

ANALISIS SENTIMEN PENGGUNA TERHADAP PLATFORM MEDIA SOSIAL X (TWITTER) PASCA AKUISISI BERBASIS WEB DENGAN METODE BERT

Bulan Attin Nurazizah¹, Agus Suharto²

^{1,2} Program Studi Sistem Informasi

Fakultas Ilmu Komputer Ilmu Komputer, Universitas Pamulang
Jl. Raya Puspitek No.11, Buaran, Serpong, Kota Tangerang Selatan, Banten 15310

E-mail: bulanattinn1001@gmail.com¹, dosen0359@unpam.ac.id²

ABSTRAK

ANALISIS SENTIMEN PENGGUNA TERHADAP PLATFORM MEDIA SOSIAL X (TWITTER) PASCA AKUISISI BERBASIS WEB DENGAN METODE BERT. Twitter mengalami perubahan signifikan setelah diakuisisi oleh Elon Musk kemudian berubah nama menjadi X pada tahun 2022, yang memicu berbagai reaksi beragam dari penggunanya. Untuk mengetahui beragam reaksi tersebut, diperlukan sistem dashboard yang dapat mengolah tweets dengan metode analisis sentimen. Penelitian ini menggunakan metode *Bidirectional Encoder Representations from Transformers* (BERT) dengan melalui proses *preprocessing*, labeling, split data, dan evaluasi model. Proses fine-tuning dilakukan menggunakan proporsi dataset sebesar 80% untuk data latih dan 20% untuk data uji. Hasil evaluasi menunjukkan model mampu mencapai akurasi sebesar 51%, dengan *precision* 29%, *recall* 36%, dan *F1-score* 32%. Secara keseluruhan model mendeteksi sentimen negatif dengan sangat baik, namun kinerja pada sentimen positif dan netral sangat rendah sehingga perlu peningkatan. Hal tersebut juga menunjukkan bahwa pandangan publik terhadap akuisisi Twitter cenderung bersifat negatif.

Kata Kunci : Analisis Sentimen, BERT, Akuisisi, Twitter, X.

ABSTRACT

USER SENTIMENT ANALYSIS TOWARDS SOCIAL MEDIA PLATFORM X (TWITTER) POST WEB-BASED ACQUISITION USING THE BERT METHOD. Twitter underwent significant changes after being acquired by Elon Musk and renamed X in 2022, which triggered various reactions from its users. To understand these reactions, a dashboard system capable of processing tweets using sentiment analysis methods is required. This study uses the *Bidirectional Encoder Representations from Transformers* (BERT) method through preprocessing, labeling, data splitting, and model evaluation processes. The fine-tuning process was carried out using a dataset proportion of 80% for training data and 20% for test data. The evaluation results showed that the model was able to achieve an accuracy of 51%, with a precision of 29%, a recall of 36%, and an F1-score of 32%. Overall, the model detected negative sentiment very well, but its performance on positive and neutral sentiment was very low and needed improvement. This also shows that public opinion about Twitter acquisition tends to be negative.

Keywords: Sentiment Analysis, BERT, Acquisition, Twitter, X.

1. PENDAHULUAN

1.1 Latar Belakang

Pada tahun 2022, Twitter mengalami transformasi besar ketika perusahaan tersebut diambil alih oleh Elon Musk. Banyak perubahan telah terjadi pada platform ini, mulai dari aturan penggunaan dan perubahan nama menjadi X. Reaksi penggunanya beragam, mulai dari dukungan terhadap inovasi hingga kritik keras mengenai kebijakan-kebijakan baru yang dianggap tidak memprioritaskan privasi, moderasi konten, atau pengetatan akses API. Sebagian pengguna menyambut positif perubahan tersebut, sementara yang lain mengungkapkan kecemasannya tentang arah masa depan platform tersebut.

Tanggapan pengguna terhadap perubahan ini dapat terlihat melalui beragam pendapat yang disampaikan di platform X tersebut. Agar dapat memahami dampak dari akuisisi tersebut, penelitian ini menerapkan analisis sentimen dengan menggunakan metode BERT (*Bidirectional Encoder Representations from Transformers*) pada tweet-tweet yang terkait. Solusi dari masalah ini adalah dengan menggunakan model BERT, untuk menganalisis dan mengklasifikasikan sentimen secara otomatis, sehingga memberikan wawasan yang lebih jelas mengenai persepsi publik terhadap perubahan di Twitter setelah akuisisi tersebut, dan menghasilkan akurasi yang tinggi.

Berdasarkan latar belakang yang telah dijabarkan, maka penulis tertarik untuk melakukan penelitian yang berjudul ANALISIS SENTIMEN PENGGUNA TERHADAP PLATFORM MEDIA SOSIAL X (TWITTER) PASCA AKUISISI BERBASIS WEB DENGAN METODE BERT. Penelitian ini diharapkan dapat memberikan gambaran yang lebih jelas tentang bagaimana persepsi dan pengalaman pengguna di platform tersebut.

1.2 Identifikasi Masalah

Berdasarkan latar belakang yang telah dipaparkan diatas, maka dapat diidentifikasi permasalahan yang ada yaitu:

a. Akuisisi platform media sosial sering memunculkan reaksi yang bervariasi dari pengguna, namun belum banyak penelitian yang mengukur dampak langsung akuisisi terhadap sentimen pengguna.

b. Belum ada alat berbasis web yang dapat menampilkan dashboard data hasil olah akuisisi twitter.

1.3 Rumusan Masalah

Dengan mempertimbangkan masalah yang diidentifikasi diatas, masalah berikut dapat dirumuskan:

- a. Bagaimana membangun sistem berbasis web untuk melakukan analisis sentimen pengguna terhadap platform media sosial X pasca akuisisi?
- b. Bagaimana model BERT dapat digunakan dalam mengklasifikasikan sentimen (positif, negatif, netral) pada tweet dengan akurasi yang baik?

