyediakan ruang bagi jutaan pengguna untuk berbagi opini yang dapat diolah menjadi data sentiment. Namun, menganalisis opini dalam jumlah besar membutuhkan model yang tepat. Penelitian ini bertujuan untuk menganalisis sentimen publik terkait topik tersebut menggunakan perbandingan model *monolingual* (IndoBERT) dan *multilingual* (mBERT), model digunakan untuk mengklasifikasikan sentiment positif, netral, dan negatif. Penelitian dilakukan dengan metode BERT yang meliputi data *crawling*, *preprocessing*, *labeling*, *data spliting*, *implementation* model dan *evaluation*. *Dataset* terdiri dari 10.140 *tweet* yang melalui proses *preprocessing* berupa *cleaning*, *case folding*, dan *tokenizing*, dan proses pelabelan sentiment (positif, netral, negatif). Hasil penelitian menunjukkan bahwa model IndoBERT memiliki akurasi tertinggi sebesar 84% dengan presisi 75%, *recall* 80%, dan *F1-Score* 78%, sedangkan mBERT mencatat akurasi 81%, presisi 69%, *recall* 78%, dan *F1-Score* 73%. Model IndoBERT terbukti lebih unggul dalam memahami konteks bahasa Indonesia dibandingkan mBERT,terutama pada sentiment positif dan negatif, itu karena keterbatasan dalam menangkap konteks khusus Bahasa Indonesia. Evaluasi menggunakan *confusion matrix* untuk mendukung pernyataan ini. Hasil penelitian ini diharapkan dapat membantu pengembangan teknologi pemrosesan bahasa alami (NLP) untuk mendukung pemahaman opini publik, terutama di sektor sosial-politik Indonesia.

***Kata kunci :*** *IndoBERT*, *mBERT*, Analisis Sentimen, Pilpres 2024, *Natural Languange Preprocessing (NLP)*, Media Sosial

## 1. PENDAHULUAN

Salah satu negara yang menganut sistem demokrasi secara ketat yakni Indonesia. Kerap kali, pemilihan umum diadakan guna memperingati perihal ini. Tujuan pemilihan umum merupakan untuk memilih kandidat otoritas negara di bermacam tingkatan, termasuk presiden serta wakil presiden. Pemilihan presiden merupakan salah satu peristiwa politik terpenting yang dapat mempengaruhi opini publik secara keseluruhan. Media sosial, khususnya Platform X (sebelumnya dikenal sebagai Twitter), memiliki peran penting dalam mencerminkan opini masyarakat mengenai isu-isu politik seperti pemilihan presiden.

Aplikasi X atau biasa dikenal sebagai Twitter menyediakan platform untuk berbagi pemikiran dengan *fitur* seperti me-*retweet*, mengambil gambar dan video, dan membagikannya ke beberapa jejaring sosial lainnya. X saat ini menjadi bagian dari kehidupan sehari-hari. Kemudahan penggunaan X dan jumlah pengikut yang tak terbatas menjadikannya platform populer untuk mengungkapkan pendapat. Dengan ruang kecil sebanyak 280 karakter memungkinkan pengguna untuk mengekspresikan tujuan dan niat mereka dengan cara yang jelas dan ringkas. X dapat memberikan sudut pandang yang tidak memihak tentang berbagai subjek, menjadikannya platform media sosial yang populer [1]. Dengan banyaknya *tweet* yang beredar di X membahas terkait Pemilu Indonesia tahun 2024, peneliti termotivasi untuk melakukan analisa sentimen dari *tweet* yang ada di X.

Analisis sentimen adalah penelitian yang meneliti bahasa tertulis individu untuk menentukan pendapat, perasaan, sikap, penilaian, emosi, subjektivitas, penilaian, atau perspektif mereka [2]. Analisis sentimen menjadi semakin penting untuk memahami opini publik karena media sosial menyediakan akses langsung dan *real-time* terhadap beragam opini publik.

Pada tahun 2018, peneliti Google AI Language mengembangkan model representasi bahasa terlatih yang disebut *Encoder Bidirectional Transformersm Representation* (BERT) [3] untuk keperluan *Natural Languange Processinng* (NLP) yakni semacam *text classification*. BERT terdiri dari transformer *encoder* yang digunakan dalam mempelajari hubungan antar kalimat ada 2 tipe BERT yang terkenal, yakni *monolingual* dan *multilingual*. *Monolingual* BERT dilatih oleh korpus bahasa tertentu dengan jumlah besar, misalnya bahasa Indonesia. Sehingga model dapat menganalisis teks berbahasa Indonesia dengan sangat baik. Sedangkan *multilingual* BERT dilatih oleh korpus dari berbagai bahasa, sehingga dapat menganalisis beberapa bahasa hanya dengan satu model. Akan tetapi model ini memiliki kekurangan

dimana hasil yang dicapai tidak sebaik model *monolingual* karena pada proses *pre-training* model tidak fokus pada bahasa tertentu.

Berdasarkan paparan diatas, IndoBERT dan mBERT dipilih sebagai model analisis sentimen karena supremasi mereka dalam memahami konteks linguistik secara menyeluruh melalui penggunaan arsitektur *Transformer*. Menjadi model monolingual, IndoBERT disesuaikan untuk bahasa Indonesia dan dapat menghasilkan temuan analisis yang lebih tepat dalam memahami ciri khas bahasa tersebut. Selanjutnya, sebagai model *multilingual* mBERT memberikan fleksibilitas untuk mengevaluasi teks dalam banyak bahasa. Keduanya menawarkan akurasi yang lebih dalam analisis sentimen dibandingkan dengan teknik yang lebih konvensional seperti SVM, KKN dan *Naive Bayes* [4], terutama untuk data rumit seperti opini publik media sosial. Karena manfaat ini, IndoBERT dan mBERT adalah pilihan terbaik untuk aplikasi Pemrosesan Bahasa Alami (NLP) terkait analisis sentimen. Penelitian ini bertujuan untuk mengevaluasi dan menentukan model yang palinng unggul dalam menganalisis sentimen teks berbahasa Indonesia pada topik pemilihan presiden. Penelitian ini secara khusus membandingkan performa IndoBERT, model *monolingual* yang dioptimalkan untuk bahasa indonesia dan mBERT, model yang menawarkan flesibilitas dalam menangani berbagai bahasa.

