

ANALISIS SENTIMEN *HATE SPEECH* MENGENAI CALON WAKIL PRESIDEN INDONESIA MENGGUNAKAN ALGORITMA BERT

Elvika Alya Junita¹⁾, Ryan Randy Suryono*²⁾.

1. Sistem Informasi, Fakultas Teknik dan Ilmu Komputer, Universitas Teknokrat Indonesia, Indonesia
2. Sistem Informasi, Fakultas Teknik dan Ilmu Komputer, Universitas Teknokrat Indonesia, Indonesia

Article Info

Kata Kunci: Analisis Sentimen; BERT; *Hate Speech*

Keywords: BERT; *Hate Speech*; *Sentimen Analysis*

Article history:

Received 29 September 2024

Revised 13 Oktober 2024

Accepted 4 November 2024

Available online 4 December 2024

DOI :

<https://doi.org/10.29100/jipi.v9i4.5544>

* Corresponding author.

Ryan Randy Suryono

E-mail address:

ryan@teknokrat.ac.id

ABSTRAK

Indonesia sebagai negara demokratis mengalami momen penting menjelang pemilihan umum, khususnya presiden dan wakil presiden. Persaingan politik yang sengit diwarnai dengan beragam pandangan dan retorika politik. Namun, munculnya *hate speech* sebagai bentuk ekstrem dari ekspresi politik mengancam stabilitas sosial dan integritas demokrasi. *Hate speech* dapat mengganggu harmoni masyarakat, mempengaruhi proses pemilihan umum dengan menyebarkan informasi palsu, dan merusak suasana politik. Analisis sentimen sangat penting dalam mendeteksi dan menangani *hate speech* terkait dengan calon Wakil Presiden Indonesia. Penelitian ini menggunakan data Twitter untuk mengeksplorasi opini masyarakat terhadap calon presiden dengan kata kunci Imin, Gibran, dan Mahfud MD sebanyak 2692 data. Hasil eksperimen menunjukkan bahwa algoritma BERT memiliki tingkat akurasi yang sangat tinggi, mencapai rata-rata 100% pada dataset Gibran Rakabuming Raka dan Mahfud MD. Analisis juga menunjukkan bahwa proporsi sentimen positif terhadap calon wakil presiden menunjukkan kecenderungan bahwa Muhaimin Iskandar mencapai tingkat akurasi 98,17%.

ABSTRACT

Indonesia, as a democratic country, experiences a crucial moment approaching the general election, particularly for the presidential and vice-presidential positions. Intense political competition is marked by various viewpoints and political rhetoric. However, the emergence of hate speech as an extreme form of political expression threatens social stability and the integrity of democracy. Hate speech can disrupt societal harmony, influence the electoral process by spreading false information, and tarnish the political atmosphere. Sentiment analysis is crucial in detecting and addressing hate speech related to the Indonesian Vice-Presidential candidates. This research utilizes Twitter data to explore public opinion on presidential candidates using the keywords Imin, Gibran, and Mahfud MD, amounting to 2692 data points. The test results indicate that the BERT algorithm achieves very high accuracy, averaging 100% on the datasets of Gibran Rakabuming Raka and Mahfud MD. The analysis also suggests that the proportion of positive sentiment towards the vice-presidential candidates indicates a tendency for Muhaimin Iskandar to have a higher proportion compared to other presidential candidates, such as Muhaimin Iskandar, who achieved an accuracy rate of 98.17%.

I. PENDAHULUAN

Indonesia sebagai negara demokratis yang mengadopsi sistem pemerintahan yang melibatkan partisipasi publik, menyaksikan momen penting dalam perjalanannya ketika mendekati pemilihan umum, terutama pemilihan presiden dan wakil presiden. Persaingan politik yang ketat sering kali menjadi ciri khas dari periode ini, dengan ruang publik dipenuhi dengan beragam pandangan, opini, dan retorika politik dari berbagai pihak. Namun, di tengah dinamika politik ini, munculnya *hate speech* atau ujaran kebencian sebagai bentuk ekstrem dari ekspresi politik menjadi ancaman serius terhadap stabilitas sosial dan integritas demokrasi [1]. *Hate speech* tidak hanya merusak harmoni masyarakat, tetapi juga dapat mempengaruhi proses pemilihan umum dengan menyebarkan informasi palsu, memicu konflik, dan memperkeruh suasana politik.

Analisis sentimen perlu dilakukan karena merupakan sarana penting dalam memahami, mendeteksi, dan menangani *hate speech* terkait dengan calon Wakil Presiden Indonesia [2]. Melalui analisis sentimen, kita dapat memperoleh pemahaman yang lebih dalam tentang pandangan dan sikap masyarakat terhadap calon tersebut, serta

mendeteksi hate speech yang dapat merusak proses demokrasi. Analisis sentimen merupakan pendekatan yang memungkinkan ekstraksi makna dari teks secara sistematis. Dengan memanfaatkan teknik-teknik analisis bahasa alami, kita dapat mengidentifikasi apakah suatu ulasan bersifat positif, netral, atau negatif [3].

Algoritma BERT (*Bidirectional Encoder Representations from Transformers*) dipilih karena keunggulannya dalam memahami konteks secara *bidirectional*, yang memungkinkan deteksi nuansa kompleks dan makna tersirat dalam kalimat. BERT, yang dilatih melalui *pre-training* dan *fine-tuning*, dapat disesuaikan dengan dataset khusus sehingga meningkatkan akurasi dalam konteks spesifik. Keunggulan BERT dalam menangani nuansa bahasa seperti sarkasme dan ironi serta mengurangi kesalahan klasifikasi membuatnya unggul dibandingkan model *unidirectional* seperti RNN atau LSTM.

Penelitian terdahulu menyoroiti masalah penggunaan kata-kata hate speech oleh netizen di media online, yang berpotensi menimbulkan masalah hukum bagi mereka. Penelitian ini fokus pada pendeteksian hate speech dalam komentar di portal berita online menggunakan Neural Language Processing dan metode Support Vector Machine (SVM). Hasilnya menunjukkan bahwa SVM dapat menghasilkan akurasi 53,88%, recall 49,69%, precision 48,77%, classification error 46,12%, dan f-measure 49,23%. Studi ini dapat menjadi panduan bagi portal berita untuk menerapkan sistem filtering guna meminimalisir kasus hate speech di masa depan [4].