1.4 Batasan Penelitian

Batasan masalah yang digunakan dalam pembahasan bertujuan untuk membuat pembahasan lebih fokus, terarah dan sesuai dengan tujuan yang akan dicapai. Adapun batasan penelitian ini ditetapkan sebagai berikut:

- a. Penelitian ini dibatasi pada tweet yang diambil dari media sosial X berjumlah sekitar 1000 data yang akan digunakan.
- b. Penelitian ini hanya menganalisis tweet berupa teks berbahasa Indonesia yang dipublikasikan setelah akuisisi.
- c. Metode yang digunakan dalam penelitian ini adalah metode BERT pada klasifikasi sentimen ulasan menjadi tiga kategori utama: positif, negatif, dan netral. Sistem yang dibangun adalah sistem berbasis web yang berfokus pada visualisasi hasil analisis sentimen, bukan pada analisis teknis atau perubahan fitur dari platform.

1.5 Manfaat Penelitian

Manfaat dari Tugas Akhir ini, diharapkan akan memberikan dampak yang positif bagi semua pihak yang berkaitan diantaranya sebagai berikut:

- a. Bagi Universitas

Manfaat penelitian bagi Universitas Pamulang yaitu berguna sebagai bahan pertimbangan dan referensi bagi mahasiswa yang akan menyusun skripsi. Selain itu, ada kemungkinan untuk memasukan referensi beserta masukan bagi pihak-pihak yang bersangkutan untuk informasi mengenai analisis sentimen.

b. Bagi Penulis

Manfaat penelitian bagi penulis yaitu mengasah kemampuan teknis penulis dari ilmu yang telah dipelajari, serta membantu penulis memahami cara kerja analisis sentimen.

2. TINJAUAN PUSTAKA

2.1 Analisis Sentimen

Sentiment Analysis merupakan perpaduan dari data mining dan *text mining*, atau sebuah cara yang digunakan untuk mengolah berbagai opini yang diberikan oleh konsumen atau para pakar melalui berbagai media, mengenai sebuah produk, jasa ataupun sebuah instansi. *Sentiment analysis* merupakan sebuah metode yang digunakan untuk memahami, mengekstrak data opini, dan mengolah data textual secara otomatis untuk mendapatkan sebuah sentimen yang terkandung dalam sebuah opini. Pada *Sentiment analysis* terdiri dari 3 jenis opini, yaitu opini positif, opini negatif dan opini netral, sehingga dengan *sentiment analysis* perusahaan atau instansi yang terkait dapat mengetahui respon masyarakat terhadap suatu pelayanan atau produk, melalui feedback masyarakat maupun para ahli (Apif, S, 2021).

2.2 BERT (*Bidirectional Encoder Representations from Transformers*)

BERT memiliki keunggulan yang signifikan, salah satunya yaitu kemampuannya memahami konteks secara mendalam karena arsitekturnya yang bidirectional, memungkinkan BERT untuk membaca kalimat dari kedua arah dan memahami makna kata berdasarkan konteks keseluruhan. Hal ini membuat BERT unggul dalam menangani nuansa sentimen dan variasi bahasa yang sering kali kompleks dalam teks, seperti tweet di Twitter. Selain itu, BERT dapat di-*pre-train* dengan data umum dan di-*fine-tune* untuk tugas spesifik seperti analisis sentimen, menghasilkan akurasi yang tinggi.

3. METODE

3.1 Metodologi Penelitian

Adapun metode pengumpulan data yang digunakan dalam penelitian ini adalah sebagai berikut:

a. Observasi

Observasi dilakukan dengan cara mendapatkan data-data dan fakta dari pengamatan terkait dengan permasalahan yang diteliti langsung pada kolom tweet di media sosial X.

b. Studi Pustaka

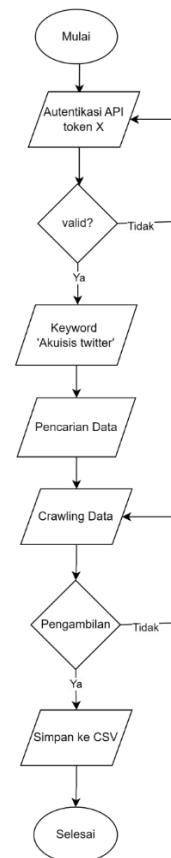
Studi pustaka melibatkan pencarian, dan menganalisis terhadap literatur, jurnal ilmiah, buku, artikel, atau sumber informasi lain yang terkait dengan topik penelitian.

3.2 Metode Pengembangan Sistem

Adapun metode pengembangan sistem yang digunakan pada penelitian ini yaitu sebagai berikut:

a. Crawling Data

Data diambil dari hasil cuitan atau twit pengguna X dengan menggunakan *tools* Tweet Harvest.

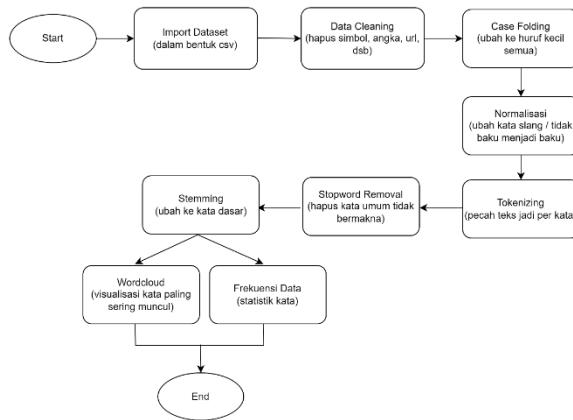


Gambar 1 Flowchart Crawling Data

b. Preprocessing Data

Setelah mendapatkan dataset yang sudah dapat digunakan, tahap penelitian berikutnya adalah *preprocessing* data. Ini dilakukan dengan mengubah data mentah atau data teks yang tidak terstruktur menjadi data yang terstruktur sehingga

ga pemrosesan sistem lebih mudah dilakukan.