## 2. TINJAUAN PUSTAKA

### 2.1. Penelitian Terkait

Klasifikasi menggunakan analisis sentimen telah dilakukan dalam sejumlah penelitian dengan menggunakan berbagai sumber data. Untuk menentukan apakah ada perbedaan substansial antara klasifikasi ujaran kebencian kedua pendekatan tersebut, Husnul Mawaddah [6] membandingkan metode fitur manual SVM dan BERT. Dari kedua metode tersebut, hasilnya menunjukkan metode BERT lebih unggul dalam mengklasifikasi dibandingkan dengan SVM yang mana hasil yang didapatkan oleh BERT adalah 98%, sedangkan SVM adalah 81%.

Pada penelitian yang dilakukan oleh wang, dkk [7] memperkenalkan metode *Extend* untuk menambah baru ke mBERT melalui penyesuaian kosakata dan pelatihan terbatas. Metode ini meningkatkan performa untuk beberapa bahasa yang tidak ada dalam mBERT asli, seperti Oromo, Amharic, dan Somali. Hasil menunjukkan peningkatan rata-rata sebesar 6% untuk bahasa yang ada dalam mBERT dan 23% untuk bahasa baru.

Herlina, dkk [4] melakukan penelitian mengenai IndoBERT, IndoBERT merupakan salah satu varian BERT yang dilatih menggunakan korpus bahasa Indonesia, menunjukkan keunggulan dalam analisis sentimen dibandingkan model tradisional seperti SVM, KNN, dan *Naive Bayes*. Penelitian ini terkait calon presiden Indonesia tahun 2024. Model ini mencapai akurasi 80% dengan skor F1 hingga 85% untuk sentimen positif. Penelitian ini membuktikan efektifitas IndoBERT dalam menangkap konteks bahasa Indonesia secara mendalam terutama dibandingkan model multibahasa atau pendekatan berbasis *lexicon*.

Berdasarkan paparan dari beberapa penelitian yang dilakukan sebelumnya peneliti bermaksud menggunakan metode BERT dalam menganalisis sentimen pengguna Twitter terhadap pemilihan umum 2024, yang mana metode BERT merupakan metode yang baru dan belum banyak digunakan dalam penelitian lain. Selain itu pada penelitian [6] dinyatakan bahwa metode BERT menunjukkan performa yang baik dalam proses klasifikasi yang dilakukan yang ditunjukkan dengan nilai akurasi yang didapatkan mencapai 98% dibanding dengan metode SVM.

### 2.2. Analisis Sentimen

Analisis sentimen adalah cabang studi komputer yang meneliti sentimen, emosi, dan opini teks. Analisis sentimen, kadang-kadang disebut sebagai penambangan opini, telah mendapatkan popularitas di NLP dan penambangan data. Tujuan utama analisis sentimen adalah untuk memproses, mengekstrak, meringkas, dan menganalisis Konten teks menggunakan sejumlah teknik untuk menyampaikan informasi subjektif penulis tentang kecenderungan emosional teks serta untuk menyimpulkan perasaan dan sudut pandang penulis [3].

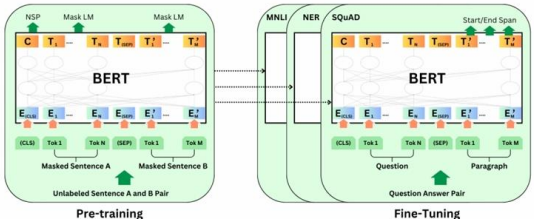
### 2.3. Aplikasi X

Aplikasi X adalah nama lain dari Twitter yang merupakan platform media sosial memungkinkan orang untuk secara terbuka mengekspresikan pikiran mereka [12].

Pada tahun 2023, X berada di peringkat 5 di antara platform media sosial yang paling sering digunakan orang Indonesia [2] untuk bereaksi terhadap berbagai tren dan sudut pandang di negara mereka.

### 2.4. *Bidirectional Encoder Representations from Transforms* (BERT)

Beberapa peneliti Google menciptakan arsitektur pembelajaran mesin BERT pada tahun 2018 [8] dengan tujuan melatih representasi dua arah yang mendalam dari teks yang tidak berlabel dengan mengkondisikan konteks kiri dan kanan di seluruh lapisan agar dapat membaca urutan kata lengkap secara bersamaan. Akibatnya, satu lapisan output tambahan dapat digunakan untuk menyempurnakan model BERT yang telah dilatih sebelumnya, membuat model yang dapat digunakan untuk segala jenis tugas PBA, termasuk tugas peringkasan teks otomatis. [9].




Gambar 1 Prosedur *pre-training* dan *Fine-tuning* pada arsitektur BERT [8]

Dari Gambar 1 merupakan *encoder* dan *decoder* pada masing-masing tingkat BERT, BERT memiliki enam tingkat *Transformers*. Pemrosesan BERT dimulai dengan kata yang memiliki representasi penyematan dari lapisan penyematan. Untuk menghasilkan representasi perantara baru, setiap lapisan menerapkan beberapa perhitungan *multihead* ke representasi kata dari lapisan sebelumnya [10].

