

Analisis Sentimen Ulasan Pengguna Aplikasi Info BMKG pada Google Play Store menggunakan *Model Transformer BERT dan RoBERTa*

Charlo Brando¹, Sajarwo Anggai², Tukiyat³

¹²³ Program Studi Teknik Informatika S-2, Program Pascasarjana, Universitas Pamulang
charlobando@gmail.com, sajarwo@gmail.com, dosen02711@unpam.ac.id

Diterima : 30 Agustus 2025
Disetujui : 20 September 2025

Abstract—. Aplikasi Info BMKG memiliki peran penting dalam menyampaikan informasi cuaca, iklim, gempa bumi, dan peringatan dini bencana kepada masyarakat. Seiring meningkatnya penggunaan perangkat mobile di Indonesia, analisis sentimen menjadi relevan untuk mengevaluasi kepuasan pengguna serta mengidentifikasi aspek yang perlu diperbaiki. Penelitian ini bertujuan untuk menganalisis sentimen ulasan pengguna terhadap aplikasi Info BMKG di Google Play Store dengan memanfaatkan model *transformer* BERT dan RoBERTa. Dataset 10.791 ulasan pengguna yang diklasifikasikan ke dalam tiga kategori sentimen meliputi positif, netral, dan negatif. Tahapan penelitian mencakup eksplorasi data awal, prapemrosesan data, serta evaluasi model. Hasil evaluasi menunjukkan bahwa model BERT memberikan performa terbaik dengan akurasi sebesar 93,14%, disusul oleh RoBERTa dengan akurasi 91,06% pada skenario pembagian data 80:10:10. Selain itu, model BERT juga unggul dalam metrik lain, yakni presisi 93,45%, recall 92,90%, dan F1-score 93,17%, dibandingkan RoBERTa dengan presisi 91,12%, recall 90,72%, dan F1-score 90,91%. Analisis lanjutan menunjukkan bahwa meskipun aplikasi mendapatkan apresiasi, pengguna juga menyoroti isu keterlambatan notifikasi gempa dan ketidakakuratan informasi. Temuan ini diharapkan dapat menjadi dasar pengembangan lebih lanjut dalam meningkatkan kualitas layanan dan efektivitas penyampaian informasi oleh BMKG.

Keywords — *Analisa Sentiment, BERT, RoBERTa, Info BMKG, Google Play Store*

I. PENDAHULUAN

Secara kumulatif, selama bulan April 2025 tercatat telah terjadi 152 bencana alam yang melanda wilayah Indonesia. Bencana tersebut didominasi oleh bencana hidrometeorologi dengan persentase mencapai 99,34%. Banjir menjadi bencana yang paling sering terjadi dengan 99 kejadian atau 65,13% dari total kejadian [1]. Kondisi intensitas hujan yang masih sering terjadi di beberapa wilayah Indonesia berimplikasi pada meningkatnya potensi akan terjadinya bencana hidrometeorologi. Kondisi tersebut menjadikan peran Badan Meteorologi, Klimatologi, dan Geofisika (BMKG) sangat krusial.

BMKG sebagai Lembaga Pemerintah Non Kementerian, memiliki tugas dan wewenang

dalam pengelolaan, penyebaran informasi, dan peringatan dini di bidang meteorologi, klimatologi, geofisika, dan kualitas udara [2] Rencana Strategis BMKG Tahun 2020-2024 menerangkan bahwa peningkatan pemahaman masyarakat terhadap informasi dan peringatan dini yang disampaikan oleh BMKG merupakan salah satu indikator kinerja utama yang dapat diwujudkan dengan menghadirkan penyampaian informasi secara cepat, tepat, dan akurat melalui *platform* berbasis *mobile* [3]. Perkembangan teknologi digital, khususnya aplikasi mobile, menjadi semakin relevan seiring meningkatnya penggunaan perangkat digital di masyarakat. APJII mencatat bahwa pada tahun 2024 sekitar 89,44% masyarakat digital Indonesia mengakses

informasi melalui perangkat *mobile* seperti gawai pintar dan tablet [4].

Perkembangan teknologi dan peran krusial BMKG dalam memberikan informasi agar dapat diakses masyarakat luas mendorong inovasi digital melalui aplikasi info BMKG. Aplikasi ini dikembangkan sebagai *platform* resmi BMKG yang menyajikan informasi cuaca, iklim, gempa bumi, kualitas udara, hingga peringatan dini bencana secara terintegrasi [5].

Pada situs layanan Google Play Store per Juni 2025, aplikasi Info BMKG telah digunakan oleh lebih dari 5 juta pengguna dengan rating 4.7 dari 81.6 ribu ulasan. Skor rating pada aplikasi akan berpengaruh terhadap perspektif pengguna lain dalam membuat keputusan untuk mengunduh dan memasang aplikasi pada perangkat mereka [6]. Rating pada setiap ulasan pengguna dapat diimplikasikan sebagai sentimen pengguna terhadap aplikasi [7]. Ulasan dan rating pengguna terhadap info BMKG menunjukkan masih terdapat celah kekurangan yang perlu diperhatikan dalam perbaikan dan pembaruan, dengan berdasar pada informasi kepuasan pengguna yang diperoleh dari ulasan yang diberikan terhadap aplikasi [8].

Masifnya ulasan yang disampaikan pengguna aplikasi Info BMKG menjadi tantangan dalam menganalisis informasi umpan balik yang diberikan pengguna [9]. Diperlukan analisis untuk menggali informasi yang termuat dalam ulasan pengguna aplikasi, pendekatan yang dapat digunakan dalam memproses ulasan berupa teks adalah dengan menggunakan model transformer [10].