Gambar 2 Flowchart Preprocessing Data

1. Import Data

Pada tahap pertama, data yang akan digunakan ke dalam sistem di *import* atau dimasukan dalam format CSV. Data ini berisi kumpulan tweet yang akan dianalisis lebih lanjut.

2. Data Cleaning

Tahap ini, teks dibersihkan dengan menghapus elemen yang tidak penting seperti angka, tanda baca, simbol, emoticon, URL, dan karakter unik. Tujuannya adalah agar menjadi bersih dan konsisten.

3. Case Folding

Selanjutnya teks diubah menjadi huruf kecil. Tahap ini dilakukan untuk membuat bentuk kata menjadi sama, sehingga tidak terjadi perbedaan makna hanya karena penggunaan huruf kapital.

4. Normalisasi

Proses mengubah kata-kata tidak baku, singkatan, atau bahasa slang menjadi bentuk baku yang lebih tepat. Misalnya "gk" menjadi "tidak".

5. Tokenizing

Tahap *tokenizing* dilakukan untuk memecah teks menjadi unit kata atau token. Misalnya "akuisisi twitter oleh elon musk" akan diubah menjadi "akuisisi", "twitter", "oleh", "elon", "musk".

6. Stopword removal

Pada tahap ini, proses menghapus kata-kata umum yang tidak memiliki makna penting. Seperti "yang", "atau", dsb. Tujuannya adalah untuk memproses hanya kata-kata bermakna yang relevan.

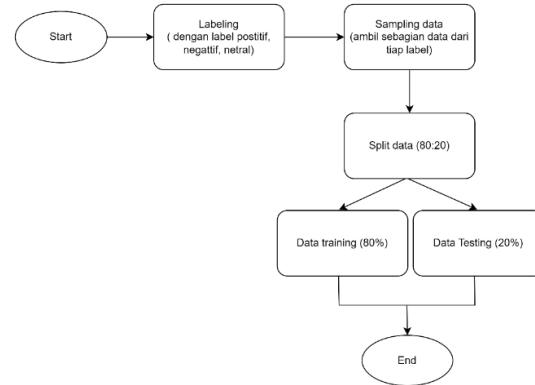
7. Stemming

Selanjutnya dilakukan *stemming* untuk mengembalikan kata ke bentuk dasarnya. Hal ini penting agar kata dengan makna sama dianggap sama.

8. Wordcloud dan Frekuensi

Data Hasil *preprocessing* kemudian divisualisasikan dalam bentuk *wordcloud*, dari kata-kata yang paling sering muncul. Selain itu, menampilkan juga frekuensi data kata dalam bentuk grafik.

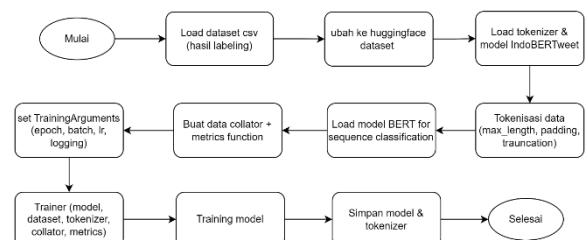
c. Labeling dan Split Data



Gambar 3 Flowchart Labeling dan Split Data

Pada tahap ini, proses pengolahan data dimulai dengan labeling, dimana setiap teks di labeli dengan kategori sentimen tertentu, seperti positif, negatif, atau netral. Label ini sangat penting karena akan digunakan oleh model untuk mengidentifikasi pola sentimen pada data. Selanjutnya, sampling data dilakukan untuk mengambil sebagian data dari setiap kategori sentimen. Tujuan dari tahap ini adalah untuk memastikan bahwa jumlah data yang terkandung dalam setiap label seimbang.

d. Implementasi Model BERT

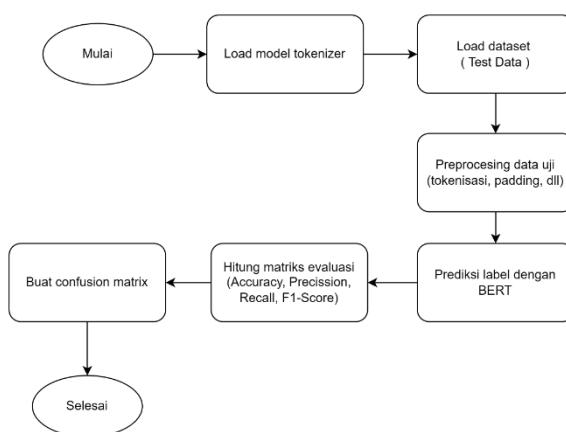


Gambar 4 Flowchart Implementasi Model BERT

Implementasi BERT dimulai dengan memuat dataset hasil pelabelan dalam format CSV. Kemudian, format ini untuk menjadi kompatibel dengan *library* Transformers, format ini diubah menjadi *HuggingFace*. Selanjutnya, model IndoBERTweet dan tokenizer digunakan sebagai dasar pemrosesan teks berbahasa Indonesia. Selanjutnya, model BERT for Sequence Classification dimuat untuk menangani klasifikasi sentimen.