Penelitian yang dilakukan oleh Atmaja dan Yustanti menggunakan model BERT pre-trained untuk analisis sentimen terhadap komentar pengguna Ruang Guru di Google Play Store. Data yang digunakan dalam penelitian ini berjumlah 5437 records. Dengan proporsi data latih dan uji 70:30, hasil evaluasi menunjukkan akurasi 99%, F1 score 98.9%, presisi 64.13%, dan recall 60.51%. Hasil ini menunjukkan bahwa penggunaan model BERT dalam analisis sentimen dapat memberikan hasil yang sangat baik dalam mengklasifikasikan komentar pengguna [5]. Penelitian selanjutnya menggunakan total 2.886.047 data review hotel dari seluruh Indonesia yang dikumpulkan dari website penyedia layanan booking hotel terbesar di Indonesia, dan menggunakan model BERT yang telah di-fine-tuning untuk menganalisis sentimen positif dan negatif dari ulasan pengunjung hotel di Indonesia, dengan akurasi mencapai 91,40% dan *f1*-score sebesar 90,51% [6].

Berdasarkan hasil analisis sentimen terhadap layanan akademik dan non-akademik Universitas Terbuka pada media sosial Twitter(X) dan TikTok, dari 685 data yang diperoleh, terdapat 281 data dengan sentimen positif dan 404 data dengan sentimen negatif. Data komentar diambil dari media sosial Twitter(X) dan TikTok menggunakan python pada Google Collab dan software WebHarvy. Hasil nilai akurasi tertinggi yang diperoleh dari analisis sentimen terhadap layanan akademik dan non-akademik Universitas Terbuka menggunakan algoritma BERT adalah 90% [7].

Penelitian terdahulu menggambarkan perubahan besar dalam interaksi manusia dengan dunia digital melalui metaverse. Pemerintah dan industri berupaya memahami perkembangan metaverse untuk menerapkan digitalisasi di berbagai sektor. Analisis sentimen publik terhadap metaverse dilakukan menggunakan data dari media sosial X, dengan dua model klasifikasi teks, Support Vector Machine dan Random Forest, serta optimasi SMOTE. Hasil menunjukkan Random Forest memiliki akurasi 91% dan Support Vector Machine 90%, menandakan keunggulan Random Forest. Optimasi SMOTE meningkatkan kemampuan kedua model dalam mengenali sentimen positif, meskipun terjadi trade-off antara recall dan precision karena keterbatasan model [8].

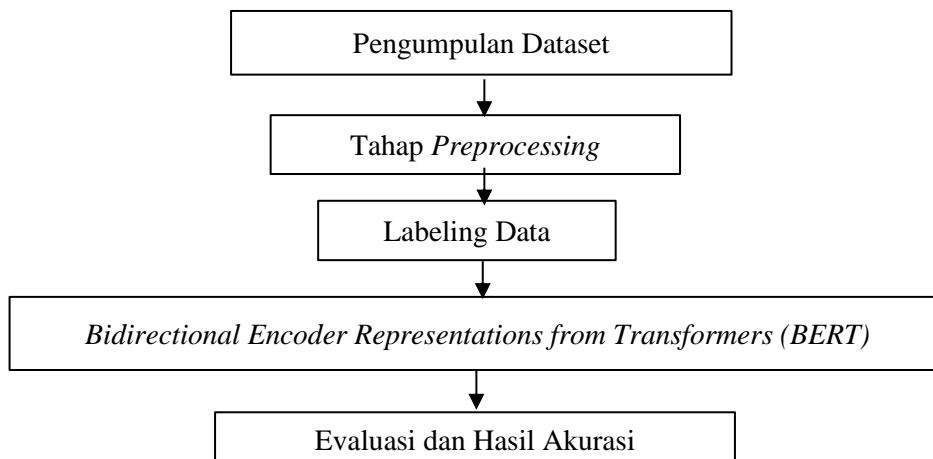
Penelitian terdahulu mengkaji respons publik terhadap kedatangan pengungsi Rohingya di Indonesia melalui analisis sentimen. Tujuan utamanya adalah memahami dinamika opini masyarakat serta membandingkan kinerja dua metode analisis sentimen, yakni Support Vector Machine (SVM) dan Naïve Bayes. Studi ini menganalisis 3350 tweet dengan pembagian data 70% untuk pelatihan dan 30% untuk pengujian. Hasilnya menunjukkan bahwa SVM memiliki akurasi 76%, sementara Naïve Bayes mencapai akurasi 70%, mengindikasikan bahwa SVM lebih efektif dalam memprediksi sentimen publik dengan tingkat kesalahan yang lebih rendah [9].

Berdasarkan latar belakang tersebut peneliti akan menganalisis sentimen *hate speech* mengenai calon Wakil Presiden Indonesia menggunakan algoritma BERT yang bertujuan untuk mendeteksi hate speech yang berkaitan dengan calon tersebut dengan akurasi tinggi, menganalisis sentimen umum masyarakat terhadapnya, memahami dinamika politik dan sosial yang terkait, mengidentifikasi tren dan pola perilaku dalam hate speech untuk pengembangan strategi penanganan yang lebih efektif, serta menyusun rekomendasi kebijakan bagi pemerintah, lembaga pengawas, dan masyarakat sipil guna mengatasi *hate speech* dan mempromosikan dialog di ruang publik digital.

II. METODE PENELITIAN

Pada tahapan ini akan diuraikan mengenai metode yang digunakan pada penelitian ini. Metode penelitian melalui beberapa langkah yang dimulai dari pengumpulan data, dengan melakukan *scrapping* data *Twitter* dan menggunakan data dari sumber pustaka sebagai bahan referensi. Langkah selanjutnya adalah pengolahan data pada tahap *preprocessing*. *Preprocessing* terdiri dari tahap *cleansing*, *case folding*, *tokenisasi*, dan *stopwords*. Setelah

itu, dilakukan klasifikasi teks dan analisis hasil klasifikasi. Berikut ilustrasi alur penelitian ini dapat dilihat pada Gambar 1.



Gambar 1. Tahapan Penelitian

2.1 Metode Pengumpulan Data

Penelitian ini mengambil data *tweet* dan komentar pengguna aplikasi sosial media Twitter. Data tweet dilakukan melalui cara *crawling* pada twitter. Data yang diperoleh berjumlah 2692 tweet dengan kata kunci Imin, Gibran, dan Mahfud. Untuk mengumpulkan data menggunakan metode *crawling* data menggunakan *library python*.