Sebanyak 110 juta parameter, 12 lapisan, 768 ukuran tersembunyi, dan 12 kepala perhatian diri membentuk BERT *Base*. BERT *Large*, di sisi lain, memiliki 340 juta parameter, 24 lapisan, 1024 ukuran tersembunyi, dan 16 kepala perhatian diri. Inilah yang memungkinkan BERT untuk memahami kosakata, tata bahasa, dan hubungan tekstual secara lebih menyeluruh.

### 2.5. *Monolingual* (IndoBERT)

IndoBERT adalah model *pretrained* berdasarkan model yang dilatih menggunakan 4 miliar kata data teks bahasa Indonesia [4] dari berbagai platform *online*, termasuk Wikipedia, berita *online*, media sosial, artikel, *subtitle* rekaman video, dan *dataset* paralel yang kemudian dijuluki Indo4B. Wikipedia bahasa Indonesia, yang menyediakan 74 juta kata, artikel dari sumber terpercaya seperti Kompas, Tempo, dan Liputan6, yang terdiri dari 55 juta kata, dan Korpus Web Indonesia, yang mencakup 90 juta kata, adalah tiga saluran utama dari kumpulan data pelatihan ini [11].

### 2.6. *Multilingual* (mBERT)

Model *multilingual* BERT telah menetapkan standar baru untuk sistem NLP. Penerapannya dalam konteks *multi*-bahasa, dengan melatih suatu model menggunakan teks dari berbagai bahasa dengan satu kosa kata, telah mendorong kemajuan dalam pembelajaran lintas bahasa dan transfer pengetahuan. *Pretraining* berbasis BERT memberikan manfaat pada tugas-tugas generasi bahasa seperti terjemahan mesin [12].

M-BERT sebelumnya telah dilatih menggunakan teks Wikipedia dari 104 bahasa teratas yaitu bahasa dengan jumlah artikel terbanyak di Wikipedia. M-BERT menggunakan tujuan *pretraining* yang sama dengan BERT, yaitu *masked language model* dan *next sentence prediction* [7].

### 2.7. *Lexicon Based*

Karena proses penggunaannya yang praktis, pendekatan berbasis *Lexicon* sering digunakan untuk menilai sentimen di media sosial [10].

Kata-kata berbasis *lexicon* adalah kata-kata yang telah diberi bobot kata yang ditentukan oleh *lexicon*. Baik sikap positif maupun negatif termasuk dalam pembobotan yang dilakukan [13].

Penerapannya dimaksudkan untuk memastikan orientasi emosional suatu kata.

### 2.8. *Confusion Matrix*

*Confusion matrix* adalah matiks yang mencakup data dengan kategori aktual dan prediksi dari sistem klasifikasi. Performa suatu model dapat dipastikan dengan menganalisis data pada matriks. *Confusion matrix* terdiri dari beberapa nilai hasil prediksi, yaitu *True Positive* (TP), *False Positive* (FP), *True Negative* (TN), dan *False Negative* (FN) berdasarkan nilai tersebut maka dapat diketahui nilai *Accuracy*, *Precision*, *Recall*, dan *F-measure* [14].

a. *Accuracy* menunjukkan seberapa dekat temuan klasifikasi dengan nilai sebenarnya. Membuat perbandingan antara data yang diklasifikasikan dengan benar dan data total menghasilkan akurasi.

$$\text{accuracy} = \frac{TP + TN}{Tp + FP + TN + FN} \times 100\% \quad (1)$$

b. *Precision* merupakan tingkat akurasi yang menunjukkan seberapa dekat nilai berbeda setiap kali pengulangan dilakukan. Kita dapat menentukan kedekatan hasil antara informasi yang diminta dan respons sistem dengan melihat nilai akurasi.

$$precision = \frac{TP}{TP + FP} \times 100\% \quad (2)$$

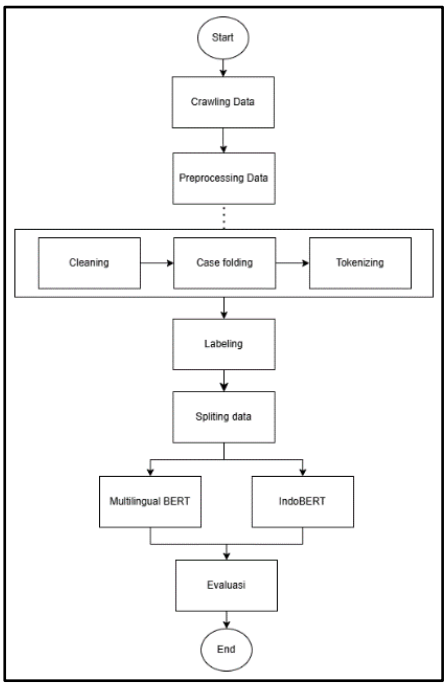
c. *Recall* adalah Nilai persentase model yang memprediksi data tidak berada di kelas, atau sensitivitas.

$$recall = \frac{TP}{TP + FN} \times 100\% \quad (3)$$

d. *F-measure* adalah komputasi yang menggabungkan presisi dan tingkat penarikan untuk menghasilkan informasi.

$$f - measure = 2 \times \frac{recall \times precision}{recall + precision} \times 100\% \quad (4)$$

## 3. METODE PENELITIAN




Gambar 2 Diagram alir metode penelitian

Pada Gambar 2. *Data crawling*, *preprocessing*, *data labeling*, *data splitting*, *model implementation*, dan *evaluation* adalah proses yang diikuti dalam siklus penelitian menggunakan teknik *Bidirectional Encoder Representation from Transformers* (BERT).