Tahap selanjutnya adalah menyiapkan data collator dan *metrics*. Parameter pelatihan ditentukan melalui *TrainingArguments*. *Trainer* menggabungkan semua komponen tersebut dan menggunakan dataset yang telah disiapkan untuk menjalankan proses pelatihan model. Setelah pelatihan selesai, model dan tokenizer disimpan agar dapat digunakan kembali untuk menguji sistem klasifikasi sentimen.

e. Evaluasi Model



Gambar 5 Flowchart Evaluasi Model

Proses ini dimulai dengan memuat model *tokenizer* dan kumpulan data uji yang sudah disiapkan. Kemudian, data uji melalui tahap *preprocessing*, termasuk tokenisasi, padding, dan penyesuaian format untuk sesuai dengan input model. Setelah data diproses, model BERT digunakan untuk menghasilkan prediksi label sentimen. Kemudian, hasil prediksi dinilai dengan menghitung matriks evaluasi, seperti akurasi, presisi, *recall*, dan *F1-score*. Selain itu, dibuat juga *confusion matrix* untuk menunjukkan secara lebih detail apakah distribusi prediksi benar atau salah. Untuk menghitung matriks evaluasi menggunakan rumus sebagai berikut:

- Accuracy*, adalah cara melihat seberapa akurat model dapat mengklasifikasikan dengan benar tingkat kedekatan prediksi yang aktual. Berikut rumus perhitungan akurasi:

$$Accuracy = \frac{TP + TN}{TOTAL} \times 100\%$$

- Precision*, yaitu prediksi benar positif dibandingkan dengan keseluruhan hasil yang diprediksi positif. Berikut perhitungan *precision*:

$$Precision = \frac{TP}{TP + FP} \times 100\%$$

- Recall*, yaitu prediksi benar positif dibandingkan dengan keseluruhan data yang benar positif keberhasilan model dalam menemukan kembali sebuah informasi. Berikut rumus perhitungan *Recall*:

$$Recall = \frac{TP}{TP + FN} \times 100\%$$

- F1-score* adalah metrik gabungan yang menggabungkan presisi dan *recall* dalam satu nilai tunggal. Berikut rumus perhitungan *F1-score*:

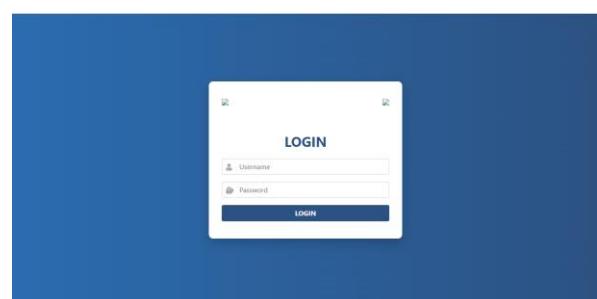
$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \times 100\%$$

4. IMPLEMENTASI DAN PENGUJIAN

4.1 Implementasi Antar Muka (*User Interface*)

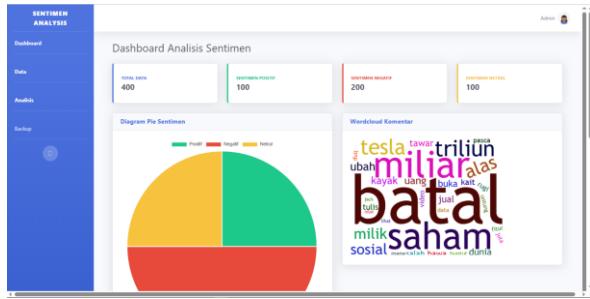
Implementasi antarmuka akan menampilkan tampilan program aplikasi yang dibangun dengan perancangan yang telah dibuat. Tampilan antarmuka dari perancangan UI yang telah dibuat dapat dilihat di bawah ini:

- Halaman *Login*



Gambar 6 Halaman *Login*

- Halaman *Dashboard*



Gambar 7 Halaman Dashboard

c. Halaman Upload Data

No	Komentar Asli	Komentar CLean	Filter Label
1	Elon Musk sempatnya ada akun twitter miliknya sendiri. Tenggolong kere ganteng apa dia? Bungkong macu bu nyepi sama view	komunikasi dengan akun twitter anak tamis ganteng yang guna bingung that reply view	negatif
2	Dianggap Lengkap Terlalu Keterlalu Pengujian Mengenai Elon Musk Tak Jadi Akhir Twitter #Investor #Investor #Investor #Investor	anggap lengkap tentu jeng manger elon musik akunis twitter	neutral
3	Orang pernah tweet oleh Elon Musk masih berlaku. Semoga memanggil pengguna drama bek hoffter elon musik lengkap proses akuisisi musik amancam batul jeng aja	drama bek hoffter elon musik lengkap	negatif
4	Elon Musk Gagal Akuisisi Twitter Tola Ternate https://t.co/2QZDf7RQ3Q	elon musik gagal akuisisi hettar tola saret	negatif
5	Elon Musk dan Eksekutif Twitter Dikarantina Bersempara Akuisisi https://t.co/fzDf09H9	elon musik akuisisi hettar kacur temu bulus	positif

Gambar 8 Halaman Upload Data

d. Halaman Analisis

Gambar 9 Halaman Analisis

e. Halaman Backup

No	Tanggal	Nama File	Deskripsi	Aksi
1	2023-08-11	test	test	Cetak PDF Hapus

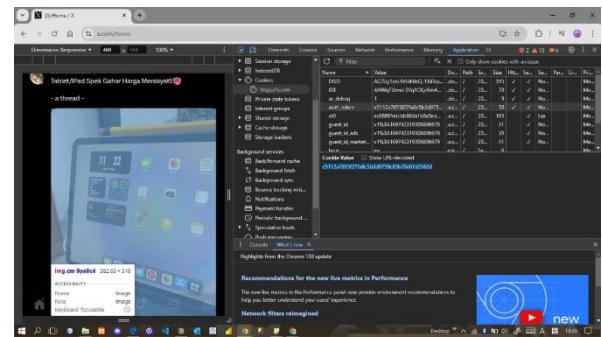
Gambar 10 Halaman Backup

4.2 Implementasi dan Evaluasi Model BERT

Proses implementasi dilakukan mulai dari pengumpulan data, *preprocessing*, labeling, implementasi BERT, hingga evaluasi model.