2.2 Tahap Preprocessing

Pre-processing merupakan tahap pengolahan data mentah sehingga menjadi data yang siap digunakan dalam prosedur *text mining* [10]. Langkah ini sangat krusial dalam *text mining*, karena bertujuan untuk mengolah dataset mentah yang belum terstruktur dan tidak siap untuk diklasifikasikan. Proses ini dilakukan untuk membersihkan data yang masih dalam kondisi tidak teratur. Tahapan ini mencakup beberapa langkah, termasuk pembersihan, *case folding*, tokenisasi, penghilangan kata umum (*stopword*), dan *stemming*. Tujuan dari tahapan ini adalah untuk menghasilkan dataset yang bersih, terstruktur, dan siap digunakan untuk analisis lanjutan [11]. Proses *cleansing* melibatkan eliminasi angka, spasi berlebih, tanda baca, dan simbol-simbol yang tidak relevan dalam konteks penelitian. Sementara itu, *case folding* adalah teknik yang digunakan untuk mengonversi semua kata dalam dokumen dataset menjadi huruf kecil, sehingga menghilangkan perbedaan antara huruf besar dan kecil yang dapat mempengaruhi konsistensi analisis [12]. Tokenisasi merupakan proses di mana teks dipecah menjadi potongan-potongan kata yang mandiri dengan menghilangkan tanda baca, sehingga menghasilkan unit-unit kata yang terpisah yang membentuk kalimat atau frase. Proses penghapusan *stopword* melibatkan eliminasi kata-kata yang dianggap tidak memiliki makna signifikan dan tidak diperlukan dalam tahap praproses. Kata-kata yang termasuk dalam kategori ini biasanya meliputi unsur-unsur seperti "yang," "di," "dan," "pada," dan sejenisnya. *Stemming* adalah langkah di mana semua imbuhan dari kata-kata dihapus sehingga kata-kata tersebut berubah menjadi bentuk dasarnya atau akar katanya dalam suatu kalimat.

A. *Cleaning Data*

Proses pembersihan teks bertujuan untuk menghilangkan berbagai elemen seperti tautan URL, nama pengguna atau mention (@username), hashtag (#), retweet, serta emotikon [13].

Tabel 1. *Cleansing Data*

<i>Tweet</i>	<i>Cleaning Data</i>
Gaya Slepert Cak Imin Terhadap Penilaian Anies ke Kemenhan: Untung Saya Enggak Jadi Menhan! #idncitizen #promedia https://t.co/dS6lH29pGo	Gaya Slepert Cak Imin Terhadap Penilaian Anies ke Kemenhan Untung Saya Engak Jadi Menhan idncitizen promedia

B. Case Folding

Proses *case folding* mengonversi seluruh teks dalam dokumen menjadi huruf kecil untuk mengatasi ketidakakonsistenan dalam penggunaan huruf kapital. Langkah ini mempertimbangkan sensitivitas terhadap perbedaan huruf besar dan kecil [14].

Tabel 2. *Case Folding*

<i>Tweet</i>	<i>Case Folding</i>
Gaya Slepet Cak Imin Terhadap Penilaian Anies ke Kemenhan Untung Saya Engak Jadi Menhan idncitizen promedia	gaya slepet cak imin terhadap penilaian anies ke kemenhan untung saya enggak jadi menhan idncitizen promedia

C. Tokenize

Tokenisasi data adalah proses mengonversi teks atau data teks menjadi serangkaian token. Token merupakan unit terkecil dalam pemrosesan teks yang memiliki makna [14]. Proses tokenisasi ini umumnya melibatkan pemisahan teks menjadi kata-kata individu atau bagian-bagian lainnya yang bermakna, seperti tanda baca atau angka.

Tabel 3. *Tokenize Data*

<i>Tweet</i>	<i>Tokenizing</i>
gaya slepet cak imin terhadap penilaian anies ke kemenhan untung saya enggak jadi menhan idncitizen promedia	['gaya', 'slepet', 'cak', 'imin', 'terhadap', 'penilaian', 'anies', 'ke', 'kemenhan', 'untung', 'saya', 'enggak', 'jadi', 'menhan', 'idncitizen', 'promedia']

D. Stopword

Penghapusan *stopword* bertujuan untuk menghilangkan kata-kata umum yang sering muncul dalam teks tetapi tidak memberikan makna penting. Metode ini memanfaatkan algoritma stoplist untuk membuang kata-kata yang dianggap tidak relevan, atau menggunakan wordlist untuk mempertahankan kata-kata yang memiliki makna signifikan [15].

Tabel 4. *Stopword*

<i>Tweet</i>	<i>Stopword</i>
['gaya', 'slepet', 'cak', 'imin', 'terhadap', 'penilaian', 'anies', 'ke', 'kemenhan', 'untung', 'saya', 'enggak', 'jadi', 'menhan', 'idncitizen', 'promedia']	['gaya', 'slepet', 'cak', 'imin', 'penilaian', 'anies', 'kemenhan', 'untung', 'enggak', 'menhan', 'idncitizen', 'promedia']

E. Stemming

Pada tahap *stemming*, tiap kata akan direduksi menjadi bentuk dasarnya dengan menghapus imbuhan [16]. Proses stemming ini memanfaatkan *library sastrawi* untuk mengimplementasikan fungsi stemming.

Tabel 5. *Stemming Data*

<i>Tweet</i>	<i>Stemming</i>
['gaya', 'slepet', 'cak', 'imin', 'penilaian', 'anies', 'kemenhan', 'untung', 'enggak', 'menhan', 'idncitizen', 'promedia']	gaya slepet cak imin hadap nilai anies ke kemenhan untung aku tidak jadi menteri tahan idncitizen promedia

2.3 Labeling Data

Setelah tahap *preprocessing* selesai, langkah selanjutnya adalah melakukan *labeling*. Tujuan dari proses ini adalah memberikan label *tweet* pengguna Twitter dengan indeks yang membedakan nilai sentimennya. *tweet* akan diberi label positif dan negatif. Selain itu, visualisasi menggunakan *wordcloud* dan *barchart* akan digunakan untuk menampilkan hasil *labeling*.

2.4 Spliting Data

Dataset dibagi menjadi dua bagian yaitu data training dan data uji. Dalam penelitian ini data dibagi dengan rasio 80% untuk data *training* dan 20% untuk data *testing*. Dengan total dataset sebanyak 2692, berarti 80% dari total data 2154 data training akan digunakan untuk melatih model, sementara 20% sisanya 538 data testing akan digunakan untuk menguji performa model.

2.5 Klasifikasi Model

Metode *Bidirectional Encoder Representations from Transformers* (BERT) adalah sebuah model bahasa yang dirancang untuk memahami dan memproses teks secara mendalam [17]. BERT menggunakan arsitektur Transformer, yang terdiri dari lapisan-lapisan pengkode (*encoder layers*) yang saling terhubung, untuk mempelajari representasi teks yang kaya dan kontekstual. BERT dilatih pada korpus teks yang sangat besar dan beragam, sehingga memperoleh pemahaman yang lebih baik tentang konteks kata-kata dalam sebuah kalimat [18]. Salah satu keunggulan utama BERT adalah kemampuannya dalam memahami kata-kata yang ambigu atau frasa yang bergantung pada konteks tertentu.