### 3.1. *Data Crawling*

Metode *data crawling* adalah proses pengumpulan data dengan mengambil informasi dari X melalui *Application Programming Interface* (API) yang telah disediakan X dan membuat kumpulan data teks yang telah diunggah di X. *Tweet* yang berisi kata kunci yang digunakan dalam proses *crawling* dianggap sebagai *tweet* yang disediakan pengguna. *Dataset* berisi *tweet* berbahasa Indonesia dengan data yang didapat 10.140 data. Kemudian akan dilakukan proses pembersihan data untuk dapat dilakukan proses *labeling* dan pengolahan *data train* dan *data test*.

### 3.2. *Preprocessing*

*Preprocessing* data memiliki beberapa tahapan seperti *cleaning, case folding, dan tokeninzing*. Pada tahap ini, proses ini berfungsi untuk merapikan *dataset* agar lebih mudah diproses pada tahap selanjutnya.

a. *Cleaning*
   Pembersihan data melibatkan penggunaan subfungsi *library* untuk membersihkan tanda baca, angka, simbol, ASCII, dan tautan. Subfungsi ini bekerja dengan mengganti *string* atau atribut kosong untuk *string* target berdasarkan pola yang ditentukan.
b. *Case folding*
   Proses ini menggunakan methode lower dari library string digunakan untuk mengubah data input menjadi huruf kecil. Jika *string* berisi karakter kapital atau huruf besar, karakter akan diubah menjadi huruf kecil.
c. *Tokenizing*
   Fungsi *word_tokenize* dari *library* nltk digunakan untuk melakukan proses token, yang memisahkan setiap kata dalam sebuah frasa.

### 3.3. *Data Labeling*

Proses *labeling* data dilakukan setelah dilakukannya tahap *preprocessing*. Dalam penelitian ini dilakukan labeling data menggunakan metode *Lexicon-based Sentiment Analisis.* Label yang diberikan adalah Posistif, Negatif, atau Netral didasari oleh nilai polaritas. Nilai polaritas pada *lexicon* memiliki rentang yang luas, pada penelitian ini rentang yang dimiliki yakni -62 sampai 39 yang dipengaruhi oleh bobot/sentimen yang dimiliki per kata, karena setiap kata memiliki makna sentimen yang berbeda berdasarkan konteks penggunaannya.

### 3.4. *Data Splitting*

Membagi *dataset* menjadi dua bagian atau lebih dikenal sebagai *data splitting*. Data pelatihan dan data pengujian adalah bentuk data yang telah di pisahkan. Sebagian dari kumpulan data yang telah dilatih untuk menghasilkan prediksi atau melakukan operasi dari algoritme lain sesuai dengan tujuan khusus mereka dikenal sebagai pelatihan data. Agar model yang dilatih dapat menemukan korelasi pada tahap evaluasi model. Sementara itu, data uji adalah data yang digunakan untuk melakukan pengujian data atau proses mengevaluasi data yang telah dilatih sebelumnya untuk menentukan seberapa efektif model dalam mengklasifikasikan sentimen baru.

### 3.5. ***Implementation*** BERT

Implementasi *Bidirectional Encoder Representation from Transformers* (BERT) memanfaatkan model dalam memahami konteks secara mendalam. Teks yang telah dibersihkan dan diberikan label akan diproses oleh model BERT untuk dilakukan *pre-trained*, yang kemudian di *fine-tune* agar dapat mengenali sentimen posistif, netral dan negatif. Proses ini melibatkan penambahan lapisan klasifikasi di atas model BERT untuk menghasilkan prediksi sentimen berdasarkan teks yang diberikan. BERT memiliki sistem tokenisasi sendiri yang meng-*encoder* data teks berdasarkan model yang digunakan, dalam hal ini menggunakan model IndoBERT dan mBERT yang kemudian dilakukan proses pengubahan format data agar dapat digunakan oleh *PyTorch.* Selanjutnya melakukan konfigurasi *hyperparameter* untuk digunakan dalam proses *train* yang menggunakan *Trainer* dari *Hugging face* berdasarkan data *PyTorch* yang telah diolah sebelumnya.

### 3.6. *Evaluation*

Pada tahap ini model yang telah dibuat dan dilatih akan dievaluasi menggunakan *Confussion Matrix*, kemudian akan menunjukkan banyaknya prediksi benar dan salah yang dihasilkan oleh model dibandingkan dengan hasil klasifikasi sebenarnya pada data uji pengukuran dengan *confussion matrix* meliputi *Accuracy, Precission,* dan *Recall.*

## 4. HASIL DAN PEMBAHASAN

### 4.1. *Data Crawling*

Proses *crawling* data memanfaatkan token dari Twitter API untuk dapat mengakses X dan melakukan *pengumpulan* data secara otomatis berdasarkan topik yang sesuai dengan topik penelitian. *Dataset* yang berhasil dikumpulkan sejumlah 10.140 data *tweet* yang tersimpan dalam format csv. Data tersebut terdiri dari 15 kolom data diantaranya, yaitu *convesation_id_str, created_at, favorite_count, full_text, ide_str, image_url, in_reply_to_screen_name, lang, location, quote_count, reply_count, retweet_count, tweet_url, user_id_str,* dan *username.*

Setelah menjalankan proses, data disimpan dalam format csv dengan nama *file* dataset.csv

| conversation_id_str | created_at | favorite_ | full_text |
|---|---|---|---|
| 1,74E+23 | Tue Jan 09 09:21:39 +0000 2024 | 1 | Calon Presiden nomor urut 1 Anies Baswedan menegaskan Indonesia harus menjadi penentu arah konstelasi global bukan hanya menjadi penonton. Demikian disampaikan Anies dalam debat ketiga Pilpres 2024 di Jakarta Minggu malam (7/1/2024). Selengkapnya: https://t.co/QhtFoRUPNq https://t.co/YL1ueOag9i |
| 1,74E+23 | Tue Jan 09 07:11:18 +0000 2024 | 3 | Sikap capres nomor 2 Prabowo Subianto diapresisi karena tidak terpancing capres lain untuk membuka data pertahanan Indonesia saat debat calon presiden (capres) Pemilihan Presiden atau Pilpres 2024 https://t.co/gufbsCUOpp Prabowo #NegarawanSejati https://t.co/kxq76itd8A |

Gambar 3 Hasil pengumpulan data

Pada Gambar 3. data yang ditampilkan berupa beberapa contoh data *yang* terdapat dalam *dataset* yang telah didapat dari proses *crawling* dengan *library tweet-harverst*.