Berdasarkan permasalahan tersebut penelitian ini akan dilakukan untuk menganalisis sentimen ulasan dari aplikasi info BMKG. Algoritma yang digunakan adalah BERT dan RoBERTa yang memiliki keunggulan dalam memahami konteks dalam teks dan mampu menangkap makna yang lebih dalam dari sebuah kalimat teks, sehingga sangat baik digunakan dalam tugas seperti analisis sentimen [11]. Penggunaan model BERT dan RoBERTa pada penelitian ini didasari oleh motivasi untuk melakukan evaluasi komparatif secara mendalam. BERT telah diakui sebagai model dasar yang revolusioner dalam pemrosesan bahasa alami (NLP) karena kemampuannya

memahami konteks dua arah (kiri ke kanan dan kanan ke kiri). Sedangkan RoBERTa dikembangkan sebagai optimisasi langsung dari arsitektur BERT, tujuannya untuk meningkatkan kinerjanya melalui modifikasi pada metode pra-pelatihan. Dengan demikian, membandingkan keduanya secara langsung pada dataset ulasan aplikasi Info BMKG menjadi relevan untuk menentukan model mana yang memberikan performa paling unggul untuk tugas analisis sentimen dalam konteks layanan publik di Indonesia.

Perbedaan fundamental antara BERT dan RoBERTa terletak pada pra-pelatihannya, terdapat tiga modifikasi utama yang diimplementasikan pada RoBERTa:

1. Penghapusan Tugas *Next Sentence Prediction* (NSP), pengembang RoBERTa menemukan bahwa tugas memprediksi kalimat berikutnya (NSP) yang ada pada BERT kurang memberikan dampak signifikan dan menghapusnya, sehingga model dapat fokus sepenuhnya pada tugas *Masked Language Modeling* (MLM).
2. Penggunaan *Dynamic Masking*, tidak seperti BERT yang menggunakan *static masking* (*token* ditutupi sekali selama pra-pemrosesan data), RoBERTa menerapkan *dynamic masking* di mana pola penutupan *token* diubah pada setiap *epoch* pelatihan. Hal ini memungkinkan model belajar dari representasi data yang lebih beragam.
3. Pelatihan dengan data dan *batch size* yang lebih besar, RoBERTa dilatih pada korpus data yang jauh lebih besar dan dengan ukuran *batch* yang lebih masif, yang secara teoretis memungkinkannya untuk mempelajari pola bahasa yang lebih besar dan kompleks.

Model RoBERTa memiliki keunggulan dibandingkan dengan model BERT terutama dalam kinerja melakukan tugas domain spesifik terutama dalam bahasa Indonesia [11]. Penelitian ini menjadi penting dalam melakukan uji empiris optimasi yang ditawarkan model BERT dan RoBERTa dapat menghasilkan model khususnya dalam melakukan analisis sentimen ulasan pengguna aplikasi info BMKG yang memiliki dampak langsung terhadap keselamatan publik.

Hasil penelitian memberikan gambaran model terbaik yang dapat digunakan untuk analisis ulasan pengguna dan dapat digunakan sebagai pengembangan aplikasi info BMKG serta berkontribusi pada penelitian yang berkaitan dengan penerapan model transformer untuk analisis sentimen berbahasa Indonesia. Objek yang akan diteliti adalah data ulasan pengguna info BMKG yang diambil dari *review* di *Google Play Store*. Hasil penelitian ini diharapkan dapat memberikan masukan dan wawasan terkait persepsi publik terhadap aplikasi info BMKG.

II. TINJAUAN LITERATUR

Penelitian mengenai analisis sentimen menggunakan metode komputasi telah berkembang pesat. Beberapa penelitian yang relevan sebelumnya menjadi landasan bagi studi ini, penelitian tentang Analisis sentimen yang oleh Putri dan Ardiansyah tentang “Analisis Sentimen Terhadap Kemajuan Kecerdasan Buatan di Indonesia Menggunakan BERT dan RoBERTa” [11] telah melakukan perbandingan langsung antara model BERT dan RoBERTa, di mana mereka menemukan bahwa BERT sedikit lebih unggul dalam menganalisis sentimen publik terhadap kemajuan kecerdasan buatan, namun objek penelitian yang diteliti berbeda dan penelitian tersebut berfokus pada sentimen terhadap kemajuan AI secara umum.

Penellitian lain yang dilakukan oleh Putu dkk, tentang “Analisis Sentimen terhadap Perundungan Siber pada Twitter menggunakan Algoritma Bidirectional Encoder Representations from Transformer (BERT)” [12] juga menerapkan model BERT dan membuktikan keefektifannya dalam mengklasifikasikan sentimen terkait perundungan siber di media sosial Twitter, namun penelitian tersebut hanya menggunakan satu metode saja yaitu BERT.

Dalam penelitian lain oleh Rihan Maulana dkk, tentang “Analisis Sentimen Ulasan Aplikasi Mypertamina Pada Google Play Store Menggunakan Algoritma Nbc” [13]. Analisis Sentimen yang digunakan adalah algoritma klasik Naive Bayes Classifier (NBC, Metode yang digunakan berbeda secara fundamental. Penelitian

tersebut menggunakan metode machine learning klasik (NBC), sedangkan penelitian ini menggunakan metode deep learning yaitu BERT dan RoBERTa yang lebih unggul dalam memahami konteks.