2.6 Evaluasi dan Hasil Akurasi

Tahap selanjutnya melibatkan pengujian model menggunakan teknik *K-fold cross-validation*, di mana kinerja model dievaluasi berdasarkan metrik kesalahan untuk menentukan tingkat akurasinya. Penilaian kinerja model dilakukan dengan menggunakan *confusion matrix* untuk memperoleh ukuran-ukuran seperti akurasi, presisi, *recall*, dan *f1-Score*. *Confusion matrix* digunakan untuk menganalisis performa model pada setiap iterasi. Ada empat elemen utama dalam *confusion matrix*, yaitu *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN) [19]. Elemen-elemen ini digunakan untuk menghitung nilai-nilai dalam *confusion matrix* yang mencerminkan performa model dalam mengklasifikasikan data. Berdasarkan matriks tersebut, dapat dihitung nilai akurasi, presisi, *recall*, dan *f1-score* dengan rumus-rumus berikut:

$$\text{Akurasi} = \frac{\text{True Positif} + \text{True Negatif}}{\text{True Positif} + \text{True Negatif} + \text{False Positif} + \text{False Negatif}} \times 100\% \quad (1)$$

$$\text{Presisi} = \frac{\text{True Positif}}{\text{True Positif} + \text{False Positif}} \quad (2)$$

$$\text{Recall} = \frac{\text{True Positif}}{\text{True Positif} + \text{False Negatif}} \quad (3)$$

$$f1 - \text{score} = \frac{2 \times \text{Presisi} \times \text{Recall}}{\text{Presisi} + \text{Recall}} \quad (4)$$

III. HASIL DAN PEMBAHASAN

A. Pengumpulan Data

Pengumpulan data untuk penelitian ini diambil melalui *crawling* data twitter dengan menggunakan aplikasi *Google Collaboratory* dengan *keyword* yang mengandung kata para calon wakil presiden seperti Imin, Gibran, dan Mahfud. Pengumpulan data ini mulai dari Desember 2023 sampai Februari 2024. Dataset yang disediakan mencakup informasi tentang jumlah sampel untuk masing-masing tokoh yang dianalisis. Untuk Muhaimin Iskandar, terdapat total 1640 sampel, dengan 52 di antaranya memiliki sentimen positif dan 1588 sentimen negatif. Gibran Rakabuming Raka memiliki 959 sampel, dengan 22 sentimen positif dan 937 sentimen negatif. Sedangkan untuk Mahfud MD, dataset memiliki total 90 sampel, dengan mayoritas 86 sentimen positif dan hanya 4 sentimen negatif. Jadi, secara keseluruhan, dataset ini terdiri dari 2689 sampel data yang memberikan informasi tentang sentimen terhadap ketiga tokoh tersebut. Tabel 6 menunjukkan hasil data *crawling* dari twitter.

Tabel 6. Hasil Pengumpulan Data

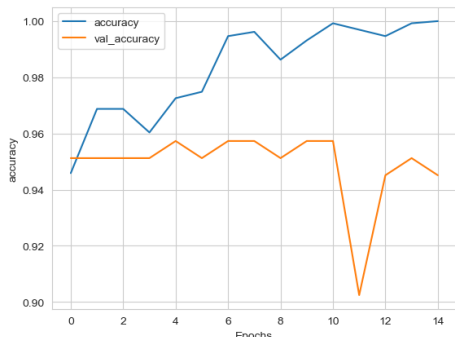
Dataset	Muhaimin Iskandar	Gibran Rakabuming Raka	Mahfud MD
Positif	52	22	86
Negatif	1588	937	4
Total	1640	959	90

B. Tahap Pengujian

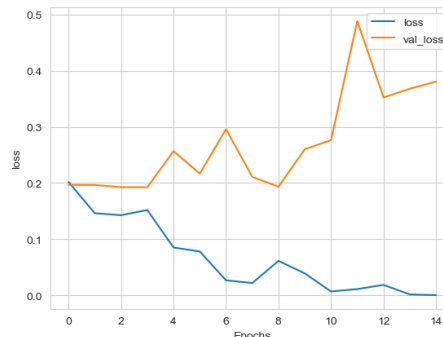
Metode *Bidirectional Encoder Representations from Transformers* (BERT) adalah sebuah model bahasa yang dirancang untuk memahami dan memproses teks secara mendalam [17]. BERT menggunakan arsitektur Transformer, yang terdiri dari lapisan-lapisan pengkode (*encoder layers*) yang saling terhubung, untuk mempelajari representasi teks yang kaya dan kontekstual. BERT dilatih pada korpus teks yang sangat besar dan beragam,

sehingga memperoleh pemahaman yang lebih baik tentang konteks kata-kata dalam sebuah kalimat [18]. Salah satu keunggulan utama BERT adalah kemampuannya dalam memahami kata-kata yang ambigu atau frasa yang bergantung pada konteks tertentu. Data train (*epoch*) yang digunakan sebanyak 15 *epoch*.