### 4.2. *Data Pre-processing*

*Pre-procesing* dilakukan dengan memanfaatkan *library* Python. *Pre-processing* data dilakukan dengan tahapa *Cleaning* , *Case folding*, dan *Tokenizing*. Sehingga menghasilkan data bersih dan siap untuk lanjut pada proses berikutnya. Berikut *library* yang digunakan dalam tahap *pre-processing*.

a. *Cleaning*

Proses pembersihan yang dilakukan seperti menghapus tautan, angka, simbol, karakter khusus, *whitespace* dan emoji. Hasil dari pembersihan dapat dilihat pada tabel 1.

Tabel 1 Hasil tahap *cleaning*

| **Sebelum *Cleaning*** | **Setelah *Cleaning*** |
|---|---|
| Aktif Cerdaskan Publik Sosok Anies Baswedan Anugerah untuk Bangsa Indonesia. Makanya dia menyayangkan Anies tidak terpilih pada Pilpres 2024 kemarin. Padahal Anies akan memberikan pengaruh yang lebih besar lagi seandainya menjadi pemimpin nasional. https://t.co/gINXE0Y1PD https://t.co/Y1OunP1ikB | Aktif Cerdaskan Publik Sosok Anies Baswedan Anugerah untuk Bangsa Indonesia Makanya dia menyayangkan Anies tidak terpilih pada Pilpres kemarin Padahal Anies akan memberikan pengaruh yang lebih besar lagi seandainya menjadi pemimpin nasional |

b. *Case Folding*

Pada tahapan ini, setiap huruf kapital yang ada dalam kalimat akan diubah menjadi huruf kecil seluruhnya, sehingga bentuk huruf dalam kalimat menjadi seragam.

Tabel 2 Hasil tahap *case folding*

| **Sebelum *Case Folding*** | **Setelah *Case Folding*** |
|---|---|
| Aktif Cerdaskan Publik Sosok Anies Baswedan Anugerah untuk Bangsa Indonesia Makanya dia menyayangkan Anies tidak terpilih pada Pilpres kemarin Padahal Anies akan memberikan pengaruh yang lebih besar lagi seandainya menjadi pemimpin nasional | aktif cerdaskan publik sosok anies baswedan anugerah untuk bangsa indonesia makanya dia menyayangkan anies tidak terpilih pada pilpres kemarin padahal anies akan memberikan pengaruh yang lebih besar lagi seandainya menjadi pemimpin nasional |

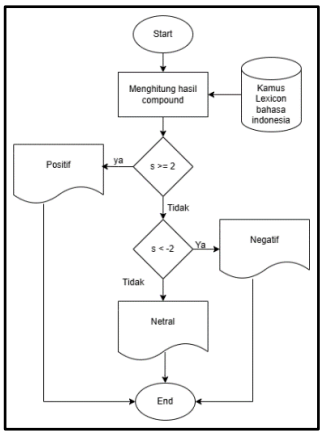
c. *Tokenizing*

Proses *tokenizing* atau tokenisasi ini menggunakan *function word*_tokenize dari *library* nltk dan dapat dilihat dari proses *tokenization* pada Tabel 3.

Tabel 3 Hasil tahap *tokenizing*

| **Sebelum *Tokenizing*** | **Setelah *Tokenizing*** |
|---|---|
| aktif cerdaskan publik sosok anies baswedan anugerah untuk bangsa indonesia makanya dia menyayangkan anies tidak terpilih pada pilpres kemarin padahal anies akan memberikan pengaruh yang lebih besar lagi seandainya menjadi pemimpin nasional | ['aktif', 'cerdaskan', 'publik', 'sosok', 'anies', 'baswedan', 'anugerah', 'untuk', 'bangsa', 'indonesia', 'makanya', 'dia', 'menyayangkan', 'anies', 'tidak', 'terpilih', 'pada', 'pilpres', 'kemarin', 'padahal', 'anies', 'akan', 'memberikan', 'pengaruh', 'yang', 'lebih', 'besar', 'lagi', 'seandainya', 'menjadi', 'pemimpin', 'nasional'] |

### 4.3. *Data Labeling*



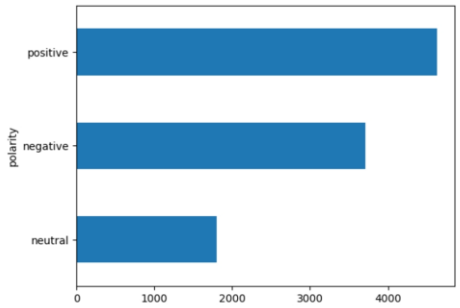

Gambar 4 Proses pelabelan data menggunakan *lexicon*

Dalam proses ini data yang digunakan yakni data yang sudah dilakukan *preprocessing*. Pada tahapan ini dibuat sebuah fungsi sentiment_analysisi_lexicon_indonesia yang berfungsi sebagai untuk mengembalikan skor sentiment dan polaritas berdasarkan kamus lexicon_positive dan kamus *lexicon_negative* [15] dengan skor yang digunakan adalah jika skor>=2, maka polaritas dianggap "*positive*", jika skor <=-2, maka polaritas dianggap "*negative*" , dan jika skor berada diantara skor tersebut makan polaritas dianggap "*neutral*". Data yang digunakan merupakan data hasil tokenisasi yang mana fungsi sentiment_analysisi_lexicon_indonesia diterapkan ke setiap baris pada kolom tokenize. Hasilnya, yang berupa skor dan polaritas yang disimpan dalam kolom *polarity_score* dan *polarity*.