Walaupun penelitian-penelitian tersebut memberikan kontribusi penting, terdapat celah yang belum terisi. Belum ada penelitian yang secara spesifik melakukan perbandingan eksperimental antara model BERT dan RoBERTa pada domain layanan informasi publik yang disediakan oleh Badan Meteorologi, Klimatologi, dan Geofisika (BMKG).

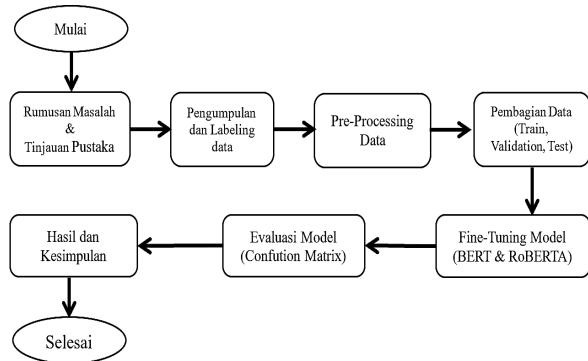
Dengan mempertimbangkan pentingnya layanan informasi publik yang disediakan oleh BMKG dalam konteks kebencanaan dan cuaca ekstrem, penelitian ini hadir untuk menjawab kebutuhan akan sistem analisis sentimen yang akurat terhadap opini masyarakat. Penelitian ini dilakukan untuk mengevaluasi kinerja dua model Transformer, yaitu BERT dan RoBERTa, dalam mengklasifikasikan sentimen ulasan pengguna terhadap aplikasi Info BMKG di Google Play Store. Ulasan tersebut mencerminkan persepsi publik terhadap kualitas layanan informasi yang diberikan, sehingga hasil analisis dapat menjadi dasar bagi pengambilan keputusan strategis. Peneliti mengidentifikasi bahwa belum terdapat studi komprehensif yang secara spesifik menguji efektivitas model NLP dalam domain layanan publik yang bersifat kritikal seperti ini. Oleh karena itu, penelitian ini tidak hanya berkontribusi secara metodologis melalui eksperimen komparatif antar model, tetapi juga secara kontekstual karena berfokus pada ranah yang memiliki dampak langsung terhadap keselamatan dan kesiapsiagaan masyarakat. Dengan pendekatan berbasis data yang dilakukan menggunakan Python, penelitian ini diharapkan mampu memberikan rekomendasi preskriptif untuk peningkatan kualitas interaksi digital antara masyarakat dan institusi layanan publik.

III. METODOLOGI PENELITIAN

A. Tahapan Penelitian

Tahapan penelitian dirancang untuk memastikan analisis sentimen ulasan pengguna

aplikasi Info BMKG di Google Play Store dapat dilakukan secara terstruktur dan sistematis.



Gambar 1. Tahapan Penelitian

B. Dataset

Metode pemilihan data dalam penelitian ini dilakukan dengan pendekatan purposive sampling, dimana pemilihan data berdasarkan kriteria tertentu yang relevan sesuai dengan tujuan penelitian. Dalam hal ini, data yang dipilih untuk penelitian adalah ulasan pengguna aplikasi Info BMKG di Google Play Store. Kriteria pemilihan sampel yang digunakan meliputi: (1) Data diambil dari *google play store*, menggunakan teknik *web scraping*; (2) Periode data ulasan sampai dengan 31 Januari 2025; (3) Relevansi teks ulasan, hanya mengandung teks deskriptif minimal terdiri dari tiga kata; (4) Bahasa yang digunakan dalam ulasan adalah bahasa Indonesia. Dataset mencakup 10.791 baris data ulasan dengan fitur *content* yang merupakan ulasan pengguna. Fitur *score* digunakan untuk proses klasifikasi ulasan pengguna ke dalam tiga kelas yaitu positif, netral, dan negatif.

C. Pra-pemrosesan Data

Pada tahap pra-pemrosesan data dilakukan untuk memastikan bahwa data dalam kondisi yang siap digunakan dalam proses pelatihan model. Beberapa langkah yang dilakukan dalam pra-pemrosesan data meliputi.

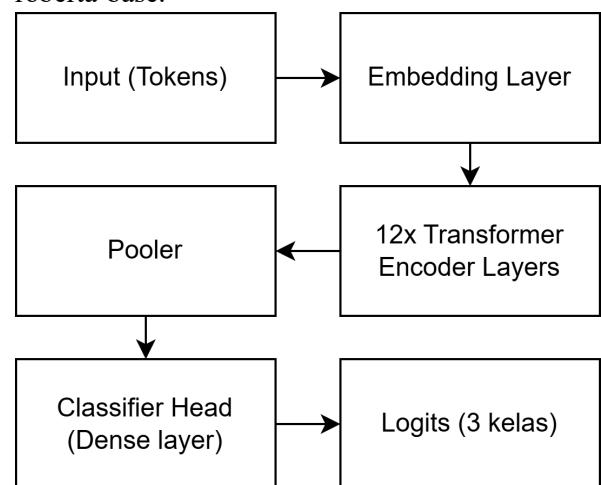
1. Normalisasi yakni proses dalam mengoreksi ejaan Bahasa Indonesia dengan memanfaatkan fitur Google AI di Google Translate. Tahap normalisasi melakukan penerjemahan dua kali pada teks ulasan, yaitu dari Bahasa Indonesia ke Bahasa Inggris, lalu diterjemahkan lagi dari Bahasa Inggris ke Bahasa Indonesia.