1. Hasil Fine Tuning Data Analisis Sentimen Muhaimin Iskandar



Gambar 2. Hasil Grafik Statistik Accuracy



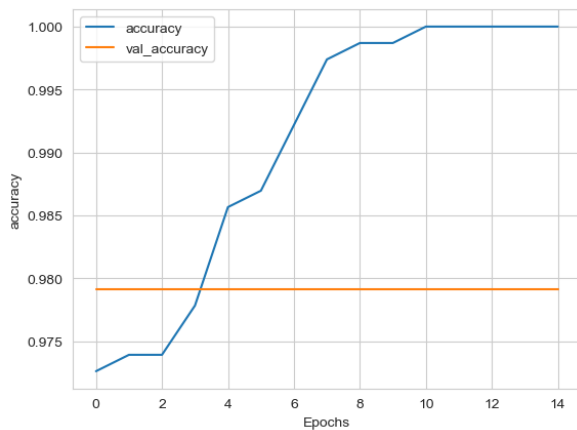
Gambar 3. Hasil Grafik Statistik Loss

Tabel 7. Hasil Performa model BERT pada setiap Epoch Dataset Muhaimin Iskandar

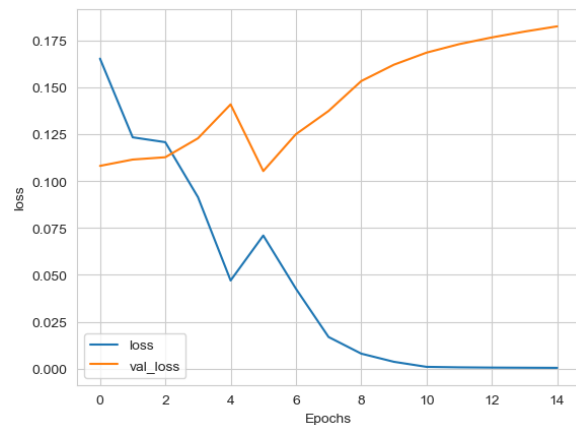
Epoch	Train Accuracy	Train Loss	Val Accuracy	Val Loss
1	94%	20%	95%	19%
2	96%	14%	95%	19%
3	96%	14%	95%	19%
4	96%	15%	95%	19%
5	97%	8%	95%	25%
6	97%	7%	95%	21%
7	99%	2%	95%	29%
8	99%	2%	95%	21%
9	98%	6%	95%	19%
10	99%	3%	95%	26%
11	99%	7%	95%	27%
12	99%	11%	90%	48%
13	99%	18%	94%	35%
14	99%	17%	95%	36%
15	100%	5%	94%	38%

Dari Tabel 7, dapat dilihat performa model BERT pada setiap epoch selama proses pelatihan. Terlihat bahwa ada tren peningkatan akurasi pada tahap awal, dengan akurasi pelatihan naik dari 94% pada epoch pertama menjadi 99% pada epoch ke-9. Namun, pada epoch ke-12, terjadi penurunan signifikan dalam akurasi validasi menjadi 90%, menunjukkan kemungkinan *overfitting* pada model. Meskipun akurasi pelatihan tetap tinggi atau terus meningkat, akurasi validasi tampak stagnan di sekitar 94-95%, dengan penurunan pada epoch terakhir. Analisis loss pelatihan dan validasi juga menunjukkan potensi masalah konvergensi dan *overfitting*. Oleh karena itu, untuk meningkatkan kinerja model, perlu dilakukan evaluasi lebih lanjut terhadap faktor-faktor seperti regulasi, penyesuaian parameter, atau penambahan data pelatihan yang lebih bervariasi. Ini menyoroti pentingnya analisis mendalam terhadap performa model pada setiap epoch untuk mengidentifikasi dan mengatasi masalah yang mungkin timbul selama pelatihan.

2. Hasil Fine Tuning Data Analisis Sentimen Gibran Rakabuming



Gambar 5. Hasil Grafik Statistik Accuracy



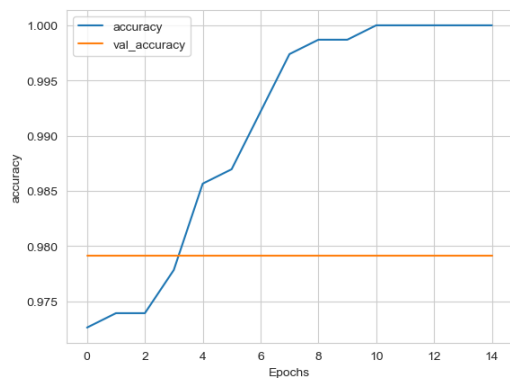
Gambar 6. Hasil Grafik Statistik Loss

Tabel 8. Hasil Performa model BERT pada setiap Epoch Dataset Gibran Rakabuming Raka

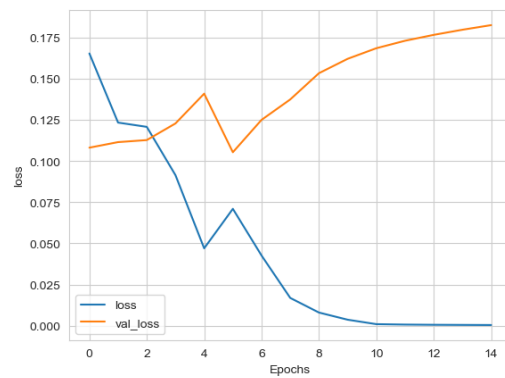
Epoch	Train Accuracy	Train Loss	Val Accuracy	Val Loss
1	97%	16%	97%	10%
2	97%	12%	97%	11%
3	97%	12%	97%	11%
4	97%	9%	97%	12%
5	98%	4%	97%	14%
6	98%	7%	97%	10%
7	99%	4%	97%	12%
8	99%	1%	97%	13%
9	99%	0,8%	97%	15%
10	99%	0,3%	97%	16%
11	100%	0,1%	97%	16%
12	100%	0,1%	97%	17%
13	100%	0,005%	97%	17%
14	100%	0,005%	97%	17%
15	100%	0,004%	97%	18%

Tabel 8 memberikan gambaran tentang performa model BERT pada setiap epoch dalam dataset yang dikhususkan untuk Gibran Rakabuming Raka. Dalam analisis yang lebih mendalam, terlihat bahwa baik train accuracy maupun val accuracy cenderung stabil pada kisaran 97-100% sepanjang epoch, menunjukkan konsistensi dalam kemampuan model untuk mempelajari pola-pola dalam data. Meskipun demikian, terdapat tanda-tanda overfitting yang terlihat dari peningkatan val loss seiring bertambahnya epoch, meskipun val accuracy tetap stabil. Namun, perbedaan antara train loss dan val loss relatif terkendali, menunjukkan bahwa overfitting belum terlalu mengganggu. Pada epoch ke-11, train accuracy mencapai 100% dengan train loss yang sangat rendah, menandakan bahwa model sudah "menghafal" data pelatihan dengan baik, namun val accuracy tetap pada 97% dengan val loss yang meningkat, menunjukkan bahwa model tidak dapat menggeneralisasi dengan baik pada data validasi. Oleh karena itu, pemantauan terus-menerus terhadap tren val loss dan val accuracy menjadi penting dalam pengembangan model, serta mempertimbangkan strategi untuk mengatasi overfitting yang mungkin terjadi.