4.2.1 Pengumpulan Dataset

Pada penelitian ini, peneliti melakukan pengumpulan dataset dengan teknik *crawling* menggunakan *tools tweet-harvest* yaitu alat yang digunakan untuk melakukan *crawling* data, dibangun menggunakan Node.js. Pertama dengan menyalin auth token dari twitter, seperti pada gambar dibawah ini:



Gambar 11 X Auth-Token

Setelah auth token di dapatkan, kemudian menginstal beberapa *package* python seperti pandas. Selain itu, dilakukan instalasi Node.js. Kemudian melakukan instalasi, skrip python dapat menggunakan *tweet-harvest* untuk mengatur proses autentikasi dengan menggunakan token akses yang telah diperoleh sebelumnya. Setelah autentikasi berhasil, tahap selanjutnya adalah melakukan *crawling* data tweet dengan mengisi keyword 'akuisisi twitter' berbahasa Indonesia dengan limit 1000 tweet. Jumlah data yang didapat yaitu berjumlah 1012 tweet.

Tabel 1 Keyword dan Parameter Crawling Data

Keyword	Limit	Hasil
Akuisisi Twitter	1000	1012

Gambar 12 Hasil *Crawling* Dataset

4.2.2 *Preprocessing Data*

Berikut tahapan *Preprocessing Data* yang dilakukan:

a. *Cleaning Data*

Berikut dataset sesudah dilakukan proses *cleaning*:

Tabel 2 *Cleaning Data*

full_text	data_cleaning
Cokkk video lawas bangett. Wayy before covid Elon wae durung akuisisi twitter hahaha Saake sing wes podo inden tekan saiki rakono kabare wkwkwkwk	cokkk video lawas bangett wayy before covid elon wae durung akuisisi twitter hahaha saake sing wes podo inden tekan saiki rakono kabare wkwkwkwk
PERSETAN ORANG YANG BIKIN ELON MUSK AKUISISI TWITTER	persetan orang yang bikin elon musk akuisisi twitter
@Power_Ranger_17 @realmadridindo1 Bro baru maen Twitter semenjak di akuisisi Elon	bro baru maen twitter semenjak di akuisisi elon
@WatchmenID Udah aku blok dari awal dia akuisisi twitter.	udah aku blok dari awal dia akuisisi twitter

b. *Case Folding*

Setelah pembersihan data selesai, case folding dilakukan yang berarti seluruh teks diubah menjadi huruf kecil. Hasilnya disimpan ke dalam kolom case_folding.

Tabel 3 *Case Folding*

data_cleaning	case_folding
cokkk video lawas banggett wayy before covid elon wae durung akuisisi twitter hahaha saake sing wes podo inden tekan saiki rakono kabare wkwkwkwk	cokkk video lawas banggett wayy before covid elon wae durung akuisisi twitter hahaha saake sing wes podo inden tekan saiki rakono kabare wkwkwkwk
persetan orang yang bikin elon musk akuisisi twitter	persetan orang yang bikin elon musk akuisisi twitter
bro baru maen twitter semenjak di akuisisi elon	bro baru maen twitter semenjak di akuisisi elon
udah aku blok dari awal dia akuisisi twitter	udah aku blok dari awal dia akuisisi twitter

c. Normalisasi

Dalam proses ini, kamus kata tidak baku dipetakan ke kata baku. Kata yang diubah dan hasil perubahannya disimpan dalam kolom hasil_normalisasi.

Tabel 4 Normalisasi

case_folding	hasil_normalisasi
cokkk video lawas bangett wayy before covid elon wae du- rung akuisisi twitter hahaha saake sing wes podo inden tekan saiki rakono kabare wkwkwkwk	cokkk video lawas banget wayy before covid elon wae du- rung akuisisi twitter hahaha saake sing wes podo inden tekan saiki rakono kabare wkwkwkwk
persetan orang yang bikin elon musk akui- sisi twitter	persetan orang yang bikin elon musk akui- sisi twitter
bro baru maen twitter semenjak di akuisisi elon	bro baru main twitter semenjak di akuisisi elon
udah aku blok dari awal dia akuisisi twit- ter	sudah aku blok dari awal dia akuisisi twit- ter

d. Tokenizing

Dalam proses tokenisasi, setiap kalimat atau teks dipisahkan menjadi daftar kata-kata berdasarkan spasi.

Tabel 5 Tokenizing

hasil_normalisasi	tokenizing
cokkk video lawas banget wayy before covid elon wae durung akuisisi twitter hahaha saake sing wes podo inden tekan saiki rakono kabare wkwkwkwk	['cokkk', 'video', 'lawas', 'banget', 'wayy', 'before', 'covid', 'elon', 'wae', 'durung', 'akuisisi', 'twitter', 'hahaha', 'saake', 'sing', 'wes', 'podo', 'inden', 'tekan', 'saiki', 'rakono', 'kabare', 'wkwkwkwk']
persetan orang yang bikin elon musk akuisisi twitter	['persetan', 'orang', 'yang', 'bikin', 'elon', 'musk', 'akuisisi', 'twitter']
bro baru main twitter semenjak di akuisisi elon	['bro', 'baru', 'main', 'twitter', 'semenjak', 'di', 'akuisisi', 'elon']
sudah aku blok dari awal dia akuisisi twi- ter	['sudah', 'aku', 'blok', 'dari', 'awal', 'dia', 'akuisisi', 'twitter']

f. Stopword removal

Proses menghapus kata-kata umum seperti "dan", "yang", "di", dan sebagainya yang tidak penting untuk analisis dikenal sebagai *stopword removal*. Untuk penelitian ini, daftar *stopword* bahasa Indonesia dari *library* NLTK digunakan. Setiap token dicocokkan dengan daftar *stopword*, sehingga hanya kata-kata penting untuk analisis sentimen yang tersedia. Tahap ini menghasilkan hasil yang disimpan dalam kolom *stopword_removal*.