2. *Case folding* yakni proses menjadi semua karakter teks ulasan menjadi huruf kecil.
3. *Filtering* merupakan proses menghilangkan karakter teks yang mengandung angka, tanda baca, URL, kode ASCII, dan emoji.
4. *Stemming* merupakan proses mencari kata dasar dengan menghapus awalan dan akhiran dari suatu kata tertentu.
5. Penghapusan *Stopwords* merupakan proses penghapusan kata umum yang tidak memiliki makna dalam teks ulasan.
6. *Re-sampling imbalanced data* teknik yang digunakan untuk menyetarakan jumlah kelas minoritas mengikuti jumlah kelas mayoritas.

D. Model Transformer

Model BERT yang digunakan dalam penelitian ini adalah BertForSequenceClassification dari ekstensi indobenchmark/indobert-base-p1. Model BertForSequenceClassification diadaptasi untuk tugas klasifikasi dengan menambahkan classification head di atas representasi [CLS]. Pretrained model tersebut telah dilatih sebelumnya menggunakan masked language modeling (MLM) objective dan next sentence prediction (NSP) objective.

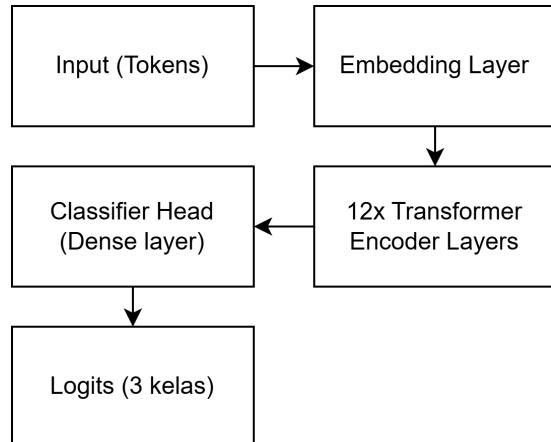
Model RoBERTa yang digunakan dalam penelitian ini adalah RobertaForSequenceClassification dari ekstensi model RoBERTa flax-community/indonesian-roberta-base.



Gambar 2. Arsitektur model BERT

RoBERTa Base Sentiment Classifier adalah model klasifikasi teks sentimen yang didasarkan pada model RoBERTa. Model ini

awalnya adalah model RoBERTa Base Indonesia yang telah dilatih sebelumnya, yang kemudian disempurnakan pada kumpulan data SmSA indonlu yang terdiri dari komentar dan ulasan berbahasa Indonesia.



Gambar 3. Arsitektur model RoBERTa

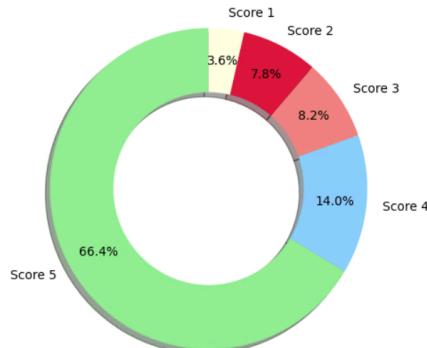
E. Evaluasi

Evaluasi model dilakukan dengan mengukur sejauh mana model mampu mengklasifikasikan data secara benar. Metrik evaluasi yang digunakan dalam penelitian ini mencakup akurasi, presisi, recall, dan F1-score.

IV. HASIL DAN PEMBAHASAN

A. Eksplorasi Data Awal

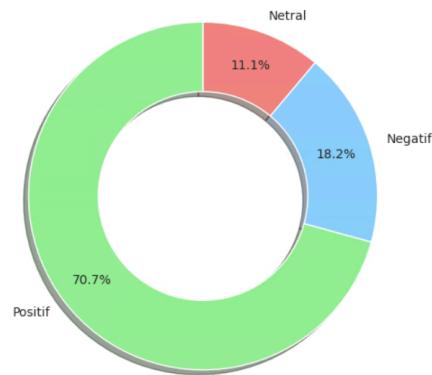
Eksplorasi data awal dilakukan untuk memahami karakteristik dataset yang digunakan penelitian. Data dikumpulkan dari ulasan pengguna aplikasi info BMKG di Google Play Store dengan rentang periode sampai dengan 31 Januari 2025.



Gambar 4. Skor rating ulasan pengguna aplikasi Info BMKG

Gambar 4 menunjukkan bahwa 66,4% ulasan pengguna memberikan Skor rating 5. Hal ini

menunjukkan bahwa sebagian besar pengguna menyatakan tingkat kepuasan yang tinggi dengan layanan dan fungsi yang ditawarkan oleh aplikasi, terutama mengenai aksesibilitas informasi meteorologi, ketepatan prakiraan, dan stabilitas keseluruhan aplikasi selama penggunaan rutin. Sebanyak 14% ulasan pengguna memberikan Skor rating 4, yang menandakan tingkat kepuasan yang cukup besar, meskipun dengan beberapa kekurangan yang telah diidentifikasi pengguna. 8,2% dari total ulasan pengguna memberikan skor rating 3, menunjukkan adanya segmen pengguna yang menganggap aplikasi cukup memuaskan. Sementara sebanyak 7,8% dari total ulasan pengguna memberikan skor rating 2 dan 3,6% memberikan skor rating 1 yang menunjukkan tingkat penilaian negatif, karena dapat menunjukkan masalah spesifik yang dihadapi oleh pengguna tertentu, termasuk keterbatasan teknis, bug perangkat lunak, atau pengalaman pengguna yang kurang optimal.



Gambar 5. Skor rating ulasan pengguna aplikasi Info BMKG

Dataset diberi label meliputi positif, negatif, dan netral berdasarkan setiap rating ulasan pengguna info BMKG [7]. Data skor rating pada gambar 4 dilakukan pelabelan sehingga proporsi label positif, negatif, dan netral ditunjukkan pada gambar 5.