3. Hasil Fine Tuning Data Analisis Sentimen Mahfud MD



Gambar 8. Hasil Grafik Statistik Accuracy



Gambar 9. Hasil Grafik Statistik Loss

Tabel 9. Hasil Performa model BERT pada setiap Epoch Dataset Mahfud MD

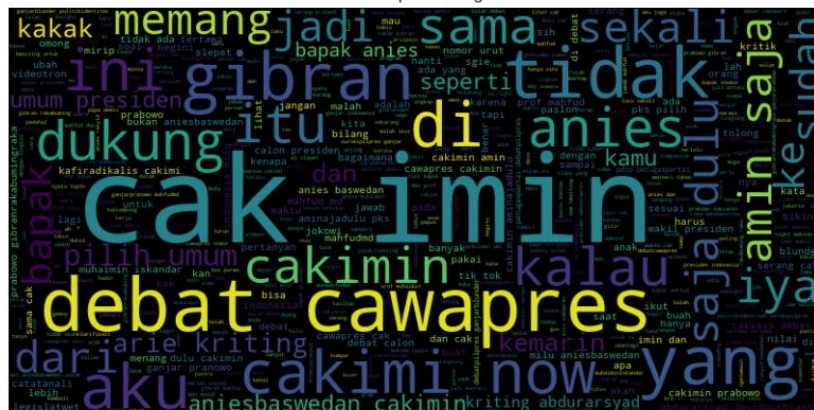
Epoch	Train Accuracy	Train Loss	Val Accuracy	Val Loss
1	97%	16%	97%	10%
2	97%	12%	97%	11%
3	97%	12%	97%	11%
4	97%	9%	97%	12%
5	98%	4%	97%	14%
6	98%	7%	97%	10%
7	99%	4%	97%	12%
8	99%	1%	97%	13%
9	99%	0,8%	97%	15%
10	99%	0,3%	97%	16%
11	100%	0,1%	97%	16%
12	100%	0,1%	97%	17%
13	100%	0,005%	97%	17%
14	100%	0,005%	97%	17%
15	100%	0,004%	97%	18%

Dalam eksperimen yang dilakukan dengan model BERT pada dataset Mahfud MD, hasil performa model pada setiap epoch di tampilkan pada Tabel 9. Pertama, dari sisi akurasi, terlihat bahwa model BERT mencapai tingkat akurasi yang sangat tinggi, mencapai 97% pada tahap awal dan meningkat secara konsisten hingga mencapai 100% pada epoch ke-11 dan seterusnya. Hal ini menunjukkan bahwa model mampu dengan sangat baik dalam mempelajari pola-pola yang ada dalam dataset Mahfud MD dan memberikan prediksi yang akurat terhadap data latih. Kedua, dari segi kerugian (loss), dapat diamati bahwa kerugian pada data latih menurun secara konsisten dari epoch ke epoch, mencapai tingkat yang sangat rendah (0,004% pada epoch ke-15). Ini menunjukkan bahwa model secara efektif meminimalkan kesalahan dalam proses pelatihan dan semakin mendekati optimumnya. Namun, pada sisi validasi, meskipun akurasi tetap konsisten pada 97% sepanjang proses, terjadi peningkatan kerugian yang cukup signifikan dari epoch ke epoch, mencapai 18% pada epoch terakhir. Hal ini menunjukkan bahwa meskipun model mampu memberikan prediksi yang akurat pada data latih, kemampuannya untuk menggeneralisasi ke data yang belum pernah dilihat sebelumnya (data validasi) mungkin menjadi lebih terbatas. Dengan demikian, analisis mendalam terhadap hasil eksperimen ini menyoroti kemampuan tinggi model BERT dalam mempelajari pola dari dataset Mahfud MD dan memberikan prediksi yang akurat pada data latih.

C. Visualisasi Data

Penggunaan *WordCloud* bertujuan untuk memberikan visualisasi yang jelas tentang dataset, memfasilitasi identifikasi kata-kata yang paling sering muncul dalam dokumen tersebut. Proses pembuatan *WordCloud* dilakukan menggunakan *library matplotlib* dalam bahasa pemrograman *Python* [20].

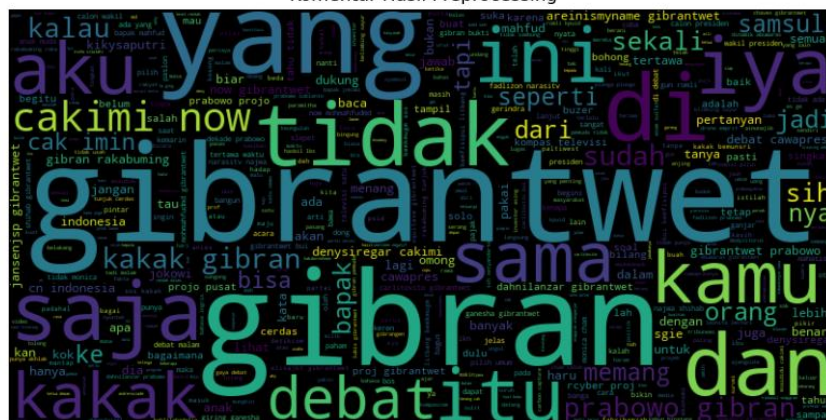
Komentar Hasil Preprocessing Model BERT



Gambar 11. Visualisasi Komentar Muhaimin

Gambar 11 menampilkan hasil visualisasi *WordCloud* dari komentar yang terkait dengan Muhaimin Iskandar menunjukkan bahwa kata "Cak Imin" adalah yang paling dominan, diikuti oleh "cawapres", "debat", "dukung", "now", dan "Anies". Dari hal ini, dapat diasumsikan bahwa topik yang sedang dibahas berkaitan dengan politik atau perdebatan, dengan perhatian khusus pada figur Cak Imin, mungkin terkait dengan peran atau dukungannya dalam sebuah debat atau konteks politik tertentu. Kemungkinan terjadi perdebatan atau dukungan terhadapnya, yang juga mungkin terkait dengan isu yang berkembang "now" atau saat ini, mungkin terkait dengan posisi politiknya atau isu-isu terkini yang sedang diperbincangkan di masyarakat. Adanya kata "Anies" juga menyarankan bahwa ada keterkaitan dengan tokoh publik atau politikus yang bernama Anies dalam konteks yang sama. Analisis lebih lanjut mungkin diperlukan untuk memahami konteks dan signifikansi kata-kata ini dalam konteks tertentu..

Komentar Hasil Preprocessing



Gambar 12. Visualisasi Komentar Gibran

Berdasarkan hasil *wordcloud* pada gambar 12 yang menampilkan kata terbesar "Gibran", dapat diasumsikan bahwa teks tersebut memfokuskan pada tokoh atau topik yang berkaitan dengan Gibran. Kemungkinan besar, teks tersebut membahas tentang seseorang yang bernama Gibran, baik sebagai tokoh publik, figur terkenal, atau subjek yang relevan dengan konteks tertentu. Selain itu, keberadaan kata-kata seperti "tweet" dan "debat" mengisyaratkan adanya keterkaitan dengan media sosial dan diskusi yang berlangsung, mungkin dalam konteks perdebatan atau pembahasan yang intens. Ditambah dengan kata "kakak", ada kemungkinan kuat bahwa teks tersebut juga menyoroti hubungan keluarga atau dinamika interpersonal yang melibatkan figur kakak sebagai elemen penting dalam naratifnya.