Tabel 6 Stopword Removal

tokenizing	stopword_removal
['cokkk', 'video', 'lawas', 'banget', 'wayy', 'before', 'covid', 'elon', 'wae', 'durung', 'akuisisi', 'twitter', 'hahaha', 'saake', 'sing', 'wes', 'podo', 'inden', 'tekan',	['cokkk', 'video', 'lawas', 'banget', 'wayy', 'before', 'covid', 'elon', 'wae', 'durung', 'akuisisi', 'twitter', 'hahaha', 'saake', 'sing', 'wes', 'podo', 'inden', 'tekan', 'saiki', 'rakono', 'kabare', 'wkwkwkwk']

'saiki', 'rakono', 'kabare', 'wkwkwkwk']	'tekan', 'saiki', 'rakono', 'kabare', 'wkwkwkwk']
['persetan', 'orang', 'yang', 'bikin', 'elon', 'musk', 'akuisisi', 'twitter']	['persetan', 'orang', 'bikin', 'elon', 'musk', 'akuisisi', 'twitter']
['bro', 'baru', 'main', 'twitter', 'semenjak', 'di', 'akuisisi', 'elon']	['bro', 'main', 'twitter', 'semenjak', 'akuisisi', 'elon']
['sudah', 'aku', 'blok', 'dari', 'awal', 'dia', 'akuisisi', 'twitter']	['blok', 'akuisisi', 'twitter']

f. Stemming

Untuk melakukan *stemming* pada kata-kata yang telah melewati tahap penghapusan stopword. Proses ini juga menggunakan fungsi *stem()* dari library *sastrawi* untuk setiap kata. Kemudian hasilnya digabungkan kembali ke dalam kalimat. *Stemming* data disimpan dalam kolom *stemming_data*.

Tabel 7 Stemming

stopword_removal	stemming
['cokkk', 'video', 'lawas', 'banget', 'wayy', 'before', 'covid', 'elon', 'wae', 'durung', 'akuisisi', 'twitter', 'hahaha', 'saake', 'sing', 'wes', 'podo', 'inden', 'tekan', 'saiki', 'rakono', 'kabare', 'wkwkwkwk']	cokkk video lawas banget wayy before covid elon wae durung akuisisi twitter hahaha saake sing wes podo inden tekan saiki rakono kabare wkwkwkwk
['persetan', 'orang', 'bikin', 'elon', 'musk', 'akuisisi', 'twitter']	persetan orang bikin elon musk akuisisi twitter
['bro', 'main', 'twitter', 'semenjak', 'akuisisi', 'elon']	bro main twitter semenjak akuisisi elon
['blok', 'akuisisi', 'twitter']	blok akuisisi twitter

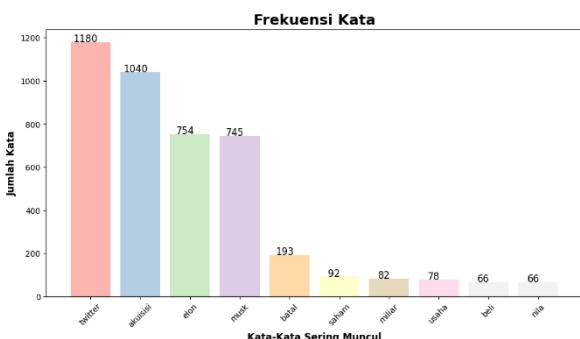
g. Wordcloud dan Frekuensi Kata

Setelah tahap *preprocessing*, yang mencakup data *cleaning*, penyusunan case, normalisasi, tokenisasi, penghapusan *stopword*, dan *stem*-

ming, selesai, visualisasi dilakukan untuk mengevaluasi kata-kata yang paling sering ditemukan pada data. Dua jenis visualisasi yang digunakan adalah sebagai berikut:



Gambar 13 Wordcloud setelah *Preprocessing*



Gambar 13 Frekuensi Kata

4.3.3 Labeling dan Split Data

Proses pemberian label dataset pada penelitian ini menggunakan metode kamus sentimen yang berisi daftar kata positif dan negatif dalam bahasa Indonesia. Kamus sentimen diam-bil dari repository github <https://github.com/fajri91/InSet>, dengan dua kamus yaitu positive.csv dan negative.csv. Dimana kata-kata yang termasuk dalam tweet akan dicocokan dengan kata-kata yang ada di dalam kamus sentimen, kemudian label sentimen untuk tweet tersebut akan dipilih berdasarkan label sentimen kata-kata yang cocok. Jika jumlah kata positif lebih banyak dari negatif, maka komentar diberi label positif. Dan jika jumlah keduanya seimbang atau tidak ditemukan kata dari kamus, maka komentar diklasifikasikan sebagai netral.

Distribusi awal label yang ditampilkan pada output menunjukkan ketidakseimbangan jumlah data, dimana label 0 (negatif) mendominasi dengan 705 data, sedangkan label 1 (positif)

dan label 2 (netral) masing-masing berjumlah 108 dan 199 data. Ketidakseimbangan ini dapat berpotensi menyebabkan lebih bias terhadap terhadap model selama proses pelatihan. Untuk mengatasi hal tersebut, dilakukan proses sampling ulang dengan jumlah yang seimbang yang mana untuk kelas negatif diambil sebanyak 200 data, sedangkan untuk kelas positif dan netral masing-masing 100 data.

Setelah mendapatkan data label yang seimbang untuk masing-masing kelas melalui sampling data, maka tahap selanjutnya adalah membagi data menjadi data latih (training) dan data uji (testing) dengan menggunakan library scikit-learn. Pembagian data dilakukan dengan proporsi 80:20 yang menghasilkan 80% data latih dan 20% data uji, dimana agar menjaga distribusi label sentimen tetap proporsional. Hasil yang diperoleh ialah sebagai berikut:

Tabel 8 Hasil *Data Splitting*

Kategori	Label	Tipe Data	Jumlah
Negatif	0	Train	160
		Test	40
Positif	1	Train	80
		Test	20
Netral	2	Train	80
		Test	20

4.3.4 Implementasi Model

Pada tahap Implementasi model BERT menggunakan pre-trained model *indolem/indobertweet-base-uncased* dari *Hugging Face* yang sudah dilatih untuk data bahasa Indonesia. Pada penelitian ini menggunakan Trainer API untuk proses pelatihan, *library Hugging Face Transformer* digunakan untuk menerapkan model BERT. Agar teks dapat diubah menjadi format numerik yang dapat diproses oleh model, dataset

terlebih dahulu melalui tahap tokenization dengan BERT tokenizer.