B. Analisis Hasil Pra-pemrosesan Data

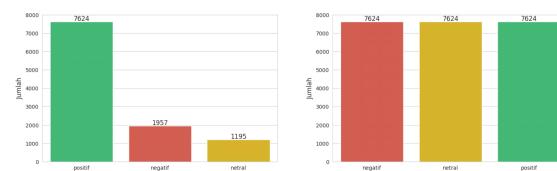
Pra-pemrosesan data merupakan tahap penting dalam pengolahan data teks, khususnya dalam konteks ulasan pengguna aplikasi. Tabel 1 menunjukkan hasil prapemrosesan terhadap data ulasan pengguna, dengan tujuan utama menyederhanakan, menormalkan, dan

menghilangkan elemen-elemen yang tidak relevan dari ulasan pengguna.

Tabel 1. Sampel hasil pemrosesan data

Masukan	Keluaran
Prakiraan cuacanya untuk daerah saya di kalimantan tengah setiap hari 90% selalu meleset. Prakiraan hujan padahal cuaca panas sampai 33 derajat	kira cuaca daerah kalimantan tengah tiap hari selalu meleset ramal hujan meski cuaca panas hingga derajat
Kok lelet sih infonya, misal pukul 19.00 terjadi gempa, pemberitahuan nya kok molor lama banget 😞 sampai berjam2	info lambat misal gempa bumi jadi pemberitahuan benar benar lama lama jam jam
Wahhh, keren banget nih apk bisa melihat cuaca dan lainnya, tapi tolong di update lagi untuk menambahkan fitur baru. Saya kasih bintang 5 langsung	sangat keren apk lihat cuaca naik tambah fitur baru suka bintang cara langsung

Hasil prapemrosesan data ulasan pengguna berhasil mentransformasikan teks mentah menjadi dataset terstruktur melalui normalisasi, case folding, filtering, stemming, re-sampling, dan label encoding. Proses ini menghilangkan noise, menstandarisasi bentuk kata, mengurangi redundansi, serta memastikan keseimbangan dan konsistensi data.



Gambar 6. Sebaran data sebelum (kiri) dan setelah (kanan) resampling imbalanced data

Proporsi sebaran data kelas yang tidak seimbang perlu dilakukan teknik re-sampling. Teknik re-sampling yang dilakukan adalah dengan over-sampling. Digunakan pustaka Imblearn dalam melakukan teknik over-sampling. Pada teknik over-sampling dilakukan penyalinan data pada kelas minoritas dan melengkapinya sebanyak kelas mayoritas. Sehingga proporsi sebaran data pada semua kelas mengikuti jumlah kelas mayoritas. Dataset yang dilakukan over-sampling, jumlah data pada masing-masing kelas sebanyak 7.634.



Gambar 7. Wordcloud ulasan berlabel positif pengguna info BMKG

Gambar 7 menunjukkan frekuensi kata-kata yang paling sering muncul dalam ulasan pengguna berlabel positif terhadap aplikasi info BMKG, khususnya yang berkaitan dengan gempa bumi. Kata-kata seperti "terima kasih," "gempa," "bumi," "aplikasi," "sangat," dan "bantu" mendominasi, yang mengindikasikan bahwa mayoritas pengguna memberikan tanggapan positif serta mengapresiasi kebermanfaatan aplikasi tersebut dalam memberikan informasi gempa bumi secara cepat dan akurat. Kemunculan kata seperti "cuaca," "baik," dan "bagus" juga mencerminkan persepsi positif pengguna terhadap fitur tambahan yang mendukung kesiapsiagaan bencana.



Gambar 8. Wordcloud ulasan berlabel netral pengguna info BMKG

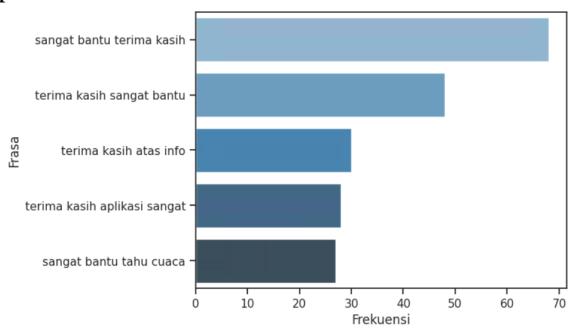
Gambar 8 menunjukkan frekuensi kata-kata yang paling sering muncul dalam ulasan pengguna berlabel netral terhadap aplikasi info BMKG. Kata-kata seperti "gempa," "bumi," "aplikasi," "baru," "info," dan "pemberitahuan" mendominasi, menunjukkan fokus pengguna pada fitur inti aplikasi, yakni pemberian informasi gempa secara real-time. Munculnya kata-kata

seperti "coba," "lihat," "harap," dan "lebih baik" mengindikasikan adanya harapan dan evaluasi dari pengguna terhadap peningkatan kinerja aplikasi, tanpa menunjukkan ekspresi emosional yang kuat.