Komentar Hasil Preprocessing



Gambar 13. Visualisasi Komentar Mahfud

Analisis lebih dalam terhadap hasil *Wordcloud* pada gambar 13 menyoroti empat kata kunci utama: "Mahfud", "Ganjar", "Bapak", dan "Prof". Kemunculan kata "Mahfud" yang dominan menandakan peran penting tokoh atau entitas dengan nama tersebut dalam konteks pembahasan. Selanjutnya, kehadiran kata "Ganjar" menunjukkan relevansi yang signifikan dari tokoh atau subjek yang terkait, seperti Ganjar Pranowo atau tokoh lainnya. Kata "Bapak" mungkin merujuk kepada figur otoritatif atau tokoh penting dalam konteks yang sedang dibahas. Sementara itu, kemunculan kata "Prof" mengisyaratkan adanya pembahasan mengenai individu dengan gelar profesor atau kaitannya dengan bidang akademik. Analisis lebih lanjut terhadap konteks teks dapat memberikan pemahaman yang lebih mendalam tentang peran dan signifikansi masing-masing kata-kata tersebut dalam wacana yang diproses.

D. Model Klasifikasi

Pada tahapan ini, hasil evaluasi model menunjukkan bahwa ketiga calon presiden yang dievaluasi dalam Tabel 10, Muhaimin Iskandar, Gibran Rakabuming Raka, dan Mahfud MD. Dengan hasil evaluasi yang menunjukkan kinerja yang sangat baik dari model klasifikasi terhadap ketiga calon presiden, langkah selanjutnya adalah melihat hasil dan pembahasan model secara lebih mendalam. Evaluasi memberikan gambaran tentang kemampuan model dalam mengidentifikasi kandidat presiden dengan akurasi yang tinggi, yang dapat menjadi landasan penting untuk pemahaman lebih lanjut tentang dinamika politik dan preferensi pemilih.

Tabel 10. Hasil akurasi menggunakan algoritma BERT

Calon Presiden	Accuracy	Precision	Recall	F1-Score
Muhaimin Iskandar	98,17%	100%	100%	100%
Gibran Rakabuming Raka	100 %	100 %	100 %	100 %
Mahfud MD	100 %	100 %	100 %	100 %

Hasil menunjukkan bahwa Muhaimin Iskandar memiliki tingkat akurasi, presisi, *recall*, dan *F1-Score* yang tinggi, mencapai 98,17%, serta 100% untuk metrik lainnya. Hal ini menunjukkan bahwa model dengan sangat baik memprediksi Muhaimin Iskandar sebagai calon presiden, tanpa kesalahan prediksi positif maupun negatif. Meskipun begitu, analisis lebih lanjut dapat mempertimbangkan apakah faktor-faktor lain seperti popularitas dan dukungan masyarakat telah tercermin dengan baik dalam prediksi ini. Selanjutnya Gibran Rakabuming Raka juga menunjukkan hasil evaluasi yang sangat baik dengan nilai 100% untuk semua metrik, menandakan bahwa model memprediksi dengan sempurna kecocokan Gibran sebagai calon presiden. Namun, dalam analisis lebih lanjut, mungkin perlu diperhatikan apakah faktor-faktor seperti pengalaman politik atau rekam jejak publik telah mempengaruhi prediksi tersebut. Calon wakil preasiden berikutnya, Mahfud MD juga menunjukkan tingkat akurasi, presisi, *recall*, dan *F1-Score* yang sempurna, mencapai 100% untuk semua metrik. Ini menandakan bahwa model dengan sangat baik mengidentifikasi Mahfud sebagai calon presiden tanpa kesalahan. Namun, analisis mendalam mungkin mempertanyakan apakah model telah memperhitungkan faktor-faktor seperti konsistensi politik atau kebijakan yang diusulkan oleh Mahfud.

Penggunaan model BERT dalam analisis teks terkait profil calon presiden Indonesia menunjukkan prediksi yang lebih akurat dan klasifikasi yang tepat. Kemampuan BERT dalam memperhitungkan konteks dan hubungan kata-kata memengaruhi tingkat akurasi identifikasi kandidat presiden. Implikasi praktisnya adalah bahwa pemangku

kepentingan politik dan masyarakat sipil dapat memanfaatkan analisis sentimen ini untuk memahami preferensi pemilih dan dinamika politik dalam pemilihan presiden, serta merancang strategi komunikasi yang lebih efektif.

IV. KESIMPULAN

Berdasarkan hasil eksperimen dapat disimpulkan bahwa penggunaan metode *Bidirectional Encoder Representations from Transformers* (BERT) dalam melakukan analisis sentimen terhadap kandidat presiden menunjukkan hasil yang sangat baik. Dari tahap pengujian yang dilakukan, terlihat bahwa model BERT mampu mencapai tingkat akurasi yang tinggi pada data latih, dengan tren peningkatan yang konsisten pada awal pelatihan. Namun, terdapat indikasi potensi *overfitting* pada beberapa tahap pelatihan, terutama pada model yang dikhususkan untuk dataset Muhaimin Iskandar. Selain itu, evaluasi terhadap performa model pada setiap epoch menyoroti pentingnya pemantauan terhadap tren akurasi dan loss baik pada data latih maupun data validasi untuk mengidentifikasi dan mengatasi masalah yang mungkin timbul selama pelatihan. Meskipun demikian, keseluruhan performa model BERT menunjukkan keunggulan dalam mempelajari pola-pola dalam teks dan memberikan prediksi yang akurat terhadap data yang diberikan. Selain itu metode BERT menunjukkan kinerja yang sangat baik dalam memprediksi kandidat presiden dengan tingkat akurasi, presisi, *recall*, dan *F1-Score* yang tinggi. Masing-masing kandidat, yaitu Muhaimin Iskandar, Gibran Rakabuming Raka, dan Mahfud MD, memiliki hasil akurasi 98,17%, 100%, dan 100% berturut-turut. Presisi, *recall*, dan *F1-Score* juga mencapai 100% untuk ketiganya. Meskipun demikian, penting untuk melakukan analisis lebih lanjut terhadap faktor-faktor eksternal yang mungkin memengaruhi hasil prediksi model. Dalam keseluruhan, penggunaan metode BERT dalam analisis sentimen terhadap kandidat presiden menjanjikan, namun membutuhkan pertimbangan terhadap faktor-faktor eksternal.