| tokenize | polarity_score | polarity | label |
|---|---|---|---|
| ['kegiatan', 'cawapres', 'cak', 'imin', 'gibran', 'mahfud', 'md', 'jelang', 'pencoblosan'] | 6 | positive | 2 |
| ['cawapres', 'nomor', 'urut', 'mahfud', 'md', 'mengaku', 'optimistis', 'dapat', 'memenangkan', 'pilpres', 'jangan', 'lupa', 'tonton', 'obrolan', 'newsroom', 'spesial', 'pemilu', 'live', 'dari', 'berbagai', 'tps', 'di', 'indonesia', 'di', 'lt', 'mahfudmd', 'pemilukcm', 'jernihmemilih'] | -6 | negative | 0 |
| ['sumpa', 'mahfud', 'md', 'ni', 'sayang', 'bangettt', 'politic', 'powernya', 'dia', 'lemah', 'banget', 'tapi', 'individual', 'power', 'op', 'banget', 'asli', 'jadi', 'kek', 'klo', 'masuk', 'bursa', 'capres', 'cawapres', 'kek', 'kurang', 'bernilai', 'padahal', 'invaluable'] | 1 | neutral | 1 |

Gambar 5 Hasil Pelabelan Data Menggunakan *Lexicon*

Dari hasil pelabelan yang dilakukan menggunakan *lexicon* didapatkan hasil sebanyak 4.619 data *positive*, 3.719 data *negative*, dan 1802 data *neutral*. Label yang dimiliki data *akan* diubah dalam bentuk angka berupa 0 "*negative*", 1 "*neutral*", dan 2 "*positive*" guna mempermudah proses implementasi pada BERT.




Gambar 6 Distribusi persebaran *dataset* setelah pelabelan

### 4.4. *Data Splitting*

Dari total 10.140 data yang digunakan, data dibagi menjadi 2 bentuk data yakni data pelatihan dan data uji. Komposisinya adalah 80% atau 8.112 data pelatihan dan 20% atau 2.028 data untuk pengujian. Data pelatihan dimaksudkan untuk melatih model BERT sehingga model dapat mengenali kelas-kelas yang ada dalam data. Data uji ditujukan untuk menguji hasil pelatihan dari model BERT yang sudah digunakan.

### 4.5. *Implementation BERT*

Setelah melakukan pelabelan, selanjutnya dilakukan proses *pemodelan* data dengan menentukan *tool* yang digunakan. Pada tahap pemodelan BERT, peneliti menggunakan *library* Transformers, yakni *BertTokenizer*, dan *BertForSequenceClassification* sebagai model. Data yang digunakan dilakukan tokenisasi berdasarkan model *BertTokenizer* dari model *IndoBERT* dan *M-BERT* yang kemudian dikonversi kedalam bentuk *PyTorch* sehingga kompatibel digunakan dengan *Trainer* dari *Hugging Face*. Proses pengubahan data ke bentuk *PyTorch* memanfaatkan token yang telah di-*encoding* (*train_encodings* dan *test_encodings*) beserta labelnya. Proses *Trainer* dari *Hugging Face* menggunakan pengaturan terperinci dari kelas *TrainningArguments* seperti jumlah maksimal *epoch* yang digunakan, ukuran *batch*, evaluasi setelah setiap *epoch* selesai. Pada proses *training* dilakukan perhitungan *embeding* dan *logits* menggunakan *library BertForSequenceClassification* dan perhitungan probabilitas dihitung menggunakan fungsi *compute* dari BERT. Hasil dari pengujian model yang dilakukan adalah mengklasifikasikan negatif, natural, dan positif dengan menggunakan algoritma *indobert-base-p2* [16] dan *bert-base-multilingual-cased* [17].Pada tabel 4. Proses implementasi BERT pada sebuah teks beserta hasil perhitungannya.

Tabel 4 Contoh implementasi BERT pada sebuah data

| Label | Nilai |
|---|---|
| *Text* | Prabowo kepemimpinannya tegas berwibawa cocok jadi presiden tapi terkadang gampang marah |
| *Token IDs* | [ 101 10975 7875 5004 2080 17710 5051 4328 8737 3981 10695 3148 |

| Label | Nilai |
|---|---|
| | 8915 12617 2022 102] |
| *Attention Mask* | [1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1] |
| *Embeddings* | [[0.37454012 0.95071431 0.73199394 ... 0.56988968 0.76245869 0.87676564] [0.34208175 0.8212573 0.11063174 ... 0.272624 0.4135491 0.12188609] [0.18114935 0.68111785 0.18143835 ... 0.57204864 0.51266773 0.29348882] ... [0.87040046 0.55823175 0.32479892 ... 0.01536625 0.42644381 0.60619354] [0.55776476 0.65133611 0.15407573 ... 0.1410598 0.70932385 0.85456517] [0.68585449 0.01787848 0.21623289 ... 0.76511151 0.76401263 0.29244546]] |
| *Logits* | [188.90950456 193.30952258] |
| *Probabilities* | [0.01212822 0.98787178] |
| *Predicted Label* | *Positive* |

### 4.6. *Evaluation*

Evaluasi dilakukan untuk mengukur performa model dengan mempertimbangkan pengaruh ukuran *batch size* dan jumlah *epoch* terhadap akurasi prediksi. Uji coba dilakuakn dengan variasi *batch size* (16, 32, dan 64) dan jumlah *epoch* (10, 20, dan 30) pada kedua model, untuk mengidentifikasi model yang menghasilkan performa terbaik.