Gambar 9. *Wordcloud* ulasan berlabel positif pengguna info BMKG

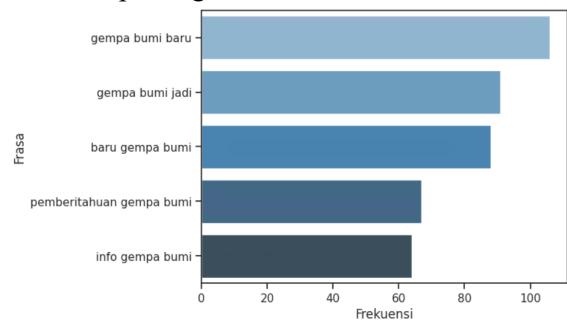
Gambar 9 menunjukkan frekuensi kata-kata yang paling sering muncul dalam ulasan pengguna berlabel negatif terhadap aplikasi info BMKG. Kata-kata seperti "gempa," "bumi," "aplikasi," dan "baru" masih dominan karena menjadi inti pembahasan, namun juga muncul kata-kata bernada kritik seperti "lambat," "gagal," "salah," "tidak," dan "buka aplikasi," yang menunjukkan keluhan terhadap performa teknis aplikasi. Istilah seperti "akurasi," "informasi," dan "lokasi" sering disandingkan dengan harapan atau ketidakpuasan, yang mengindikasikan bahwa ketidakakuratan data dan keterlambatan notifikasi menjadi permasalahan utama.



Gambar 10 Frasa paling sering muncul pada ulasan label positif

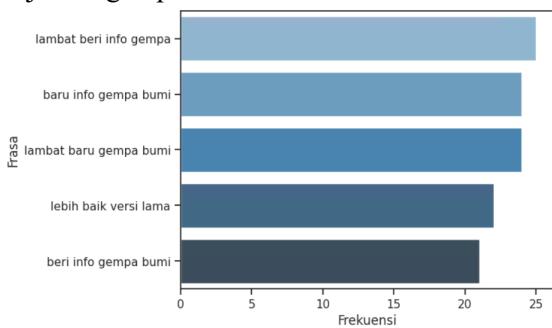
Gambar 10 menunjukkan frasa yang paling sering muncul dalam ulasan pengguna berlabel positif terhadap aplikasi info BMKG. Frasa seperti "*sangat bantu terima kasih*," "*terima kasih sangat bantu*," dan "*terima kasih atas info*" menunjukkan apresiasi pengguna atas informasi yang diberikan

oleh aplikasi, khususnya terkait kebencanaan. Frasa lainnya seperti "*terima kasih aplikasi sangat*" dan "*sangat bantu tahu cuaca*" menegaskan bahwa aplikasi dianggap bermanfaat dan membantu pengguna dalam memperoleh informasi penting.



Gambar 11 Frasa paling sering muncul pada ulasan label netral

Gambar 11 menunjukkan frasa yang paling sering muncul dalam ulasan pengguna berlabel netral terhadap aplikasi info BMKG. Frasa-frasa seperti "*gempa bumi baru*," "*gempa bumi jadi*," "*baru gempa bumi*," "*pemberitahuan gempa bumi*," dan "*info gempa bumi*" bersifat informatif dan deskriptif tanpa mengandung ekspresi emosi yang kuat, baik positif maupun negatif. Frasa-frasa ini umumnya menggambarkan fakta atau pengamatan pengguna terhadap fungsi utama aplikasi, yaitu memberikan informasi mengenai kejadian gempa bumi



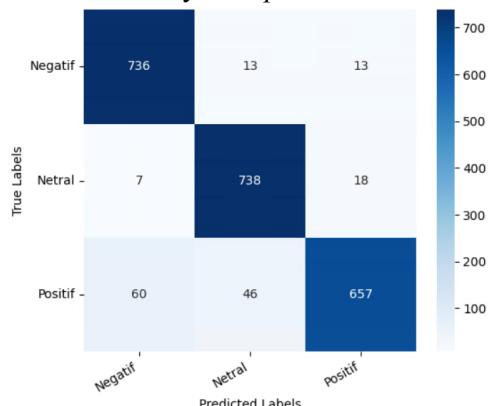
Gambar 12 Frasa paling sering muncul pada ulasan label negatif

Gambar 12 menunjukkan frasa yang paling sering muncul dalam ulasan pengguna berlabel negatif terhadap aplikasi info BMKG. Frasa seperti "*lambat beri info gempa*," "*baru info gempa bumi*," dan "*lambat baru gempa bumi*" menunjukkan keluhan terkait keterlambatan dalam pemberian informasi, yang merupakan

aspek krusial dalam konteks aplikasi kebencanaan. Sementara itu, frasa "*lebih baik versi lama*" mencerminkan ketidakpuasan terhadap versi aplikasi terbaru dibandingkan versi sebelumnya.

C. Evaluasi Model

Penelitian ini dilakukan dua skenario pembagian dataset untuk mengevaluasi pengaruh proporsi data latih, validasi, dan uji terhadap performa model. Skenario 1: 80% data latih, 10% data validasi, dan 10% data uji (80:10:10) dan skenario 2: 70% data latih, 20% data validasi, dan 10% data uji (70:20:10). Dengan keseluruhan model dilatih sebanyak 3 epoch.



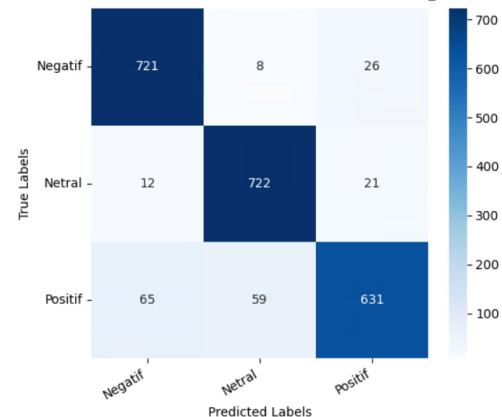
Gambar 12 Konfusi matriks model BERT skenario (80:10:10)

Gambar 12 menunjukkan matriks konfusi dari hasil evaluasi model BERT pada skenario pembagian data 80:10:10 untuk pelatihan, validasi, dan pengujian. Model menunjukkan performa klasifikasi yang tinggi dengan jumlah prediksi benar sebesar 736 untuk label negatif, 738 untuk label netral, dan 657 untuk label positif.