DAFTAR PUSTAKA

- [1] B. D. Adine Nayla, Casi Setianingsih, "Deteksi Hate Speech Pada Twitter," *E-Proceeding Eng.*, Vol. 10, No. 1, Pp. 256–262, 2023.
- [2] A. N. Ulfah, M. K. Anam, J. T. Informatika, H. Speech, And S. V. Machine, "Analisis Sentimen Hate Speech Pada Portal Berita Online Menggunakan Support Vector Machine (Svm)," Vol. 7, No. 1, 2020.
- [3] E. E. Amelia And I. Yustiana, "Analisis Sentimen Pada Ulasan Produk Uniqlo Dengan Algoritma Naive Bayes," Vol. 8, Pp. 141–148, 2024.
- [4] A. Sadat, H. Lawelai, A. Suherman, U. M. Buton, And U. M. Buton, "Analisis Sentimen Media Sosial: Hate Speech Kepada Pemerintah Di Twitter," Vol. 10, Pp. 69–76, 2021.
- [5] R. M. R. W. P. K. Atmaja And W. Yustanti, "Analisis Sentimen Customer Review Aplikasi Ruangguru Dengan Metode Bert (Bidirectional Encoderrepresentations From Transformers)," *Jeisbi (Journal Emerg. Inf. Syst. Bus. Intell.*, Vol. 02, No. 3, Pp. 55–62, 2021.
- [6] Vidya Chandradev, I Made Agus Dwi Suarjaya, And I Putu Agung Bayupati, "Analisis Sentimen Review Hotel Menggunakan Metode Deep Learning Bert," *J. Buana Inform.*, Vol. 14, No. 02, Pp. 107–116, 2023, Doi: 10.24002/Jbi.V14i02.7244.
- [7] R. Fatmasari, R. K. Septiani, T. H. Pinem, D. Fabiyanto, And W. Gata, "Implementasi Algoritma Bert Pada Komentar Layanan Akademik Dan Non Akademik Universitas Terbuka Di Media Sosial," *Sains, Apl. Komputasi Dan Teknol. Inf.*, Vol. 5, No. 2, P. 96, 2024, Doi: 10.30872/Jsakti.V5i2.13915.
- [8] P. K. Sari And R. R. Suryono, "Komparasi Algoritma Support Vector Machine Dan Random Forest Untuk Analisis Sentimen Metaverse," *J. Mnemon.*, Vol. 7, No. 1, Pp. 31–39, 2024, Doi: 10.36040/Mnemonic.V7i1.8977.
- [9] D. Ananda And R. R. Suryono, "Analisis Sentimen Publik Terhadap Pengungsi Rohingya Di Indonesia Dengan Metode Support Vector Machine Dan Naive Bayes," Vol. 8, No. April, Pp. 748–757, 2024, Doi: 10.30865/Mib.V8i2.7517.
- [10] H. Saputra, "Analisis Sentimen Pada Vaksin Booster Menggunakan Algoritma Support Vector Machine Multiclass Di Twitter," Vol. 3, No. 10, Pp. 1–26, 2023.
- [11] B. Rais, S. Suhardi, And R. P. I. P. Sari, "Critical Discourse Analysis Of Additional Information In Skincare Products," *Eralingua J. Pendidik. Bhs. Asing Dan Sastra*, Vol. 4, No. 1, P. 89, 2020, Doi: 10.26858/Eralingua.V4i1.12519.
- [12] C. F. Hasri And D. Alita, "Penerapan Metode Naive Bayes Classifier Dan Support Vector Machine Pada Analisis Sentimen Terhadap Dampak Virus Corona Di Twitter," *J. Inform. Dan Rekayasa Perangkat Lunak*, Vol. 3, No. 2, Pp. 145–160, 2022.
- [13] V. D. Yunanda And N. Hendrastuty, "Perbandingan Kernel Polynomial Dan Rbf Pada Algoritma Svm Untuk Analisis Sentimen Skincare Di Indonesia," Vol. 8, No. April, Pp. 726–735, 2024, Doi: 10.30865/Mib.V8i2.7425.
- [14] T. T. Widowati And M. Sadikin, "Analisis Sentimen Twitter Terhadap Tokoh Publik Dengan Algoritma Naive Bayes Dan Support Vector Machine," *Simetris J. Tek. Mesin, Elektro Dan Ilmu Komput.*, Vol. 11, No. 2, Pp. 626–636, 2021, Doi: 10.24176/Simet.V11i2.4568.
- [15] A. I. Purnamasari And I. Ali, "Analisis Sentimen Komentar Berita Detik . Com Menggunakan Algoritma Suport Vektor Machine (Svm)," Vol. 8, No. 3, Pp. 3175–3181, 2024.
- [16] G. Sahanji, Julian, And H. Syah, "Wfraud Alert Sebagai Prediksi Pesan Penipuan Whatsapp Menggunakan Naive

- Bayes,” *Tekno Kompak*, Vol. 18, No. 1, Pp. 113–125, 2024.
- [17] J. U. S. Lazuardi And A. Juarna, “Analisis Sentimen Ulasan Pengguna Aplikasi Joox Pada Android Menggunakan Metode Bidirectional Encoder Representation From Transformer (Bert),” *J. Ilm. Inform. Komput.*, Vol. 28, No. 3, Pp. 251–260, 2023, Doi: 10.35760/Ik.2023.V28i3.10090.
- [18] N. P. V. D. Saraswati, N. Yudistira, And P. P. Adikara, “Analisis Sentimen Terhadap Perundungan Siber Pada Twittermenggunakan Algoritma Bidirectional Encoder Representations Fromtransformer (Bert),” *J. Pengemb. Teknol. Inf. Dan Ilmu Komput.*, Vol. 7, No. 2, Pp. 909–916, 2023, [Online]. Available: [Http://J-Ptiik.Ub.Ac.Id](http://J-Ptiik.Ub.Ac.Id)
- [19] A. F. Setyaningsih, D. Septiyani, And S. R. Widiyari, “Implementasi Algoritma Naïve Bayes Untuk Analisis Sentimen Masyarakat Pada Twitter Mengenai Kepopuleran Produk Skincare Di Indonesia,” *J. Teknol. Inform. Dan Komput.*, Vol. 9, No. 1, Pp. 224–235, 2023, Doi: 10.37012/Jtik.V9i1.1409.
- [20] F. Sodik And I. Kharisudin, “Analisis Sentimen Dengan Svm , Naive Bayes Dan Knn Untuk Studi Tanggapan Masyarakat Indonesia Terhadap Pandemi Covid-19 Pada Media Sosial Twitter,” *Prisma*, Vol. 4, Pp. 628–634, 2021.