4.3.5 Evaluasi Model

Setelah model BERT diimplementasikan, selanjutnya adalah mengevaluasi kinerjanya. Metrik-metrik seperti *precision*, *recall*, *accuracy*, dan *F1score* digunakan untuk menilai kinerja dan performa model. Perhitungan dilakukan dengan menggunakan fungsi *classification_report* dari library Scikit-learn. Untuk memudahkan membaca pembagian nilai True Positif (TP), True Negative (TN), False Positive (FP), dan False Negative (FN) dari setiap kelas, berikut adalah tabel ringkasan *confusion matrix*. Kelas dengan indeks 0 menunjukkan sentimen negatif, 1 menunjukkan positif, dan 2 menunjukkan netral.

Tabel 9 Tabel Metrik

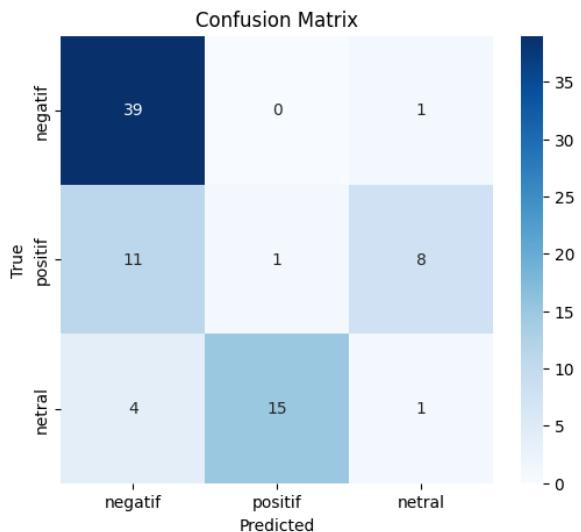
Kelas	True Positive (TP)	True Negative (TN)	False Positive (FP)	False Negative (FN)
0 (Negatif)	39	25	25	1
1 (Positif)	1	45	15	19
2 (Netral)	1	51	9	19

Classification Report:				
	precision	recall	f1-score	support
negatif	0.72	0.97	0.83	40
positif	0.06	0.05	0.06	20
netral	0.10	0.05	0.07	20
accuracy			0.51	80
macro avg	0.29	0.36	0.32	80
weighted avg	0.40	0.51	0.45	80

Gambar 14 Classification Report

Dari gambar 14 hasil klasifikasi menunjukkan model menghasilkan akurasi keseluruhan sebesar 51% dengan *precision* 29%, *recall* 36%, dan *F1-score* 32%. Pada tingkat perkelas, sentimen negatif memiliki *precision* 72%, *recall* 97%,

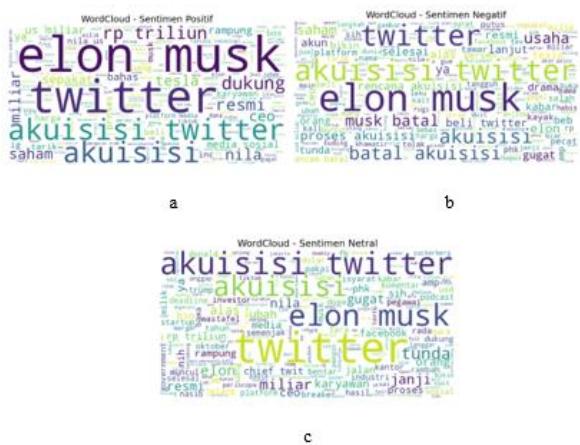
F1score 83%. Sentimen positif memiliki *precision* 6%, *recall* 5%, *F1-score* 6%. Dan sentimen netral memiliki *precision* 10%, *recall* 5%, dan *F1-score* 7%, menunjukkan model cukup akurat dan dapat menemukan sentimen negatif dengan cukup konsisten, tetapi masih kesulitan untuk menemukan sentimen positif dan netral.



Gambar 15 Confusion matrix

Pada gambar 15, model menghasilkan logits yaitu nilai mentah sebelum fungsi aktivasi softmax. Pada logits dilakukan proses argmax untuk menemukan indeks kelas dengan probabilitas tertinggi.

Evaluasi model memanfaatkan *confusion matrix* untuk menampilkan perbandingan antara hasil prediksi dan nilai sebenarnya. Dari 40 data di kelas negatif, 39 berhasil diprediksi dengan benar, dan hanya 1 yang salah dikategorikan sebagai netral. Dari 20 data di kelas positif, hanya 1 data yang terprediksi dengan benar, sedangkan 11 data salah diklasifikasikan sebagai negatif dan 8 data sebagai netral. Dari 20 data di kelas netral, model hanya memprediksi benar 1 data, sementara 15 data salah dikategorikan sebagai positif dan 4 data sebagai negatif. Pola ini menunjukkan bahwa model memiliki kemampuan prediksi yang sangat baik pada kelas netral dan negatif. Namun, sentimen positif dari negatif masih sulit dibedakan.



Gambar 16 Wordcloud Hasil Evaluasi Model

Gambar a menunjukkan Wordcloud untuk sentimen positif yang terdiri dari kata-kata seperti “twitter”, “elon”, “musk”, “akuisisi” dan “resmi”, ini menunjukkan bahwa pengguna mendukung atau memiliki perspektif yang positif tentang masalah yang dibahas. Gambar b menunjukkan Wordcloud untuk sentimen negatif, dimana kata-kata seperti “batal”, “gugat”, “tuntut”, dan “ancam” yang menunjukkan bahwa komentar negatif cenderung mengandung kekhawatiran, ketidakpuasan, atau reaksi keras terhadap masalah akuisisi Twitter. Gambar c menampilkan Wordcloud untuk sentimen netral, dimana kata-kata seperti “resmi”, “triliun”, “saham”, dan “karyawan” menunjukkan informasi faktual dan netral tentang proses akuisisi, seperti akor yang terlibat, dan dampak umum terhadap perusahaan, semuanya tanpa opini emosional.