Namun, masih terdapat sejumlah kesalahan klasifikasi, seperti 60 data positif yang diklasifikasikan sebagai negatif, serta 46 data positif yang salah diklasifikasikan sebagai netral. Hal ini menunjukkan bahwa meskipun model mampu mengenali pola dengan cukup baik, masih terdapat ambiguitas antar kelas sentimen, khususnya dalam membedakan antara sentimen positif dengan netral dan negatif.

Gambar 13 menunjukkan matriks konfusi dari hasil pengujian model BERT pada skenario pembagian data 70:20:10 untuk pelatihan, validasi, dan pengujian. Hasil klasifikasi

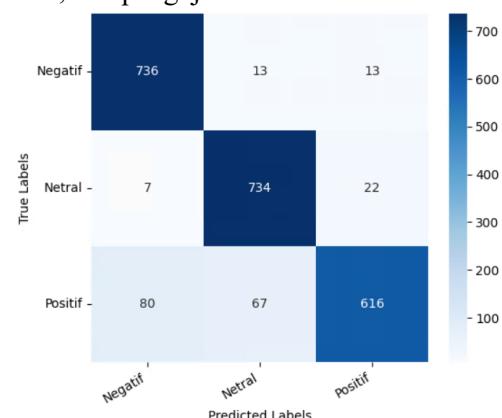
menunjukkan bahwa model berhasil mengidentifikasi sebagian besar data dengan benar, yaitu 721 data untuk kelas negatif, 722 untuk kelas netral, dan 631 untuk kelas positif.



Gambar 13 Konfusi matriks model BERT skenario (70:20:10)

Namun, model masih mengalami kesalahan klasifikasi, terutama pada kelas positif, dengan 65 data salah diklasifikasikan sebagai negatif dan 59 sebagai netral. Selain itu, 26 data negatif juga salah diklasifikasikan sebagai positif. Hal ini menunjukkan bahwa meskipun model cukup akurat secara keseluruhan, tantangan dalam membedakan sentimen positif dari negatif dan netral masih ada.

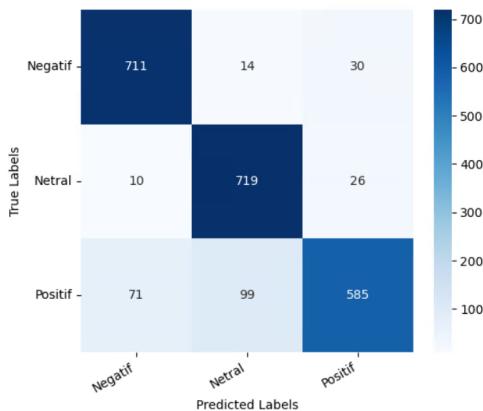
Gambar 14 memperlihatkan matriks konfusi hasil evaluasi model RoBERTa pada skenario pembagian data 80:10:10 untuk pelatihan, validasi, dan pengujian



Gambar 14 Konfusi matriks model RoBERTa skenario (80:10:10)

Model menunjukkan performa yang cukup baik dengan jumlah prediksi benar sebanyak 736 untuk label negatif, 734 untuk label netral, dan 616

untuk label positif. Meski demikian, kesalahan klasifikasi masih terjadi, terutama pada kelas positif, di mana 80 data salah diklasifikasikan sebagai negatif dan 67 sebagai netral. Sementara itu, jumlah kesalahan pada kelas netral dan negatif relatif kecil.



Gambar 15 Konfusi matriks model RoBERTa skenario (70:20:10)

Gambar 15 menunjukkan confusion matrix dari model RoBERTa pada skenario pembagian data 70:20:10 (pelatihan:validasi:pengujian). Matriks ini mengevaluasi kinerja model dalam mengklasifikasikan sentimen ulasan ke dalam tiga kategori: Negatif, Netral, dan Positif. Model berhasil mengklasifikasikan data dengan benar untuk kelas Negatif sebanyak 711 instance, Netral sebanyak 719, dan Positif sebanyak 585. Namun, terlihat pula adanya kesalahan klasifikasi, seperti 71 data Positif yang diprediksi sebagai Negatif dan 99 data Positif sebagai Netral.

Tabel 2 menunjukkan hasil evaluasi model BERT dan RoBERTa dalam dua skenario pembagian data (80:10:10 dan 70:20:10). Secara umum, model BERT menunjukkan performa yang sedikit lebih unggul dibandingkan RoBERTa pada semua metrik evaluasi, yaitu akurasi, presisi, recall, dan F1-score, di kedua scenario

Tabel 2. Hasil evaluasi model

Parameter Evaluasi	BERT		RoBERTa	
	Skenario 80:10:10	Skenario 70:20:10	Skenario 80:10:10	Skenario 70:20:10
Akurasi	93,14%	91,57%	91,17%	88,96%
Presisi	93,25%	91,64%	91,14%	89,15%
Recall	93,14%	91,57%	91,17%	88,96%
F1-Score	93,08%	91,48%	91,03%	88,76%

Terlihat bahwa perubahan proporsi data pelatihan dan pengujian berdampak pada penurunan performa model, terutama pada skenario 70:20:10, di mana RoBERTa mengalami penurunan signifikan (misalnya F1-score turun menjadi 88,76%). Hal ini mengindikasikan bahwa RoBERTa lebih sensitif terhadap pengurangan data pelatihan dibanding BERT, yang relatif lebih stabil. Evaluasi ini memberikan wawasan bahwa meskipun RoBERTa merupakan pengembangan dari BERT, performanya tidak selalu lebih baik, tergantung pada proporsi data dan konteks penggunaannya.