Tabel 5 Perbandingan model IndoBERT dan mBERT

| ***Epoch*** | **IndoBERT Batch** | | | **mBERT Batch** | | |
|---|---|---|---|---|---|---|
| | 16 | 32 | 64 | 16 | 32 | 64 |
| 10 | 82 | 83 | 81 | 82 | 81 | 81 |
| 20 | 83 | 82 | 81 | 82 | 82 | 81 |
| 30 | 82 | 84 | 83 | 82 | 81 | 81 |

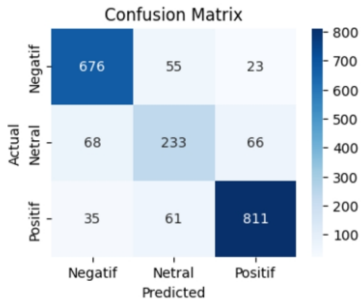
Pada Tabel 5. tersebut menunjukkan akurasi model IndoBERT dan mBERT dalam analisis sentimen berdasarkan jumlah *batch size* (16, 32, 64) dan *epoch* (10, 20, 30). Untuk IndoBERT, akurasi tertinggi dicapai pada *batch size* 32 dan *epoch* 30 dengan nilai 84%, menunjukkan bahwa kombinasi tersebut memberikan hasil terbaik dalam memahami data berbahasa Indonesia. Namun, akurasi IndoBERT menunjukkan sedikit ketidakstabilan, terutama pada *batch size* 16 dan 64, dengan akurasi berkisar antara 81-84%. Sementara itu, mBERT memiliki akurasi yang lebih stabil pada kisaran 81-82% tanpa perbedaan signifikan antar *batch size* dan *epoch*, yang menunjukkan bahwa model ini tidak terlalu dipengaruhi oleh perubahan parameter. Secara keseluruhan, IndoBERT lebih unggul dalam akurasi dibandingkan mBERT, menjadikannya pilihan yang lebih tepat untuk tugas analisis sentimen dalam bahasa Indonesia, karena model ini dioptimalkan khusus untuk bahasa tersebut.

Setelah dilakukan perbandingan performa akurasi pada berbagai kombinasi *batch size* dan jumlah *epoch*, evaluasi lebih lanjut difokuskan pada *hyperparameter fine-tuning batch size* 32 dengan 30 *epoch* untuk mengetahui nilai *confusion matrix*.

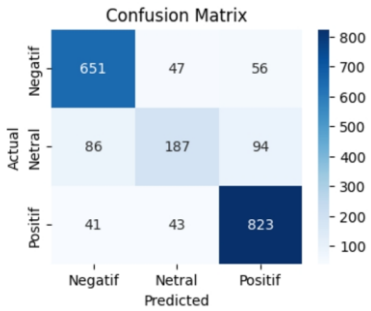
Tabel 6 *Classification Report*

| Model | Akurasi | Presisi | *Recall* | Skor F1 |
|---|---|---|---|---|
| *IndoBERT* | 0.84 | 0.75 | 0.80 | 0.78 |
| *mBERT* | 0.81 | 0.69 | 0.78 | 0.73 |

Berdasarkan Tabel 6. menunjukkan IndoBERT memiliki akurasi yang lebih baik daripada mBERT, menunjukkan bahwa lebih akurat dalam memprediksi label sentimen secara keseluruhan, menurut tabel yang menampilkan metrik evaluasi model. Akurasi model IndoBERT juga lebih tinggi dari mBERT, menunjukkan keunggulan IndoBERT dalam memprediksi kelas positif, negatif, dan netral secara akurat dengan tingkat kesalahan yang berkurang. Dengan daya ingat yang sedikit lebih besar daripada mBERT, IndoBERT menunjukkan kemampuan yang unggul untuk menangkap data nyata yang relevan di setiap kategori sentimen. Fakta bahwa skor F1 IndoBERT lebih besar dari mBERT menunjukkan bahwa skor ini menyeimbangkan ingatan dan presisi dengan lebih baik.




Gambar 7 *Confusion Matrix IndoBERT*




Gambar 8 *Confusion Matrix mBERT*

Melihat hasil *confusion matrix*, IndoBERT cenderung memiliki lebih banyak prediksi yang benar di setiap kelas sentimen (positif, negatif, netral) dibandingkan mBERT, yang terlihat dari nilai lebih tinggi di diagonal utama dan nilai lebih rendah pada

kesalahan klasifikasi. Meskipun mBERT juga memiliki hasil yang cukup baik, performanya sedikit lebih rendah karena sifatnya yang dilatih untuk multibahasa, yang membuatnya kurang optimal untuk bahasa Indonesia dibandingkan IndoBERT yang dioptimalkan secara khusus untuk bahasa tersebut.

IndoBERT memberikan performa yang lebih baik dalam analisis sentimen bahasa Indonesia dibandingkan mBERT, dengan hasil prediksi yang tepat lebih banyak dan nilai akurasi, presisi, *recall*, dan skor F1 yang lebih tinggi. Oleh karena itu, IndoBERT lebih direkomendasikan untuk tugas-tugas analisis sentimen dalam konteks bahasa Indonesia.