5 KESIMPULAN

Berdasarkan penelitian yang dilakukan mengenai analisis sentimen pengguna terhadap platform media sosial x (twitter) pasca akuisisi berbasis WEB dengan metode BERT, penulis dapat mencapai kesimpulan berikut:

- Sistem dibangun menggunakan model bert yang telah di fine-tuning untuk mengklasifikasi sentimen (positif, negatif, netral) dari tweet berbahasa Indonesia. Sistem ini dilengkapi antarmuka web untuk unggah data, klasifikasi otomatis, serta visualisasi hasil berupa *confusion matrix*, pie chart, dan *wordcloud*.
- Model BERT digunakan dengan melakukan fine-tuning pada dataset tweet berbahasa Indonesia yang telah diberi label sentimen. Pada penelitian ini, proses fine-tuning dilakukan

menggunakan proporsi dataset sebesar 80% data latih dan 20% data uji. Hasil evaluasi menunjukkan model mampu mencapai akurasi sebesar 51%, dengan *precision* 29%, *recall* 36%, dan F1score 32. Pada tingkat perkelas, sentimen negatif memiliki *precision* 72%, *recall* 97%, *F1-score* 83%. Sentimen positif memiliki *precision* 6%, *recall* 5%, *F1-score* 6%. Dan sentimen netral memiliki *precision* 10%, *recall* 5%, dan *F1-score* 7%. Secara keseluruhan model mendeteksi sentimen negatif dengan sangat baik, namun kinerja pada sentimen positif dan netral sangat rendah sehingga perlu peningkatan. Hal tersebut juga menunjukkan bahwa opini masyarakat cenderung negatif terhadap suatu kejadian akuisisi twitter.

DAFTAR PUSTAKA

- Abdurrazaq, M. A., & Tjiong, E. L. (2022). Analisis Sentimen KUHP Baru Pada Data Twitter Menggunakan Model BERT. *Jurnal Komunikasi, Sains dan Teknologi*.
- Alizah, M. D., & Nugroho, A. (2020). Sentimen Analisis Terkait Lockdown pada Sosial Media. *IJSE – Indonesian Journal on Software Engineering*.
- Aljabar, A., & Karomah, B. M. (2024). Mengungkap Opini Publik: Pendekatan BERT-based caused untuk Analisis Sentimen pada Komentar Film. *Journal of System and Computer Engineering (JSCE)*.
- Atmajaya, D., Febrianti, A., & Darwis, H. (2023). Metode SVM dan Naive Bayes untuk Analisis Sentimen ChatGPT di Twitter. *Indonesian Journal of Computer Science*.
- Budianto, I., & Anwar, S. N. (2022). ANALISIS SENTIMENT PENGGUNA TWITTER MENGENAI PROGRAM VAKSINASI COVID-19 MENGGUNAKAN ALGORITMA NAÏVE BAYES. *Jurnal Teknologi Informasi*.
- Elhan, A., Hardhienata, M. K., Herdiyeni, Y., Wijaya, S. H., & Adisantoso, J. (2022). Analisis Sentimen Pengguna Twitter terhadap Vaksinasi COVID-19 di Indonesia menggunakan Algoritme Random Forest dan BERT. *Jurnal Ilmu Komputer Agri-Informatika*.
- Hafiz, Y. A., & Anggoro, D. A. (2024). ANALISIS SENTIMEN MASYARAKAT TERHADAP ISU KELANGKAAN BERAS BERDASARKAN MEDIA SOSIAL X

DENGAN PENDEKATAN BIDIRECTIONAL ENCODER REPRESENTATION FROM TRANSFORMERS (BERT). Universitas Muhammadiyah Surakarta Library.

- [8]. Helmiyah, S., & Verdian, A. (2024). Analisis Sentimen Terhadap Minat Belajar pada Tayangan Acara CoC by Ruangguru Berdasarkan Tweets Menggunakan Metode NLP dan Model BERT. JURNAL PENDIDIKAN STKIP ROSALIA LAMPUNG.
- [9]. Pamungkas, A. (2024). Penerapan BERT (Bidirectional Encoder Representations from Transformers) dengan MATLAB. Retrieved from pemrogramanmatlab.com: <https://pemrogramanmatlab.com>
- [10]. Rahman, S., Sembiring, A., Siregar, D., Khair, H., Prahmana, I. G., Puspadi, R., & Zen, M. (2023). PYTHON : DASAR DAN PEMROGRAMAN BERORIENTASI OBJEK. Tahta Media Group.
- [11]. Saraswati, N. P., Yudistira, N., & Adikara, P. P. (2023). ANALISIS SENTIMEN TERHADAP PERUNDUNGAN SIBER PADA TWITTER MENGGUNAKAN ALGORITMA BIDERCTIONAL ENCODER REPRESENTATION TRANSFORMER (BERT). Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer.
- [12]. Sari, S. (2021). Sentiment Analysis Against Beauty Shaming Comments on Twitter Social. IJIRSE: Indonesian Journal of Informatic Research and Software Engineering.
- [13]. Sjoraida, D. F., Guna, B. W., & Yudhakusuma, D. (2024). Analisis Sentimen Film Dirty Vote Menggunakan BERT (Bidirectional). Jurnal JTIK (Jurnal Teknologi Informasi dan Komunikasi) .
- [14]. Sofi, N., Sulistyorini, T., & Nazarudin, M. (2023). Analisis Sentimen Masyarakat Pengguna Media Sosial Twitter Terhadap Motogp Mandalika Lombok Menggunakan Metode Bidirectional Encoder Representation From Transformers (BERT) . Jurnal Informasi, Sains, dan Teknologi