Berdasarkan hasil pengujian, model BERT menunjukkan performa lebih konsisten dibanding RoBERTa, dengan skor akurasi, presisi, recall, dan F1-score yang lebih tinggi di kedua skenario. Performa RoBERTa mengalami penurunan cukup signifikan pada skenario 70:20:10, khususnya pada klasifikasi sentimen positif, yang ditunjukkan oleh tingginya jumlah kesalahan klasifikasi pada confusion matrix. Hal ini mengindikasikan bahwa RoBERTa lebih sensitif terhadap perubahan proporsi data pelatihan dibandingkan BERT. Secara keseluruhan, model BERT lebih andal dalam konteks klasifikasi sentimen ulasan pengguna aplikasi Info BMKG, baik dalam hal akurasi maupun stabilitas performa terhadap variasi proporsi data pelatihan. Temuan ini dapat menjadi dasar dalam pengembangan sistem analisis opini publik berbasis NLP di sektor layanan public

V. KESIMPULAN

Penelitian ini dilakukan untuk menganalisis sentimen ulasan pengguna terhadap aplikasi Info BMKG yang diambil dari Google Play Store. Aplikasi ini berperan penting dalam menyampaikan informasi cuaca, iklim, gempa bumi, dan peringatan dini bencana secara terintegrasi kepada masyarakat. Dengan meningkatnya pengguna digital di Indonesia, penting untuk mengevaluasi kepuasan pengguna melalui analisis terhadap ulasan yang tersedia.

Penggunaan pendekatan *Natural Language Processing* (NLP) dan model transformer seperti BERT dan RoBERTa terbukti efektif dalam memahami konteks kalimat dan

mengklasifikasikan sentimen pengguna ke dalam kategori positif, negatif, maupun netral. Model ini memungkinkan analisis dilakukan secara otomatis dan dalam skala besar dengan hasil yang lebih akurat dibandingkan pendekatan tradisional.

Hasil analisis menunjukkan bahwa meskipun mayoritas pengguna memberikan ulasan positif terhadap aplikasi Info BMKG, masih terdapat sejumlah masukan berupa keluhan yang perlu diperhatikan untuk peningkatan kualitas aplikasi, terutama terkait keakuratan data dan pengalaman pengguna.

DAFTAR PUSTAKA

- [1] Badan Nasional Penanggulangan Bencana, “Info Bencana BNPB April 2025,” *Data dan Informasi Kebencanaan Bulanan Teraktual*, vol. 6, no. 4, 2025.
- [2] Republik Indonesia, “Undang-Undang Republik Indonesia Nomor 31 Tahun 2009 Tentang Meteorologi, Klimatologi, dan Geofisika,” 2009.
- [3] Badan Meteorologi Klimatologi dan Geofisika, “Rencana Strategis BMKG Tahun 2020-2024,” 2020.
- [4] APJII, “Survey Penetrasi dan Perilaku Internet,” 2024.
- [5] P. A. Fajriyah and A. P. Hadi, “Pengaruh Penggunaan Aplikasi InfoBMKG Terhadap Sikap Tanggap Bencana Mahasiswa Universitas Mataram Pasca Gempa Bumi Lombok,” *JCOMMSCI-JOURNAL OF MEDIA AND COMMUNICATION SCIENCE*, vol. 2, pp. 1–16, 2019.
- [6] M. Khalid, M. Asif, and U. Shehzaib, “Towards Improving the Quality of Mobile App Reviews,” *International Journal of Information Technology and Computer Science*, vol. 7, no. 10, pp. 35–41, Sep. 2015, doi: 10.5815/ijitcs.2015.10.05.
- [7] S. Fransiska and A. Irham Gufroni, “Sentiment Analysis Provider by.U on Google Play Store Reviews with TF-IDF and Support Vector Machine (SVM) Method,” *Scientific Journal of Informatics*, vol. 7, no. 2, pp. 2407–7658, 2020.
- [8] K. Dwi Pratama, D. Wisnu Brata, and W. Purnomo, “Analisis Sentimen Ulasan Pengguna Aplikasi Info BMKG pada Google Play Store di Indonesia,” 2023.
- [9] M. J. Zaki Wagner Meira Jr, “Data Mining and Analysis: Fundamental Concepts and Algorithms,” 2014.
- [10] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding,” Oct. 2018.
- [11] N. A. R. Putri and Ardiansyah, “Analisis Sentimen Terhadap Kemajuan Kecerdasan Buatan di Indonesia Menggunakan BERT dan RoBERTa,” *Jurnal Sains dan Informatika*, vol. 9, no. 2, pp. 136–145, Nov. 2023, doi: 10.34128/jsi.v9i2.649.
- [12] N. Putu, V. D. Saraswati, N. Yudistira, and P. P. Adikara, “Analisis Sentimen terhadap Perundungan Siber pada Twitter menggunakan Algoritma Bidirectional Encoder Representations from Transformer (BERT),” 2023.
- [13] R. Maulana, A. Voutama, and T. Ridwan, “Analisis Sentimen Ulasan Aplikasi Mypertamina Pada Google Play Store Menggunakan Algoritma Nbc,” *Jurnal Teknologi Terpadu*, vol. 9, no. 1, pp. 42–48, 2023.