## 5. KESIMPULAN DAN SARAN

Penelitian menunjukkan IndoBERT lebih unggul dari mBERT untuk analisis sentimen *tweet* Bahasa Indonesia di platform X, dengan akurasi 84% vs. 81%. IndoBERT juga mencatat presisi 75%, *recall* 80%, dan F1-Score 78%, dibandingkan mBERT yang meraih 69%, 78%, dan 73%. Keunggulan IndoBERT terletak pada pemahaman konteks Bahasa Indonesia. Hal ini menunjukkan IndoBERT lebih baik untuk analisis sentimen terkait konteks spesifik Bahasa Indonesia, Pengaturan *hyperparameter* seperti *epoch*, *batch size*, *learning rate*, dan dropout rate memengaruhi performa. *Fine-tuning* IndoBERT, penambahan data pelatihan, serta pengujian pada *f*lain dapat meningkatkan hasil. IndoBERT cocok untuk analisis sentimen di platform X, sementara mBERT tetap dapat dioptimalkan.

## DAFTAR PUSTAKA


[1] R. Vindua and A. U. Zailani, "Analisis Sentimen Pemilu Indonesia Tahun 2024 Dari Media Sosial Twitter Menggunakan Python," *JURIKOM (Jurnal Riset Komputer)*, vol. 10, no. 2, p. 479, Apr. 2023, doi: 10.30865/jurikom.v10i2.5945.

[2] A. Jaya, "Analisis Sentimen Pandangan Public Profesi PNS (Pegawai Negeri Sipil) dari Twiter menerapkan indonesian Roberta Base Sentiment Classifier," *Indonesian Journal of Data and Science*, vol. 4, no. 1, pp. 38–44, 2023, doi: 10.56705/ijodas.v4i1.66.

[3] B. Kurniawan, A. Ari Aldino, and A. Rahman Isnain, "SENTIMEN ANALISIS TERHADAP KEBIJAKAN PENYELENGGARA SISTEM ELEKTRONIK (PSE) MENGGUNAKAN ALGORITMA BIDIRECTIONAL ENCODER REPRESENTATIONS FROM TRANSFORMERS (BERT)," 2022. [Online]. Available: http://jim.teknokrat.ac.id/index.php/JTSI

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

[5] Vidya Chandradev, I Made Agus Dwi Suarjaya, and I Putu Agung Bayupati, "Analisis Sentimen Review Hotel Menggunakan Metode Deep Learning BERT," *Jurnal Buana Informatika*, vol. 14, no. 02, pp. 107–116, 2023, doi: 10.24002/jbi.v14i02.7244.

[6] K. Mawaddah, "Analisis Dan Perbandingan Klasifikasi Hate Speech Menggunakan Metode Fitur Manual Svm Dan Bert," 2023.

[7] Z. Wang, K. K, S. Mayhew, and D. Roth, "Extending Multilingual BERT to Low-Resource Languages," Apr. 2020, [Online]. Available: http://arxiv.org/abs/2004.13640

[8] J. Khatib Sulaiman, M. Putri, T. Edy Sutanto, and S. Inna, "Studi Empiris Model BERT dan DistilBERT: Analisis Sentimen pada Pemilihan Presiden Indonesia," *Indonesian Journal of Computer Science Attribution*, vol. 12, no. 5, pp. 2023–2972.

[9] F. V. P. Samosir, H. Toba, and M. Ayub, "BESKlus : BERT Extractive Summarization with K-Means Clustering in Scientific Paper," *Jurnal Teknik Informatika dan Sistem Informasi*, vol. 8, no. 1, Apr. 2022, doi: 10.28932/jutisi.v8i1.4474.

[10] R. Mas, R. W. Panca, K. Atmaja1, and W. Yustanti2, "Analisis Sentimen Customer Review Aplikasi Ruang Guru dengan Metode BERT (Bidirectional Encoder Representations from Transformers)," *Jeisbi*, vol. 02, no. 3, pp. 55–62, 2021.

[11] P. Sayarizki and H. Nurrahmi, "Implementation of IndoBERT for Sentiment Analysis of Indonesian Presidential Candidates," *Journal on Computing*, vol. 9, no. 2, pp. 61–72, 2024, doi: 10.34818/indojc.2024.9.2.934.

[12] S. Wu and M. Dredze, "Are All Languages Created Equal in Multilingual BERT?," May 2020, [Online]. Available: http://arxiv.org/abs/2005.09093

[13] Mahira Putri, T. E. Sutanto, and S. Inna, "Studi Empiris Model BERT dan DistilBERT Analisis Sentimen pada Pemilihan Presiden Indonesia," *Indonesian Journal of Computer Science*, vol. 12, no. 5, pp. 2972–2980, 2023, doi: 10.33022/ijcs.v12i5.3445.

[14] J. Homepage *et al.*, "MALCOM: Indonesian Journal of Machine Learning and Computer Science Sentiment Analysis of Online Lectures in Indonesia from Twitter *Dataset* Using InSet Lexicon Analisis Sentimen terhadap Perkuliahan Daring di Indonesia dari Twitter *Dataset* Menggunakan InSet Lexicon," vol. 1, pp. 24–33, 2021.

[15] angelmetanosaa, "dataset," GitHub. Accessed: Dec. 02, 2024. [Online]. Available: https://github.com/angelmetanosaa/dataset

[16] IndoBenchmark, " IndoBERT-base-P2," Hugging Face. Accessed: Dec. 02, 2024. [Online]. Available:

https://huggingface.co/indobenchmark/indobert-base-p2
[17] google-bert, "bert-base-multilingual-cased," Hugging Face. Accessed: Dec. 02, 2024. [Online]. Available: https://huggingface.co/google-bert/bert-base-multilingual-cased